Genet. Res., Camb. (1996), **68**, *pp.* 233–240

Mixed model approaches for diallel analysis based on a bio-model

JUN ZHU[1] AND BRUCE S. WEIR[2]*

[1] *Department of Agronomy, Zhejiang Agricultural University, Hangzhou, Zhejiang, China*
[2] *Program in Statistical Genetics, Department of Statistics, North Carolina State University, Raleigh, NC 27695-8203, USA*

(*Received 14 September 1993, and in revised form 9 February 1995 and 21 June 1996*)

Summary

A MINQUE(1) procedure, which is minimum norm quadratic unbiased estimation (MINQUE) method with 1 for all the prior values, is suggested for estimating variance and covariance components in a bio-model for diallel crosses. Unbiasedness and efficiency of estimation were compared for MINQUE(1), restricted maximum likelihood (REML) and MINQUE(θ) which has parameter values for the prior values. MINQUE(1) is almost as efficient as MINQUE(θ) for unbiased estimation of genetic variance and covariance components. The bio-model is efficient and robust for estimating variance and covariance components for maternal and paternal effects as well as for nuclear effects. A procedure of adjusted unbiased prediction (AUP) is proposed for predicting random genetic effects in the bio-model. The jack-knife procedure is suggested for estimation of sampling variances of estimated variance and covariance components and of predicted genetic effects. Worked examples are given for estimation of variance and covariance components and for prediction of genetic merits.

1. Introduction

Estimating genetic variance is of importance for quantitative genetic research as well as for plant and animal breeding. Estimation of genetic variance components is generally accomplished by the method of Cockerham (1963). A mating design is used to generate sets of relatives tested in one or more environments and an analysis of variance (ANOVA) can be constructed for estimating variance components which are then translated into covariances of relatives. These covariances of relatives can also be interpreted in terms of genetic and environmental components and hence estimators of genetic variance components can be derived. Among various mating designs, the nested (Design I) and factorial (Design II) designs (Comstock & Robinson, 1952; Hallauer & Miranda, 1981) and the diallel designs (Yates, 1947; Hayman, 1954; Griffing, 1956; Matzinger & Kempthorne, 1956; Gardner & Eberhart, 1966) are most used by plant and animal breeders. Cockerham & Weir (1977) proposed a bio-model for diallel crosses. This model partitions extranuclear effects into maternal effects and paternal effects, and is more representative of the biological situation. Since the ANOVA method cannot give separate estimates for variance components of maternal effects and paternal effects, mixed linear model approaches need to be employed for unbiased estimation of variance components.

In this study, Monte Carlo simulations are used to evaluate estimation methods of minimum norm quadratic unbiased estimation (MINQUE) (Rao, 1970, 1971) and restricted maximum likelihood (REML) (Patterson & Thompson, 1971; Corbeil & Searle, 1976). A MINQUE procedure is suggested for estimating covariance components between two traits with equal design matrices. Unbiasedness and efficiency for estimating variance components and covariance components are tested by Monte Carlo simulations for MINQUE and REML procedures. A method of adjusted unbiased prediction (AUP) for random genetic effects is compared with the best linear unbiased prediction (BLUP) procedure. A worked example is presented to illustrate the use of these new methods of analysis.

2. Model and methodology

Cockerham & Weir's (1977) bio-model of diallel crosses provides a way of estimating maternal and paternal variance components. If other higher-order interaction effects are not included in the model, the

* Corresponding author.

bio-model for an observation in the kth block of the cross between line i (maternal) and line j (paternal) can be expressed as

$$y_{ijk} = \mu + n_i + n_j + t_{ij} + m_i + p_j + b_k + \epsilon_{ijk}, \quad (1)$$

where y_{ijk} is the average phenotypic value of individuals from line $i \times$ line j in block k; μ is the population mean; n_i is the effect of nuclear contribution of maternal line $i, n_i \sim (0, \sigma_n^2)$; n_j is the effect of nuclear contribution of paternal line $j, n_j \sim (0, \sigma_n^2)$; t_{ij} is the interaction effect of nuclear contributions of lines $i \times j, t_{ij} \sim (0, \sigma_t^2)$; m_i is the extranuclear maternal effect of line $i, m_i \sim (0, \sigma_m^2)$; p_j is the extranuclear paternal effect of line $j, p_j \sim (0, \sigma_p^2)$; b_k is the effect of block, $k, b_k \sim (0, \sigma_b^2)$; ϵ_{ijk} is the residual effect, $\epsilon_{ijk} \sim (0, \sigma_e^2)$.

