Indirect Prediction of Genetic Values[1])

By T. L. White and G. R. Hodge

Department of Forestry, Institute of Food and Agricultural Sciences, University of Florida, Gainesville, FL, 32611, USA

(Received 26th February 1990)

Summary

This paper describes the application of selection index theory to indirect prediction of target genetic values (such as parental breeding values of harvest-age volume) from separate traits measured in indirect genetic tests (such as growth room, greenhouse, nursery bed and young field tests). These standard selection index procedures can be used to predict a single target genetic value or any linear function of several target genetic values from any number of indirect observations (which could, for example, be individual measurements, family means and/or clonal means on several traits).

The correlation between predicted genetic values and true genetic values ($Corr(g_t,\hat{g}_t)$) is described as a convenient measure of assessing precision of predictions; it approaches 1 as the predictions of the target genetic values approach their true, but always unknown, values. $Corr(g_t,\hat{g}_t)$ can be used to assess the value of adding new indirect observations to an index by judging whether the increase in precision (as measured by an increase in $Corr(g_t,\hat{g}_t)$) is enough to warrant the modification. Once certain parameters have been estimated reflecting the quality and quantity of indirect data and its relationship to the target traits being predicted, index coefficients and $Corr(g_t,\hat{g}_t)$ can be calculated in the absence of actual indirect data; this makes $Corr(g_t,\hat{g}_t)$ useful for planning purposes. For example, $Corr(g_t,\hat{g}_t)$ can be calculated for alternative testing strategies, experimental designs and proposed indirect traits to facilitate planning efficient indirect testing schemes. Limitations of the selection index approach are described, but in many situations, it appears a useful method both for evaluating alternative indirect testing strategies and for predicting target genetic values from a series of indirect observations.

Key words: Selection index, best linear prediction, breeding value, early testing, genetic correlation.

Introduction

Rotation ages of most forest trees make direct genetic testing for harvest-age traits a very long-term process. Also, these long-term genetic tests in field environments (herein called direct tests) are quite costly to maintain and measure over long time periods and are high risk in the sense that the tests can be quite variable (reducing the precision of the testing) and subject to partial or complete loss due to unforeseen catastrophes. Thus, there has been much emphasis in developing methods of indirect genetic testing in forest tree improvement programs. The goal of indirect testing is to rank genetic entries (e. g., clones, parents, full-sib families and/or individuals) for a target trait (such as harvest volume at rotation age) from data collected on one or more different traits (such as height, phenology and/or physiological data) measured in one or more indirect environments (such as greenhouses, growth rooms, nursery beds and young field tests).

By this definition, indirect genetic testing is already in use by nearly all tree improvement programs; progeny test evaluations and advanced-generation selections are nearly always done prior to rotation age with measurements on "young" traits used to rank the genetic entries even though the main interests (target traits) are performance at harvest age (Lambeth, 1980; Lambeth *et al.*, 1983). In addition there has been considerable research into the use of indirect testing using traits measured in non-field environments such as growth rooms, greenhouses and nursery beds (Cannell *et al.*, 1978; Robinson and Van Buijtenen, 1979; Cotterill and Nambiar, 1981; Waxler and Van Buijtenen, 1981; Williams, 1987; Adams *et al.*, 1989).

For all uses and types of indirect testing, three questions arise: 1) How are data from various relatives and traits measured in indirect environments most effectively combined into a single prediction of genetic value of the target trait?, 2) What method(s) can be used to evaluate relative efficiency of indirect testing compared to both direct testing and to other types of indirect testing? and 3) What factors affect efficiency of indirect testing? Standard linear theory for use of one indirect trait to predict a single target trait (sometimes called correlated response) is described by Falconer (1981, p. 286) and has been used by foresters to assess efficiency of early selection based on a single indirect trait (Lambeth, 1980, 1983; Talbert and Lambeth, 1986). In addition, Stonecypher and Arbez (1976) stated that selection indices could be developed from measurements on multiple indirect traits for predicting a single target trait and Baradat (1976) sketched selection index theory for doing this. Binet (1965) and Falconer (1981, p. 295) present the basic theory for indirect prediction from multiple traits and give examples for simple cases.

The objectives of this expository paper are to 1) apply general selection index theory to the prediction of genetic values of one or more target traits from measurements of one or more indirect traits on a variety of relatives, 2) provide a framework for assessing the relative precision (i.e., efficiency) of the indirect predictions, 3) investigate the factors affecting the efficiency of the predictions, and 4) describe some limitations of this overall approach.

Selection Index Theory for Indirect Prediction

Specification of the Problem and Definitions

Selection index theory is well-described in the literature (Henderson, 1963; Binet, 1965; Lin, 1978; Nordskog, 1978; Namkoong, 1979; Falconer, 1981; Bulmer, 1985; White and Hodge, 1989, chapters 9 and 10) and there have been several descriptions of selection indices for direct prediction in forest tree improvement programs (Stonecypher and Arbez, 1976; Bridgwater and Stonecypher, 1979; Burdon,

[1]) This is Journal Series No. R-00360 of the Institute of Food and Agricultural Sciences, University of Florida, Gainesville, FL, 32611.

1979, 1982; Shelbourne and Low, 1980; Cotterill and Jackson, 1981, 1985; Bridgwater *et al.*, 1983; Christophe and Birot, 1983; Dean *et al.*, 1983, 1986; Talbert, 1984, 1986; Harvey and Townsend, 1985; Bridgwater and Squillace, 1986; Land *et al.*, 1987; Cotterill and Dean, 1989). All of the standard selection index theory usually applied to direct prediction also applies to indirect prediction (see Binet, 1965; Henderson, 1977, 1984 p. 47); thus, it is only necessary here to highlight the most important aspects and describe the application of this theory to indirect prediction.

We make the following definitions using boldface type to indicate vectors and matrices.

g is a q x 1 vector of unobservable genetic values of the target traits to be predicted for each candidate (where a candidate is defined as a genetic entry, such as a clone, parent, individual or full-sib family, for which a prediction is being made).

a is a q x 1 vector of known economic weights for each of the q target traits.

y is an m x 1 vector of phenotypic, indirect observations pertaining to a single candidate (this may include several types of observations such as individual measurements and family means measured on several indirect traits in a variety of indirect environments).

α is an m x 1 vector of expected values (i.e., means) of the observed data in **y**.

b is an m x 1 vector of index coefficients to be estimated.

Var(**y**) = **V** is an m x m matrix of phenotypic variances and covariances among the indirect observations on a single candidate.

Cov(**y**,**g**) = **C** is an m x q matrix containing covariances between the observations in **y** and genetic values in **g**. Each column contains covariances between the observations and a given target genetic value. For example, the first column contains m covariances, one for each of the m observations with the genetic value of the first target trait. These are genetic covariances because only genetic effects correlate between observations and target genetic values, but the exact form (e.g., 1/4 of the additive genetic covariance or 1/2 of the total genetic covariance) depends on the exact nature of both **y** and **g**.

Var(**g**) = **G** is a q x q matrix of variances and covariances among the q genetic values being predicted. If the genetic values are breeding values, **G** contains additive variances and covariances of the target traits (since by definition for a given trait, the variance of breeding values is the additive genetic variance). If total (i.e., clonal) genetic values of the target traits are being predicted, **G** contains total genetic variances.

