

# Monte Carlo Simulation Models of Breeding-Population Advancement

By J. N. KING<sup>1)</sup> and G. R. JOHNSON<sup>2)</sup>

Forest Research Institute, Private Bag 3020,  
Rotorua, New Zealand

(Received 10th January 1992)

## Summary

- 
- <sup>1)</sup> British Columbia Ministry of Forests, Research Branch, 31  
Bastion Square, Victoria, B. C., V8W 3E7, Canada  
<sup>2)</sup> U. S. National Arboretum, A. R. S., U. S. D. A., 3501 New  
York Ave. N. E., Washington, D. C., 20002, USA

Five generations of population improvement were modelled using Monte Carlo simulations. The model was designed to address questions that are important to the

development of an advanced generation breeding population. Specifically we addressed the effects on both gain and effective population size of different mating schemes when creating a recombinant population for recurrent selection. Five different mating schemes were modelled with recurrent selection, 4 control-pollinated (CP) and one open-pollinated (OP) scheme. The CP schemes were: 2 random mating (RM) designs, 1 with 2 crosses per parent and 1 with 8 crosses per parent; and imbalanced parental contribution schemes including assortative mating (AM), and random mating with an 'elite' nucleus (EN). Genetic gain is maximized in the mating scheme that uses random mating and increases the selection differential for recombinant cross selection. The imbalanced designs (AM and EN) increased gain only slightly over random mating. The model was also used to look at the erosion of the breeding-population base when applying different restrictions on selection to the CP treatments. RM offered the highest levels of gain for a given effective population size. AM was little different to RM but there was quite rapid attrition of the genetic base for on-going breeding with the EN scheme. The trade-off between maximizing short-term gain with family selection and maintaining genetic diversity for long-term potential within the framework of a fixed resource is discussed.

*Key words:* Assortative mating, computer modelling, Monte Carlo simulation, genetic gain, genetic diversity.

### Introduction

The breeding and domestication of forest trees is a comparatively recent development. Tree breeding has concentrated primarily on broad-based population improvement through recurrent selection of general combining ability (RS/GCA) (BURDON and SHELBORNE, 1971). This strategy is designed to progressively increase the frequency of genes in the breeding population which exhibit desirable additive effects, by means of cycles of selection and crossing. The breeding population is part of a hierarchical population structure ranging from a broad-based gene resource population through the more narrowly based pedigree breeding population (open-pollinated (OP) and/or control-pollinated (CP)), to the highly selected parents used in the seed production population.

This paper uses computer modelling and Monte Carlo simulation of breeding population advancement over several generations. We have particularly looked at different mating schemes for their contribution to per generation gain and their effect (with selection) on the genetic diversity of the breeding population. The delivery of genetic gains from the breeding population to production population and the utilization of non-additive genetic variance is part of another paper. The details in the model are based on the New Zealand radiata pine improvement program. Results of the simulation should, however, have inferences for most programs using RS/GCA.

Genetic gains are accumulated each generation in the breeding population through selection. Mating schemes for population advancement (to create a new generation for recurrent selection) may be used to improve these gains. Assortative mating (AM) leads to an expectation of increased genetic variance (McBRIDE and ROBERTSON, 1983). Assortative mating refers to the mating of individuals of the same phenotype more often than would occur by chance (FALCONER, 1981). Positive assortative mating applies when the phenotypic value of both mates deviates from the population mean in the same direction (GIANOLA, 1982). The expectation of increased genetic variance will be

diminished if heritabilities are low (thus the phenotype less accurately represents the genotype) and/or only partial assortative mating occurs, i. e. best is mated with best but worst is not mated with worst (McBRIDE and ROBERTSON, 1963).

LINDGREN (1986) proposed promoting some of the features of assortative mating by the use of imbalanced family contribution for breeding populations. This comprises having both more representatives of good families and more cross combinations with better genotypes. One of the ways of allowing better parents to contribute more is to have the breeding population divided into a hierarchy with an 'elite' group recombined in addition the 'main' population matings. A top down hierarchy with genetic change being directed through an elite nucleus is an important feature of some animal improvement programs (JAMES, 1989) and has been proposed for the breeding populations of forest trees (COTTERILL et al., 1988). MAHALOVICH (1990) found in a similar simulation exercise that an elite nucleus offered significant extra gain over an undivided breeding population. Another method of using unequal contributions to favour better parents is to assign a mating frequency in accordance with an individual's breeding value. LINDGREN (1986) proposed that mating frequency be assigned in a linear relationship to a parent's breeding value. This kind of selection for mating frequency provides another level of selection over the initial truncation selection (KANG and NAMKOONG, 1988). Imbalanced breeding schemes like assortative or 'nucleus' type matings, appear to offer a means of increasing gain over random mating for a given selection differential. This increased gain from imbalanced breeding should be examined and compared for their efficiencies to random mating schemes with increased selection differentials.

Different selective breeding schemes need to be compared not only for genetic gain, but also for their impact on genetic diversity within the breeding population. An effective population size involving the reciprocal of the sum of squares of each parent's proportional contribution to the breeding population offers a simple and effective way of evaluating genetic diversity (ROBERTSON, 1961; KANG and NAMKOONG, 1988; LINDGREN, 1991). Truncation selection without restrictions will maximize genetic gain but will be expected to have a high cost in terms of reduced effective population size. Within family selection and balanced mating will maximize effective population size but genetic gains will be minimal and it would be unlikely to be worth the effort of continued breeding. LINDGREN (1986) argues that imbalanced mating (where higher breeding value parents are used in more crosses) may have an adverse effect on overall genetic diversity and allele loss but it would be balanced by counteracting allele loss in the very best genotypes. The highest breeding value parents are likely to offer the greatest potential for breeding population improvement and it is therefore worth concentrating the investment in genetic improvement on these genotypes. KANG and NAMKOONG (1988) showed that in some circumstances imbalanced designs where mating frequency is in proportion to an individual's breeding value could even generate larger effective population sizes than truncation selection followed by balanced mating.

