# MODELING LOBLOLLY PINE AGE-AGE CORRELATION FOR HEIGHT USING THE DEGREE OF NON-DETERMINATION.

F. H. Kung[1]

Abstract.--Growth is a collective process. If we can assume that the error component of predicting growth from year to year is homogenous and additive, then the degree of non-determination (1-r2) between height in two years is a linear function of the age difference. We used data from the South-wide Loblolly Pine Provenance Test to calculate the age-age correlations for height growth from age 3, 5, 10, 15, 20 and 25 years. The degree of non-determination (DON) then was computed from each correlation. The DON model and its derived non-linear model were then compared with two other models: Lambeth model and response surface model. The DON model had the greatest F value (416.56) and the largest r-square (0.956). After transformation from DON to correlation, the predictive correlation from the DON model proved to be closer to the observed correlation than other models. The DON model suggests that about 2.5% of accountability in height growth among loblolly pine provenances is lost in each subsequent year.

Keywords: Pinus taeda L., juvenile-mature correlation, selection.

## INTRODUCTION

Tree improvement programs involve long-term investment. We would like to make selections as early as possible to reduce the breeding cycle and to maximize efficiency of land use. For example, knowing that oleoresin yields of various slash pine progeny correspond closely with the yields of their parents, enables us to use short-term progeny test of 3-year rotation and 3 feet spacing for parental selection (Squillace and Gansel 1968). The efficacy of early selection is related to the correlation between early and late assessment of the trait being improved (Kung 1975). Early selection is usually less effective, but may be more efficient than late selection in terms of genetic gain per unit of area and per unit of time (Kung 1973, Bohren 1975).

Given N repeated measurements at various ages, N*N correlations are possible in the age-age correlation matrix. For example, the original South-wide Loblolly Pine Seed Source Study (Wells and Wakeley, 1966, Nance and Wells, 1981) was measured at age 3, 5, 10, 15, 20, and 25 years (Table 1), and therefore, there are 36 correlations in the matrix (Table 2). Because the matrix is symmetrical and the values on the diagonal are unity, usually only a triangular matrix is reported in literature. The age-age correlation becomes smaller as the difference between two ages becomes greater. Taking a logarithm transformation of age, Lambeth (1980) found a linear relationship between the difference of two transformed ages and its age-age correlation. The Lambeth model is expressed as follows:

$$r = a + b\,(LAR)$$

Where
- $r$ = correlation coefficient
- $a,b$ = regression coefficient
- $LAR$ = logarithm of age ratio = log(young age/old age) = log(young age)-log(old age)

---

[1] Professor, Department of Forestry, Southern Illinois University, Carbondale, IL. 62901-4411

Table 1. Mean height of loblolly pine provenances at various ages.

| No. | Provenance | 3 | 5 | 10 | 15 | 20 | 25 |
|---|---|---|---|---|---|---|---|
| | | Mean height at age (years) | | | | | |
| | | -cm- | | | | | |
| C-301 | E. Maryland | 137 | 331 | 875 | 1272 | 1606 | 1856 |
| C-303 | SE. N. Carolina | 138 | 332 | 886 | 1305 | 1642 | 1932 |
| C-305 | E. N. Carolina | 145 | 346 | 920 | 1353 | 1701 | 1953 |
| C-307 | W. S. Carolina | 119 | 288 | 782 | 1190 | 1547 | 1800 |
| C-309 | SE. Georgia | 146 | 341 | 873 | 1269 | 1624 | 1921 |
| C-311 | NE. Georgia | 118 | 286 | 792 | 1205 | 1568 | 1817 |
| C-315 | N. Alabama | 131 | 317 | 827 | 1245 | 1564 | 1830 |
| C-317 | NE. Alabama | 116 | 282 | 767 | 1164 | 1502 | 1787 |
| C-319 | N. Alabama | 139 | 322 | 850 | 1253 | 1598 | 1849 |
| C-321 | NE. Mississippi | 114 | 275 | 775 | 1188 | 1516 | 1837 |
| C-323 | SE. Mississippi | 131 | 328 | 883 | 1293 | 1620 | 1878 |
| C-325 | E. Texas | 133 | 319 | 851 | 1225 | 1525 | 1782 |
| C-327 | SW. Arkansas | 125 | 302 | 808 | 1152 | 1476 | 1731 |
| C-329 | W. Tennessee | 119 | 284 | 794 | 1151 | 1472 | 1719 |
| C-331 | NW. Georgia | 105 | 275 | 787 | 1193 | 1497 | 1813 |

