EFFECTS OF SAMPLING ON OPEN-POLLINATED BIVARIATE PROGENY TESTS

L.A. Apiolaza[1], RD. Burdon[2] and D.J. Garrick[1]

[1]Department of Animal Science, Massey University, Palmerston North, New Zealand
[2]Forest Research Institute, Private Bag 3020, Rotorua, New Zealand

SUMMARY

Bivariate open-pollinated progeny tests were simulated for 80 combinations of genetic parameters (heritability for trait 1, heritability for trait 2 and genetic correlation between the two traits) and percentage of phenotypes recorded for one trait (10, 30, 50 and 100%). The tests had a completely random design, 200 families with 30 individuals each, no genetic relationship between the families, phenotypic variance of 1 for both traits and an environmental correlation of 0. The effect of sampling one trait on estimation of genetic parameters, prediction of breeding values and estimation of expected response to selection was investigated. The use of the lowest proportion of sampling generated higher magnitudes of bias, a poor representation of the distribution and larger standard errors for the estimates. The correlation between the breeding values predicted only with the sampled individuals and those predicted with all the information was substantially lower than 1. Additionally, the direct and correlated response for the sampled trait was underestimated as accuracy of selection was lower. Incrementing the sampled proportion presented diminishing positive consequences, with no significant effect over 15 individuals per family (50%).

Keywords: sampling, simulation, half-sibs, REML, tree-breeding.

INTRODUCTION

Tree breeding programs have historically focused on the improvement of growth and form traits (e.g. m^3/ha, stem straightness) rather than in increasing the profit of the wood-processing industry. This approach seems to ignore that most of the profits of the overall forestry sector come from the end-products and not from the production of logs, and the potential importance of wood properties improvement. However, in the last few years there has come an increasing interest in linking tree breeding programs with industrial activities, mainly in pulp and paper production (e.g. Borralho *et al.* 1993; Greaves et *al.* 1997).

One limiting factor in the effective incorporation of wood properties into tree breeding programmes is the lack of adequate genetic parameters. Estimates of these parameters are obtained from analysis of progeny-test data. Additionally, these tests are used for predicting the breeding values (BV) of individuals and estimation of expected response to selection (White 1987). While growth traits (usually of low heritability) are generally cheap and easy to measure, precise assessment of wood properties (usually highly heritable) typically requires the use of wood samples, which involves high costs. Because of cost constraints it is necessary to assess relatively small samples of individuals for the latter traits. The objective of this study is to explore, through simulation, the effects of different sampling intensities on accuracy of estimation of genetic parameters, consistency of rankings by BV based on best linear unbiased prediction (BLUP) and estimates of expected response to selection.

MATERIALS AND METHODS

The simulation experiment consisted of the full factorial combination of heritability of trait 1 ($h_1^2 = 0.1$ and 0.3), heritability of trait 2 ($h_2^2 = 0.4$ and 0.8), genetic correlation between the two traits ($r_g = -0.6, -0.3, 0, 0.3$ and 0.6) and sampling intensity of trait 2 (10, 30, 50 and 100% of the observations). Trait 1 was always considered with 100% sampling. One hundred bivariate progeny tests were simulated for each combination of levels of the factors. The tests were assumed to have a completely random design, 200 open-pollinated families with 30 individuals per family, and 100% survival. Also assumed were: no genetic relationship between the families, phenotypic variance of 1 for both traits and an environmental correlation of 0.

The bivariate observations with desired variance-covariance matrices were obtained using Cholesky decomposition (Van Vleck 1994). The sampling schemes were accomplished by randomly deleting observations of trait 2 from the complete simulated test, and reanalysing the test every time. Variance and covariance components were estimated for each simulated test using restricted maximum likelihood (REML, Patterson and Thompson 1971). An iterative average information algorithm was applied to maximise the likelihood function using AIREML (Johnson and Thompson 1995). For each simulated combination of parameters and sampling the statistical significance of skewness and biases was tested. This study considered the prediction of breeding values for backward (or parental) selection. The breeding values were obtained as the solutions of multivariate mixed model equations. The effects of sampling on the prediction of breeding values was quantified using the correlation ($r_{BVis,BVi}$) between breeding values for trait *i* predicted with a sample of the data (BV_is) and those predicted using all the data (BV_i). The expected direct (ΔG_i) and correlated ($\Delta_c G_j$) response to selection was estimated using standard formulae.

RESULTS AND DISCUSSION

Sample distribution. The REML estimates for heritabilities and genetic correlations presented skewed distributions, especially for low sampling intensities and extreme heritabilities. The REML property of not allowing results out of the parameter space tended to create a concentration of estimates close to the lower or upper bounds. Significant changes of sign for skewness while increasing sampling intensities were commonplace, indicating that the lowest sampling intensities did not give a reliable representation of the shape of the distribution of the estimates. The means of the genetic parameter estimates varied slightly according to the different sampling schemes. The magnitude of the bias of the estimates was higher for the lowest sampling intensities. The largest deviations were for r_g, followed by h^2

of traits 2 and 1. Increasing the sampled percentage from 10 to 30% reduced the bias, but further increments had no substantial effect on its magnitude. The observed standard deviations of the estimates for each combination of parameters were considered as the true standard errors. As expected, an increment in the proportion of sampled trees produced a reduction in the standard error for all parameters estimates. However, using more than 50% sampling gave only a small reduction in estimated standard errors. The averages of the standard errors generated by the REML analysis were overestimating the true standard errors in near 50% of the cases.

