

Selection Age and Risk: Finding the Compromise

By S. MAGNUSSEN¹⁾ and A. D. YANCHUK²⁾

(Received 1st July 1992)

Abstract

Selection ages for height growth in coastal Douglas-fir on four sites on Vancouver Island were determined by stochastic simulation of age-to-age correlations. Stochastic perturbations of expected age-to-age correlations at both family and within-family levels were derived from bootstrapping more than 4000 records of 23 years of observed height growth. Stochastic dominance principles were used to isolate stochastic efficient (risk averse) selection ages in both family and forward selection schemes. The optimum age for family selection depended strongly on the number of progenies tested in each family and on the test site. Early family selection (age <15) required at least 20 trees per family; very early family selection (age <10) is feasible at a low risk at family sizes above 40. Selection ages in forward selection schemes depend on gain expectations from the within-family selection and on the number of trees selected per family. Reliance on gain from within-family selection pushed the stochastic efficient selection age upwards. Selecting fewer trees per family, but more families, increased the recommended selection ages. Age 17 appears to be minimum 'safe' age for conventional forward selection schemes.

Key words: Early selection, risk, simulation stochastic dominance, bootstrap.

Introduction

The potential of a big pay-off maintains the interest and efforts to reduce selection age in forest tree improvement programs (ZOBEL and TALBERT, 1984). It is now clear that the efficiency of accelerated selection varies according to the material and test conditions in question (FOSTER, 1986; HÜHN, 1987; LAMBETH et al., 1983; STEINHOFF, 1974; WAKELEY, 1971). Two approaches are currently used to enhance the efficiency of early selection: i) the search for juvenile indicators (physiological, morphological, or biochemical) of mature performance (JIANG et al., 1989; BRIDGWATER and WILLIAMS, 1987; LI et al., 1989), and ii) the development of predictive models of age-to-age correlations or future relative performance (LAMBETH, 1980; MAGNUSSEN, 1988, 1989a, 1989b, 1991).

A common feature in these approaches is the absence of an adequate risk assessment. This poses a serious problem inasmuch as the risk increases as the selection age decreases (REHFELDT, 1983). The need for quantifying the uncertainties surrounding a selection option is, therefore, essential for a responsible choice of selection age. Approximations of the possibility of obtaining less gain per unit time in early selections than expected from selections at maturity can be obtained by Monte-Carlo simulations or bootstrapping (EFRON, 1982; RIPLEY, 1987). Combining this possibility with a utility function that evaluates the gain produced (DILLON and HARDAKER 1977) allows a preferential ordering of all the selection options.

¹⁾ Forestry Canada, Petawawa National Forestry Institute, P. O. Box 2000, Chalk River, Ontario K0J 1J0, Canada

²⁾ B. C. Ministry of Forests, Research Branch, 31 Bastion Square, Victoria, B. C. V8W 3E7, Canada

In this study we demonstrate how to quantify the uncertainties of predicted age-to-age correlations by bootstrap techniques and how applications of stochastic dominance principles (WHITMORE and FINDLAY, 1978) can assist in finding a prudent compromise between risk and pay-off. Data for the example were obtained from a coastal Douglas-fir (*Pseudotsuga menziesii* (MIRB.) FRANCO) progeny test growing on four test sites on Vancouver Island (ORR-EWING, 1966; ORR-EWING et al., 1972; ORR-EWING and YEH, 1978).

Modeling genetic gain in early selection

Expected genetic gain in a single mature trait (G_M) arising from selection and subsequent random mating of selected trees can be expressed as the sum of the gains achieved by the implied family and within-family selection (ZOBEL and TALBERT, 1984). The gain equation reads:

$$[1] \quad G_M = i_f \cdot h_f^2 \cdot \sigma_f + i_w \cdot h_w^2 \cdot \sigma_w$$

where i = selection intensity applied to mature trait

h^2 = heritability of mature trait(s) under selection (additive)

σ = phenotypic standard deviation of mature trait

subscripts: f = family, w = within family

In early selection the candidate trees are screened on their expected mature performance, which is their current performance multiplied by the expected age-to-age correlation of the trait under selection (MAGNUSSEN, 1988). Hence, the gain in the mature trait arising from early selection ($G_{J:M}$), can be computed as:

$$[2] \quad G_{J:M} = i_f \cdot r_{f(J:M)} \cdot h_f^2 \cdot \sigma_f + i_w \cdot r_{w(J:M)} \cdot h_w^2 \cdot \sigma_w$$

where $r_{(J:M)}$ = juvenile mature correlation of the trait in question (other symbols as in [1]). Note that $G_{J:M}$ is the predicted mature gain arising from an early selection with identical intensities as those assumed for the selection at maturity.

Expected pay-off from an early selection tree improvement programme is generally expressed as the ratio of the expected achieved annual gain in the mature trait arising from juvenile selection to the corresponding gain arising from selections at maturity. Mathematically this gain ratio GR(J:M) equals:

$$[3] \quad GR(J:M) = \frac{(M + d)}{(J + d)} \cdot (r_f + WGR \cdot r_w) / (1 + WGR)$$

where M = age of maturity

J = age at early selection

d = delay between selection and seed production (here 5 years)

and

$$WGR = \frac{i_w \cdot h_w^2 \cdot \sigma_w}{i_f \cdot h_f^2 \cdot \sigma_f}$$

This simplification is achieved by noting that all parameters except the age-to-age correlations are expected values of the mature population. Note that other equally relevant formulations of the annual gain ratio based upon for example, a net present value principle are possible. They are, however, outside the scope of this paper. Applications of [3] to gauge the benefits of early selections require estimates of $r_f(J:M)$, $r_w(J:M)$, and WGR. Age-to-age correlation models will furnish estimates of the former two parameters while the third parameter WGR must be provided extrinsically. The WGR factor in equation [3] indicates the importance of the gain arising from within-family selection relative to that expected from family selection. The breeder may have some estimates of mature heritabilities and phenotypic standard deviations that, combined with the intended selection intensities, can provide a range of reasonable expectations to the value of WGR. Substitution of a range of likely WGR-values generates a measure of sensitivity of the gain ratio (GR) to changes in WGR. It should be stressed that identical WGR values may cover contrasting scenarios. In the examples given we have used WGR values of 0, 0.2, 0.5, 0.7, 1.0, 1.5, and 2.0 in the belief that most actual WGR values will lie well within these limits.

Equation [3] expresses the expected gain ratio derived from model predictions, and it follows that the gain ratio for fixed M, J, and d is dominated by r_f when WGR is close to zero and by r_w for large WGR values. In any particular early selection event the actual gain ratio will deviate from the expected value (= model prediction) due to stochastic perturbations of the underlying growth process. It is evident from equation [3] that the stochastic uncertainties surrounding GR (J:M) are mainly associated with the stochastic properties of the involved age-to-age correlations. Although the WGR ratio is not known exactly, any error in WGR will not influence the quality of decisions about selection age. Poor estimates of the WGR ratio may introduce an error in the gain ratio (GR). The distribution of this error will be age dependent.

We consider negative stochastic deviations of the gain ratio (GR) from its expectation as the risk of early selection. Quantification of this risk was done by first modelling the probability distributions of both $r_f(J:M)$ and $r_w(J:M)$ and then carrying out a Monte Carlo simulation of an early selection programme. After simulation of the various outcomes (gain ratios) of early selection we ordered the realizations according to their utility (based on some utility function of the gain ratios) to screen out risky prospects and to choose a selection age that offered more protection against adverse results than a selection age that only optimized expectations. Stochastic dominance theory, in conjunction with a general 'universal' utility function, was used to solve this problem.

Selection scenarios

The above principles of finding a stochastically efficient selection age were applied to a series of family and forward (family and within-family) selection schemes for a single trait. Selection for height growth in Douglas-fir will provide examples of the modelling approach. Only information about age-to-age correlations is needed to identify the stochastic efficient selection ages. Height at

ages 3, 6, 11, 14, and 22 were measured in four B.C. Ministry of Forests progeny tests on Vancouver Island (Test sites 13, 15, 16 and 17) from 12 to 14 full-sib families (wide crosses from 1965; ORR-EWING, 1966; ORR-EWING, et al., 1972; ORR-EWING and YEH, 1978) and 2 to 4 open pollinated (half-sib) families per site. These observations served as the data source for the age-to-age correlation prediction models (see Appendix A). The data was assumed to delineate representative genetic entries at a 'family' level and a 'within' family level, even though families are not wholly unrelated.

Selection was carried out between age 1 and age 50 (rotation age) on four sites on Vancouver Island ("San Juan", "Taylor", "Ucluelet", "Oktwanch"). Family selections ($WGR = 0$) were simulated for family sizes of 4, 8, 12, 16, 20, 40, and 60 trees in the progeny test while forward selections ($WGR > 0$) were simulated for the scenarios in which 2 or 3 trees were selected from each of 30 and 45 selected families, respectively (family size and thus i_w remained unconstrained). For each selection age (J), selection design (type, family size, number of selections), and site we drew at random (with replacement) age-to-age correlations from appropriate probability distribution models (of family and within-family correlations, see next section) and calculated the gain ratio GR(J:M). By repeating this process 400 times for each scenario we obtained distributions of gain ratios. These distributions of gain ratios formed the basis for assessing the risk of early selections and finding stochastically efficient selection ages (see section on stochastic dominance). No attempts were made to integrate site effects into an overall choice of selection age. The site effects merely demonstrate their overriding importance when considering a stochastically efficient selection age.

Age-to-age correlations and their stochastic properties

It was argued that gain ratios (GR) in early selections would deviate from the expected ratio due to perturbations of the expected age-to-age correlations. There are two kinds of perturbations: bias in the expectations and random deviations from the bias-adjusted expectations. Both will be unknown but estimates of their probability distribution (around means of zero) can be obtained and used in the assessment of risk. A quantification of the risk (negative deviation from the expected value) in early selection requires a model that encapsulates these perturbations. We can express this stochastic process in the following way:

$$[4] \quad r(J:M) = \hat{r}(J:M) + \varepsilon(\hat{r}) + \varepsilon(J:M)$$

where $r(J:M)$ = realized (actual) age-age correlation

$\hat{r}(J:M)$ = expected age-age correlation estimated from models (see equation [5])

$\varepsilon(\hat{r})$ = random perturbations due to errors in the parameters of the prediction model (bias)

$\varepsilon(J:M)$ = random perturbations of bias corrected expectations ($r(J:M) + \varepsilon(r)$)

In the following we will describe how we arrived at quantifying $\hat{r}(J:M)$, $\varepsilon(\hat{r})$, $\varepsilon(J:M)$, and thus $r(J:M)$.

