PROGENY TEST DATA SUMMARIZATION PROCEDURES IN THE WESTERN GULF FOREST TREE IMPROVEMENT PROGRAM

W. J. Lowe and J. P. van Buijtenen[1]

Abstract. One of the major objectives of progeny tests in an operational tree improvement program is parental evaluation. To accomplish this objective, data usually needs to be consolidated from progeny tests of different ages, located on different sites, and which contain different variances. Percentage data is subject to scale effects and usually results in a greater weighting of data from younger tests. It also overestimates genetic gain because phenotypic values are used instead of estimates of genetic values. The Western Gulf Forest Tree Improvement Program (WGFTIP) has developed a technique which utilizes standardized performance scores and the coefficient of genetic prediction to estimate parental breeding values at a common age. In addition to reporting an estimated breeding value instead of phenotypic superiority, this technique weights data from older progeny tests heavier than young tests and adjusts for different variances among tests. The incorporation of this technique in the WGFTIP's slash and loblolly pine progeny testing programs is described.

Keywords: *Pinus taeda* L., *Pinus elliottii* Engelm., breeding value, indirect selection.

INTRODUCTION

Genetic tests can be established to satisfy numerous objectives (McKinley, 1983); however, parental evaluation and ranking is one of the major functions of genetics tests in operational tree improvement programs. Parental evaluations are used to rogue seed orchards, design new seed orchards, and evaluate advanced generation selections. Standard statistical procedures are available for data analysis in progeny tests that contain parents in a balanced mating design and are replicated across time and space (Lowe et al, 1983). In operational tree improvement programs these conditions are seldom satisfied. Different mating and test designs may be used by cooperators within the same breeding zone. Parents must be evaluated that are established in different genetic tests which are located on different sites and planted in different years. Test precision and measurement age will also vary.

Data summary procedures have been developed to obtain unbiased

---

[1] Associate Geneticist, Texas Forest Service and Assistant Professor, Department of Forest Science, Texas Agricultural Experiment Station; and Head, Reforestation Department, Texas Forest Service and Professor, Department of Forest Science, Texas Agricultural Experiment Station, College Station, TX.

parental evaluations that utilize specific genetic test data as a fixed effect or predicts parental breeding values as a random effect. Hatcher et al (1981) and Cotterill et al (1983) describe various techniques to obtain unbiased estimates of parental performance as a fixed effect. The Best Linear Unbiased Prediction described by White et al (1981) and White and Hodge (1989) is a procedure to estimate breeding values from genetic test data.

The objective of this paper is to describe the procedures developed by the Western Gulf Forest Tree Improvement Program (WGFTIP) to estimate parental breeding values for loblolly pine (*Pinus taeda* L.) and slash pine (P. *elliottii* Engelm.) from genetic test data.

MODEL DEVELOPMENT

Progeny tests are routinely measured at young ages (i.e. 5 years); however, parental performance at a mature age needs to be estimated. The general formula for response to indirect selection from Falconer (1981) is:

$$R_m = i\ h_m\ h_j\ \sigma_{pm}\ r_a \qquad (1)$$

Where R_m is mature response, i is selection intensity, h_m is square root of mature heritability, h_j is square root of juvenile heritability, σ_{pm} is mature phenotypic standard deviation and r_a is the genetic correlation between the juvenile and mature trait.

Baradat (1976) defined the coefficient of genetic prediction (CGP) between two traits as:

$$CGP = Cov(A_1A_2)\ /\sigma_{p1}\sigma_{p2} \qquad (2)$$

Where $Cov(A_1A_2)$ is the additive covariance between two traits,and σ_{p1} and σ_{p2} are the phenotypic standard deviations of trait 1 and trait 2, respectively. When the phenotopic mean of trait 1 is shifted one standard deviation, the breeding value of trait 2 is shifted by the CGP multiplied by the phenotypic standard deviation of trait 2. The two traits may be measures of the same variable (i.e. volume) at different ages. In this situation, trait 1 can be a juvenile measurement and trait 2 can be a mature measurement of the same variable.

To use the CGP to calculate response to indirect selection the performance of the juvenile trait must be expressed as standard deviations. The "z" score changes the sample mean to zero and adjusts the variance to one. It is calculated as:

$$z = \frac{(x-X)}{s} \qquad (3)$$

where x is the observed trait, X is the sample mean, and s is the standard deviation.

