Selection Indices using information from Multiple Sources for the Single-Trait Case

By R. D. Burdon
Forest Research Institute,
Private Bag, Rotorua, New Zealand[1])

(Received 17th December 1981)

Summary

Combined sib-family plus individual selection indices can be considered as a special case of using information from more than one class of relatives. In addition, however, ostensibly non-genetic items in a classification can be given weightings in an index, either as built-in corrections for environmental effects, or to utilise genetic effects which are effectively confounded with these items.

A complete selection index is developed for selecting individuals, for a single trait, within a half-sib progeny trial with multi-tree plots and block replicates, and special emphasis is given to the problems of accommodating the information provided by plot means with small numbers of trees per plot. The index is derived both for inputting absolute phenotypic values of individuals and plot means, etc., and for inputting effects (i. e., departures from class and subclass means).

Some incidental points of index construction, namely, finite population corrections and allocation of economic weights, are also covered. Approaches for coping with unbalanced classifications are considered and some suggestions made.

Key words: Selection indices, advanced-generation selection, combined selection, finite population corrections.

Zusammenfassung

Selektionsindizes, welche die Geschwister-, Familien- und Individualleistung zusammenfassen, können als Spezialfall, der die Information von mehr als einer Klasse von Verwandten nutzt, aufgefaßt werden. Es können jedoch zusätzlich auch den scheinbar nicht-genetischen Größen in einer solchen Klassifizierung in einem Index Gewichte gegeben werden, entweder durch die Anbringung von Korrekturen für die Umwelteffekte oder durch die Verwendung genetischer Effekte, die mit diesen Größen vermengt sind.

Es wird hier für die Selektion von Individuen auf der Basis eines Merkmals, innerhalb eines Versuchs mit Halbgeschwister-Familien mit Mehrbaumparzellen und wiederholten Blöcken, ein vollständiger Selektionsindex entwikkelt. Besonderes Gewicht wird auf das Problem der Zuordnung von Information, die aus den Mittelwerten kleiner Parzellen gewonnen wird, gelegt. Der Index ist sowohl für die Verwendung absoluter phänotypischer Werte (Individuen, Parzellenmittel etc.) als auch für die Eingabe von Effekten (z. B. Abweichungen vom Klassen- bzw. Unterklassenmittel) abgeleitet worden.

Einige Nebenpunkte der Indexkonstruktion, namentlich die Korrektur für endliche Populationsgröße und die Zuordnung ökonomischer Gewichte, werden ebenfalls behandelt. Für die Bewerkstelligung nicht ausgewogener Klassifizierungen werden Näherungslösungen in Betracht gezogen und einige Vorschläge gemacht.

Introduction

Combined sib-family plus individual index selection can be considered as a special case of using information from more than one class of relatives (cf. Hazel, 1943; Namkoong, 1966; Becker, 1967); in this case the individual itself is viewed as representing one class and the sib relatives represent another class. In principle, at least, there is no restriction on the number or types of class which might be used to provide information: one could, for instance, include information from both parents and grandparents and/or from both half-sib and full-sib relatives.

Selection indices using information from more than one class of relative may be constructed by regarding the phenotypic value for each class as representing a separate trait. In individual plus half-sib family index selection, for example, the individual tree value for height would effectively be handled as a separate trait from the half-sib family mean for height.

This approach, however, can be generalised to all effects in an experimental classification. An index can be constructed so as to incorporate corrections for non-genetic effects. Moreover, because finite populations must be used in experiments every effect (genetic and non-genetic) in the classification is contaminated by some genetic sampling error, and these genetic error components may become important enough to be worth accommodating in the index. The paper covers the formulation of a selection index so as to accommodate ostensibly non-genetic effects, in the case of selecting individuals for a single trait within a half-sib progeny trial.

Selection indices can be applied to absolute phenotypic values or to effects (deviations from class and subclass means) (Stonecypher and Arbez, 1976). The two approaches can be expected to give equivalent results, but although it is slightly more straightforward to construct an index for effects it is normally more convenient to apply an index to absolute phenotypic values. The selection index formulated in this report is done so both for absolute phenotypic values and for effects.