Quantifying $\hat{r}(J:M)$

Several commonly used age-to-age correlation models (BURDON, 1989; LAMBETH, 1980; MAGNUSSEN, 1989b) were tested but rejected due to poor fit on one or more sites or in certain age ranges. Lack of model flexibility was the main

cause for their poor performance. Instead, two new non-linear models were developed. Expected age-to-age correlations ($\hat{r}(J:M)$) were derived from two non-linear regression models:

$$[5] \quad \hat{r}_f(J:M) = \left(1 + \frac{M-J}{J \hat{\alpha}_f} \right) \hat{\beta}_f / n_f \hat{\delta}_f$$

$$\hat{r}_w(J:M) = \left(1 + \frac{M-J}{J \hat{\alpha}_w} \right) \hat{\beta}_w$$

where n_f = family size (trees per family)
 $\hat{\alpha}, \hat{\beta}, \hat{\delta}$ = estimated regression coefficients
 $M = 50, J = 1, 2, \dots, 45$
subscripts f = family, w = within family

The 'built-in' autocorrelation arising from the fact that the same tree is measured on 2 occasions (i. e. the height at age J is part of the height at age M) is captured by the α parameter. As J increases the autocorrelation will increase for any given value of $M-J$ simply because more height is shared ($\alpha > 1$). The parameters $\alpha, \beta,$ and δ in [5] were estimated by non-linear least-squares regression using 10 unique pairs of J and $M-J$; [from the observed heights we constructed the following age pairs: $(J, M-J) = \{ (4, 3); (4, 8); (4, 14); (4, 19); (7, 5); (7, 11); (7, 16); (12, 6); (12, 11);$

(18, 5) }. Correlations for age combinations outside these datum points were obtained as model predictions. The site specific dependence of r_f on n_f was explored by generating 1000 random bootstrap samples (EFRON, 1982) of n_f -observed records (trees) of height growth from each family on a site ($n_f = 1, 2, 3, 4, 5, 10, 20, 40, 60$) and then calculating the age-to-age correlations of the family means prior to the non-linear estimation of the regression parameters. The fact that some families had fewer trees than sampled by the bootstrap poses no problem for the statistical properties of the results due to sampling with replacement (EFRON, 1982). Estimated parameters of age-to-age correlation models of family means are listed separately for each site in table 1. Inspection of residuals revealed no particular bias in the overall fit. A single model for the four sites was not tenable due to statistically significant differences among regression parameters obtained at different sites (F-test, $P < 0.001$). A graphical interpretation of the model for the Ucluelet site is given in figure 1. The effect of sample size and autocorrelation ($r \rightarrow 1$ for $J \rightarrow \infty$ regardless of the magnitude of $M-J$) is quite evident. It is also clear that $r_f \rightarrow r_{\text{genetic}}$ for $n_f \rightarrow \infty$ due to the removal of environmental effects (r_{genetic} is the additive genetic correlation).

The within-family age-to-age correlation (\hat{r}_w) could be predicted [via the second model in (5) with an error of less than 5% of the correlations derived from the height measurements. Regression models were fitted individually

Table 1. — Estimated regression parameters and asymptotic statistics of the model

$$\hat{r}_f = \left[1 + \frac{(M-J)}{J \hat{\alpha}_f} \right] \hat{\beta}_f / n_f \hat{\delta}_f$$

Parameter	San Juan	Taylor	Ucluelet	Oktwanch
$\hat{\alpha}_f$	1.4074	1.0000	1.1031	1.0000
$\hat{\beta}_f$	-0.81158	-0.5971	-0.5157	-0.6000
$\hat{\delta}_f$	0.0668	0.1097	0.1647	0.2000
$\hat{\sigma}^2(\hat{\alpha}_f)$	$6.6594 \cdot 10^{-4}$	$8.6816 \cdot 10^{-3}$	$2.6272 \cdot 10^{-3}$	$6.7379 \cdot 10^{-3}$
$\hat{\sigma}^2(\hat{\beta}_f)$	$8.4497 \cdot 10^{-4}$	$5.1729 \cdot 10^{-3}$	$1.1597 \cdot 10^{-3}$	$4.1847 \cdot 10^{-3}$
$\hat{\sigma}^2(\hat{\delta}_f)$	$6.8549 \cdot 10^{-5}$	$8.5500 \cdot 10^{-4}$	$2.9387 \cdot 10^{-4}$	$6.9961 \cdot 10^{-4}$
$\hat{\sigma}(\hat{\alpha}_f, \hat{\beta}_f)$	$-6.3817 \cdot 10^{-4}$	$-5.6673 \cdot 10^{-3}$	$-1.4697 \cdot 10^{-3}$	$-4.5886 \cdot 10^{-3}$
$\hat{\sigma}(\hat{\alpha}_f, \hat{\delta}_f)$	$2.7840 \cdot 10^{-5}$	$1.6910 \cdot 10^{-4}$	$4.7375 \cdot 10^{-5}$	$2.5725 \cdot 10^{-4}$
$\hat{\sigma}(\hat{\beta}_f, \hat{\delta}_f)$	$-1.1033 \cdot 10^{-4}$	$-1.0287 \cdot 10^{-3}$	$-2.7903 \cdot 10^{-4}$	$-8.6593 \cdot 10^{-4}$
$\hat{\sigma}_{\text{res}}^2$	$6.4697 \cdot 10^{-4}$	$8.6434 \cdot 10^{-3}$	$1.970 \cdot 10^{-3}$	$5.2842 \cdot 10^{-3}$
R^2	0.9802	0.7447	0.9034	0.7822

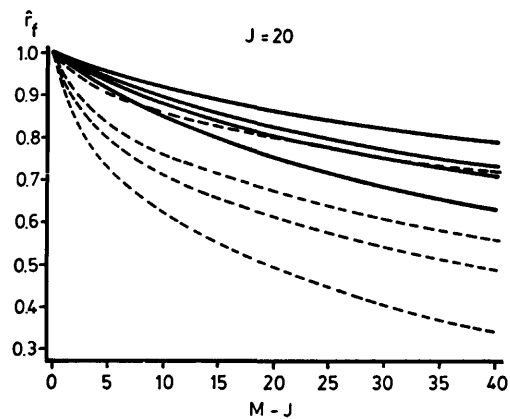
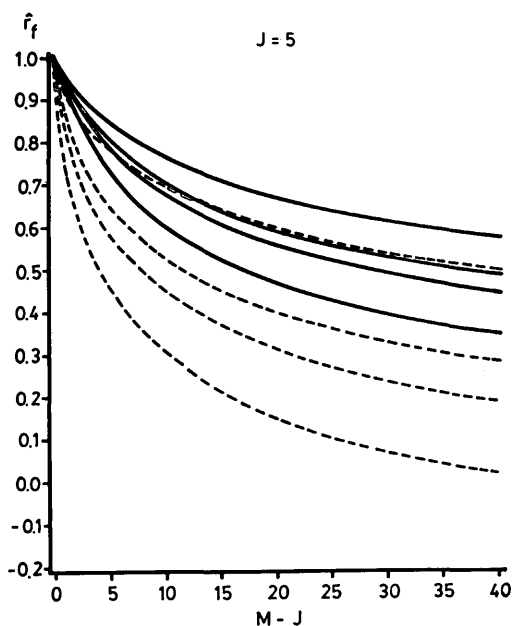


Figure 1. — Predicted family age-to-age correlations at the Ucluelet site. Full lines are mean correlations based on (top to bottom) 50, 10, 5, and 1 tree per family. Dashed lines are the corresponding lower 5 percentiles of the underlying distribution of the correlations (estimated from the 1000 bootstrap estimates).

to all the 66 combinations of family x site in the original data set because significant differences among the correlations from different families and/or sites were manifest (Table 2). Site averages of the within-family correlation models are listed in table 3, together with the among-family standard deviation of the regression parameters. Statistically significant differences existed among parameters from different sites (t-test, all $P < 0.10$). An impression of the dispersion of family-specific within-family age-to-age correlations at the Ucluelet site is provided in figure 2, in which the average within-family correlations are plotted along with the curves for the families with the highest and the lowest within-family correlations for

any given $M-J$ value. Within family autocorrelation effects (\hat{a}_w) are visualized for a single family at the Ucluelet site in figure 3.

Quantifying $\varepsilon(\hat{r})$

Random perturbations due to errors in the parameters of the prediction model (\hat{r}) were generated by assuming that the regression parameters in [5] were multivariate normal distributed with a variance-covariance structure equal to the asymptotic estimates derived from the non-linear regressions (cf. Table 1 and 3). A Cholesky decomposition of the estimated covariance matrix of the regression parameters multiplied to the drawn vector of random independent and identical distributed ($N(0,1)$) vectors (RIPLEY, 1987) created the needed multivariate perturbations $\varepsilon(\hat{r})$.

Table 2. — Homogeneity test of within-family age-to-age correlations (r_w). Test: BARTLETT'S U test ON FISHER'S z-transformed and bias corrected observed age-to-age correlations (SNEDECOR and COCHRAN, 1971).

J, (M-J)	San Juan	Taylor	Ucluelet	Oktwanch
4,3	0.00	0.09	0.11	0.00
4,8	0.00	0.63	0.02	0.07
4,14	0.02	0.48	0.21	0.40
4,19	0.00	0.03	0.35	0.21
7,5	0.00	0.88	0.00	0.33
7,11	0.10	0.38	0.00	0.06
7,16	0.02	0.04	0.01	0.02
12,6	0.00	0.24	0.18	0.08
12,11	0.00	0.28	0.02	0.24
18,5	0.16	0.03	0.00	0.44

Table entries: Probability that all age-to-age correlations are equal ($n_w > 40$ in all crosses).

Table 3. — Average within family age-to-age regression parameters of the model $\hat{r}_w = (1 + \frac{\beta_w}{J\hat{\alpha}_w})^{M-J}$

	San Juan	Taylor	Ucluelet	Oktwanch
$\hat{\alpha}_w$	1.1908	1.3429	1.0440	0.8713
$\hat{\beta}_w$	-0.7275	-0.9895	-0.5956	-0.5845
$\hat{\sigma}_1^2(\hat{\alpha}_w)$	0.0220	0.0417	0.0253	0.0678
$\hat{\sigma}_1^2(\hat{\beta}_w)$	0.0278	0.0586	0.0268	0.0313
$\hat{\sigma}_1(\hat{\alpha}_w, \hat{\beta}_w)$	-0.0201	-0.0412	-0.0210	-0.0352
$\hat{\sigma}_2^2(\hat{\alpha}_w)$	0.0834	0.0881	0.0457	0.0467
$\hat{\sigma}_2^2(\hat{\beta}_w)$	0.1279	0.2474	0.0835	0.0975
$\hat{\sigma}_{res}^2$	0.0233	0.0477	0.0313	0.0795
\bar{R}^2	0.9077	0.8456	0.8801	0.7072

$\hat{\sigma}_1^2$ = avg. variance of regression estimates
 $\hat{\sigma}_2^2$ = among family variance of regression parameter.

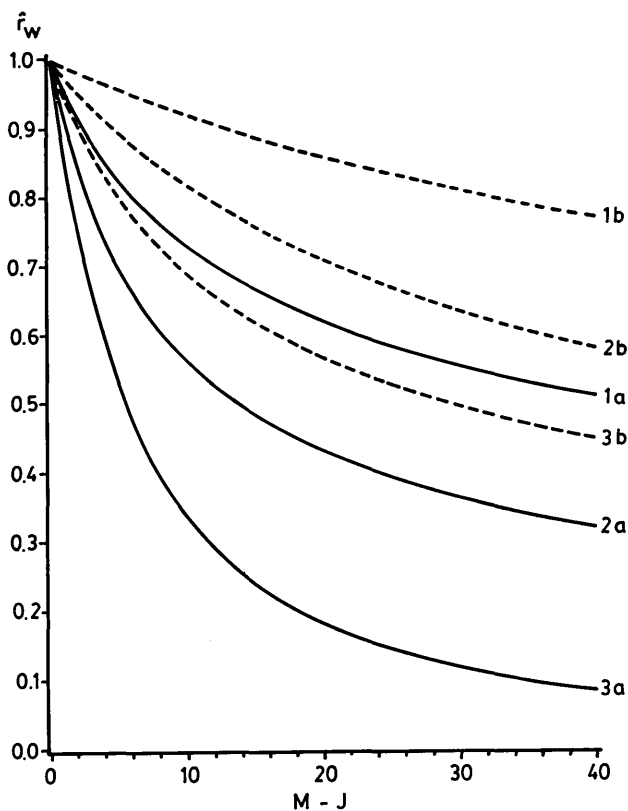


Figure 2. — Within-family age-to-age correlations at the Ucluelet site

J = younger age, M = older age, M - J = time lag.
 1 = family with maximum within-family age-to-age correlations;
 2 = average within-family age-to-age correlations;
 3 = family with minimum within-family age-to-age correlations.
 a J = 5 years; b J = 20 years.

