

Efficiency of including first-generation information in second-generation ranking and selection: results of computer simulation

Terrance Z. Ye · Keith J. S. Jayawickrama
G. R. Johnson

Received: 30 April 2006 / Revised: 24 August 2006 / Accepted: 30 October 2006 / Published online: 20 December 2006
Springer-Verlag 2006

Abstract Using computer simulation, we evaluated the impact of using first-generation information to increase selection efficiency in a second-generation breeding program. Selection efficiency was compared in terms of increase in rank correlation between estimated and true breeding values (i.e., ranking accuracy), reduction in coefficient of variation of correlation coefficients (i.e., ranking reliability), and increase in realized gain, with best linear unbiased prediction (BLUP). The test populations were generated with varying parameters: selection strategy (forward vs backward selection of parents); number of parents (24–96); number of crosses per parent (1–8); heritability (0.05–0.35); ratio of dominance to additive variance (0–3); ratio of additive-by-site to additive variance (0–3); and ratio of dominance-by-site to additive variance (0–3). The two selection strategies gave distinct results. When parents of the second-generation crosses had been selected via backward selection, adding first-generation information markedly increased selection efficiency. Conversely, when parents had been selected via forward selection, first-generation information provided little increase in efficiency. The amount of increase depended more on heritabilities in both generations and less on dominance and genotype-by-environment effects. Including first-generation

information helped more when there were many parents and few crosses per parent in the second generation. Only in the case of extremely low first-generation heritabilities was there no benefit to adding first-generation information in terms of improved ranking reliability and accuracy.

Keywords Stochastic simulation · Selection efficiency · BLUP · First-generation information · Second-generation selection

Introduction

Most forest-tree breeding programs are based on recurrent selection (e.g., White et al. 1993; McKeand and Bridgwater 1998; Jayawickrama and Carson 2000), with the premise that gains will be cumulative over generations. In each generation, genetic testing is one of the most important and expensive facets of the breeding programs. Although there are many reasons for genetic tests, estimating genetic parameters, ranking and selecting parents (i.e., backward selection), and ranking and selecting trees (i.e., forward selection) are the three primary reasons (Libby 1973; White 1987, 1996; Lindgren 1991).

Obtaining precise estimates or ranks requires many families and adequate numbers of trees per family (White 1996). The actual scale of the test, however, generally is dictated by the quantity of seed available and economic resources. Many breeding programs operate with austere budgets, so breeders must limit the number of test sites and genetic materials.

As programs move to their second and subsequent generations, testing data from previous generations are usually available. Such data were often obtained with great

Communicated by Sally Aitken

T. Z. Ye (✉) · K. J. S. Jayawickrama
Department of Forest Science, Oregon State University,
321 Richardson Hall,
Corvallis, OR 97331, USA
e-mail: Terrance.Ye@vregonstate.edu

G. R. Johnson
USDA Forest Service, Pacific Northwest Research Station,
3200 SW Jefferson Way,
Corvallis, OR 97331, USA

cost and effort. For example, first-generation tests for coastal Douglas-fir in the Pacific Northwest included around 4.2 million progeny (Lipow et al. 2003). At the same time, recent improvements in computer hardware and software make it possible to use large data sets efficiently. It is therefore timely for breeders to investigate the merits of integrating data from previous generations that may increase the accuracy and reliability of performance estimates.

In theory, the efficiency of such combined analysis depends on genetic relatedness between generations as well as the degree of genetic relationships that already exists in the current generation test. Historically, animal breeders rely heavily on using records of the performance of ancestors to increase the accuracy of estimating an animal's breeding value (Skjervold and Odegard 1959; Young 1961; Lasley 1987). However, unlike in forestry, they are limited in the number of progeny that can be evaluated.

Best linear unbiased prediction (BLUP) is routinely used in forest tree improvement programs (White and Hodge 1989; White 1996; Dutkowski et al. 2002). In BLUP, related individuals contribute to the predictions of each other; it therefore is believed that integrating data from all generations would improve the efficiency in predicting genetic values by increasing the effective number of observations available for evaluating each genotype (White and Hodge 1989; Falconer and Mackay 1996; Kerr et al. 2004). Johnson (1998), however, found that including historical parental information actually increased the efficiency of estimating breeding values very little in a computer simulation, because the correlation coefficient and percent increase in orchard gain increased little in all scenarios. Although his study provided useful information about across-generation combined analysis, it was based on one specific combination of genetic parameters; therefore, the results were limited in scope. To our knowledge, no systematic study has been conducted to address the issue.

The objective of this study was to determine whether including first-generation information markedly enhances the accuracy and reliability of second-generation selection. Varying selection strategies and combinations of genetic parameters were examined with stochastic data sets generated by computer simulation.

Materials and methods

Genetic models and simulation processes

An initial first-generation base population of 30,000 individuals from 300 unrelated half-sib families (i.e., 100 individuals per family) was generated, from which N_1 individuals were selected based on their estimated breeding

values. The simulation used a parameter-based model in which the assumed additive genetic variance and heritability (h_{ai}^2) of individual trees were used to generate independent normal distributions of genotypic values and environmental deviations (King and Johnson 1993; Johnson 1998). Experimental design was ignored in the base population, primarily, for simplicity. The following genetic model was used for estimating breeding values for the base population:

$$P = A + E \quad (1)$$

where P is the phenotypic value; A is the additive genetic effect; E is the residual deviation (where environmental effects dominate).

