

Best linear prediction of breeding values in a forest tree improvement program

T. L. White and G. R. Hodge

Department of Forestry, University of Florida, Gainesville, FL 32611, USA

Received January 15, 1988; Accepted March 30, 1988

Communicated by L. D. Van Vleck

Summary. Best Linear Prediction (BLP) was used to predict breeding values for 1,396 parents from progeny test data in an operational slash pine breeding program. BLP rankings of parents were compared to rankings of averaged standard scores, a common approach in forestry. Using BLP rankings, selection of higher ranking parents tends to choose parents in a larger number of more precise progeny tests. The trend is the opposite with standard scores; higher ranking parents tend to be those in fewer, less precise tests. BLP and a related methodology, Best Linear Unbiased Prediction (BLUP), were developed by dairy cattle breeders and have not been used widely outside of animal breeding for predicting breeding values from messy progeny test data. Application of either of these techniques usually requires simplifying assumptions to keep the problem computationally tractable. The more appropriate technique for a given application depends upon which set of assumptions are better for the given problem. An assumption of homogeneous genetic and error variances and covariances, generally made by animal breeders when applying BLUP, was inappropriate for our data. We employed an approach that treated fixed effects as known and treated the same trait measured in different environments as different traits with heterogeneous variance structures. As tree improvement programs become more complex, the ease with which BLP and BLUP handle messy data and incorporate diverse sources of information should make these techniques appealing to forest tree breeders.

Key words: Progeny testing – Selection index – Slash pine – Random models

Introduction

As in breeding programs of many species, most forest tree improvement programs around the world employ progeny tests to estimate breeding values of selections made each generation. Offspring from the selected parents are planted in randomized, replicated tests that are usually established in a number of different years and field site locations. Parents whose progeny perform better, on the average across all tests, are considered genetically superior and the parental rankings are used in numerous ways to enhance genetic progress.

Because the parental rankings from these progeny tests are critical to genetic progress, the predicted breeding values must be precise and accurate. However, often in forestry (Cotterill et al. 1983; Lowe et al. 1983; White et al. 1986), the analyses are complicated, because the data are messy¹: (1) often, only a subset of the parents are represented in a given test; (2) parents are represented in different numbers of tests; (3) tests are measured at different ages; and (4) tests vary dramatically in the level of precision.

While many analytical approaches have been used in forestry to develop parental rankings from these types of messy data, most (if not all) have treated breeding values as fixed effects to be estimated from the data (class I model of Eisenhart 1947) rather than as random effects. The data from a given test are often transformed in one or two ways prior to combining data across tests (Hatcher et al. 1981; Lowe et al. 1983; White et al. 1986). Commonly, an estimated environmental mean is subtracted from the data in an attempt to adjust for field location and year effects that differ from test to test and

¹ The term “messy” derives from “Analysis of Messy Data” (Milliken and Johnson 1984)

the data are then standardized in an attempt to achieve similar or equal variances among family means in all tests. These standard scores for a given parent are then averaged over all tests in which that parent is represented (Hatcher et al. 1981; Cotterill et al. 1983; Lowe et al. 1983). This approach results in all tests and measurement ages having the same relative weight in determining parental rankings and the variance among averaged scores will be smaller for parents represented in more progeny tests (White et al. 1986).

Dairy cattle breeders, also faced with messy progeny test data, developed analytical approaches that consider breeding values as random effects to be predicted rather than fixed effects to be estimated (Henderson 1963, 1973, 1977, 1984). These analytical methods, Best Linear Prediction (BLP) and Best Linear Unbiased Prediction (BLUP), have many desirable properties in some situations. However, to our knowledge these methods have not been widely used outside of animal breeding and most applications within animal breeding have assumed homogeneous genetic and error variances (Hill 1984; Garrick and Van Vleck 1987).

This paper describes the application of BLP for predicting parental breeding values in an operational slash pine (*Pinus elliottii* var. *elliottii*) improvement program in the southeastern United States. BLP is the selection index first used for plants and animals by Smith (1936) and Hazel (1943), respectively. However, as commonly applied, selection index assumes equal amounts and quality of information for each parent and a single vector of weights are developed and used all parents. BLP, as applied in this paper, develops a different vector of coefficients for each parent, specifically to reflect the nature of the progeny test data available for that parent.

The specific objectives of this research were to: (1) estimate the variances and covariances needed to implement BLP; (2) develop best linear predictions of breeding values for two important traits (volume growth and resistance to fusiform rust, *Cronartium quercum* f. sp. *fusiforme*) for 1,396 slash pine parents in 364 wind-pollinated progeny tests; and (3) compare the BLP breeding value predictions to those developed using an analytical method (averaged standard scores) currently common in forest tree improvement.

Materials and methods

Progeny test data

The Cooperative Forest Genetics Program at the University of Florida directs a cooperative breeding program for 15 private corporations and government agencies that are improving slash pine for approximately 5 million hectares of timberlands in the lower coastal plain in Florida, Georgia and Alabama. Mass selection of superior trees growing in natural stands began in the 1950's with emphasis on volume growth, tree form and freedom

from disease (Goddard 1980). After each cooperator made 50–200 selections, scions from these genotypes were grafted into seed orchards to produce improved seed for operational reforestation.

Wind-pollinated seed from seed orchards was used to establish 364 progeny tests from 1960 to 1980. Each test is a randomized complete block design with 5–10 trees (average = 8) from each wind-pollinated family represented in a row-plot within each block. The number of parents represented in a given test ranges from 5–160 (average = 33) and the number of blocks from 3–10 (average = 6). The number of trees planted in each test ranges from 500–2,400 (average = 1,600) and survival averaged 78% at 15 years.