All the effects except the constant μ are independent random effects. Terms n and t are for nuclear genetic effects. They refer to effects associated with genes transmitted from parent to offspring. The extranuclear effects are not due to the individual's genotype but to the maternal (m) or paternal (p) influence. The assumptions for the bio-model are: (a) regular diploid segregation; (b) parents randomly sampled from a reference population; (c) no epistatic effects; and (d) no genotype by environment interaction. There is no correlation between nuclear effects and extranuclear effects in the bio-model. The maternal effects consist mainly of maternal genetic effects and/or cytoplasmic effects. The paternal effects are mainly paternal genetic effects. The maternal (or paternal) common environmental effects can be eliminated by special experimental designs, such as randomized complete block design. Reciprocal crosses are needed to separate genetically determined variation from maternally or paternally determined variation.

The bio-model of diallel crosses can be written in a matrix form of the mixed linear model,

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{U}_n \mathbf{e}_n + \mathbf{U}_t \mathbf{e}_t + \mathbf{U}_m \mathbf{e}_m + \mathbf{U}_p \mathbf{e}_p + \mathbf{u}_b \mathbf{e}_b + \mathbf{e}_\epsilon$$
$$= \mathbf{1}\mu + \sum_{u=1}^{5} \mathbf{U}_u \mathbf{e}_u + \mathbf{U}_6 \mathbf{e}_6.$$

The vector $\mathbf{y}$ is distributed with mean $\mathbf{1}\mu$ and variance $\mathbf{V} = \sum_{u=1}^{5} \sigma_u^2 \mathbf{U}_u \mathbf{U}_u' + \sigma_6^2 \mathbf{I}$. The constant μ is the population mean, and $\mathbf{U}_u$ is the known incidence matrix relating to the random vector $\mathbf{e}_u \sim (\mathbf{0}, \sigma_u^2 \mathbf{I})$.

The MINQUE method was proposed by Rao (1970, 1971) for estimating variance components. Variance components of the bio-model can be estimated by solving the following MINQUE equations for $u, v = 1, 2, \ldots 6$:

$$[\mathrm{tr}(\mathbf{U}_u' \mathbf{Q}_\alpha \mathbf{U}_v \mathbf{U}_v' \mathbf{Q}_\alpha \mathbf{U}_u)][\sigma_u^2] = [\mathbf{y}' \mathbf{Q}_\alpha \mathbf{U}_u \mathbf{U}_u' \mathbf{Q}_\alpha \mathbf{y}], \quad (2)$$

where the trace tr is the sum of diagonals of a matrix, and

$$\mathbf{Q}_\alpha = \mathbf{V}_\alpha^{-1} - \mathbf{V}_\alpha^{-1} \mathbf{1}(\mathbf{1}' \mathbf{V}_\alpha^{-1} \mathbf{1})^{-1} \mathbf{1}' \mathbf{V}_\alpha^{-1}$$

$$\mathbf{V}_\alpha = \sum_{u=1}^{6} \alpha_u \mathbf{U}_u \mathbf{U}_u'$$

and $\mathbf{V}_\alpha^{-1}$ is the inverse of $\mathbf{V}_\alpha$ with prior values α_u in place of σ_u^2 in $\mathbf{V}$.