If equation 2 is substituted into equation 1 and i is replaced by z, the formula for response to indirect selection can be shown as:

$$R_m = z \text{ CGP } \sigma_{pm} \qquad (4)$$

Equation 4 predicts the breeding value of a mature trait (R_m) in response to indirect selection on a juvenile trait when juvenile selection is expressed as a standard deviation from the sample mean and the CGP is estimated between the juvenile and mature traits. Because the z score changes the sample mean to zero, the mean for the mature trait (M_m) needs to be added to the predicted response to estimate the breeding value of the mature trait. This can be expressed as a percent as:

$$PBV_m(\%) = [(M_m + z \text{ CGP } \sigma_{pm})/M_m] * 100 \qquad (5)$$

where PBV_m (%) is predicted breeding value of the mature trait in percent and the other terms are as previously defined.

The model described in equation 5 is valid to predict the breeding value of a mature trait in response to selection on a juvenile trait if each trait is normally distributed. The distribution of most measured traits in forest trees is assumed to be normal. The use of a standard score (z) adjusts for widely differing variances among different progeny tests.

MODEL IMPLEMENTATION IN THE WGFTIP

Members of the WGFTIP have maintained progeny tests for at least 20 years. In many cases, height and diameter measurements were taken at ages 5, 10, 15 and 20 years. An initial review concentrated on all progeny tests that were at least 20 years old. Because very few slash pine tests had reached that age, a second review was made of all slash pine tests that were at least 15 years old. Progeny tests were located in South Arkansas, Louisiana and East Texas. Both open and control—pollinated progeny tests were used in the analysis. Test design varied widely among tests: replications varied from two to twelve and the number of trees per plot ranged from four to one—hundred. If the progeny test had been thinned, the removed volume was added to later measurements to obtain total volume production per acre.

To develop the information needed for equation 5, the data from 31 loblolly pine progeny tests that had reached age 20 were utilized. For slash pine, data from 14 progeny tests that were 15 years old were used. Average progeny test height was used to estimate site index for each test. Site index was used to predict mature volume and the standard deviation among families for volume at age 20 for loblolly pine. Survival or the percent infection by fusiform rust affected the slash pine equations for volume at age 15 and the standard deviation among families for volume production. CGP's were calculated across ages for each test in which significant differences (alpha = 0.1) occurred among genetic entries. Non-significant tests and tests which were not measured on the scheduled interval were not used to estimate CGP's or prediction equations.

Loblolly Pine

In 1983, data from 22 loblolly pine progeny tests that had reached age 20 were used to estimate the parameters required for equation 5. An additional nine progeny tests had reached age 20 by 1986 and the parameters were reestimated using the larger data base. Since no meaningful differences were evident between the parameter estimates for the two data sets, only the results of the last set of analyses will be presented.

Site index (base age 50) was calculated for each progeny test using 20-year mean height by the following equation:

$$\text{Log SI} = \text{Log(Mean Height)} - 5.54757\ (1/50 - 1/\text{age})$$

Estimated site indexes at age 20 ranged from 61 to 100 for the 31 progeny tests. Mean progeny tests heights at ages 5, 10 and 15 were used to develop regression equations to predict the site index for each age. Site index was used as the independent variable to predict mean plantation volume at age 20 and the standard deviation among families for volume production. Appendix 1 shows the equations developed for each measurement age.

The coefficient of genetic prediction for volume was calculated between each juvenile age (5, 10 and 15 years) and the mature age (20 years) (Table 1). Approximately 90 percent of the genetic gain in volume growth is obtained by indirect selection for volume at age 10 as compared to direct selection at age 20.

Table 1. Family coefficients of genetic prediction for 20-year volume production with loblolly pine.

Juvenile Age	Mature Age	Coef. of Genetic Pred.
5	20	0.58
10	20	0.63
15	20	0.66
20	20	0.70

Table 2 presents an example of a five-year-old loblolly pine progeny test showing both phenotypic superiorities and estimated breeding values for volume production at age 20. The mean plantation height was 4.45 meters. From the equations in Appendix 1, the site index was estimated to be 88, estimated total volume at age 20 was 16.25

m^3/ha/yr. and the phenotypic standard deviation for volume among families was 2.71 m^3/ha/yr. The volume production of each family was expressed as a standard deviation of the plantation average (z) and the coefficient of genetic prediction was 0.58.

