MULTIVARIATE APPROACH TO THE ESTIMATION OF GENETIC PARAMETERS

Mark J. Dieters', Susan F. Jarvis[2], and Arthur R. Gilmour3

Abstract:-- Data collected from genetic tests of forest trees is typically multivariate in nature: multiple traits are measured and assessed, and the same traits are measured at different ages (on the same trees) and in different environments. Nevertheless, this type of test data has usually been analysed using univariate statistical techniques to estimate variance components, from which genetic parameters such as heritability are then derived. In this paper, results are presented from the analysis of progeny test data from *Araucaria cunninghamii* (hoop pine) using both univariate and multivariate implementations of restricted maximum likelihood (REML) estimation using the program ASREML.

Heritability estimates from univariate and multivariate analyses, when averaged across 11 different tests, were very similar in magnitude and precision (as assessed by the size of the standard errors). However, when half of the data was deleted for one trait (tree height at 12 years of age) and the data reanalysed, univariate methods proved to yield less accurate estimates of heritability, with larger standard errors. Therefore, it appears that multivariate estimation will provide more reliable heritability estimates for traits that are a) measured only on a subset of the trees in a test, and b) correlated to other traits assessed on all individuals in the test.

Keywords: Genetic parameters, heritability, multivariate, univariate, REML

INTRODUCTION

Genetic testing is a basic component of all improvement programs with forest tree species. These genetic tests are established to serve a variety of purposes including: estimation of genetic parameters, prediction of breeding values, and selection of parents for the next generation of breeding. Information is collected in these tests for a number of traits, at multiple ages, and from the same trait measured in different environments, for example: growth and rust incidence in *Pinus elliottii* (White et al. 1993); growth, form, wood density and disease resistance in *P. radiata* (Boomsma 1997); and, growth, form and reduced crown defects in *P. elliottii* (Toon et al. 1996). Multivariate data of this type has been collected for over 40 years from progeny tests of *Araucaria cunninghamii,* a native conifer of Queensland (Dieters et al. 1990).

Although forestry data is typically multivariate in nature, univariate approaches have commonly been used to estimate genetic parameters from this type of progeny data, primarily because suitable statistical software has not been readily available. Multivariate techniques based on analysis of variance methods for the estimation of variance components is problematic where data is unbalanced or missing: software such as Proc GLM in SAS (SAS 1989, p. 910) eliminate observations from the analysis where any data is missing, and generally analysis of variance based estimation is considered to be inferior to restricted maximum likelihood (REML) techniques where data is unbalanced (Khuri and Sahai 1985, Huber 1993). In genetic tests of forest tree species mortality commonly results in missing data as tests increase

Queensland Forestry Research Institute and Cooperative Research Centre for Sustainable Production Forestry, MS 483 Fraser Rd. Gympie Qld. 4570, Australia
[2] Pastoral & Veterinary Institute, Hamilton VIC 3300, Australia
[3] New South Wales Agriculture, Orange NSW 2800, Australia

in age, and tests are often unbalanced. Further sometimes traits that are expensive to measure are only assessed on a subset of the trees in a test, or tests are thinned so later-age data is only available from the surviving trees. In cases such as these where large amounts of data are missing, the use of information from correlated traits should increase the precision of genetic parameter estimates, particularly if the incomplete data is from a selected subset of the data and the trait used to make the selection is also in the analysis.

Therefore, in many situations the use of multivariate (REML) estimation procedures can be expected to improve the precision of variance component estimates (and hence that of derived parameters such as heritability). Further, simultaneous estimation of genetic and residual correlations directly addresses problems of non-conformity of variance-covariance matrices and the correlated error structures between traits measured on the same trees. In this paper we demonstrate the application of multivariate REML procedures using ASREML (Gilmour et al. 1999) to data from tests of half-sib and full-sib progeny of *A. cunninghamii* (hoop pine) in Queensland for the estimation of narrow sense heritability. The impacts of missing data on the reliability of heritability estimates is also examined.

MATERIALS AND METHODS

Univariate and multivariate analyses were performed on data from eleven progeny tests of hoop pine growing on sites in south-east Queensland (Table 1) using either a parental model (for open pollinated tests) or an individual model (for mixed open and control pollinated tests). For the multivariate analyses, unstructured variance-covariance matrices were fitted for the family (or individual), replicate and residual effects. The traits analysed were: tree height at 1, 4, 8 and 12 years; diameter at breast height (1.3 m above ground level) at 8 and 12 years; stem straightness (6-point scale, 1=crooked, 6=straight) at 8 years; and, internode length (average length of top two internodes) at 8 years. However, not all traits were measured in all tests.

Table 1. Details of hoop pine progeny tests used in the analyses when the tests were measured for diameter at breast height (1.3m) at 8 years of age.