Methods of estimating covariance components were proposed by Rao & Kleffe (1980) for the MINQUE procedure. Those procedures involve extensive computations and have been put to little use in practice. A much simpler MINQUE procedure for estimating covariance components can be derived for any number of traits for the genetic model. If two variables $\mathbf{y}_1$ and $\mathbf{y}_2$ have equal design matrices, they have covariance matrix $\mathbf{V}_{1/2} = \sum_{u=1}^{6} \sigma_{1_u/2_u} \mathbf{U}_u \mathbf{U}_u'$. The expectation of the quadratic function $\mathbf{y}_1' \mathbf{Q}_\alpha \mathbf{U}_u \mathbf{U}_u' \mathbf{Q}_\alpha \mathbf{y}_2$ is

$$\mathrm{tr}(\mathbf{Q}_\alpha \mathbf{U}_u \mathbf{U}_u' \mathbf{Q}_\alpha \mathbf{V}_{1/2}) = \sum_{v=1}^{6} \sigma_{1_u/2_u} \mathrm{tr}(\mathbf{U}_u' \mathbf{Q}_\alpha \mathbf{U}_v \mathbf{U}_v' \mathbf{Q}_\alpha \mathbf{U}_u).$$

By MINQUE theory, the invariant and unbiased estimators of covariance components can then be obtained by solving the following system of equations for $u, v = 1, 2, \ldots, 6$:

$$[\mathrm{tr}(\mathbf{U}_u' \mathbf{Q}_\alpha \mathbf{U}_v \mathbf{U}_v' \mathbf{Q}_\alpha \mathbf{U}_u)][\sigma_{1_u/2_u}] = [\mathbf{y}_1' \mathbf{Q}_\alpha \mathbf{U}_u \mathbf{U}_u' \mathbf{Q}_\alpha \mathbf{y}_2]. \quad (3)$$

Although the estimates of variances and covariances depend on prior values α, they are unbiased, provided that the choice of α does not depend on the data. Because α is a vector of known values, it may be chosen from prior experiments, from iterations or theoretical considerations. If the parameter values are known and used for $[\alpha_u]$ ($\alpha_u = \sigma_u^2$ or $\sigma_{1_u/2_u}$), this is the MINQUE(θ) which will give the minimum variance, invariant, unbiased estimators for linear functions of variance components under the normality assumption (Rao, 1972). If the user has no basis for selecting α, MINQUE(1) with $\alpha = 1$ was suggested by Giesbrecht (1985). If α_u are replaced by the iterated estimates with the restriction that they are within the parameter space, REML estimates can be obtained by iteration until converging.

By the best linear unbiased prediction (BLUP) method (Henderson, 1963), the uth vector of random genetic effects in the bio-model can be obtained by

$$\hat{\mathbf{e}}_{u(\theta)} = \sigma_u^2 \mathbf{U}_u' \mathbf{Q}_\theta \mathbf{y},$$

where $\mathbf{Q}_\theta$ is $\mathbf{Q}_\alpha$ with the prior values replaced by parameter values. The BLUP needs known variances. Since the true variances are unknown in practice, estimates are usually used in prediction. Then the genetic effects can be predicted as

$$\hat{\mathbf{e}}_{u(\theta)} = \hat{\sigma}_u^2 \mathbf{U}_u' \mathbf{Q}_{\hat{\theta}} \mathbf{y}.$$

Such prediction is just a so-called 'BLUP' since it has already lost the linearity and the guarantee for unbiasedness by using estimated variances. The unknown variances can also be replaced by prior values from prior experiment or from reasonable guesses. Therefore the genetic effects can be predicted by choosing prior values α as in the case of MINQUE method:

$$\hat{\mathbf{e}}_{u(\alpha)} = \alpha_u \mathbf{U}_u' \mathbf{Q}_\alpha \mathbf{y}.$$

https://doi.org/10.1017/S0016672300034200 Published online by Cambridge University Press

The linear unbiased prediction (LUP) can be accommodated by the estimated variance to give an adjusted unbiased prediction (AUP):

$$\hat{\mathbf{e}}^{A}_{u(\alpha)} = \kappa \hat{\mathbf{e}}_{u(\alpha)},$$

where $\kappa = \sqrt{[(df_u \hat{\sigma}^2_u)/(\hat{\mathbf{e}}'_{u(\alpha)} \hat{\mathbf{e}}_{u(\alpha)})]}$, with constraint $\hat{\sigma}^2_u \geqslant 0$, and df is the uth vector size minus one.

The jack-knife procedure (Miller, 1974; Efron, 1982) can be used for estimating sampling variances of estimated variance or covariance components and of predicted genetic effects. In this study genetic entries serve as jack-knifing units, leaving out one genotype for all replicates at a time. When jack-knife estimates and standard errors are obtained, the null hypotheses of zero parameter values can be tested by a t-test.

