Construction and Interpretation of Path Diagrams for Height and Increment from Loblolly Pine Provenances Tests in Time Series

By F. H. KUNG[1])

(Received 19th June 1992)

Summary

Growth is an accumulation process. The total height model is described by the previous height and the last increment. The periodic increment model is described by three causes: (1) the previous height, (2) all prior periodic increments, and (3) other factors or random error. In this paper, height and increment models at age 5, 10, 15, 20, and 25 years were built by using data from the Southwide Loblolly Pine Provenance Test. The regression coefficients from these models were converted into path coefficients and then they were assembled into path diagrams. A synthesized path diagram can show the cause and effect relationship in any combinations of heights and periodic increments. Variances and correlations can be partitioned by tracing separate routes in a path diagram so that important factors may be identified. After construction, the path diagram was verified by comparing correlations calculated from the path diagram and those calculated from the data set. They were identical. For loblolly pine, paths leading to total height are all positive, so there is persistent dominance of tall and fast growing trees. Paths leading to periodic increment may be positive, negative, or may be greater than unity, it shows that there may exist a counterbalance between total height and preceding increment in forming subsequent increment. There seems to be a growth phase shift and a change of direction in counterbalance about 15 years of age. Applications of path analysis to provenance selection are discussed.

Key words: Path analysis, growth analysis, variance components.

Introduction

According to BRUCE and WENSEL (1988), there are 2 types of growth models: "process" and "empirical" models. The "process" models simulate the biological process that converts carbon dioxide, nutrients, and moisture into biomass through photosynthesis. "Empirical" models are based on periodical measurements and make no attempt to measure every factor that may affect tree growth. This paper is an attempt to use path analysis to build an empirical growth model.

To construct an empirical growth model, estimates (e. g., tree diameter, height, volume, number) are made of the changes with time, and with other driving functions based on tree species, age, site, climate, and management history. In this paper, only height growth over time is considered.

Most growth models use point estimates of growth to provide mean performance under a given set of driving functions. The growth model presented here deals with the scale estimate, the variance of growth at a given time in a time series.

The objective of this paper is to demonstrate (1) how a path model may be constructed from data in a longitudinal time series, (2) how the paths are interconnected, (3) how variance and correlation may be partitioned according to the path model, and (4) how the contribution of each component may be transmitted over time.

[1]) Professor, Dept. of Forestry, Southern Illinois University, Carbondale, IL. 62901, USA

Throughout this paper, algebraic expressions are formulated in the computer language format. For example, The symbol * is for multiplication, / for division, () for subscript inclusion, and finally, SQRT(X) is for the square root function of variable X.

Path Structure

In a longitudinal study, let us assume that the investigation period starts with point (a), passing through points (b) and (c), and ends at point (d). Furthermore, let us denote H for height, I for increment and E for error.

Mathematically, the total height at any given time is completely determined by two parts: (i) the last cumulative height, i. e., the last observation, and (ii) the antecedent increment, or difference between the last observation and the current observation. For example, H(b) is the sum of two values H(a) and I(b-a), while H(c) is the sum of H(b) and I(c-b). Except the rounding error, it assumed that there are no errors in adding the 2 parts to the whole.

The periodic increment may be modelled by three types of factors, (1) the last cumulative growth from which the current increment was initiated, (2) all prior increments, and (3) an error. For example, I(b-a) being the very first increment, it will have only two factors, H(a) and E(b-a), in the model. The second increment in *figure 1, I(c-b)*, is determined by H(b), I(b-a), and E(c-b), while the third increment, I(d-c), is determined by H(c), both I(b-a) and I(c-b), and E(d-c).

If we place the dependent variable at the head of an arrow and the independent variable at the tail of that arrow, the direction of the arrow then represents a cause-and-effect relationship between the 2 variables. As a

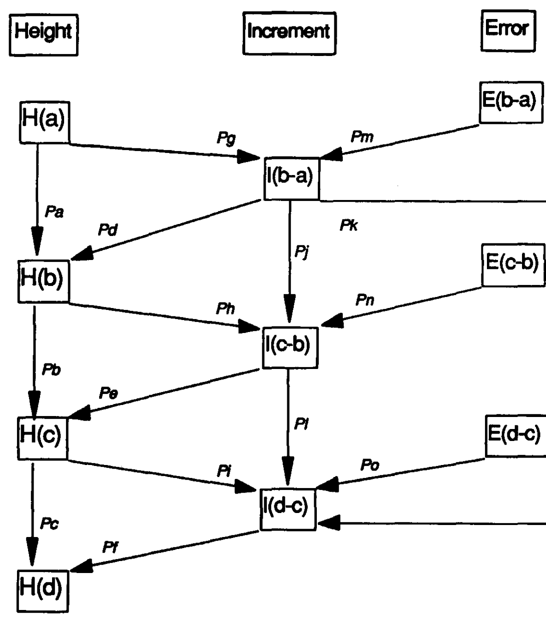

Figure 1. — Path diagram for height and increment interrelationship.

response variable may respond to several stimuli, graphically it can be the target of many arrows. Because in a chain event a certain variable may serve as response variable in one part of the chain and as predictor variable in another part, that variable can be situated in the middle of two arrows. Thus, the interrelationships among heights, increments and errors from point (a) to point (d) in time may be represented by the structured path diagram in *figure 1*.

