

Choosing Selection Intensities for Seedling Seed Orchards

By GENE NAMKOONG¹⁾

(Received for publication July 5, 1968)

In the establishment of certain seedling seed orchards, a mass selection is made among wild trees, and then wind-pollinated seeds of the selected trees are collected. The progeny are grown in the orchard and rogued according to their own and their families' performance. Those that are retained become seed producers for commercial plantings and are treated to maximize seed yield.

If a certain number of wild trees (N_1) can be reasonably well examined, how many of these should the breeder choose in the first stage of selection and how many of the progeny should he cull in the second, to obtain a given final seed orchard density with maximum gain per unit cost? The answer to this question depends partly on how effectively the mother trees can be selected. If the appropriate heritability for woods selection is low, then selection following seedling testing must be intensive to provide much gain, and hence many mother tree families will have to be brought into seedling tests. If the woods heritability is high, selection can be much more intensive in the woods and the breeder can plan on only a light culling of individual seedlings. The solutions for optimum allocation of selection intensities given by FINNEY (1957) or by HANSON and BRIM (1963) cannot be extended to this kind of selection system or the associated cost functions.

To assess the selection intensities required, it is convenient to assume that the individual seedlings are placed to avoid breeding among sibs of the same family. It would then be necessary to construct breeding units that contain only one tree per family, and to construct as many such units as can be afforded for a given seed-production requirement. The problem attacked in this paper is the number of individuals to be brought into the test-breeding units and the degree to which they should be culled.

Initially, it is desirable to include several sibs per family in each unit. They may be arranged randomly or into family plots within replicated "testing-breeding" blocks with a predetermined number of seedlings per plot and an appropriate number of blocks. If, to avoid inbreeding, only one individual per family is allowed to remain within a breeding block, the effect of any within-family selection is a limited, additional gain constant for all family selection allocations. This gain is maximized by including as many seedlings within families as spacing restrictions permit. Following the initial woods selection of a proportion, say p_1 , of the wild population, each family would be grown in each block on single-tree plots or small family plots. After performance has been assessed and family plots have been thinned to a single tree, a proportion of the plots, say p_2 , in each block will be allowed to produce seed. The final, completely culled orchard is thus conceived as being composed of as many breeding blocks as required for seed production, and with each block containing freely inter-mating, unrelated trees that have survived culling on the

basis of their own as well as their parents' and half-sib family's performance. As the blocks are culled independently, the mixture of families will probably vary from block to block in the final orchard.

The distribution of genetic values, except for the mean, is expected to be largely regenerated by mating the woods trees to produce orchard seedlings, and hence selection in the first stage would not be expected to affect variances of the seedling generation greatly. If selection in the second stage is on the basis of an index, however, there are some complicating effects of first-stage selection on the variances used to derive expected second-stage gains. At woods heritabilities less than 0.8 these would be negligible, and therefore the selection stages can be treated as essentially additive. Total genetic gain is thus formulable as:

$$G = i_1 \sigma_1 h^2_1 + i_2 \sigma_2 h^2_2 \quad (1)$$

where i_1, i_2 = selection differentials of the first and second stages,
 σ_1, σ_2 = standard errors of the tree value in the two stages,
 h^2_1, h^2_2 = heritability "coefficients" applicable to the respective stages.

If the second stage will be based on an index including the mother tree's performance, the open-pollinated and presumably half-sib family mean, and the individual's own performance, the weights for each index item would be calculated as:

$$w = (\sigma_{x_i x_j})^{-1} \sigma_{g x_i} = \begin{pmatrix} \sigma^2_{x_1} & \sigma_{x_1 x_2} & \sigma_{x_1 x_3} \\ \sigma_{x_1 x_2} & \sigma^2_{x_2} & \sigma_{x_2 x_3} \\ \sigma_{x_1 x_3} & \sigma_{x_2 x_3} & \sigma^2_{x_3} \end{pmatrix}^{-1} \begin{pmatrix} \sigma_{g x_1} \\ \sigma_{g x_2} \\ \sigma_{g x_3} \end{pmatrix}$$

where $(\sigma_{x_i x_j})$ = Phenotypic covariance matrix of the three data sources

$\sigma_{g x_i}$ = Vector of covariances between the phenotypic measure (x_i) and genotypic value of the individual (g) (HENDERSON, 1963).

