

GROWTH MODEL EVALUATION: SOUTH-WIDE LOBLOLLY PINE
SEED SOURCE STUDY

Fan H. Kung

ABSTRACT.--A growth model useful to the forest geneticists should have the following five P's: (1) perfection of fit to the data presented, (2) predictability of growth potentials, (3) possibility of inference based on parameter estimates, (4) power of discrimination among seed sources, and (5) persistency of regression coefficients over time.

To illustrate, the function of $\ln(\text{HT}) = B_0 + B_1/\text{AGE} + B_2 \ln(1 + 1/\text{AGE})$ was fitted to the height growth of 15 seedlots of loblolly pine in a south-wide study. The degree of determination was 0.999 at least. The five-year and the ten-year projections were low by 3% and 5% respectively. The regression coefficients B_0 , B_1 and B_2 were highly significant among seedlots. When the three growth periods were compared, the coefficient of variation for the regression coefficients was less than 5%.

Additional Key Words: Growth projection, Model verification, Growth curve discrimination

INTRODUCTION

The mathematical characterization of growth is among the oldest scientific pursuits. Indispensable long-term planning in forestry requires reliable information about the growth of forest stands. A wide array of growth and yield models ranging from whole stand models to individual tree models has been developed for southern species.

Genetic field tests are subjected to many uncontrolled disturbances. However, height of dominant-codominant trees is much less dependent on density and therefore is a better measure of inherent growth differences. Growth and yield models can be used to translate differences in dominant-codominant height into stand differences expected in the absence of uncontrolled disturbances (Nance and Wells 1981).

Foresters have long used the mean height of dominant-codominant trees for site index, as a universal measure of the potential

^{1/} Professor, Department of Forestry, Southern Illinois University, Carbondale, IL 62901. The author is grateful to Dr. Osborn O. Wells and the Southern Forest Experiment Station for providing the data for this study, and to Dr. Marilyn A. Buford for fruitful discussions.

productivity of forest land. Thus, a further improvement in precision for measuring genetic differences in growth rate would be removing the site effect and provenance by site interaction by using the regional mean from many plantations.

Suppose that we have determined to use the range-wide mean height to assess genetic differences, the question still remains with us as to which growth model should we use as a base for comparison.

Before growth models can be evaluated, one must outline the criteria for model selection. We believe a useful growth model should have the following characteristics:

1. perfection of fit to the data presented,
2. predictability on growth potentials,
3. possibility of inference based on parameter estimates,
4. power of discrimination among seed sources, and
5. persistency of regression coefficients over times.

In this paper, we use data from the loblolly pine south-wide study to illustrate such desirable properties of a growth function.

SOUTH-WIDE LOBLOLLY PINE SEED SOURCE STUDY

Complete details of the Loblolly pine experiment are given by Wells and Wakeley (1966). Fifteen seed sources are represented and 16 plantings survived after 25 years in the field.

Measurements of total height were made at various ages from 1 to 27 years. However instead of using only the dominant and codominant trees, we used all available trees measured in 3, 5, 10, 15, 20, and 25 years to calculate the mean height growth for each seed source. The reason for this selection is based on statistical and not on silvicultural ground. The statistical property of dominant and codominant trees is close to the extreme number distribution while the average height is close to the normal distribution. The latter is much easier for data analysis.

The south-wide regional means of the 15 seed sources are listed in Table 1. The mean standard deviation and the coefficient of variation are also presented.

GROWTH FUNCTION

Nonlinear growth functions have been proposed for total height. The monomolecular function is useful for site index curves for age 20 and older which show no point of inflection (Lundgren and Dolid 1970). Richards' curve gives a good fit for both height and volume growth of Douglas-fir provenances (Namkoong, Usanis, and Silen 1972). The Weibull function can describe the growth of trees and stands (Yang, Kozak and Smith 1978). However, we are in favor of linear models over nonlinear models because linear models give rise to unbiased, normally distributed, minimum variance estimators, whereas nonlinear

Table 1.--South-wide regional means of height growth among loblolly pine seed sources.