**Table 2.--Correlation matrix and degree of non-determination matrix for tree heights at age 3, 5, 10, 15, 20 and 25 years.**

***Correlation Matrix***

| Age | 3 | 5 | 10 | 15 | 20 | 25 |
|---|---|---|---|---|---|---|
| | Age in Years | | | | | |
| 3 | 1.00000 | 0.96739 | 0.90097 | 0.79722 | 0.79064 | 0.65974 |
| 5 | 0.96739 | 1.00000 | 0.96333 | 0.87443 | 0.83263 | 0.71456 |
| 10 | 0.90097 | 0.96333 | 1.00000 | 0.92528 | 0.85869 | 0.75366 |
| 15 | 0.79722 | 0.87443 | 0.92528 | 1.00000 | 0.96427 | 0.91275 |
| 20 | 0.79064 | 0.83263 | 0.85869 | 0.96427 | 1.00000 | 0.93072 |
| 25 | 0.65974 | 0.71456 | 0.75366 | 0.91275 | 0.93072 | 1.00000 |

***Degree of Non-determination Matrix***

| Age | 3 | 5 | 10 | 15 | 20 | 25 |
|---|---|---|---|---|---|---|
| | Age in Years | | | | | |
| 3 | 0.00 | 0.06 | 0.19 | 0.36 | 0.37 | 0.56 |
| 5 | 0.06 | 0.00 | 0.07 | 0.24 | 0.31 | 0.49 |
| 10 | 0.19 | 0.07 | 0.00 | 0.14 | 0.26 | 0.43 |
| 15 | 0.36 | 0.24 | 0.14 | 0.00 | 0.07 | 0.17 |
| 20 | 0.37 | 0.31 | 0.26 | 0.07 | 0.00 | 0.13 |
| 25 | 0.56 | 0.49 | 0.43 | 0.17 | 0.13 | 0.00 |

When we examine a square correlation matrix (Table 2) , we can also see that the correlation becomes smaller as the distance from a cell to the diagonal becomes greater. Therefore, the correlation matrix may be viewed as a symmetrical response-surface model with a ridge on the diagonal and with slopes incline toward two corners. A general response-surface model in analytical geometry is usually expressed as

$$Z = aX^2 + bXY + cY^2 + dX + eY + f$$

Because of symmetry, it is necessary to have coefficients a=c and d=e. Furthermore, contrary to mathematicians, statisticians like to order variables by ascending power. Thus, a response-surface regression model for age-age correlation may be expressed as follows:

$$r = a + b\,(X+Y) + c\,(X^2 + Y^2) + d\,(X*Y)$$

Where
r = correlation between age X and age Y
a = intercept
b = coefficient for linear terms
c = coefficient for quadratic terms
d = coefficient for crossproduct term.

In a previous study, we have successfully fitted a symmetrical response-surface model to stem volume data obtained from stem-analysis of 51 cryptomeria trees between age 3 to 30 years (Yang and Kung 1987). Since we have found that coefficient b was not significant, and that coefficient d was twice the size of coefficient c but had a negative sign, the above 4-coefficient model can be shorten to a 2-coefficient model with little reduction in the coefficient of determination:

$$r = a + b\,(X - Y)^2$$

If we define difference in ages (DA) as X - Y, then

$$r = a + b\,(DA)^2 \qquad (2)$$

We will call this model by the name of DA2 model in this paper, representing linear relationship with the second power of age differences (DA).