The goal is to use the indirect data in **y** to predict one or more of the genetic values in **g** or a linear function of the values in **g**, say **a'g** (where ' is the transpose operator). **G**, **V**, **C**, α and **a** are assumed to be vectors and matrices of constants that are known exactly (no error), while **y** and **g** are random vectors (i.e., vectors containing random variables). While the first and second moments of **y** and **g** (i.e., **G**, **V**, **C**. α and **a**) are assumed known in the following discussion of selection index theory, in practice, these must be estimated from experimental data: data on the indirect observations to estimate α and **V**; data on the target traits to estimate **G** and **a**; and data relating the indirect observations (in **y**) to the target genetic values (in **g**) to estimate **C**. For details on construction and use of these vectors and matrices, see White and Hodge (1989).

Predicting Genetic Values of Target Traits

The classical derivation of selection index formulae usually frames the problem in terms of predicting a single aggregate genetic worth, w, defined as a linear function of the genetic values in **g**, weighted by their respective economic weights:

$$\begin{aligned} w &= a_1g_1 + a_2g_2 + \ldots + a_qg_q \\ &= \Sigma\, a_ig_i \\ &= \mathbf{a'g} \ . \end{aligned}$$

It is important to note that the economic weights are defined with respect to the genetic values of the target traits and not the indirect traits being measured.

Then, if only linear functions of the observed data are considered, the solution that both minimizes the error variance of the prediction and maximizes the correlation between true and predicted genetic worth, i.e., corr ($w,\hat{w}$) is

$$\begin{aligned} \hat{\mathbf{w}} &= \mathbf{a'C'V^{-1}(y - \alpha)} \\ &= \mathbf{b'(y - \alpha)} \end{aligned} \qquad (1)$$

where $\mathbf{b'} = \mathbf{a'C'V^{-1}}$ is the row vector (of dimension 1 x m) of index coefficients.

Standard selection index theory assumes that a single index is appropriate for all candidates. Thus, the coefficients for the index are usually calculated once and then applied to all candidates as $\mathbf{b'(y - \alpha)}$ where **y** is the observed data and changes for each candidate. For example, to use progeny test data in the form of family means collected on a variety of indirect traits to predict breeding values of several harvest-age traits for 100 parents, the index coefficients are assumed the same for all 100 parents. The coefficients are calculated once and the data vector of indirect data, **y**, for a given parent is used to calculate $\hat{w}$ for that parent; this is done 100 times.

If a genetic value of only a single target trait is predicted, then q = 1 and the q x 1 **g** vector reduces to a scalar value (written g_t, the genetic value of the target trait) which is predicted as

$$\begin{aligned} \hat{g}_t &= \mathbf{c'V^{-1}(y - \alpha)} \\ &= \mathbf{b'(y - \alpha)} \end{aligned} \qquad (2)$$

where **c** is an m x 1 vector of covariances between the data values in **y** and the single target genetic value being predicted (**c** can be thought of as the column of **C** that applies to this particular target genetic value), there is only one economic weight in **a** so this is arbitrarily set to 1, and finally **b** is still an m x 1 vector of index coefficients used for each candidate to predict g_t as a linear function of the m observations for that candidate.

Assessing Precision of Predictions

Once calculated, it is often useful to estimate the precision of a given prediction. Precision is a measure of how closely predicted genetic values of a target trait (or aggregate genetic worth) cluster, upon repeated sampling, around the true, but unknown, value of that target trait (or genetic worth). Predictions based on a large amount of data on several indirect traits that are strongly related to the target genetic values result in more precise predictions than predictions made from small amounts of data on only a few indirect traits not strongly related to the target genetic values. All other things being equal, indi-

rect tests that result in more precise predictions allow more genetic gain to be made from selecting higher ranking candidates on the basis of their predicted values. Two of the most important uses of estimates of precision are: 1) assessing whether additional indirect traits or more data results in a modified index that is sufficiently more precise to warrant the modification and 2) comparing the precisions of alternative testing schemes (e.g., comparing precision of early field testing, greenhouse testing and harvest-age field testing).

Three measures of precision of a given prediction are error variance of prediction, gain possible from different types of selection, and correlation of true and predicted genetic values. These measures have two things in common. First, none require actual indirect data (i.e, **y**) for calculation. Rather, precision depends on the quality and quantity of data used. That is, data (**y**) are not actually required to assess precision as long as the parameters in **a, C, V** and **G** are known; these parameters can be estimated from previous data sets. This is analogous to the usual prediction of gain in the indirect, single trait case (Namkoong, 1979, p. 124; Falconer, 1981, p. 286): if heritabilities of the direct and indirect traits, the genetic correlation and the selection intensity are known, then expected gain can be calculated. Second, since precision depends only on parameter estimates in these vectors and matrices (and not actual data), all measures of precision are assumed to be the same for all predictions calculated from the same set of index coefficients. If one index is used to predict genetic values for 100 candidates, all 100 predictions have the same estimated precision.

As an easily interpreted measure of precision (Falconer, 1981, p. 222), the correlation between true and predicted aggregate genetic worth is calculated as (White and Hodge, 1989, p. 220)

$$\text{Corr}(\mathbf{w},\hat{\mathbf{w}}) = [(\mathbf{a'C'V^{-1}Ca})/(\mathbf{a'Ga})]^{1/2} = [\text{Var}(\hat{\mathbf{w}})/\text{Var}(\mathbf{w})]^{1/2} \quad (3)$$

If the proper variances, covariances and economic weights needed for the above vectors and matrices can be estimated, then Corr(w,$\hat{w}$) can be estimated and is interpreted as the correlation between the true genetic worths and predicted genetic worths expected upon repeated calculations using data of the quality and quantity defined by the vectors and matrices in (3).

For precision of predicting a genetic value of only a single target trait (say harvest-age volume) from an index of one or more indirect traits,

$$\text{Corr}(g_t,\hat{g}_t) = [(\mathbf{c'V^{-1}c})/\sigma^2_{g,t}]^{1/2} \quad (4)$$

where $\sigma^2_{g,t}$ is the genetic variance of the target trait being predicted and **c** is defined in (2).

To predict gain from selecting candidates on the basis of their index values (i.e., on the basis of their predicted genetic worths),

$$\text{Gain}_w = i \bullet (\mathbf{a'C'V^{-1}Ca})^{1/2} = i \bullet [\text{Var}(\hat{w})]^{1/2} \quad (5)$$

where i is the intensity of selection (Falconer, 1981) and Gain_w is predicted in the units of the index (not%) as a deviation from zero (White and Hodge, 1989, p. 221). For predicting a single target trait, gain (in units of the target trait as a deviation from the unselected mean) is:

$$\text{Gain}_{g,t} = i \bullet (\mathbf{c'V^{-1}c})^{1/2} \quad (6)$$

Both of these measures of precision (correlation and gain) are useful and are in fact similar (compare (3) with (5) and (4) with (6)). For predicting a single target trait note

$$\text{Corr}(g_t,\hat{g}_t) = (\text{Gain}_{g,t})/(i \bullet \sigma_{g,t}) \quad (7)$$

so that $\text{Corr}(g_t,\hat{g}_t)$ can be interpreted as a standardized gain (i.e., gain per unit of selection intensity and per unit of genetic variance in the target trait). Thus, for comparisons of different testing procedures for which selection intensities are not known, the correlation between true and predicted genetic values may be the better measure of precision.