Monte Carlo simulation with computer modelling offers a valuable tool for the empirical testing of alternatives. It has been widely used in animal breeding research (MA-

Table 1. — Genetic parameters used in the simulated breeding populations.

Parameters	Intermediate heritability	Low heritability	High heritability
$\sigma_f^2$	140	62	300
$\sigma_A^2$	560	248	1200
$\sigma_e^2$	2235	2235	2235
$\sigma_{pi}^2$	2795	2483	3435
$\sigma_{pw}^2$	2515	2359	2835
$h_i^2$	0.20	0.10	0.35
$\sigma_{pf}^2$	174	93	340
$h_f^2$	0.80	0.67	0.88

where:

- $\sigma_f^2$  = half-sib family variance component.
- $\sigma_A^2$  = additive genetic variance =  $4\sigma_f^2$ .
- $\sigma_e^2$  = true environmental variance.
- $\sigma_{pi}^2$  = individual phenotypic variance =  $\sigma_e^2 + \sigma_A^2$ .
- $\sigma_{pw}^2$  = phenotypic variance of individuals within full-sib families =  $\sigma_e^2 + \frac{1}{2}\sigma_A^2$ .
- $h_i^2$  = individual tree (narrow sense) heritability =  $\sigma_A^2/\sigma_{pi}^2$ .
- $\sigma_{pf}^2$  = half-sib family mean phenotypic variance =  $\sigma_f^2 + (\frac{3}{4}\sigma_A^2 + \sigma_e^2)/K$ .
- K = no. of individuals used to estimate family mean = 78.
- $h_f^2$  = family repeatability.

DALENA and HILL, 1972; CROSBY, 1973; DEMPFLE, 1974; FERNANDO and GIANOLA, 1986). MAHALOVICH (1990) developed gene effect models (additive and partial dominance) to evaluate breeding strategies in forestry programs. Our study involves a Monte Carlo simulation of breeding population advancement in order to investigate features of the breeding population structure. Questions addressed with the computer simulation were:

1. Can substantial extra gain be achieved by using imbalanced matings schemes where better phenotypes are used more often? This includes a scheme using assortative mating (AM) features and one using an elite nucleus with random mating in the main population (EN).
2. Within the framework of a fixed parent and fixed offspring resource, how do these imbalanced mating schemes compare to increasing recombinant family selection differentials with balanced random mating (RM)?
3. Would a simpler mating design such as OP breeding be a worthwhile alternative to CP designs for breeding-population advancement?

4. How is effective population size in the breeding population affected by different levels of restriction on combined index selection and how does the relationship between effective population size and gain vary between random and imbalanced mating schemes?

#### The Model

The modelling uses a Monte Carlo simulation of the advancement of the breeding-population for five generations of recurrent selection for a single trait. Unlike MAHALOVICH (1990), who used a gene effects model, we used a parameter based model. Estimates of genetic and environmental variances were used to produce independent normal distributions of true genotypic values and environmental deviations. These were used to construct a simulated breeding-population. The genetic parameters used in the model and the formulae for their estimation are shown in table 1. Underlying genetic assumptions are that the genetic values for the trait derive from a poly-

Table 2. — Index selection procedures used.

Index	half-sib		full-sib		
$P =$	$\sigma_{pf}^2$	$\frac{1}{4} \sigma_A^2$	$\sigma_{pf}^2$	0	0
	$\frac{1}{4} \sigma_A^2$	$\sigma_{pi}^2$	0	$\sigma_{pf}^2$	0
			0	0	$\sigma_{pw}^2$
$G =$	$\frac{1}{4} \sigma_A^2$	$\frac{1}{4} \sigma_A^2$	$\frac{1}{4} \sigma_A^2$	0	0
	$\frac{1}{4} \sigma_A^2$	$\sigma_A^2$	0	$\frac{1}{4} \sigma_A^2$	0
			0	0	$\frac{1}{2} \sigma_A^2$
$a =$	$\begin{pmatrix} 0 \\ 1 \end{pmatrix}$			$\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$	
	$b = P^{-1} G a$				
for intermediate heritability					
$b_1$	0.67		1.60	(0.80 each for male and female GCA)	
$b_2$	0.167		0.110		

where:  $b_1$  is the index weight of families.  
 $b_2$  is the index weight of individuals.

genic and purely additive gene effect model. Selection of the breeding population employed combined index selection with some restrictions. Table 2 shows how the parameters were used in index selection procedures. A heritability of 0.20 was used — this value is commonly estimated for growth traits of New Zealand radiata pine — and the actual numbers (Table 1) came from the recent volume assessment of a disconnected diallel (used for a third-generation of selection) at 7.5 years (KING et al., in MS). Lower and higher heritability values were constructed by changing the additive genetic variance. The progression of events that are modelled is outlined in table 3. Although the details of the simulation were designed for the advanced generation breeding program of radiata pine in New Zealand (SHELFOURNE et al., 1989) it has many features common to the genetic development in many commercial forest tree improvement programs.

One feature common to most programs is the fact that the first generation of improvement consists of a wide base of phenotypic selections (plus trees) followed by progeny testing of these selections. Subsequent generations utilize selections based on progeny test information (usually selection indices) followed by breeding and recurrent selection. As generations of recurrent selection and breeding progress, especially in a closed breeding population, related crosses will eventually occur which will be at a distinct disadvantage if tested alongside non-inbred (or less-inbred) families. Sublining and the use of 2 complementary mating designs are being used to over-

come this problem in many programs (SHELBOURNE et al., 1989; McKEAND and BRIDGEWATER, 1992; HODGE, G., pers. comm.; LOWE and VAN BUITENEN, 1989). Sublining accepts inbreeding within the breeding population sublines, while outcrossing is assured in commercial seed production by crossing between unrelated sublines (BURDON, 1986). The complementary mating design includes: (1) an out-crossing GCA test of breeding population parents, and (2) pair-crosses between parents within breeding population sublines in order to produce a new generation for recurrent selection.