The underlying purpose of this paper is twofold:

(i) To call attention to how the omission of ostensibly non-genetic information could cause a material loss of efficiency; this would arise when trees of individual progenies are grouped into small plots;

(ii) Incidental to (i), to illustrate various points of index construction which in total are not readily abstracted from the literature.

The progeny trial and the model

The Progeny Trial

Consider a half-sib progeny trial with c families each with one fully randomised plot of n trees in each of q block replicates. I will later focus attention on the case of n being small, both as an illustrative example and as a situation that is likely to be of practical importance.

[1]) Final revision completed at Genetics Department, North Carolina State University, Raleigh, N. C. 27650, U.S.A.

The Linear Model

Applying a linear model, (1) a:

$$Y_{ijk} = \mu + a_j + d_k + (ad)_{jk} + e_{i:jk} \quad \ldots\ldots\ldots\ldots (1)$$

where Y_{ijk} = the absolute phenotype of the i^{th} individual of the j^{th} family in the k^{th} replicate

μ = the general mean

a_j = the effect of the j^{th} family (assumed random)

d_k = the effect of the k^{th} replicate (assumed random)

$(ad)_{jk}$ = (formally) the effect of the interaction between the j^{th} family and the k^{th} replicate, which is actually deemed to be the random environmental effect of the jk^{th} plot

$e_{i:jk}$ = the effect of the i^{th} tree within the jk^{th} plot

Analysis of Variance

The analysis of variance is of the form of (1) b:

Item	Degrees of freedom	Expected mean square
Replicates (Reps)*	q-1	$\sigma_w^2 + n\sigma_p^2 + nc\sigma_r^2$
Families	c-1	$\sigma_w^2 + n\sigma_p^2 + nq\sigma_f^2$
Reps. x families (plots)	(q-1)(c-1)	$\sigma_w^2 + n\sigma_p^2$
Individuals within plots	qc(n-1)	σ_w^2

where σ_w^2 = trees-within-plots variance

σ_p^2 = plots variance

σ_r^2 = replicates variance

and σ_f^2 = families variance

*σ_p^2 could appropriately be omitted from this expected mean square since the plots, whose effects are assumed to dominate the formal interaction, represent 100% of the replicates. This, however, would not affect the development of the selection index and would create complications in later formulations that could obscure some of the points I want to illustrate.

Assuming an additivity + dominance genetic model, and ignoring for the present the finite nature of class and subclass sizes.

$$\sigma_f^2 = 1/4\sigma_A^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)*$$

$$\sigma_w^2 = 3/4\sigma_A^2 + \sigma_D^2 + \sigma_s^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots (3)$$

where σ_A^2 and σ_D^2 are additive and dominance genetic variances respectively and σ_s^2 = within-plot environmental variance.

*For simplicity of presentation I am ignoring the contingency of the parents showing less genetic variation among each other, as a result of selection, than a random population sample, such that $\sigma_f^2 < 1/4\sigma_A^2$ (FINNEY, 1956; NAMKOONG et al., 1966).

The estimation of variance components from actual mean squares is self-evident.

Variances and Covariances of Individuals and Means

Variances of means will show the following expectations.

Variance of family means ($\sigma_{\bar{f}}^2$) = (expected families mean square)/nq

$$= \sigma_f^2 + \frac{\sigma_p^2}{q} + \frac{\sigma_w^2}{nq}$$

$$= 1/4\sigma_A^2 + \frac{\sigma_p^2}{q} + (3/4\sigma_A^2 + \sigma_D^2 + \sigma_s^2)/nq \quad (4)$$

Variance of rep. means ($\sigma_{\bar{r}}^2$) = (expected reps. mean square)/nc

$$= \sigma_r^2 + \frac{\sigma_p^2}{c} + \frac{\sigma_w^2}{nc}$$

$$= \sigma_r^2 + \frac{\sigma_p^2}{c} + (3/4\sigma_A^2 + \sigma_D^2 + \sigma_s^2)/nc \quad \ldots (5)$$