Path Formulation

For a linear regression model, the dependent variable Y is determined by a linear combination of weighted independent variables X. The weight is the regression coefficient. In a path model, the path coefficient is a regression coefficient with respect to standardized variable LI, 1975, p. 111).

To illustrate the relationship between regression model and path model for total height, let us start with the regression model for H(b). In equation 1, the height at the end, H(b), is the sum of initial height, H(a) and the increment I(b-a):

$$H(b) = H(a) + I(b-a) \qquad \ldots (1)$$

Because we have assumed that there are no errors in the above equation, the variable H(a) is completely determined by variables H(a) and I(b-a). There is no use to run regression analysis for H(b), using H(a) and I(b-a) as input variables, because the model is not a full rank model. However, since it takes one observation of the last height and one observation of the subsequent increment to get the total height, equation (1) may be viewed as a regression model:

$$H(b) = 1.0{*}H(a) + 1.0{*}I(b-a) + 0.0 \qquad \ldots (2)$$

In other words, the regression coefficient for H(a) and I(b-a) are both unity and the error terms is fixed at zero. Therefore, Pa in *figure 1* is simply the standard deviation of H(a) divided by the standard deviation of H(b). Let S denote standard deviation, we have

$$Pa = SH(a) / SH(b) \qquad \ldots (3)$$

Similarly,

$$Pd = SI(b-a) / SH(b) \qquad \ldots (4)$$

Again, regression analysis is not needed for evaluating paths Pb and Pe, Pc and Pf. They can be calculated by their respective ratio of standard deviations.

On the contrary, the path model for increment must to be developed from its regression model. Let us take increment I(c-b) for example. To obtain the value for Ph, Pj, and Pn in *figure 1*, we need to develop a regression model for the increment I(c-b) as follows:

$$I(c-b) = B_0 + B_1{*}H(b) + B_2{*}I(b-a) + E(c-b) \qquad \ldots (5)$$

Then we need to standardize the regression coefficients so that

$$Ph = B_1{*}SH(b)/SI(c-b) \qquad \ldots (6)$$

and

$$Pj = B_2{*}SI(b-a)/SI(c-b) \qquad \ldots (7)$$

From the regression model we can also obtain the degree of determination, or the so called "rsquare" for a given model. The path Pn from error to increment can be computed by the square root of the quantity (1-rsquare):

$$Pn = SQRT(1-rsquare) \qquad \ldots (8)$$

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

Provenance No.	Region		Mean height at age (years) 3	5	10	15	20	25
			-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. — Standard deviations of height and increment at various ages.

Age	Height Variable	Std. Dev.	Increment Variable	Std. Dev.
Years		cm.		cm.
3	H(3)	12.32		
			I(5-3)	13.53
5	H(5)	25.08		
			I(10-5)	25.30
10	H(10)	48.55		
			I(15-10)	23.91
15	H(15)	60.17		
			I(20-15)	18.40
20	H(20)	67.21		
			I(25-20)	25.15
25	H(25)	67.88		

Following the same procedure, paths Pg and Pm are calculated from a model using I(b-a) as response variable and H(a) and E(b-a) as predictor variables; paths Pi, Pl and Po are computed from a model with I(d-c) as dependent variable and H(c), I(c-b), and I(b-a) as independent variables.

The 3 path models for height (i. e., H(b), H(c) and H(d)) and the three path models for increment (i. e., I(b-a), I(c-b) and I(d-c)) can be assembled together to form a single synthesized path diagram. *Figure 1* represent a 3-level model, more layers can be added. Method of estimation is the same for later paths.

Empirical Path Calculation

The original South-wide Loblolly Pine Seed Source Study was described by Wells and Wakeley (1966). Later plantation management and measurements were summarized by Nance and Wells (1981). The mean height at age 3, 5, 10, 15, 20, and 25 years for each of the 15 provenances are given in *table 1*. Each provenance mean was computed from about 500 trees.