Using Standardized Data

Sometimes it may be desirable to standardize indirect data by dividing each observation in **y** by its phenotypic standard deviation (Hill, 1984). When this is appropriate, the variance matrix of the observations, **V**, becomes a correlation matrix: all diagonal elements are one since the data have been standardized to have a variance of one, and all off-diagonal elements are phenotypic correlations among the observations. There is no theoretical value to using standardized data because selection index coefficients automatically scale all observations properly (Gianola, 1986). However, for the purposes of investigating what factors are important in indirect selection, or comparing the precisions of alternative testing strategies, it may be conceptually more attractive to use correlations instead of variances and covariances.

Numerical Example

To illustrate the construction and use of **C** and **V**, consider an example (using unpublished data from the Ph. D. dissertation research of Sonia De Souza, Dept. of Forestry. University of Florida) where the goal is to predict breeding values of slash pine parents for a single target trait of field resistance to fusiform rust (measured in % infection) from three indirect traits measured on young seedlings after artificial inoculation in the greenhouse. The three greenhouse symptom traits are: SYMNO, the presence (1) or absence (0) of a purple spot on the stem (presence is indicative of a thwarted infection); ADVEN, the presence (1) or absence (0) of adventitious shoots from a gall; and THIN, indicative of a rust gall that is more slender (1) than normal (0).

Open-pollinated (considered half-sib) seedlings from say each of 100 parents are sown in the greenhouse and after inoculation are placed in a randomized complete block design with 20 seedlings from each family in each of of 6 blocks. The observations in **y** are chosen to be the 3 family means (averaged over all 120 seedlings of a family) for each trait. Thus, **y** is a 3 x 1 vector containing family means for each of three indirect traits. To construct the 3 x 3 **V** matrix, note that the diagonal elements are variances of family means for each of the three indirect traits and can be modeled (for each trait) as

$$\text{Var}(y) = \sigma^2_f + \sigma^2_{fe} + \sigma^2_p/6 + \sigma^2_w/120 \quad (8)$$

where for each different symptom trait, σ^2_f is the variance due to true half-sib family effects, σ^2_{fe} is the variance

Table 1. — Parameter estimates for three greenhouse symptom traits after artificial inoculation of young slash pine seedlings with fusiform rust and their genetic correlations with the target trait: rust resistance (in %) of slash pine measured in field environments [1].

Trait	Mean[2]	Heritability[3] individual (h^2)	Heritability[3] family (h^2_f)	Variance of family means[4] (V(y))	Gen. Corr. with target[5] $r_{gt:gi}$
SYMNO	0.04	0.10	0.58	0.001781	-0.321
ADVEN	0.74	0.10	0.52	0.008977	0.567
THIN	0.11	0.13	0.63	0.004941	-0.607

[1]) Unpublished data from the Ph.D. Dissertation of Sonia Maria De Souza, Dept. of Forestry, University of Florida, Gainesville, FL, 32611.

[2]) The mean for each trait is the proportion of seedlings with this symptom trait in an average greenhouse test.

[3]) For each trait, $h^2 = 4\sigma^2_f/(\sigma^2_f + \sigma^2_{fe} + \sigma^2_p + \sigma^2_w)$.
For each trait, $h^2_f = \sigma^2_f/V(y)$.

[4]) For each trait, $V(y) = \sigma^2_f + \sigma^2_{fe} + \sigma^2_p/6 + \sigma^2_w/120$.

[5]) For each trait, $r_{gt:gi} = Cov_A(g_t,g_i)/(\sigma^2_{g,t}\cdot\sigma^2_{g,i})^{1/2}$ where the numerator is the additive genetic covariance between the indirect symptom trait and the field resistance, and the denominator terms are the respective additive genetic variances.

due to family by environment interaction, σ^2_p is the variance due to block by family interactions (which is divided by 6, the number of blocks) and σ^2_w is the variance due to within-plot effects (i.e., among the 20 seedlings from each family within a block pooled across blocks). Assuming balanced data, an estimate of the variance of family means for each greenhouse symptom trait can be obtained by either estimating variance components from an analysis of variance and calculating (**8**) or by calculating the variance among the 100 observed family means observed for that trait. By greenhouse testing in different years and seasons, it is possible to get variance component estimates and heritabilities that are free from bias due to family by environment interactions (*Table 1*).

The off-diagonal elements of **V** are phenotypic covariances of family means between pairs of indirect traits and can be estimated either through analysis of covariance or through family mean correlations (White and Hodge, 1989, Topic 5 of Chapter 7). For these data, estimates of family mean correlations are: -0.23, 0.61 and -0.53 for SYMNO with ADVEN, SYMNO with THIN and ADVEN with THIN, respectively, and the appropriate covariances are obtained from these correlations using the variances of family means from *Table 1*. Then,

$$\mathbf{V} = \begin{bmatrix} 0.001781 & -0.000916 & 0.001810 \\ -0.000916 & 0.008977 & -0.003556 \\ 0.001810 & -0.003556 & 0.004941 \end{bmatrix}$$

The 3 covariances needed for the 3 x 1 **c** vector are covariances between the phenotypic observations (in this case family means) of the indirect traits and the breeding value of the target trait being predicted. For this example, open-pollinated progeny from 100 parents were tested in both greenhouse (indirect) and field (direct) environments and these data used to estimate the additive genetic correlations between each indirect trait and field rust incidence (*Table 1*). Then, assuming only common parental (i.e., genetic) effects correlate between the field and greenhouse environments, each element of **c** is calculated as:

$$1/2\ Cov(g_t,g_{yi}) = 1/2\ r_{gt:gi}\ (\sigma^2_{g,t} \cdot \sigma^2_{g,i})^{1/2} \qquad (9)$$

where for each of the three indirect traits (i = 1, 2, 3), $r_{gt:gi}$ is the additive genetic correlation from *Table 1*, $\sigma^2_{g,t}$ is the additive genetic variance of the target trait (0.05 in this case), and $\sigma^2_{g,i}$ is the additive genetic variance of the i^{th} indirect trait (0.00416, 0.01876, and .01240 for SYMNO, ADVEN and THIN, respectively). The covariance is multiplied by ½ because the covariance between half-sib family effects and the parental breeding values is half the additive genetic covariance (White and Hodge, 1989, p. 89). The elements of **c** can be filled in and **c'** = [-0.002315 0.008682 -0.007557].