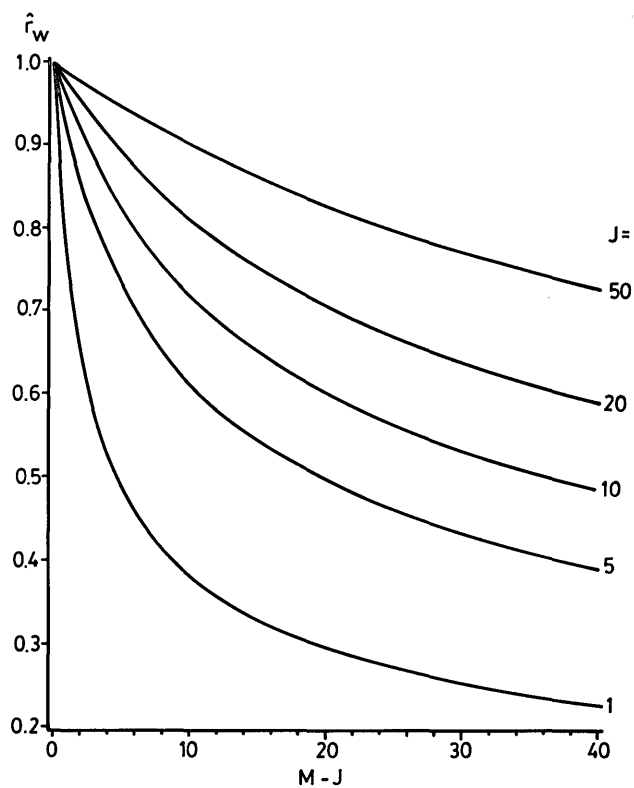


Figure 3. — Within-family age-to-age correlations curves for different initial age (J). The given example is family 492 x AR on the Ucluelet site. J = younger age, M = older age, M - J = time lag.

Quantifying $r(J:M)$

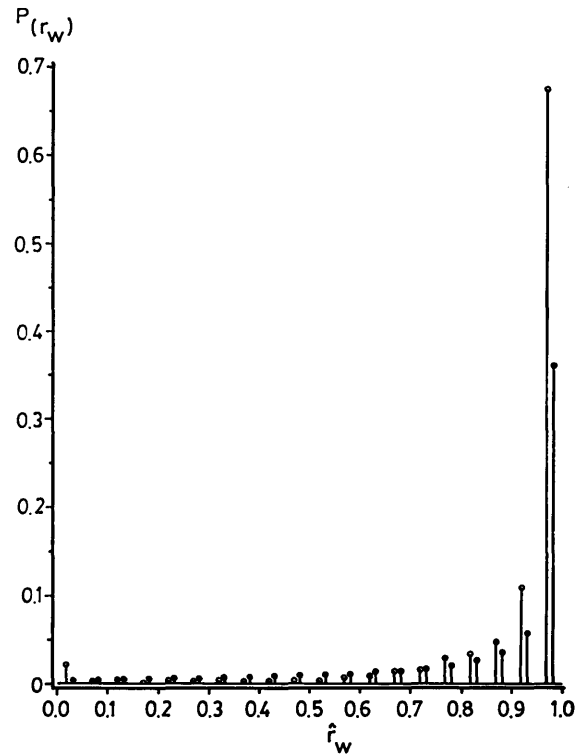
The realized age-to-age correlation ($r(J:M)$) for a single selection scenario was found by random draws of uniformly distributed numbers u_i ($0 \leq u_i < 1.0$). For a given u_i

we found the correlation $r(J:M)$ for which the underlying cumulative distribution function (cdf) for the considered selection scenario and bias-corrected expectation ($=\hat{r}(J:M) + \varepsilon(\hat{r})$) takes the value u_i . Generalized beta distributions generated the needed cumulative distribution functions. An earlier study (MAGNUSSEN, 1990) indicated that this type of distribution is suitable for age-to-age correlations. Parameters needed to determine these beta distributions were the mean ($=\hat{r}(J:M) + \varepsilon(\hat{r})$), the maximum (r_{max}), the minimum (r_{min}), and the variance ($\hat{\sigma}^2(r(J:M) + \varepsilon(\hat{r}))$) of the underlying distributions (JOHNSON and KORZ, 1971, p. 44).

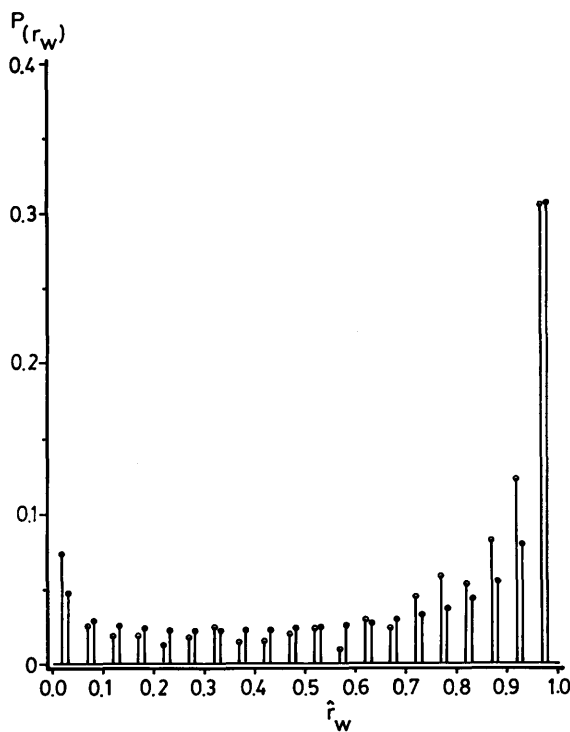
Formulae for calculating the beta-distribution parameters are given in Appendix B. Bootstrapping of observed height records (with replacements) and calculations of age-to-age correlations (r_t or r_w) after each cycle of bootstrapping (1000 cycles per scenario and age) generated empirical probability distributions of age-to-age correlations. These bootstrap generated distributions furnished support for the beta model and they provided the necessary input parameters for the generalized beta distribution.

An impression of how well the expected beta distribution followed the empirical bootstrap distributions of age-to-age correlations is provided in figures 4a to f. Overall, the agreement between the actual bootstrap distributions and what was expected was satisfactory, as illustrated in the statistical summary of the KOLMOGOROV-SMIRNOV tests (Figure 5). The agreement between the bootstrap-generated and expected distributions was especially good for the family-mean correlation, where 80% of the K_D values were below the 5% critical value under the null hypothesis of no difference and about 97% were below the 1% critical value. At the within-family correlation level the agreement was less tight, although the majority (56%) of the distributions were deemed equal to their beta expectations

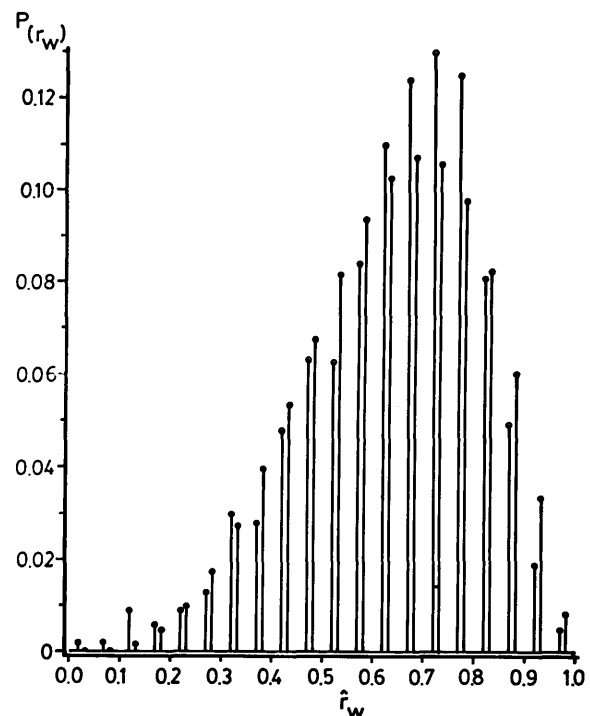
at the 5% risk of a type I error. The few major discrepancies ($K_D > 0.2$) could always be traced to a lack of fit in the two uppermost correlation classes in inverse J-shaped distributions ($E(\hat{r}) > 0.8$). No bias of practical significance arose from this fitting problem. It was concluded that the provided framework allowed a realistic estimation of the



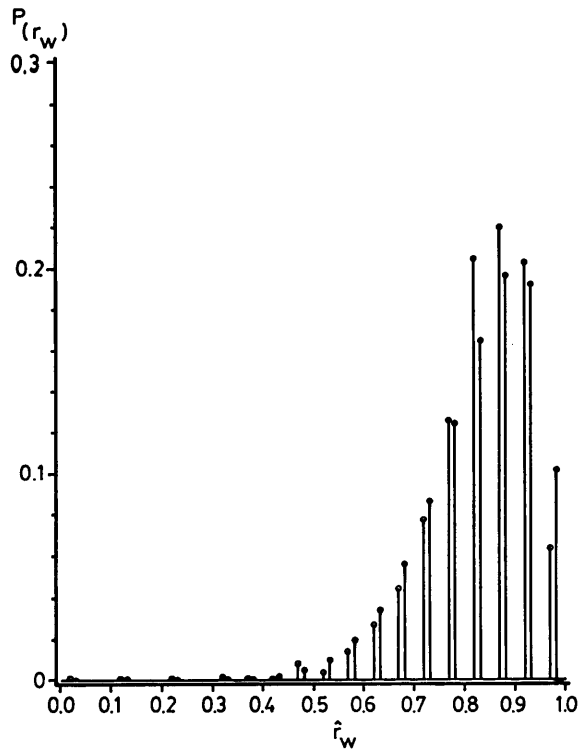
b) Within-family age-to-age correlations (\hat{r}_w). Family = '492xAR', $J=18$, $M=23$, $M-J=5$, $n_w=3$, site = Ucluelet.



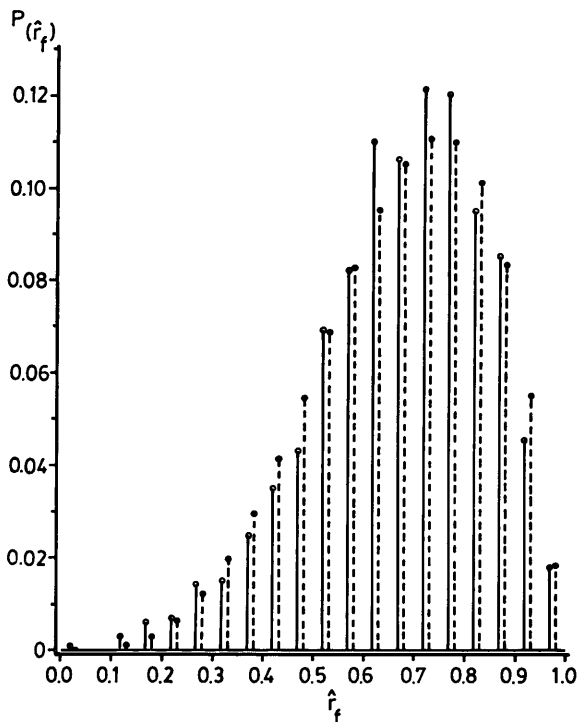
a) Within-family age-to-age correlations (\hat{r}_w). Family = '492xAR', $J=4$, $M=23$, $M-J=19$, $n_w=3$, site = Ucluelet.



c) Within-family age-to-age correlations (\hat{r}_w). Family = '492xAR', $J=4$, $M=23$, $M-J=19$, $n_w=20$, site = Ucluelet.



d) Within-family age-to-age correlations (\hat{r}_w). Family = '492xAR', $J=18$, $M=23$, $M-J=19$, $n_w=20$, site = Ucluelet.