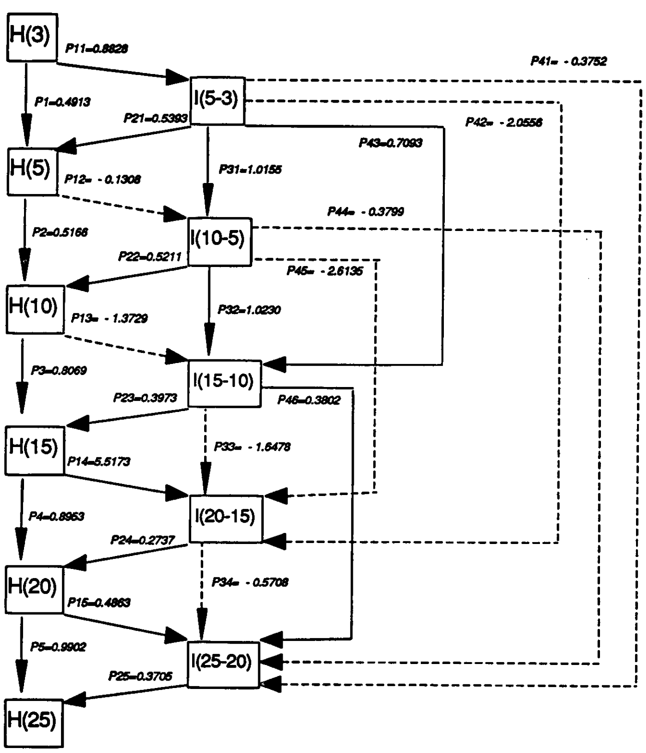

Figure 2. — Path diagram for loblolly pine from age 3 to 25. Negative paths are represenaed by dash lines.

Periodic height increment was computed from the difference between two consecutive cumulative heights (e. g., I(5-3) = H(5) — H(3)). The standard deviation of height and of increment at various age are presented in *table 2*. With these standard deviations, we can immediately construct the paths leading to cumulative height in *figure 2*. For example, according to equation (3) and (4) we have the two paths leading to H(5):

P1 = 12.32/25.08 = 0.4913,
P21 = 13.53/25.08 = 0.5393,

Similarly, we can compute the two paths leading to H(10) as:

P2 = 25.08/48.55 = 0.5166,
and P22 = 25.30/48.55 = 0.5211.

The rest of path pairs P3 and P23, P4 and P24, P5 and P25 can be computed directly from the associated standard deviations in *table 2*. The numbering of paths in *figure 2* increase from left to right and from top to bottom.

Regression coefficients for the periodic increment models are presented in *table 3*. They were used together with standard deviations listed in *table 2* to form the paths leading to the periodic increment. For example, given coefficients for Model 2 in *table 3*, and according to equation (6) and (7), we have:

P12 = —0.13197 * 25.08/25.30 = —0.1308
P31 = 1.89944 * 13.53/25.30 = 1.0155.

Similarly, given Model 3 in *table 3*, we have:

P13 = —0.67611 * 48.55/23.91 = —1.3729
P32 = 0.96678 * 25.30/23.91 = 1.0230
P43 = 1.25351 * 13.53/23.91 = 0.7093

The degree of determination given for each increment model in *table 3* can be used to compute the path from error to increment. For example, given the R-square for Model 2 in *table 3* as 0.7898, the path Pn in *figure 1* can be computed by equation (8) as:

Pn = SQRT(1—0.7898) = 0.4585.

Similarly, with Model 3 we have Po in *figure 1* as:

Po = SQRT(1—0.1352) = 0.9299.

In order to reduce overcrowding in the graph, the paths from error to increment were not presented in *figure 2*. Thus, *Figure 2* is a partial diagram only to show the interrelationships among growths and increments. *Figure 2*

Table 3. — Results of regression analysis for the increment models. Data of increment were generated from *table 1*.

Dependent Variable	Independent Variables	Regression Coefficient	T for H0	Prob >|T|	R-square
		-----MODEL1-----			
I(5-3)					0.7794
	Intercept	56.50582	3.067	0.0090	
	H(3)	0.96899	6.776	0.0001	
		-----MODEL2-----			
I(10-5)					0.7898
	Intercept	220.04815	4.984	0.0003	
	H(5)	-0.13194	-0.237	0.8160	
	I(5-3)	1.89938	1.843	0.0883	
		-----MODEL3-----			
I(15-10)					0.1352
	Intercept	229.59649	1.528	0.1524	
	H(10)	-0.67611	-0.609	0.5542	
	I(10-5)	0.96678	0.800	0.4392	
	I(5-3)	1.25351	0.543	0.5971	
		-----MODEL4-----			
I(20-15)					0.4649
	Intercept	261.49896	2.518	0.0286	
	H(15)	1.68719	2.366	0.0374	
	I(15-10)	-1.26804	-1.799	0.0995	
	I(10-5)	-1.90074	-2.424	0.0337	
	I(5-3)	-2.79539	-1.892	0.0851	
		-----MODEL5-----			
I(25-20)					0.3084
	Intercept	409.20184	1.926	0.0830	
	H(20)	0.18199	0.127	0.9011	
	I(20-15)	-0.78026	-0.443	0.6669	
	I(15-10)	0.39990	0.306	0.7659	
	I(10-5)	-0.37765	-0.239	0.8162	
	I(5-3)	-0.69737	-0.252	0.8065	

is valid for displaying the variance of total height and for displaying correlations among heights and increments, but it is not a closed system for the variance of increment. Please keep in mind that a closed system for increment should be completed with all the error paths as illustrated in *figure 1*.

Model Validation

The complete model constructed from data in *tables 2* and *3* is shown in *figure 2*. According to path analysis (Li, 1975), the correlation between 2 variables is the sum or product of all paths between them: paths arranged in series are multiplicative while paths in parallel are additive. Some other common rules of tracing the paths in a path diagram are:

(i) A path is represented by an arrowhead at one end of a line, while a correlation is represented by arrowheads at both ends of a line. Thus, a path is similar to a one-way street while a correlation is similar to a two-way street.