Two selection strategies were applied to the base population for comparison: (1) forward selection (i.e., selecting the best individual from each of the top N_1 families), and (2) backward selection (i.e., selecting the best N_1 parents). The selections were then used as second-generation parents and crossed in a disconnected $N_2 \times N_2$ factorial mating design. The progeny were assumed to be tested on six sites with 20 trees per cross per site and single-tree plots.

In the second-generation population, the phenotypic value of an individual was defined as the sum of genetic effects, environmental effect, and genotype-by-environment interactions:

$$P = \text{Site} + A + D + A \times \text{Site} + D \times \text{Site} + E \quad (2)$$

where P , A , and E are the same as in Eq. 1; D is the dominance effect; Site is the site effect; $A \times \text{Site}$ and $D \times \text{Site}$ are the additive-by-site and dominance-by-site effect, respectively. Although a randomized complete block design was assumed at each site, the block effect was ignored, as it can be removed from tree estimates and does not affect relative breeding value estimation and family ranking.

Each genetic effect in Eq. 2 has two components, among-family and within-family, as detailed below. For the additive genetic effect, the among-family component is the mean of parental breeding values. The within-family component is the Mendelian sampling (Bulmer 1985), which is normally distributed with a mean of zero and a variance that is half the additive variance.

$$A = \frac{1}{2}(A_F + A_M) + N(0, \frac{1}{2}\sigma^2)$$

$$= \frac{1}{2}(A_F + A_M) + k_0 \quad (3)$$

where A_F and A_M are the breeding values of female and male parents, respectively; k_0 is a normal deviate drawn from a distribution of random numbers with a mean of 0 and variance of 1; σ^2 is the additive variance.

For dominance (or dominance-by-site) effect, both the among-family and within-family components are drawn randomly from normal distributions, with mean equal to zero and variances σ_D^2 (or $\sigma_{D \times E}^2$) and $\sigma_{D \times E}^2$ (or $\sigma_{D \times E}^2$), respectively.

$$D = N(0, \sigma_D^2) + N(0, \sigma_{D \times E}^2) \tag{4}$$

$$D \times E = N(0, \sigma_{D \times E}^2) + N(0, \sigma_{D \times E}^2) \tag{5}$$

where k_1, k_2, k_3 , and k_4 are random normal deviates; σ_D^2 and $\sigma_{D \times E}^2$ are the dominance and dominance-by-site variance, respectively.

Similarly, additive-by-site effect can be simulated by using the formula

$$AE = N(0, \sigma_{AE}^2) + N(0, \sigma_{AE}^2) \tag{6}$$

where k_5 and k_6 are random normal deviates; σ_{AE}^2 is the additive-by-site variance.

Environmental effect is assumed to be normally distributed with mean zero and variance σ_E^2 :

$$E = k_7 \tag{7}$$

where k_7 is a random normal deviate; σ_E^2 is the environmental variance, which is assigned and adjusted by specifying the narrow-sense heritability (h^2).

All simulation data were generated with a program written in SAS[®] Macro (SAS Institute Inc. 2004).

Assessment and ranking

Estimates of genetic and phenotypic parameters for the population and breeding values for each parent and individual were obtained by using an individual-tree BLUP model (Henderson 1984) with either second-generation data only or data from both generations, as well as their pedigree information:

$$y = Xb + Zu + e \tag{8}$$

where y is the vector of observations; b is a vector of fixed effects with its incidence matrix X ; u is a vector of random

effects with its incidence matrix Z ; and e is the vector of random residual terms. The random effects (u and e) were assumed to have independent multivariate normal distributions with zero means and common variances/covariances. The model included a fixed effect for site and random effects for genetic (additive and dominance) merits and their interaction with site. When analyzing combined data from both generations, we treated the generation effect as a genetic group effect and incorporated it into the mixed model. Parents and individuals were ranked on the basis of estimated breeding values. The breeding values and the variance components used in the BLUP analyses were estimated with ASReml (Gilmour et al. 2002).

Simulation scenarios

One hundred and eighty scenarios were tested with various combinations of the parameters in the simulation. The following parameters or selection strategies were assumed. These parameters were initially based on the cooperative Douglas-fir breeding program in the Pacific Northwest and extended to fit the genetic patterns of growth traits in most coniferous species.

For the first-generation population:

- Initial additive variance was fixed at 10;
- Individual-tree heritability h^2 : 0.05, 0.15, 0.25, and 0.35;
- Number of selections N_1 : 24, 48, 72, and 96;
- Two selection strategies: strategy I-forward selection (i.e., family selection including the top N_1 families and then the best individual from each of these families) and strategy II-backward selection (i.e., selection of the best N_1 parents);

For the second-generation population:

- Disconnected $N_2 \times N_2$ factorial mating design, where N_2 was varied: 1 (single-pair mating), 2, 3, 4, 6, and 8.
- Individual-tree heritability h^2 : 0.05, 0.15, 0.25, and 0.35;
- Ratio of dominance to additive variance σ_D^2/σ_A^2 : 0, 0.5, 1, 2, and 3;
- Ratio of additive-by-site to additive variance σ_{AE}^2/σ_A^2 : 0, 0.5, 1, 2, and 3;
- Ratio of dominance-by-site to dominance variance $\sigma_{D \times E}^2/\sigma_D^2$: 0, 0.5, 1, 2, and 3.