Therefore, the index $I = w' x$, where x = the vector of phenotypic values, and hence $V(I) = w' (\sigma_{x_i x_j}) w$. Since heritability can be defined on this index basis as

$$h^2_I = \frac{\text{Cov}(g, I)}{V(I)}$$

and $\text{Cov}(g, I) = \text{Cov}(g, w' x_i)$
 $= w' \text{Cov}(g, x_i)$
 $= w' (\sigma_{x_i x_j}) w$ since $w = (\sigma_{x_i x_j})^{-1} \text{Cov}(g, x_i)$
 $= V(I)$,

then $h^2_I = 1 = h^2_2$, and $\sigma_2 = [w' (\sigma_{x_i x_j}) w]^{1/2}$.

The values of h_i and σ_{ij} are assumed to be essentially constant in any program.

When the costs of the various stages of the program are computed, a linear model for the total cost may be constructed. That is, the cost for each stage probably would be approximately proportional to the number of trees or families handled in each phase. This assumption is accurate only at high numbers but it will be used here as a working model. The total cost will be proportional to:

$$C_t = N_1 C_1 + N_1 p_1 C_2$$

¹⁾ Geneticist, Institute of Forest Genetics, Southern Forest Experiment Station, USDA Forest Service, and the Department of Genetics and the School of Forestry, North Carolina State University at Raleigh. Paper published with the approval of the Director of Research, North Carolina Agricultural Experiment Station, as No. 2438 of the Journal Series. The computing services were supported by N. I. H. Grant No. FR-00011, held by the Institute of Statistics, North Carolina State University at Raleigh.

where N_1 = the number of wild trees examined,
 p_1 = the proportion of wild trees selected
and C_1, C_2 = the per-unit cost of first- and second-stage tests, respectively.

It may be assumed that the cost for each of the original wild trees examined can range between a low of \$ 0.00004 (ZOBEL, North Carolina State University, personal communication) to \$ 0.03 (SZIKLAI, University of British Columbia, personal communication) and up to as high as \$ 0.50 (USDA Forest Service, unpublished data). The second-stage costs would include the expense of planting and thinning more seedlings than are used for commercial seed production and for measurement and analysis. If the value of extra seedlings for increased precision and intensity of selection or for assurance of sufficient seed is not included, the second-stage costs may vary from a few cents to hundreds of dollars, depending on how many extra seedlings per plot and how many extra family units are planted but not actually used for seed production. Therefore, C_2/C_1 may range from fractions of around $\frac{1}{2}$ up to values in the hundreds or thousands.

From equation (1) and the definitions of index selection gain given above, the expected gain per unit can be formulated as:

$$\frac{\Delta G}{C_t} = (h^2_1 \sigma_1 i_1 + \sigma_2 i_2) \cdot \frac{1}{C_t} \quad (2)$$

with the constraint that $p_1 \cdot p_2 = \text{constant}$ (q) and therefore $q - [p_1 \cdot p_2] = 0$, where p_2 is the fraction retained after the second-stage selection. While this function does not necessarily reflect a true economic evaluation of the cost-return function for breeding, it may serve as a reasonable guide for allocation of breeding effort. More complicated cost-return functions that include time costs are direct extensions of these results.

The cost function is maximized by equating to zero the partial differentiation of $\frac{\Delta G}{C_t} - \lambda (q - p_1 p_2)$ with respect to the culling levels for the two stages. In this equation, $(q - p_1 p_2)$ is the constraint function and λ is the Lagrangian multiplier. This is a standard optimization technique and is detailed in any advanced calculus text. The differentiation results in independent solutions for maximization of the gain/cost function with respect to the allocation of test materials into plots and replications. The latter allocation is optimally determined by equations such as those of MARCUSE (1949). The solution of the above differentiation implies that:

$$i_a \left[i_b - b - h^2_1 \frac{\sigma_1}{\sigma_I} (i_a - a) - \frac{C_2}{C_1 + p_1 C_2} Z_a h^2_1 \frac{\sigma_1}{\sigma_I} \right] = \frac{C_2}{C_1 + p_1 C_2} Z_a i_b \quad (3)$$

where i_a and $i_b = \frac{Z_a}{p_1}$ and $\frac{Z_I}{p_2}$ respectively, and Z_a and Z_I

are the heights of the ordinate of the normal curve at truncation points a and I . The pairs of a and b values to be tested in the equation are graphed in Figure 1 for various levels of final selection intensities. For a final selection proportion of q , the normal deviates would have to be truncated at pairs of a, b values defined by the curves. E. g., for $q = 0.01$, truncation points for the normal deviates $a = 1.5$, and $b = 1.05$ would give $p_1 = 0.067$, $p_2 = 0.147$ and $p_1 p_2 = q = 0.01$.