Both the Lambeth model (1) and DA2 model (2) show the relationship between the correlation and the age difference but not why. A systematic model therefore is proposed to clarify this relationship.

## THE DON MODEL

Growth is a collective process. The total height of a tree is completely determined by the previous height and the last increment, but the system controlling increment may have two types of elements: order and chaos. The orderly elements are the previous height and all earlier increments. The chaos element is a random and independent distribution of residual errors within a given specific growing interval.

The square of a correlation coefficient is called the R-square. The R-square in a regression model is called the degree of determination, and the quantity of 1-(R-square) is call the degree of non-determination. The former indicates how much variance in terms of sum of squares (SSQ) can be explained by the regression model, and the latter, cannot be explained. The degree of non-determination (DON) therefore is the ratio of error SSQ divided by the total SSQ.

Because we cannot change history, we may assume that previous deviation from orderly growth is interminable. For example, if we compare a tree with accidental damage to the terminal bud with other non-damaged trees, the damaged tree will deviate from the orderly pattern of height growth. The damage may be repaired but can never be denied. The deviation is recorded and cannot be erased.

Total growth is the accumulation of all previous annual growth and total error is the sum of all previous errors. Can we then assume that the degree of non-determination is also additive? In Table 2 we find in several cases that the additive property seems to be valid. For example, the DON from age 3 to age 5 is 0.06, from age 5 to age 25 is 0.49, the sum of these two is 0.55 which is close to the value 0.56 given by the DON between age 3 and age 25. In the second example, the fitting errors from age 3 to age 20 (DON=0.37) is the sum of those form age 3 to age 5 (DON=0.06), and from age 5 to age 20 (DON=0.31). In the third example, we have 0.07 (DON for age 5 to 10) + 0.43 (DON for age 10 to 25) = 0.50, the sum is again very close to the value of 0.49 (DON for age 5 to 25).

Further examination of the DON matrix in Table 2 suggests that DON has a linear relationship with age differences (DA). The average DON for a 5-year time lag is .13, for a 10-year lag is .22, and for a 15-year lag is .37. Therefore, the DON model can be expressed as:

$$DON = a + b\ (DA)\ . \tag{3}$$

Because the DON model minimizes the error variance of DON, not the original correlation, would the following nonlinear model, transformed from the DON model fit the original correlation better?

$$r = \text{sqrt}\ [a + b\ (DA)\ ] \tag{4}$$

We will call this nonlinear model as NLIN in this paper.

## STATISTICAL COMPARISON OF THE FOUR MODELS

The above four models were formulated by PROC REG and PROC NLIN (SAS Institute, Inc. 1987). The regression coefficients and their standard errors (listed below the coefficients) of the four models are as follows:

| | | | |
|---|---|---|---|
| DON Model | $(1-r^2)$ = | $-0.004164$ + | $0.02473$ * DA |
| | | [0.01246] | [0.001211] |
| or, | r = | (1.004164 | $-$ 0.02473 * DA) $^{0.5}$ |
| DA2 Model | r = | 0.971385 | $-$ 0.000699 * $DA^2$ |
| | | [0.009257] | [0.00005325] |
| LAMBETH Model | r = | 0.996399 | $-$ 0.13703 * LAR |
| | | [0.01440] | [0.01459] |
| NLIN Model | r = | (1.005177 | $-$ 0.024924 * DA) $^{0.5}$ |
| | | [0.014143] | [0.001191] |

When there is no age difference (i. e. DA=0 or LAR=1), the true correlation is 1.000. The DON and the NLIN models overestimate this correlation by 0.0020 and 0.0025, the Lambeth and the DA2 model underestimate by 0.0036 and 0.0286 respectively. It is also interesting to know from the DON model that each year about 2.5% (0.02473) of unexplained variance is accumulated.

