

## Efficiency of Using First-generation Information during Second-generation Selection: Results of Computer Simulation

T. Z. Ye<sup>1</sup>, K. J. S. Jayawickrama<sup>1</sup>, and G. R. Johnson<sup>2</sup>

*BLUP* (Best linear unbiased prediction) method has been widely used in forest tree improvement programs. Since one of the properties of *BLUP* is that related individuals contribute to the predictions of each other, it seems logical that integrating data from all generations and from all populations would improve both the precision and accuracy in predicting genetic values by increasing the effective number of observations on each genotype (White and Hodge 1989; Kerr et al. 2004). However, some studies based on computer simulation (e.g. Johnson 1998) and field data (e.g. Panter and Allen 1995) showed that including historical parental information actually did little to increase the efficiency of estimating breeding values under some circumstances.

The objective of this study was to determine whether the inclusion of first-generation information sufficiently enhance the accuracy and precision of second-generation selection under different selection strategies and combinations of genetic parameters using stochastic data sets generated by computer simulation.

### Simulation Scenarios

- We assumed that the 1<sup>st</sup>-generation population consisted of 300 open-pollinated families with 100 trees per family. The individual-tree heritability  $h^2(1)$  was set in the range of 0.05 to 0.35.
- Two different selection strategies were adopted for the 1<sup>st</sup>-generation selection. (1) Strategy 1 (backward selection) – select top 48 parents; (2) Strategy 2 (forward selection) – select the best progeny from each of the top 48 families. All selections were based on estimated breeding values (*EBVs*).
- The 48 selections were then used as 2<sup>nd</sup>-generation parents and crossed in a disconnected  $2 \times 2$  factorial mating design which resulted in 48 crosses in total.
- The 2<sup>nd</sup>-generation progeny trials had six test sites with 20 trees per cross per site. Different levels of heritability ( $h^2(2) = 0.05, 0.15, 0.25, 0.35$ ), ratio of dominance to additive genetic variance ( $V_D/V_A = 0, 1, 2, 3$ ), and genotype-by-environment interaction ( $V_{G \times E}/V_A = 0, 1, 2, 3$ ) were assigned.

### Simulation Method

- King and Johnson's (1993) method was used to generate 500 independent stochastic data sets for each combination of genetic parameters and selection method (scenario).
- For each data set, we estimated breeding values for all the 2<sup>nd</sup>-generation parents and progeny using *BLUP* with and without integration of the 1<sup>st</sup>-generation information.

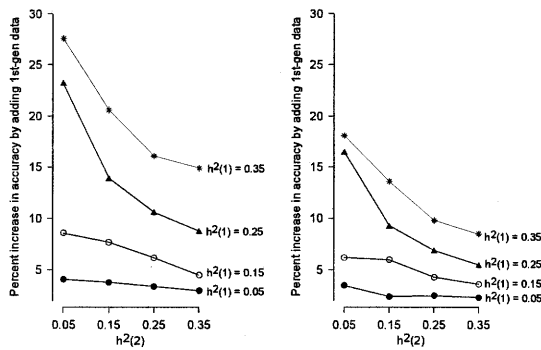
<sup>1</sup> Department of Forest Science, Oregon State University, 321 Richardson Hall, Corvallis, OR 97331, USA

<sup>2</sup> USDA Forest Service, Pacific Northwest Research Station, 3200 SW Jefferson Way, Corvallis, OR 97331, USA

- Kendall's rank correlation coefficients ( $\tau$ ) between the true and estimated breeding values for both parental and tree selections were calculated. The mean and coefficient of variation of the 500 correlation coefficients from each scenario were used to quantify the accuracy and precision of selection, respectively.
- The increase in accuracy was measured as the average percent improvement in the correlation when the 1<sup>st</sup>-generation data were included for 500 simulations. Increased precision was measured as the percent reduction in the coefficient of variation of the 500 simulations.

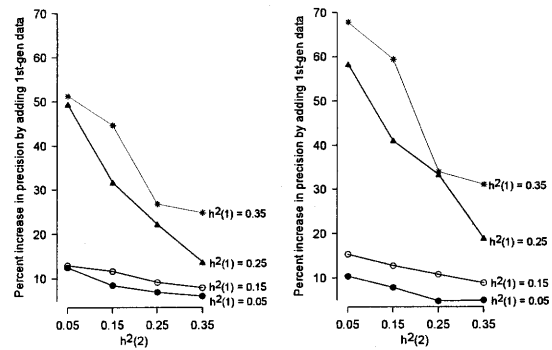
## Results and Discussion

When backward selection was used in the 1<sup>st</sup>-generation, including the 1<sup>st</sup>-generation information helped increase both the accuracy and the precision of 2<sup>nd</sup>-generation selection in all scenarios. However, the amount of increase varied and generally depended more on the heritabilities in both generations (Figures 1 and 2) and less on the size of dominance and G×E effects. The value of adding 1<sup>st</sup>-generation data was high when  $h^2(1)$  is high and  $h^2(2)$  is low and decreased as  $h^2(1)$  decreased and  $h^2(2)$  increased. This is expected since the 1<sup>st</sup>-generation data contain more genetic information than random noise when  $h^2(1)$  is high.