This solution is satisfactory if the allocation of plant materials does not result in the possibility of significant inbreeding by selection of more than one member of any single family in a breeding block. Since the possibility of inbreed-

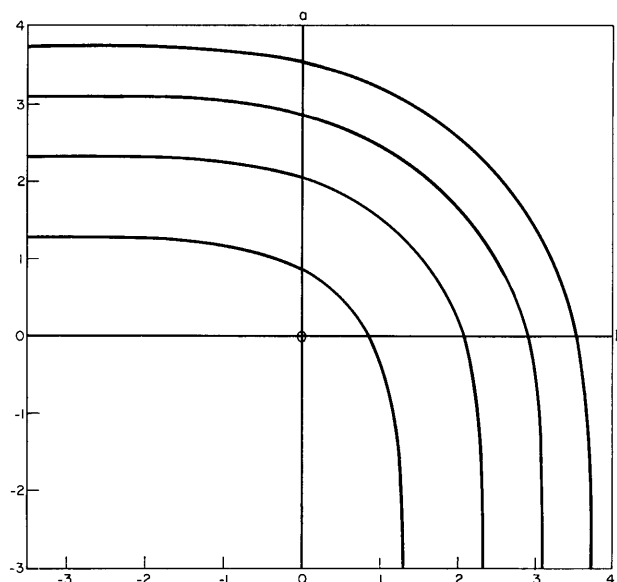


Figure 1. — Truncation points which sequentially produce final selection proportions (q) of 0.1, 0.01, 0.001, 0.0001 in an independent, bivariate normal distribution.

ing depression depends on nonadditive genetic effects, the model is inadequate to determine optimum allocations. If several seedlings are grown in each family plot, it is possible that several trees of the same family will score well on the selection index and if left in the orchard will cause potential inbreeding. An absolute ban on intrafamily mating would prevent such inbreeding; the extra members of the family included for precise evaluation would be culled on the basis of intrafamily comparisons. If such a culling procedure is adopted, a constant selection differential would exist, and one would compare the best individuals in each family plot with others that are the best within their own family plots. The effect of this within-family selection would be constant for all allocations of selection intensities between families.

The solution is therefore directly derivable for any given heritability and cost situation by applying equation (3). The specific solution depends only on the cost ratio and the level of the final selection intensity. To solve the equation, points along the lines drawn in Figure 1 were used and a computer program was written to find all critical values for different values of the cost ratio and h^2_1 . Four curves are drawn in Figure 2 for two heritabilities and two final selection proportions. As might be expected, they clearly show that high test costs force one to select more heavily in the woods at the expense of selection following testing.

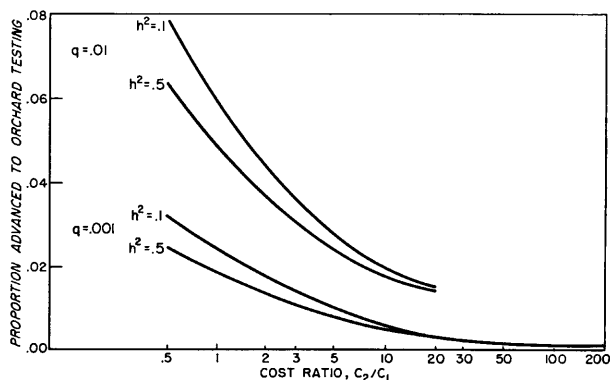


Figure 2. — First-stage selection proportions.

The decrease in the utility of selection following testing is less drastic at high cost ratios, but there are limits in the cost ratio beyond which no testing is feasible for maximizing gain per unit cost. In actuality, sib testing costs are expected to be low and to be offset by seed production benefits; hence, the cost ratios will often be small.

If it becomes feasible to decrease the time to flowering significantly — as by cultivation, fertilization, or irrigation — the precision of the sibling tests may be affected. If no such speeding possibilities exist, the above calculations are sufficient for the situation. However, the breeder may well be faced with choosing between a quicker seed yield and a more precise second-stage selection. A new optimum level of selection will have to be found according to the relative decrease of h^2_1 by the ratio of time factors $\frac{t_2}{t_1}$ or other appropriate economic functions of time costs.

Also, the variance ($\sigma^2_{x_2}$) of the progeny test may be increased by confounding interaction effects caused by the flowering treatments, in which case a heritability factor (h^2_2) would enter into the gain formulation (equation 2) and into the σ_{1_2} part of the optimizing equation (3). Both factors would therefore tend to increase the selection intensity of the first stage.