Seed Source	Regional Mean at Age					
	3	5	10	15	20	25
	- - - - - 0.01 ft. - - - - -					
301	451	1089	2872	4174	5270	6092
303	455	1092	2907	4283	5388	6339
305	479	1137	3021	4440	5582	6410
307	391	948	2566	3907	5078	5907
309	481	1120	2866	4166	5331	6305
311	388	940	2599	3955	5146	5964
315	432	1043	2716	4085	5133	6007
319	457	1057	2790	4113	5243	6069
321	377	903	2545	3898	4976	6028
323	433	1079	2897	4244	5317	6162
325	438	1048	2793	4020	5006	5849
327	412	992	2654	3782	4843	5681
329	393	934	2608	3779	4831	5643
331	346	905	2585	3915	4912	5949
Mean	421	1014	2729	4039	5132	6018
St. Dev.	40	82	159	197	220	222
C.V. %	9.5	8.1	5.8	4.9	4.3	3.7

regression models tend generally to do so only as the sample size becomes very large (Ratkowsky 1983). With only a total of six data points we are not very comfortable with the results of a nonlinear model. Therefore, to illustrate that the five desirable properties are obtainable, we use the following curve with an intrinsically linear combination of parameters:

$$\ln(HH) = B0 + B1(1/AGE) + B2 \ln(1 + 1/AGE),$$

where \ln is the natural logarithm transformation and $B0$, $B1$, and $B2$ are coefficients to be determined by regression analysis.

The function provides precise description of observed data points and provides trustworthy predictions. The function is differentiable and is applicable to many growth and yield characters in life sciences (Kung 1984).

PERFECTION OF FIT

Growth is a continual process but it may be subdivided into stages. Tree physiologists indicated that growth consists of division elongation, differentiation and maturation of cells (Kozlowski 1971). Forest geneticists divided stand development into juvenile-genotypic, mature-genotypic and codominance-suppression phase (Franklin 1979). Forest biometricians believe that a growth curve begins at the value of zero, climbs slowly at first and then more steeply. After a turning point, the increment diminishes and then asymptotically moves towards some final value

(Prodan 1968). Thus a growth function should act like an adjustable ship curve used in drafting that fits all data points through various stages equally well and not just for a single stage. For example, a polynomial, as well as the simple exponential function may fit a part of a growth series better than the more complex nonlinear models, but may have a poor fit elsewhere.

In the paper, three periods were used for comparison: (1) age 3 to 15, (2) 3 to 20, and (3) 3 to 25. The F value, the root mean square of error and the coefficient of determination are used to judge the fit of the model.

The F values ranged from a minimum of 3040 to a maximum of 999999 (Table 2). All models were significant at the 0.0001 probability of error.

Table 2. Fit between data and model among seed sources in three time periods at age 3 to 15, 3 to 20 and 3 to 25 years.

Seed Source	Model Statistics					
	F value			Root mean square		
	3 - 15	3 - 20	3 - 25	3 - 15	3 - 20	3 - 25
301	18,909	13,895	12,333	.009	.012	.015
303	58,791	32,671	9,387	.005	.008	.016
305	32,812	32,243	38,561	.007	.008	.008
307	29,514	4,855	6,166	.007	.022	.022
309	123,317	5,170	3,040	.003	.019	.029
311	999,999	7,173	10,099	.001	.018	.017
315	10,901	11,113	7,760	.011	.014	.012
317	70,568	7,852	4,596	.005	.017	.025
319	45,775	11,704	12,122	.006	.013	.014
321	483,648	72,226	5,834	.002	.006	.022
323	29,266	35,172	20,694	.007	.008	.011
325	4,926	13,345	10,270	.017	.012	.016
327	2,844	4,473	4,068	.023	.022	.029
329	2,104	5,331	7,232	.027	.020	.020
331	115,717	318,691	6,114	.004	.003	.022
Mean				.009	.013	.019
St. Dev.				.008	.006	.006
C.V. %				85	45	32

The root mean square of fitting errors ranged from a minimum of .001 to a maximum of .029. On the average the fitting error was .009 for the 3 to 15 year period and increased to .019 for the 3 to 25 year period. The average for the three periods was .014. For small values of e , we have approximation of $\text{Exp}(e) = 1 + e$. Therefore, the small size of the error term represents the relative error. In other words, the percent of error for the 3 - 15 year period would be only .9 percent, corresponding to the average RMSE of .009. The average of the relative error for the study is 1.4 percent.

The coefficient of determination ranged from 0.9997 to 1.000 among 45 regression models. It may lead one to wonder whether a perfect fit has been achieved.