With these values of **c** and **V**, the index coefficients from (**2**) are **b'** = [0.2346 0.4916 -1.2615] and these are the weights that would be applied to the family means for the three indirect traits from a given parent to predict that parent's breeding value for rust resistance in the field. The precision of these predictions is estimated from (**4**) as $Corr(g_t,\hat{g}_t) = 0.515$ and is interpreted as the expected value of the correlation between true and predicted parental breeding values for field rust resistance. Other indirect testing procedures or direct testing procedures that resulted in higher correlations would be preferred, all other things (length and cost of tests, selection intensity, etc.) being equal. Also, the value of adding or deleting one or more indirect traits can be investigated. For example, the correlation could be recalculated after deleting column and row 1 in **V** and the first element of **c** to see the effects of omitting SYMNO from the index. When this is done, the reduced index containing THIN and ADVEN results in $Corr(g_t,\hat{g}_t) = 0.514$ indicating that the full index with three indirect traits is not enough better than this two-trait index to justify measurement of SYMNO.

Factors Affecting the Precision of Indirect Predictions

When comparing different testing strategies or investigating which indirect traits to consider measuring, it is useful to understand how different factors affect the precision of the predictions as measured by Corr $(g_t,\hat{g}_t)$. Even for field testing (whether measurements are taken at harvest age or before), precision of the predicted genetic values can be calculated from (3) or (4) and depends on the quality and quantity of field data (e.g., age, number of tests, precision of tests, etc.). For predicting slash pine parental breeding values for field rust resistance (% incidence), we have used field measurements on open-pollinated offspring to predict breeding values of 2245 parents represented in varying numbers and qualities of field tests (WHITE and HODGE 1988). Each of the 2245 predictions is based on a different selection index that reflects the quality and quantity of data for that parent; hence, $Corr(g_t,\hat{g}_t)$ is different for each prediction. Corr $(g_t,\hat{g}_t)$ has been estimated for each of the 2245 predictions and, for each parent, depends on the experimental designs of the field tests (more blocks and trees per block increase precision), the average rust incidence in the test (too much or too little rust reduces precision), and, in particular, the number of sites a parent's offspring are represented in (unpublished data). Averaged over predictions of breeding values for 2245 parents, $Corr(g_t,\hat{g}_t)$ is 0.63, 0.73 0.79 and 0.84 for parents with offspring in 1, 2, 3 to 5, and more than 5 tests, respectively. Thus, predictions from an indirect (say greenhouse) test with $Corr(g_t,\hat{g}_t)$ = 0.6 are approximately equivalent (in terms of precision) to those from one field test and an indirect test with $Corr(g_t,\hat{g}_t)$ = 0.8 is as good as many field tests.

In the following sections, we solve (4) for several simple cases of predicting a single target trait from several indirect traits to understand the factors causing $Corr(g_t,\hat{g}_t)$ to change.

Predicting a Single Target Value from One Indirect Observation

For the simplest case, consider $Corr(g_t,\hat{g}_t)$ for predicting the genetic value of a single target trait from measurement of a single indirect trait. Equation (4) applies to prediction of any type of genetic value (e.g., breeding value, total clonal value) from any type of measurement (e.g., on plantlets, seedlings). For this simple case, let y_1 be the phenotypic value of the one indirect trait (e.g., a half-sib family mean for the indirect trait), then

$$\begin{aligned} Corr(g_t,\hat{g}_t) &= [(c'V^{-1}c)/\sigma^2_{g,t}]^{1/2} \\ &= [(Cov(g_t,y_1)\bullet Var(y_1)^{-1}\bullet Cov(g_t,y_1))/\sigma^2_{g,t}]^{1/2} \\ &= [Cov(g_t,y_1)^2/(\sigma^2_{g,t}\bullet Var(y_1))]^{1/2} \\ &= abs[Cov(g_t,y_1)/(\sigma_{g,t}\bullet\sigma_{p,y1})] \\ &= abs(r_{gt:y1}) \\ &= abs(r_{gt:g1})\bullet h_{y1} \end{aligned} \qquad (10)$$

where $\sigma_{p,y1}$ = phenotypic standard deviation for the indirect trait y_1, "abs" indicates the absolute value, $r_{gt:y1}$ is defined (by lines 4 and 5 in the derivation above) to be the *genophenotypic correlation* (i.e., the correlation between the genetic value being predicted and the phenotypic value measured,) $r_{gt:g1}$ is the genetic correlation between the target trait and measured trait and h_{y1} is the square root of the heritability of the indirect observation (see also BINET, 1965). Note that the phenotypic standard deviation and heritability must be defined with respect to the observation, y_1, which may be an individual measurement, an open-pollinated family mean, a clonal mean, etc. HANSON (1963) describes the concept of heritability applied to different types of observations (or "reference units" by HANSON's terminology).

The precision of prediction of a single genetic value from any single indirect observation depends only on the genophenotypic correlation between the genetic value of the target trait being predicted and the phenotypic value of the predictor. As heritability of the indirect observation approaches 1, Corr $(g_t,\hat{g}_t)$ and the genophenotypic correlation both approach the absolute value of the genetic correlation. Because higher heritability of the phenotypic observation on the indirect trait increases the precision of the prediction of target trait, indirect test environments should be as uniform as possible. Also, when the indirect observation is a mean (say family or clonal mean), increased replication in the indirect environment increases the precision of the prediction. So, depending upon the situation, it may be better (i.e., higher precision) to collect more data on a single indirect trait (say by replicating in space or time or by measuring the trait in different treatments) measured in a carefully controlled test environment, than to increase the number of indirect traits in the index.

For the numeric example from the previous section (*Table 1*), the precision (Corr $(g_t,\hat{g}_t)$) for predicting field rust resistance from the single best indirect trait (family mean of THIN) is $abs(-0.607)\cdot(0.63)^{1/2}$ = 0.481 and is nearly as precise as the index with all three indirect traits ($Corr(g_t,\hat{g}_t)$ = 0.515). The family heritability for THIN increases to 0.84 if the greenhouse test is repeated three different seasons (compared to family heritability of 0.63 in Table 1 when tested only once)and is obtained by calculating (8) with the proper denominators to reflect 3 tests instead of one. Then, $Corr(g_t,\hat{g}_t)$ = $abs(-0.607)\cdot(0.84)^{1/2}$ = 0.56 which is considerably higher than the precision when the family mean for THIN is obtained from a single test, and is near the absolute value of the genetic correlation between THIN and the target trait (—0.607, *Table 1*).

Predicting a Single Target Value from Two Indirect Observations

When two observations (such as family means on two indirect traits) are used to predict the genetic value of a single target trait, (4) reduces to the following scalar form

$$Corr(g_t,\hat{g}_t) = [(r^2_{gt:y1} + r^2_{gt:y2} - 2r_{gt:y1}r_{gt:y2}r_{y1:y2})/(1-r^2_{y1:y2})]^{1/2} \qquad (11)$$

where y_1 and y_2 are the two phenotypic observations, $r_{y1:y2}$ is the phenotypic correlation between indirect observations and all other terms are defined analogously to the terms in (10). Here, precision of the predictions, Corr $(g_t,\hat{g}_t)$, depends on three factors: the phenotypic correlation between y_1 and y_2 and the two genophenotypic correlations (g_t with y_1, and g_t with y_2). If the two observations have high heritabilities (say because they are measured with high precision), then the genophenotypic correlations approximate the two genetic correlations, see (10).