Monte Carlo simulation was used to address how the latter mating with selection affects both genetic gain and effective population size in the advancement over generations of the breeding population sublines. In the model, breeding subline size was maintained at 60 parents, except the first generation which started with 240 selections — in order to simulate gain from large-scale family screening of the first generation. Sixty second-generation selections from the 240 OP families provided a common base from which the different mating and selection schemes were applied. Later generations were kept closed — i. e., no new genetic material was added. Selections for each successive generation of breeding were made with combined index selection using family (half or full-sib) and individual (or within-family) phenotypic values.

Five different mating schemes were investigated. One hundred individuals per cross were generated from each mating (except that in one scheme 25 individuals were

Table 3. — Parameters used and values generated in model development.

	Step	Input parameters	Designation
<b>Generation 1-2</b>			
	a - Generate 240 half-sib genetic values	$\frac{1}{4} \sigma_A^2$	G1
	b - Generate 240 random error deviations	$(\frac{3}{4} \sigma_A^2 + \sigma_e^2)/78$ (1)	E1
	c - Calculate 240 phenotypic values	G1 + E1	P1
1st phase	d - Select top 60 families	P1	
	e - Generate 100 individual genetic values within each of the 60 families	$\frac{3}{4} \sigma_A^2 + G1$	G2
	f - Generate 100 individual environmental deviations within each of the 60 families	$\sigma_e^2$	E2
	g - Calculate 2nd-generation phenotypic values	P2 = G2 + E2	E2
2nd phase	h - Select best individual per family	P2	
	i - Index rank 2nd-generation breeding parents	$\sum b_1 P1 + b_2 P2$	I2
<b>Generation 2-3</b>			
	j - Combine family random error variation with genetic value for GCA ranking	G2/2 + E1	P2'
	k - Cross 60 2nd-generation selections	G2	
	l - Generate 100 (4) individual genetic values within families	$\frac{1}{2} \sigma_A^2 + G2$ (2)	G3
	m - Generate 100 (4) individual environmental deviations within families	$\sigma_e^2$	E3
	n - Calculate 3rd-generation phenotypic values	P3 = G3 + E3	E3
	o - Index selection for next generation	$\sum b_1 P2' + b_2 P3$ (3)	I3
	p - Select best 60 index-ranked individuals for breeding population (with constraint of maximum of 4 per family)	I3	
Generation 3-4 and 4-5; repeat Generation 2-3			

<sup>1)</sup> 78 = no. of individuals used to estimate family mean in GCA testing.

<sup>2)</sup> In OP scheme is  $\frac{3}{4} \sigma_A^2 + \sigma_e^2$  + pollen contribution.

<sup>3)</sup> Half-sib index for OP scheme and full-sib index for CP schemes (Table 2).

<sup>4)</sup> 25 individuals generated in the 240 full-sib cross scheme.

generated in order to maintain a fixed number of trees in the breeding population resource). The population was generated by using random, independently assigned genotypic values (G) and environmental deviations (E) for each individual in the breeding population (P (phenotypic value) = G + E). The simulation was modelled with the SAS software (SAS Institute, 1985). Details of the simulation are presented in table 3. The progression of events were:

1. OP families of 240 first-generation selections were generated and underwent a two-phase selection; first, family selection including the top 60 families (Table 3; steps a to d), and then the best individual from each of

these families (Table 3; steps e to h). These 60 second-generation breeding parents were then ranked on a combined index (half-sib family + individual) (Table 3, step i) to provide a common basis for the simulation of breeding population advancement. All five alternative mating designs investigated in the model started with the same 60 index-ranked second generation parents for each run of the model.

2. Five different mating schemes for recurrent selection were investigated for these 60 index-ranked parents (Table 4);

(a) 60 OP families with single-tree plot field layout. This was the low-cost option, with dual function mating

Table 4. — Mating schemes for the 60 index-ranked selected parents.

Scheme	Function	Field layout	No. of families	Index <sup>1</sup> selection	No. of parents selected
(a) OP-60	Dual-function: GCA testing and forwards selection	STP 100 trees per family	60	Individual + half-sib family	60
(b) RM-60 <sup>2</sup>	Single function: forwards selection	Family block 100 trees/block	60	Within full-sib family + 2 half-sib family	60
(c) AM-60 <sup>2</sup>	"	"	60	"	60
(d) EN-60 <sup>2</sup>	"	"	60	"	60
(e) RM-240 <sup>2</sup>	"	Family block 25 trees/block	240	"	60

STP = Single-tree-plot.

<sup>1</sup>) See table 2 for details.

<sup>2</sup>) Also includes separate GCA test of parents

design and field test. It was assumed there was no pollen contamination from outside the trial, (OP-60).

(b) 60 random mated double pair-crossed families (2 crosses per parent) (RM-60).

(c) Assortative mating scheme. The breeding parents were divided into 3 groups based on their index score. Parents in the top third were designated 'A' parents and crossed 3 times, parents in the middle third ('B' parents) were crossed twice, and parents in the bottom third ('C' parents) were crossed only once (an average of 2 crosses per parent or 60 full-sib families). The feature of this scheme is the use of the 'better' parents in more crosses. Although this does not have all the features of positive assortative mating (i. e. poor performers are not crossed together CxC) it has some of its practical features. The basic crossing strategy was AxA, AxB, BxB, and AxC (AM-60).

(d) 'Elite' nucleus mating scheme. In addition to random double pair-crossed families (RM-60), the 'top-six' selections were crossed in a half-diallel 'elite' nucleus to try to ensure that more families from combinations of the best parents were made available for recurrent selection. Seventy five families were generated in this scheme (60 random and 15 in the nucleus). Sixty of the 75 families were randomly chosen to be tested in order to ensure a comparable family truncation levels as in treatments a through c (EN-60).

(e) 240 full-sib families were generated from random mating (8 crosses per parent). This scheme examined the effect on increasing family selection intensity while keeping the parent population size constant by increasing the number of crosses per parent. Because the number of families was increased in this scheme the number of individuals per family was decreased in order to maintain a fixed resource (RM-240).

