Genet. Res., Camb. (1999), **74**, *pp.* 179–187. Printed in the United Kingdom

Estimation of additive and dominance variance components in finite polygenic models and complex pedigrees

F.-X. DU, I. HOESCHELE* AND K. M. GAGE-LAHTI

Departments of Dairy Science and Statistics, and Interdepartmental Genetics Program, Virginia Polytechnic Institute & State University, Blacksburg, VA 24061-0315, USA

(*Received 16 September 1997 and in revised form 21 October 1998 and 16 February 1999*)

Summary

Estimation of variance components with the finite polygenic model (FPM) was evaluated. Phenotypic data for a 6300-pedigree simulated under a wide range of additive genetic models were analysed with constant homozygote difference across loci using deterministic Maximum Likelihood (DML) and a Bayesian method implemented via Gibbs sampling (BGS). Results indicate that under no selection, both DML and BGS accurately estimated the variance components, with a FPM of 5 loci or more. When both analysis methods were applied to equivalent data sets on populations that had undergone selection, the DML method produced upward biased estimates of additive genetic variation and heritability due to its use of pedigree loop cutting, while BGS provided more accurate estimation. BGS was extended to non-additive FPMs with variable homozygote differences and dominance effect across loci. This method was used to analyse data simulated under two genetic models with positive, completely dominant gene action at all loci. Results indicate that the estimates of additive and dominance variances slowly increase as the number of loci in the FPM for analysis increases, while accuracy of predicting individual breeding values and dominance deviations remains unaffected. For the simulated pedigree structure, a FPM with 10 loci or slightly fewer appears to be appropriate for variance component estimation in the presence of dominance.

1. Introduction

A finite polygenic model (FPM) was first proposed by Thompson & Skolnick (1977) for estimating the heritability of longevity in complex human pedigrees. In contrast to the infinitesimal model, in which a trait is assumed to be influenced by a normally distributed polygenic component consisting of an infinite number of additive genes of small effect, a FPM fits a finite number of unlinked polygenic loci to describe the genetic covariance among pedigree members. The additive FPM results in a computationally more simple likelihood than the infinitesimal model, under the mixed model of inheritance containing both a major quantitative trait locus (QTL) and a polygenic component (Fernando *et al.*, 1994).

The inclusion of dominance variation under the infinitesimal model is difficult for theoretical and computational reasons, especially when inbreeding is present (DeBoer & Hoeschele, 1993). Under inbreeding, with no epistasis, it is necessary to fit extra parameters, including dominance variance and covariance between additive and dominance effects in a completely inbred population with allele frequencies equal to those in the base population, and the effect of inbreeding depression, in the model for phenotypes (Gillois, 1964; Harris, 1964; DeBoer & Hoeschele, 1993). These complications to the genetic covariance structure under the infinitesimal model motivated investigation of an alternative model, the FPM.

The mixed models of inheritance of Fernando *et al.* (1994) and Stricker *et al.* (1995*b*), which contain finite polygenic loci along with a segregating major QTL,

* Corresponding author: Department of Dairy Science, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061-0315, USA. Telephone: +1 (540) 231 4760. Fax: +1 (540) 231 5014. e-mail: inah@vt.edu

appeared accurately to estimate major gene effects and polygenic heritability. In this contribution, the ability of the FPM to estimate the narrow-sense heritability in the absence of a major gene was investigated first. Data simulated under various additive FPMs and an infinitesimal model, all either with or without selection, were analysed with additive FPMs assuming different numbers of polygenic loci as suggested by Thompson & Skolnick (1977). The additive analysis was performed using two methods: a deterministic Maximum Likelihood method (DML) implemented via the Segregation and Linkage Analysis for Pedigrees package (SALP) by Stricker *et al.* (1995*a*), and a Bayesian approach implemented via Gibbs sampling (BGS). For non-additive analysis, the BGS program was modified to sample genotypes at individual loci, and to fit variable homozygote differences and dominance deviations in the FPM for each biallelic locus, allowing estimation of both the additive and dominance variances across loci, as well as the prediction of individual breeding values (BVs) and dominance deviations (DVs).

2. Methodology

(i) *Finite polygenic model with polygenic numbers*

In the additive FPM, each polygenic locus contains two alleles with equal, fixed gene frequency of 0·5 and equal but unknown homozygote difference, as outlined in Fernando *et al.* (1994) but without the major gene. Under this FPM, the additive genotypic value of an individual can be completely described as a function of its polygenic number v defined as the total number of favourable alleles across all loci, and constant homozygote difference a. For k loci, there are $2k+1$ possible polygenic numbers with values from 0 to $2k$. Assuming single records, the linear model for phenotype is

$$\boldsymbol{y} = \boldsymbol{X}\beta + (\boldsymbol{v} - \boldsymbol{1}k)\,a + \boldsymbol{e}, \tag{1}$$

where $\boldsymbol{y}$ is a vector of phenotypes; $\boldsymbol{X}$ is a known design-covariate matrix relating observations in $\boldsymbol{y}$ to the vector of fixed effects (β); $\boldsymbol{v}$ is a vector of polygenic numbers; k is the number of biallelic loci; $\boldsymbol{1}$ is a vector of ones; a is the homozygote difference which is constant across loci; and $\boldsymbol{e}$ is a vector of residuals.

