Implementation of a Multivariate BLUP Model for Genetic Evaluation of *Eucalyptus globulus* in Australia

S.F. Jarvis[1], N.M.G. Borralho[2] and B.M. Potts[2]

[1] Southern Tree Breeding Association, PO Box 1811, Mt Gambier, South Australia, 5290, Australia.
[2] CRC Temperate Hardwood Forestry, Plant Science Dept., University of Tasmania, GPO Box 252C, Hobart, Tasmania, 7001, Australia.

Abstract

A BLUP genetic evaluation for diameter of *Eucalyptus globulus* Labill. ssp. *globulus* base population across four states in Australia is described. The data set consisted of over 52,000 open pollinated progeny from 589 unrelated parents, selected across the range of natural distribution of the subspecies. The model included effects of block within each site, stand type and plot. A multiple trait procedure was used, assuming diameter in each state as a different trait. Estimated breeding values for diameter in each state were calculated and used to compare different breeding and deployment strategies. A cooperative breeding strategy, running a single national breeding population, with separate deployment populations for each state is suggested. BLUP selection leads to in an increase in co-selection of sibs from the best families, but a restriction of 10 selections (parent or progeny) per family gave an acceptable rate of inbreeding (less than 1% per generation) and coefficient of variation of response (CV=9%) while resulting in substantial gains in volume.

Introduction

Best Linear Unbiased Prediction (BLUP) of breeding values, as described by Henderson (1975), has become the method of choice for genetic evaluation of livestock throughout the world, for species as diverse as dairy and beef cattle, swine, horse, sheep and fish. The use of an individual model BLUP has some distinct advantages (Woolliams, 1990; Borralho, 1995) including the unbiased prediction of breeding values, the simultaneous estimation of genetic and environmental fixed effects, thus correctly accounting for genetic differences between levels of a fixed effect, the accounting of all available genetic information however distant genetically or geographically and the provision of breeding values for comparison across age groups, so that selection of the best individuals can be made, regardless of age or generation.

The use of an individual model is relatively recent in tree breeding, although BLP using a parent model has been used for some time (White and Hodge, 1989). The reluctance to adopt the individual model seems to stem from concern about its computational difficulty. However, developments such as the discovery of a rapid method for generating the inverse of the numerator relationship matrix (Henderson, 1976), improvement in computing power and algorithms make individual model BLUP an easy and flexible tool for prediction of breeding values even for large numbers of individuals. This paper reports the results of a national BLUP genetic evaluation of *Eucalyptus globulus* ssp *globulus* in Australia for diameter at 3-5 years. The implications for breeding and deployment under alternative strategies are discussed.

Material and Methods

The data used in this analysis were based on measurements of 52,172 open-pollinated progeny from 589 unrelated parents. These parents were selected across the range of natural distribution of *Eucalyptus globulus* Labill. ssp. *globulus* as part of two range wide collections made by the Australian Tree Seed Centre in 1987 and 1988. The data set included a total of fourteen trials; four in Victoria, established by APM; one in South Australia, established by Primary Industries of SA (Forestry); four in Western Australia, established by Bunnings Treefarms; five in Tasmania, established by North Forest Products. Site conditions were variable but characterised by low altitude and moderate fertility. Rainfall varied from high in Tasmania to low in Western Australia. All sites were ploughed prior to planting and fertiliser was applied at planting. Experimental layout was an incomplete block design with between 20 and 25 families per block in Tasmania and Western Australia, randomised complete blocks in South Australia and a randomised complete block within sublines in Victoria. Plot sizes ranged from single tree plots to five tree row plots, with each family represented by up to 40 trees at each site.