Table 2. Estimated breeding values for volume production at age 20 and percent volume superiorities for a five-year-old loblolly pine progeny test

Parent	Percent Sup.	Breeding Value (%)	Parent	Percent Sup.	Breeding Value (%
S2PT25	59	80	LBWLob2	101	101
S2PT23	77	89	C25B	104	102
U18A	77	89	CKLOT #2	105	102
C23A	84	93		107	103
U25A	86	93	C21A	114	107
CKLOT #1	86	93	52Lob5	118	108
36Lob8	87	94	S3PT22	122	111
C18B	90	95	T261-2	131	115
S2PT24	92	96	28Lob1	137	118
K-195	94	97			

This procedure can be used for either open or control-pollinated data. For control-pollinated data, the estimated breeding value of each cross is used to determine the general combining ability of each parent in the test. This procedure offers a number of advantages in data summarization. Family ranks within a test are not changed by this procedure and adjustments are made for different site qualities and variances among tests. Because the CGP increases with age, superior growth in older tests receives greater emphasis than in younger tests. Also, reported values are based on estimated breeding values adjusted to a common age instead of phenotypic superiorities at variable measurement ages.

Slash Pine

Fourteen slash pine progeny tests, which were 15 years old, were used in 1989 to develop a similar technique for the prediction of breeding values based on volume at age 15.

Site index (base age 25) was calculated for each progeny test using the 15-year mean test height by the following equation:

$$\text{LogSI} = \text{Log(Mean Height)} - 4.767429\ (1/25 - 1/\text{age})$$

Estimated site indices at age 15 ranged from 52 to 68 for the 14 progeny tests. Mean test height at ages 5 and 10 were used to estimate the age 15 site index. Because of mortality caused by fusiform rust, site index alone was not a good predictor of mean test volume at age 15 or the standard deviation among families for volume production. Either average test survival or the percent of trees infected by fusiform rust was added to the model to predict total volume and the phenotypic standard deviation. The equations developed for each measurement age are shown in Appendix 2.

Table 3 presents the coefficients of genetic prediction used in the procedure. Because of rust associated mortality, CGP's were calculated for both rust infection and volume as juvenile selection traits with volume growth at age 15. At age 5, for tests with rust infection rates of 30 percent or greater, the amount of rust infection is a better predictor of age 15 volume than five year volume. This is because rust associated mortality has not affected rapid growing families that are severely infected by fusiform rust. At age 10, volume growth is a better predictor of age 15 volume than rust infection. By this time, mortality has started to affect family rankings for volume.

Table 3. Family coefficients of genetic prediction for 15 year volume production with slash pine.

Juvenile Age	Juvenile Trait	Mature Age	Coef. of Genet. Pred.
5	Rust Infection	15	-0.48
10	Volume	15	0.54
15	Volume	15	0.62

The formulas in Appendix 2 were used to develop the estimated breeding values for volume at age 15 for a slash pine test at age 5 (Table 4). At age 5 the test had a mean height of 3.90 meters, average infection by fusiform rust of 74 percent and 92 percent survival. Site index was estimated as 60, volume at age 15 as 10.26 m^3/ha/yr., and the standard deviation among families as 2.33 m^3/ha/yr. The percent infection by fusiform rust for each family was standardized by the z score. Family K-6 had an above average volume superiority (105 percent) but a below average breeding value because of its high fusiform rust infection rate (91 percent). By age 10, the average survival for K-6 had decreased from 85 to 55 percent because of rust associated mortality. This resulted in the volume superiority dropping to 74 percent.

Table 4. Estimated breeding values for volume production at age 15, percent volume superiority and percent rust infection at age 5 for a slash pine progeny test.

	Age 5		Age 15		Age 5		Age 15
Parent	% Sup.	Rust Infec.	Breed. Value (%)	Parent	% Sup.	Rust Infec.	Breed. Value (%)
K-141	52	97	85	01S-5	64	72	101
K-163	92	92	88	BSS-9	91	72	101
MFCS-1	68	92	88	S5PC1	106	67	104
BSS-6	80	92	89	K-211	104	64	106
K-6	104	91	89	01S-4	123	58	109
01S-3	105	87	92	K-179	105	56	111
K-202	97	86	92	BSS-10	143	50	114
K-142	78	84	94	BSS-13	125	41	120
CKLOT	111	82	95	C-103	145	41	120
OIS-1	100	78	97				

This procedure has many of the same characteristics as discussed previously for loblolly pine. However, for it to be effective, fusiform rust infection must be severe enough to detect genetic differences among families. In geographic areas where fusiform rust is not a problem, this procedure is not applicable. In the WGFTIP, rust infection must be greater or equal to 30 percent before the data is used in summarization procedures. Also, the family ranks can change because different traits are used for selection at different ages.