For simplicity, the v of an individual is assumed to be conditionally independent of the v of any ancestor or sibling given the v values of the parents, although under Mendelian inheritance only approximate conditional independence holds. Occasionally, the polygenic numbers of full-sibs are not conditionally independent given the polygenic numbers of their parents, as shown in a counter-example in Fernando *et al.* (1994).

(*a*) *Deterministic ML*

The likelihood of a pedigree is given by

$$f(\boldsymbol{y}\,|\,\beta, a, \sigma_e^2) = \sum_v \left[\prod_{i=1}^{n_b} f(v_i) f(y_i\,|\,v_i, \beta, a, \sigma_e^2) \times \prod_{i=n_b+1}^{n} f(v_i\,|\,v_s, v_d) f(y_i\,|\,v_i, \beta, a, \sigma_e^2) \right], \tag{2}$$

where σ_e^2 is the residual variance; n_b is the number of base (founder) individuals in the population; n is the total number of individuals; and $f(y_i\,|\,v_i, \beta, a, \sigma_e^2)$ is the penetrance function with normal density. The transition probability that a non-base individual will have polygenic number v_i given the polygenic numbers of its sire (v_{s_i}) and dam (v_{d_i}) is denoted by $f(v_i\,|\,v_{s_i}, v_{d_i})$, and can be calculated recursively according to appendix C of Fernando *et al.* (1994). Probability $f(v_i)$ that a base individual will have polygenic number v_i is equal to the population frequency of v_i. The summation is over all possible vectors $\boldsymbol{v}$ of polygenic numbers.

For evaluation of likelihood, the ML program SALP cuts pedigree loops according to the method of Stricker *et al.* (1995*c*), in which an additional founder with an identical phenotype is inserted at the site of the cutting, resulting in an approximation to the likelihood. After loops are cut, the likelihood of the data is computed using the recursive algorithm of Fernando *et al.* (1993):

$$f(y\,|\,\beta, a, \sigma_e^2) = \prod_{\text{pedigrees}} \sum_{v_i} \left[\alpha_i(v_i) f(y_i\,|\,\beta, a, \sigma_e^2, v_i) \prod_{j \in S_i} p_{i,j}(v_i) \right], \tag{3}$$

where $\alpha_i(v_i)$ denotes the anterior probability of individual i having polygenic number v_i, $p_{i,j}(v_i)$ the posterior probability of i through mate j, and S_i the set of individual i's mates. Further, i represents an arbitrary individual in a pedigree, and the likelihood equals the product of likelihood for all independent pedigrees within the data set. The likelihood in (3) is then deterministically maximized using the Downhill Simplex algorithm of Nelder & Mead (1965). After SALP converged the first time, it was restarted with those estimates to ensure that convergence had occurred.

(*b*) *Bayesian analysis via Gibbs sampling*

The joint posterior density of the parameters and polygenic numbers can be written as

$$f(\beta, a, \sigma_e^2, \boldsymbol{v}\,|\,\boldsymbol{y}) \propto \prod_{i=1}^{n} f(y_i\,|\,v_i, \beta, a, \sigma_e^2) \prod_{i=1}^{n_b} \times f(v_i) \prod_{i=n_b+1}^{n} f(v_i\,|\,v_{s_i}, v_{d_i}) f(\beta) f(a) f(\sigma_e^2), \tag{4}$$

https://doi.org/10.1017/S0016672399003948 Published online by Cambridge University Press

where $f(\beta)$, $f(a)$ and $f(\sigma_e^2)$ are priors for β, α and σ_e^2 respectively.

Flat but bounded priors were assumed for β, a and σ_e^2. The effects in β and a were sampled from fully conditional, normal distributions, and the error variance was sampled from an inverse-χ^2 distribution derived in Wang *et al.* (1993). The polygenic number of individual i was sampled from the distribution with probability density

$$f(v_i \mid \boldsymbol{y}, \beta, a, \sigma_e^2, \boldsymbol{v}) \propto f(y_i \mid v_i, \beta, a, \sigma_e^2)\,[\boldsymbol{I}(i) f(v_i) + (1 - \boldsymbol{I}(i)) f(v_i \mid v_{s_i}, v_{d_i})] \cdot \prod_{o_i} f(v_{o_i} / v_i, v_{m_i}), \quad (5)$$

where $\boldsymbol{I}(i)$ is an indicator function that equals 1 if individual i is a base animal and 0 otherwise.

Parameters β, a and σ_e^2 were estimated as averages of samples generated using the Gibbs sampler. Additive genetic variance (σ_a^2) was calculated as $0{\cdot}5\, ka^2$ and narrow-sense heritability (h^2) as

$$\frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2}$$

in each Gibbs cycle using that cycle's samples of α and σ_e^2.