Interestingly, the phenotypic correlation, not the genetic correlation, between the indirect observations occurs in (11). Phenotypic correlations can differ from genetic correlations in magnitude and even sign. Consider two indirect phenotypic observations that both have positive geno-

phenotypic correlations with the target genetic value (so that increases in both y_1 and y_2 imply positive change in g_t). If g_t, y_1 and y_2 are all quantitative traits controlled by many genes, it is possible that the genes in common between g_t and y_1 are different than the genes in common between g_t and y_2. Then y_1 and y_2 could be genetically uncorrelated or even negatively correlated. Further, since we are interested in the phenotypic correlation between y_1 and y_2, environmental effects on the phenotypic expression of y_1 and y_2 may be negatively correlated, possibly contributing to low or negative phenotypic correlation between the two indirect observations.

In order to investigate the manner in which the three correlations in (**11**) interact to influence the precision of the predictions from a two trait index, it is first necessary to develop some concepts regarding *systems* of correlations. Consider a system among the three variables, y_1, y_2 and g_t where the absolute value of the genophenotypic correlations between both indirect observations and g_t are 0.5, and the absolute value of the phenotypic correlation between y_1 and y_2 is 0.3. There are eight possible systems of correlations shown in *figure 1* (2^3 = 8, three correlations of two possible signs, either positive or negative). If both y_1 and y_2 are positively correlated with g_t and if

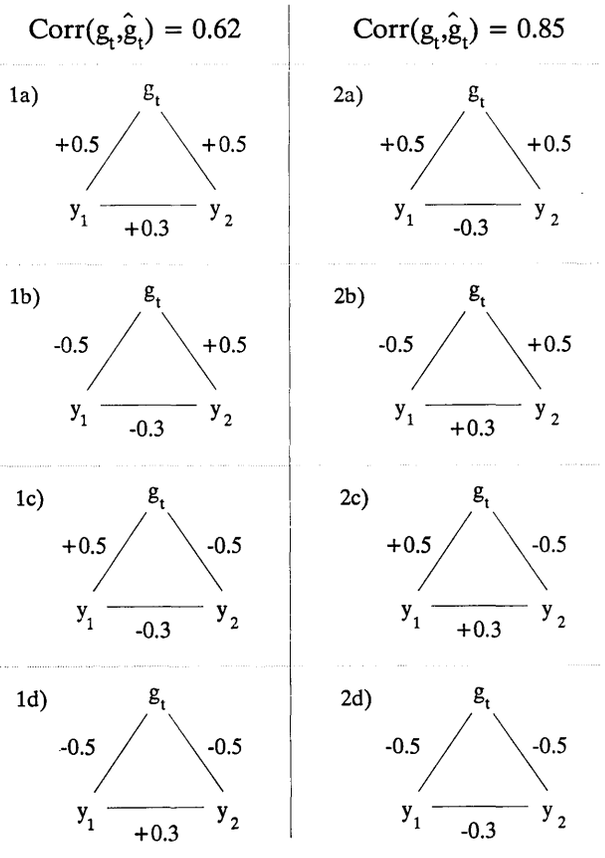

Figure 1. — Systems of correlations for the case where two indirect traits, y_1 and y_2, are used to predict a target trait g_t. Four of the systems (left column) result in a correlation between predicted and true genetic value (Corr(g_t,$\hat{g}_t$)) of 0.62. Four of the systems (right column) result in Corr (g_t,$\hat{g}_t$) = 0.85. The four systems within the left column are equivalent, i.e., any one system can be made into another simply by multiplying observations on y_1, y_2 or g_t by -1 prior to calculation of the correlation, and similarly for the systems within the right column. However, it is not possible to multiply y_1, y_2, or g_t by -1 to change a system from the left column into a system from the right column.

y_1 and y_2 are positively correlated with each other (system 1a in *Figure 1*), then Corr(g_t,$\hat{g}_t$) = 0.62 (from (**11**)). However, if y_1 and y_2 are positively correlated with g_t, but are negatively correlated with each other (system 2a), then the precision of the index is substantially higher (Corr (g_t,$\hat{g}_t$) = 0.85). Initially, this may seem counter-intuitive: that only the absolute value of the correlation between y_1 and y_2 should be important, but not the sign. The sign can be changed simply by multiplying all measurements of either y_1 or y_2 by -1 before calculating their correlation, and this does not change the degree of relationship between the two variables. However, this considers only the pair of indirect observations y_1 and y_2, while we must be concerned with the *system* of 3 correlations among the three random variables y_1, y_2 and g_t.

The remaining six possible systems of correlations in *figure 1* are all equivalent to one of the two just described (1a or 2a) and can be obtained by multiplying one of the three random variables by —1 prior to calculating the correlations. For example, if the correlations between y_1 and g_t, y_2 and g_t, and y_1 and y_2 are —0.5, +0.5, and —0.3, respectively (system 1b, *Figure 1*) then Corr (g_t,$\hat{g}_t$) = 0.62. If all observations of y_1 were multiplied by —1, the resulting system of correlations would be equivalent to the system with all positive correlations among all three traits (system 1a, *Figure 1*). Note, however, that no manipulation of variables can make system 1a equivalent to system 2a; they are fundamentally different, as reflected in the fact that Corr (g_t,$\hat{g}_t$) is different. An important implication of this for investigating the interaction of the three correlations in (**11**) is that without loss of generality, we can arbitrarily assume that both indirect observations are positively correlated with g_t, and thereby only need to investigate systems 1a and 2a in *Figure 1*; these are the only ones fundamentally different.

One way to investigate the factors influencing (**11**) is to graph Corr(g_t,$\hat{g}_t$) across the ranges of each of the three correlations on the right-side of the equation; however, some hypothetical sets of correlations among y_1, y_2 and g_t do not define a system of correlations in the sense that that set can never occur among three random variables. As an extreme example, if both y_1 and y_2 have correlations of 1 with g_t, then a correlation of 1 betwen y_1 and y_2 is the only permissible value that defines a system of correlations. If both y_1 and y_2 have correlations of 0.9 with g_t, then only correlations greater than 0.62 between y_1 and y_2 are permissible. To be a system of correlations, the symmetric matrix formed by putting ones on the diagonals and the correlations in the appropriate off-diagonals must be positive definite (Searle, 1966) and therefore have only positive eigenvalues. Thus, for this example, the 3 x 3 matrix with the three correlations above and below the diagonal of ones can be checked for positive eigenvalues to ensure that the set defines a permissible system of correlations.

For two indirect observations in an index, we defined (without loss of generality) both to have positive genophenotypic correlations with the target genetic value and investigated Corr(g_t,$\hat{g}_t$) for permissible systems of correlations among the three random variables. *Figure 2* shows the pattern for the case when the genophenotypic correlation between y_1 and g_t is 0.5 and the pattern is similar for other levels of this particular correlation. Note that many systems of correlations result in an index with precise predictions; for example, seven systems shown in

Figure 2 result in $Corr(g_t,\hat{g}_t) = 0.8$. Remember that these systems are theoretically possible, but some may be biologically unlikely.