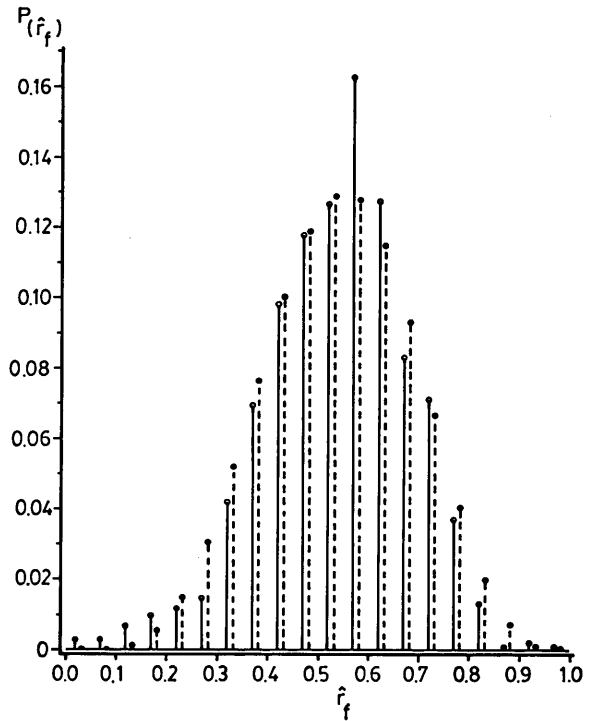


e) Family mean age-to-age correlations (\hat{r}_f). $J=4$, $M=23$, $M-J=19$, site = Ucluelet, $n_f=1$ (trees per family).

functional form of the probability distributions of age-to-age correlations (\hat{r}_f and \hat{r}_w).

Stochastic dominance

The Monte Carlo technique described above gave each selection age (J) and selection scenario a probability dis-



f) Family mean age-to-age correlations (\hat{r}_f). $J=4$, $M=23$, $M-J=19$, site = Ucluelet, $n_f=40$ (trees per family).

Figure 4. — Bootstrap-generated and "predicted" distributions of age-to-age correlations. Predictions were generated from a beta distribution model with parameters p and q determined via models in table 4.

$P(\bullet)$ = probability that a generated age-to-age correlation falls into one of 20 correlation classes (width = 0.05).

"Observed" = open circles (leftmost in each pair); "Predicted" = filled circles (rightmost in each pair).

All distributions have been scaled to $[0,1]$ via the formula

$$r = \frac{(r - r_{\min})}{(r_{\max} - r_{\min})}$$

tribution of gain ratios (GR), say $P_J(\text{GR})$. The mean gain ratio ($\overline{\text{GR}}_J$) for selection at age J is computed as:

[6]

$$\overline{\text{GR}}_J = \int_{-\infty}^{\infty} \text{GR} \cdot P_J(\text{GR}) \cdot d\text{GR}, \text{ for } J = 1, 2, 3, \dots, M$$

A preference for the selection age with the highest mean gain ratio ($\overline{\text{GR}}_J$) is only rational if all potentially possible gain ratios³⁾ have the same utility (value) to the breeder. A more realistic approach is to specify a utility function $U(\text{GR})$ that equates a gain ratio (GR) with a value to the breeder, and to calculate the average utility (\overline{U}_J) of all possible gain ratios for the selection age under consideration. Specifically

[7]

$$\overline{U}_J = \int_{-\infty}^{\infty} U(\text{GR}) \cdot P_J(\text{GR}) \cdot d\text{GR}, \text{ for } J = 1, 2, 3, \dots, M$$

Preference for the selection age (J) that maximizes \overline{U}_J gives, on an average, the highest utility (value) to the

³⁾ The range of possible gain ratios can be obtained by setting $r(J:M) = \pm 1$ in equation [3].

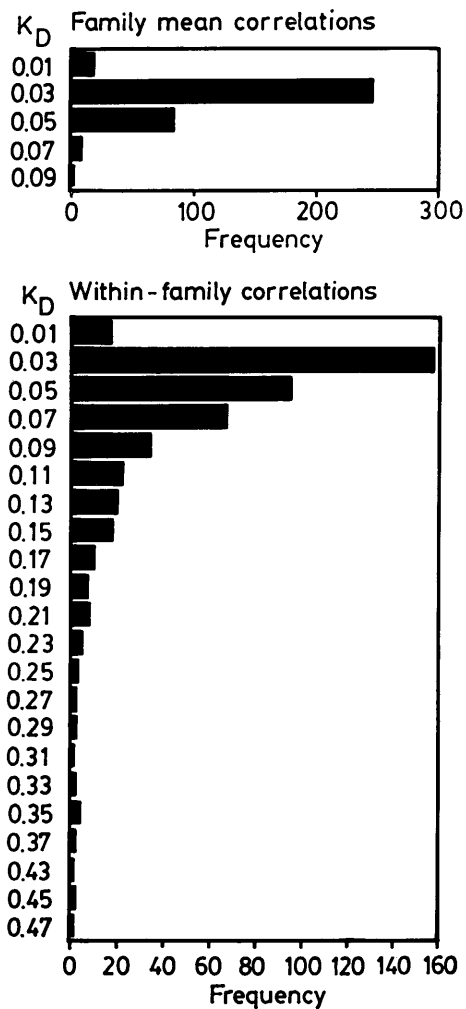


Figure 5. — Frequency distributions of K_D (the maximum absolute difference between the observed and expected distribution) in the Kolmogorov-Smirnov two-sample test of the hypothesis $H_0 =$ the bootstrap generated distribution of r_t and r_w follows a beta distribution with parameters p , and q determined from the minimum, maximum, and second moment of the simulated distribution. Critical ($P = 0.05$) value (for $n_{sim} = 1000$) = 0.062. (SIEGEL, 1956).

breeder. Figure 6 portrays one possible utility function ($U(GR)$).

Often, however, exact knowledge of the utility function of a gain ratio is rare (ANDERSON, 1974; MENZ, 1980; POPE and ZIEMER, 1984) which impedes the use of [7] as a means for finding the most efficient selection age. Instead of exact knowledge of $U(GR)$ we shall assume some general 'universal' properties of $U(GR)$. Let GR and GR' be two different gain ratios, then:

- 1) if $GR > GR'$ then $U(GR) > U(GR')$ (first derivative of U is positive, i.e. $U' = \partial U / \partial GR > 0$);
- 2) the utility of a unit increase in gain ratio (GR) decreases with the magnitude of the gain ratio (i.e. it is more important to improve GR from, say, 0 to 1 than it is to improve GR from 7 to 8). In other words the utility function is convex. $U'' = \partial^2 U / \partial GR^2 < 0$;
- 3) negative implications arising from low or even negative gain ratios accelerates as the gain ratio decreases further (i.e., $U''' = \partial^3 U / \partial GR^3 > 0$).

The utility function in figure 6 has these 3 'universal' properties. By accepting these ($U' > 0$, $U'' < 0$, and $U''' > 0$)

properties as inherent to the utility function U in tree improvement we can develop stochastically efficient criteria for choosing a selection age that is an optimal balance of gain and risk for all utility functions with the abovementioned 'universal' properties. Three sequential stochastic dominance criteria (DILLON and HARDAKER, 1977) will be used to identify selection ages that maximizes gain for the underlying 'universal' utility function and the relevant probability distribution of gain ratios (GR).

1) Selection age J is preferred over selection age K if

$$\int_{-\infty}^{GR} P_J(GR) \cdot dGR \leq \int_{-\infty}^{GR} P_K(GR) \cdot dGR$$

for all gain ratios (GR) with a positive probability of realization. We say that selection age J is first degree stochastic dominant over selection age K . This criterion is logical. If selection age J is first degree stochastically dominant over selection age K there will be a lower probability of lower gain ratios and a higher probability of higher gain ratios by opting for selection age J rather than selection age K . All selection ages and all selection scenarios were scanned for first degree stochastically efficient selection ages. When no first degree stochastically efficient selection age could be identified (i.e., the cdf's are crossing) we proceeded to separate selection ages based on the criteria of second degree stochastic dominance.

2) Selection age J is preferred over selection age K if

$$\int_{-\infty}^{GR} \int_{-\infty}^{GR} P_J(GR) \cdot dGR \cdot dGR \leq \int_{-\infty}^{GR} \int_{-\infty}^{GR} P_K(GR) \cdot dGR \cdot dGR$$

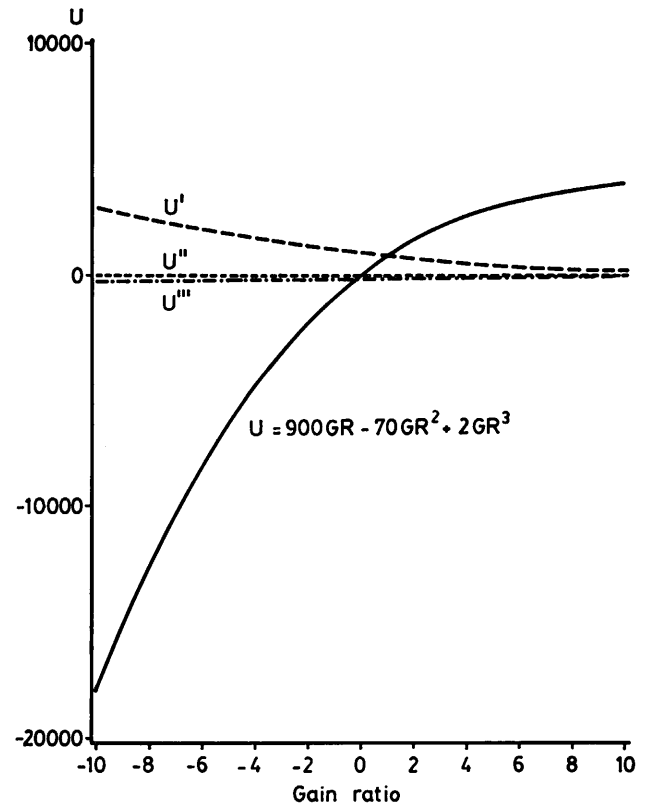


Figure 6. — An example of an utility function (U) with the desired 'universal' properties ($U = \partial U / \partial GR > 0$, $U'' = \partial^2 U / \partial GR^2 < 0$, $U''' = \partial^3 U / \partial GR^3 > 0$)

for all gain ratios (GR). We say that selection age J is second degree stochastically dominant over selection age K. This second (sequential) criterion compares the risk and the gain associated with selection ages J and K. Because neither age dominates the other in the first degree, the two cumulative probability functions of gain ratios must intersect at least once. Intersecting cumulative probability functions indicates that one selection age has a higher potential of generating both relatively better and worse results than the other. By integrating the cumulative probabilities one balances the attractive features (higher gain) against the negative aspects (less gain) of choosing one selection age over the other. Whenever a subset of selection ages was declared second degree stochastic dominant we used the 'Mean-Gini' algorithm (BUCCOLA and SUBAEL, 1984) to find the best selection age in the subset. This algorithm compares the average dispersal adjusted means ($\overline{GR}_{J:M} - \Gamma_{J:M}$), where $\Gamma_{J:M}$ is the average dispersal of gain ratios at selection age J) in the subset. The highest "risk-adjusted" gain-ratio was preferred. When no selection age emerged as second degree stochastic dominant then a search for third degree dominant selection age was carried out (see 3).

3) Selection age J is preferred over selection age K if

$$\int_{-\infty}^{GR} \int_{-\infty}^{GR} \int_{-\infty}^{GR} P_J(GR) \cdot dGR \cdot dGR \cdot dGR \leq \int_{-\infty}^{GR} \int_{-\infty}^{GR} \int_{-\infty}^{GR} P_K(GR) \cdot dGR \cdot dGR \cdot dGR$$

for all gain ratios (GR). We say that selection age J is third degree stochastic dominant over selection age K. This third criteria for choosing a selection age is arrived at by simply repeating the arguments and rationale for second degree stochastic dominance (criteria 2). Further separation could proceed by iterated integration of the underlying cumulative probability functions; however, all decisions regarding selection age in the presented results could be made based on 1 of the 3 degrees (1 to 3) of stochastic dominance.