(ii) *Finite polygenic model with genotypes at individual loci*

To generalise the FPM to also incorporate dominance variation, the model becomes

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Z}_G \boldsymbol{a} + \boldsymbol{W}_G \boldsymbol{d} + \boldsymbol{e}, \quad (6)$$

where $\boldsymbol{y}$, $\boldsymbol{X}$, $\boldsymbol{\beta}$ and $\boldsymbol{e}$ are defined as before, $\boldsymbol{a}$ is a vector of k variable homozygote differences at k biallelic loci; $\boldsymbol{d}$ is a vector of k variable dominance deviations at k biallelic loci; $\boldsymbol{Z}_G$ is a design matrix with k columns containing coefficients of -1, 0 and 1 corresponding to the three genotypes at a biallelic locus; and $\boldsymbol{W}_G$ is a design matrix with k columns containing coefficients of 1 for the heterozygous genotype and 0 for the homozygous genotypes. Both $\boldsymbol{Z}_G$ and $\boldsymbol{W}_G$ can be constructed based on known $\boldsymbol{G}$, a matrix of genotypes of all individuals at all k biallelic loci. The allele frequency at each of the k biallelic loci was fixed at 0·5. The unknowns now include the parameters in $\boldsymbol{\beta}$, $\boldsymbol{a}$, $\boldsymbol{d}$, σ_e^2 and the nk genotypes in $\boldsymbol{G}$. The joint posterior density of all unknowns can be written as

$$f(\boldsymbol{\beta}, \boldsymbol{a}, \boldsymbol{d}, \sigma_e^2, \boldsymbol{G} \mid \boldsymbol{y}) \propto \prod_{i=1}^{n} f(y_i \mid \boldsymbol{\beta}, \boldsymbol{a}, \boldsymbol{d}, \sigma_e^2, \boldsymbol{G}_i) \prod_{i=1}^{n_b} \prod_{j=1}^{k} f(g_{ij}) \prod_{i=n_b+1}^{n} \prod_{j=1}^{k} \times f(g_{ij} \mid g_{s_{ij}}, g_{d_{ij}}) \cdot f(\boldsymbol{\beta}) f(\boldsymbol{a}) f(\boldsymbol{d}) f(\sigma_e^2), \quad (7)$$

where $\boldsymbol{G}_i$ is row i of $\boldsymbol{G}$, $\boldsymbol{G}_i = \{g_{ij}\}$ with g_{ij} being the genotype of individual i at locus j, and $f(g_{ij})$ and $f(g_{ij} \mid g_{s_{ij}}, g_{d_{ij}})$ are population frequency and transition probability of genotype g_{ij}, respectively.

Flat but bounded priors were assumed for $\boldsymbol{\beta}$, $\boldsymbol{a}$, $\boldsymbol{d}$ and σ_e^2. Parameters in $\boldsymbol{\beta}$, $\boldsymbol{a}$ and $\boldsymbol{d}$ were sampled from fully conditional, normal distributions, and the error variance was sampled from an inverse-χ^2 distribution derived in Wang *et al.* (1993). Genotypes of non-parent individuals at single loci were sampled from standard, fully conditional distributions. For sampling genotypes of parents, the sampling scheme of Janss *et al.* (1995) was extended to sample both sires and dams unconditionally on their final offspring, which further improves mixing of genotype states.

Parameters $\boldsymbol{\beta}$, $\boldsymbol{a}$, $\boldsymbol{d}$ and σ_e^2 were estimated as averages of samples generated using the Gibbs sampler. Additive genetic variance (σ_a^2) and dominance variance (σ_d^2) were calculated as

$$0{\cdot}5 \sum_{i=1}^{k} a_i^2 \quad \text{and} \quad 0{\cdot}25 \sum_{i=1}^{k} d_i^2,$$

respectively, where a_i and d_i are the ith element of vector $\boldsymbol{a}$ and $\boldsymbol{d}$, respectively. Individual BVs and DVs were predicted as $\boldsymbol{z_i a}$ and $\boldsymbol{w_i d}$, where $\boldsymbol{z_i}$ and $\boldsymbol{w_i}$ are the ith row of design matrix $\boldsymbol{Z}_G$ and $\boldsymbol{W}_G$, respectively. Correlation coefficients between true and predicted BVs and between true and predicted DVs were estimated to evaluate the accuracy of the prediction.

(iii) *Data and analysis*

The simulated population structure consisted of $n = 6300$ individuals over one base generation and three discrete offspring generations. Every generation, 50 males and 250 females were randomly selected, with each male randomly mated to five females. Females produced 8-offspring litters divided equally in half by sex, giving each sire 40 progeny. Twenty independent replications under the same population structure were generated. Unintentional inbreeding and the mating structure of the unselected population resulted in 945 cuts of inbreeding and mating loops in the pedigree by the deterministic ML program, SALP. Under selection, the size and structure of the pedigree were maintained. In each generation, the 50 males and 250 females with the highest breeding values as predicted by Best Linear Unbiased Prediction (BLUP) with the animal model program JAA (Misztal, 1989) were selected to be parents. This scheme resulted in a selected proportion of 5% for males and 25% for females. There were on average 985 cuts made to pedigree loops.