At measurement ages of 5, 10 or 15 years, each tree was scored for the presence (score = 1) or absence (score = 0) of fusiform rust and measured for height and diameter at breast height (DBH). The height and DBH measurements were converted into an estimated volume for each tree (Goddard and Strickland 1968). A linear model for a measured trait at a given age is:

$$x_{ijkl} = \mu + E_i + B_{ij} + f_k + fe_{ik} + p_{ijk} + e_{ijkl} \quad (1)$$

where

μ = a general mean,

E_i = fixed effect of i^{th} test environment, $i = 1, 2, \dots, t$,

B_{ij} = fixed effect of j^{th} block in i^{th} test, $j = 1, 2, \dots, b_i$,

f_k = random effect of k^{th} family, $k = 1, 2, \dots, p_i$, $E(f_k) = 0$,
 $\text{Var}(f_k) = \sigma_f^2$,

fe_{ik} = random interaction of k^{th} family and i^{th} test, $E(fe_{ik}) = 0$,
 $\text{Var}(fe_{ik}) = \sigma_{fe}^2$,

p_{ijk} = random plot error of k^{th} family in j^{th} block of i^{th} test,
 $E(p_{ijk}) = 0$, $\text{Var}(p_{ijk}) = \sigma_p^2$, and

e_{ijkl} = tree error of l^{th} tree in ijk^{th} plot, $E(e_{ijkl}) = 0$, $\text{Var}(e_{ijkl}) = \sigma_e^2$.

The covariances between all effects are assumed to be zero. E_i and B_{ij} are treated as fixed effects since the purpose here is to compare and rank families and it is necessary to adjust family means for the specific fixed set of blocks and test environments in which they occur (see Henderson 1973). Wind-pollinated families are considered half-sib families.

Computation of best linear predictions

We wish to use y , the $n \times 1$ vector of observed data records from wind-pollinated progeny tests to predict g , a non-observable $2p \times 1$ vector of two breeding values (volume growth and rust resistance) for each of p parents. The derivation of BLP does not require that the exact form of the joint distribution between y and g be known, but assumes we know the following first and second moments (Henderson 1963, 1973, 1977, 1984):

$E(y) = \alpha$, the expected value of the $n \times 1$ vector of data records,

$\text{Cov}(y, g') = C$, an $n \times 2p$ matrix of genetic covariances between the observed data and the breeding values being predicted, and

$\text{Var}(y) = V$, an $n \times n$ matrix of variances and covariances among the observations.

If these parameters are known and only linear functions of the observed data records are considered, then the solution that minimizes error variance of predictions, $E(\hat{g}_i - g_i)^2$, among the class of all linear functions is:

$$\hat{g} = C' V^{-1} (y - \alpha). \quad (2)$$

Because C is the covariance between g and y and V is the variance of y , $C' V^{-1}$ is, heuristically, $\text{cov}(g, y)$ "divided by"

var(y) and can be thought of as multiple regression coefficients relating y to g (Thompson 1979). A different set of coefficients is developed for each parent depending upon the exact ages, precisions and number of tests that the parent's offspring are planted in. The statistical properties of these predictions are well-documented, assuming that the first and second moments are known parameters (Henderson 1973, 1977, 1984); in practice, estimates of these parameters are used in the computations.

In our situation, parents whose breeding values are being predicted are assumed unrelated and observed progeny test data from a parent contribute information about only that parent (since the progeny are considered a maternal half-sib family). Therefore, the V matrix is block diagonal and the C matrix contains non-zero elements only when progeny test data are from the corresponding parent being predicted (White et al. 1986). Thus, observations on family k do not contribute to breeding value predictions of family k' and breeding values were computed individually for each of the 1,396 parents from a program written in SAS Proc Matrix (SAS Institute 1982) as follows:

$$\hat{g}_k = C'_k V_k^{-1} (y_k - \alpha_k) \quad (3)$$

where

$\hat{g}_k =$ a 2×1 vector of breeding value predictions for two traits (rust and volume) on the k^{th} parent, and C'_k, V_k, y_k are the portions of the respective matrices or vectors from Eq. (2) that correspond to the k^{th} parent.

The estimated variance among predicted breeding values was calculated for each of the two predicted breeding values for each parent as (Henderson 1973, 1984):

$$\text{Var}(\hat{g})_k = C'_k V_k^{-1} C_k \quad (4)$$

Defining the g and y vectors

For the g_k vectors, we chose to predict parental breeding values for rust resistance at 5 years of age (BLPRUST) and for tree volume at 15 years (BLPVOL). Breeding values for both traits are predicted in the units of measurement (untransformed) and are predicted as deviations from an unimproved population mean of zero [that is, $E(g) = 0$].

The data records in the y_k vectors could have been individual tree measurements, plot means (the average of all trees from a given family within a block) or family means (the mean of all plot means of a given family within a test). We chose to use family means as records for the y vector because: (1) for many of the 364 progeny tests, this was the only information in machine readable form; (2) little information is lost, since in our tests survival was generally high and the number of missing plots low; (3) rust is scored as a Bernoulli trait and family means based on approximately 50 trees will be approximately normally distributed; and (4) computations are much more tractable (13,630 records representing family means for two traits from 364 tests instead of approximately 1 million individual tree records).