For each combination of parameters (i.e., scenario), 1,000 independent runs were conducted. If any estimated parameter from a simulated data set deviated from the predefined parameter by more than one standard deviation, this data set was excluded. This approach reduced the extensive overlap in some estimated variance compo-

nents among scenarios. As a result, the actual number of independent runs for each scenario was 500 or above.

Summarization

Various statistics were computed to measure the increase in efficiency of the second-generation selection by including the first-generation information.

Kendall's rank correlations (tau-b) between the estimated breeding values (EBV) and the true breeding values (TBV) for both parental and individual selections were calculated for each scenario. Kendall's tau-b has the advantages of a known sampling distribution and a direct interpretation in terms of probabilities of observing concordant and discordant pairs (Conover 1980). The mean and coefficient of variation of the correlations' coefficients from each scenario were used to quantify the accuracy and reliability of selection, respectively. The increase in accuracy obtained by adding the first-generation information was measured as the average percent improvement in the mean of rank correlation coefficients. Similarly, increased reliability was measured as the percent reduction in the coefficient of variation of correlation coefficients.

Realized gains were estimated for seed orchard roguing by reselecting the top 30% of the second-generation parents and for forward selection by choosing the best progeny from each of the top 30% of second-generation families. They were calculated as percentages of the mean TBV of the EBV-based selections over that of the TBV-based selections; i.e., they represented the percentage of the

potential gain that could be achieved for each selection scenario.

Results and discussion

Impact of selection strategy for second-generation parents

Tree breeders typically use forward (strategy I) or backward (strategy II) selections, or both, from the previous generations to establish breeding populations for the current generation. Figure 1 shows the comparison between these two selection strategies based on a typical scenario with $h_{(1)} = h_{(2)} = 0.25$, $\sigma_{D \times E} = 0.5\sigma_A$, $\sigma_{X_E} = \sigma_A^2$, $N_1=48$, and $N_2=2$, which conforms to the genetic patterns and breeding strategy found in the cooperative Douglas-fir breeding programs in the Pacific Northwest. When strategy II was employed, including first-generation information increased the accuracy by 8.5%, reliability by 22.0%, and realized gain by 5.4% during the second-generation parental ranking and reselection. The added efficiency decreased markedly, however, when the second-generation parents were initially selected via forward selection (strategy I). The corresponding increases in ranking accuracy, reliability, and realized gain were 1.4, 6.2, and 0.9%, respectively. This is consistent with the observation of Johnson (1998) who estimated that adding first-generation information when only forward selection was used increased Pearson's correlation between true and estimated breeding values by less than 2% and orchard gain from roguing by less than 1%.

Fig. 1 Percent increase in accuracy, reliability, and realized gain by adding first-generation information for $h_{(1)}^2 = h_{(2)}^2 = 0.25$; $\sigma_{D \times E} = 0.5\sigma_A$; $\sigma_{X_E} = \sigma_A^2$; a disconnected 2x2 factorial mating design in second-generation with 48 parents. The baseline gains (based on second-generation data only) for reselecting the top 30% of the second-generation parents and for selecting the best progeny from each of the top 30% of second-generation families were 67% and 56%, respectively, for strategy I, and 71% and 53%, respectively, for strategy II

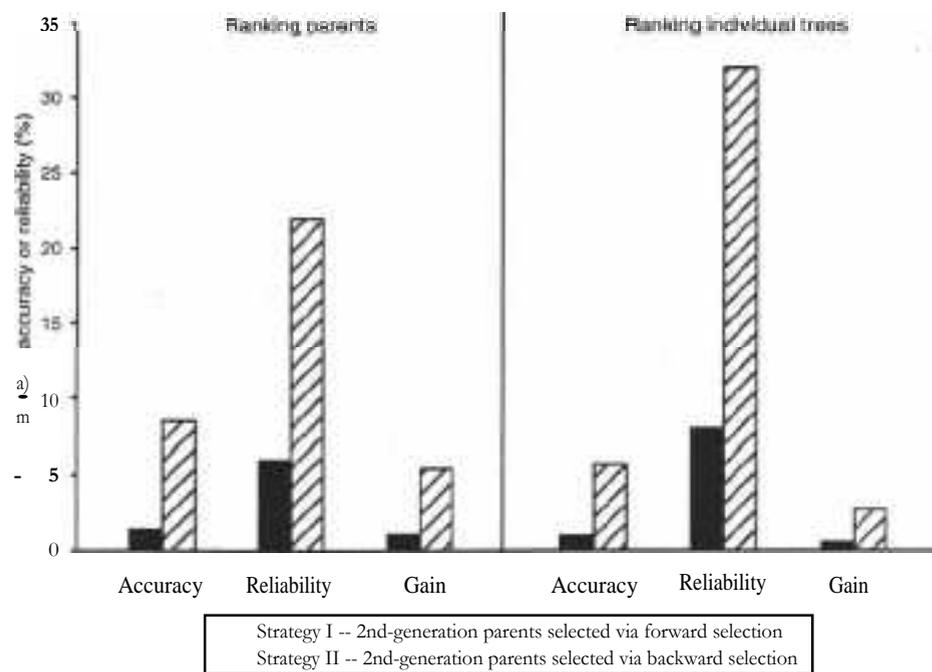


Table 1 Kendall's rank correlation coefficient (r) between the true and estimated breeding values and their coefficient of variation (in parentheses) for both parental and tree selections, based on >500 simulations per scenario, for $h_{a_{II}} = h \sim \sim_1 = 0.25$; $\sim O = a_{2D_{3E}} = 0.5c$; $e_{A \times E} = aA$

