THE USE OF BEST LINEAR PREDICTION TO OBTAIN BREEDING VALUES FOR HEIGHT AND SURVIVAL IN 37 FULL-SIB PROGENY TESTS OF SHORTLEAF PINE (PINUS ECHINATA MILL.) ON THE OUACHITA AND OZARK-ST.FRANCIS NATIONAL FORESTS

T. La Farge and J. E. Gates [1]/

U.S.D.A. Forest Service, Southern Region

Abstract.--Since 1978 the Tree Improvement Program of the Southern Region (Region 8) of the Forest Service has been progeny testing seed orchard clones of shortleaf pine. Breeding Population 1, located on the Ouachita and Ozark-St. Francis National Forests in Arkansas and Oklahoma, comprises 25 6 X 6 diallel crossing groups totalling 150 parents.

After 12 years of progeny test establishment, there is a great deal of imbalance in the data. Each year full-sib progeny tests were established at as many as five progeny test locations, but not all families could be included at each location, and most diallel crossing groups were less than complete. There were also several test failures due to summer droughts. Hence, the prospects of obtaining balanced ANOVAs and unbiased estimates of breeding values were formidable until the recent availability or Best Linear prediction (BLP) as a method of data analysis.

This paper reports predicted breeding values for height and survival of shortleaf pine in the National Forests in Arkansas and Oklahoma. It discusses advantages and disadvantages of BLP and some limitations encountered in applying this method by means of the Statistical Analysis System for Personal Computers (PC SAS).

Keywords: Pinus echinata Mill., Best Linear Prediction, full-sib, progeny test.

INTRODUCTION

In 1978 the Tree Improvement Program of the Southern Region (Region 8) of the Forest Service began establishing progeny tests of seed orchard clones in five species of the southern pines. Shortleaf pine (Pinus echinata Mill.) is an important species in the National Forest System, especially in the Piedmont and Mountain provinces. In the Region 8 Tree Improvement Program, Breeding Population 1, representing the Ouachita and Ozark-St. Francis National Forests in Arkansas and Oklahoma, is the largest of six breeding populations of

---

[1]/ Eastern Zone Geneticist, U.S.D.A. Forest Service, Region 8, Atlanta, Georgia, and Western Zone Geneticist, U.S.D.A. Forest Service, Region 8, Alexandria, Louisiana.

shortleaf pine. The mating scheme comprised full-sib crosses in 25 6 X 6 diallel crossing groups totalling 150 parents. All clones of this breeding population are located in the Ouachita Seed Orchard near Hot Springs, Arkansas.

Because of logistic problems in controlled breeding and several test failures resulting from summer droughts, there is a great deal of imbalance in the data available after 13 years of progeny test establishment. Each year full-sib progeny tests were established at *as* many as five progeny test locations, but not all families could be included at each location, and most diallel crossing groups were less than complete. Hence, the prospect of obtaining balanced ANOVAs and unbiased estimates of breeding values were formidable until the recent availability of Best Linear Prediction (BLP) as a method of data analysis (White et al. 1986; White and Hodge 1989).

Previously La Farge (1989) compared two methods of applying BLP. One method was designed for predicting breeding values for different target environments when genotype x environment interactions were involved or suspected. In the present case the simpler model is assumed, which is that there is only one breeding population, or zone, and there are no significant genotype x environment interactions.

METHODS

Progeny Test Design

The field designs of all progeny tests conformed to the standard randomized complete-block design used in the Region 8 Tree Improvement Program. In this design the tests are laid out in 10-tree row plots so that all plots within a block followed the same contour. Most tests contained five replicates, but three had only four. The 37 tests comprising this analysis were established in the eight years from 1978 through 1985. One test was established in 1978 and one in 1979. Five tests were planted in 1981, six in 1982, five in 1983, eleven in 1984, and eight in 1985.

The maximum number of tests in which any one crossing group was replicate(in any one year was three, but in some years groups were included in only one or two tests. After 1978, all crossing groups were replicated in two years. This is the only breeding population in the Region 8 Program in which replication by years was attempted. However, in the present analysis year of establishment was not included as an effect.