Thus, the assumed linear model for each element in a y_k vector is:

$$y_{ik} = \bar{x}_{i.k} = \mu + E_i + f_k + fe_{ik} + \frac{P_{i.k}}{b_i} + \frac{e_{i.k}}{m_{ik}} \quad (5)$$

where the dot notation indicates summation over that subscript, y_{ik} is the family mean for the k^{th} family in the i^{th} test, b_i is the number of blocks, m_{ik} is the number of trees in the k^{th} family, all other terms are defined in Eq. (1), and block effects sum to zero.

The n_k family means in each y_k vector are ordered with the first two records being family means for volume and rust in the

first progeny test. If a progeny test was measured at more than one age (5, 10 or 15 years), only the oldest data were used.

Estimating the fixed effects, α_k

The fixed effects are assumed known in BLP and in our case are associated with the effects of test environments. These are used to express the n_k family means, y_k , as deviations from the test means, α_k , in Eq. (3). An ordinary least squares estimate of the test environment average for each progeny test was obtained by averaging across all family means and the means of the one to four extra wind-pollinated families (called genetic check lots) that were standard across most progeny tests (White et al. 1986).

Defining the C_k and V_k matrices

Each element of an $n_k \times 2$ C_k matrix is a covariance between an observation y_{ik}^a , and an unobserved breeding value, g_k^b , where the superscripts indicate that these may be for the same ($a = b$) or different traits. Because the breeding values are unobservable, the covariances must be calculated by a combination of statistical estimation and genetic theory. First, the linear model from Eq. (5) is substituted for y_{ik}^a and since all non-genetic terms are uncorrelated:

$$\text{cov}(y_{ik}^a, g_k^b) = \text{cov}(f_k^a, g_k^b).$$

Then for a genetic model that assumes that wind-pollinated families are maternal half-sib families, the family effects represent half the breeding value of the common parent, $f_k^a = \frac{1}{2} g_k^a$ and:

$$\text{cov}(y_{ik}^a, g_k^b) = \frac{1}{2} \text{cov}(g_k^a, g_k^b).$$

When $a = b$, then the covariance is one-half of the additive genetic variance of the trait and when $a \neq b$, then the covariance is one-half the additive covariance between traits a and b .

The $n_k \times n_k$ V_k matrix for each parent contains variances and covariances among its observed family means. There are three distinct types of elements (see White et al. 1986 for details):

(1) The diagonal elements are variances of family means for the six different measured traits (volume and rust at three ages);

(2) One type of off-diagonal element is the covariance of family means between rust and volume measured on the same trees in the same test (this corresponds to a type A covariance of Burdon 1977); and

(3) Other off-diagonal elements are covariances of family means for the same or different traits measured in different tests (type B covariances).

All three types can be expressed in terms of the linear model in Eq. (5) (White et al. 1986). Because only genetic effects are common between family means measured in different tests, the type B covariances are identical to those needed for C_k matrices.

Estimating elements of the C_k and V_k matrices

Whereas the fixed effects for both rust and volume were estimated for each of the 364 tests, the variances and covariances for the C_k and V_k matrices were estimated from a subset of 28 of the tests. All 28 had measurements at ages 10 and 15 years, and 18 had measurements at 5 years. These estimates were then applied to all other tests in the three steps described below. First, to estimate variance components, each of the 28 tests was subjected to an unweighted analysis of plot means for both rust and growth at each age (Snedecor and Cochran 1967, p 475). Estimates of within-plot variances were calculated from the individual tree measurements and divided by the harmonic mean to put them on a plot mean basis. SAS Type I observed mean squares (Freund and Littell 1981) were then equated to their expected values and the resulting system of equations solved for the estimated variance components. This is analogous to Hen-

derson's Method III (Milliken and Johnson 1984, p 231). Preliminary analyses indicated that for tests with rust incidences between 10% and 90%, arcsin transformations of the rust plot means did not materially affect the relative values of variance components for rust (see Rockwood and Goddard 1973; Sohn and Goddard 1979 for similar results) and no transformations were made.

Second, estimates of the variances of family means for each trait-age combination were then obtained for each of the 28 tests by assuming that all tests had 6 blocks and 8 trees per family in each block [$b_i=6$ and $m_{ik}=48$ in Eq. (5)]. Most tests do not deviate much from these average values and this assumption of constant test design greatly simplified estimation. Thus,

$$\text{Var}(y_{ik}) = \text{Var}(\bar{x}_{i,k}) = \hat{\sigma}_F^2 + \hat{\sigma}_p^2/6 + \hat{\sigma}_e^2/48 \quad (6)$$

where terms are defined in (1) and (5) except $\sigma_F^2 = \text{var}(f_k + fe_{ik})$.

Type A covariances of family means for the \mathbf{V} matrix were estimated similarly by conducting analyses of covariance between rust and volume for each of the 84 age-test combinations. SAS Type I cross products from multivariate analyses of variance were equated to their expected values to estimate covariance components. To estimate Type B covariances of family means, tests that contained common families were paired and covariances among family means obtained for all possible pairs of traits and ages.

Finally, it was apparent from initial analyses that variances and covariances varied dramatically both from age to age and across the 28 tests at a given age. Thus, the assumption usually made in animal breeding (Henderson 1984; Hill 1984; Garrick and Van Vleck 1987) of homogeneous second moments across all tests was not appropriate, and we developed linear regression equations from these 28 tests to predict the second moments as functions of easily measured independent variables (the regressors in Table 1) that would be available for all 364 tests. Then for each test and pair of tests, the elements needed for the \mathbf{C}_k and \mathbf{V}_k matrices could be predicted from the measured independent variables.