Number of parents	Number of crosses per parent	Selection in first-generation	Data used in analysis	Second-generation population	
				Parents	Progeny
24	2	Backward	First-generation data	0.477 (25.358)	
			Second-generation data	0.549 (16.248)	0.472 (9.698)
			Both	0.595 (14.768)	0.498 (8.977)
		Forward	First-generation data	0.201 (74.164)	
			Second-generation data	0.510 (18.687)	0.441 (10.348)
			Both	0.520 (17.971)	0.446 (10.259)
48	2	Backward	First-generation data	0.511 (14.381)	
			Second-generation data	0.553 (11.330)	0.492 (6.354)
			Both	0.600 (9.909)	0.520 (5.565)
		Forward	First-generation data	0.205 (46.253)	
			Second-generation data	0.497 (13.258)	0.450 (7.431)
			Both	0.504 (12.924)	0.454 (7.110)
72	2	Backward	First-generation data	0.540 (10.408)	
			Second-generation data	0.563 (8.439)	0.487 (5.427)
			Both	0.615 (7.266)	0.518 (4.648)
		Forward	First-generation data	0.230 (33.641)	
			Second-generation data	0.512 (10.363)	0.444 (6.287)
			Both	0.522 (9.987)	0.448 (6.169)
96	2	Backward	First-generation data	0.565 (8.154)	
			Second-generation data	0.569 (7.327)	0.491 (4.707)
			Both	0.624 (6.172)	0.525 (3.941)
		Forward	First-generation data	0.227 (26.611)	
			Second-generation data	0.509 (8.785)	0.451 (5.285)
			Both	0.519 (8.541)	0.456 (5.219)
48	1	Backward	First-generation data	0.520 (13.845)	
			Second-generation data	0.530 (12.705)	0.462 (7.781)
			Both	0.579 (10.825)	0.504 (6.692)
		Forward	First-generation data	0.219 (44.829)	
			Second-generation data	0.437 (15.453)	0.447 (8.419)
			Both	0.443 (15.529)	0.453 (8.515)
	2	Backward	First-generation data	0.511 (14.381)	
			Second-generation data	0.553 (11.330)	0.492 (6.354)
			Both	0.600 (9.909)	0.520 (5.565)
		Forward	First-generation data	0.205 (46.253)	
			Second-generation data	0.497 (13.258)	0.450 (7.431)
			Both	0.504 (12.924)	0.454 (7.110)
	3	Backward	First-generation data	0.512 (14.532)	
			Second-generation data	0.584 (9.856)	0.493 (6.115)
			Both	0.633 (8.642)	0.516 (5.564)
		Forward	First-generation data	0.208 (46.719)	
			Second-generation data	0.556 (10.616)	0.456 (6.258)
			Both	0.564 (10.508)	0.460 (6.323)
4	Backward	First-generation data	0.511 (14.934)		
		Second-generation data	0.606 (8.557)	0.505 (5.184)	
		Both	0.656 (7.452)	0.524 (4.823)	
	Forward	First-generation data	0.209 (46.569)		
		Second-generation data	0.589 (9.834)	0.463 (5.781)	
		Both	0.597 (9.365)	0.467 (5.980)	
5	Backward	First-generation data	0.511 (14.094)		
		Second-generation data	0.616 (8.405)	0.513 (4.610)	
		Both	0.668 (7.134)	0.530 (4.205)	

Table I (continued)

Number of parents	Number of crosses per parent	Selection in first-generation	Data used in analysis	Second-generation population	
				Parents	Progeny
		Forward	First-generation data	0.206 (45.332)	
			Second-generation data	0.619 (8.787)	0.472 (4.885)
			Both	0.628 (8.249)	0.477 (5.234)
		Backward	First-generation data	0.510 (14.707)	
			Second-generation data	0.615 (9.083)	0.516 (4.807)
			Both	0.669 (7.026)	0.533 (4.455)
		Forward	First-generation data	0.201 (48.014)	
			Second-generation data	0.640 (8.376)	0.478 (4.741)
			Both	0.648 (7.949)	0.483 (5.136)

Contribution of historical information depends on the quality and the quantity of the related genetic information. In our simulations, the relationship between the first-generation progeny and the second-generation parents is between offspring and parents when backward selection is applied and between half sibs when forward selection is applied. In other words, adding first-generation information in strategy II is like adding more progeny to estimate a parent's breeding value and only adding half-sibs in strategy I. From quantitative genetic theory, the genetic covariance of offspring and one parent is half the additive genetic variance of the parents, whereas the genetic covariance among half sibs is only a quarter of the additive variance. Thus, with backward selection, the second-generation population maintains a higher level of genetic connection with the population of the previous generation. In contrast, the genetic correlation between both generations is expected to be low with forward selection. As shown in Table 1, the rank correlation (r) between the true breeding values (TBV) and estimated

breeding values (EBV) based on first-generation data for each combination of number of parents and crosses was much higher for strategy II ($r=0.50$) than that for strategy I (0.21).