For ranking 2<sup>nd</sup>-gen parents  
For ranking 2<sup>nd</sup>-gen parents

Fig. 1. Effect of heritability on the *accuracy* of ranking when *backward* selection was used in the 1<sup>st</sup>-generation selection; assuming  $V_D = V_{G \times E} = V_A$



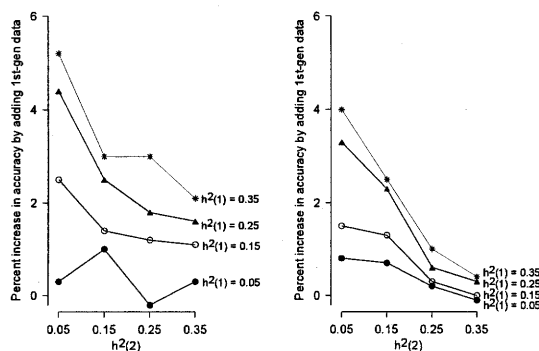
For ranking 2<sup>nd</sup>-gen progenies  
For ranking 2<sup>nd</sup>-gen progenies

Fig. 2. Effect of heritability on the *precision* of ranking when *backward* selection was used in the 1<sup>st</sup>-generation selection; assuming  $V_D = V_{G \times E} = V_A$

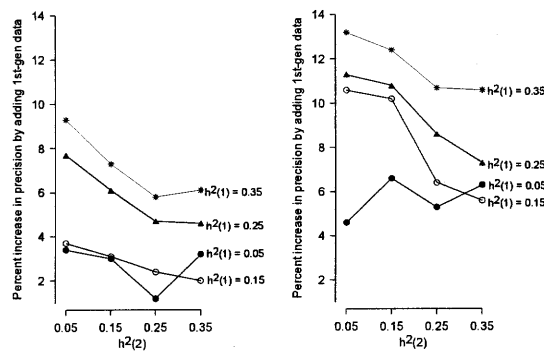
The effects of dominance and G×E on selection efficiency followed the expected pattern. The impact of adding the 1<sup>st</sup>-generation data increased as  $V_D$  or  $V_{G \times E}$  increased, but the percent increase was usually less than 5% when  $V_D = V_A$  or  $V_{G \times E} = V_A$ .

The 1<sup>st</sup>-generation information helped more in increasing the accuracy of re-selecting the 2<sup>nd</sup>-generation parents (backward) than that of selecting the 2<sup>nd</sup>-generation progenies (forward). This would be expected since the coefficient of relatedness between the 1<sup>st</sup>-generation progeny and the 2<sup>nd</sup>-generation parents is larger than the relationship between the 1<sup>st</sup>- and 2<sup>nd</sup>-generation progeny.

When forward selection was used in the 1<sup>st</sup>-generation, the 1<sup>st</sup>-generation information did little to increase the accuracy of ranking in the 2<sup>nd</sup>-generation (Figures 3 and 4). Because within-family selection is relatively imprecise, the rank correlation between true and estimated parental breeding values was quite low ( $\tau = 0.22$  when  $h^2(1) = h^2(2) = 0.25$ ;  $V_D = V_{G \times E} = 0.5V_A$ ) (Table 1). When  $h^2(1)$  was very low, the 1<sup>st</sup>-generation data simply added random noise for the 2<sup>nd</sup>-generation selections. The precision of ranking could increase up to 10~14% probably due to the increase the effective number of observations with high  $h^2(1)$  and low  $h^2(2)$ .



For ranking 2<sup>nd</sup>-gen parents  
For ranking 2<sup>nd</sup>-gen parents



For ranking 2<sup>nd</sup>-gen progenies  
For ranking 2<sup>nd</sup>-gen progenies

Fig. 3. Effect of heritability on the *accuracy* of ranking when *forward* selection was used in the 1<sup>st</sup>-generation selection; assuming  $V_D = V_{G \times E} = V_A$

Fig. 4. Effect of heritability on the *precision* of ranking when *forward* selection was used in the 1<sup>st</sup>-generation selection; assuming  $V_D = V_{G \times E} = V_A$

Table 1. Kendall's rank correlation coefficient ( $\tau$ ) between the true and estimated breeding values and the coefficient of variation (in parenthesis) for both parental and tree selections based on 500 simulations per scenario; assuming  $h^2(1) = h^2(2) = 0.25$ ;  $V_D = V_{G \times E} = 0.5V_A$ .

Selection in 1 <sup>st</sup> -generation	$\bar{\tau}$ & $CV(\tau)$	1 <sup>st</sup> -gen parents	1 <sup>st</sup> -gen progeny	2 <sup>nd</sup> -gen parents	2 <sup>nd</sup> -gen progeny
Backward	with 1 <sup>st</sup> -gen data	0.77 (2.2%)	0.44 (3.4%)	0.52 (14.6%)	
	with 2 <sup>nd</sup> -gen data			0.55 (12.7%)	0.48 (8.3%)
	with both			0.61 (10.4%)	0.52 (6.5%)
Forward	with 1 <sup>st</sup> -gen data	0.77 (2.2%)	0.43 (3.8%)	0.22 (44.0%)	
	with 2 <sup>nd</sup> -gen data			0.50 (14.8%)	0.45 (9.6%)
	with both			0.51 (14.2%)	0.45 (8.9%)

## References

1. White, T. L. and Hodge, G. R. 1989. Predicting Breeding Values with Applications in Forest Tree Improvement. Kluwer Academic Publishers, Dordrecht, The Netherlands.

2. Kerr, R. J., Dieters, M. J., Tier, B., and Dungey, H. S. 2004. Simulation of hybrid forest tree breeding strategies. *Can. J. For. Res.* 34: 195-208.
3. Johnson, G. R. 1998. Parental GCA testing: How many crosses per parent? *Can. J. For. Res.* 28: 540-545.
4. Panter, D. M. and Allen, F. L. 1995. Using Best Linear Unbiased Predictions to enhance breeding for yield in Soybean: II. Selection of superior crosses from a limited number of yield trials. *Crop Sci.* 35: 405-410.
5. King, J. N. and Johnson, G. R. 1993. Monte Carlo simulation models of breeding-population advancement. *Silvae Genet.* 42: 68-78.