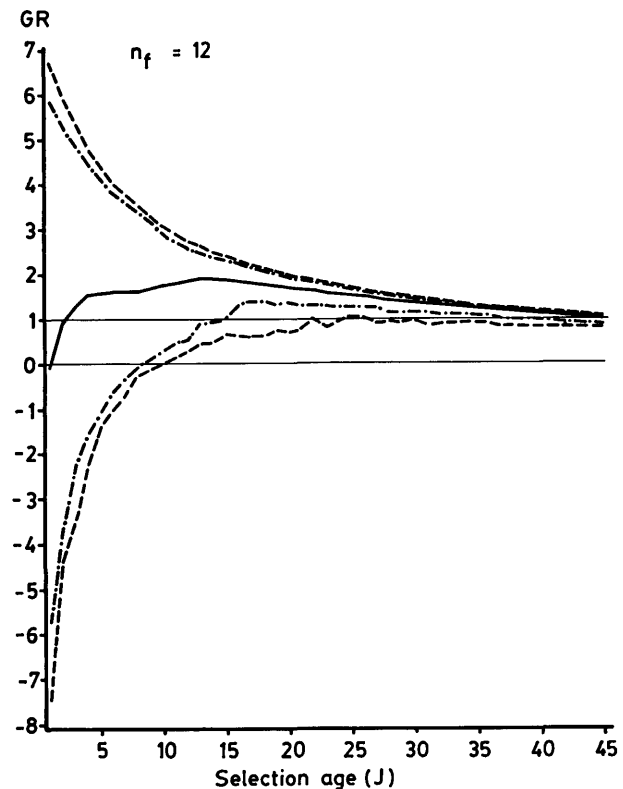
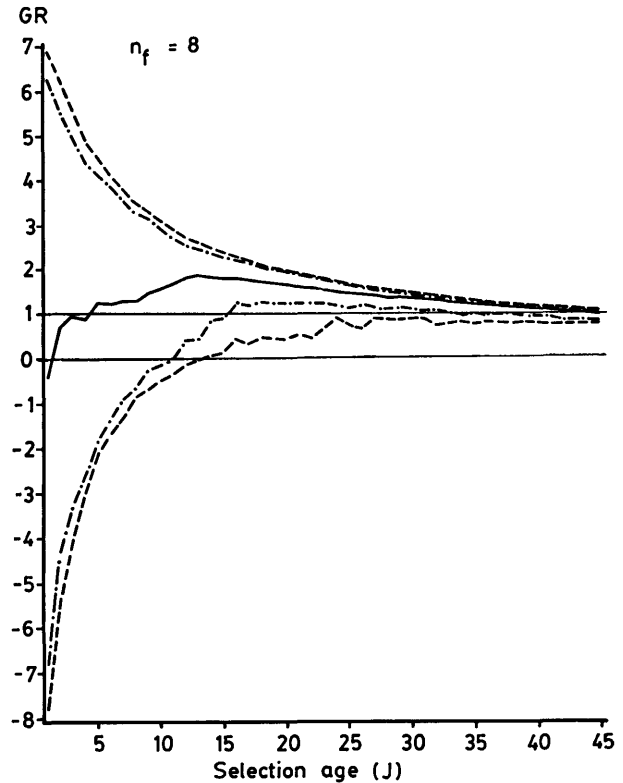
Results

Family selections (WGR = 0)

Four hundred simulations of early family selections based on the prediction model of \hat{r}_f in [5] plus generated stochastic perturbations of model parameters $\varepsilon(\hat{r})$ and random perturbations (ε) were computed for each of the four sites and for family sizes (n_f) of 4, 8, 12, 16, 20, 40, and 60. Based on the three ordering criteria of stochastic dominance a stochastically efficient selection age could be found for each of the above combinations of site and n_f . A graphical image of the simulation output is displayed in figure 7 for $n_f = 8, 12, 20,$ and 40 on the Ucluelet site. The curve for the average gain ratio in the simulations assumes a maximum towards younger ages as the family sample size (n_f) increases. A distinct peak in the gain ratio was only apparent for $n_f > 40$. At lower family sizes the best average selection age could be anywhere within a plateau of 5 to 10 years. It was also apparent that the magnitude of positive and negative (risk) deviations from the expected gain ratio declined exponentially with selection age. For this reason no first degree stochastic efficient selection age could be identified in any selection scenario; however, all selection scenarios contained a single most efficient (2nd degree stochastic dominant) selection age. For small family sizes ($n_f < 20$) a real risk of negative

gain ratios (at Ucluelet) persisted up to an age of about 10 years.

For those decision makers guided by worst case scenarios (the curve of minimum gain ratio (GR)), early selection is only an option for $n_f \geq 40$. Trends in selection ages based on optimum mean (no utility function) criteria, stochastic dominance (2nd degree), and the optimum lower



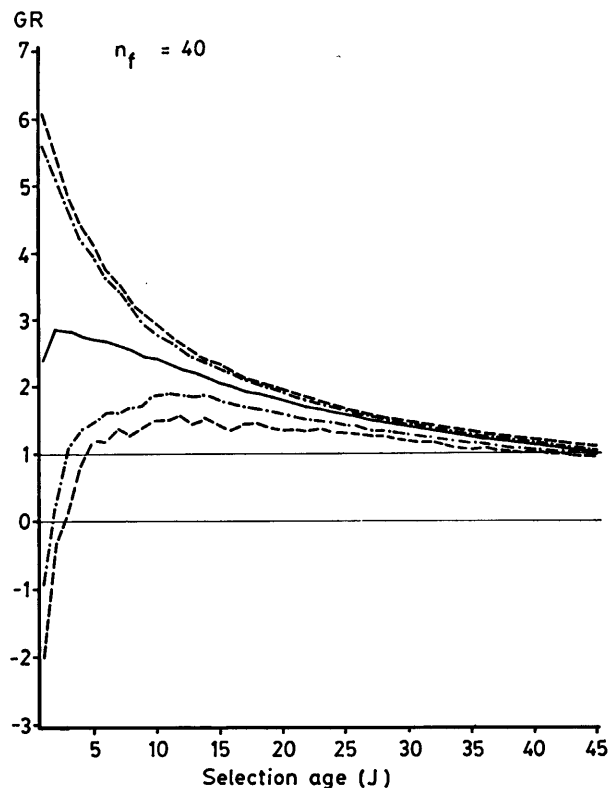
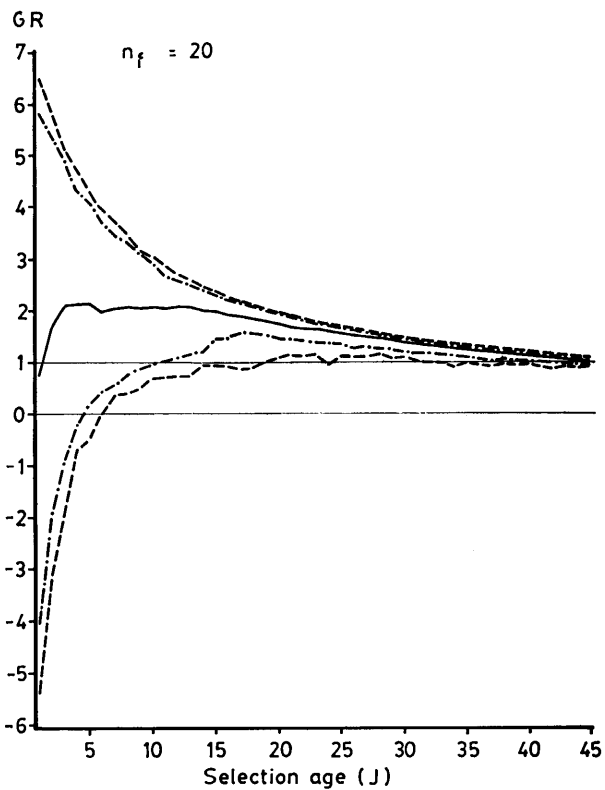
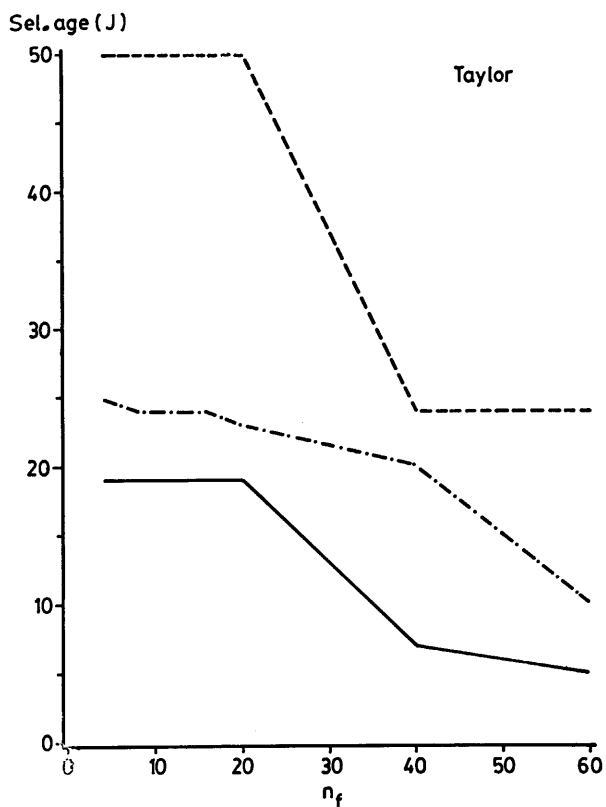
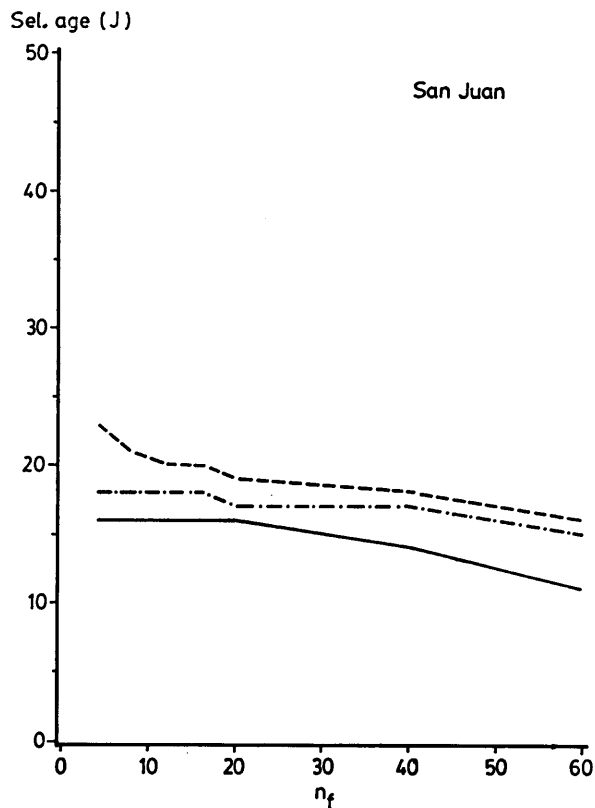


Figure 7. — Gain ratios (GR) of early family selections ($WGR=0$) at Ucluelet. Full line = mean gain ratio in 400 simulations. Dash line = minimum and maximum of the 400 simulated gain ratios. Dash-dot line = lower and upper 5% limits of the simulated gain ratios. Thin lines = the 0 and 1 thresholds of gain ratios.