Both the parents selected per generation (60) and total number of offspring produced for recurrent selection (6000) were kept fixed for all mating scheme treatments. The true genotypic values of parents were combined to produce (for each full-sib cross) a true genotypic cross value (in the case of OP matings this was the half-sib

value from the parent and the pollen was assumed to be from the unselected breeding population of the previous generation) (Table 3; step k).

3. Besides the mating schemes for population advancement, each of the CP schemes (b to e above; and Table 4) also had a concurrent outcrossing GCA test of the parents. For each parent an environmental deviation was generated and combined with the genotypic value to produce a phenotypic half-sib family mean value for this GCA testing phase (Table 3, step j).

4. Genotypic values and environmental deviations for 100 individuals (25 individuals for the 240 random full-sib families) were generated within each full-sib cross (half-sib in the case of the OP families), using one-half the additive genetic variance (three-quarters of the additive genetic variance for OP families) and the environmental variance (Table 1; Table 3, steps l and m). A random-number generator was used to independently produce normally-distributed samples of genotypic values and environmental deviations using the variances in table 1. These values were combined to represent the 6000 individuals of the third generation (Table 3; step n).

5. The individuals of this third generation were then ranked using combined index selection (Table 3; step o). The full-sib mating index used estimates for: half-sib mean of parent 1 + half-sib mean of parent 2 from the GCA test; and phenotypic deviation of the individual within the full-sib family blocks (Table 2). The OP index used the half-sib mean and individual phenotypic value, assuming a stand-alone single-tree-plot field design identical to the GCA testing design (Table 2).

6. Sixty individuals were selected from the index ranking for each of the 5 mating schemes (Table 3; step p). Because these selections are for the third-generation breeding population, and because of the concerns regarding maintaining genetic diversity in the breeding population, the restriction that only 4 individuals could be taken from each full-sib cross was added to the pure index selection.

7. Simulated crossing, production of a new generation, and selection were carried on for 2 more generations (gen-

erations 4 and 5). The within-family genotypic and environmental variances used to generate populations in each generation were kept constant and the index weight also remained the same.

8. Monte Carlo simulations were made for 50 independent runs of the intermediate heritability ( $h^2_i = 0.20$ ) and 25 runs each of the high and low heritabilities ( $h^2_i = 0.35$  and  $0.10$ , respectively) for each of the mating scheme treatments. Each run was independent over the full series of generations after the second.

9. The attrition of the genetic base of the breeding subline was also recorded in the simulation between generation 3 and 4. Inbreeding effective population size (KANG and NAMKOONG, 1988) was calculated using ROBERTSON'S (1961) formula of:

$$N_e = \frac{(\sum u_i)^2}{\sum u_i^2} \quad (1)$$

Where  $u_i$  represents the contribution of the  $i^{\text{th}}$  parent. This figure was converted to a relative effective population size as outlined by LINDGREN (1991), but is reported as a percentage of the original parents used in crossing (60 parents). Five restrictions to selection were compared for this exercise: a maximum of 6, 4, 3, 2 and 1 individual per full-sib cross. These 5 different levels of family selection restrictions were applied to all the CP treatments between generations 3 and 4.

Genetic gain ( $\Delta G$ ), based on the cumulative genotypic values over successive generations of the breeding population, was calculated for each mating scheme treatment and for each independent run. Besides means over independent runs, standard errors were obtained which provided the basis for significance testing for comparisons of different mating scheme treatments. In the exercise that looked at the genetic base of the breeding population (9. above) effective population size ( $N_e$ ) was regressed

on gain ( $\Delta G$ ) for each of the different CP treatments in order to examine the functional relationship between gain and effective population size.

## Results and Discussion

### Mating-Scheme Comparisons for Genetic Gain

The results of the Monte Carlo simulations of 5 generations of breeding population development for the intermediate heritability ( $h^2_i = 0.20$ ) are shown in table 5. The differences in the breeding-population gains between the mating schemes were significant ( $P < 0.05$ ). Comparisons with the higher and lower heritabilities were similar to the intermediate heritability and differed only by scale. The major contrast was between the OP scheme and the CP matings; the OP mating had an average mean population gain that was 70% of those of the CP matings.

Even though gain is less for OP mating, it can offer some major advantages. Generation turnover time can be saved. Once a selection is made seed can be collected for the next generation of recurrent selection, thus saving the time of making crosses in the field or the even greater delay of collecting scions, grafting and waiting several years for the crosses to be made. Another factor that is favourable for the OP scheme, is that the fixed resource model that we have used is only for the recurrent selection population. The CP schemes modelled all required an additional GCA test; as the resources for GCA testing are saved in the OP scheme more families can be tested and be available for selection, thereby increasing gains. Because of these advantages OP options have been used in the past and provided the first of the advanced generation series for New Zealand radiata pine (SHELBOURNE et al., 1986; JOHNSON and BURDON, 1990). OP breeding populations have been used in Eucalyptus improvement (FRANKLIN and MESKIMEN, 1983; KING and WILCOX, 1988) and have been advocated by COTTERILL (1986) and COTTERILL et al. (1988).

Table 5. — Simulated gains in the breeding sublimes for 5 mating scheme treatments modelled at the intermediate heritability with maximum of 4 individuals selected per family.

Mating scheme (Table 4)	Cumulative breeding population gain at generation				Percentage breeding population <sup>1</sup> gain per generation				
	2	3	4	5	1-2	2-3	3-4	4-5	Mean <sup>3</sup>
OP-60	34.3	46.6	70.0	87.2	(15.5)	4.8	8.7	5.9	6.5
RM-60	34.3	62.2	87.3	112.6	(15.5)	10.9	8.8	8.2	9.3
AM-60	34.3	64.7	92.6	119.7	(15.5)	11.9	9.7	8.6	10.1
EN-60	34.3	67.0	94.0	122.5	(15.5)	12.8	9.3	9.0	10.4
RM-240	34.3	67.7	97.6	129.4	(15.5)	13.0	10.3	9.9	11.1
Standard error <sup>2</sup> (50 runs)	± 0.5	± 0.8	± 1.0	± 1.1					

<sup>1</sup>) Percentage breeding population gains presents gain as a percentage increase over the previous

$$\text{generation} = \frac{(X_{(i+1)} - X_i)}{X_i} \times 100\%$$

where  $X_i$  = mean of population at generation  $i$ ;  
and  $X_1 = 222$ .