Phenotypic data were simulated under four additive models with true values of $\sigma_e^2 = 50$ and $\sigma_a^2 = 50$ for residual and additive genetic variance (across loci), respectively, and a mean (μ) equal to zero as the only element of β. Model 1 was the infinitesimal model, while models 2, 3 and 4 were FPMs of 40 loci of equal effect, 18 loci of diminishing effect and 5 loci of equal

https://doi.org/10.1017/S0016672399003948 Published online by Cambridge University Press

Table 1. *Genetic models used for data simulation*

Model	No. of loci	*a*	*d*
1	Infinitesimal	—	0
2	40	Constant	0
3	18	Diminishing	0
4	5	Constant	0
5	18	Diminishing	Positive complete
6	20	Constant	Positive complete

a is the homozygote difference and *d* is the dominance deviation across loci. The 18 loci in models 3 and 5 included one locus with additive genetic variance 25, 2 loci with variance 5, 5 loci with variance 2, and 10 loci with variance 0·5.

effect, respectively, with all loci being unlinked and biallelic with $P = 0{\cdot}5$ (Table 1). The 18 loci in model 3 included one locus with variance 25, two loci with variance 5, five loci with variance 2, and 10 loci with variance 0·5.

Two genetic models with dominance were also simulated for this 6300-pedigree (Table 1). Model 5 was simulated by modifying the model 3 structure with positive complete dominance at all 18 loci. Model 6 contained 20 loci with equal additive genetic and positive complete dominance effects across loci. In both models, $\sigma_e^2 = 50{\cdot}0$, $\sigma_a^2 = 50{\cdot}0$ and $\sigma_d^2 = 25{\cdot}0$, where σ_d^2 denotes dominance variance. Ten replications were simulated for each non-additive model.

For analyses under additive gene action, data sets from the four genetic models were all analysed by both the DML and the BGS methods, all using additive FMPs of 3, 4 or 5 loci. Starting values of $\mu = 5$, $\sigma_e^2 = 65$ and $\sigma_a^2 = 35$ were arbitrarily chosen, and the starting homozygote difference a was calculated as $2\sqrt{\sigma_a^2/2k}$, resulting in $a = 4{\cdot}83$, $4{\cdot}18$ or $3{\cdot}74$ for the three different analysis models. Under selection, the data sets were analysed by the 5-loci FPM only. Data sets generated under the dominance models were analysed with non-additive FPMs of 5, 10 or 20 loci, with variable homozygote differences and dominance deviations across loci. Arbitrary starting values of $\mu = 0$, $\sigma_e^2 = 10$, $\boldsymbol{a} = 1$ and $\boldsymbol{d} = 0$ were used for all non-additive FPMs. For all additive and non-additive FPMs, the starting value for $\boldsymbol{G}$ was heterozygous at all loci for all individuals.

To evaluate the accuracy of BV and DV prediction for all non-additive FPMs, pedigree members were divided into three groups based on their offspring numbers: > 10, 1–10, and no offspring. The correlation coefficients between the true and predicted BVs and between the true and predicted DVs were estimated for each group and the entire pedigree.

3. Results

(i) *Additive genetic FPM: unselected populations*

For analysis under the additive genetic model, chains of 10000 Gibbs cycles with 50 burn-in cycles were used, and the average Monte Carlo standard errors were below 0·5% for variance components and heritability (data not shown). These were calculated as suggested by Geyer (1992) using the square root of the variance of a sample mean found in Sorensen *et al.* (1995). Despite large numbers of offspring for some individuals, the univariate sampling distribution for the polygenic number in (5) allows fast mixing of the sampler, unlike the slow mixing which occurs due to large numbers of offspring when sampling genotypes (Janss *et al.*, 1995). The $2k+1$ polygenic numbers are more finely discretized than genotypes at individual loci, with many different genotype combinations having the same polygenic number, causing more rapid mixing.

Table 2. *Average parameter estimates* (*and standard errors*) *obtained from 20 replicates of an additive 3-loci FPM Bayesian Gibbs Sampler* (*BGS*) *or SALP analysis with constant additive effect for data simulated without selection under various genetic models* (*see Table 1*)

		Genetic models			
		Infinitesimal	40 Equal	18 Diminishing	5 Equal
BGS	μ	−0·22 (0·13)	−0·02 (0·12)	−0·19 (0·10)	−0·23 (0·14)
	σ_e^2	50·71 (0·53)	49·65 (0·57)	49·93 (0·44)	50·84 (0·35)
	σ_a^2	49·10 (0·87)	51·16 (1·08)	50·49 (0·62)	49·28 (0·74)
	h^2	0·49 (0·007)	0·51 (0·008)	0·50 (0·005)	0·49 (0·005)
SALP	μ	0·09 (0·22)	0·05 (0·19)	−0·18 (0·19)	−0·08 (0·12)
	σ_e^2	51·6 (0·67)	51·9 (0·93)	50·3 (0·84)	49·9 (0·65)
	σ_a^2	47·4 (1·09)	46·7 (1·46)	50·9 (1·51)	49·4 (1·31)
	h^2	0·48 (0·01)	0·47 (0·01)	0·50 (0·01)	0·50 (0·01)