DATA SUMMARIZATION

The procedures previously described are utilized to obtain a breeding value estimate for each parent in a progeny test. Breeding value estimates are averaged across progeny tests to obtain estimated parental evaluation. Open and control-pollinated tests are weighted equally in the final average determined for a clone. For example, a loblolly pine selection that is in two open-pollinated progeny tests with estimated breeding values of 99 and 114 and has general combining ability estimates from three control-pollinated progeny tests of 111, 114 and 121 has an estimated breeding value of 111.8 based on five progeny tests. To be included in the final evaluation each progeny test must have significant differences (alpha = 0.1) among genetic entries. Any test not showing significant differences is deleted prior to averaging parental breeding values across all progeny test. Fusiform rust infection must be equal to or greater than 30 percent in slash pine tests to be used in the data summarization procedure.

SUMMARY

These data summary procedures have been incorporated into the Western Gulf Forest Tree Improvement Program. The equations are easily adapted into an analysis program and the data from each progeny test is expressed as estimated breeding values instead of phenotypic superiorities after analysis. The use of a standardized score adjusts for different variances among tests. Increased growth in older progeny tests receives greater emphasis because the coefficient of genetic prediction increases with age. Data summarization across tests is relatively easy in an incremental procedure. As new test results become available, parental breeding values are calculated for these tests and new averages are computed across all tests. The procedure described above can be developed for any program which has a sufficient data base to develop the required equations.

LITERATURE CITED

Cotterill, P. P., R. L. Correll and R. Boardman. 1983. Methods of estimating the average performance of families across incomplete open-pollinated progeny tests. Silvae Genet. 32(1-2):28-32.

Hatcher, A. V., F. E. Bridgwater and R. J. Weir. 1981. Performance level-standardized score for progeny test performance. Silvae Genet. 30(1-6):184-7.

Lowe, W. J., R. Stonecypher, and A. V. Hatcher. 1983. Progeny test data handling and analysis: In Progeny Testing of Forested Trees. S. Coop. Series Bull. No. 275. p. 51-67.

McKinley, C. R. 1983. Objectives of progeny tests: In Progeny Testing of Forest Trees. S. Coop. Series Bull., No. 275. p. 2-5.

White, T. L., G. R. Hodge and M. A. Delorenzo. 1987. Best linear prediction of breeding values in forest tree improvement: In Statistical Considerations in Genetic Testing of Forest Trees. S. Coop. Series Bull. No. 324. p. 99-121.

White, T. L. and G. R. Hodge. 1989. Predicting Breeding Values with Applications in Forest Tree Improvement. Kluwer Academic Publishers, Norwell, MA. 367 pp.

Appendix 1. Equations to estimate loblolly pine breeding values for volume at age 20.

A. Site Index (SI)

Age 5: $SI = 41.855 + 10.276$ (Mean Height)
$R^2 = 0.64$

Age 10: $SI = 13.962 + 6.976$ (Mean Height)
$R^2 = 0.87$

Age 15: $SI = 6.041$ (Mean Height)
$R^2 = 0.83$

Age 20: $SI = 4.813$ (Mean Height)
$R^2 = 0.99$

B. Total Volume (Age 20)

$\text{Volume} = -7.034 + 0.266\ SI$
$R^2 = 0.58$

C. Standard Deviation Among Families (Age 20)

$\text{Std. Dev.} = 0.031\ SI$
$R^2 = 0.23$

Appendix 2. Equations to estimate slash pine breeding values for volume at age 15.

A. Age 5

Site Index (SI) = 37.695 + 5.805 (Mean Height)
$R^2 = 0.81$

Volume (Age 15) = -12.780 + 0.225 SI + 0.103 (Mean Survival)
$R^2 = 0.74$

Std. Dev. Among Families (Age 15) = -6.777 + 0.151 SI
$R^2 = 0.71$

B. Age 10

Site Index (SI) = 26.886 + 3.624 (Mean Height)
$R^2 = 0.64$

Volume (Age 15) = -15.043 + 0.277 SI + 0.102 (Mean Survival)
$R^2 = 0.82$

Std. Dev. Among Families (Age 15) = -9.883 + 0.2317 SI -0.027 (Mean Infection)
$R^2 = 0.83$

C. Age 15

Site Index (SI) = 0.930 + 4.334 (Mean Height)
$R^2 = 0.99$

Volume **(Age** 15) = -17.924 + 0.337 SI + 0.108 (Mean Survival)
$R^2 = 0.88$

Std. Dev. Among Families (Age 15) = -9.8836 + 0.232 SI -0.027 (Mean Infection at Age 10)
$R^2 = 0.83$