USE OF A FINITE LOCUS MODEL TO ESTIMATE GENETIC PARAMETERS IN UNSELECTED POPULATIONS

R. J. Kerr, J. M. Henshall and B. Tier

Animal Genetics and Breeding Unit[*], University of New England, Armidale, NSW 2351.

SUMMARY
A finite locus model has been developed which models an individual's genetic value as being the aggregate effect of a series of loci. Established Monte Carlo Markov Chain (MCMC) techniques are used to sample fixed effects and the error variance, while a new sampling method is used to sample genotypes and locus effects and is based on sampling all genotypes in a pedigree jointly. The model is tested using unselected populations.
Keywords: Finite locus models, Gibbs sampler, genetic variance

INTRODUCTION
A mixed linear model that includes an individual's additive genetic effect is the basic model for current genetic evaluation systems. It is generally assumed that the inheritance of the additive genetic effect is controlled through many genes each of small effect (infinitesimal model). Genetic effects are considered as random effects and are predicted by assuming a covariance structure based on knowledge of relationships between animals.

Quantitative geneticists are now facing the challenge of including genetic marker information and non-additive genetic effects into genetic evaluation systems. The addition of these types of genetic effects into linear models complicates the genetic covariance structure to such a degree that in nearly all but trivial situations exact solutions cannot be obtained. Several authors (Fernando et al. 1994; Goddard 1998; Pong-Wong et al. 1998) have proposed using alternative modeling of the genetic effects to provide more tractable solutions. In these "gene based" models an individual's genetic value is the aggregate effect of a finite number of loci. The segregation and the genotypic effects are analysed at each locus. Variances for each type of genetic effect can then be calculated directly from the individual locus estimates. As Goddard (1998) has pointed out one or more of the loci can be considered "major loci" and may be identified genes and the remaining loci are used to explain a polygenic component. Both Goddard (1998) and Pong-Wong et al (1998) used Markov Chain Monte Carlo (MCMC) techniques to sample genotypes and genetic values at each locus. Estimates of genetic and environmental parameters were calculated from the sampled posterior distributions. Fernando et al (1994) were concerned with formulating a model to be used with maximum likelihood procedures. They were able to simplify considerably the likelihood calculations for the series of loci which explain the polygenic component. The algorithm assumes the loci are unlinked, bi-allelic and have equal and only additive effects and have equal gene frequencies.

We agree with Goddard (1998) that MCMC methods are well suited to providing solutions in a finite locus setting. The assumptions that are required in the algorithms of Fernando *et al* (1994) seem overly restrictive. As far as the authors of the present study are aware previous MCMC studies have used sampling schemes in which an individual's genotype at a single locus is sampled using a conditional probability based on the individual's nearest relatives. It is well known that this technique, when used in complex pedigrees, is susceptible to "getting stuck" in particular configurations (Thompson 1994). Tier and Henshall (pers. comm.) have described a new method of sampling genotypes, which is exact and is not susceptible to mixing problems. The purpose of the present study was to implement this method, along with current MCMC techniques, into a finite locus model. Computer simulation was used to check the accuracy of the method in estimating genetic and environmental parameters.

MATERIALS AND METHODS

The simulated trait was controlled by 100 unlinked loci to mimic an infinitesimal model. In one treatment, the deviation a_i, from the mean value of the two homozygotes at the ith locus was equal to $\sqrt{2}$ for all i. In another treatment the value for a_i, was sampled from a gamma distribution with scale and shape parameters both set to 1. This treatment mimics the situation of few genes of moderate to large effect and many genes each of small effect. Maximum values for a_i ranged between 6 and 8 over the replicates. The dominance deviation was assumed to be zero in all cases and epistasis was not modeled. An individual's observed genotypic value (G) was the sum of the contributions from all loci, $G = \Sigma g_i$. The base population consisted of 10 and 150 unrelated and non-inbred males and females, respectively. Sampling of genotypes for base animals was done on the basis of gene frequencies at each locus equaling 0.5. The expectation for the observed genetic variance for a population in perfect Hardy-Weinberg and linkage equilibrium for both treatments was 100. For each of 4 mating cycles randomly selected parents were randomly mated to produce 2 progeny per family, with males mated once. For all descendents an observation was simulated by adding a contemporary group and an individual residual effect to G. Both effects were sampled from a normal distribution with zero mean and variance = 250.

Variance components and breeding values were estimated at each cycle using both infinitesimal and finite locus models. Under the infinitesimal model variance components were estimated using the ASREML (Gilmour 1998) software package. Under the finite locus model, variance components were estimated using software, which modeled 4, 8 or 16 loci. For details on the strategy used for sampling genotypes see Tier *et al* (1999). Details of the MCMC method are very similar to that outlined in Sorenson (1998). Following a burn in period of 500 samples, 1000 samples were generated from which means and sampling variances were calculated for each parameter of interest. The genetic variance for the current population was estimated empirically from vector containing all the $\hat{G}$ values. Ten replicates of each treatment combination were completed.