5% gain ratio (\approx 'minimax' principle) are displayed in figure 8. Note the site specificity of the results. Also, the optimum average gain ratio criteria suggested lower (2 to

15 years) selection ages than did the criteria of second degree stochastic dominance. Selection ages based on mean gain ratio criteria are, therefore, not efficient in the presence of an utility function with 'universal' attributes ($U' > 0$, $U'' < 0$, $U''' > 0$). The discrepancy of selection ages based on the mean gain ratio and those that are second degree stochastic efficient between the Oktwanch



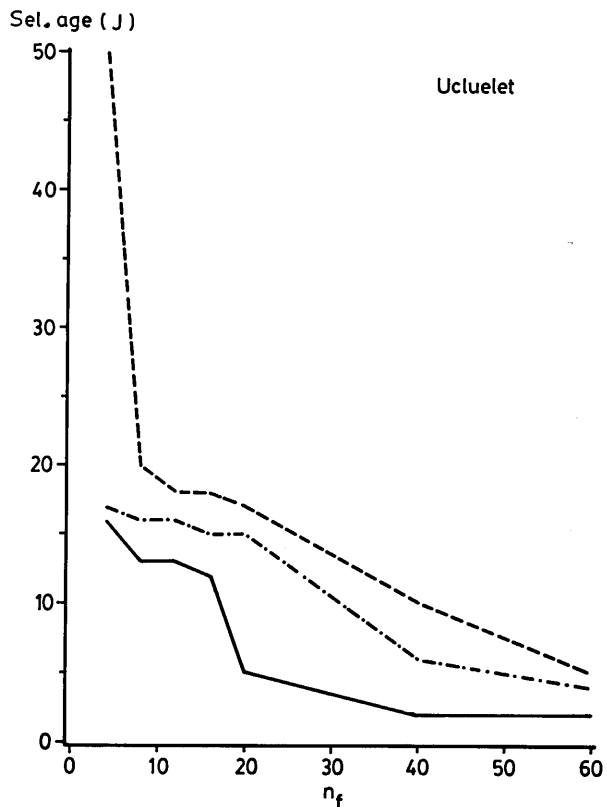
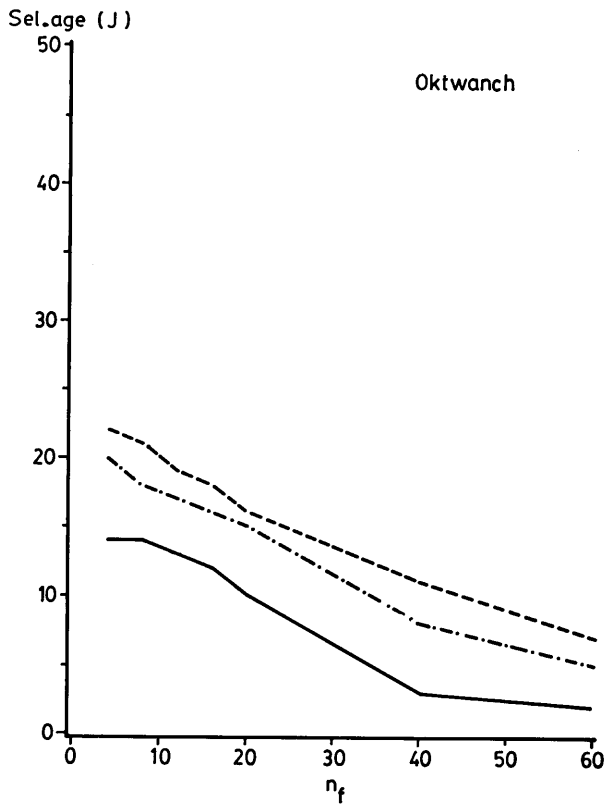


Figure 8. — Optimum family selection ages based on different criteria versus family sizes (n_f). Full line = optimum selection age for average gain ratio. Dash line = second degree stochastic efficient selection age. Dash-dot line = optimum selection age for the lower 5% gain ratio.

and Taylor sites for $30 < n_f < 50$ calls for concern about selection ages dictated by mean gain ratios alone. Risk adverse breeders may opt for the best outcome of the

lower 5% realized gain ratios. If this is the case, they would have to postpone selections by an additional 1 to 25 years over and above the second degree stochastic dominant selection age.

Family test size (n_f) had little impact on results at the San Juan site whereas a pronounced effect was evident on

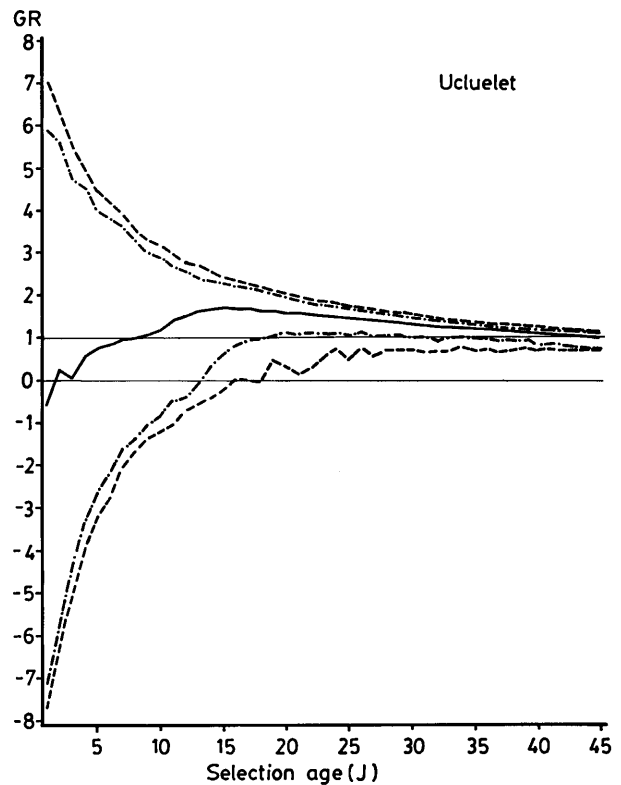
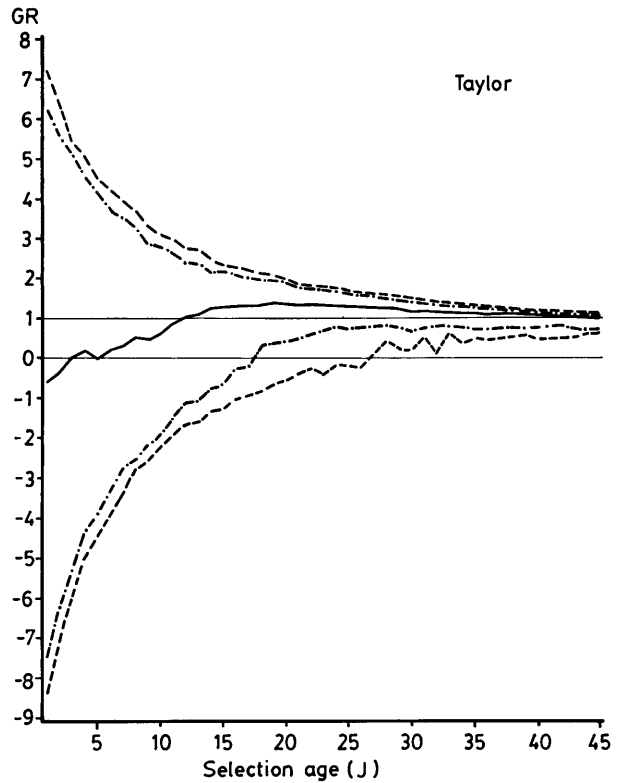


Figure 9. — Gain ratios (GR) of early forward selections (WGR=0.2) at Ucluelet and Taylor. Selection of 2 trees from each of 45 families.

the remaining sites. For an average of 30 trees per family it seems that a selection for height at an age of about 17 years is 'safe' (for all utility functions with the 'universal' properties) on the San Juan site, while 12 years seems to be sufficient at Oktwanch and Ucluelet. More than 20 years of height growth are needed on the Taylor site before safe family selections can be made for $n_f = 30$. A doubling of the family size would allow family selections about 5 years earlier on these sites (Taylor, Ucluelet, and Oktwanch) while the selections on San Juan appear to be independent of family size.

Forward selections

Individual gain ratios in the simulated early forward selections fluctuated within a broad range of values and the mean ratios peaked relatively late (depending on the WGR ratio, number of families, and trees within families) with no clearly defined maximum (Figure 9). Optimum selection ages often fluctuated by 2 to 4 years between repeated sets of 400 simulations. The results shown in figures 10 and 11 have been "smoothed" in order to avoid inconsistent details caused by random perturbations. No selection age was first degree stochastic dominant but one age could always be found to dominate all other ages by the second degree stochastic dominance criteria⁴⁾.

⁴⁾ And a subsequent Mean-Gini separation

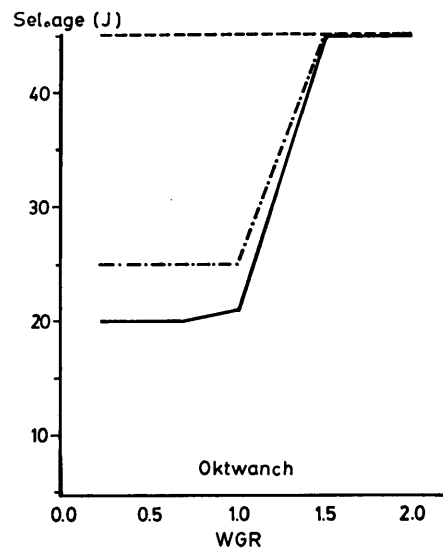
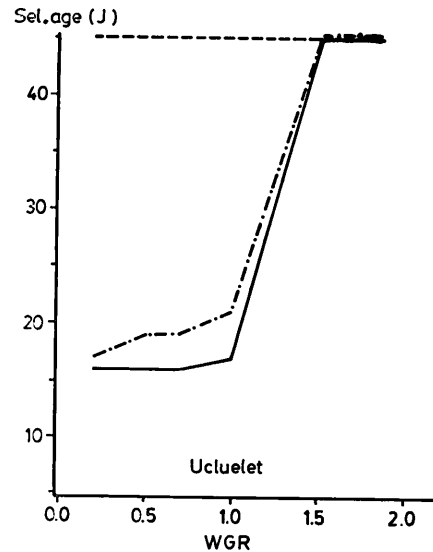
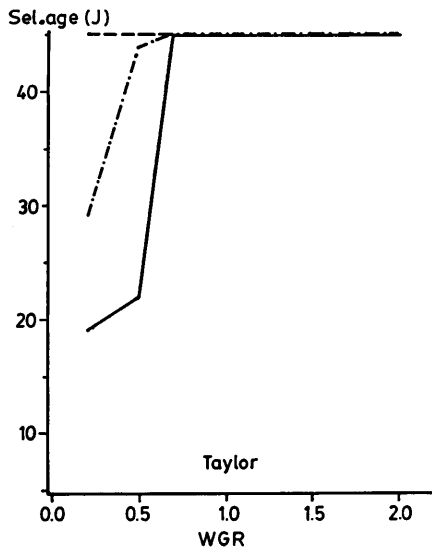
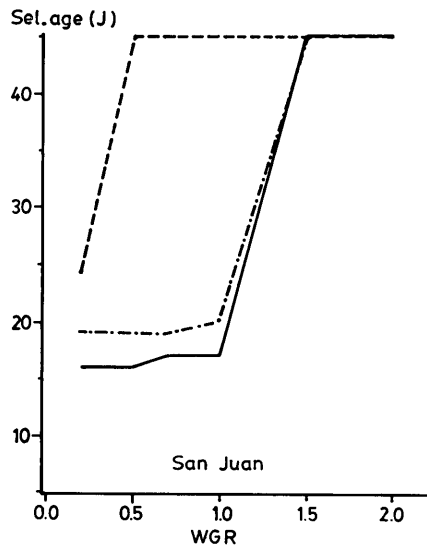


Figure 10. — Optimum selection ages in a forward selection scheme of selecting 2 trees from each of 45 families. Full line = optimum selection age for average gain ratio criteria. Dash line = second degree stochastic efficient selection age criteria. Dash-dot line = optimum selection age for the lower 5% gain ratio criteria.