While the pattern is somewhat complicated, several conclusions can be drawn. First, adding a second trait to an index always results in better (i.e., more precise) predictions than using one. This can be seen by noting that when using one indirect trait, the $Corr(g_t,\hat{g}_t)$ is equal to the genophenotypic correlation (see **10** above) and that all values of $Corr(g_t,\hat{g}_t)$ in *Figure 2* are above 0.5 (the genophenotypic correlation between y_1 and g_t which is held constant for all points in *Figure 2*). Second, when the phenotypic correlation between y_1 and y_2 is zero (zero on the x axis), the value of adding y_2 to the index results in progressively more precise predictions as y_2 (the second indirect observation) becomes more strongly correlated with the target genetic value. Third, when both indirect observations have equal genophenotypic correlations with the target genetic value (as can be seen be tracing the curve for $r_{gt;y2} = 0.5$ in *Figure 2*), $Corr(g_t,\hat{g}_t)$ increases as the phenotypic correlation between the two indirect observation decreases. This is analogous to the multicollinearity problem in multiple linear regression: when two regressors are highly correlated, adding both variables to the regression model does not greatly increase the coefficient of determination. Heuristically, variables that are both correlated with the target and highly correlated with each other are 'redundant'. Two variables which are uncorrelated with each other are not 'redundant', i.e., the second variable contains only 'new' information about the target trait. A fourth concept (seen by tracing the curve for $r_{gt;y2} = 0.0$ in *Figure 2*) is that adding a second indirect observation that is uncorrelated with the target genetic value can increase the precision of the predictions if it is correlated (and therefore provides information about) the first indirect observation. Therefore, indirect observations with low genophenotypic correlations with the target genetic value should not be discarded automatically; they may contribute significantly to the predictive ability of the index.

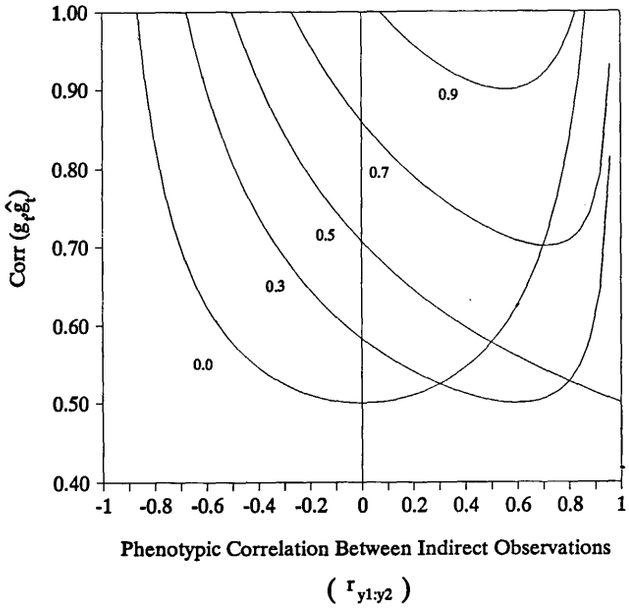

Figure 2. — Correlation between true and predicted genetic values of the target trait ($Corr(g_t,\hat{g}_t)$) for different levels of phenotypic correlation between the indirect observations ($r_{y1:y2}$). For simplicity of display, the genophenotypic correlation between g_t and y_1 ($r_{gt:y1}$) is assumed 0.5. The curves represent different levels of genophenotypic correlation between g_t and y_2 ($r_{gt;y2}$).

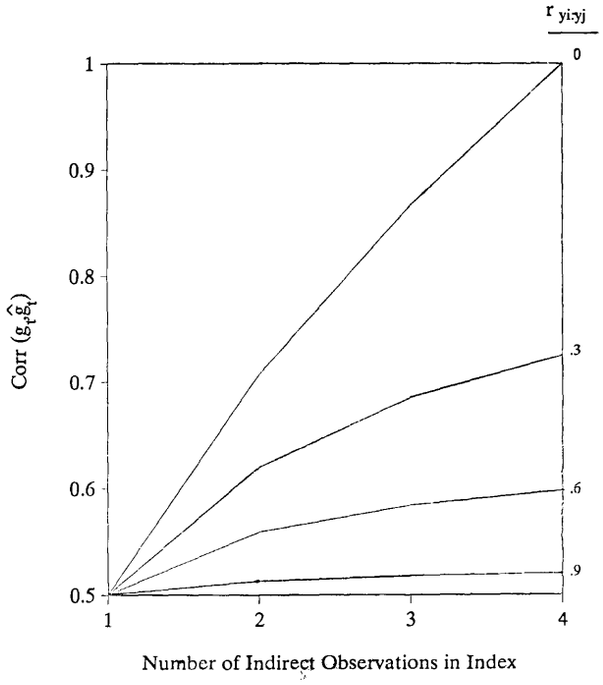

Figure 3. — Correlation between true and predicted genetic values of the target trait ($Corr(g_t,\hat{g}_t)$) for different numbers of indirect observations in an index and different phenotypic correlations ($r_{yi:yj}$) among the indirect traits. For simplicity of display, it assumed that all indirect observations have a genophenotypic correlation of 0.5 with the genetic value of the target trait and that all phenoypic correlations are equal among pairs of indirect traits.

Predicting One or More Target Genetic Values from Multiple Indirect Observations

As more indirect observations are added to the index, the concepts are similar to those observed with two indirect traits. For example, *Figure 3* illustrates the restricted case of all indirect observations having genophenotypic correlations of 0.5 with the genetic value of the target trait. When three indirect observations are included in an index, the precision is extremely high, $Corr(g_t,\hat{g}_t)$ is nearly 0.9, if the traits are uncorrelated. However, adding measurements of three or four indirect traits to the index is not appreciably better than an index with one indirect observation if the three or four traits are highly correlated. The importance of low phenotypic correlations among the indirect traits to be measured may have important implications for breeders searching for good combinations of traits to include in an indirect selection index. It may be worthwhile to consider target traits as functions of processes which are independent of (and therefore uncorrelated with) one another. For example, dry matter production in loblolly pine may be a function of both nutrient uptake efficiency and nutrient utilization efficiency (Li *et al.*, 1989). These two processes may be controlled by different genes, and could be essentially uncorrelated with one another.

To this point, we have considered factors affecting the precision of the prediction of a single genetic value, g_t.

When predicting a linear combination of target genetic values (an aggregate genetic worth, w), the interrelationships among the factors contributing to precise predictions become even more complex. Thus, it may be difficult, if not impossible, to determine *a priori* the relative values of several indirect observations. It will likely be necessary to try all possible combinations of these in a systematic fashion to determine an index with a combination of observations that results in precise predictions. In fact, there may be several combinations of indirect observations that all result in nearly the same precision. In this instance, the determining factors may be cost or time of measuring different traits, the ease of obtaining precise values of different observations, the precision of associated parameter estimates, or whether the prediction equation seems biologically meaningful.