(ii) In tracing paths from one variable to another, direction of travel is first against the arrow, and then follow the arrow if needed. One and only one change in direction is permitted. Against the arrow trekking is necessary because it traces backward from a result to a cause. Trekking with one change of direction is similar to finding a common cause for 2 results.

According to these rules, there is no correlation between H(a) and E(c-b) in *figure 1*, because starting from either point the direction of travel is following the arrow and not against the arrow. Also, there is no single path and no multi-path with "against-the-arrow" traveling routes between them. The logic is that both variables H(a) and E(c-b) have no prior common denounminators and they are initial causes to later growths. On the contrary, there is a correlation between H(b) and E(b-a) because H(b) is the result and E(b-a), the cause. Trekking with one change of direction is illustrated in the next example. The correlation between H(b) and I(b-a) includes two routes: Pd and Pa*Pg. Starting from H(b), the direct route Pd represents that H(b) is affected by I(b-a), whereas the indirect route Pa*Pg, with a change of direction at H(a), symbolizes that H(a) is a prior cause for both variables H(b) and I(b-a).

After learning the method of tracing paths on a path diagram, we can verify that a path diagram is a complete and exact model for correlations. The correlation of a variable with itself is unity. This unity can be viewed as complete variance for that variable. Each variance of total height can be partitioned by tracking the round trips starting and ending at that particular total height. For example, in *figure 2*, H(5) is composed of the following 4 routes:

[Example 1]

(i)	*via* H(3)	P1*P1	= 0.4913*0.4913	= 0.2414
(ii)	*via* H(3), I(5-3)	P1*P11*P21	= 0.4913*0.8828*0.5393	= 0.2339
(iii)	*via* I(5-3)	P21*P21	= 0.5393*0.5393	= 0.2908
(iv)	*via* I(5-3), H(3)	P21*P11*P1	= 0.5393*0.8828*0.4913	= 0.2339
			sum	= 1.0000

Therefore, the variance of H(5) is the sum of three types of contribution: (1) the direct effect of initial height growth H(3), the contribution from item (i) is 24%; (2) the direct effect of periodic increment I(5-3), it contributes 29% from item (iii); and finally, (3) the joint effect of both factors, the subtotal is 47% from items (ii) and (iv).

Variance partitioning for H(10) is similar to above procedure. However, it is necessary to add a new rule here for tracing the path diagram: no intermediate cause may be counted more than once within a route. After leaving the starting point, each station in the round trip can be visited once and only once. In other words, there is no "figure-8 double round trips," no "in-and-out-and-in-and-out stop-over." For example, in *figure 2*, from H(10) to H(3) the trip P2*P1*P1*P2 is not allowed because it stops over H(5) twice and form a *"figure-8* double round trip." For the same reason, the trip P2*P1*P11*P21*P12*P22 cannot be counted. Therefore, the legitimate pathways for variance component in H(10) are those passing through the following contributing factors or causes:

[Example 2]

(i)	H(5):	P2*P2	= 0.2669
(ii)	I(10-5):	P22*P22	= 0.2715
(iii)	H(5), I(10-5):	P2*P12*P22 + P22*P12*P2	=-0.0704
(iv)	H(5), I(5-3), I(10-5):	2*P2*P21*P31*P22	= 0.2949
(v)	H(5), H(3), I(5-3), I(10-5):	2*P2*P1*P11*P31*P22	= 0.2371
		sum	= 1.0000

Variance for increment can be partitioned in the same manner. The components and pathaway for the variance of I(10-5), excluding the error contribution, are as follows: follows:

[Example 3]

(i)	H(5):	P12*P12	= 0.0171
(ii)	I(5-3):	P31*P31	= 1.0312
(iii)	H(5), I(5-3):	2*P12*P21*P31	=-0.1433
(iv)	H(5), H(3), I(5-3):	2*P12*P1*P11*P31	=-0.1152
		sum	= 0.7898

The resulting sum is identical to the R-square for Model 2 in *table 3*. Thus, the total variance in variable I(10-5) is unity when the contribution from error is added. The path model for increment, therefore, is validated.

After the special case for r=1 is verified, we need to show that the same path diagram is also applicable to general case of correlation partitioning. *Table 4* presents correlation matrices among heights and increments. They will be used for further validation of the path model.