Variance among all plot means, pooling reps, families and plots (or 'interaction') effects, ($\sigma_{\bar{p}'}^2$):

$$\sigma_{\bar{p}'}^2 = \text{(expected plots overall mean square)/n}$$

$$= \sigma_p^2 + \sigma_f^2 + \sigma_r^2 + \frac{\sigma_w^2}{n}$$

$$= \sigma_p^2 + 1/4\sigma_A^2 + \sigma_r^2 + (3/4\sigma_A^2 + \sigma_D^2 + \sigma_s^2)/n \quad \ldots (6)$$

Variance of individuals, pooling all effects ($\sigma_{\bar{e}'}^2$) is given by:

$$\sigma_{\bar{e}'}^2 = \text{expected overall mean square}$$

$$= \sigma_w^2 + \sigma_p^2 + \sigma_f^2 + \sigma_r^2 \quad \ldots\ldots\ldots\ldots\ldots (7)$$

The foregoing variances are formulated assuming a large experiment that has sampled a very large population. The selection index, however, must be used within an actual experiment, so we must consider the variances (V) *within the finite subset that the experiment represents* (cf. FRASER, 1958, Ch. 7, 2.13).

Variance of family means ($V_{\bar{f}}$) can be given by:

$$V_{\bar{f}} = \frac{(c-1)}{c}\left(\frac{\sigma_w^2}{nq} + \frac{\sigma_p^2}{q} + \sigma_f^2\right) \quad \ldots\ldots\ldots\ldots\ldots\ldots (8)$$

(cf Eqn. 4)

Variance of rep. means ($V_{\bar{r}}$) can be given by:

$$V_{\bar{r}} = \frac{(q-1)}{q}\left(\frac{\sigma_w^2}{nc} + \frac{\sigma_p^2}{c} + \sigma_r^2\right) \quad \ldots\ldots\ldots\ldots\ldots\ldots (9)$$

(cf Eqn. 5)

Variance of plot means overall ($V_{\bar{p}'}$) can be given by:

$$V_{\bar{p}'} = \frac{(cq-1)}{cq}\left(\frac{\sigma_w^2}{n} + \sigma_p^2\right) + \frac{(c-1)}{c}\sigma_f^2 + \frac{(q-1)}{q}\sigma_r^2 \quad \ldots\ldots\ldots (10)$$

(cf Eqn. 6)

Variance of individuals overall ($V_{e'}$) can be given by:

$$V_{e'} = \frac{(cqn-1)}{cqn}\sigma_w^2 + \frac{(c-1)(q-1)}{cq}\sigma_p^2 + \frac{(c-1)}{c}\sigma_f^2 + \frac{(q-1)}{q}\sigma_r^2 \quad \ldots\ldots (11)$$

(cf Eqn. 7)

Clearly, the finite population corrections would be of moment only if any of n, c, or q are small. In practice where c is small some errors of estimating variances can become very large, although this does not negate the theoretical principles illustrated here.

The covariances of individual tree values with their family means will be equal to $V_{\bar{f}}$, as will be the *covariances between plot means and their family means.* Likewise, *the covariances of both individual tree values and plot means with their rep. means* will be equal to $V_{\bar{r}}$. The *covariance of individual tree values with their* plot means will be equal to $V_{\bar{p}}$.

However, there is *zero covariance between family means and replicate means* since there is a crossed classification with respect to these items.

Variances of Effects

Considering the variances of the effects (V_a = variance of the families effect, etc.) allowance must be made for the fact that effects are measurable within finite subsets (cf. FRASER, *loc. cit.*) giving the following expectations:

$$V_e = \frac{(n-1)\sigma^2_w}{n} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (12)$$

$$V_{(ad)} = \frac{(c-1)(q-1)(\sigma^2_p + \frac{\sigma^2_w}{n})}{cq} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (13)$$

$$V_a = \frac{(c-1)(\sigma^2_f + \frac{\sigma^2_p}{q} + \frac{\sigma^2_w}{nq})}{c} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (14)$$

$$V_d = \frac{(q-1)(\sigma^2_r + \frac{\sigma^2_p}{c} + \frac{\sigma^2_w}{nc})}{q} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (15)$$