Many different transformations and re-expressions of these regressors were tried in an attempt to linearize relationships and homogenize variances (Mosteller and Tukey 1977). Both unweighted and weighted (by the number of families in each test) regressions were used depending upon whether the number of families was inversely proportional to the variances of the observations. Screening was done with all-possible-combinations regression and then the most promising models (high R^2 and biologically interpretable) were subjected to further testing. To be adopted as a final predictive equation, a model had to be biologically interpretable, be significant at $P=0.01$, and have each regressor in the model significant at $P=0.05$. After detailed testing of each model over the ranges of the regressors, limits on either predictions or regressors were sometimes specified.

Before modeling, both type A and type B covariances were transformed to family mean correlations by dividing by the appropriate standard deviations of family means and models were fit to the correlations. The modeling of correlations was felt more appropriate because they are bounded by -1 and 1 . The predicted type A and B correlations for each pair of tests were then multiplied by the square roots of predicted family variances for those tests to obtain predicted covariances.

Target planting environments were specified for prediction of breeding values because the genetic expression of both rust and growth varies with the test environment (Table 2). These specifications of target environments were used to calculate elements of the \mathbf{C}_k matrices that require estimates of the covariance between the breeding value in the specified target environment and the observed family means in their test environments (White et al. 1986). For 15 year volume, the target environment

Table 1. Definitions and abbreviations of independent variables used in building regression models

Abbreviation	Definition
R	mean rust infection level of a given test, $0 \leq R \leq 1$
RDIF	absolute value of the difference in rust between two tests, $= \text{abs}(R_i - R_j)$
RIMS	block by family interaction mean square from an analysis of variance of plot means for rust in a given test
RSUM	sum of mean rust infection levels for a pair of tests, $= (R_i + R_j)$
V	mean volume of a test, dm^3
VCV	coefficient of variation of a test for volume, $= (\text{sqrt}(VIMS))/V$
VCVSUM	sum of volume coefficients of variation for a pair of tests, $= VCV_i + VCV_j$
VDIF	absolute difference of standardized volumes for two tests, $= \text{abs}(VSTAN_i - VSTAN_j)$
VIMS	block by family interaction from an analysis of variance for volume of a test
VSTAN	mean test volume expressed as a standardized deviate from the overall mean for that age; so for each age, the mean volumes from the 28 tests were standardized to have mean = 0 and variance = 1

for predicting breeding values was chosen to be that of average test volume and CV, while for rust the target environment was chosen to be that which would have an average rust infection level of 50% if unimproved seedlings were planted.

Results and discussion

Estimation of second moments

Regression models developed from the subset of 28 test show that specific estimates of the second moments vary substantially for different tests and pairs of tests (Table 2). This approach treats a single trait such as volume growth at 5 years old as multiple traits when expressed in multiple test environments (Burdon 1977; Gianola 1986).

For volume growth, variance of family means was greatly influenced by test age and to lesser extents by the average volume of the test within an age class and its coefficient of variation (Model 1, Table 2). Thus, the estimated variances of family means for volume to use on the diagonals of \mathbf{V}_k matrices vary by as much as 50,000 fold, depending upon the age and characteristics of the tests in which the family means are observed. As is common in the biological sciences (Snedecor and Cochran 1967), variances for volume increased dramatically as the age and therefore mean volume of the test increased (Table 2). Coefficients of variation for volume growth were similar across age classes (0.26, 0.23 and 0.19 for 5, 10 and 15

Table 2. Back-transformed regression models and their fit statistics for estimating the second moments needed for the C_k and V_k matrices; all models are significant at $P=0.001$ and the reported R^2 values are for the final back-transformed model accounting for the imposed limits; abbreviations for regressor variables are defined in Table 1

Model no. and dependent variable ^a	Age	Mean	Regression model	Imposed limits	R^2	n^b
1 Variance of family, means for volume (dm^3)	all				0.78	76
	5	1	$3.17 [10 \exp(80.53 + 2.89(\text{VCV}) - 83.04(\text{V exp} - 0.01))]$	$0.02 \leq \hat{y} \leq 4.00$		
	10	85	$3.17 [10 \exp(80.90 + 1.40(\text{VCV}) - 83.04(\text{V exp} - 0.01))]$	$2.41 \leq \hat{y} \leq 280$		
	15	600	$3.17 [10 \exp(81.14 + 0.93(\text{VCV}) - 83.04(\text{V exp} - 0.01))]$	$32.1 \leq \hat{y} \leq 1350$		
2 Variance of family, means for rust $[(\%/100)^2]$	all	0.015	$0.00355 + 0.06517(\text{R}) - 0.06296(\text{R}^2) - 0.4750(\text{RIMS}) + 1.9715(\text{R})(\text{RIMS}) - 1.9048(\text{R}^2)(\text{RIMS})$	$0.1 \leq \text{R} \leq 0.9$ $0.001 \leq \hat{y} \leq 0.025$	0.56	78
3 Type A correlation, volume and rust	5	0.06	No significant model found	NA	NA	18
	10	0.13				28
	15	0.14				28
4 Type B correlation, rust in two tests	all	0.50	$0.2471 + 0.6414(\text{R SUM}) - 0.2680(\text{R SUM}^2) - 0.5006(\text{RDIF})$	$0.1 \leq \text{R} \leq 0.9$	0.25	316
5 Type B correlation, volume in two tests	5: all	0.15	$0.7288 - 1.029(\text{VCV SUM}) - 0.0931(\text{VDIF})$	$0.00 \leq \hat{y}$	0.20	115
6 Type B correlation, volume in two tests	10:10	0.34	$0.8213 - 0.6924(\text{VCV SUM}) - 0.2110(\text{VDIF})$	$0.00 \leq \hat{y}$	0.24	171
	10:15	0.40	$0.8547 - 0.6924(\text{VCV SUM}) - 0.2110(\text{VDIF})$			
	15:15	0.44	$0.8700 - 0.6924(\text{VCV SUM}) - 0.2110(\text{VDIF})$			