First, let us look at correlation between heights. The correlation between H(3) and H(5) is given as 0.9674 in *table 4*. If we trace the paths between H(3) and H(5), we will find two routes starting from H(5) and their values are:

[Example 4]

(i)	P1		= 0.4913
(ii)	P21*P11	= 0.5393*0.8828	= 0.4761
		sum	= 0.9674

To go down another step, the correlation between H(3) and H(10) given in *table 4* is 0.9010. From the path diagram this correlation can be partitioned into 5 routes starting from H(10) as follows:

[Example 5]

(i)	P2*P1	= 0.2538
(ii)	P22*P12*P1	= -0.0335
(iii)	P22*P31*P11	= 0.4672
(iv)	P2*P21*P11	= 0.2460
(v)	P22*P12*P21*P11	= -0.0325
		------------
	sum	= 0.9010

In the next example, we will verify the correlation between total height, H(3), and periodic increment, I(10-5), given in *table 4* as 0.7699. The procedure is to trace and to add up the following routes going from I(10-5) to H(3):

[Example 6]

(i)	P12*P1	= -0.0643
(ii)	P12*P21*P11	= -0.0623
(iii)	P31*P11	= 0.8965
		------------
	sum	= 0.7699

Finally, the correlation between increment I(10-5) and increment I(5-3) is the sum of the following 3 terms:

[Example 7]

(i)	P31	= 1.0155
(ii)	P12*P21	= -0.0705
(iii)	P12*P1*P11	= -0.0567
		------------
	sum	= 0.8883

Thus, all correlations listed in *table 4* can be constructed from path diagram in *figure 2*. However, the closer the interval between two variables, and the earlier the age of the variables, the simpler is the task of finding all allowable connecting paths. When the routes are many and complex, one may be tempted to start from two variables and try to find some common junctions for them to meet. Such strategies usually give incomplete enumerations. Given two variables, I find it is unfailing to start with the lower one in the path diagram and search for all immediate predecessors, then move up to each predecessor and repeat the process until the target variable is found. For example, tracing the paths between H(3) and I(15-10) can be facilitated by finding and expanding the following three routes coming to I(15—10): (i) P13*correlation between H(3) and H(10), (ii) P32*correlation between H(3) and I(10—5) and (iii) P43*P11. The reader may want to verify that the sum of these 3 routes is the same as the sum of the following 9 trips:

[Example 8]

five trips from (i):	(1) P13*	P2	*P1
	(2) P13*	P22*P12	*P1
	(3) P13*	P22*P31	*P11
	(4) P13*	P2*P21	*P11
	(5) P13*	P22*P12*P21	*P11
three trips from (ii):	(6) P32*	P12	*P1
	(7) P32*	P12*P21	*P11
	(8) P32*	P31	*P11
one trip from (iii):	(9) P43		*P11

Discussion

An interesting question about the error term was raised by a reviewer. During the data collection phase, tree heights were measured with errors and increments were obtained by substraction; then, why should all errors in the path model be allocated to the increment factor and not to the height factor? The reader should keep in mind that the path model is subjective "cause and effect model", not an objective procedure model. The path model tries to explain what could happen from the view point described by the modeler. The modeler wants to show how the data might be related and not how the data were collected. If indeed we had actually measured three variables (initial height, increment, and final height) separately, we could have the following regression model:

$$\text{Final height} = b_0 + b_1 \text{ (initial height)} + b_2 \text{ (increment)} + \text{error} \quad \ldots (9)$$

Since we have measured only initial heights and final heights, therefore, the choice of model is either

$$\text{(1) final height} = b_0 + b_1 \text{ (initial height)} + \text{residual fitting error} \quad \ldots (10)$$

or,

$$\text{(2) final height} = b_0 + b_1 \text{ (initial height)} + \text{increment.} \quad \ldots (11)$$

The path model takes the second alternative. By setting b_0 to 0, equation (11) is changed to equation (2). Therefore, total heights are considered to be error free because they were obtained simply by adding 2 quantities together. On the other hand, although the increment is computed from difference between 2 heights during data acquisition, the 2 heights are not included as independent variables at the same time during regression analysis for a increment model, so an error term must be added.

The path model may be viewed as an "input-output model." The following input-output rules are useful in constructing any height-increment path models at any length. The very first height supplys two outputs to the system and the final height acquires two inputs from the system. All other heights receive two inputs and produce 2 outputs. Although the increments have variable numbers of inputs and oputs, the number of inputs equals the order of that increment in the series, and the sum of these 2 numbers is always a constant, which equals to the total number of increments in the system plus one. The one is

Table 4. — Correlation among heights and periodic increments in loblolly pine.

Variable	Variable									
	H(3)	H(5)	H(10)	H(15)	H(20)	H(25)	I(5-3)	I(10-5)	I(15-10)	I(20-15)
	Correlation among heights									
H(5)	0.9674									
H(10)	0.9010	0.9633								
H(15)	0.7972	0.8744	0.9253							
H(20)	0.7906	0.8326	0.8587	0.9643						
H(25)	0.6597	0.7146	0.7537	0.9128	0.9307					
	Correlation between height and increment						Correlation among increments			
I(5-3)	0.8828	0.9730	0.9655	0.8952	0.8237	0.7240				
I(10-5)	0.7699	0.8572	0.9640	0.9087	0.8224	0.7379	0.8882			
I(15-10)	0.1767	0.2444	0.2979	0.6377	0.6831	0.7667	0.2922	0.3294		
I(20-15)	0.2810	0.1819	0.1107	0.2525	0.4994	0.4149	0.0813	0.0322	0.4095	
I(25-20)	−0.3323	−0.2966	−0.2607	−0.1135	−0.1604	0.2117	−0.2472	−0.2063	0.2438	−0.2149

for the last total height. It can be easily verified that in *figure 2* the are 1 input and 5 outputs for the first increment, 2 inputs and 4 outputs for the second increment, and finally, 5 inputs and 1 output for the 5th and the final increment. When there are 5 increments in the path diagram, the total number of inputs and outputs adding together is always equal to 6.