<sup>2</sup>) Standard error of mean pooled over mating schemes.

<sup>3</sup>) Mean of generations 2 to 5.

Within the context of the assumptions for the model, the OP scheme appears to be a reasonable option since 70% of the CP gain can be achieved at either a lower cost or in a shorter time interval. A number of aspects about the model should be noted however:

1) The high OP gains between generation 3 to 4 (8.7% compared with 4.8% in generation 2 to 3) is not a true representation of OP gains. The gains from generation to generation reflect gains of the female side resulting from the previous generation of selection (generation 2 in this case) and gain in the pollen cloud which comes from 2 generations previous (generation 1). This added boost of gain was therefore not a function of only the 60 OP families but also included the effect of the 240 families in generation 1. If only generations 2 to 3 and 4 to 5 are examined to compare gain, one finds that the OP option results in only 53% of RM-60 gain.

2) The model assumes that female flowering in the progeny test is not a constraint thus affecting selection intensity or worst that there is not an adverse correlation between early flowering and rotation-age performance.

3) Another and probably highly fallacious assumption is that pollen within a trial will be equivalent to that of the breeding population of the previous generation. Evidence from realized genetic gain trials comparing New Zealand radiata pine CP and OP seedlots demonstrated that at least 50% of pollen in OP fertilization came from unimproved surround trees rather than the material within the seed orchard. In other words, the assumption that pollen comes from the un-selected breeding population is likely to overestimate OP gains by 25%. Also pollen from within the trial will be primarily from precocious trees which may exhibit the adverse correlation noted in 2) above.

Another feature of the OP scheme is that inbreeding is not as easily controlled as in the CP matings. This might make OP mating less attractive for long-term breeding potential.

The next biggest contrast in genetic gain between the mating schemes resulted from the higher selection intensity offered by the 240 random crosses (RM-240) (8 crosses per parent) over the 60 random crosses (RM-60) (2 crosses per parent) (11.1 vs 9.3; Table 5). This was higher than the contrast between the 'elite' nucleus (EN-60) and RM-60 mating scheme (10.4 vs 9.3; Table 5) and the assortative (AM-60) and RM-60 mating scheme (10.1 vs 9.3; Table 5). Within the limitations of a fixed resource (in this case

60 parents and 6000 progeny) increasing recombinant family selection can be more effective for realizing gains than assortative mating. Therefore, having more crosses from which to select can provide more gain than the use of imperfectly estimated breeding values to practice assortative mating.

The imbalanced schemes, i.e., the 'elite' nucleus (EN-60) and AM-60 matings; showed similar levels of gain with EN-60 showing slightly higher gains over AM-60 (though not significantly). The imbalanced schemes showed an average of 9% more gain per generation over RM (Table 5). BAKER (1973) states that increase in response for AM over RM is rarely expected to exceed 10% and most animal and plant breeding programs would normally achieve only 4% to 5% more gain. Our heritabilities are higher than those presented by BAKER (1973) so our figures fall very much in the range expected for AM. In a simulation study by FERNANDO and GIANOLA (1986) gains for AM over RM could be expected to be up to 20% higher but this required a much higher portion of the population selected than was the case in our study. This result is also confirmed by TALLIS and LEPPARD (1987) who state that the increase in efficiency for AM over RM will be at most 5% if the proportion selected is less than 0.2. Our schemes did not follow positive assortative mating in the true sense (we did not have the expected variance enhancing by crossing poor phenotypes together), rather it provided an enhancing of the breeding population by having a higher mating frequency with good phenotypes.

MAHALOVICH (1990) showed that a breeding population with an elite nucleus offered substantial extra gain (over 100% in some circumstances) over the single unhierarchical mating scheme. Adding positive assortative mating (PAM) to the populations (which varied in size from 12 to 48) did not offer substantial gains over random mating in the elite nucleus with the additive gene model. However with partial dominance, gains for PAM over random mating in the elite nucleus could be up to 10%.

#### Gains and Genetic Base Under Different Restrictions

Although a scheme that intensifies recombinant family selection will maximize per generation gains in the short-term, long-term gain is aided by the maintenance of a large effective population size in breeding sublines. The greatest long-term response to selection has been predicted

Table 6. — Effective population size ( $N_e$ ), relative effective population size ( $N_{er}\%$ ) and generation gain under different restrictions to combined index selection.

Mating Scheme	Restriction - maximum individuals per full-sib cross														
	1			2			3			4			6		
	$N_e$	$N_{er}\%$	$\Delta G\%$	$N_e$	$N_{er}\%$	$\Delta G\%$	$N_e$	$N_{er}\%$	$\Delta G\%$	$N_e$	$N_{er}\%$	$\Delta G\%$	$N_e$	$N_{er}\%$	$\Delta G\%$
RM-60	60	100	100	37	62	151	28	46	167	22	37	175	17	28	182
AM-60	51	86	109	32	54	159	24	41	173	20	33	181	15	25	190
EN-60	41	69	110	25	42	161	19	32	176	16	27	187	13	21	196
RM-240	29	48	172	19	32	189	15	25	198	13	22	201	11	18	209

$N_e$  = Effective population size =  $(\sum u_i)^2 / \sum u_i^2$ , where  $u_i$  = number of contributions of parent  $i$  (ROBERTSON, 1961).

$N_{er}\%$  = Relative effective population size =  $N_e$  expressed as percentage of parents used (60)

$\Delta G\%$  = Gain from generation 3 to 4; simulation for all contrasts started with same 3rd generation values derived by random mating in second generation (a different set of runs than Table 5). Gain is expressed as a percentage of gain for RM-60 at restriction of one individual per family



to occur when half the population is selected each generation (ROBERTSON, 1960). Besides the estimation of gains, the attrition of the genetic resources in the breeding subline by selection was also simulated. Different levels of family selection were applied to the four CP mating schemes in generation 3 to 4. This was done by changing the restrictions to combined index selection. Restrictions of 1, 2, 3, 4 and 6 individuals per full-sib family were applied in separate runs of the CP mating schemes (Table 6). In the random mating with 60 full-sib families and a restriction of only one individual per family, only within family selection is applied and  $N_e$  remains at 60. Comparisons of response to selection are made with this treatment (RM-60 at a restriction of one) representing 100%.