Breeding Population 1 of shortleaf pine is subdivided into 25 crossing groups, of which 21 have been tested and measured so far. Each crossing group consists of six parents which are intercrossed to form a 6 x 6 element diallel. This design is a disconnected partial diallel, and all crossing groups are unrelated. Although the potential number of crosses in a 6 x 6 element diallel is 15, the largest number of crosses in any single crossing group in the current test is 14.

Site quality varied greatly, resulting in large differences in growth rate and survival among sites. Average test height at age 5 years ranged from 5.2 to 13.2 feet, and average percentage survival varied from 69.1 to 99.9.

The BLP Equation

All progeny tests included in this analysis were full-sib tests, and both the theory and methods of analysis used in the present investigation are described by White and Hodge (1989). A primary utility of BLP is that it can be used to predict parental genotypes in full-sib progeny tests.

A principal assumption of BLP is that the first and second moments are known (Henderson 1984, White and Hodge 1989). The second moments are specified in the C and V matrices. The C matrix, which is nonsymmetric, defines the genetic relationships between the observed full-sib family means at each site and the true yet unknown breeding values, g. Each column of C represents a parental breeding value to be predicted, and the elements comprising C are estimated from genetic theory.

The V matrix, which is symmetric, represents the variances and covariances between the observed phenotypic values. The main diagonal consists of variances of family means for each planting location. In full-sib progeny tests the covariances in the off-diagonals not equal to zero are covariances between family means which refer to either: (1) different tests with two common parents; (2) different tests with one common parent; or (3) the same test with one common parent (White and Hodge 1989). When these matrices have been constructed, the following formula can be used to predict the breeding values:

$$\hat{g} - C'V^{-1}y \qquad (1)$$

where C and V are defined above, y - a vector of data representing observed deviations of the family means at each test location from the test location mean, and $\hat{g}$ - the breeding values to be predicted.

Procedures Used

All tests were measured at five years of age. The traits measured were height in feet and survival. The survival percentages were analyzed as the arcsin X the square root of percentage expressed in radians. In the Region 8 Program, diameter and straightness are not measured until the tenth year.

Space does not allow a complete discussion of the theory or procedures involved in BLP, but discussions of the theory and construction of C and V matrices are given by White et al. (1986) and White and Hodge (1989), and examples of operational C and V matrices are provided by La Farge (1989).

To generate the second moments needed to construct the C and V matrices, it is necessary to perform a combined ANOVA on the data. This was accomplished by means of the VARCOMP Procedure of the Statistical Analysis System for Personal Computers and and the MIVQUEO Method (SAS 1987). The variance components generated by Proc VARCOMP were then combined in appropriate equations to produce the second moments needed to construct the C and V matrices. The equations that are needed for this conversion are provided by White and Hodge (1989). The second moments were loaded into the C and V matrices by means of language provided in the SAS IML Guide for Personal Computers (1985).

The process of attempting to analyze such a large number of progeny tests with Proc VARCOMP presents data management problems on PC SAS. Proc VARCOMP generates an error reading when too much data are entered. Usually the largest data set that it will allow includes data from only five tests. Hence analysis of the data from 37 tests on PC SAS is out of the question. One solution is to divide the data into 5-test subsets and then obtain an average value for each variance component estimated (G. R. Hodge, personal communication).

The averaging of variance components was the method used. Although this method appears to have no theoretical basis, it seems to work. However, in the process of averaging variance component estimates, another problem arose. Which data sets should be averaged to make these estimates? In other Region 8 breeding populations, as many as four or five progeny tests are planted each year, and these tests comprise the same set of one or two crossing groups, including up to five check lots. However, as mentioned above, the tests in Shortleaf Breeding Population 1 were replicated by years as well as by sites or by locations. When these tests are grouped only by common years, many crosses and crossing groups are not common among tests. Such unmatched groupings tend to produce weak variance compents.