In the forward selection scenarios presented (Figure 10 and 11), with 2 trees selected from each of 45 families or three trees selected from each of 30 families, a selection age around 17 appears to be the minimum stochastic efficient age (2nd degree stochastic dominance) and 15 years appears to be the minimum selection age based on mean gain ratio expectations. More conservative (risk adverse) selection ages were substantially higher. Selection of 3 trees in each of the 30 best families led to lower (1 to 10 years) recommended selection ages than obtained in schemes in which only 2 trees were selected in each of the best 45 families. The increased uncertainty associated with the prediction of the average performance of 2 trees rather than three was the main cause for this upward shift in selection age as the number of selected trees per family declined. With only two trees selected per family, the selection age with a 5% risk of a more adverse gain ratio remained, in most cases, close to the rotation age. Reliance on gain from within-family selection (high

WGR values) had a negative influence on selection age. Higher gain expectations from within family selections ($WGR > 1.0$) rendered early selection in the scenarios of two selections per family so risky and unattractive that the possibilities for accelerated breeding for height must be questioned.

In scenarios where only 20% to 40% of the gain can be expected to come from within-family selections a selection age around 16 to 25 years was, in most cases, 2nd degree stochastic efficient. Site effects were less pronounced in forward selections due to the masking effects of the variation in within-family correlation patterns.

Discussion

Age-to-age correlations plays a pivotal role in determining optimal selection age for any given breeding programme. Although several prediction models of age-to-age correlations exist (BURDON, 1989; LAMBETH, 1980), their stochastic properties and thus the risks associated with early selection are little known. Stochastic properties of age-to-age correlations derived from normal approximations are clearly inappropriate when the number of selected entities (families, trees per family) is less than about 30 (JOHNSON and KOTZ, 1971; MAGNUSSEN, 1990). It should be stressed that the stochastic properties are de-

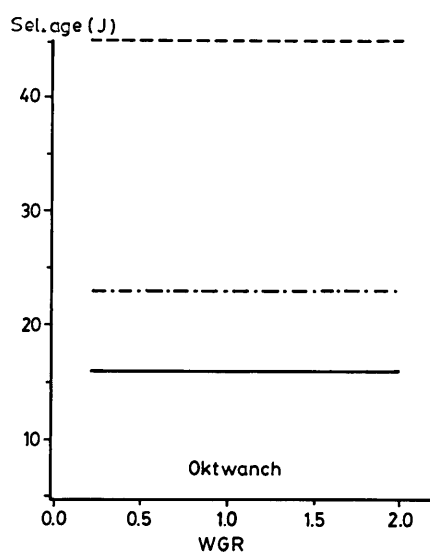
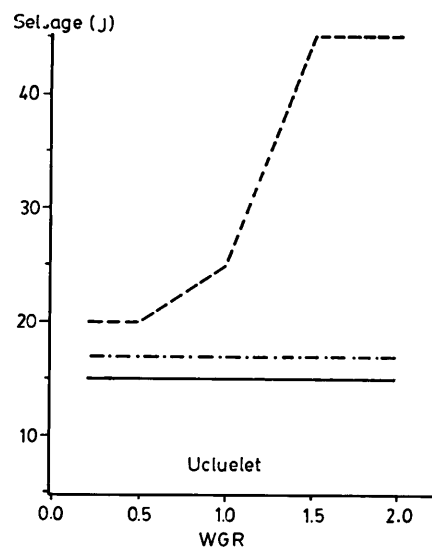
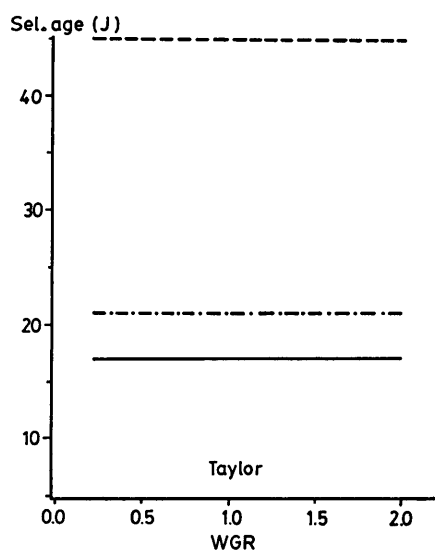
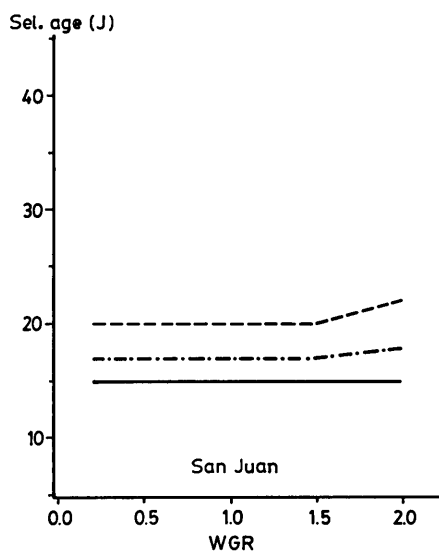


Figure 11. — Optimum selection ages in a forward selection scheme of selecting three trees from each of 30 families. Full line = optimum selection age for average gain ratio criteria. Dash line = second degree stochastic efficient selection age criteria. Dash-dot line = optimum selection age for the lower 5% gain ratio criteria.

termined by the trees selected and not by the entire base population. The practitioner will often operate below these limits, especially when sublimes or sets are improved (BURDON and VAN BUIJTENEN, 1990; BURDON and NAMKOONG, 1983; COTTERILL, 1986). Our use of the bootstrap technique (EFRON, 1982) to generate from a single data set the probability distribution of age-to-age correlation is one simple but powerful way of estimating the error structure of complex population statistics. A more elegant but data-demanding way of obtaining the same results is through time series analyses (Fourier transforms) as shown by MAGNUSSEN (1990). The fact that early selections are based on predictions entails the additional risk that the prediction model itself is biased. Not only must we consider the risk of not getting the expected value due to random perturbations of the growth process but also the possible impact of bias in the expected values themselves. Estimates of both sources of perturbations have been used in this study in an attempt to map the probabilities of all possible gain ratios in early selections. Recognition of family spe-

cific within-family correlation patterns added further realism to our simulations.

Given the high cost of tree improvement (CORNELIUS and MORGENSTERN, 1988; LOFGREN, 1988; PORTERFIELD et al., 1975; THOMSON et al., 1989) the risk of producing negative gain must be given serious consideration and efforts have to be made to give the best possible guarantees to managers who support tree breeding activities. Reliance on simple expectations of averages is no longer appropriate (LAMBETH et al., 1983). Complete knowledge of the risks of early selection will never be attained and the approach taken here only covers certain risk aspects. Genotype x environment interactions, for example, are manifest in this progeny test (ORR-EWING and YEH, 1978) but no attempt was made to integrate them in the overall question of selection age. Deployment of "improved" seedlots at several sites, each with a slightly different but generally unknown age-to-age correlation structure will, for a given risk level at individual sites, lead to higher selection ages in order to avoid local failures. Multi-trait selections will also tend to raise the selection age until the risk of adverse results is acceptable in the trait with the weakest age-to-age correlations.

A potentially serious weakness of most age-to-age correlation models (including that presented here) is the built-in reliance on unverified extrapolations from growth records restricted to the exponential or perhaps linear phase of the growth curve trajectory. In this study the extrapolation was done from age 22 to age 50. Height growth curves up to age 22 were, with a few exceptions, linear. Extrapolations of the expected age-to-age correlations will always be positive under these circumstances. However, if observations of growth are continued beyond the inclination point on a sigmoid growth curve, correlations may indeed become negative (FRANKLIN, 1979; NAMKOONG et al., 1972). There is no immediate solution to this dilemma because, even if complete juvenile-mature growth records of progenies were available for modeling, questions about intergenotypic competition and repeatability of the growth records would still linger.

Choosing the most appropriate prediction model is not only very important but also very difficult. We tried both the "LAMBETH" (LAMBETH, 1980) model and 2 models developed by MAGNUSSEN (1988, 1989b) but these models were either too inflexible or some underlying assumptions, regarding stationarity of variances or homogeneity of covariances, were excessively violated. It became clear, from this exercise, that a proper model choice cannot be made from only two very young (age 3 and 6) measurements because the validity of the model assumptions remains unknown. Thus, unless a validated prediction already exists, early selection based on shaky predictions may be a very risky proposition. The fact that the uncertainties associated with the family:within-family gain ratio (WGR) hasn't even been addressed here only stresses this point further. Preliminary calculations of WGR (at age 20) for the B.C. Douglas-fir programme suggest that this ratio will vary considerably among sites. A simple deterministic simulation of all possible outcomes (MAGNUSSEN, 1989a) may be the only acceptable "modeling" approach under these circumstances.

Whether the assumption of an 'universal' utility function with some desirable attributes and the use of stochastic dominance principles will lead to better decisions is hard to say but, by facing the risk and actively

formulating an attitude, we believe that some progress in the quality of our decisions regarding selection age can be made. A known utility function would simplify matters considerably. The 'universal' attributes of the utility function make "common sense" in agriculture (FLOOD et al., 1983; MENZ, 1980). Thus, an even more conservative attitude ought to prevail in forest decision-making. What is needed is an expansion of the information base on age-to-age correlations so that we eventually will have sufficient data and models to make near optimal selection-age decisions. A recent application of stochastic dominance as an aid to decision making demonstrated its usefulness in screening among genotypes to be rogued from a seed orchard (NEWBOLD and WEAVER, 1991).

Our results also have implications for choosing breeding strategies. COTTERILL (1986), for example, compared several breeding options without considering the effect of selection age on the gain per unit time. We have shown that a safe selection age can be lower when more gain arises from family selection and when more trees are selected per family. Experimental designs will also have to reflect the significance of family size for selection age. The optimum family size for comparison of means recommended by COTTERILL and JAMES (1984) may render a test unsuited for early selections.

Conclusion

The risk associated with early selection can be substantial and the expected genetic progress may be reduced or even reversed from the expected gain. Given site specific and family specific patterns of age-to-age correlations the task of finding an acceptable selection age is not straightforward. Stochastic simulation and modeling are essential tools to achieve a sound compromise between the desire to optimize gain and reduce the risks to acceptable levels. The stochastic dominance theory offers an intuitively appealing substitute for the utility function in decision making. Our results indicate that a prudent attitude to risk will raise the selection age anywhere from 1 to 15 years above the solution found with deterministic models.

Acknowledgments

Valuable comments and critiques of an earlier version of this paper were provided by Drs. J. LOO-DINKINS, J.-C. BASTIEN, and C. HEAMAN.