It is well known that heights are affected by various genetic and environment variances. How does this affect the path model? Since the path diagrams included 3 types of effects: (1) height, (2) increment and (3) error, the genetic and environmental variannces are all hidden, concealed by the fitting error in the increment model. The coefficient of alienation (1-rsquare) indicates how strong the other variances affect the increment model. Because a path diagram describes correlations completely, no more and no less, so until the other genetic and environment factors are observed and presented in the correlation matrices, they cannot be resolved in the path diagram.

The beauty of variance partitioning by path diagrams is that it presents contributions not only by independent variables alone, but also by combinations of correlated variables; it displays not only direct effects, but also compounded indirect effects. The total contribution of a cause to the variance of a dependent variable is the sum of each contribution that the cause has involved. In *example 2*, the total contribution from H(5) to H(10) is the sum of four terms which included H(5) as one of the factors: 0.2669 — 0.0704 + 0.2949 + 0.2371 = 0.7285. Accordingly, other causes and each total contributions are:

I(10-5) = 0.7331
I(5-3) = 0.5320
H(3) = 0.2371

From the example above, we can see that H(5) and I(10-5) have about equal contribution to H(10), they also contribute about 3 times that of H(3).

The separate contribution for an effect standing alone or joined with other effects is interesting, specially in the presence of some negative paths. In *example 3*, the effect of H(5) standing alone will add about 2% to the variance of I(10-5), but when it is joined with the other 2 factors, H(3) and I(5-3), it will reduce the variance of I(10-5) by 26%. The net loss of variance therefore, is 24%. The variance reduction as revealed by path analysis is particular informative and is not immediately available from other commonly used statistical procedures such as forward or backward stepwise regression analysis.

A path coefficient, unlike a correlation or a partial correlation coefficient, may take values greater than unity. As given in *Figure 2* and in example 3, the path from I(5-3) to I(10-5), P31, is greater than unity, and its direct contribution to the variance of I(10-5) then is also greater than unity. Without looking at the whole picture it is difficult to accept that a cause can contribute more than 100% to a circumstance. However, because the path P12 is negative, the total contribution due to I(5-3), as calculated by summing up the separate contributions of P31*P31, 2*P12*P21*P31, and 2*p12*p1*p11*p31 is only 0.7727. Because the total degree of determination cannot exceed unity, the presence of a path coefficient greater than unity in a causal system is a sure signal for the existence of a compensatory mechanism (negative influence) in that system. When viewed this way, the property that a path coefficient may take values greater than unity is not only meaningful but of practical importance (LI, 1975, p. 170).

The probability of the absolute t-value for the regression coefficients suggests that not all coefficients in *Table 3* are statistically significant at the 5% level. Should we delete these non significant paths? For example, the variables (H(5) and I(5-3) are not significant in Model 2, *table 3*; should we delete P12 and P31 in *figure 2?* The answer is no.

The correlation between two variables will no longer be true if any path leading to that dependent variable is deleted after the complete path model is built. Following the previous example, if P12 and P31 were deleted from the path diagram, then there is only one route left to