To compensate for this source of error, new groupings were assembled based on their common crossing groups. Six such groupings were found and were averaged to produce a set of variance components for each trait. One grouping was discarded for survival because the survival in the three tests comprising that grouping was so high that the variation was insufficient to produce strong variance components. The total subset of tests from which variance components were derived for height was 24; for survival the total was 21 tests. These variance components are listed in Table 1.

RESULTS

Breeding values for height and survival were predicted for a total of 123 parents, three of which were check lots common to most progeny tests. The breeding values of the best and worst of these parents and checks are given in Table 2 for height and survival. The breeding values for survival have been reconverted from radians to percentages.

Also included in Table 2 are the estimated correlations between the true and predicted genetic values (CRGG), which are obtained by means of the following equation:

$$\text{CRGG} = \text{Corr}(\hat{g}, g) = (\text{Var}(\hat{g})/\text{Var}(g))^{1/2} \qquad (2)$$

(White and Hodge 1989), where Var(g) is the variance of the predicted breeding values and Var(g) is the variance of the true but unknown breeding values. Since best Linear Prediction provides no direct method of detecting significances of differences among treatment means, the estimation of a CRGG value for each breeding value is a useful if imperfect measure of the strength of the breeding value. Although there is no predetermined level of acceptability for values of CRGG, the closer each value is to unity, the better. Our experience with many breeding values suggest that they should not be less than 0.5.

Table 1. Variance component estimates based on averages of six test subsets for height and five test subsets for survival in 37 shortleaf pine progeny tests on the Ouachita and Ozark National Forests.

Source of variation	Variance component for: Height	Survival
Var_F 1/	0.057200	0.0021490
Var_M	0.056577	0.0017991
Var_{FS}	0.013258	0.0022249
Var_{MS}	0.026943	0.0009755
Var_{FM}	0.045690	0.0016402
Var_{FMS}	0.017522	0.0056500
$Var_{FB(S)}$	0.012493	0.0012195
$Var_{MB(S)}$	0.003973	0.0008943
$Var_{FMB(S)}$	0.235810	0.0202050
Var_e	2.339910	0.3173500

1/ Var_F, Var_M, Var_{FS}, Var_{MS}, Var_{FM}, Var_{FMS}, $Var_{FB(S)}$, $Var_{MB(S)}$, $Var_{FMB(S)}$, and Var_e are, respectively, variances due to female, male, female X site, male X site, female X male, female X male X site, female X block-within-site, male X block-within-site, female X male X block-within-site, and tree-within-plot.

It is obvious from Table 2 that the breeding values predicted for height are stronger than those predicted for survival. The range of variation among predicted values for height is almost two feet, that for survival 10.9 percent. Further, the values for the CRGGs are generally much higher for height, although all values for survival are 0.57 or higher. Individual and family heritabilities, given in Table 3, also show the relative strengths of these statistics, as do the realized genetic gains in Table 4. The realized gains are calculated on the basis of the General Forest Area (G.F.A.) Mix. In the Region 8 Program G.F.A. check lots are usually represented by a sample of stand collections within a breeding zone.

Table 2. Rank comparisons of the best and worst shortleaf pine parents for height and survival in 37 progeny tests at age five on the Ouachita and Ozark National Forests.

Parental I.D.	Crossing group	Height	Height rank	CRGG[1] for height	Survival	Survival rank	CRGG for survival
		Feet			Percentage		
			BEST HEIGHT				
243	14	8.98	1	0.785	90.0	113	0.635
127	11	8.84	2	.774	91.8	89	.619
136	17	8.80	3	.819	94.1	37	.682
230	18	8.74	4	.784	93.6	44	.634
			BEST SURVIVAL				
332	18	8.15	48	.790	97.7	1	.638
104	9	8.41	18	.869	96.7	2	.753
142	7	8.24	39	.660	96.3	3	.660
150	22	8.16	47	.767	96.2	4	.603
			POOREST HEIGHT				
135	14	7.19	123	.875	88.2	122	.761
343	10	7.30	122	.767	91.6	91	.604
201	17	7.38	121	.873	90.3	108	.760
226	18	7.39	120	.829	91.6	95	.685
			POOREST SURVIVAL				
137	6	7.44	116	.798	86.8	123	.654
135	14	7.19	123	.875	88.2	122	.761
206	21	8.01	71	.760	88.5	121	.571
244	13	7.90	88	.758	88.7	120	.571
			CHECK LOTS				
S.P.A. Ozark[2]	51	7.62	108	.969	91.5	97	.920
G.F.A. Mix[3]	54	7.51	113	.977	89.3	116	.922
S.P.A. Ouachita	51	7.44	115	.974	93.1	62	.914