These variances sum to $(\frac{ncq-1}{ncq})\sigma^2_w + (\frac{cq-1}{cq})\sigma^2_p + (\frac{q-1}{q})\sigma^2_r + (\frac{c-1}{c})\sigma^2_f$ . . (16)

Subtracting this from $\sigma^2_w + \sigma^2_p + \sigma^2_f + \sigma^2_r$ gives a closely approximate expectation, under a fully random experimental model, of the sampling variance of the mean (σ^2_μ), namely

$$\sigma^2_\mu = \frac{\sigma^2_w}{cqn} + \frac{\sigma^2_p}{cq} + \frac{\sigma^2_r}{q} + \frac{\sigma^2_f}{c} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (17)$$

A general expression for the expected variance of effects of any item, within a balanced experimental classification, is (sum of squares) ÷ N, N being the grand total number of individuals.

The Index

Using Absolute Phenotypic Values

The full least squares index is for the form (18):

$$I_{ijk} = b_1X_{ijk} + b_2\bar{X}_{jk} + b_3\bar{X}_j + b_4\bar{X}_k \quad \ldots\ldots\ldots\ldots\ldots\ldots (18)$$

where I_{ijk} = the index value of the i^{th} tree of the j^{th} family in the k^{th} replicate

X_{ijk} = absolute phenotypic value of the i^{th} tree of the j^{th} family in the k^{th} replicate.

$\bar{X}_{jk}$ = mean absolute phenotypic value of the plot for the j^{th} family in the k^{th} replicate

$\bar{X}_j$ = mean absolute phenotypic value of the j^{th} family

$\bar{X}_k$ = mean absolute value in the k^{th} replicate

and the b's are the weightings to be given to the respective phenotypic values, and are determined so as to give expected genetic gains of maximum economic worth.

Assuming that the variances and covariances are satisfactorily estimated, it is possible to use matrices to arrive at a set of optimal b's. A general formulation of the solution which also applies to multi-trait selection, is given by Young (1964), this being:

$$\mathbf{b} = \mathbf{P}^{-1}\mathbf{A}\mathbf{a} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (19)$$

where **b** is the vector of appropriate weightings

P is the phenotypic variance-covariance matrix

A is the additive genetic variance-covariance matrix

a is the column vector of economic weights (which will be designated according to the same scheme of subscripting as the b's), such that

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad \ldots\ldots\ldots\ldots (20)$$

The economic weight of 1 for the value of an individual tree is arbitrary. Since we are concerned entirely with the merit of the individual no *economic* weight is attached to the other items, although they can be expected to contribute information giving non-zero solutions for their b's. In fact, multiplication by the **a** vector means that the **Aa** vector product is simply the first column of the **A** matrix.

The form of the matrix **P**, incorporating finite population corrections, is given by:

$$\mathbf{P} = \begin{bmatrix} V_{e'} & V_{\bar{p}'} & V_{\bar{f}} & V_{\bar{r}} \\ V_{\bar{p}'} & V_{\bar{p}'} & V_{\bar{f}} & V_{\bar{r}} \\ V_{\bar{f}} & V_{\bar{f}} & V_{\bar{f}} & 0 \\ V_{\bar{r}} & V_{\bar{r}} & 0 & V_{\bar{r}} \end{bmatrix} \quad \ldots\ldots\ldots\ldots\ldots (21)$$

which can be expanded from Equations (8) to (11).