Expt. Series	Location Code[†]	Number of Individuals	Number of Reps.[‡]	Number of Families[§]	Family Type[††]	Mean (cm)	Min. (cm)	Max. (cm)
545	im	3433	48	76	OP	11.4	0.8	20.0
	bk	4825	64	77	OP	14.3	1.3	23.5
568	im	1510	64	25	OP	12.5	2.2	20.6
	bk	1293	64	29	OP	11.2	3.0	18.0
607	im1	1427	56	26	OP	13.9	3.3	21.2
	yr	1609	56	31	OP	13.7	3.4	24.5
	im2	1131	48	24	OP	14.9	4.0	25.3
641	im	1940	64	25	OP & CP	12.4	1.3	20.2
	ev	1923	64	25	OP & CP	11.3	0.6	18.3
653	im	1343	36	34	OP & CP	14.1	2.8	21.6
	yr	1328	36	34	OP & CP	12.5	5.0	18.9

[†] Tests replicated across locations in south-east Queensland
[‡] Number of complete blocks, families represented by one tree in each block
[§] Number of families for OP tests, and number of parents of mixed OP & CP tests
[††] OP = Open Pollinated, CP = Control Pollinated Families

Narrow sense heritabilities were estimated as:

$$h^2 = \frac{4\sigma^2_{Fam.}}{\sigma^2_{Rep.} + \sigma^2_{Fam.} + \sigma^2_{Resid.}} \quad \text{, or}$$

$$h^2 = \frac{\sigma^2_A}{\sigma^2_{Rep.} + \sigma^2_A + \sigma^2_{Resid.}}$$

for variance components estimated from parental and individual model respectively. Note that: a) open pollinated families were assumed to be half-sib families, and the variance among half-sib families was assumed to be one quarter of the additive genetic variance (σ^2_A); b) residual variances estimated from parental and individual models are not equivalent; and c) a completely random model was fitted to the data.

When companng the results from univariate and multivariate analyses, because the actual underlying variance components are unknown, it is impossible to determine whether or not estimates differ in bias or precision unless simulated data is used. Further, it is not possible to compare parameter estimates based on the size of the standard errors because the standard error is partially related to the magnitude of the parameter being estimated. Therefore, for the purposes of this paper, heritability estimates derived from the multivariate analyses are assumed to be the most reliable and univariate estimates are presented as deviations from these estimates.

In order to examine the effect of missing data that might result from the assessment of a trait on only a subset of a test (such might result if a trait is expensive to measure), data for 12 year height was deleted from half of the replicates in Expt. Series 607. Univariate and multivariate analyses were repeated, and the heritability estimates for this trait in the three tests is compared to the estimates from the multivariate analyses using data from all replicates.

RESULTS AND DISCUSSION

Heritability estimates obtained from the multivariate analyses for height and diameters are slightly higher than those previously reported for univariate analyses of hoop pine data (Dean et al. 1988, Dieters et al. 1990), while those for straightness and diameter are similar to previously published estimates (Table 2). Although the maximum absolute deviation between heritability estimates from multivariate and univariate analyses was 0.19 (data not reported), overall there were relatively minor differences in both the heritability estimates and their standard errors that were obtained from the two analysis methods (Table 2). Therefore, it appears that either method, on average, will yield similar estimates of heritability where there is relatively little missing data.

When half the data for 12 year height was deleted and the data reanalysed, substantially different heritability estimates were obtained (Figure 1). Heritability estimates obtained using multivariate analyses with only half the height data were very similar to those using all the data; however, estimates from univariate analyses differed by as much as 0.13 from multivariate analyses using all the data (Figure 1). Similarly, the standard errors of the heritability estimates were always larger for the univariate analyses (with half the height data) than multivariate analyses (Figure 2), and this trend did not reflect the relative sizes of the heritability estimates (cf. Figures 1 and 2). Therefore, it appears that multivariate analyses can improve the accuracy and precision of heritability estimates obtained from traits that are a) only assessed on a subset of the individuals in a genetic test, and b) strongly correlated to other traits that have been assessed on all individuals in the test.

Table 2. Average heritability estimates (h^2 ± average standard errors) obtained from multivariate analyses, and the average difference in heritability estimates (+ average difference in standard errors) obtained from the multivariate and univariate analyses of the data from hoop pine progeny tests in south-east Queensland.

Trait	Age (years)	Number of Tests	Average h^2	Average h2 Difference
Height (m)	1	9	0.27 ± 0.06	0.004 ± -0.009
	4	10	0.24 f 0.06	0.016 ± -0.006
	8	8	0.29 ± 0.07	0.042 ± -0.004
	12	3	0.27 ± 0.08	-0.002 ± -0.008
Diameter (cm)	8	11	0.28 + 0.06	0.021 ± -0.009
	12	6	0.36 + 0.06	-0.013 ± -0.030
Straightness (1-6)	8	10	0.20 ± 0.06	0.012± 0.014
Internode length (m)	8	7	0.54 ± 0.12	0.025 ± -0.021