True values are mean (μ) = 0·0, error variance (σ_e^2) = 50·0, additive variance (σ_a^2) = 50·0 and heritability (h^2) = 0·50.

https://doi.org/10.1017/S0016672399003948 Published online by Cambridge University Press

Table 3. *Average parameter estimates* (*and standard errors*) *obtained from 20 replicates of an additive 5-loci FPM Bayesian Gibbs Sampler* (*BGS*) *or SALP analysis with constant additive effect for data simulated without selection under various genetic models* (*see Table 1*)

		Genetic models			
		Infinitesimal	40 Equal	18 Diminishing	5 Equal
BGS	μ	−0·26 (0·11)	0·24 (0·12)	−0·25 (0·11)	−0·22 (0·11)
	σ^2_e	50·53 (0·44)	49·41 (0·60)	49·88 (0·44)	49·32 (0·67)
	σ^2_a	49·08 (0·76)	51·01 (1·08)	50·31 (0·59)	49·32 (0·67)
	h^2	0·49 (0·006)	0·51 (0·008)	0·50 (0·005)	0·49 (0·005)
	a	—	—	—	4·44 (0·03)
SALP	μ	0·11 (0·20)	0·34 (0·23)	−0·14 (0·20)	−0·18 (0·13)
	σ^2_e	50·7 (0·79)	49·9 (0·73)	50·8 (0·68)	51·3 (0·74)
	σ^2_a	48·1 (1·09)	48·8 (1·31)	49·2 (1·22)	48·2 (1·44)
	h^2	0·49 (0·01)	0·49 (0·01)	0·49 (0·01)	0·48 (0·01)
	a	—	—	—	4·38 (0·06)

True values are mean (μ) = 0·0, error variance (σ^2_e) = 50·0, additive variance (σ^2_a) = 50·0, heritability (h^2) = 0·50 and half of the homozygote difference (a) = 4·48.

Table 4. *Average parameter estimates* (*and standard errors*) *obtained from 20 replicates of an additive 5-loci FPM Bayesian Gibbs Sampler* (*BGS*) *or SALP analysis with constant additive effect for data simulated with selection under various genetic models* (*see Table 1*)

		Genetic models			
		Infinitesimal	40 Equal	18 Diminishing	5 Equal
BGS	μ	0·98 (0·14)	0·63 (0·07)	−0·16 (0·10)	0·12 (0·14)
	σ^2_e	52·9 (0·28)	51·8 (0·31)	48·8 (0·30)	49·7 (0·67)
	σ^2_a	58·8 (0·67)	55·0 (0·68)	48·7 (0·66)	51·2 (0·76)
	h^2	0·53 (0·004)	0·52 (0·004)	0·50 (0·01)	0·51 (0·01)
	a	—	—	—	4·52 (0·04)
SALP	μ	9·09 (0·19)	8·92 (0·14)	8·28 (0·13)	8·51 (0·19)
	σ^2_e	50·4 (0·47)	51·0 (0·48)	48·9 (0·56)	47·3 (0·94)
	σ^2_a	66·6 (1·04)	60·9 (0·88)	55·4 (0·98)	59·2 (1·05)
	h^2	0·57 (0·01)	0·56 (0·02)	0·53 (0·01)	0·56 (0·01)
	a	—	—	—	4·86 (0·04)

True values are mean (μ) = 0·0, error variance (σ^2_e) = 50·0, additive variance (σ^2_a) = 50·0, heritability (h^2) = 0·50 and half of the homozygote difference (a) = 4·48.

Posterior mean estimates of the parameters of interest and empirical standard errors for the unselected populations are given in Tables 2 and 3 for the 3- and 5-loci additive FPMs, respectively. Analysing FPM models with 3-loci or more using BGS produced accurate estimates of population mean, additive genetic variance and heritability (Tables 2 and 3 and results from the 4-loci FPM analysis not shown). For data simulated under models 3 and 4, SALP estimates for the additive genetic variance and heritability appeared to be accurate. For data simulated under the infinitesimal or 40-equal-polygenic-loci models, SALP estimates of heritability equalled 0·479 and 0·474 using the 3-loci FPM, and 0·476 and 0·475 using the 4-loci FPM, respectively (Table 2 and results not shown). Tested with a two-sided t-test at the 0·05 significance level, these heritability SALP estimates significantly underestimated heritability, but all other SALP estimates appear to be accurate (Tables 2 and 3 and data not shown). When the genetic model of 5-loci was used as the analysis FPM, the effect of the favourable allele at each locus a was estimated accurately with both DML and BGS (Table 3).