Monte Carlo simulations were performed in this study to evaluate the estimation of variances and covariances of the bio-model by MINQUE(θ), MINQUE(1) and REML. The unbiasedness and efficiency of prediction with BLUP, AUP and 'BLUP' methods were also compared by Monte Carlo simulations. Pseudo-random normal deviates with zero mean and unit variance were generated by the method of Kinderman & Monahan (1977). For each case 200 simulations were run to obtain sample means of estimates, bias and mean squared error (MSE). If the absolute value of bias is less than 10% of the parameter value, the parameter is said to be well estimated. In cases where the parameter value of a variance or covariance component is zero, bias $< 1\%$ of the sum of variances and covariances is considered to be negligible.

For simplicity, randomized complete block designs with three replications were used in this study. Ten parents were used for constructing an unbalanced diallel mating design with reciprocals by assuming that parents 1–6 could not mate with each other and crosses from parent 1 to parents 9 and 10 are missing. Parents were not included. The 56 genetic entries were assigned at random within each block. Since estimates of block variance and covariance are usually not of much concern for diallel analyses, simulation results of block effects are not presented in this paper.

3. Monte Carlo simulation results

(i) *Estimation of variance and covariance components*

Simulation results for bias and MSE are summarized in Table 1 for variance components and in Table 2 for covariance components. Variance and covariance of residual effects can always be efficiently estimated without bias by MINQUE(θ), MINQUE(1) and REML methods. But estimation of genetic variance and covariance components is not equally efficient for these three methods. Both MINQUE(θ) and MINQUE(1) can give unbiased estimates for variance components and covariance components no matter what values are set for parameters. There are no apparent differences of bias and MSE between these two MINQUE methods. Unbiased estimates are obtained by the REML method for variance and covariance components of nuclear effects. Variance and covariances for maternal and paternal effects tend to be slightly over-estimated by the REML method. REML may give a smaller MSE for σ^2_n, σ^2_m and σ^2_p but a larger MSE for σ^2_t. Since REML requires more computations due to iterations, there is no apparent advantage of REML over MINQUE(1). By the MINQUE(1) method, the bio-model is quite robust for estimating variance and covariance components even though there are no paternal and maternal effects.

When the procedure of jack-knifing over genotypes is used in MINQUE(1) estimation, unbiased estimates are obtained for both variance components and covariance components. The values of bias and MSE are very close to those of MINQUE(1) in Tables 1 and 2. Power values (the probabilities of rejecting the null hypotheses of no variation) are over 90% for testing significant variance of nuclear effects and residual effects. Power values are relatively low (around 50%) for detecting significant variances of maternal and paternal effects if they exist. If there are no paternal effects ($\sigma^2_p = 0$), correct conclusions of no paternal effects can be drawn with a probability over 90%. When there are no maternal and paternal effects ($\sigma^2_m = 0$ and $\sigma^2_p = 0$), non-significance can be detected with 99% probability for these two variance components. The power values of hypothesis tests for covariance components are relatively low compared with the variance estimation. It is indicated that more genetic entries are needed to detect significance of covariance components between two traits.

(ii) *Prediction of genetic effects*

Two prediction methods – 'BLUP' $\hat{\mathbf{e}}_{u(\hat{\theta})}$ using REML estimates and AUP $\hat{\mathbf{e}}^{A}_{u(1)}$ with $\alpha = \mathbf{1}$ – were compared with the BLUP $\hat{\mathbf{e}}_{u(\theta)}$ using parameter values. Two hundred simulation runs were conducted for estimating bias in predicted effects and distance between predictor vector $\hat{\mathbf{e}}_u$ and sampling vector $\tilde{\mathbf{e}}_u$. The distance is defined as $\|\hat{\mathbf{e}} - \tilde{\mathbf{e}}\| = \sqrt{[\sum_u (\hat{e}_u - \tilde{e}_u)]^2}$.