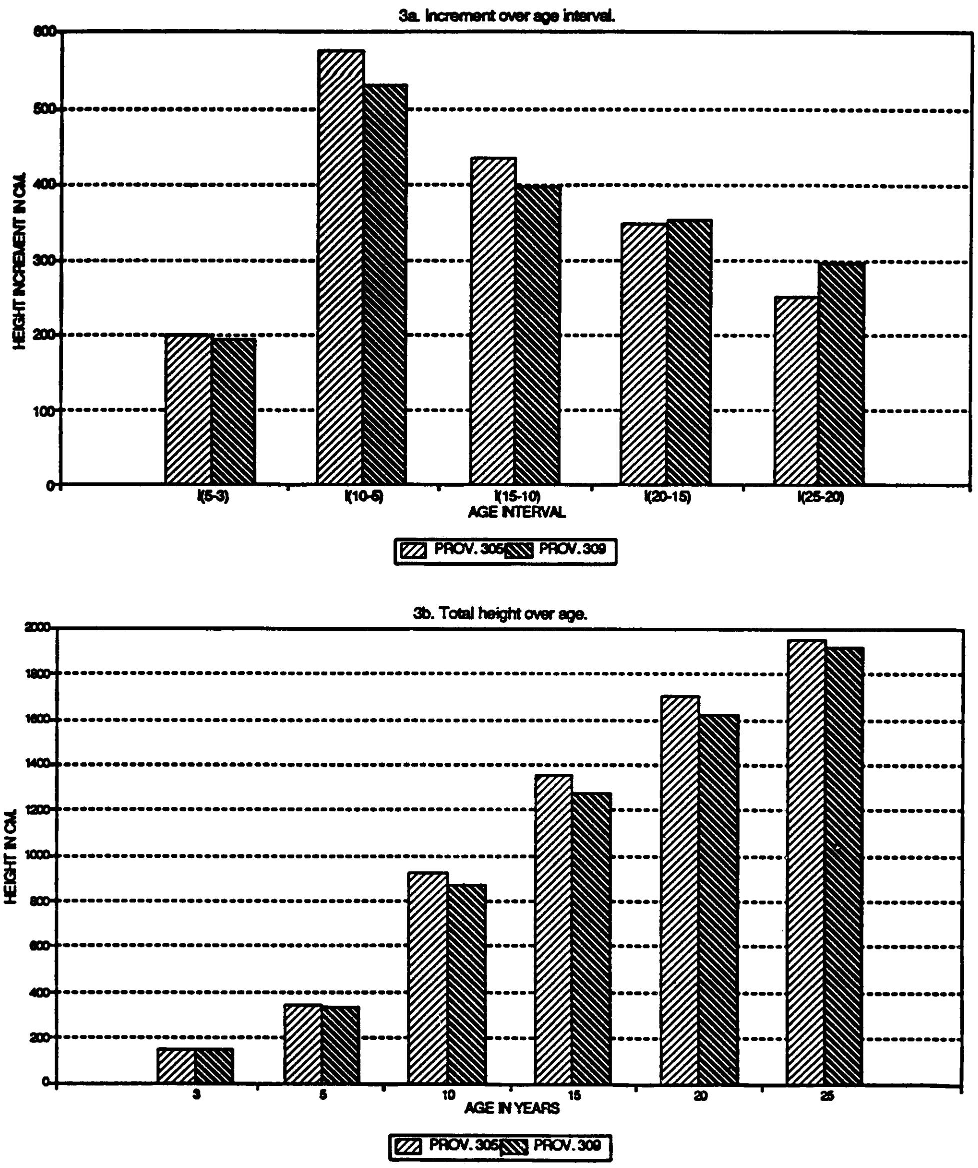

Figure 3. — Contrast of height and increment growths in 2 provenances.

between H(5) and H(10): that is going through the path P2. Since the value of P2 is 0.5166 while the correlation between H(5) and H(10) is 0.0633, such discrepancy is too large to be ignored.

The path diagram shows a consistent trend in the paths connecting two heights. From top to bottom, paths on the left (P1 through P5) keep increasing with time. Because the square of a path is the direct contribution of an effect, the contribution of the preceding height to current height is also increasing with age. The direct contribution from the preceding height at age 10 is only 27%, it doubles at age 15 to 65%, increases to 80% at age 20, finally it reaches 98% at age 25. Thus, the later the age of selection, the more effective is the selection for the final height. Furthermore, effects of paths in series are multiplicative, given each path connecting cumulative heights is less than unity, the direct contribution of early height to successive heights becomes smaller and smaller as trees are growing older and older. Fox example, direct contribution from H(3) to H(10) is 6%, to H(15) is 4%, to H(20) and H(25) is only 3%. It shows that for early selection, there are diminishing returns with increasing age interval. Thus, early selection is not as effective as later selection.

Contributions from an increment to the next height were all positive. Direct contribution was about 28% before age 10 and about 15% to 7% from age 15 to 25. This suggests that early increments may be useful for selecting early height but later increment may not be effective for selecting matur height.

In forming subsequent height increment, there seems to be a counterbalance between preceding increment and total height. Notice that in *figure 2*, before age 15, if an input path (e. g., P12) from a height (e. g., H(5)) to an increment (e. g., I(10-5)) was negative then the path (e. g., P31) from the last increment (e. g., (I(5-3)) to this assigned increment (e. g., I(10-5)) would be positive. The opposite was found after age 15. As the input path from height to this increment (P14 or P15) became positive, the input path from the last increment to this increment (P33 or P34) became negative.

Age 15 also seems to be a switching zone for the sign of ouput path from one increment to other increments. Take the output paths from I(5-3) in *figure 2* as example, P31 and P43 are positive, but P41 and P42 are negative. Again, we see that starting from I(10-5), path P32 is positive but paths P44 and P45 are negative. In the present study the timing of separation between two age groups by the reversal of path signs agrees with age break that separates juvenile and mature genotype phases commonly observed in 4 North American Conifers (FRANKLIN, 1979).

The positive output path may be an indication of competition advantage of fast growing trees. On the contrary, the negative output path may symbolize "early burn-out" of fast growing provenances and "late catch-up" of the slow starters. It has been shown that there are genetic differences in relative growth rate and acceleration of growth rate among loblolly pine seed sources (KUNG, 1987). For example, provenance C-305, from Pamlico County in eastern North Carolina, has the greatest relative growth rate but the least acceleration, while provenance C-309 from WILCOX and CRISP counties in southwestern Georgia, has the least relative growth rate but the highest acceleration. Compare the height increments of the two provenances in *Figure 3a,* provenance C-305 grew faster than provenance C-309 before age 15, but became slower after age 15. The slope of change is steeper for provenance C-305 than the slope for provenance C-309. The height over age curves for these 2 provenances diverge from age 3 and converge at the end (*Figure 3b*). The divergence causes positive and the convergence causes negative incremental paths.