PREDICTABILITY OF GROWTH POTENTIALS

One of the most rigorous tests of a fitted equation is cross verification with a second sample taken at another time (Daniel and Wood 1980). Because this is impossible, we have used a longitudinal verification for the growth curve. First, we developed regressions based on 3 to 15 and 3 to 20 year periods for each seed source. The second step was to project the height at age 25 from each regression. The final step was to compare the projected and the observed height. The results are presented in Table 3.

Table 3.--Comparison between projected and observed height at 25 years of age.

Seed Source	Observation at age 25	Projected from ages		Error rate	
		3-15	3-20	3-15	3-20
		----- 0.01 ft. -----		----- % -----	
301	6,092	5,858	5,944	-3.9	-2.4
303	6,339	6,051	6,114	-4.5	-3.5
305	6,410	6,278	6,337	-2.1	-1.1
307	5,907	5,571	5,741	-5.7	-2.8
309	6,305	5,800	5,963	-8.1	-5.4
311	5,964	5,699	5,847	-4.4	-2.0
315	6,007	5,715	5,805	-4.8	-3.4
317	5,865	5,456	5,588	-7.0	-4.7
319	6,069	5,825	5,932	-4.0	-2.2
321	6,028	5,681	5,726	-5.8	-5.0
323	6,162	5,977	6,027	-3.0	-2.2
325	5,849	5,671	5,686	-3.0	-2.8
327	5,681	5,333	5,444	-6.1	-4.2
329	5,643	5,461	5,510	-3.2	-2.4
331	5,949	5,646	5,641	-5.1	-5.1
Mean	6,018	5,735	5,821	-4.7	-3.3
St. Dev.	222	228	239	1.6	1.3

All projections were lower than the observed heights. The 10 year projection was low by 2.8 feet or 4.7 percent; while the 5 year projection was low by 2.0 feet or 3.3 percent. However the standard deviation of the three groups were the same. The projections were as precise as the observations even though the accuracy may be off a little. The standard deviation for the relative error rate were 1.6 and 1.3 percent respectively for the 10-year and 5-year projections.

The bias in projection is not the fault of using transformation. The transform of the expected value does not equal the

expected value of the transform, although they are crude approximations of each other (Kruskal 1978). The percent of bias calculated according to the formula given by Wiant and Harner (1979) was less than 0.04 percent. Because of the small error variance in the regressions, the adjustment of the prediction (Baskerville, 1972) offered little reduction in the bias. This large and consistent bias needs to be rectified by adding 3 percent to the 5-year forecasting and 5 percent to the 10-year forecasting.

POSSIBILITY OF INFERENCE

Growth potential and growth rate are of interest to the forest manager. The function $\ln(\text{HT}) = B_0 + B_1(1/\text{AGE}) + B_2 \ln(1 + 1/\text{AGE})$ indicates that the asymptotic height should be near the value of $\text{Exp}(B_0)$. For example, the maximum value (9.334) for B_0 among 15 seed sources was found in provenance 305, therefore the asymptotic height would be $\text{Exp}(9.334) \times .01 \text{ ft.} = 113 \text{ ft.}$ On the other hand, a minimum of B_0 in seed source 327 indicated that the average asymptotic height could be 92 ft. Notice that the average height at age 25 for the complete study is 60 feet, the average site index for the complete experiment could be estimated as 90 ft. at age 50. Which is the average site for loblolly pine. Trees at that site grow slowly after 50 years of age, hence our estimate of asymptotic height seems to be reasonable.

The function can be differentiated. The instantaneous growth rate of $\log(\text{Ht})$ at age x is:

$$\begin{aligned} d(\ln\text{HT})/dx &= -B_1/x^2 - B_2/(x^2 + x) \\ &= (-x(B_1 + B_2) - B_1)/(x^3 + x^2) \end{aligned}$$

Using the coefficients developed from the 3 to 25 year as an example we found in Table 4 that seed source 321 would have the greatest growth rate of $\log(\text{Ht})$ at age x as $((17.15x - 33.90)/(x^3 + x^2))$, while seed source 325 would have the smallest growth rate as $((14.81x - 24.58)/(x^3 + x^2))$ at the age of x years. Because of the extremely high correlation between B_1 and B_2 ($r = 0.997$), selection of growth rate can be simplified as selection for B_1 .