The only trait included in the analysis was diameter over bark at breast height (1.3 m) when the trees were between three and five years. Trees with any apparent disease, stem or crown damage or with more than one stem were removed from the analysis. Seedlots were classified into four stand types as described by Borralho and Potts (1995). Heterogeneous variance across sites within each state (due to different ages or growth rates)

Table 1. Summary of the data set in the national analysis, divided by states

	VIC	SA	WA	TAS
No Records	12543	3113	14182	22334
No Parents	220	168	370	563
Site.Rep.Block	4	4	452	625
Diam[1] mean	11.02	13.14	12.37	9.55
std. dev.	2.83	2.58	3.38	3.08

[1] After transformation

was corrected by multiplying diameter by the ratio between the pooled phenotypic standard deviation and the trial phenotypic standard deviation, as described by Visscher *et al.*, (1991). Transformed diameter at each state was then treated as a separate trait. The data are summarised in Table 1.

Statistical model

The model used was (in matrix notation):

$$y = Xb + Z_u u + Z_p p + e$$

where y is the vector of individual tree observations, b is the vector of unknown parameters for all fixed effects, including block within replicate within trial and stand type, u is the vector of breeding values of trees, p is the vector of plot effects and e is the vector of random residuals, X is the incidence matrix connecting the observations to the fixed effects, Z_u and Z_p are the incidence matrices connecting the observations to the random effects (tree and plot, respectively). Means and variances of these effects are:

$$
\begin{bmatrix} y \\ e \\ p \\ u \end{bmatrix} \approx N \left(\begin{bmatrix} Xb \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} V & R & Z_P G_P & Z_u G_A \\ R & R & 0 & 0 \\ G_P Z_P' & 0 & G_P & 0 \\ G_A Z_u' & 0 & 0 & G_A \end{bmatrix} \right)
$$

where $V = R + Z_p G_p Z_p' + Z_u G_A Z_u'$, and R is the residual (co)variance matrix, $G_p = I \otimes G_{po}$ where I is the identity matrix, G_{po} is the plot (co)variance matrix, $G_A = A \otimes G_o$ where G_o is the additive genetic (co)variance matrix and A is the numerator relationship matrix and $\otimes$ is the direct or Kronecker product. The variances and covariances in R, G_{po} and G_o were estimated from the same data set using a derivative-free restricted maximum likelihood method (Meyer, 1991)

following an individual tree model (Borralho *et al.*, 1995). The solutions for fixed and random effects were calculated using the PEST program (Groeneveld *et al.*, 1990).

Results and discussion

The multivariate analysis of 52,172 trees and 589 parents was run on a 486 PC. It used 43% of 16MB of RAM and took 10 minutes for data preparation and 3 hours, 11 minutes to set up and solve the Mixed Model Equations (MME).

A BLUP of breeding value for diameter in each state (denoted EBV) was obtained for every tree and the best 30, 300 and 600 trees were selected on the basis of the average EBV across the four states. Table 2 shows the number of parents and progeny selected in the top 30, 300 and 600 trees.

An important feature of BLUP is that it allows direct comparison between parent and progeny EBVs. For deployment populations (top 30 trees), half of the selected trees were parents in native stands and half were progeny established in field trials. In the breeding populations there was a much higher proportion of progeny than parents (88% and 93% were progeny out of the top 300 and 600 trees, respectively).

It is apparent from Table 2 that the selected progeny came from only a few families. The use of information from relatives in BLUP analysis results in an increase in the co-selection of sibs from the best families. The effect is more marked with traits of low heritability, such as diameter, because more weight is put on the information from relatives. Even with a breeding population of 600, the 556 progeny selected came from only 63 families

Table 2. Number of parents and progeny selected in the top 30, 300 and 600 trees

No. selected	No. of parents[1]	No. of progeny[2]	No. of families[3]
30	15 (11)	15 (5)	16
300	37 (12)	263 (34)	46
600	44 (3)	556 (60)	63