All these prediction methods will give extremely low bias for predicted mean genetic effects. Hence these predictors are unbiased for random genetic effects. The variances of predicted random genetic effects are always smaller than the true variances for both BLUP and 'BLUP' methods (Table 3). It is shown by Monte Carlo simulations that these two methods yield predictions with unbiased means but under-estimated variances for all the random effects. In plant and animal breeding, 'BLUP' is mostly used by breeders for evaluating breeding values of genetic materials.

Table 1. *Bias and MSE of variance components estimated by MINQUE(θ), MINQUE(1) and REML for unbalanced diallel crosses*

		MINQUE(θ)		MINQUE(1)		REML	
Parameter	Value	Bias	MSE	Bias	MSE	Bias	MSE
σ^2_n	50	0·78	1279	0·70	1267	1·00	1260
σ^2_t	30	0·86	158	0·88	159	0·85	161
σ^2_m	20	1·60	506	1·56	508	2·88[a]	449
σ^2_p	20	−1·89	362	−1·92	361	−0·51	286
σ^2_e	30	0·00	13	0·00	13	−0·01	13
σ^2_n	50	2·20	1105	2·13	1096	2·33	1105
σ^2_t	30	0·86	158	0·88	159	0·83	159
σ^2_m	20	0·24	315	0·15	319	0·61	304
σ^2_p	0	−0·01	184	−0·02	187	5·07[a]	96
σ^2_e	30	−0·00	13	0·01	13	−0·00	13
σ^2_n	50	2·37	878	2·33	872	2·53	877
σ^2_t	30	0·87	158	0·88	159	0·81	158
σ^2_m	0	0·16	21	0·30	32	2·22[a]	17
σ^2_p	0	−0·10	23	−0·24	31	2·07[a]	15
σ^2_e	30	0·01	13	0·01	13	−0·01	13

[a] Bias > 10% of the true value.
If the true value = 0, bias > 1% of the total variance value.

Table 2. *Bias and MSE of covariance components estimated by MINQUE(θ), MINQUE(1) and REML for unbalanced diallel crosses*

		MINQUE(θ)		MINQUE(1)		REML	
Parameter	Value	Bias	MSE	Bias	MSE	Bias	MSE
$\sigma_{n/n}$	25	−0·15	669	−0·15	668	1·11	563
$\sigma_{t/t}$	15	0·79	73	0·78	74	1·72	96
$\sigma_{m/m}$	10	0·26	239	0·24	240	2·33[a]	172
$\sigma_{p/p}$	10	−0·37	281	−0·35	282	2·06[a]	212
$\sigma_{e/e}$	15	0·01	10	0·01	10	0·01	10
$\sigma_{n/n}$	25	0·11	521	0·08	524	0·79	461
$\sigma_{t/t}$	15	0·76	73	0·77	73	1·42	104
$\sigma_{m/m}$	10	0·06	157	0·05	159	1·14[a]	128
$\sigma_{p/p}$	0	−0·27	98	−0·31	99	3·91[a]	55
$\sigma_{e/e}$	15	0·01	10	0·01	10	0·00	10
$\sigma_{n/n}$	25	0·34	490	0·38	491	1·36	429
$\sigma_{t/t}$	15	0·76	73	0·76	73	1·14	90
$\sigma_{m/m}$	0	0·44	12	0·47	16	1·68[a]	10
$\sigma_{p/p}$	0	−0·49	12	−0·57	15	1·17[a]	5
$\sigma_{e/e}$	15	0·00	9	0·01	10	−0·00	10

[a] Bias > 10% of the true value.
If the true value = 0, bias > 1% of the total variance value.

Under-estimating variances of 'BLUP' predictors mean that the absolute values of predicted genetic effects will be smaller than the real values. When adjusted by estimated variances, AUP will give predictors with unbiased means and variances for random genetic effects if they exist. Since adjustor κ needs the constraint of $\hat{\sigma}^2_u \geqslant 0$, variances of AUP predictors are slightly over-estimated if there are no random effects.