As expected, increasing family selection intensity, by using more recombinant crosses (240 compared to 60), or by reducing the restrictions to within-family selection (2, 3, 4, or 6 instead of 1), increases gain and reduces effective population size. Non-random mating (AM or EN) also increases gain and reduces effective population size over random mating (RM) for a given level of restriction. EN has slightly higher levels of gain over AM but often at marked reduction in effective population size.

Regressions of  $N_e$  on gain were made for each mating scheme treatment (Table 7) using the different restrictions in order to tell if different functional relationships between gain and  $N_e$  existed among mating schemes. RM (60 and 240) slopes and intercepts were the same differing only by selection intensity. The AM slope and intercept did not differ statistically from RM ( $\alpha = 0.05$ ), even though the equation shows it to be less efficient. The EN mating scheme differed both in intercept and slope, and for the range of selection intensities examined more gain could be made for a given level of  $N_e$  with random mating. The lower slope of the EN may lead one to believe that this scheme is more efficient at higher selection intensities, however at these high selection intensities the plots of  $N_e$  to gain become increasingly non-linear.

#### Producing Gain and Maintaining Long Term Genetic Diversity

There is no single solution to the trade-off between maintaining genetic diversity and maximizing gain within the framework of a fixed resource. Even maximizing effective population size by practising only within-family

Table 7. — Intercepts and slopes for the regression of effective population size ( $N_e$ ) on gain  $\Delta G$  for each mating scheme at 5 different restriction levels. Mating scheme parameters are tested against RM-60 with intercept ( $b_0$ ) = 113 and slope ( $b_1$ ) = -3.7.  $N_e = b_0 + b_1 \Delta G$ .

Mating Scheme	Intercept	Test <sup>1</sup>	Slope	Test
AM-60	100	*	-3.2	*
EN-60	78	****	-2.4	****
RM-240	116	NS	-3.7	NS

NS non-significant  $P > 0.1$

\*  $P < 0.1$

\*\*  $P < 0.01$

\*\*\*  $P < 0.001$

\*\*\*\*  $P < 0.0001$

<sup>1</sup>) Tests hypotheses comparing model parameters (Intercept and Slope) against those of RM-60.

selection (the double pair-cross random mating scheme with a selection of one individual per family; Table 6) will provide a 25% probability of loss of particular alleles (LINDGREN, 1986). This loss is before selection and is due to the fact that any particular allele is not transmitted by either gamete involved in the two crosses (50% probability in each cross). LINDGREN (1986) pointed out that the very best advanced generation genotypes constitute a large investment and the major potential of a breeding program and this potential allele loss in the top parents should be considered unacceptable. The promotion of more recombinant crosses with better genotypes (AM or EN schemes) provides higher probabilities of maintaining favorable alleles in the resulting progeny.

The model results show moreover that not all superior genotype combinations come from just the top parents. Having more recombinant crosses across the entire selected breeding population emphasizes that it is through selection (with higher selection intensity), rather than assortative breeding, that gains are made. Promoting more recombinant crosses for selection can result in a greater potential for allele loss through the population as a whole, but as the modelling demonstrated this added loss (reduction in  $N_e$ ) is the cost of increased gain. Allele loss can be controlled through monitoring a diversity index such as effective population size. It is in the selection process that control of genetic gain and diversity are best made. Having more recombinant crosses allows for more and better choices in the balancing between gain and diversity.

A large recombinant population also provides more complex levels of coancestry and hence better inbreeding control. Having many different relatives offers better choices for inbreeding avoidance in further cycles of selective breeding (choices of distant cousins rather than sibs or near cousins). There are arguments to be made however that inbreeding should be promoted rather than avoided. Inbreeding releases genetic variance and helps to unmask recessive alleles. LINDGREN (1986) proposed a strategy that avoids inbreeding overall, yet promotes inbreeding in some of the subline populations. A large recombinant population offers the best way of inbreeding avoidance, while a small highly selected nucleus with selfs offers the best way of inbreeding promotion. Ultimately no single solution exists. The model by examining simple options has in fact highlighted that the breeding populations should be managed for multiple objectives including both short term gains and long term diversity.

#### Assumptions Used and Consequences of Simplifications

Certain assumptions and simplifications were used that require discussion and potential pitfalls in the model need elaboration.

##### 1. Only additive gene effects were modelled

The common breeding strategy for most tree improvement programs is one of population improvement using recurrent selection for GCA (RS-GCA). In this strategy the cumulative genetic improvement sought in the breeding population is based on additive gene effects. Current evidence, for radiata pine and many other species indicates that additive gene effects account for most of the genetic variation and these additive genes are expressed favourably over a variety of sites. Non-additive genetic variances and genotype-environment interaction variances can be significant, but not substantial, compared to additive ge-

netic variance (CARSON, 1986; COTTERILL et al., 1987; JOHNSON, 1992). Where these variances are substantial they can be accounted for in the RS-GCA strategy by a lower heritability (increasing the phenotypic variance). Indications from the lower heritability case ( $h^2_1 = 0.10$ ) are that the treatment comparisons modelled give similar results. Certain seed-production systems can use non-additive genetic variances (tested crosses — CARSON, 1986; clonal selection) but these systems can also be modelled and are accounted for elsewhere (KING and JOHNSON, in MS).

## 2. The trait used is truly polygenic

This is a simplifying assumption, which implies that in over 5 generations there is not truncation of variance due to gene fixation after recombination. Although we worked with this basis polygenic additive model, 2 simplifying (and likely false) assumptions were used.