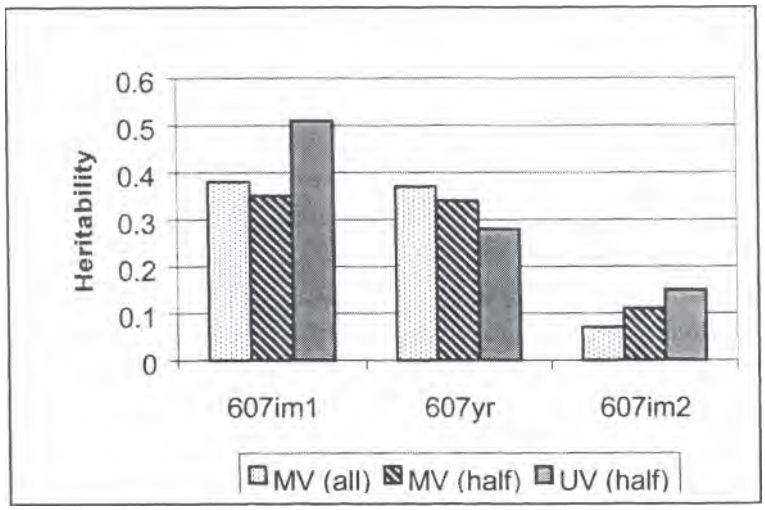

Figure 1. Heritability estimates obtained for 12 year height from multivariate (MV) and univariate analyses of data from one series of hoop pine progeny tests, using height data from all the replicates (all) and half of the replicates (half)

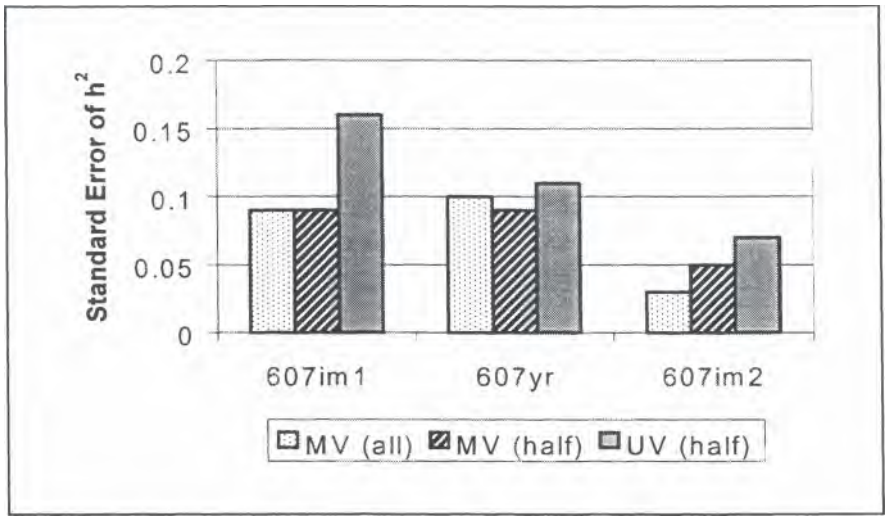

Figure 2. Standard errors of heritability (h^2) estimates obtained for 12 year height from multivariate (MV) and univariate analyses of data from one series of hoop pine progeny tests, using height data from all the replicates (all) and half of the replicates (half).

ACKNOWLEDGEMENTS

All data used in the preparation of this paper is the product of over 40 years research into the genetic improvement of hoop pine in Queensland by staff of the Queensland Forestry Research Institute (and its antecedents). The authors gratefully acknowledge their contribution and labour.

LITERATURE CITED

Boomsma, D. 1997. Development of the Southern Tree Breeding Association radiata pine breeding population. In: IUFRO '97 Genetics of Radiata Pine, Burdon, R.D. et al. (Eds.), NZFRI Bulletin No. 203, pp. 211-216.

Dean, C.A., D.G. Nikles, and K. J. Harding. 1988. Estimates of genetic parameters and gains expected from selection in hoop pine in south-east Queensland. Silvae Genetica 37: 243-247

Dieters, M.J., R.R. Woolaston, and D.G. Nikles. 1990. Internode length of hoop pine: genetic parameters and prospects for development of a long-internode breed. NZ. J. For. Res. 20:138-47.

Gilmour, A.R., B.R. Cullis, S.J. Welham, and R. Thompson. 1999. ASREML Reference Manual. NSW Agriculture Biometric Bulletin No. 3. NSW Agriculture, Orange, NSW, Australia, 210pp.

Huber, D.A. 1993. Optimal Mating Designs and Optimal Techniques for Analysis of Quantitative Traits in Forest Genetics. Ph.D. Dissertation, University of Florida, Gainesville FL., 151p.

Khuri and Sahai 1985. Variance components analysis: a selective literature survey. International Statistical Review. 53: 279-300

SAS Institute Inc. 1989. SAS/STAT User's Guide, Version 6, Fourth Edition, Volume 2. Cary NC, SAS Institute Inc., 846 p.

White, T.L., G.R. Hodge, and G.L. Powell et al. 1993. An advanced-generation tree improvement planfor slash pine in the south eastern United States. Silvae Genetica 42: 359-371.