A few breeding programs use controlled crosses in the first-generation progeny tests. In these programs, the added efficiency with forward selection is expected to be higher, as full sibs are related by 0.50. To prove this, we replaced the 300 unrelated half-sib families with 150 unrelated full-sib families in the initial base population (300 parents, single-pair mating, $h^2=0.25$). The rank correlation (r) became much closer between strategy II (0.54) and strategy I (0.446).

Adding first-generation information reduced variation of the estimate of rank correlation between EBV and TBV and therefore increased ranking reliability (Fig. 1). The percent increase in ranking reliability was 6-8% for strategy I, which is lower than the 22-32% increase for strategy II, but is still not negligible. Increases in ranking reliability likely were primarily due to the advantage of incorporating the

Fig. 2 Effect of heritability on the accuracy and reliability of ranking second-generation parents by adding first-generation information for $h^2_{11} = \mathcal{I}^2_{121} = 0.25$; $Q_U = \sigma_{U_rE} - 0.5aA$; $a = 3$; a disconnected 2×2 factorial mating design in second-generation with 48 parents selected via backward selection (strategy II)

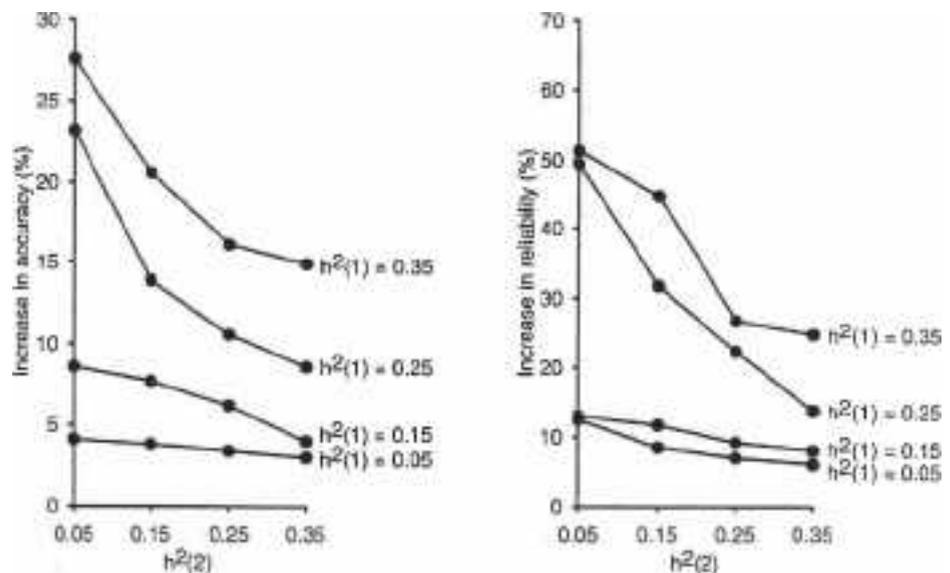


Table 2 Effect of dominance and additive G × E interaction on the accuracy and reliability of ranking second-generation parents by adding first-generation information in a disconnected 2 × 2 factorial mating design with 48 parents selected via backward selection

$h^2(1)$	$h^2(2)$	σ^2		Percent increase in	
				Accuracy	Reliability
0.05	0.05	0	1	2.9	9.2
				4.1	11.7
		2	1	7.3	9.8
				7.7	12.5
0.15	0.15	0	1	10.2	21.3
				10.6	22.7
		2	1	12.4	22.4
				14.9	24.8
0.25	0.25	0	1	14.1	43.9
				27.6	51.3
		2	1	30.0	55.7
				31.7	56.3
0.35	0.35	0	1	2.7	6.5
				4.1	11.7
		2	1	6.0	8.1
				7.7	12.5
0.35	0.05	0	1	10.1	22.3
				10.6	22.7
		2	1	12.0	23.3
				14.9	24.8
0.35	0.15	0	1	22.4	33.4
				27.6	51.3
		2	1	39.6	65.8
				42.5	79.6

genetic relationship matrix, as that increases the number of related observations for an entry.

Our results also indicated that including first-generation information helped more in reselecting parents than in

selecting progeny in the second-generation population. This is not surprising, as the coefficient of relatedness between the first-generation progeny and the second-generation parents is twice that of the relationship between the progeny across generations.

Impact of genetic parameters

When second-generation parents were selected via strategy II, the percent increase in accuracy and reliability varied greatly among different combinations of genetic parameters. It generally depended more on the heritabilities in both generations and less on dominance and genotype-by-environment (G × E) effects.