A) First, the model assumed that the family variance components  $\sigma^2_f$  in matrices used to calculate index weights, b's, were always  $1/4 \sigma^2_A$  and did not change from generation to generation. Truncation brought about by selection in a previous generation will actually reduce  $\sigma^2_f$  such that  $\sigma^2_f < 1/4 \sigma^2_A$ , and this truncation effect will be greatest in the first generation of selection. The result of this assumption is that the index weights are sub-optimal and biased downwards. Gains from combined family-plus-individual selection, however, are usually robust with respect to variations in index weights. True positive assortative designs, where proportions selected are kept high, may offset this truncation phenomena to some degree.

B) Inbreeding did not affect the within-family variation (family selection used a separate outcrossing GCA-test — thus should not be affected). The segregational variation was kept at  $1/2 \sigma^2_A$  and did not change from generation to generation. The better families (from assortative crosses) are likely to contain higher amounts of inbreeding. Ignoring this effect in the model thus biases gains from these selections upwards.

It would be hoped that these 2 simplifying assumptions offset each other to some degree. The comparison between the mating schemes should not be markedly affected by departures from these assumptions.

## 3. The model used only a single trait

Breeding in many commercial forest tree species has been for multiple traits. Although the parameters (Table 1) originated from a volume assessment of an advanced generation progeny test, the result could be applied to any trait with a heritability in the range from 0.10 to 0.35. The results presented here should also apply to a multiple-trait index with a heritability within this range. Where genetic correlations are strong, further refinements to the simulation model would be required.

## 4. The model used only forward selection of offspring as candidates for the next generation breeding population

It is likely that a breeding population will also contain an element of backward selection (re-selected parents) as well. Outstanding genotypes are likely to remain several generations in breeding populations before they are replaced (LINDGREN, 1986). McBRIDE and ROBERTSON (1963) characterized assortative mating as being more powerful when the heritability is high. Because the breeding values of these reselected parents will be better known than their offspring (selection made by progeny mean vs the

individual phenotype and family information of the index): AM should be more effective with these outstanding reselected genotypes. It is likely that imbalanced options and assortative mating features will be more fruitful than our comparisons indicated if they also include the very most outstanding parental genotypes along with the best progeny.

## 5. The model used a closed breeding population

In reality it is likely that the breeding population will remain open. Augmentation of the breeding population from a gene resource population (in many cases this will be natural stands, provenance trials, etc) is planned by many tree improvement programs. Problems are likely to arise when the mean of the breeding population becomes very much higher than the gene resource population. BURDON (1988) offers a systematic way of breeding population augmentation. Constant augmentation will bring new alleles in to the population and enlarge the effective population size. Any augmentation system, however, should not reduce the validity of the comparisons made in our simulation.

## 6. The model used only the objective of crossing and selecting for the breeding population

In reality the production population would also be selected from the same breeding population. The production population, as a forest plantation, does not have to be a depository of genetic diversity; it is not expected to reproduce and it's only requirement is that it has enough genetic buffering to healthily survive until commercial rotation age. Most parents in the breeding population have undergone one to several generations of testing for growth rate, health, and survival. It is very likely therefore that offspring from highly selected seed production parents will have more than adequate genetic resources to thrive for a single commercial rotation. The implications of this are that concerns for maintaining a large effective breeding population size are not needed for production parents. Mating schemes such as EN that are less efficient in maintaining  $N_e$ , but deliver higher gains would be more attractive when the objective also includes production population selections.

The other aspect that cannot be overlooked in any paper that tries to draw conclusions about breeding alternatives is that practical and biological considerations should be paramount in devising any breeding strategy. Precocious flowering species offer more and different opportunities than do recalcitrant species. The open-pollinated option will be more favourable in insect pollinated species where pollen contamination would be expected to be less of a problem and control pollinations can be difficult and expensive. If the resources and capabilities of a program are not able to carry out complicated breeding programs or the delays will be too long then any extra gain offered can be lost and simpler alternatives should be pursued (COTTERILL, 1986). Per generation gains may make OP alternatives attractive if too much time is taken in generation turnround even accounting for the negative assumptions that might have to be made with OP.

## Conclusion

Non-random, imbalanced mating designs in which better parents are crossed more often does increase the per generational gains by nearly 10%. However there is a cost associated in terms of reduced effective population size.

Balanced random mating designs with recombinant family selection provide higher levels of genetic gain with reduced losses to effective population size than the non-random designs. Among the non-random designs the elite nucleus was especially damaging on  $N_e$  whereas the assortative scheme showed results not too dissimilar from the random scheme. It would appear that gain and diversity (using effective population size) are best managed in the breeding population through the selection process. A mating scheme that produces many recombinant crosses for selection offers more choices for this management than a scheme that attempts to assortatively recombine parents. An elite or assortative element for the breeding population may best fit when the objectives are combined with production population selections where gain should be maximized and diversity is less of a concern. Simpler schemes such as OP mating may have a place in breeding population development depending on the assumptions and constraints that a breeding program can live with.

#### Acknowledgements

We would like to thank the scientists and technical officers of the Forest Research Institute in Rotorua whom we had the great pleasure of being involved with in the advanced generation planning of the radiata pine program. We would like to acknowledge PAUL COTTERILL for questioning some early assumptions we held thus spurring us on to do this project. We would like to express thanks for the constructive reviews and help provided by ROWLAND BURDON, MIKE CARSON, DAG LINDGREN, BILL LOWE, MARY MAHALOVICH, TONY SHELBOURNE and ALVIN YANCHUK. ROWLAND BURDON and DAG LINDGREN especially provided many thoughtful comments on this work.