Limitations of Selection Index Approach for Indirect Prediction

Since no new theory has been developed, the limitations and disadvantages of using selection indices for indirect prediction are the same as for direct prediction. Foremost is the fact that economic weights, and the genetic parameters (phenotypic and genetic variances and covariances) in **C** and **V** are assumed known, when in fact they must be estimated. Most research on the effects of using parameter estimates in place of true parameters has been based on mass selection for multiple traits (Williams, 1962a, 1962b; Harris, 1963, 1964; Sales and Hill, 1976; Bulmer, 1985; Baker, 1986). Simulation studies comparing index selection using indices with "true" parameters to indices with "estimated" parameters indicate that relative rankings of candidates are more stable to the use of estimated parameters than are the predicted genetic values themselves. That is, the candidates may be ranked quite well even when the predicted genetic values are very different from the true genetic values. Thus, the actual selections made are less affected (and therefore the actual gain achieved is less affected) than the amount of gain predicted based on the estimated parameters and predicted values. Because most previous research has been based on indices from mass selection, new research is needed to investigate the efficiency of indices based on parameter estimates from experimental designs important in forest tree improvement, especially indirect selection.

Selection index and best linear prediction maximize $Corr(g_t,\hat{g}_t)$ only among all possible linear combinations of the data (Henderson, 1963, 1984; Bulmer, 1985; White and Hodge, 1989). There may be non-linear functions of the data that predict genetic values with a higher correlation with the true genetic values and their use would result in more gain. However, if the joint distribution of **y** and **g** is multivariate normal, then selection index is best among all combinations of the data, both linear and non-linear (Henderson, 1963, 1984; Bulmer, 1985). Many of the variables used in forest tree breeding are measured on a continuous phenotypic scale (such as height), and often family means are used as observations which are approximately normally distributed according to the central limit theorem. Thus, for many traits in forestry, this is not likely to be an area of concern. However, some problems may occur with traits which are not normally distributed, for example certain biomial traits such as disease resistance. Even with binomial traits measured 0 or 1 on an individual tree basis, if parental breeding values are predicted from family means averaged over many individuals, the distribution of these family means approaches normality. The problem would occur in predicting the breeding (or total genetic) value of a specific individual with 0,1 data.

Derivation of selection index formulae assumes balanced data, i.e., that the exact same quantity and quality of information is available for each candidate for which a genetic value is being predicted; thus, the same index is assumed appropriate for all candidates. This is quite unlikely in direct tests for a trait such as harvest-age volume in the field, since mortality within tests, and even loss of entire tests, is possible. Balanced data are more likely in indirect selection, for example in short-term tests in a greenhouse; but data may vary in precision among runs of the greenhouse test. If data are unbalanced or vary in quality, best linear prediction (BLP) can be used to calculate a different selection index for each candidate or class of candidates having different kinds of data (White and Hodge, 1989). For the purposes of planning testing strategies and examining which traits to include in an index, an assumption of balanced data can be made. Then, alternative testing strategies can be compared for how precisely they predict the target trait by use of a single index for each alternative.

Another limitation of using the correlation between predicted and true genetic values ($Corr(g_t,\hat{g}_t)$) is that it does not consider all important factors affecting the use of indirect prediction and selection, and therefore should not be the only criterion used in evaluating alternatives. The breeder must consider factors such as length of testing, cost of various alternatives, the selection intensity which could be applied, etc. For example, the selection intensity achievable in an indirect (say greenhouse) test may be considerably higher than in a long-term field (i.e., direct) test because more families can reasonably be screened. However, use of $Corr(g_t,\hat{g}_t)$ provides the basis for an objective technique to evaluate alternative strategies using a criteria of primary importance, i.e., precision of prediction. It can be calculated for prediction of a single target trait (**4**) or any linear function of several target traits (**3**).

Conclusions

Selection index and best linear prediction theory both apply equally well to direct or indirect testing for prediction of genetic values. If selection index values are to be used to rank candidates and make selections, then a variety of different testing schemes can be compared on an equal basis. Since all testing schemes are directed to predicting the same target genetic value, all schemes can be compared using the genetic gain or the correlation between true and predicted genetic values ($Corr(g_t,\hat{g}_t)$). $Corr(g_t,\hat{g}_t)$ can also be thought of as the standardized gain, i.e., gain per unit of selection intensity and per unit of genetic variance in the target trait. The marginal advantage of adding additional indirect traits to an index can be determined by calculating the marginal increase in $Corr(g_t,\hat{g}_t)$.

From this study, the precision of an index of indirect observations (i.e., $Corr(g_t,\hat{g}_t)$) can be increased substantially by adding traits that are either 1) weakly correlated with indirect traits already in the index and strongly correlated with the target genetic value, or 2) strongly

correlated with indirect traits already in the index and weakly correlated with the target genetic value. Heuristically, adding this second type of trait to an index indirectly provides information about g_t through additional information on the other indirect trait. For multiple trait situations, the complex interrelationships among variables can make it difficult to evaluate *a priori* the marginal increase in $Corr(g_t,\hat{g}_t)$ from the addition of a particular trait, thus increasing the value of an objective measure of precision of prediction. It may become necessary to calculate $Corr(g_t,\hat{g}_t)$ with all traits, and systematically delete traits or combinations of traits to identify an optimum index. This is analogous to a stepwise or all-possible combination approach to building regression models. This method has limitations, but it does allow assessment of the precision afforded by a wide range of testing strategies, and can serve as the basis for more complex analyses.

Acknowledgements

We thank W. T. Adams, Michael DeLorenzo, Dudley Huber, Sonia De Souza, Steen Magnussen and Don Rockwood for their critical reviews of this manuscript.