POWER OF DISCRIMINATION

If simple functional differences among genotypes were existent or nonexistent, a growth function should have the power of discrimination to prove or to disprove the differences in growth form. By fitting the mean height from age 5 to 55 years old of 13 populations of Douglas-fir in Wind River, Oregon, Namkoong, Usanis and Silen (1972) found that the parameters A , C , and m of the Richards' function were nonsignificant among populations. Using the Weibull function to quantify sweetgum germination, it was found that the coefficients b and c were significant among families within the stand, but all of the three coefficients (a , b , c) were not significant among stands (Rink et al. 1979). The significant difference in a given parameter indicated that selection in that characteristics of the growth curve may be possible. On the other hand, no growth rate can be selected if all growth curves are the same.

Table 4.--Summary of regression coefficients in growth models for 15 loblolly pine seed sources during three time period.

Seedlot	Coefficient											
	B0				B1				B2			
	3-15	3-20	3-25	C.V.	3-15	3-20	3-25	C.V.	3-15	3-20	3-25	C.V.
301	9.222	9.248	9.267	.24	22.58	23.89	25.02	5.17	-36.98	-38.58	-39.96	3.87
303	9.272	9.289	9.318	.25	24.66	25.56	27.22	5.03	-39.53	-40.62	-42.65	3.87
305	9.309	9.324	9.384	.13	25.04	25.85	26.38	2.62	-39.93	-40.91	-41.56	2.01
307	9.217	9.267	9.290	.40	27.36	29.96	31.28	6.75	-42.98	-46.16	-47.77	5.34
309	9.206	9.253	9.298	.50	23.41	25.82	28.40	9.64	-37.65	-40.51	-43.74	7.50
311	9.252	9.296	9.311	.33	28.58	30.80	31.72	5.32	-44.55	-47.27	-48.38	4.22
315	9.212	9.239	9.266	.29	24.46	25.82	27.40	5.68	-39.26	-40.93	-42.85	4.38
317	9.196	9.236	9.275	.43	27.13	29.20	31.44	7.37	-42.75	-45.28	-47.99	5.78
319	9.240	9.270	9.289	.27	26.65	28.22	29.28	4.72	-41.71	-43.63	-44.92	3.72
321	9.268	9.281	9.322	.30	30.84	31.53	33.90	5.00	-47.33	-48.17	-51.05	3.99
323	9.249	9.264	9.281	.17	22.09	22.83	23.85	3.86	-36.65	-37.55	-38.79	2.85
325	9.193	9.197	9.220	.16	23.04	23.26	24.58	3.53	-37.51	-37.79	-39.39	2.65
327	9.125	9.160	9.194	.38	22.15	23.83	25.80	7.64	-36.35	-38.52	-40.92	5.92
329	9.189	9.204	9.223	.19	26.65	27.42	28.53	3.43	-42.06	-43.01	-44.35	2.67
331	9.235	9.236	9.277	.26	24.70	24.72	27.13	5.48	-40.41	-40.43	-43.35	4.10
Mean	9.226	9.250	9.278	.29	25.29	26.58	28.13	5.41	40.37	41.95	43.84	4.19
St. Dev.	.043	.042	.040	.10	2.54	2.80	2.93	1.83	3.18	3.46	3.62	1.45
C.V. %	.47	.45	.43		10.06	10.55	10.41		7.89	8.25	8.26	

The functions presented in this paper have great power of discrimination among seed sources of loblolly pine. All three coefficients B0, B1 and B2 are significant beyond the 0.001 probability of error (Table 5). The repeatability or provenance heritability calculated from the F value (Kung and Bey 1978) was 0.98 for each coefficient.

Table 5.--Analysis of variance of the regression coefficients.

Source	d.f.	Coefficient					
		B0		B1		B2	
		MS	F	MS	F	MS	F
Seed source	14	.0050	40 ^a	22.20	63 ^a	34.19	65 ^a
Age period	2	.0101	82 ^a	30.31	86 ^a	45.22	86 ^a
Error	28	.0001		.35		.52	
Total	44						

^aSignificant at the 0.0001 level.

PERSISTENCY OF COEFFICIENT

A good growth curve should be relatively independent of the range of data base. If the growth curve developed from the juvenile growth period were the same as that developed from the complete life span, we would be more successful in making early selection.

The persistency of coefficient in the three periods (3 to 15,

3 to **20**, and 3 to 25 years of age) is evident in Table 4. The average coefficient of variation among three periods was only .29 % for **BO**, 5.4 % for **B1** and 4.2 % for **B2**.