The value of adding first-generation information was relatively high when $h^2(1)$ was high and $h^2(2)$ was low and diminished as $h^2(1)$ decreased and $h^2(2)$ increased (Fig. 2). For example, with $h^2(1) = 0.25$ and $\sigma^2 = \sigma^2_{A \times E} = \sigma^2_{ixE}$, the increase in ranking accuracy for second-generation parents decreased from 23.2 to 8.6% as $h^2(2)$ increased from 0.05 to 0.35; however, the corresponding decrease was only from 8.6 to 4.0% when $h^2(1) = 0.15$. First-generation data have a high ratio of genetic information to random noise when $h^2(1)$ is high and, therefore, provides a higher contribution to second-generation evaluation. This trend is even more pronounced at a low value of $h^2(2)$ (Fig. 2).

Dominance and the additive G × E interaction showed a similar pattern of effect on changes in the efficiency of selection. The impact of adding first-generation information increased as σ^2_D or $\sigma^2_{O \times E}$ increased, but the percent increase in ranking accuracy and reliability was usually less than 5% when $\sigma^2_D < 0$ or $\sigma^2_{A \times E} < \sigma^2_A$ (situations typical of many breeding programs). There were, however, some exceptions. For example, with very high $h^2(1)$ and very low σ^2_D ,

Fig. 3 Effect of heritability on the accuracy and reliability of ranking second-generation parents by adding first-generation information for $h^2(1) = h^2(2) = 0.25$;

$\sigma^2_{ixE} = 0.5\sigma^2_R$; $\sigma^2_D = 0$; a disconnected 2 × 2 factorial mating design in second-generation with 48 parents selected via forward selection (strategy I)

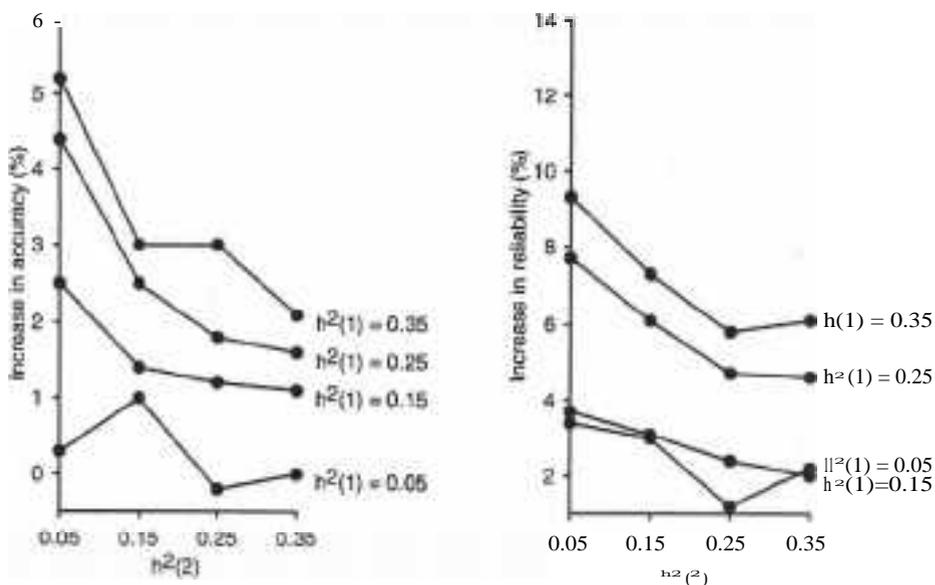
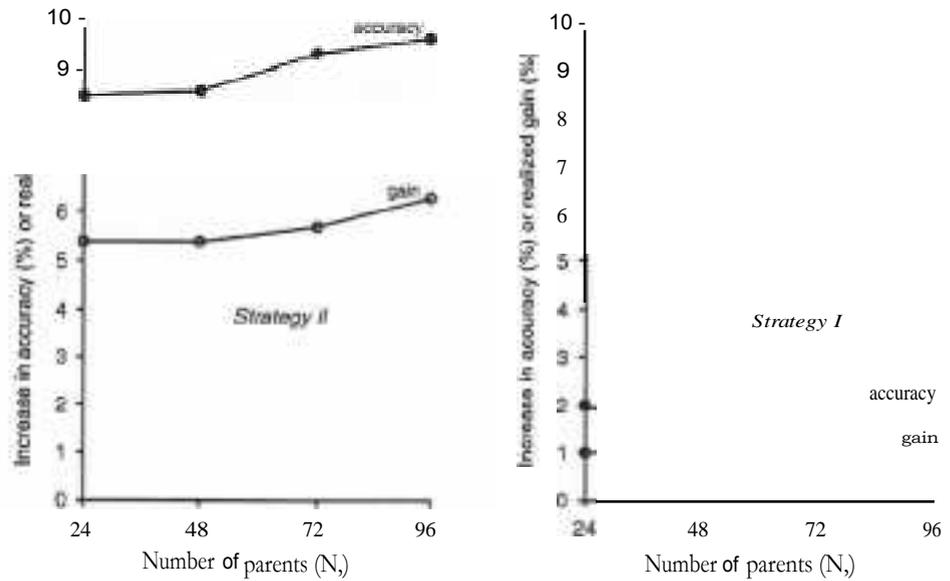


Fig. 4 Percent increase in accuracy and realized gain for ranking second-generation parents with different numbers of parents by adding first-generation information for $h_{11} = h_{22} = 0.25$; $c = 0.50$; $\sigma_{D \times E} = 0.50$; $\sigma_{A \times E} = 0$; a disconnected 2×2 factorial mating design in second-generation with parents selected via backward selection (strategy II) or forward selection (strategy I)



when c increased from 0 to aA , the percent increase in ranking accuracy could be up to 14% (Table 2).