Literature Cited

ANDERSON, J. R.: Risk efficiency in the interpretation of agricultural production research. *Rev. Market. Agric. Econ.* 42: 131-184 (1974). — BRIDGWATER, F. E. and WILLIAMS, C. G.: Early testing and juvenile selection in loblolly pine. Pages 1-7. *In Proc. 30th Northeastern Forest Tree Improvement Conference. Univ. of Maine, Orono, July 1986* (1987). — BUCCOLA, S. T. and SUBAEI, A.: Mean-Gini efficiency and weak risk aversion. *Austr. J. Agr. Econ.* 28: 77-86 (1984). — BURDON, R. D.: Early selection in tree breeding: principles for applying index selection and inferring input parameters. *Can. J. For. Res.* 19: 499-504 (1989). — BURDON, R. D. and VAN BUIJTENEN, J. P.: Expected efficiencies of mating designs for reselection of parents. *Can. J. For. Res.* 20: 1664-1671 (1990). — BURDON, R. D. and NAMKOONG, G.: Multiple populations and sublines. *Silvae Genet.* 32: 221-222 (1983). — CORNELIUS, J. P. and MORGENSTERN, E. K.: An economic analysis of black spruce breeding in new Brunswick. *Can. J. For. Res.* 16: 476-483 (1988). — COTTERILL, P. O.: Genetic gains expected from alternative breeding strategies including low cost options. *Silvae Genet.* 35: 212-223 (1986). — COTTERILL, P. P. and JAMES, J. W.: Number of offspring and plot-sizes required for progeny testing. *Silvae Genet.* 33: 203-209 (1984). — DILLON, J. L. and HARDAKER, J. B.: *Agricultural decision analysis.* Iowa State University Press (1977). — EFRON,

B.: The jackknife, the bootstrap and other resampling plans. Soc. Industr. Appl. Math. Philadelphia (1982). — FLOOD, M., McCAMLEY, F. and SCHNEEBERGER, K.: Stochastic dominance with respect to a function as an approach to evaluate the rationality of farmer variety choices. In: Proc. Western Agric. Econ. Assoc. Meeting, Laramie, Wyoming, July 1983 (1983). — FOSTER, S.: Trends in genetic parameters with stand development and their influence on early selection for volume growth in loblolly pine. Forest Sci 32: 944–959 (1986). — FRANKLIN, E. C.: Model relating levels of genetic variance to stand development of four North American conifers. Silvae Genet. 28: 207–212 (1979). — HÜHN, M.: Clonal mixtures, juvenile-mature correlations and necessary number of clones. Silvae Genet 36: 83–92 (1987). — JIANG, I. B.-J., JONSSON, A. and ERIKSSON, G.: Within- and between-population variation in growth of *Pinus contorta* var. *latifolia*: A combined study of growth-chamber and field-trial experiments. Silvae Genet. 38: 201–211 (1989). — JOHNSON, N. L. and KOTZ, S.: Continuous univariate distributions. Houghton Mifflin Co. (1971). — 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). — LI, B., McKEAND, S. E. and ALLEN, H. L.: Early selection of loblolly pine families based on seedling shoot elongation characters. Pages 228–234. In: Proc. 20th Southern Forest Tree Improvement Conf. June 26–30, 1989, Charleston, SC. (1989). — LOFGREN, K.-G.: On the economic value of genetic progress in forestry. Forest Sci. 34: 708–723 (1988). — MAGNUSSEN, S.: Minimum age-to-age correlations in early selections. Forest Sci. 34: 928–938 (1988). — MAGNUSSEN, S.: Determination of optimum selection ages: a simulation approach. Pages 269–285. In: Proc. 20th Southern Tree Improvement Conference. June 26–30, 1989, Charleston, SC (1989a). — MAGNUSSEN, S.: Age-to-age correlations in growth processes with fixed and random effects. Silvae Genet. 38: 49–55 (1989b). — MAGNUSSEN, S.: Stochastic dynamics of age-to-age correlations. Pages 4.86–4.94. In: Proceedings Joint Meeting of Western Forest Genetics Association and IUFRO Working Parties S2.02–05, 06, 12, 14, Olympia, WA, Aug. 20 to 24, (1990). — MAGNUSSEN, S.: Efficiency of early selections for stem-volume and predictions of size distributions of selections in a red pine spacing trial. Forest Sci. 37: 593–612 (1991). — MENZ, K. M.: A comparative analysis of wheat adaptation across international environments using stochastic dominance and pattern analysis. Field Crop Res. 3: 33–41 (1980). — NAMKOONG, G., USANIS, R. A. and SILEN, R. R.: Age-related variation in genetic control of height growth in Douglas-fir. Theor. Appl. Genet. 42: 151–159 (1972). — NEWBOLD, R. A. and WEAVER, G. H.: Stochastic dominance as an aid to genetic roguing. Silvae Genet. 40: 156–160 (1991). — ORR-EWING, A. L.: Inter- and intraspecific crossing in Douglas fir, *Pseudotsuga menziesii* (MIRB.) FRANCO. Silvae Genet.

15: 121–126 (1966). — ORR-EWING, A. L., FRASER, A. R. and KARLSSON, I.: Interracial crosses with Douglas Fir. Early field results. Research Note, No. 55. British Columbia Forest Service (1972). — ORR-EWING, A. L.: and YEH, F. C.: Survival and growth traits of racial crosses with Douglas fir. Research Note, No. 85. Min. of For., B. C. (1978). — POPE, R. D. and ZIEMER, R. F.: Stochastic efficiency, normality and sampling errors in agricultural risk analysis. Am. J. Agr. Econ. 66: 31–40 (1984). — PORTERFIELD, R. L., ZOBEL, B. J. and LEDIG, F. T.: Evaluating the efficiency of tree improvement programs. Silvae Genet. 24: 33–44 (1975). — REHFELDT, G.: Veracity of short-term testing. Pages 118–146. In: Proc. USDA Forest Serv. Servicewide Genetics Workshop, Charleston, SC (1983). — RIPLEY, B. D.: Stochastic simulation. John Wiley & Sons (1987). — SIEGEL, S.: Nonparametric statistics for the behavioral sciences. McGraw-Hill Book Company (1956). — SNEDECOR, G. W. and COCHRAN, W. G.: Statistical methods. Iowa State University Press, Ames, Sixth ed. (1971). — STEINHOFF, R. J.: Juvenile-mature correlations in Ponderosa and western white pine. Pages 243–251. In: Proc. IUFRO Joint Meeting of Working Parties on Population and Biological Genetics, Breeding Theory and Progeny Testing, Stockholm (1974). — THOMSON, T. A., LESTER, D. T., MARTIN, J. A. and FOSTER, G. S.: Using economic and decision making concepts to evaluate and design a corporate tree improvement programme. Silvae Genet. 38: 21–28 (1989). — WAKELEY, P. C.: Relation of thirtieth-year to earlier dimension of southern pines. Forest Sci. 17: 200–209 (1971). — WHITMORE, G. A. and FINDLAY, M. C.: Stochastic dominance. An approach to decision-making under risk. Lexington Books (1978). — ZOBEL, B. and TALBERT, J.: Applied Forest tree improvement. John Wiley & Sons (1984).

Appendix A (the data source)

Individual tree height and age data gathered in four progeny tests formed the basis for the age-to-age correlation models. Heights were measured at ages 3, 6, 11, 14, and 22 on 14 to 18 families per test (12 to 14 full sibs and 2 to 4 half sibs; 14 families were common to all 4 tests, 16 families were common to 3 tests, and 18 families were common in 2 tests). The test design was one of randomized complete blocks (two replications) with 50-tree row plots planted in 1968 with one plus one transplants at a spacing of 10 x 10 feet. Further details about the test sites and the family structure has been published by ORR-EWING et al. (1972). A summary of the height measurements are given below:

Site	# trees in analysis	# families	trees per family	height (m) and coefficient of variation* (in brackets)				
				3	6	11	14	22
San Juan	1372	18	61-88	0.9(28)	3.0(17)	8.1(10)	13.8(8)	18.2(7)
Taylor	852	14	50-75	0.8(22)	1.6(22)	4.1(21)	8.0(16)	10.8(14)
Ucluelet	1210	18	48-80	1.6(27)	3.9(16)	8.0(11)	12.8(10)	17.2(8)
Oktwanch	789	16	24-60	1.0(22)	1.9(23)	5.6(15)	10.8(11)	14.8(9)

$$\bullet) \frac{\bar{x}}{\sigma} \bullet 100$$

Family effects were highly significant at all ages and on all sites (min $F_{fam} > 6.0$, $P \leq 0.01$). Heights generally followed a Gaussian distribution (4 of 5 Chi-square tests of 'goodness of fit' yielded non-significant departures at the 5% level of significance). Trees used in the analyses were all free of any visible damage. Causes of differential number of trees per family were: 1) root rots (*Armillaria mellea*, *Phellinus weirii*), 2) animal damage, 3) shortage of stock.

Appendix B (the generalized beta distributions of age-to-age correlations)

The appropriate beta distribution for a given expected mean age-to-age correlation (=model prediction plus a

random effect due to bias) was determined via 2 parameters (\hat{p} , and \hat{q}). See also JOHNSON and KOTZ (1971), chapter 24, page 44 equations 16 to 20 for further details.

These 2 parameters were, in turn, related to otherwise obtainable parameters in the following way:

$$\hat{p} = \frac{(\bar{r} - r_{\min})}{(r_{\max} - r_{\min})} \bullet \left[1 - \frac{(r_{\max} - r_{\min})}{(r_{\max} - r_{\min})} \right] \bigg/ \left[\frac{s^2(\bar{r})}{(r_{\max} - r_{\min})^2} - \frac{(\bar{r} - r_{\min})}{(r_{\max} - r_{\min})} \right]$$

$$\hat{q} = [\hat{p} \bullet (r_{\max} - r_{\min})] / (\bar{r} - r_{\min}) - \hat{p}$$

where $\hat{r} = \hat{r}(J:M) + \epsilon(\hat{r})$ as predicted from eq [8] and perturbed with a random disturbance term.

Table 4. — Regression models used to determine the maximum (\hat{r}_{\max}), the minimum (\hat{r}_{\min}), and the variance ($\hat{\sigma}_r^2$) of an age-to-age correlation (r). The estimates are used, in turn, to obtain parameters for the probability distribution (beta) of age-to-age correlations (see Appendix C).

1. Variance of age-age correlations

Family mean correlations:

$$\hat{\sigma}_{r_f}^2 = (1 - \hat{r}_f^2) / n_f \cdot (0.2600 - 0.2572 \cdot e^{-0.1791 \cdot n_f})$$

(coeff. of determination = 0.91, df = 89, s.e. (regr.) = 0.00355)

Within family correlations:

$$\hat{\sigma}_{r_w}^2 = (1 - \hat{r}_w^2) / n_w \cdot (0.7834 - 0.4018 \cdot e^{-2.3099 \cdot n_w})$$

where n_w = number of trees selected per family (=2,3)

(coeff. of determination = 0.9835, df = 479, s.e. (regr.) = 0.0146)

2. Maximum of age-age correlations

Family mean correlations:

$$\hat{r}_{f_{\max}} = \hat{r}_f + (1 - \hat{r}_f) \cdot (1 - e^{-9.5480 \cdot \hat{\sigma}_{r_f}})$$

(coeff. of determination = 0.6318, df. = 359, s.e. (regr.) = 0.042)

• Within-family correlations:

$$\hat{r}_{w_{\max}} = \hat{r}_f + (1 - \hat{r}_f) \cdot (1 - e^{-9.5480 \cdot \hat{\sigma}_{r_f}})$$

(coeff. of determination = 0.6648, df = 479, s.e. (regr.) = 0.030)

3. Minimum age-age correlations

Family mean correlations:

$$\hat{r}_{f_{\min}} = 2 \cdot e^{-3.1806 \cdot cv_f}, \quad cv_f = \hat{\sigma}_{r_f} / \hat{r}_f$$

(coeff. of determination = 0.9072; df = 359, s.e. (regr.) = 0.118)

Within-family correlations:

$$\hat{r}_{w_{\min}} = 2 \cdot e^{-4.1283 \cdot cv_w}, \quad cv_w = \hat{\sigma}_{r_w} / \hat{r}_w$$

(coeff. of determination = 0.8432; df = 479, s.e. (regr.) = 0.2107)

where r_{\min} = maximum age-age correlation (see Table 4 for method of predicting this parameter).

where r_{\max} = maximum age-age correlation (see Table 4 for method of predicting this parameter).

$\sigma^2(\bar{r})_y$ = variance of age-age correlation (see Table for method of predicting this parameter).

Simple regression models to predict the minimum, maxi-

imum, and variance of the empirical bootstrap-generated distributions of age-to-age correlations from the bias-corrected expected age-to-age correlation ($\hat{r}(J:M) + \varepsilon(\hat{r})$), and sample size (Table 4) were computed from the bootstrap results in order to generate, in a general way, the necessary parameters of the assumed beta-distribution of $r(J:M)$.