The form of the corresponding matrix A is given by:

$$\mathbf{A} = \begin{bmatrix} V_{a1} & V_{a2} & V_{a3} & V_{a4} \\ V_{a2} & V_{a2} & V_{a3} & V_{a4} \\ V_{a3} & V_{a3} & V_{a3} & 0 \\ V_{a4} & Va4 & 0 & V_{a4} \end{bmatrix} \quad \ldots\ldots\ldots\ldots\ldots (22)$$

where $V_{a1} = \frac{(ncq-1)}{ncq}3/4\sigma^2_A + \frac{(c-1)}{c}1/4\sigma^2_A$

$$V_{a2} = \frac{(cq-1)}{cq}\frac{3/4\sigma^2_A}{n} + \frac{(c-1)}{c}1/4\sigma^2_A$$

$$V_{a3} = \frac{(c-1)}{c}(\frac{3/4\sigma^2_A}{nq} + 1/4\sigma^2_A)$$

$$V_{a4} = \frac{(q-1)}{q}\frac{3/4\sigma^2_A}{nc}$$

Given that c is large (which, anyway, is a precondition for the trial giving a satisfactory estimate of σ^2_f), and hat q or nq is large (which is necessary for satisfactory estimates of family means), a satisfactory approximation for these matrices will be:

$$\mathbf{P} = \begin{bmatrix} \sigma^2_w + \sigma^2_p + \sigma^2_f + \sigma^2_r & \frac{\sigma^2_w}{n} + \sigma^2_p + \sigma^2_f + \sigma^2_r & \sigma^2_f & \sigma^2_r \\ \frac{\sigma^2_w}{n} + \sigma^2_p + \sigma^2_f + \sigma^2_r & \frac{\sigma^2_w}{n} + \sigma^2_p + \sigma^2_f + \sigma^2_r & \sigma^2_f & \sigma^2_r \\ \sigma^2_f & \sigma^2_f & \sigma^2_f & 0 \\ \sigma^2_r & \sigma^2_r & 0 & \sigma^2_r \end{bmatrix} \quad \ldots\ldots (23)$$

$$\mathbf{A} = \begin{bmatrix} \sigma^2_A & \frac{3/4\sigma^2_A}{n} + 1/4\sigma^2_A & 1/4\sigma^2_A & 0 \\ \frac{3/4\sigma^2_A}{n} + 1/4\sigma^2_A & \frac{3/4\sigma^2_A}{n} + 1/4\sigma^2_A & 1/4\sigma^2_A & 0 \\ 1/4\sigma^2_A & 1/4\sigma^2_A & 1/4\sigma^2_A & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \ldots\ldots (24)$$

The solution will give a negative weight to replicate means in order to give full adjustment for replicate effects. If the phenotypic values have been previously adjusted for replicate effects then replicate arrays will of course vanish from the matrices and the replicate values will not enter into the index. A somewhat more approximate solution, which would also avoid incorporating rep. effects in the actual index, would be to apply the index separately within each replicate.

The weighting given to the means of the plots will represent a balance between correcting for between-plot environmental effects (requiring negative weighting) and using the effective heritability of the plot effects (requiring

positive weighting). An alternative solution, which would avoid incorporating plot effects in the index and which offers potential advantages in coping with imbalance in the experimental classification, would be to adjust individual phenotypic values for the expected environmental component of plot effects. This would effectively remove plot effects from the model, and an adjusted phenotypic value (Y'_{ijk}) would be estimated as:

$$Y'_{ijk} = Y_{ijk} - ad_{jk}(1-h^2_{(ad)}) \quad \ldots\ldots\ldots\ldots\ldots\ldots (25)$$

where $h^2_{(ad)}$ is the heritability of plot effects

$$= \frac{3/4\sigma^2_A/n}{\sigma^2_p + \sigma^2_w/n} = \frac{3/4\sigma^2_A/n}{\sigma^2_p + (3/4\sigma^2_A + \sigma^2_D + \sigma^2_S)/n}$$

Yet another approach, which could also eliminate plot effects from the model, would be to correct phenotypic values for microenvironment effects as estimated by trend surface analysis (cf. JEFFERS, 1975).

Using Effects

The full index is of the form:

$$I^*_{ijk} = b^*_1 e_{i:jk} + b^*_2 (ad)_{jk} + b^*_3 a_j + b^*_4 d_k \ldots\ldots (26)$$

The solution for b*'s is given by (cf Eqn. 19):

$$\mathbf{b^* = (P^*)^{-1} A^* a^*} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (27)$$

where **b*** is the vector of least squares estimates of the appropriate weights

P* is the phenotypic variance-covariance matrix for the effects

A* is the additive genetic variance-covariance matrix for the effects

a* is the vector of economic weights for the effects.