^a Models 1 and 2 were used to estimate variances for use on the diagonal of the V_k matrices and, in conjunction with correlation estimates, for estimating covariances

^b The number of observations, n , is less than 84 (28 tests \times 3 ages) in models 1 and 2 because of missing measurements at age 5

years, respectively), but varied dramatically within age, classes (especially within the 5 year age class where CVs ranged from 0.13–0.50). The models predict that more variable tests (higher CVs) within an age class have higher variances of family means and this effect is most pronounced in the 5 year age class. In general, tests with higher CVs are those that are more variable due to higher environmental variances (due to a heterogeneous field site or other experimental factors). If the effect of higher variances in more variable tests is ignored (i.e., if a constant variance of family means for volume growth is applied to all tests regardless of CV), there will tend to be a larger variance among breeding values predicted for parents in variable tests; this can lead to errors of selection due to the choice of higher ranking parents represented in tests with large environmental variances (Hill 1984; White et al. 1986).

For rust, the variance of family means was affected by both the mean level of rust infection in the test and by RIMS, the interaction mean square for rust infection (Model 2, Table 2). As found by others (Sohn and Goddard 1979), the variance is a quadratic function of the mean rust level with tests of 40%–70% rust infection having the largest variances. This is expected of family

means derived from individual tree Bernoulli scores. The quadratic was modified by the RIMS in a manner similar to the influence of CV on the variances for volume growth: tests of a given rust level with higher interaction mean squares (more variable) are predicted to have higher variances of family means for rust.

Several different types of correlations had to be modeled for calculation of covariances needed for the C_k matrices and the off-diagonal elements of the V_k matrices (Table 2 and “Materials and methods”). For Type B rust correlations (family mean correlations for rust observed in two different test environments), the regression models predict higher correlations when both tests have similar levels of rust and both are in the 40%–70% range for mean level of infection (Model 4, Table 2). For example, when wind-pollinated offspring from the same set of parents are planted in two different test environments, the predicted correlation of family means is 0.65 when both tests average 65% rust, but only 0.32 when the two tests have 20% and 80%.

For tree volumes in two different tests, the model predicts higher type B family mean correlations between tests that are older and have low CVs (Models 5 and 6, Table 2). For volume measured in one test and rust infec-

tion in another test, no adequate regression models were found and the average observed correlation did not significantly differ from 0, which was used. Finally, for family mean correlations between rust and volume measured in the same test (Model 3, Table 2), no adequate regression models were found, so the average values of the correlations observed in the subset of 28 tests (which differed significantly from zero) were used for all tests.

Comparison of BLP breeding values to standard scores

Using the regression models described above to estimate the appropriate second moments in Eq. (3), breeding values for rust resistance (BLPRUST) and 15 year volume (BLPVOL) were predicted for 1,396 slash pine parents. Averages of the standard scores (Hatcher et al. 1981; Cotterill et al. 1983) for both rust (SSRUST) and volume (SSVOL) were also calculated for each parent for comparison. BLPVOL and BLPRUST are expressed in the units of measurement (dm^3 and %, respectively), while standard score averages (hereafter called standard scores) are unitless.

For the 1,396 parents, there was a fairly good correspondence between the breeding value predictions and the standard scores ($r^2 = 0.73$ for BLPVOL, SSVOL and $r^2 = 0.83$ for BLPRUST, SSRUST). This indicates that 73% of the variance in the 1,396 parental breeding value predictions for volume could be accounted for by the corresponding averages of standard scores. The level of correspondence decreases at the extremes of the distributions (high and low breeding values) and this is critical to selection programs choosing elite parents (high breeding values).

To examine the influence of different selection intensities on gain from using BLP breeding values versus standard scores, we calculated the expected relative genetic gain from use of the two methods as follows. The breeding values developed from BLP are estimates of genetic value, not phenotypic value; that is, the breeding value predictions are regressed back towards the mean genotypic value (zero in our case) by the multiple regression coefficients (CV^{-1}) applied to the observed (phenotypic) family means (Thompson 1979). As such, for certain types of selection the breeding value predictions of the parents selected can simply be averaged to estimate the expected genetic gains. A common practice in forestry is that a certain number of the very best parents are grafted into a seed orchard from which wind-pollinated seed will be used for operational reforestation (Namkoong et al. 1966; Shelbourne 1969). If parents are selected for the orchard based on BLP breeding values, the expected breeding value of trees grown from orchard seed (i.e., the expected genetic gain) is the average of BLP breeding values of parents in the orchard (for a monoecious species assuming equal representation and fecundity of all par-

ents). If parents for the orchard are selected on the basis of having the top standard scores for a particular trait instead of on predicted breeding values, the expected genetic gains will be less if BLP more efficiently ranks parents according to true breeding values. However, expected progress from selection based on standard scores can still be estimated by averaging BLP breeding values of the parents actually selected. Then, for a given selection intensity, we define expected relative gain as expected genetic gain from parental selection based on standard scores divided by expected genetic gain from selection based on breeding value predictions. Note that this approach implicitly assumes that BLP is more efficient than standard scores.