#### References

- BAKER, R. J.: Assortative mating and artificial selection. *Heredity* 31: 231–238 (1973). — BURDON, R. D.: Clonal forestry and breeding strategies — a perspective. In: Proceedings, IUFRO Working Parties on Breeding Theory, Progeny Testing, and Seed Orchards. Williamsburg, Virginia (1986). — BURDON, R. D.: Recruitment for breeding populations: objectives, genetics and implementation. In: WEIR, B. S., EISEN, E. S., GOODMAN, M. M. and NAMKOONG, G. (eds). Proc. of the Second Int. Conf. on Quantitative Genetics. Sinauer, Sunderland, MA 01375, USA (1988). — BURDON, R. D. and SHELBOURNE, C. J. A.: Breeding populations for recurrent selection: conflicts and possible solutions. *New Zealand J. For. Sci.* 1: 174–193 (1971). — CARSON, M. J.: Control-pollinated seed orchards of best general combiners — a new strategy for radiata pine improvement. In: Plant Breeding Symp DSIR 1986. *Agron. Soc. NZ Spezial Publ.* No. 5 (1986). — COTTERILL, P. P.: Breeding strategy: Don't underestimate simplicity. In: Proc. IUFRO Joint Meeting of Working Parties on Breeding Theory, Progeny Testing and Seed Orchards, Williamsburg, VA, pp 8–23 (1986). — COTTERILL, P. P., DEAN, C., CAMERON, J. and BRINDSBERGS, M.: Nucleus breeding: a new strategy for rapid improvement under clonal forestry. In: Proc IUFRO Meeting on Breeding Tropical Trees, Pattaya, Thailand, 1988 (1988). — COTTERILL, P. P., DEAN, C. A. and VAN WYCK, G.: Additive and dominance genetic effects in *Pinus pinaster*, *P. radiata* and *P. elliotti* and some implications for breeding strategy. *Silv. Genet.* 36: 221–232 (1987). — CROSBY, J. L.: Computer simulation in genetics. John Wiley and Sons Ltd, N. Y., 477 pp. (1973). — DEMPFLE, L.: A note on increasing the limit of selection through selection within families. *Genet. Res.* 24: 127–135 (1974). — FALCONER, D. S.: Introduction to quantitative genetics. Longman, Inc., 340 pp. (1981). — FERNANDO, R. L. and GIANOLA, D.: Effect of assortative mating on genetic change due to selection. *Theor. Appl. Genet.* 72: 395–404 (1986). — FRANKLIN, E. C. and MESKIMEN, G.: Choice of species and provenances in cold summer rainfall climates. *Proc. IUFRO Colloque Inter. sur les eucalyptus resistant au froid*. Afocel, Nangis, France, pp 341–356 (1983). — GIANOLA, D.: Assortative mating and genetic correlation. *Theor. Appl. Genet.* 62(3): 225–231 (1982). — JAMES, J. W.: Design of animal breeding programmes. In: *Evolution and Animal Breeding* (edited by W. G. HILL and T. F. C. MACKAY). Wallingford, Oxon, U. K., C. A. B., pp. 189–195. (1989). — JOHNSON, G. R. and BURDON, R. D.: Family-site interaction in *Pinus radiata*: implications for progeny testing strategy and regionalised breeding in New Zealand. *Silvae Genet.* 39: 55–62 (1990). — JOHNSON, I. G.: Family-site interactions in radiata pine families in New South Wales, Australia. *Silvae Genet.* 41: 55–60 (1992). — KANG, H. and NAMKOONG, G.: Inbreeding effective population size under some artificial selection schemes. I. Linear distribution of breeding values. *Theor. Appl. Genet.* 75: 333–339 (1988). — KING, J. N., CARSON, M. J. and JOHNSON, G. R.: Selections from a disconnected diallel series for a third generation of breeding for New Zealand radiata pine (in MS). — KING, J. N. and JOHNSON, G. R.: Computer simulation for production- population gain including clonal selection options. (in MS). — KING, J. N. and WILCOX, M. D.: Family tests as a basis for the genetic improvement of *Eucalyptus nitens* in New Zealand. *N. Z. Jour. of Forestry Science* 18: 253–266 (1988). — LINDGREN, D.: How should breeders respond to breeding values? In: Proc. IUFRO Joint Meeting of Working Parties on Breeding Theory, Progeny Testing and Seed Orchards. Williamsburg, VA, pp 361–372 (1986). — LINDGREN, D.: Optimal utilization of genetic resources. *Forest Tree Improvement* 23: 49–67 (1991). — LOVE, W. J. and VAN BUIJTENEN, J. P.: The incorporation of early testing procedures into an operational tree improvement program. *Silvae Genetica* 38(5–6): 157–252 (1989). — MADALENA, F. E. and HILL, W. G.: Population structure in artificial selection programmes: simulation studies. *Genet. Res.* 20: 75–99 (1972). — MAHALOVICH, M. F.: Modeling positive assortative mating and elite populations in recurrent selection programs for general combining ability. PhD Thesis. N. C. State University, Raleigh N. C. 104pp (1990). — MCBRIDE, G. and ROBERTSON, A.: Selection using assortative mating in *Drosophila melanogaster*. *Genetical Res.* 4: 356–369 (1963). — MCKEAN, S. E. and BRIDGWATER, F. E.: Third-generation breeding strategy for the North Carolina State University-Industry Cooperative Tee Improvement Program. Proc: IUFRO Section S2.02-08, Breeding Tropical Trees, Resolving Tropical Forest resource Concerns through Tree Improvement, Gene Conservation and Domestication of New Species. Cartagena, Colombia. October 9–18, 1992. (1992). — ROBERTSON, A.: A Theory on the limits of artificial selection. *Proc. of the Royal Society of London. B.* 153, 234–249 (1960). — ROBERTSON, A.: Inbreeding in artificial selection programmes. *Genet. Res.* 2: 189–194 (1961). — SAS INSTITUTE: SAS user's guide. SAS Institute, Cary NC, 584 pp (1985). — SHELBOURNE, C. J. A., CARSON, M. J. and WILCOX, M. D.: New techniques in the genetic improvement of radiata pine. *Commonw. For. Rev.* 68: 191–201 (1989). — SHELBOURNE, C. J. A., BURDON, R. D., CARSON, S. D., FIRTH, A. and VINCENT, T. G.: Development plan for radiata pine breeding. Ministry of Forestry, Forest Research Institute, Rotorua, NZ. 142 pp (1986). — TALLIS, G. M. and LEPPARD, P.: The joint effects of selection and assortative mating on a single polygenic character. *Theor. Appl. Genet.* 75: 41–45 (1987).