While F-test is not applicable to the nonlinear model, among the remaining three models, the DON model has the greatest F-value (Table 3). The Don model also has the greatest R-square when compared with other models. On the other hand, the Lambeth model has the smallest F-value, the largest MSQ error and SSQ error, and the largest predicted residual sum of squares.

The smallest error mean square and error sum_of squares are found in the NLIN model. The DON model should not be compared here because it is not based on the original correlation but on the degree of non-determination.

Table 3.--Comparison of statistical analysis of four models.

| Statistics | Regression Model | | | |
|---|---|---|---|---|
| | DON | DA2 | LAMBETH | NLIN |
| F-value | 416.56 | 172.55 | 88.27 | na |
| MSQ model | 0.61214 | 0.19569 | 0.17876 | 8.56043 |
| MSQ error | 0.00147 | 0.00113 | 0.00203 | 0.00052 |
| R-square | 0.9564 | 0.9008 | 0.8229 | 0.9548* |
| SSQ model | 0.61214 | 0.19569 | 0.17876 | 17.12086 |
| SSQ error | 0.02792 | 0.02155 | 0.03848 | 0.00983 |
| SSQ uncorrected total | | | | 17.13068 |
| SSQ corrected total | 0.64007 | 0.21724 | 0.21724 | 0.21724 |
| Predicted Resid SS | 0.0331 | 0.0263 | 0.0467 | |

na: not applicable
*: calculated from 1-(SSQ error/SSQ corrected total)

## PERFORMANCE COMPARISON OF THE FOUR MODELS

In order to have a valid evaluation of the four models, we should compare how good are the predictions. By inserting the ages into the four regression models and solving for the correlation, the predicted age-age correlations are listed in Table 4. If we rank the predictions from 1 to 4, 1 being the nearest to the observation, and 4, the remotest from the observation, then the DON model with a mean score of 1.667 is the best model (Table 5). The NLIN model is better than the Lambeth model, and the DA2 model is between the NLIN and the Lambeth models.

Using the absolute deviation between predicted and observed values as input variables in a two-way analysis of variance, we find no differences between DON and NLIN models, no differences between DA2 and Lambeth models, but differences between these two groups are highly significant. The mean absolute deviations are 0.016 for the DON and NLIN models, 0.028 for the DA2 model and 0.032 for the Lambeth model respectively.

## CONCLUSIONS

Based on the South-wide Loblolly Pine Provenance Tests data, age-age correlations can be modelled successfully using the degree of non-determination (DON). The DON model is superior to the acclaimed Lambeth model because it has smaller fitting errors. The DON model indicates that the error component is accumulated at a rate about 2.5 percent each year.

Table 4.--Comparison of observed and predicted correlations in the four models.