The vector **a*** is composed of identical economic weights for all traits (arbitrarily assumed to equal 1), since the phenotype is specified by the sum of all the component effects.

The matrices are diagonal, because the effects being orthogonal to each other all have zero covariances. The **P*** and **A*** matrices are given by:

$$\mathbf{P^*} = \begin{bmatrix} \frac{(n-1)}{n}\sigma^2_w & 0 & 0 & 0 \\ 0 & \frac{(c-1)(q-1)}{cq}\left(\frac{\sigma^2_w}{n} + \sigma^2_p\right) & 0 & 0 \\ 0 & 0 & \frac{(c-1)}{c}\left(\frac{\sigma^2_w}{nq} + \frac{\sigma^2_p}{q} + \sigma^2_f\right) & 0 \\ 0 & 0 & 0 & \frac{(q-1)}{q}\left(\frac{\sigma^2_w}{nc} + \frac{\sigma^2_p}{c} + \sigma^2_r\right) \end{bmatrix} \quad \ldots (28)$$

$$\mathbf{A^*} = \begin{bmatrix} \frac{(n-1)3/4\sigma^2_A}{n} & 0 & 0 & 0 \\ 0 & \frac{(c-1)(q-1)}{cq}\frac{3/4\sigma^2_A}{n} & 0 & 0 \\ 0 & 0 & \frac{(c-1)}{c}\left(\frac{3/4\sigma^2_A}{nq} + 1/4\sigma^2_A\right) & 0 \\ 0 & 0 & 0 & \frac{(q-1)}{q}\frac{3/4\sigma^2_A}{nc} \end{bmatrix} \quad \ldots (29)$$

Invoking the condition that c at least is large, we have:

$$\mathbf{P^*} \approx \begin{bmatrix} \frac{(n-1)}{n}\sigma^2_w & 0 & 0 & 0 \\ 0 & \frac{\sigma^2_w}{n} + \sigma^2_p & 0 & 0 \\ 0 & 0 & \sigma^2_f & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \ldots\ldots (30)$$

$$\mathbf{A^*} \approx \begin{bmatrix} \frac{(n-1)3/4\sigma^2_A}{n} & 0 & 0 & 0 \\ 0 & \frac{3/4\sigma^2_A}{n} & 0 & 0 \\ 0 & 0 & 1/4\sigma^2_A & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \ldots\ldots (31)$$

With the zero arrays for replicates the solution for b*'s would simplify to the form of:

$$\begin{bmatrix} b^*_1 \\ b^*_2 \\ b^*_3 \end{bmatrix} = \begin{bmatrix} \frac{(n-1)\sigma^2_w}{n} & 0 & 0 \\ 0 & \frac{\sigma^2_w}{n} + \sigma^2_p & 0 \\ 0 & 0 & \sigma^2_f \end{bmatrix} -1 \begin{bmatrix} \frac{(n-1)3/4\sigma^2_A}{n} & 0 & 0 \\ 0 & \frac{3/4\sigma^2_A}{n} & 0 \\ 0 & 0 & 1/4\sigma^2_A \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \quad \ldots (32)$$

which simplifies to $\begin{bmatrix} b^*_1 \\ b^*_2 \\ b^*_3 \end{bmatrix} \approx \begin{bmatrix} h^2_e \\ h^2_{(ad)} \\ h^2_a \end{bmatrix} \quad \ldots\ldots\ldots\ldots (33)$

where h^2_e = heritability of trees-within-plots effects $\approx \frac{3/4\sigma^2_A}{\sigma^2_w}$

$h^2_{(ad)}$ = heritability of plot effects (see Eqn. 25)

h^2_a = heritability of family effects $\approx \frac{1/4\sigma^2_A}{\sigma^2_f + \frac{\sigma^2_p}{q} + \frac{\sigma^2_w}{nq}}$

(which tends to unity with q large (unless $h^2 \to 0$)

$$\text{so } I^*_{ijk} = h^2_e e_{i:jk} + h^2_{(ad)}(ad)_{jk} + h^2_a a_j \ldots\ldots\ldots\ldots\ldots (34)$$

which is just an extension of the well-known solution for single-trait selection using a within- and between-family selection index based on effects (STONECYPHER and ARBEZ, 1976).