The BLUP $\hat{\mathbf{e}}_{u(\theta)}$ should give the smallest distance for the predicted genetic effects among all unbiased linear predictions (Henderson, 1979). The distances of 'BLUP' $\hat{\mathbf{e}}_{u(\hat{\theta})}$ and AUP $\hat{\mathbf{e}}^A_{u(1)}$ are a little larger than those of BLUP $\hat{\mathbf{e}}_{u(\theta)}$. The difference between distances of 'BLUP' $\hat{\mathbf{e}}_{u(\hat{\theta})}$ and AUP $\hat{\mathbf{e}}^A_{u(1)}$ is negligible in most cases. It is concluded that AUP can be used for predicting genetic effects. Since the distribution of AUP is unknown, standard errors of predictors can be obtained by the jack-knife procedure as in the case of estimating variances and covariances. A t-test can then be used for detecting significance of specific genetic effect.

https://doi.org/10.1017/S0016672300034200 Published online by Cambridge University Press

Table 3. *Prediction of genetic effects by BLUP, AUP and 'BLUP' for unbalanced diallel crosses*[a]

	BLUP $\hat{\mathbf{e}}_{u(\theta)}$		AUP $\hat{\mathbf{e}}^{A}_{(1)}$		'BLUP' $\hat{\mathbf{e}}_{u(\hat{\theta})}$	
Parameter[b]	Variance	Distance	Variance	Distance	Variance	Distance
$\sigma^2_m = \sigma^2_p = 20$						
n	36·7	12·6	50·9	13·4	39·9	13·9
t	18·7	18·3	30·8	19·4	19·6	18·8
m	10·7	10·2	22·7	12·2	16·4	11·6
p	10·2	10·0	19·4	11·6	13·2	11·4
$\sigma^2_m = 20, \sigma^2_p = 0$						
n	43·5	10·3	52·1	12·1	40·7	11·6
t	18·9	18·2	30·9	19·4	19·6	18·7
m	16·9	6·9	20·5	9·6	13·4	8·3
p	0·0	0·0	5·0	4·2	3·0	3·0
$\sigma^2_m = \sigma^2_p = 0$						
n	44·2	10·0	52·3	10·3	43·7	10·5
t	18·9	18·2	30·8	19·3	19·5	18·6
m	0·0	0·0	2·27	2·93	0·60	1·43
p	0·0	0·0	2·03	2·62	1·08	3·48

[a] Absolute bias for mean prediction of genetic effects is $10^{-5} \sim 10^{-7}$ for these three predictions.
[b] Parameter values are set to $\sigma^2_n = 50, \sigma^2_t = 30$ for nuclear effects.

4. Example

Data for *Nicotiana rustica* plants from Hayman (1954) were presented in appendix C of Cockerham & Weir (1977) and are also used as an example here. They are the mean flowering times of five plants per plot in two complete blocks. Variance components were estimated by the MINQUE(1) method, and their standard errors were estimated by the jack-knife procedure from cell means of eight-parent diallel crosses. The estimates of variance components and their standard errors are listed in Table 4. The estimates of variance components are very close to those obtained by the ANOVA method (Cockerham & Weir, 1977) although those authors did not give standard errors for their estimates. Significance for variance components is detected by the t-test for σ^2_n and σ^2_e. As an example, data from blocks I and II were used for estimating covariance components. The estimated genetic covariance components are very close to the estimates of genetic variance components (Table 4). There was strong nuclear additive correlation for observations between blocks I and II.

Table 4. *Jack-knife estimates of variance components by cell means and covariance components between block I and block II with the MINQUE(1) method for diallel crosses from Hayman's data*

Estimate	n	t	m	p	e
σ^2	5·106**	2·386	1·913	3·308	5·830*
S.E.$(\hat{\sigma}^2)$	(2·125)	(3·601)	(2·211)	(3·236)	(3·235)
$\hat{\sigma}_{1/2}$	4·855*	2·727	1·938	3·501	3·873
S.E.$(\hat{\sigma}_{1/2})$	(2·152)	(3·892)	(2·231)	(3·280)	(3·541)

* Significantly different from zero at 5% level.
** Significantly different from zero at 1% level.

Random genetic effects were predicted by the jack-knife procedure with the AUP method for these eight parents regarded as a sample from a reference population. Predictors $\hat{\mathbf{e}}^{A}_{u(1)}$ and their standard errors are listed in Table 5. Parental genetic merits can be evaluated mainly by predicted nuclear effects $\hat{n}_i$. Both parents 1 and 3 can be used in crosses for delaying flowering time, while parents 2, 7 and 8 may have effects on earliness. Although significant values of extranuclear variances are not observed, a significant positive maternal effect and a negative paternal effect are detected for parent 4. Parent 4 may have very little nuclear contribution but strong extranuclear effects for flowering time.