(ii) *Additive genetic FPM: selected populations*

Table 4 gives parameter estimates under selection, with the 5-loci additive FPM as the analysis model. The Bayesian method produced accurate parameter

https://doi.org/10.1017/S0016672399003948 Published online by Cambridge University Press

Table 5. *Parameter estimates* (*and standard errors*) *averaged across 10 replicates of a 5-loci FPM Bayesian Gibbs sampler* (*BGS*) *analysis with variable dominance effects for data simulated without selection under model 4* (*see Table 1*)

	Additive effects	
	Constant	Variable
μ	0·09 (0·08)	0·05 (0·09)
σ^2_a	50·8 (1·22)	51·1 (1·25)
σ^2_d	3·7 (0·43)	3·9 (0·39)
σ^2_e	47·0 (0·71)	46·8 (0·70)
h^2	0·52 (0·009)	0·52 (0·009)

True values are mean (μ) = 0·0, additive genetic variance (σ^2_a) = 50·0, dominance genetic variance (σ^2_d) = 0·0, error variance (σ^2_e) = 50·0 and heritability (h^2) = 0·50.

estimates for the two genetic models with the fewest numbers of loci, models 3 and 4, but overestimated additive genetic variance and heritability for data simulated under models 1 and 2. SALP produced biased estimates for all four genetic models and greatly overestimated the mean. There appeared to be a slight trend with both methods of increasingly overestimating genetic variance as the number of loci in the genetic model increased. Therefore, the selected data sets were reanalysed using a 20-loci FPM for the Bayesian approach and a 14-loci FPM for SALP (14 was the maximum number of loci fitted by SALP). However, increasing the number of loci in the analysis model did not significantly improve the SALP estimates of either genetic variance or the mean, and the Bayesian estimates were virtually unchanged as well (results not shown).

(iii) *Non-additive genetic FPM*

As a preliminary step in investigating the estimation of both additive and dominance variances, 10 data sets simulated under model 4 were analysed with 5-loci non-additive FPMs using the BGS method modified to sample genotypes at individual loci and fit both homozygote differences and dominance deviations at individual loci. The results, shown in Table 5, indicate that constancy versus variability of homozygote differences across loci has virtually no effect on accuracy of parameter estimation, and that the analysis correctly estimates a dominance variance near zero when no dominance variation exists.

When dominance variance was non-zero in the simulation models, analysis with non-additive FPMs required larger numbers of Gibbs cycles to produce sufficiently small Monte Carlo standard errors. Gibbs chains of length 100 000 with 2000 burn-in cycles were used in the analyses reported in Tables 6 and 7. The Monte Carlo standard errors for all variance components were below 0·5% of the genetic variance (results not shown). Despite large correlations among the parameters, joint sampling of μ, $\boldsymbol{a}$ and $\boldsymbol{d}$ did not significantly improve MC standard errors (results not shown).

As shown in Table 6, variance component estimation was affected by the choices of both analysis and simulation models. Variable homozygote differences and dominance deviations across loci were assumed in all analysis FPMs. Accurate estimates of additive genetic variance and downward biased estimates of dominance variance, with the extent of bias depending on the genetic model, was found when the 5-loci FPM was used for analysis. As the number of loci in the analysis FPM increased, the estimate of dominance variance slowly increased. When 20 loci were included in the analysis FPM, additive genetic variance and dominance variance were slightly and

Table 6. *Variance component estimates* (*and standard error*) *obtained from analysis of FPM with variable additive and dominance effects using Bayesian Gibbs sampler* (*BGS*) *for data simulated under models 5 and 6* (*see Table 1*). *Estimates are averaged across several replicates* (*10 for FPMs with 5 and 10 loci and 5 for FPMs with 20 loci*)

Genetic model	No. of analysis loci	σ^2_e	σ^2_a	σ^2_d
Model 5	5	51·2 (0·91)	51·9 (1·34)	23·8 (0·97)
	10	46·8 (0·96)	52·7 (1·36)	28·3 (1·03)
	20	39·5 (1·83)	56·0 (2·70)	35·6 (1·69)
Model 6	5	54·4 (0·92)	51·0 (0·94)	20·9 (1·11)
	10	48·6 (1·05)	51·9 (1·02)	26·7 (1·21)
	20	42·8 (2·18)	53·8 (0·93)	32·9 (2·75)

True values are error variance (σ^2_e) = 50·0, additive variance (σ^2_a) = 50·0 and dominance variance (σ^2_d) = 25·0.

https://doi.org/10.1017/S0016672399003948 Published online by Cambridge University Press

Table 7. *Accuracy* (*and standard errors*) *for predicting individual breeding values and dominance deviations from analysis of FPMs with variable additive and dominance effects using Bayesian Gibbs sampler* (*BGS*) *for data simulated under Models 5 and 6* (*see Table 1*). *Estimates are averaged across several replicates* (*10 for FPMs with 5 and 10 loci and 5 for FPMs with 20 loci*)

Genetic model	No. of analysis loci	Breeding value				Dominance deviation			
		r_{AT}	r_{A1}	r_{A2}	r_{A3}	r_{DT}	r_{D1}	r_{D2}	r_{D3}
Model 5	5	0·764 (0·005)	0·905 (0·005)	0·810 (0·004)	0·753 (0·005)	0·534 (0·007)	0·676 (0·009)	0·570 (0·008)	0·524 (0·007)
	10	0·763 (0·005)	0·905 (0·005)	0·810 (0·005)	0·753 (0·005)	0·532 (0·006)	0·663 (0·010)	0·568 (0·008)	0·523 (0·007)
	20	0·767 (0·004)	0·907 (0·006)	0·815 (0·010)	0·756 (0·003)	0·537 (0·004)	0·675 (0·016)	0·569 (0·012)	0·527 (0·005)
Model 6	5	0·746 (0·004)	0·881 (0·007)	0·783 (0·004)	0·737 (0·004)	0·503 (0·005)	0·562 (0·022)	0·515 (0·009)	0·499 (0·004)
	10	0·748 (0·003)	0·884 (0·007)	0·783 (0·005)	0·739 (0·003)	0·508 (0·004)	0·568 (0·022)	0·516 (0·009)	0·504 (0·004)
	20	0·749 (0·003)	0·874 (0·009)	0·789 (0·009)	0·739 (0·002)	0·502 (0·006)	0·537 (0·034)	0·503 (0·011)	0·500 (0·006)