Elimination of plot or rep. effects, prior to contructing and applying the index, can be done as indicated for applying the index to absolute phenotypic values.

Discussion

Unbalanced Classification

The formulations given assume a balanced experimental classification, an ideal rather than the usual reality. It is appropriate to consider the problems raised by the imbalance and alternative approaches to overcome it, although no attempt is made to give formulations.

Imbalance causes especial problems in applying this type of index selection. There are several solutions in the literature (which I will not go into) for estimating variance components in unbalanced analysis of variance. These

would leave no special problems with the **A** matrix. Estimates can be made of family means adjusting for unbalanced representation of families among replicates, although such estimates could be much less robust than the actual analysis of variance with respect to heterogeneity of variance structures among replicates.

It is with the **P** matrix, within the framework which I have described, that the worst troubles could arise. Depending on the representation of a family or a plot within the trial the effective phenotypic variances of family or plot means will vary. This would in effect demand a different set of b vectors for each family, according to how it was represented, with the requirement that they give commensurate index ratings to all individuals.

A solution which effectively meets these requirements has been developed and refined for animal breeding situations (*see* HENDERSON, 1975, 1977; THOMPSON, 1979). This is called the Best Linear Unbiased Predictor (BLUP) which gives commensurate estimates of the individual breeding values in complex and unbalanced classifications. However, it has apparently not been applied with forest trees. Its practicability, for this sort of progeny trial, especially with multi-trait selection, remains to be established.

Adjustment of effects according to the representation of the classes and subclasses within the experiment (cf. EFRON and MORRIS, 1977; KUNG, 1979) offers the prospect of giving in effect a phenotypic variance-covariance structure that would apply to the entire experiment. This would be intricate, and is likely to require iterative estimation of the various effects. Reducing family and plot effects according to the heritabilities of the individual effects would, in terms of Equation 27 $\mathbf{[b^* = (P^*)^{-1}A^*a^*]}$, make $\mathbf{P^*}$ and $\mathbf{A^*}$ the same, so the index value for an individual would simplify just to the sum of the reduced effects. Such an approach may in fact be equivalent to BLUP.

Unbalanced representation of families should probably be ignored in deriving index values, unless it is bad enough to cause serious errors, because there seems to be no easy option for actually overcoming the problem arising from the imbalance.

The greatest imbalance is likely to arise at the level of plots, particularly if plot size is small. This might be circumvented by procedures that effectively remove plot effects from the model, thereby confining imbalance to levels in the classification at which the imbalance is minor, although such procedures would probably sacrifice some between-trait covariance information in multi-trait selection. Theoretically at least, one could adjust individual phenotypic values by adapting Equation 25 [Y_{ijk} . . .] such that $h^2_{(ad)}$ is of the form $h^2_{(ad)jk}$, calculated by using n_{jk} (the number of trees in the particular plot) in place of n. Trend surface analysis, although it would have its own complications, should cope with imbalance at this level. Apart from giving possible absolute gains in efficiency it could overcome certain conflicts in layout requirements for efficient selection of families (or parent clones), on the one hand, and individual offspring on the other.

Concluding Remarks

The practical applications of the refinements introduced by this more generalised selection index formulation can be expected to lie in the case of small-plot trials, for which Equations 23 and 24 should be satisfactory approximations. It is with small plots (other than single-tree plots) that plot heritabilities are potentially important, so that ignoring them could incur appreciable inefficiency. At the same time small plots are especially prone to significant imbalance in the classification. Yet tree breeders are now tending to use smaller plots in order to get satisfactory estimates of family means.