The impact of the dominance $G \times E$ interaction on the added efficiency was similar to that of the additive $G \times E$ interaction but was about 50% smaller (details not shown).

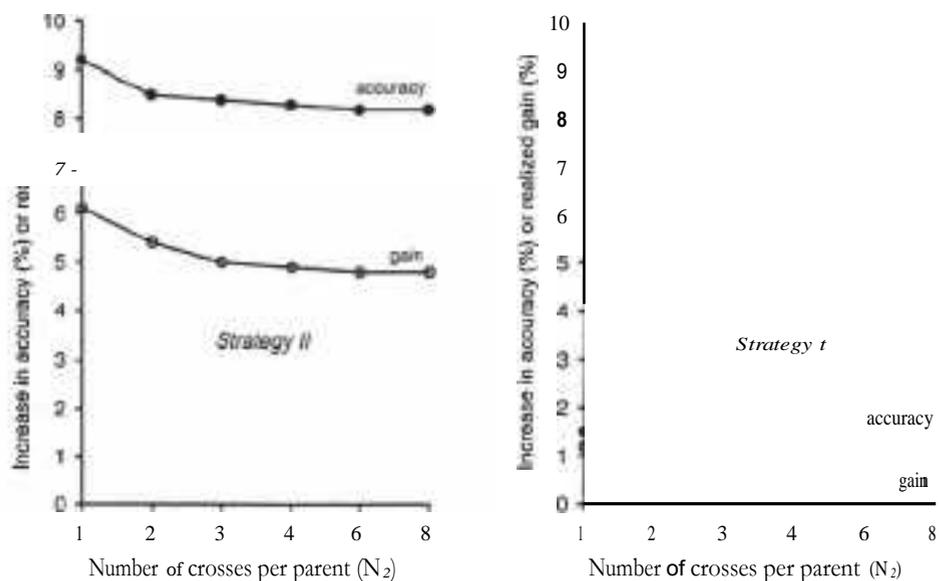
When the second-generation parents were selected via strategy I, the impacts of genetic parameters followed similar patterns as via strategy II, but overall, the first-generation information did little to increase ranking accuracy in the second-generation (Fig. 3). For example, the percent increase dropped to one-sixth to one-third of that in strategy II at $h_{11} = h_{22} = 0.25$. Note that when h_{11} was very low, the first-generation data simply added random noise for the second-generation progeny selections. The ranking reliability of second-generation parents could

still increase up to 10-14%, probably due to the increase in the relative quality of the first-generation data at high h_{11} and low c .

Impact of number of parents and number of crosses per parent

With strategy II, ranking accuracy and realized gain improved consistently as more first-generation parents were selected (Fig. 4). The percent increase in accuracy of ranking the second-generation parents was 8.3% for 24 parents and up to 9.6% for 96 parents in a typical scenario where $h_{11} = h_{22} = 0.25$, $c = 0.50$, $\sigma_{D \times E} = 0.50$, $\sigma_{A \times E} = 0.6A$, and $N_2 = 2$. The corresponding increase in realized gain was from 5.3 to 6.3%. Trends for ranking and selecting the

Fig. 5 Percent increase in accuracy and realized gain for ranking second-generation parents with different numbers of crosses per parent by adding first-generation information for $q_{11} = h_{22} = 0.25$; $\sigma_{D \times E} = 0.50$; $\sigma_{A \times E} = 0$; a disconnected factorial mating design in second-generation with 48 parents selected via backward selection (strategy II) or forward selection (strategy I)



second-generation progeny were similar. Such increases were apparently in concordance with progressive increase in rank correlation between generations. For example, as number of parents increased from 24 to 96, rank correlation increased from 0.477 up to 0.565 (Table 1). These increased correlations are, in part, due to the impact of the truncated population on correlations. Correlations tend to weaken as smaller proportions of a population are selected.

Although the number of parents (N_1) affected the value of adding first-generation information with strategy II, it did not show the same pattern with strategy I. With strategy I, the change in N_1 only caused small fluctuations, with less than 0.5% changes in both ranking accuracy and realized gain (Fig. 4). This is understandable, as adding first-generation information generally had little effect on the ranking with strategy I, and changing N_1 from 24 to 96 was not enough to make a significant difference.

In the simulation, we assumed that the second-generation population was a truncated population, whereby all the second-generation parents were selected from the first-generation population. Second-generation populations, however, may also include trees that were selected elsewhere (infusions) and not tested in the first generation. Thus, including first-generation data increase effective family size markedly only for those parents commonly tested across generations; others are unchanged. This eventually creates or exacerbates imbalance in the combined data sets. To test the effect of such imbalance of data on ranking changes, subsets of the first-generation data were created by randomly dropping half of the parents before the data were integrated with the second-generation data. Compared to the integration of the full data set from the first-generation, the percent increases in accuracy and realized gain were reduced by 16 and 10%, respectively, with strategy II, but there was almost no change ($< \pm 2\%$) with strategy I (details not shown). The reason was that BLUP with the integration of the first-generation data provided more conservative estimates than when only the second-generation data are used. Hill and Rosenberger (1985) and White and Hodge (1989) have indicated the conservative nature of BLUP values when there are few observations. The estimated breeding values of the parents tested only in the second generation were more adjusted toward the overall mean, which occasionally changed the ranking among parents.