| Case | Age2 | Age1 | Obser. | DON | DA2 | LAMBETH | NLIN |
|---|---|---|---|---|---|---|---|
| | Age | | Correlation | | | | |
| | | | | Predicted by Model | | | |
| 1 | 3 | 3 | 1.00000 | 1.0020* | 0.9714 | 0.9964 | 1.0025 |
| 2 | 5 | 3 | 0.96739 | 0.9770 | 0.9686* | 0.9264 | 0.9774 |
| 3 | 10 | 3 | 0.90097 | 0.9116 | 0.9371 | 0.8314 | 0.9114* |
| 4 | 15 | 3 | 0.79722 | 0.8410 | 0.8707 | 0.7758* | 0.8402 |
| 5 | 20 | 3 | 0.79064 | 0.7640 | 0.7692* | 0.7364 | 0.7625 |
| 6 | 25 | 3 | 0.65974 | 0.6783 | 0.6328 | 0.7058 | 0.6759* |
| 7 | 5 | 5 | 1.00000 | 1.0020* | 0.9714 | 0.9964 | 1.0025 |
| 8 | 10 | 5 | 0.96333 | 0.9383 | 0.9539* | 0.9014 | 0.9383 |
| 9 | 15 | 5 | 0.87443 | 0.8699* | 0.9014 | 0.8458 | 0.8694 |
| 10 | 20 | 5 | 0.83263 | 0.7957 | 0.8140* | 0.8064 | 0.7945 |
| 11 | 25 | 5 | 0.71456 | 0.7138* | 0.6916 | 0.7758 | 0.7118 |
| 12 | 10 | 10 | 1.00000 | 1.0020* | 0.9714 | 0.9964 | 1.0025 |
| 13 | 15 | 10 | 0.92528 | 0.9383* | 0.9539 | 0.9408 | 0.9383 |
| 14 | 20 | 10 | 0.85869 | 0.8699 | 0.9014 | 0.9014 | 0.8694* |
| 15 | 25 | 10 | 0.75366 | 0.7957 | 0.8140 | 0.8708 | 0.7945* |
| 16 | 15 | 15 | 1.00000 | 1.0020* | 0.9714 | 0.9964 | 1.0025 |
| 17 | 20 | 15 | 0.96427 | 0.9383 | 0.9539 | 0.9569* | 0.9383 |
| 18 | 25 | 15 | 0.91275 | 0.8699 | 0.9014* | 0.9264 | 0.8694 |
| 19 | 20 | 20 | 1.00000 | 1.0020* | 0.9714 | 0.9964 | 1.0025 |
| 20 | 25 | 20 | 0.93072 | 0.9383* | 0.9539 | 0.9658 | 0.9383 |
| 21 | 25 | 25 | 1.00000 | 1.0020* | 0.9714 | 0.9964 | 1.0025 |

* Best predictor in each row.

Table 5.--Grouping of the four models by Duncan's multiple-range test.

| Duncan Grouping | | Mean Rank | Model |
|---|---|---|---|
| 1] Using rank, ranking order 1 = prediction is the closest to observation, 4 = prediction is the remotest from observation. | | | |
| | A | 3.190 | LAMBETH |
| | A | | |
| B | A | 2.857 | DA2 |
| B | | | |
| B | | 2.286 | NLIN |
| | C | 1.667 | DON |
| 2] Using absolute deviation between prediction and observation. | | | |
| | A | 0.03159 | LAMBETH |
| | A | | |
| | A | 0.02787 | DA2 |
| | B | 0.01596 | NLIN |
| | B | | |
| | B | 0.01576 | DON |

## LITERATURE CITED


Bohren, B. B. 1975. Designing artificial selection experiments for specific objectives. Genetics 80: 205-220.

Kung F. H. 1973. Development and use of juvenile-mature correlations in a black walnut tree improvement program. Proc. So. For. Tree Improv. Conf. 12: 243-249.

Kung F. H. 1975. The efficacy of early or indirect truncated selection. Proc. Cent. St. Tree Improv. Conf. 9: 35-41.

Lambeth, C. C. 1980. Juvenile-mature correlations in Pinaceae and implications for early selection. For. Sci. 26: 571-580.

Nance W. L. and 0. 0. Wells. 1981. Site index models for Height growth of planted loblolly pine *(pinus teada* L.) seed sources. South. For. Tree Improv. Conf. 16: 86-96.

SAS Institute, Inc. 1987. SAS/STAT guide for personal computer, Ver. 6 ed. SAS Institute, Inc., Cary, NC. 378 p.

Squillace A. E. and C. R. Gansel 1968. Assessing the potential oleoresin yields of slash pine progenies at juvenile ages. USDA For. Serv. Res. Note SE-95, 4 pp.

Wells, 0. 0. and P. C. Wakeley. 1966. Geographic variation in survival, growth and fusiform-rust infection of planted loblolly pine. For. Sci. Monogr. 11, 40 pp.

Yang C. Y. and F. H. Kung. 1987. Symmetrical response-surface models for age-age autocorrelation in volume growth of cryptomeria trees. in Forest Growth Modelling and Prediction. USDA For. Serv. Gen. Tech. Rep. NC-120. Vol 1:572-579.