r_{ij}: correlation coefficient between the true and predicted i in group j: $i = A$ and D for breeding value and dominance deviation, respectively; $j = T$, 1, 2 and 3, for the entire pedigree, groups of animals with 10, 1–10 and no offspring, respectively.

greatly overestimated, respectively, coupled with severe underestimation of residual variance. For data simulated under model 6, 10 loci in the analysis FPM appear to be appropriate. For data from model 5 in which a major gene is segregating, a number between 5 and 10 loci in the analysis FPM appears to be optimum. In contrast, the accuracy of predicting both BVs and DVs is virtually constant across different FPMs with various numbers of loci (Table 7). For DV prediction, the results show that progeny information improves the prediction of parental DVs. The segregation of a major gene in model 5 appears to improve the accuracy of BV prediction slightly and that of DV prediction considerably. As the genetic model was changed from model 6 to model 5, the correlation between predicted and true BVs was increased by approximately 2%, and the correlation between predicted and true DVs was increased by 5–25%. Therefore, the accuracy of prediction of BVs and DVs depends on the genetic model of a trait, while it appears to be unaffected by the choice of analysis FPM.

4. Discussion

In the case of no selection, the results obtained in this study show that BGS with a FPM of three loci or more gives accurate estimates of additive variance and narrow-sense heritability under a wide range of additive genetic models. In contrast, DML analyses of 3- or 4-loci FPMs produced underestimation of additive genetic variance and heritability for data simulated under some genetic models. This underestimation appears to be corrected as the number of loci in the analysis model increases to five, despite an approximation to the likelihood that resulted from cutting pedigree loops in SALP. Hence, it would be prudent to fit at least five loci in an additive FPM. Under intense selection, severe overestimation of additive variance and heritability was found for the DML analysis, with degree of bias depending on the genetic model. The overestimation probably resulted from the practice of loop cutting. When a loop was cut and an individual was replaced by a founder with the same phenotype, that pseudo-founder had a higher probability of having a 'poor' genotype at each locus than the original individual. As a consequence, the overall mean may be biased upwards, and the additive genetic variance increased. Recently, Hagger & Stricker (1998) reported similar estimation of additive genetic variance by SALP in their FPM analysis of egg weights from a multigeneration selection experiment with chicken. Although some upward bias was also detected when analysing data simulated under the infinitesimal and forty-equal-polygenic-loci models, the BGS method gave much more accurate estimates of additive genetic variance and heritability than DML for data simulated under all genetic models. Therefore, sampling-based algorithms, which do not require loop-cutting, should be preferred.

For data simulated under non-additive models, estimation of additive, dominance, and residual variances appear to be accurate only when a non-additive FPM model with an appropriate number of loci is fitted. However, as the number of loci in the analysis FPM with variable dominance effects

https://doi.org/10.1017/S0016672399003948 Published online by Cambridge University Press

increased, the estimates of dominance and additive variances slowly increased, coupled with a steady decrease in the estimate of residual variance. Although the optimal FPM for analysis depends on the genetic model of a trait, the 5- or 10-loci FPMs give reasonably accurate estimates for the simulated population structure. Moreover, the number of loci has little effect on the accuracy of individual BV and DV predictions. Upward biased estimates of dominance variance were obtained using the 20-loci FPM for analysis, and the overestimation did not significantly diminish when the data were analysed using the 20-loci FPM with constant additive effects and variable dominance effects across loci. While more research is needed to clarify the cause(s) of bias, one possible reason for the overestimation is that the number of parameters in an analysis FPM is too large relative to the sample size and (or) number of founders.

To compare our results with the findings of Pong-Wong *et al*. (1998), their genetic model and population structure were simulated. When the data were analysed with FPMs that include constant homozygous difference and constant dominance deviation across loci as in their analysis, very similar results were obtained (data not shown). However, the variation in the estimates among replications is unacceptably large for such a small population. Moreover, the direction of dominance action might vary across loci (positive and negative) in the true genetic model. Therefore, it is not optimal to fit constant dominance deviation for analysing data with an unknown genetic model. The analysis of this small data set with non-additive FPMs with variable dominance deviations shows a more rapid increase in the estimate of dominance variance as the number of loci increases (data not shown), suggesting size and structure of the pedigree play a role in the overestimation.