As the range of ages becomes wider, the absolute values of the **BO**, **B1** and **B2** also increases. The differences are significant beyond the 0.001 probability of error (Table 5). Although one would like to have a constant regression coefficient throughout the years if possible, the second best would be that for any given parameter, it may differ from one period to another, but it should vary in a predictable manner.

The persistency of coefficient can be shown also by the correlation coefficient for **BO**, **B1** and **B2** among three periods (Table 6). All correlations are significant at the 0.001 level. From the period, of 3 to 15 years to the period of 3 to 20 years, the **B1** increases by 6 % and the **B2** increases by 4 %; while from the period of 3 to **20** years to the period of 3 to 25 years both **B1** and **B2** increases only 2 %.

Table 6.--Correlations among three periods for parameters **BO**, **B1** and **B2**.

Correlation between periods		Correlation coefficients for parameters		
		BO	B1	B2
3 - 20	3 - 25	.97	.98	.98
3 - 15	3 - 20	.94	.96	.96
3 - 15	3 - 25	.89	.94	.94

DISCUSSION

Many growth and yield models are available to tree improvement workers. From a practical point of view we recommend the 5-P criteria for model selection. However, we have not assigned any weight to each criterion which may differ from one program to another.

We use the range means of the seed source and not the individual tree in hope that if all the environmental errors could be averaged out we would have a more accurate evaluation of the performance of the growth model as well as the genetic difference among seed sources.

The function used for illustration is almost ideal in the 5-P criteria. However, we are still searching for a perfect curve which is totally independent of the age range. Would it be possible to have a model which regression coefficients are changeless between any period range from 1 to 100 years? or is it an impossible dream?

LITERATURE CITED

- Baskerville, G. L. 1972. Use of logarithmic regression in the estimation of plant biomass. *Can. J. For. Res.* 2:49-53.
- Daniel C., and F. S. Wood. 1980, *Fitting Equations to Data*. 2nd ed., 458 p. Wiley, New York.
- Franklin, E. C. 1979. Model relating levels of genetic variance to stand development of four North American conifers. *Silvae Genetica* 28:207-212.
- Kozlowski, T. T. 1971. *Growth and Development of Trees, Vol. I, Seed Germination, Ontogeny and Shoot Growth*. 443 p. Academic Press, New York and London.
- Kruskal, J. B. 1978. Transformations of data. In *International Encyclopedia of Statistics, Vol. II*. 1,350 p. The Free Press, New York.
- Kung, F. H., and C. F. Bey. 1978. Heritability construction for provenance and family selection. *Lake States For. Tree Impro. Conf.* 13:136-146.
- Kung, F. H. 1984. A law-like empirical growth function. *Abstract Am. Stat. Ass. Meeting* 1984:234.
- Lundgren, A. L., and W. A. Dolid. 1970. Biological Growth Functions Describe Published Site Index Curves for Lake State Timber Species. *U.S.D.A. For. Serv. Res. Paper NC-36*. 9 p.
- Namkoong, G., R. A. Usanis, and R. R. Silen. 1972. Age-related variation in genetic control of height growth in Douglas-fir. *Theo. Appl. Genet.* 42:151-159.
- Nance, W. L., and O. O. Wells. 1981. Estimating volume potential in genetic tests using growth and yield models. *South For. Tree Impro. Conf.* 16:39-46.
- Prodan, M. 1968. *Forest Biometrics*. 447 p. Pergamon Press, New York.
- Ratkowsky, D. A. 1983. *Nonlinear Regression Modeling*. 276 p. Marcel Dekker, Inc., New York and Basel.
- Ring, G., T. R. Dell, G. Switzer, and F. T. Bonner. 1979. Use of the Weibull function to quantify Sweetgum germination data. *Silvae Genetica* 28:9-12.
- Wells, O. O., and P. C. Wakeley. 1966. Geographic variation in survival, growth and fusiform-rust infection of planted loblolly pine. *For. Sci. Monogr.* 11:40.

Wiant, H. V., and E. J. Harner. 1979. Percent bias and standard error in logarithmic regression. *For. Sci.* 25:167-168.

Yang, R. C., A. Kozak, and J. H. G. Smith. 1978. The potential of Weibull-type functions as flexible growth curves. *Can. J. For Res* 8:424-431.