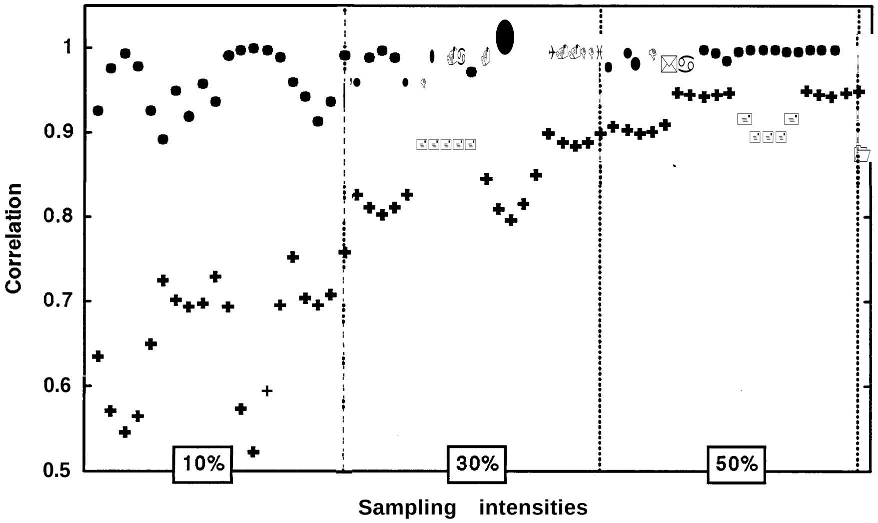

Figure 1. Correlation between breeding values predicted using 10, 30, and 50% of the observations and those predicted with 100%, considering all the combinations of genetic parameters, for $r_{BV1s,BV1}$ (•) and $r_{BV2s,BV2}$ (+). Each sampling intensity considers 20 combinations of genetic parameters, the first 10 for h_1^2 = 0.1 and the following 10 for h_1^2 = 0.3. From each group of 10 the first 5 are for h_2^2 = 0.4 and the others for h_2^2 = 0.8; and in each group of 5 there is one observation for r_g = -0.6, -0.3, 0, 0.3 and 0.6.

Breeding values. A central aspect in the breeding process is the selection of the parents for the next generation. The effect of Sampling on the correlation $r_{BVis.BVi}$ had a direct relationship with sampling intensity and heritability of the trait (Fig. 1), es&ally for trait 2. Even when trait 1 was not subject to sampling, the correlations $r_{BV1s,BV1}$ were affected because the multivariate analysis increased the accuracy of prediction when including information from trait 2 (Thompson and Meyer 1986). The effect of including trait 2 depended on the value of h_2^2 and r,; thus the highest $r_{BV1s,BV1}$ were when trait 2 had low heritability and poor association with trait 1. The effect of r_g was symmetric; i.e. an increment of r_g in either way produce the same change on the correlation $r_{BVis,BVi}$. Even for a sampling intensity of 10%, the magnitude of $r_{BV1s,BV1}$ was high (over 0.9), indicating a close agreement of the selected parents using the different sampling intensities.

Response to selection. Sampling affected the estimation of direct response for trait 2 and the correlated response for both traits. Direct response was very dependant on the heritability of the trait and, to a lesser extent, on the genetic correlation between the traits, which weighted the contribution of other traits in the estimation of response. The effect of genetic correlation was, as in the prediction of breeding values, symmetric. In trait 2, low sampling intensities consistently reduced the predicted response. The effects on ΔG_1 when sampling trait 2 were not important. The magnitude of the correlated response of trait 1 when selecting for trait 2 ($\Delta_c G_1$) was directly dependent on the heritability of trait 1; with higher gains for $h_1^2 = 0.3$. Low sampling intensities overestimated the expected gain. The correlated response of trait 2 when selecting for 1 ($\Delta_c G_2$) follows the same pattern of $\Delta_c G_1$; but depending on the heritability of trait 2, with higher gains for $h_2^2 = 0.8$. The magnitude of $\Delta_c G_2$ was always superior to $\Delta_c G_1$ when $r_g \neq 0$.

Even though the combinations were not exhaustive, they covered a range of situations in tree breeding. The use of the lowest proportion of sampling showed to be generally inadequate for producing reliable estimates of genetic parameters, prediction of breeding values and estimations of expected response to selection, all necessary tools to make the decisions involved in a breeding program. Incrementing the sampled proportion presented diminishing positive consequences, with no significant effect over 50%. If the interest is only the estimation of genetic parameters, it is possible to use elliptical selection (Cameron and Thompson 1986). This method would concentrate the sampling on the extremes of the distribution, but would not generate enough information for predicting the breeding values for all the parents. Furthermore, the chosen sampling intensity will be heavily influenced by the economic importance of the traits under study and the population size.

REFERENCES

Borralho, N.M.G., Cotterill, P.P. and Kanowski, P. J. (1993) *Can. J. For. Res.* **23**:648-656.
Cameron, N.D. and Thompson, R. (1986) *Theor. Appl. Genet.* **72**:466-476.
Greaves, B.L., Borralho, N.M.G. and Raymond, C.A. (1997) For. *Sci.* **43**:*(in press).*
Johnson, D.L. and Thompson, R. (1995) *J. Dairy Sci.* **78**:449-456.
Patterson, H.D. and Thompson, R. (197 1) *Biometrika* **58**:545-554.
Thompson, R. and Meyer, K. (1986) *Livest. Prod. Sci.* **15**:299-3 13.
Van Vleck, L.D. (1994) *Braz. J. Genet.* **17**:53-57.
White, T.L. (1987) New. *For.* **1**:325-342.