References

Adams, W. T., Joyce, D., Li, P., ST. Clair, B. and Vargas-Hernandez, J.: Pacific Northwest Tree Improvement Research Cooperative Annual Report. Forest Research Lab., Oregon St. Univ., Corvallis, OR, 97331. 34 pp., (1989). —— Baker, R. J.: Selection Indices in Plant Breeding. CRC Press, Boca Raton, FL. 218 pp., (1986). —— Baradat, P.: Use of juvenile-mature relationships and information from relatives in combined multitrait selection. In: Proc. IUFRO Joint Meeting on Adv. Gen. Breeding. June 14—18, Bordeaux, France. pp. 121—138 (1976). —— Binet, F. E.: On the construction of an index for indirect selection. Biometrics **21**, 291—299 (1965). —— Bridgwater, F. E. and Squillace, A. E.: Selection indexes for forest trees. In: Advanced Generation Breeding of forest Trees. Southern Coop. Series Bull. 309. Louisiana Ag. Exp. Stn., Baton Rouge, LA. pp. 17—20 (1986). —— Bridgwater, F. E. and Stonecypher, R. W.: Index selection for volume and straightness in a loblolly pine population. In Proc. 15th. South. For. Tree Improve. Conf. June 19—21, Miss State, MS. pp. 132—139 (1979). —— Bridgwater, F. E., Talbert, J. T. and Jahromi, S.: Index selection for increased dry weight in a young loblolly pine population. Silvae Genet. **32**, 157—161 (1983) —— Bulmer, M. G.: The Mathematical Theory of Quantitative Genetics. Clarendon Press, Oxford, GB. 254 pp., (1985). —— Burdon, R. D.: Generalisation of multi-trait selection indices using information from several sites. New Zealand J. For. Sci. **9**, 145—152 (1979). —— Burdon, R. D.: Selection indices using information from multiple sources for the single-trait case. Silvae Genet. **31**, 81—85 (1982) Cannell, M. G. R., Bridgwater, F. E. and Greenwood, M. S.: Seedling growth rates, water stress responses and root-shoot relationships related to eight-year volumes among families of *Pinus taeda* L. Silvae Genet. **27**, 237—249 (1978). —— Christophe, C. and Birot, Y.: Genetic structures and expected genetic gains from multitrait selection in wild populations of Douglas-fir and Sitka spruce. II. Practical application of index selection on several populations. Silvae Genet. **32**, 173—181 (1983). —— Cotterill, P. and Jackson, N.: Index selection with restrictions in tree breeding. Silvae Genet. **30**, 106—108 (1981). —— Cotterill, P. P. and Dean, C. A.: Handbook of Applied Index Selection. CSIRO, Div. Forestry, Canberra, Australia. (In Press) (1989). —— Cotterill, P. P. and Jackson, N.: On index selection. I. Method of determining economic weights. Silvae Genet. **34**, 56—63 (1985). —— Cotterill, P. P, and Nambiar, E. K. S.: Seedling physiology of three radiata pine families with parents of contrasting growth. Aust. For. Res. **11**, 13—22 (1981). —— Dean, C. A. Cotterill, P. P. and Cameron, J. N.: Genetic parameters and gains expected from multiple trait selection of radiata pine in eastern Victoria. Aust. For. Res. **13**, 271—278 (1983). —— Dean, C. A., Cotterill, P. P. and Eisemann, R. L.: Genetic parameters and gains expected from selection in *Pinus caribaea* var. *hondurensis* in northern Queensland, Australia. Silvae Genet. **35**, 229—236 (1986). —— Falconer, D. S.: Introduction to Quantitative Genetics. Longman and Co., NY. 340 pp., (1981). —— Gianola, D.: On selection criteria and estimation of parameters when the variance is heterogeneous. Theor. Appl. Genet. **72**, 671—677 (1986). —— Hanson, W. D.: Heritability. In: Statistical Genetics and Plant Breeding. Eds: W. D. Hanson and H. F. Robinson. NAS-NRC Pub. No. 982, Washington, DC. pp. 125—140 (1963). —— Harris, D. L.: The influence of errors of parameter estimation upon index selection. In: Statistical Genetics and Plant Breeding. Nat. Acad. Sciences, Nat. Research Council, Washington, DC. pp. 491—500 (1963). —— Harris, D. L.: Expected and predicted progress from index selection involving estimates of the population parameters. Biometrics **20**, 46—72 (1964). —— Harvey, W. R. and Townsend, A. M.: Selection on a combination of individual, family, and stand merit in provenance tests. Forest Sci. **31**, 813—821 (1985). —— Henderson, C. R.: Selection index and expected genetic advance. In: Statistical Genetics and Plant Breeding. Eds. W. D. Hanson and H. F. Robinson. NAS-NRC Pub. No. 982, Washington, DC. pp 141—163 (1963). —— Henderson, C. R.: Best linear prediction of breeding values not in the model for records. J. Dairy Sci. **60**, 783—787 (1977). —— Henderson, C. R.: Applications of Linear Models in Animal Breeding. University of Guelph, Guelph, Ontario, CAN. 462 pp., (1984). —— Hill, W. G.: On selection among groups with heterogeneous variance. Anim. Prod. **39**, 473—477 (1984). —— Lambeth, C. C.: Juvenile-mature correlations in *Pinaceae* and implications for early selection. Forest Sci. **26**, 571—580 (1980). —— Lambeth, C. C., van Buijtenen, J. P., Duke, S. D. and McCullough, R. B.: Early selection is effective in 20-year-old genetic tests of loblolly pine. Silvae Genet. **32**, 210—215 (1983). —— Land, S. B., Bongarten, B. C. and Toliver, J. R.: Genetic parameters and selection indices from provenance/progeny tests. In: Statistical Considerations in Genetic Testing of Forest Trees. South. Coop. Series Bull. 324. University of Florida, Gainesville, FL. pp. 59—74 (1987). —— Li, B. McKeand, S. E. and Allen, H. L.: Early selection of loblolly pine based on growth, morphological, physiological and nutritional characteristics. In: Twentieth South. For. Tree Improve. Conf. June 26—30, Charleston, SC. (1989). —— Lin, C. Y.: Index selection for genetic improvement of quantitative characters. Theor. Appl Genet. **52**, 49—56 (1978). —— Namkoong, G.: Introduction to Quantitative Genetics in Forestry. U.S. Dept. Agric. Tech. Bull. No. 1588, (1979), 342pp. —— Nordskog, A. W.: Some statistical properties of an index of multiple traits. Theor. Appl. Genet. **52**, 91—94 (1978). —— Robinson, J. F. and van Buijtenen, J. P.: Correlation of seed weight and nursery bed traits with 5-, 10- and 15-year volumes in a loblolly pine progeny test. Forest Sci. **25**, 591—596 (1979). —— Sales, J. and Hill, W. G.: Effect of sampling errors on efficiency of selection indices. Anim. Prod. **22**, 1—17 (1976). —— Searle, S.: Matrix Algebra for the Biological Sciences. John Wiley and Sons, NY. 296pp., (1966). —— Shelbourne, C. J. A. and Low, C. B.: Multitrait index selection and associated genetic gains of *Pinus radiata* progenies at five sites N. Z. J. For. Sci. **10**, 307—324 (1980). —— Stonecypher, R. and Arbetz, M.: Methods of selection. In: Proc. IUFRO Conf. Advanced Generation Breeding. June 14—18, Bordeaux, France. pp. 31—46 (1976). —— Talbert, C. B.: An Analysis of Several Approaches to Multiple-Trait Index Selection in Loblolly Pine (*Pinus taeda L.*). Ph. D Thesis. North Carolina St. Univ., Raleigh, NC. 105 pp., (1984) —— Talbert, C. B.: Multicriterion index selection as a tool for operational tree improvement. In: Proc. IUFRO Conf. on Breeding Theory, Progeny Testing and Seed Orchards. October 13—17, Williamsburg, VA. pp. 228—238 (1986). —— Talbert, C. B. and Lambeth, C. C.: Early testing and multi-stage selection. In: Advanced Generation Breeding of Forest Trees, South Coop. Series Bull. No. 309. Louisiana Agric. Exp. Stn., Baton Rouge, LA 70803. pp. 43—52 (1986). —— Waxler, M. S. and van Buijtenen, J. P.: Early genetic evaluation of loblolly pine.. Can. J. For. Res. **11**, 351—355 (1981). —— White, T. L. and Hodge, G. R.: Best linear prediction of breeding values in a forest tree improvement program. Thor. Appl. Genet. **76**, 719—727 (1988). —— White, T. L. and Hodge, G. R.: Predicting Values with Applications in Forest Tree Improvement. Kluwer Academic Pub., Dordrecht, The Netherlands. 367pp., (1989). —— Williams, C. G.: The influence of shoot ontogeny on juvenile-mature correlations in loblolly pine. Forest Sci. **33**, 411—422 (1987). —— Williams, J. S.: Some statistical properties of a genetic selection index. Biometrika **49**, 325—337 (1962a). —— Williams, J. S.: The evaluation of a selection index. Biometrics **18**, 375—393 (1962b).