RESULTS AND DISCUSSION

Table 1 presents the mean estimate for the additive genetic variance, across 10 replicates for the various treatments. Generally all methods gave unbiased estimates. It can be seen that for both the infinitesimal model (ASREML) and finite locus models (FIN4, FIN8 and FIN16) the addition of data

with each new mating cycle resulted in more precise estimates, as seen in the reduced standard errors of the means (SEM). Mean estimates of the residual error variance are not shown but again all methods gave unbiased estimates. Estimates were not statistically different from the simulated value of 250 and standard errors of the means were of the same order as that for additive genetic variance and ranged between 12 and 52. The correlation between true and estimated genotypic values $r_{G,\hat{G}}$, calculated under the finite locus model and the correlation between true and estimated additive genetic values $r_{a,\hat{a}}$, calculated under the infinitesimal model were also computed and compared. Both types of correlation were consistently equal or close to 0.6, which is expected considering the simulated heritability was 0.29.

Table 1. Mean (SEM) of estimates of additive genetic variance using an infinitesimal model (ASREML) and a finite locus model (FIN4, FIN8 or FIN16), for two distributions of gene effects

Mating cycle	ASREML	FIN4	FIN8	FIN16
		$Va = 100\ (a_i = \sqrt{2}\ \text{for}\ i = 1,...,100)$		
1	89. (37)	119 (57)	116 (33)	118 (32)
2	105 (37)	94 (25)	96 (34)	100 (28)
3	114 (26)	95 (18)	85 (29)	92 (21)
4	107 (21)	101 (22)	85 (27)	93 (13)
		$Va = 100\ (a_i \sim \text{GAMMA}(1,1))$		
1	118 (48)	99 (47)	128 (61)	153 (65)
2	113 (41)	110 (43)	104 (45)	123 (43)
3	117 (27)	109 (30)	98 (35)	100 (37)
4	113 (24)	110 (23)	100 (35)	112 (35)

Under the conditions simulated it would appear that 4 hypothetical loci are adequate for a finite locus model to accurately estimate genetic variances and individual aggregate genotypic values. From Table 1 it can be seen that though the estimates remain statistically not different from 100, there is some real evidence of a trend for the FIN8 and FIN16 treatments to underestimate the additive genetic variance in the later mating cycles. Inbreeding and genetic drift can contribute to increased covariances between additive genetic values in an unselected population and cause a reduction in the additive genetic variance. Generally, there is a need to develop criteria to determine the optimum number of hypothetical loci needed to explain a polygenic component.

In Figure 1 two plots are presented which describe the distribution of estimated additive gene effects (a_i) across all replicates and mating cycles for the FIN16 treatment, when the actual simulated distribution was either a uniform or gamma distribution. The effects of each genotype class were sampled using conditional normal distributions. To a large degree the distribution of estimated gene effects reflects this model – estimated effects have roughly a normal distribution. However the distribution on the right, which corresponds to the case of the simulated distribution being a gamma distribution, is noticeably more skewed. The same comparison for the FIN4 treatment did not reveal any noticeable differences. From these results it is apparent that 16 or more assumed hypothetical

loci might offer the potential to indicate any underlying distribution of gene effects. Of course a gamma distribution could have been used to sample gene effects. This will need to be investigated.
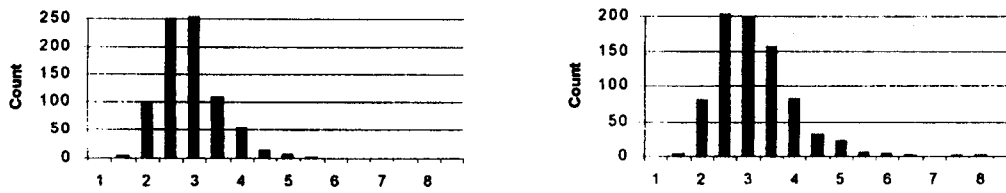

Figure 1. Distribution of estimated gene effects under a finite locus model assuming 16 loci, when the simulated distribution was a uniform (left plot) or gamma (right plot) distribution.

The main difference between this study and that of Goddard (1998) has been the method in sampling genotypes. To avoid problems of slow convergence Goddard sampled simultaneously the genotypes of sires and progeny from terminal families, as well as introducing a mutation rate and retaining samples that showed no mutation. The sampling scheme used in the present study samples all genotypes in the pedigree jointly. A mutation rate could also be considered, though for the present study, was considered unnecessary. Goddard also assumed the error variance and gene frequencies in the base population known whereas in the present study all parameters were considered unknown.

In conclusion the study has demonstrated that a finite locus model is at least comparable to a standard REML analysis in estimating an additive genetic variance of a polygenic trait in an unselected population. In future work, selected populations will be used and models will be developed to include non-additive gene effects.

REFERENCES

Fernando, R.L., Strickcr, C. and Elston, R.C. (1994) *Theor. Appl. Genet.* **88**:573

Gilmour, A. (1999) "ASREML users guide". NSW Agriculture, Orange, NSW

Goddard, M.E. (1998) *Proc. 6th World Congr. Genet. Appl. Livest. Prod.* **26**:33

Pong-Wong, R., Shaw, F. and Woolliams, J.A. (1998) *Proc. 6th World Congr. Genet. Appl. Livest. Prod.* **26**:41

Sorenson, D. (1998) "Gibbs sampling in quantitative genetics". Danish Institute of Agricultural Sciences, Copenhagen

Tier, B., Henshall, J.M. and Kerr, R.J. (1999) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **13**:

Thompson, E.A. (1994) Phil. Trans. R. Soc. Lond. B. 34: 355