---

[1] The estimated correlations between true and predicted genetic values.

[2] Open-pollinated seed from a seed-production area.

[3] A mixture of seed collected from the general forest area.

DISCUSSION AND CONCLUSIONS

While the range of breeding values for each trait is not large, the differences for height seem to be great enough to achieve satisfactory gains. The percentage gain in height growth is in close agreement with those reported for other species with data combined from many tests, e.g. Kraus and La Farge (1982). The corresponding statistics for survival might be stronger if most of the progeny tests had had poorer survival than they did. Fortunately the poor survival in some tests is probably attributable to certain years with very dry summers. Because of the uncertainty of predicting years of severe summer drought or sites that may be prone to drought, selection for high survival or drought resistance probably has little practical value.

Table 3. Individual and family heritabilities for height and survival in 37 full-sib progeny tests of shortleaf pine on the Ouachita and Ozark National Forests.

	Heritability, h^2	
	Individual	Family
Height	0.08	0.96
Survival	0.02	0.53

Table 4. Realized genetic gains for height and survival in an unrogued shortleaf pine seed orchard based on 37 full-sib progeny tests in the Ouachita and Ozark National Forests.

Trait	Realized genetic gain
	Percentage
Height	7.52
Survival	3.96

The breeding values generated from these analyses actually have greater utility than merely for making backward selections and roguing the poorest parents. Although some roguing will be done, their primary purpose will be for use in making selections of the best trees in those families with the highest parental breeding values by means of the method discussed by Hodge et al. (1989). Also, since some of the 37 tests have reached age 10, diameter and straightness measurements will permit selection indexes for more traits to be utilized in the BLP procedures.

The primary utility of these methods is that they allow the combined analyses of data from many progeny tests over a period of time. Within limits BLP tolerates messy data, but poorly estimated breeding values occasionally need to be discarded. The CRGG values provide some guidance as to which breeding values are poorly estimated, and inspection of the data will usually indicate the reason for the low CRGG values (e.g., inadequate replication).

LITERATURE CITED

Henderson, C. R. 1984. Applications of linear models in animal breeding. University of Guelph, Guelph, Ontario, Canada. 462 p.

Hodge, G. R., G. L. Powell and T. L. White. 1989. Establishment of the second-generation selected population of slash pine in the Cooperative Forest Genetics Research Program. P. 68-74 in Proc. 20th Southern For. Tree Imp. Conf. Charleston, SC.

La Farge, T. 1989. Applications of best linear prediction to the analysis of five full-sib loblolly pine progeny tests. P. 315-322 in Proc. 20th Southern For. Tree Imp. Conf. Charleston, SC.

SAS Institute. 1985. SAS/IML guide for personal computers, version 6 edition. SAS Institute, Inc., Cary, North Carolina. 244 p.

SAS Institute. 1987. SAS/STAT guide for personal computers, version 6 edition. SAS Institute, Inc., Cary, North Carolina. 1028 p.

White, T. L., G. R. Hodge and M. A. Delorenzo. 1986. Best linear prediction of breeding values in forest tree improvement. P. 99-122 in Statistical Considerations in Genetic testing of Forest Trees, Proc. 1986 Workshop Southern Regional Information Exchange Group 40. Southern Coop. Series Bull. 324. Univ. of Florida, Gainesville, FL.

White, T. L. and G. R. Hodge. 1989. Predicting breeding values with applications in forest tree improvement. Kluwer Academic Publishers, Dordrecht, The Netherlands. 367 p.