In practice the gains in selective efficiency provided by the refinements may not be large compared with the efficiencies of more rough-and-ready procedures. Against that is the consideration that after the first generation of selection the availability of genetic material becomes a limiting factor in the breeding population. Moreover, marginal genetic gains accruing from some analytical refinements could be well worthwhile against the background of very heavy fixed costs in progeny trial establishment and assessment and in seed orchard operations.

The principles illustrated in this paper and a preceding one (BURDON, 1979) should provide the framework for developing selection indices in more complex situations, although the usual caveats (NAMKOONG, 1969; ARBEZ *et al.*, 1974) must still apply.

Acknowledgements

Thanks are extended to Drs. P. P. COTTERILL and M. D. WILCOX for their critical comments on the draft, and to Dr. G. NAMKOONG and Mr. J. H. ROBERDS for helpful discussions. Final revision was facilitated by the author's tenure of a David Henry Scholarship, administered by New Zealand Forest Products Ltd. Also acknowledged are the efforts of Mss. B. MCCUTCHAN and M. GUINON and Mr. M. FREI in providing a Translation of the Summary.

References

ARBEZ, M., BARADAT, PH., MAUGÉ, J. P., MILLIER, C., and BADIA, J.: Some problems related to the use of selection indices in forest tree breeding. Pp. 97—116 *in* Proc. Joint IUFRO Meeting S2.04.01-03, Stockholm, 1974, Royal of Forestry, Stockholm (1974). — BECKER, W. A.: Manual of Procedures in Quantitative Genetics, 2nd Ed. Program in Genetics. Washington State University (1967). — BURDON, R. D.: Generalisation of multi-trait selection indices using progeny test information from several sites. N. Z. J. For. Sci. **9**: 145—52 (1979). — EFRON, B. and MORRIS, C.: Stein's paradox in statistics. Scientific American **236** (5): 119—27 (1977). — FRASER, D. A. S.: Statistics — An Introduction. John Wiley & Sons, N. Y./Chapman & Hall, London (1958). — FINNEY, D. J.: The consequences of selection for a variate subject to errors of measurement. Revue de l'Institut Internationale de Statistique **24**: 1—10 (1956). — HAZEL, L. N.: The genetic basis for constructing selection indices. Genetics **28**: 476—90 (1943). — HENDERSON, C. R.: Best linear unbiased estimation and prediction under a selection model. Biometrics **31**: 432—47 (1975). — HENDERSON, C. R.: Prediction of future records. Pp. 615—38 in Proc. Int. Conf. Quant. Genet., Aug. 16—21, 1976, Ed. E. Pollak, O. Kempthorne, T. B. Bailey. Iowa Sta. Univ. Press, Ames, ISBN 0-8138-1895-8 (1977). — JEFFERS, J. R. N.: An application of trend surface analysis to the interpretation of varietal trials. Institute of Terrestial Ecology, Merlewood Research and Development Paper 63 (1975). — KUNG, F. H.: Improved estimates for provenance breeding values. Silvae Genet. **28**: 114—6 (1979). — NAMKOONG, G.: Family indices for seed orchard selection. Joint Proc. 2nd Genet. Workshop Soc. Am. Foresters and 7th Lake States For. Tree Impr. Conf., USFS Res. Pap. NC-6: 7—12 (1966). — NAMKOONG, G.: Problems of multiple-trait breeding. Paper 7/4 2nd World Consultation on Forest Tree Breeding, Washington, D. C., 7—16 Aug. (1969). — NAMKOONG, G., SNYDER, E. B. and STONECYPHER, R. W.: Heritability and gain concepts for evaluating breeding systems such as seedling orchards. Silvae Genet. **15**: 76—84 (1966). — STONECYPHER, R., and ARBEZ, M.: Methods of Selection. Pp. 31—46 *in* Proc. IUFRO Joint Meeting of Genetic Working Parties on Advanced Generation Breeding, Bordeaux, 1976, INRA Laboratoire d'Amélioration des Conifères — 33610 Cestas, France (1976). — THOMPSON, R.: Sire evaluation. Biometrics **35**: 339—53 (1979). — YOUNG, S. S. Y.: Multi-stage selection for genetic gain. Heredity **19**: 130—45 (1964).