[1] Number of parents with no progeny selected shown in parenthesis
[2] Number of different half-sib families among progeny shown in parenthesis
[3] Number of families meaning either parents with no progeny, parents and half-sib progeny or half-sib families without parents
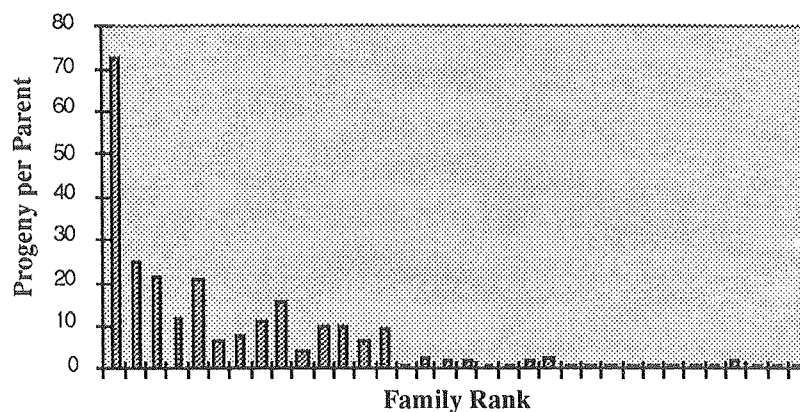
Figure 1. Number of progeny per family in the top 300 trees.

(Table 2). Figure 1 clearly shows that unrestricted use of BLUP will result in a few families dominating the next generation of breeding. For example, the best family contributed 74 progeny in the top 300 trees.

The co-selection of sibs under BLUP has important implications for the breeding strategy in terms of effective population size, N_e, and therefore rate of inbreeding and variance of response. A common strategy is to place some constraint on the number of representatives per family (Cotterill and Dean, 1990). The impact of these constraints on genetic gain, effective population size, rate of inbreeding and variance of response are shown in Table 3.

Table 3 shows that placing no restrictions on family size maximised gains but lead to a dramatic reduction in N_e. As a consequence, rates of inbreeding were around 7% per generation (or about 1% per year, assuming a generation length of 7 years) for both deployment and breeding populations. This is considered acceptable for avoiding the worst impact of inbreeding depression (Nicholas, 1987), but the effect of such a small effective population on variance of response is sufficient justification to prefer the use of a strategy where there is a limit on co-selection of relatives. Restricting family size to less than 10 representatives resulted in only 5% reduction in gain but a substantial increase in N_e (from 7 to 67), a reduction in the rate of inbreeding from 6.8% to 0.8% per generation, and a decrease in the CV of response from 25% to 9%. Restricting the co-selection of relatives to 5 or even to 2 trees resulted in large increases in N_e (to 194 and 546 respectively), but this is not likely to compensate for the reduction in gains (of 10% and 24% respectively). Restricting co-selection of relatives to a maximum of 10 trees per family seems to be appropriate for the breeding population, resulting in a

Table 3. Effect of placing limitations on family size on effective population size (N_e), change in rate of inbreeding (ΔF), gain in diameter, variance of response and coefficient of variation of response (shown in parenthesis).

Strategy	No. selected	No. of families	No. of parents	N_e [1]	ΔF (%) [2]	Gain	Variance of response [3]
No limit	30	16	15	6.68	7.49	2.40	0.227 (19.8)
	300	46	37	7.35	6.80	1.84	0.206 (24.7)
	600	63	44	9.46	5.29	1.69	0.160 (23.7)
10 per family	30	17	15	9.35	5.35	2.39	0.162 (16.8)
	300	66	45	67.03	0.75	1.74	0.023 (8.6)
	600	114	67	130.10	0.38	1.54	0.012 (7.0)
5 per family	30	18	17	17.45	2.87	2.36	0.087 (12.5)
	300	90	58	193.60	0.26	1.65	0.008 (5.4)
	600	181	88	405.99	0.12	1.42	0.004 (4.3)
2 per family, no parents	30	18	0	54.00	0.93	2.03	0.028 (8.2)
	300	174	0	545.50	0.09	1.40	0.003 (3.8)
	600	326	0	1124.60	0.04	1.16	0.001 (3.2)

[1] N_e is effective population size, calculated using $N_e = 4N_c/(V_k+2)$ where N_c is the number of cohorts (N minus no. of parents) and V_k is the variance of family size (Falconer, 1989).
[2] ΔF is the change in inbreeding per generation, calculated using $F=1/2N_e$
[3] Variance of response after one generation is calculated as V_a/N_e (Falconer, 1989), where $V_a = 1.513$ (Borralho *et al.*, 1995)

Table 4. Gains in diameter at three to five years in each state and overall for selecting the best 300 trees, for four alternative breeding strategies and the best 30 trees for alternative deployment strategies. Overall gains when parents are excluded and percentage gains in volume (in parenthesis) are also given.