If the heritabilities listed by CAMPBELL (1964) for various growth traits are a guide, a reasonable heritability for such traits appropriate for woods selection might well be around 0.10. If the progeny test plantings can raise the heritability for family selection to 0.5, the following calculations can be made according to the formulas given previously on the basis of $\sigma^2_{x_1} = 1$:

$$\sigma_{x_1 x_j} = \begin{pmatrix} 1.00 & 0.05 & 0.05 \\ 0.05 & 0.20 & 0.025 \\ 0.05 & 0.025 & 1.00 \end{pmatrix}$$

$$\sigma_{g x_1} = \begin{pmatrix} 0.05 \\ 0.025 \\ 0.100 \end{pmatrix}$$

where x_1 = mother tree phenotypic value
 x_2 = half-sib family mean
 x_3 = individual seedling value

The index ($w'x$) would be:

$$0.040 x_1 + 0.101 x_2 + 0.083 x_3$$

The value of σ_1 would be 0.1188. With these figures inserted into formula (3), the optimum selection proportions for a cost ratio (C_2/C_1) of 1 would be approximately (0.06, 0.17) for first- and second-stage selections, respectively, for a final combined selection intensity of 0.01. If the selection is on the order of 20 out of 2,000, the per-unit gain/cost ratio is 177.

It is instructive to note that the gain/cost ratio for other allocations of selection intensity for these same parameters is not very bad within the neighborhood of the optimum selection. In the breeding situation described above, selection proportions of 0.1, 0.1, respectively, give a gain/cost of 175. Allocations much different, say 0.02, 0.5, do however result in a significant loss in gain/cost, down to 165.

Summary

The allocation of selection intensities in a seedling seed orchard program can be optimized on the basis of gain per unit cost. Ordinary maximization procedures result in equations that can be solved for the proper selection intensities under given heritability and cost conditions for any desired final selection intensity. The solutions indicate that, as cost ratios rise, reliance should be increasingly placed on selection of suitable trees in the woods, rather than on selection following sib testing.

Literature Cited

- CAMPBELL, R. K.: Recommended traits to be improved in a breeding program for Douglas fir. Weyerhaeuser Co. Forest. Res. Note 57, 19 pp. (1964). — FINNEY, D. J.: Statistical problems of plant selection. Bull. de l'Inst. Int. de Statist. 36: 242–268 (1957). — HANSON, W. D., and BRIM, C. A.: Optimum allocation of test material for two-stage testing with an application to evaluation of soybean lines. Crop Sci. 3: 43–49 (1963). — HENDERSON, C. R.: Selection index and expected genetic advance. In: HANSON, W. D., and H. F. ROBINSON (eds.): Symp. on Statistical Genetics and Plant Breeding. Publ. 982, Nat. Acad. Sci.-Nat. Res. Council. 1961: 141–163. Washington, D. C. (1963). — MARCUSE, S.: Optimum allocation and variance components in nested sampling with an application to chemical analysis. Biometrics 5: 189–206 (1949).

Growth Patterns of *Pinus sylvestris* L. Provenances in Minnesota*)

By MUHAMMAD A. K. KHALIL

(Received for publication August 2, 1968)

Introduction

Results of correlation analyses of the phenological data of *Pinus sylvestris* L. provenances grown in Minnesota with the object of determining the nature of variation in the species have been discussed in a separate paper (KHALIL, 1968). The conclusion was made that clinal variation exists from north to south with reference to (1) the number of days after April 15 to the time of the maximum rate of

height growth, (2) the number of days after April 15 to the cessation of height growth, and (3) the total height in early age. The variation was non-clinal with respect to the remaining eight growth characteristics.

Clinal variation from west to east was found only for two characters, viz. the number of days after April 15 to the commencement of height growth and to the time of the maximum rate of height growth. The variation was found to be non-clinal with reference to the remaining nine growth characteristics studied.

Clinal variation with altitude was found to be absent for all the eleven characters studied.

These results show that a large part of the variation in the species is non-clinal, discontinuous or ecotypic, result-

*) This paper is based in part on a dissertation submitted in December 1967 to the Faculty of the Graduate School, University of Minnesota in partial fulfillment of the requirements for the degree of Doctor of Philosophy. — Published as Scientific Journal Series Paper No. 6610 of the University of Minnesota Agricultural Experiment Station.