For 15 year volume growth, selection of the top half of the parents (699 out of 1,396) based on their standard scores, SSVOL, results in 91% of the expected genetic progress compared to selection on the basis of BLP breeding values, BLPVOL (Fig. 1). As selection intensity increases, the disparity increases and for a selection of the top 1%, selection using SSVOL is only 55% as efficient as selection on the basis of BLPVOL. Thus, for programs and selection strategies applying low to moderate selection intensities, both methods of ranking parents result in similar expected genetic gains; however, the methods diverge in parental rankings and expected genetic progress as programs intensify.

Another type of comparison between the two analytical systems is how each deals with the three main sources of "messiness" in the progeny test data: different test precisions, different test ages and parental representation in different numbers of tests. To demonstrate these effects for 15 year volume, we modeled the variance among predicted breeding values [estimated for each parent's BLPVOL from Eq. (4)] as a function of test age, precision (VCV) and the number of tests. The estimated variance

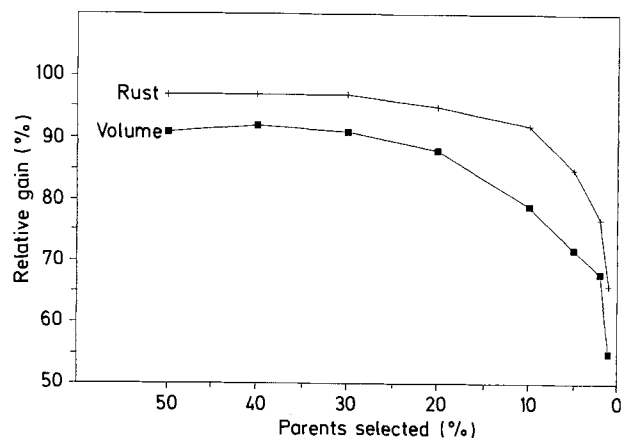


Fig. 1. Relative genetic gain efficiency from parental selection based on standard score versus BLP breeding values for different levels of selection intensity

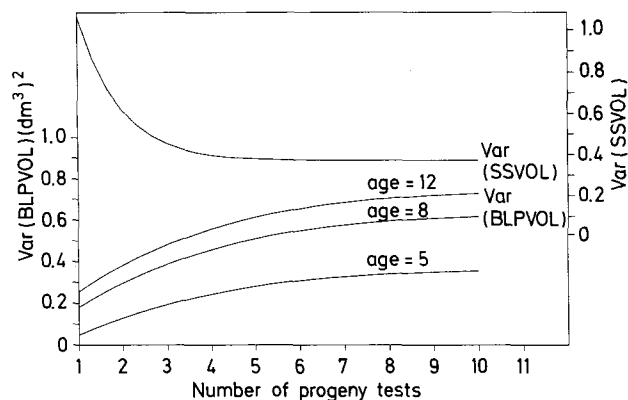


Fig. 2. Predicted variance among standard scores, Var (SSVOL), for parents in different numbers of tests and variance among BLP breeding values, Var (BLPVOL), for parents in different numbers of tests and in tests of different age; for Var (SSVOL), $\hat{y}=f$ (number of tests) and $R^2=0.94$; for Var (BLPVOL), $\hat{y}=f$ (number of tests and their average age and precision) and $R^2=0.82$

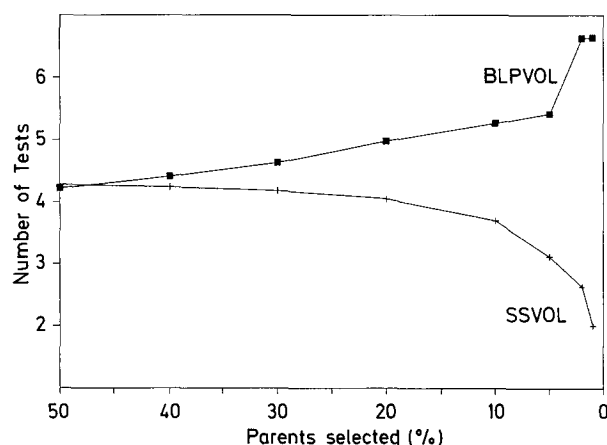


Fig. 3. For volume, the average number of progeny tests in which parents are represented for different levels of selection intensity based on standard score (SSVOL) and predicted breeding values (BLPVOL)

among predicted breeding values is larger when parents are in older, more precise and larger numbers of progeny tests (Fig. 2). That is, when the quality and quantity of data are high, BLP tends to spread out the predictions more so that there will be more very high and very low breeding value predictions. If parents are ranked on the basis of these predictions, a larger fraction of the higher ranking parents will be those with higher quality and quantity of test data (White et al. 1986). For comparison, note that by virtue of the method of calculation, standard scores treat all ages and precisions of test data equally. Furthermore, the trend with numbers of tests is opposite to that of BLP (Fig. 2) and the use of standard scores will result in a larger fraction of the high ranking parents being represented in fewer progeny tests.

That the impact of the effects with numbers of tests is magnified when selection intensity is high is shown by examining, for volume, the average number of progeny tests in which parents are represented for various selection intensities (Fig. 3). At a low selection intensity of 50%, the 699 parents retained are in an average of about 4.2 progeny tests, regardless of whether the selection is based on SSVOL or BLPVOL. However, if the top 1% of the parents are selected based on standard scores, those 14 parents are in an average of only 2 progeny tests. When parents are in fewer tests, the variance among standard scores is larger and thus it is more likely to find very high ranking parents. This is a property of ordinary least squares estimators that have similar characteristics as standard scores (White et al. 1986). Just the opposite is true for BLP, which tends to result in the highest (and lowest) ranking parents being better tested. When the top 1% of the parents are selected based on BLPVOL, those 14 parents are in an average of 6.6 tests. Cotterill et al. (1983) noted the same trend with their use of shrunken least squares. Also, the same results have been observed in dairy breeding where sires with fewer progeny are more likely to be selected using least squares estimates of sire breeding values, but the reverse is true using BLUP predictions (Henderson 1973).