Additional interpretation for the negative paths may be that the fast growing trees, after gaining dominance in height, are changing emphasis from vegetative growth to reproductive growth. Althought there were no actual reproduction data available from this study, it has been shown that individual loblolly pine trees occasionally produce flowers and viable seed at less than 10 years of age (RIGHTER, 1939; WEDDELL, 1935). Seed production increases gradually until the trees are 30 to 50 years old (POMEROY, 1949). Negative correlations were seen where growth tradeoffs, growth strategies, and defensive investments were involved (LOEHLE and NAMKOONG, 1987). Height growth of loblolly pine may also be subject to a population from of developmental canalization, similar to the examples given in Douglas-fir (NAMKOONG et al, 1972) and in ponderosa pine (NAMKOONG and CONKLE, 1976).

When the output path from height to increment is positive, the fast growing seed sources will increase faster than the slow growing seed sources. The differentiation among provenances increases with age. It may be worthwhile to delay early selection and to wait for the coming of the highest selection differential. But when the path becomes negative, differences among provenances are diminishing with time. In the latter case, selection should not be delayed any further but should start soon. Otherwise, selection differential and genetic gain would be reduced with waiting.

Conclusions

The path diagram is an exact model for growth and increment correlations. It can be used to partition total variance of a variable, and to partition correlation coefficient between two variables ,into separate additive components. Because single direct effect and the joint contribution of causes can be seen, the path diagrams offer better interpretation and deeper insight to the growth system analysis. Input paths forming total height are all positive, this signifies the chronic dominance of tall and fast growing trees in the test plantation. Paths forming increment may be positive, negative, or may be greater than unity. The negative paths are especially useful because they show the existence of a compensatory mechanism (negative influence) in that system.

In producing the subsequent increment, the counterbalance between total growth and preceding increment seesawed at age 15. There seems to be a growth phase shift around age 15 for loblolly pine provenance tests. Thus, early selection would not be effective and should be discouraged.

Acknowledgements

The author is grateful to Dr. OSBORN O. WELLS and the Southern Forest Experiment Station, U. S. Forest Service for providing data for this study. The author would like to thank Dr. ANTHONY E. SQUILLACE, Dr. GEORGE RINK, Dr. E. K. MORGENSTERN and Dr. J. P. VAN BUIJTENEN for reviewing the manuscript and giving many editorial improvements to this paper

Literature Cited

BRUCE, D. and WENSEL, L. C.: Modelling forest growth: Approaches, definitions, and problems. P. 1—8, in: IUFRO Forest Growth Modelling and Prediction Conf. Ek., A. R.. SHIFLEY, S. R. and BURK, T. E. (eds.). USDA For. Serv. Gen. Tech. Rep. NC—120. 1149 p. (1988). —— FRANKLIN, E. C.: Model relating levels of genetic variance to stand deveopment of four North American conifers. Silvae Genetica **28**: 207—212 (1979). —— KUNG, F. H.: Differences in growth rate and in acceleration of growth rate among loblolly pine rangewide seed sources confirmed. South. For. Tree Impro. Conf. **19**: 407—414 (1987). —— LI, C. C.: Path Analysis — a primer. Boxwood Press, Pacific Grove, Calif. 346 pp. (1975). —— LOEHLE, C. and NAMKOONG, G.: Constraints on tree breeding: growth tradeoffs, growth strategies, and defensive investments. For. Sci. **33**: 1089—1097 (1987). —— NAMKOONG, G. and CONKLE, M. T.: Time trends in genetic control of height growth in ponderosa pine. For. Sci. **22**: 2—12 (1976). —— NAMKOONG, G., USANIS, R. A. and SILEN, R. R.: Age-related variation in genetic control of height growth in Douglas fir. Theo. Appl. Genet. **42**: 151—159 (1972). —— NANCE, W. L. and WELLS, O. O.: Site index models for Height growth of planted loblolly pine (*Pinus taeda* L.) seed sources. South. For. Tree Impro. Conf. **16**: 86—96 (1981). —— POMEROY, K. B.: Loblolly pine seed trees: selection, fruitfulnes, and mortality. U. S. For. Serv. Southeast. For. Expt. Sta., Paper 5, 17 pp. (1949). —— RIGHTER, F. I.: Early flower production among the pine. Jour. For. **37**: 935—938 (1939). —— WEDDELL, D. J.: Viable seed from nine-year-old southern pine. Jour. For. **33**: 902 (1935). —— WELLS, O. O. and WAKELEY, P. C.: Geographic variation in survival, growth and fusiform-rust infection of planted loblolly pine. For. Sci. Monogr. **11**, 40 pp. (1966).