Instead of using the kinship coefficients as in the infinitesimal model, the FPM approach uses the transmission of alleles at artificially created loci to model the resemblance among relatives. Similar to the traditional mixed linear model methodologies, the objective of FPM analysis is to estimate variance components and to predict individual BMs and DVs, rather than to identify the true genetic model of a trait. With pedigree and phenotypic data only, genetic effects of individual polygenes are not identifiable.

It is straightforward to modify additive FPMs to include non-additive components such as dominance and (or) epistasis. Moreover, the results from this study suggest that the FPM approach has the potential for improved estimation of non-additive genetic variance components due to the severe limitations of analyses under the infinitesimal model (see Section 1). However, the comparison of analyses under FPMs and infinitesimal genetic models for different pedigree structures and different genetic models is currently hampered by lack of reliable software for non-additive analysis under the infinitesimal model for general pedigrees with or without inbreeding.

Further investigations of the FPM for parameter estimation in non-additive genetic models are warranted and needed. A future contribution will consider epistasis. Alternative genotype sampling schemes should be investigated, e.g. sampling all individuals at a given locus jointly or sampling several loci jointly. In the present study we found no indication of insufficient mixing in genotype sampling. Furthermore, the effect of estimating gene frequencies at the biallelic loci, rather than fixing them at 0·5, on accuracy of estimation of genetic parameters and genetic merits has not yet been evaluated. Finally, the dependence of accuracy of parameter estimation on the number of loci in a FPM is of concern and requires further study.

Financial support provided by the National Science Foundation (DBI-9696265/9740519), the US Department of Agriculture (92-01732), the US Holstein Association and Infigen Inc. to I.H. is gratefully acknowledged. We thank Chris Stricker for assistance with the use of SALP.

References

De Boer, I. J. M. & Hoeschele, I. (1993). Genetic evaluation methods for populations with dominance and inbreeding. *Theoretical and Applied Genetics* **86**, 245–258.

Fernando, R. L., Stricker, C. & Elston, R. C. (1993). An efficient algorithm to compute the posterior genotypic distribution for every member of a pedigree without loops. *Theoretical and Applied Genetics* **87**, 89–93.

Fernando, R. L., Stricker, C. & Elston, R. C. (1994). The finite polygenic mixed model: an alternative formulation for the mixed model of inheritance. *Theoretical and Applied Genetics* **88**, 573–580.

Geyer, C. J. (1992). Practical Markov Chain Monte Carlo. *Statistical Science* **7**, 467–511.

Gillois, M. (1964). Calcul de coefficients d'identité d'aprés les liens de parenté. In: *Communication aux Journées d'études de la Commission de Génétique de la Fédération européenne de Zootechnie*. Lisbon.

Hagger, C. & Stricker, C. (1998). Infinitesimal and segregation analyses of a trait under selection. In: *Proceedings of the Sixth World Congress on Genetics Applied to Livestock Production*, Armidale **26**, 37–40.

Harris, D. L. (1964). Genotypic covariances between inbred relatives. *Genetics* **50**, 1319–1348.

Janss, L. L. G., Thompson, R. & VanArendonk, J. A. M. (1995). Application of Gibbs sampling for inference in a mixed major gene–polygenic inheritance model in animal populations. *Theoretical and Applied Genetics* **91**, 1137–1147.

Misztal, I. (1989). JAA: mixed model program using iteration on data with support for animal model, version 2.0. Computer program package, University of Georgia.

Nelder, J. A. & Mead, R. (1965). A simplex method for function minimization. *Computer Journal* **7**, 308–313.

Pong-Wong, R., Shaw, F. & Woolliams, J. A. (1998). Estimation of dominance variation using a finite-locus model. In: *Proceedings of the Sixth World Congress on Genetics Applied to Livestock Production*, Armidale **26**, 41–44.

Sorensen, D. A., Andersen, S., Gianola, D. & Korsgaard, I. (1995). Bayesian inference in threshold models using Gibbs sampling. *Genetics*, *Selection*, *Evolution* **27**, 229–249.

Stricker, C., Fernando, R. L. & Elston, R. C. (1995*a*). SALP: Segregation and Linkage Analysis for Pedigrees, release 2.0. Computer program package. Zurich, Switzerland: Federal Institute of Technology ETH, Institute of Animal Sciences.

Stricker, C., Fernando, R. L. & Elston, R. C. (1995*b*). Linkage analysis with an alternative formulation for the mixed model of inheritance: the finite polygenic mixed model. *Genetics* **141**, 1651–1656.

Stricker, C., Fernando, R. L. & Elston, R. C. (1995*c*). An algorithm to approximate the likelihood for pedigree data with loops by cutting. *Theoretical and Applied Genetics* **91**, 1054–1063.

Thompson, E. A. & Skolnick, M. H. (1977). Likelihoods on complex pedigrees for quantitative traits. In: *Proceedings of the International Conference on Quantitative Genetics*, (ed. E. Pollack, O. Kempthorne & T. B. Bailey, Jr), pp. 815–818. Ames, Iowa: Iowa State University Press.

Wang, C. S., Rutledge, J. J. & Gianola, D. (1993). Marginal inferences about variance components in a mixed linear model using Gibbs sampling. *Genetics*, *Selection*, *Evolution* **25**, 41–62.