General discussion

BLP versus standard scores

It is often difficult to quantify the advantages and disadvantages of two different methodologies. Sometimes direct analytical methods are available, but more often the methods are compared by either simulation techniques or by application of both methods to specific data sets (Henderson 1975). By employing the latter technique, we have implicitly assumed that BLP is superior to the use of standard scores. This is in fact true if C , V and α are exactly known, because BLP minimizes the error variance of predictions among all linear combinations of the family means (Henderson 1973, 1977, 1984). Further, if the joint distribution between the breeding values being predicted and the observed family means is multivariate normal, then BLP predictions are also best predictions, i.e., the lowest error variance of prediction among all possible combinations of the observed family means (Henderson 1973, 1977, 1984).

In most, if not all cases, C , V and α are estimated rather than known and the BLP predictions are only approximate. In general, rankings based on single trait predictions are relatively insensitive to errors involved in estimating C and V (Henderson 1984). Extrapolation from previous simulation studies of selection indexes (Harris 1963, 1964; Sales and Hill 1976; Bulmer 1985)

indicates that for the levels of genetic parameters and sample sizes in our data base, actual efficiency losses in gains from selection appear to be less than 2% below that expected using the optimal index. However, work is needed on efficiency loss when the **C** and **V** parameters are estimated by regression models as in our case (the estimation of fixed effects, α , is discussed in the next section).

Random (BLP) and mixed (BLUP) models

While both BLP and BLUP assume that genetic values are unobservable random effects, BLP assumes the fixed effects are known constants and BLUP (and the equivalent solutions to the mixed model equations) estimates these fixed effects simultaneously with the prediction of the random genetic effects (Henderson 1973, 1977, 1984). BLUP has been the overwhelming method of choice in dairy breeding (for which the technique was developed and has been refined) because fixed effects (herd, year and season effects) are, for several reasons, difficult to estimate precisely and unbiasedly: (1) subclass cell means are often more than 90% missing; (2) genetic trends in the data (due to selection progress) make contemporary or herdmate comparisons biased; and (3) records are subject to culling such that there is more data on better animals (Henderson 1973, 1974; McDaniel 1974; Iloje and Wilcox 1981; Wilcox and DeLorenzo 1983).

For our slash pine progeny test data, we believe that ordinary least squares (OLS) estimates of fixed effects (test environmental means) are both unbiased and precise. First, because the parents represented on each site approximate a random sample from the inference population and 2–6 average genetic checklots are included at every location, the estimated test environment mean should be unbiased. Secondly, on the average site with 33 parents and 4 checklots, the error variance of the estimated test environment mean is approximately 2% of that associated with a given family mean.

Both BLP and BLUP assume that **C** and **V** are known and since this is never the case in actual application, both techniques are approximations to BLP and BLUP. By using OLS estimates of fixed effects, we could allow complete generality in the **C** and **V** matrices and still keep the computations tractable. This was especially important for our data since both variances and covariances were quite heterogeneous.

Multiple trait best linear prediction

Our formulation of BLP treats a given variable observed on different sites as different traits. As such, the variance of family means at the two different sites may differ and the correlation of family means between the two sites is

a measure of genotype by environment interaction (Burdon 1977; Gianola 1986). This is formally incorporated into the model by the use of the predictive regression models in Table 2. This multiple trait approach means that selection based on predicted breeding values is indirect selection (Falconer 1981).

The long generation time in forestry makes progeny testing to harvest age (25–100 years) extremely expensive and time consuming. Thus, there is tremendous economic pressure to use data from young field progeny tests (Lambeth 1980; Lambeth et al. 1983) and also to develop controlled environment tests that are predictive of field performance at harvest (Robinson and van Buijtenen 1979; Waxler and van Buijtenen 1981; Williams 1987). A big advantage of BLP and BLUP is that diverse sources of data can be directly incorporated into predictions.

We believe that as tree improvement programs mature, diverse sources of data (from controlled environment tests, previous generations and other relatives) will become more commonplace and that the parents being evaluated will have divergent quality and quantity of data. BLP and BLUP are particularly well suited to handle these situations and will probably find more use in forest tree breeding.

Acknowledgements. We thank Drs. Paul Cotterill, Mike DeLorenzo and Barbara McCutchan and two anonymous reviewers for their thorough reviews of this manuscript. This is Journal Series No. 9245 of the Institute for Food and Agricultural Sciences, University of Florida, Gainesville, FL 32611, USA.

References

- Bulmer MG (1985) *The mathematical theory of quantitative genetics*. Clarendon Press, Oxford, 254 pp
- Burdon RD (1977) Genetic correlation as a concept for studying genotype-environment interaction in forest tree breeding. *Silvae Genet* 26:168–175
- Cotterill PP, Correll RL, Boardman R (1983) Methods of estimating the average performance of families across incomplete open-pollinated progeny tests. *Silvae Genet* 32:28–32
- Eisenhart C (1947) The assumptions underlying the analysis of variance. *Biometrics* 3:1–21
- Falconer DS (1981) *Introduction to quantitative genetics*. Longman, London, 340 pp
- Freund RJ, Littell RC (1981) *SAS for linear models*. SAS Institute, Cary/NC, 231 pp
- Garrick DJ, Van Vleck LD (1987) Aspects of selection for performance in several environments with heterogeneous variances. *J Anim Sci* 65:409–421
- Gianola D (1986) On selection criteria and estimation of parameters when the variance is heterogeneous. *Theor Appl Genet* 72:671–677
- Goddard RE (1980) The University of Florida cooperative genetics research program. In: Proc 15th North Am Quant For Genet Group Workshop. August 6–8, Coeur d'Alene/ID, pp 31–42
- Goddard RE, Strickland RK (1968) Volume and weight tables for five-year-old plantation grown slash pine. School of Forestry Res rep no. 14. University of Florida, Gainesville/FL, 8 pp