Strategy	VIC	SA	WA	TAS	Average[1,2] (parents included)	Average[1,2] (parents excluded)
Breeding						
Coop. National	2.105	1.750	1.605	1.904	1.841 (16.8)	1.780 (16.3)
Coop. State	2.125	1.828	1.700	1.934	1.897 (17.4)	1.836 (16.8)
State 1	1.949	1.559	1.593	1.838	1.735 (15.9)	1.634 (14.9)
State 2	1.840	1.377	1.607	1.749	1.643 (15.0)	1.592 (14.6)
Deployment						
Coop. National	2.776	2.258	2.083	2.464	2.395 (21.9)	2.164 (19.8)
Coop. State	2.794	2.327	2.195	2.493	2.452 (22.4)	2.246 (20.5)
State 1	2.762	2.272	2.184	2.465	2.421 (22.1)	2.155 (19.7)
State 2	2.563	1.958	2.215	2.280	2.254 (20.6)	2.110 (19.3)

[1] Calculated as the average of the gains in diameter over the four states.
[2] Percentage gain in volume calculated using assumed genetic variance for volume of 2025 cm^2 and a genetic correlation between diameter and volume of 0.75

low rate of inbreeding (about 0.15% per year) and a low coefficient of variation of response (CV=9%), thus ensuring that predicted gains are likely to be achieved.

National versus regional breeding strategies

A national tree improvement program covering a wide range of environmental conditions and silvicultural practices is likely to have to accommodate some degree of genotype-environment interaction (GxE). Although GxE seems moderate in *E. globulus* (Kube *et al.*, 1995), genetic correlations between states were around 0.7 (Borralho *et al.*, 1995) and growth in each state was treated as a different trait in the analysis. Thus, every tree has an EBV for each state. These EBVs can be used to explore the outcomes from alternative national and regional breeding strategies. The four alternative breeding strategies are:

Cooperative breeding - National program
- where both genetic material and information are shared and trees are selected on the basis of the average of their EBVs across four states. There is only one breeding or deployment population;

Cooperative breeding - State program
- where both genetic material and information are shared, trees are selected on the basis of the EBVs within each state, thus resulting in four different (but jointly managed) breeding or deployment populations, one for each state;

State Breeding 1
- where information is shared but not genetic material. Trees are selected on the basis of their EBVs within each state but the four breeding or deployment populations will only include trees from each state and programs will be managed independently;

State Breeding 2
- where there is no sharing of either information or genetic material. Trees are selected on the basis of EBVs from a univariate analysis for each state. The four breeding or deployment populations will only include trees from each state and programs are managed independently.

The expected gains for the four alternative breeding programs based on a breeding population of 300 trees and a deployment population of 30 trees are shown in Table 4.

The predicted gains from selecting the best 30 or 300 trees including and excluding parents demonstrate the benefits of across generation selection using BLUP. The gains from selection of the best trees, regardless of whether they are parents or progeny, were always higher than the gains when the selection was from progeny only. Although differences were small, parents trees are more likely to be flowering at the time of selection, and so can be used in a crossing program sooner. This will have important implications in reducing the generation interval and deployment lag time.