Changing the number of crosses per parent (N_2) affected the efficiency of including the first-generation data as well (Fig. 5). For strategy II, percent increase in accuracy of ranking parents decreased from 9.2 to 8.5% as N_2 changed from single cross to double cross, and then became almost stable for $N_2=3-8$. The change in realized gain followed a similar pattern. Such reduction in added efficiency, as the number of crosses increase, probably arises because high

degrees of genetic relationships already existed in the second-generation population with more than two crosses per parent. This result was in line with that of Johnson (1998) who found that the added efficiency of making more crosses per parent dropped markedly after only two crosses. In comparison, increases in both ranking accuracy and realized gain remained almost unchanged with the change in N_2 when strategy I was used.

Conclusion

The added benefit of combined analysis depends on whether there is strong genetic relatedness between both generations and whether a high degree of genetic relationship is already present in the current (i.e., second) generation test. When second-generation parents were selected via backward selection, adding first-generation information could markedly increase the accuracy and reliability in ranking, as well as the realized gain from selection. The increases depended more on the heritabilities in both generations and less on the size of dominance and GXE effects. The number of parents and crosses per parent also were important factors. Including first-generation information helped more when there were many parents and few crosses per parent in the second generation.

When second-generation parents were selected via forward selection, the advantage of adding first-generation information was less. Tree breeders will need to determine whether such increases in ranking efficiency and gain are sufficient to justify the extra effort involved in mass computation.

Acknowledgment We would like to thank two anonymous reviewers for helpful reviews and comments on the manuscript. Terrance Ye and Keith Jayawickrama are funded by the Northwest Tree Improvement Cooperative. Our thanks to Alison Moldenke for editorial comments.

References

- Bulmer MG (1985) The mathematical theory of quantitative genetics. Clarendon, Oxford, UK
- Conover WJ (1980) Practical non-parametric statistics, 2nd edn. Wiley, New York
- Dutkowski GW, Costa e Silva J, Gilmour AR, Lopez GA (2002) Spatial analysis methods for forest genetic trials. Can J For Res 32:2201–2214
- Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics, 4th edn. Prentice Hall, Harlow, UK
- Gilmour AR, Gogel BJ, Cullis BR, Welham SJ, Thompson R (2002) ASRem1 user guide release 1.0. VSN International, Hemel Hempstead, HP1 IES, UK
- Henderson CR (1984) Application of linear models in animal breeding. University of Guelph, Guelph, ON, Canada
- Hill Jr RR, Rosenberger JL (1985) Methods for combining data from germplasm evaluation trials. Crop Sci 25:467–470

- Jayawickrama KJS, Carson ML (2000) A breeding strategy for the New Zealand Radiata Pine Breeding Cooperative. *Silvae Genet* 49:82-90
- Johnson GR (1998) Parental GCA testing: how many crosses per parent? *Can J For Res* 28:540-545
- Kerr RI, Dieters MJ, Tier B, Dungey HS (2004) Simulation of hybrid forest tree breeding strategies. *Can J For Res* 34:195-208
- King JN, Johnson GR (1993) Monte Carlo simulation models of breeding-population advancement. *Silvae Genet* 42:68-78
- Lasley JF (1987) *Genetics of livestock improvement*, 4th edn. Prentice Hall, Englewood Cliffs, NJ, USA
- Libby WJ (1973) Domestication strategies for forest trees. *Can J For Res* 3:265-276
- Lindgren D (1991) Progeny testing. In: *Genetics of Scots pine development in plant genetics and breeding*. Elsevier Science Publishers, Amsterdam, pp 191-203
- Lipow SR, Johnson GR, St Clair JB, Jayawickrama KJ (2003) The role of tree improvement programs for ex situ gene conservation of coastal Douglas-fir in the Pacific Northwest. *For Genet* 10:111-120
- McKeand SE, Bridgwater FW (1998) A strategy for the third breeding cycle of loblolly pine in the south-eastern US. *Silvae Genet* 47:223-234
- SAS Institute Inc. (2004) SAS[®] 9.1 Macro language: reference. SAS Institute, Cary, NC
- Skjervold H, Odegard AK (1959) Estimation of breeding value on the basis of the individual's own phenotype and ancestors' merits. *Acta Agric Scand* 9:341-354
- White TL (1987) A conceptual framework for tree improvement programs. *New For* 4:325-342
- White TL (1996) Genetic parameter estimates and breeding value predictions: issues and implications in tree improvement programs. In: Dieters MI et al (eds) *Proc. tree improvement for sustainable tropical forestry*. QFRI-IUFRO Conf. pp 110-117
- White TL, Hodge GR (1989) *Predicting breeding values with applications in forest tree improvement*. Kluwer, Dordrecht, The Netherlands
- White TL, Hodge GR, Powell GL (1993) An advanced-generation tree improvement plan for slash pine in the southeastern United States. *Silvae Genet* 42:359-371
- Young SSY (1961) The use of sire's and dam's records in animal selection. *Heredity* 16:91-102