- Harris DL (1963) The influence of errors of parameter estimation upon index selection. In: Hanson WD, Robinson HF (eds) *Statistical Genetics and Plant Breeding*. Natl Acad Sci, Natl Res Council, Washington/DC, pp 491–500
- Harris DL (1964) Expected and predicted progress from index selection involving estimates of the population parameters. *Biometrics* 20:46–72
- Hatcher AV, Bridgwater FE, Weir RJ (1981) Performance level-Standardized score for progeny test performance. *Silvae Genet* 30:184–187
- Hazel LN (1943) The genetic basis for constructing selection indexes. *Genetics* 28:476–490
- Henderson CR (1963) Selection index and expected genetic advance. In: Hanson WD, Robinson HF (eds) *Statistical genetics and plant breeding*. NAS-NRC Publ No. 982, Washington/DC, pp 141–163
- Henderson CR (1973) Sire evaluation and genetic trends. In: *Anim breed genet symp in honor of J. Lush*. Anim Sci Assoc Am, Champaign/IL, pp 10–41
- Henderson CR (1974) General flexibility of linear model techniques for sire evaluation. *J Dairy Sci* 57:963–972
- Henderson CR (1975) Comparison of alternative sire evaluation methods. *J Anim Sci* 41:760–770
- Henderson CR (1977) Prediction of future records. In: Pollack E, Kempthorne O, Bailey T (eds) *Proc Int Conf Quan Genet*. Iowa State University Press, Ames/IA, pp 615–638
- Henderson CR (1984) Applications of linear models in animal breeding. University of Guelph, Ontario, Canada, 462 pp
- Hill WG (1984) On selection among groups with heterogeneous variance. *Anim Prod* 39:473–477
- Iloje MU, Wilcox CJ (1981) Best linear unbiased prediction: A method of estimating breeding values. *Dairy Sci Inf Sheet* No. DY80-71. University of Florida, Gainesville/FL, 9 pp
- Lambeth CC (1980) Juvenile-mature correlations in Pinaceae and implications for early selection. *For Sci* 26:571–580
- Lambeth CC, van Buijtenen JP, Duke SD, McCullough RB (1983) Early selection is effective in 20-year-old genetic tests of loblolly pine. *Silvae Genet* 32:210–215
- Lowe WJ, Stonecypher R, Hatcher AV (1983) Progeny test data handling and analysis. In: *Progeny testing of forest trees*. South Coop Ser Bull No. 275. Texas A & M University, College Station/TX, pp 51–68
- McDaniel BT (1974) Why new sire summaries are needed. *J Dairy Sci* 57:951–955
- Milliken GA, Johnson DE (1984) Analysis of messy data. I. Designed experiments. Lifetime Learning, Belmont/CA, 473 pp
- Mosteller F, Tukey JW (1977) *Data analysis and regression*. Addison-Wesley, London, 588 pp
- Namkoong G, Snyder EB, Stonecypher RW (1966) Heritability and gain concepts for evaluating breeding systems such as seedling orchards. *Silvae Genet* 15:76–84
- Robinson JF, van Buijtenen JP (1979) Correlation of seed weight and nursery bed traits with 5, 10 and 15-year volumes in a loblolly pine progeny test. *For Sci* 25:591–596
- Rockwood DL, Goddard RE (1973) Predicted gains for fusiform rust resistance in slash pine. In: *Proc 12th South For Tree Improv Conf*, June 12–13, Baton Rouge/LA, pp 31–37
- Sales J, Hill WG (1976) Effect of sampling errors on efficiency of selection indices. *Anim Prod* 22:1–17
- SAS Institute (1982) *SAS user's guide*. SAS Institute, Cary/NC, 584 pp
- Shelbourne CJA (1969) Tree breeding methods. *For Res Int Tech paper no. 55*. NZ For Serv, Wellington, New Zealand
- Smith H (1936) A discriminant function for plant selection. *Ann Eug* 7:240–250
- Snedecor GW, Cochran WG (1967) *Statistical methods*. Iowa State University Press, Ames/IA, 593 pp
- Sohn S, Goddard RE (1979) Influence of infection percent on improvement of fusiform rust resistance in slash pine. *Silvae Genet* 28:173–180
- Thompson R (1979) Sire evaluation. *Biometrics* 35:339–353
- Waxler MS, van Buijtenen JP (1981) Early genetic evaluation of loblolly pine. *Can J For Res* 11:351–355
- White TL, Hodge GR, Delorenzo MA (1986) Best linear prediction of breeding values in forest tree improvement. In: *Proc Genet Breed Southern Pines*, June 25–26, Gainesville/FL, pp 99–122
- Wilcox CJ, Delorenzo MA (1983) Progeny testing in dairy cattle. In: *Proc 1st Int Symp Anim Prod*, Nov. 15–19, Riberao Preto, Sao Paulo, Brazil, pp 101–113
- Williams CG (1987) The influence of shoot ontogeny on juvenile-mature correlations in loblolly pine. *For Sci* 33:411–422