Gains in diameter at 3-5 years were consistently higher for the two cooperative breeding strategies. These gains in diameter correspond to a gain in volume at rotation end of between 15% and 17%. A cooperative state breeding strategy, in which both information and genetic material are shared, resulted in the greatest gains for each state. However this strategy required four different breeding populations. The cooperative national strategy, with a single breeding population, had gains nearly as large as a cooperative state strategy but required only one quarter of the resources. The extra cost of breeding *E. globulus* separately in each state would negate the value of the very small amount of extra gain that could

be made. Gains from State 1 and State 2 strategies were generally inferior to the cooperative national strategy. The superiority of the cooperative strategies over State 1 (where information is shared but genetic material is not) was due to an increase in selection intensity, as more trees became available for selection. The reduction in gain for State 1 was greatest for South Australia as the proportion selected increased from 300/52,761 to 300/3702. Gains from State 2 strategy (where there is no exchange of information or material) were generally lower than gains from State 1 as a consequence of the lower selection accuracy of the former (due to fewer progeny per family).

The same general conclusions can be drawn about the alternative deployment strategies. However, selecting four different deployment groups of 30 trees, one for each state, using the information from all four states, is not going to cause the large escalation of costs that running four separate breeding programs will have. Results suggest that a common breeding population should be used, but that specific deployment populations (e.g. seed orchards) should be defined for each state. This seems particularly important for Western Australia, where gains from deployment were greatest under State 2 strategy (Table 4).

Conclusions

BLUP analysis is readily applicable to national genetic evaluation schemes in tree breeding. The use of BLUP for the national analysis of *Eucalyptus globulus* in Australia has enabled the prediction of breeding values for over 52,000 progeny and parent trees. The ability to compare breeding values of parents and progeny resulted in greater genetic gains in diameter for breeding and deployment than would selection within a discrete generation. Results also suggest that a cooperative national strategy for breeding will lead to the greatest gain for the least cost and that a cooperative state strategy for deployment is optimal for most states.

Acknowledgments

References

Borralho, N.M.G., 1995. The impact of individual tree mixed model methods (BLUP) in tree breeding strategies. These proceedings.

Borralho, N.M.G. and Potts, B.M., 1995. Accounting for native stand characteristics in genetic evaluations of open-pollinated progeny from *Eucalyptus globulus* base population. New Forest, submitted.

Borralho, N.M.G, Jarvis, S.F. and Potts, B.M., 1995. Multivariate Restricted Maximum Likelihood estimation of genetic parameters for diameter at 3 to 5 years in *Eucalyptus globulus* base population in Australia. These proceedings.

Cotterill, P.P. and Dean, C.A., 1990. Successful Tree Breeding with Index Selection. CSIRO Publication.

Groeneveld, E.M., Kovac, M. and Wang, T., 1990. PEST, a general purpose BLUP package for multivariate prediction and estimation. 4th World Congr. Genet. Appl. Livest. Prod., XIII: 488-491.

Henderson, C.R., 1975. Best linear unbiased estimation and prediction under a selection model. Biometrics, 31: 423-447.

Henderson, C.R., 1976. A simple method for computing the inverse of a numerator relationship matrix used in the prediction of breeding values. Biometrics, 32: 69-83.

Kube, P.D., Matheson, A.C., Raymond, C.A. and Boomsma, D.B., 1995. Provenance variation in *Eucalyptus globulus*. These proceedings.

Meyer, K., 1991. DFREML version 2.0. Programs to Estimate Variance Components by Restricted Maximum Likelihood Using a Derivative-Free Algorithm. User Notes. Animal Genetics and Breeding Unit, University of New England, Armidale, NSW (Mimeo)

Nicholas, F.W., 1987. Veterinary Genetics. Oxford University Press.

Visscher, P.M., Thompson, R. and Hill, W.G., 1991. Estimation of genetic and environmental variances for fat yield in individual herds and an investigation into heterogeneity of variance between herds. Livest. Prod. Sci., 28: 273-290.

White, T.L. and Hodge, G.R., 1989. Predicting Breeding Values with Applications in Forest Tree Improvement. Forest Sciences vol. 33. Kluwer Academic Publishers, The Netherlands, 367 pp.

Woolliams, J.A., 1990. Strategies to maximise selection progress in dairy cattle. 4th World Congr. Genet. Appl. Livest. Prod. XIV: 15-24.