5. Discussion

Parents were not included in diallel analysis for the original bio-model. This is preferred for the estimation of genetic variance and covariance components, especially for cross-pollinated plants and domestic animals. For diallel mating of completely outbred parents, additive variance (or covariance) can be estimated by $V_A = 4\sigma^2_n$ (or $C_A = 4\sigma_{n/n}$) and dominance variance (or covariance) by $V_D = 4\sigma^2_t$ (or $C_D = 4\sigma_{t/t}$). Additive variance refers to the sum of the contributions of each allele separately to the variance, while dominance variance refers to the joint contributions of the two alleles at a locus.

Since breeding pure line cultivars is the major purpose of breeding programmes for self-pollinated plants, diallel analysis including inbred parents may give plant breeders a chance to evaluate the potential usefulness of the selected parents as breeding materials

https://doi.org/10.1017/S0016672300034200 Published online by Cambridge University Press

Table 5. *Predicted genetic effects* $\hat{\mathbf{e}}^{A}_{u(1)}$ *and standard errors (in parentheses) by jack-knife procedure for diallel crosses from Hayman's (1954) data*

Parent no.	Genetic effect $\hat{n}_i$	$\hat{t}_{i2}$	$\hat{t}_{i3}$	$\hat{t}_{i4}$	$\hat{t}_{i5}$	$\hat{t}_{i6}$	$\hat{t}_{i7}$	$\hat{t}_{i8}$	$\hat{m}_i$	$\hat{p}_i$
1	3·01**	−2·71*	9·04	−0·01	−1·13	−0·56	−0·13	−1·65	1·20	1·49
	(0·85)	(1·35)	(4·58)	(0·74)	(0·90)	(3·14)	(0·99)	(1·26)	(0·93)	(1·06)
2	−1·73*		0·13	0·72	−0·04	0·25	1·84	−1·85	−1·48	0·05
	(0·66)		(2·25)	(0·81)	(0·56)	(0·85)	(1·66)	(1·30)	(0·87)	(0·70)
3	3·75**			0·66	−2·95	3·75	−3·58	−3·44	−0·92	4·19
	(1·08)			(0·96)	(1·43)	(3·80)	(2·32)	(2·71)	(0·97)	(2·17)
4	0·31				−3·17*	1·06	−0·16	1·22	3·03*	−3·24*
	(0·58)				(1·44)	(1·03)	(1·36)	(1·89)	(1·44)	(0·89)
5	−0·48					0·33	1·25	5·25*	0·84	−1·41*
	(0·53)					(0·92)	(1·08)	(2·24)	(0·50)	(0·65)
6	0·28						−2·74	−1·72	1·24	−0·88
	(0·84)						(1·47)	(1·10)	(1·06)	(1·14)
7	−3·26**							0·35	−1·80	−0·96
	(0·79)							(0·83)	(1·09)	(0·79)
8	−1·89*								−2·11	0·76
	(0·74)								(1·21)	(0·81)

* Significantly different from zero at 5% level.
** Significantly different from zero at 1% level.

or directly as cultivars. Under the bio-model, we assume no epistasis (joint contributions from alleles at more than one locus). When inbred parents are included in the bio-model, covariances of relatives involving inbred parents consist of quadratic components D_1, D_2^* as well as an additive genetic variance component (Cockerham & Weir, 1984). With an additive and dominance model for only nuclear gene effects, the covariances of full sib families are

$$\begin{aligned}\mathrm{Cov}(y_{iik}, y_{iik'}) &= 4\sigma_n^2 + \sigma_t^2 \\ &= 2V_A + 4D_1 + D_2^*\end{aligned}$$

for inbred parents, and

$$\begin{aligned}\mathrm{Cov}(y_{ijk}, y_{ijk'}) &= 2\sigma_n^2 + \sigma_t^2 \\ &= V_A + V_D\end{aligned}$$

for F_1 organisms, where D_1 is the covariance of additive and homozygous dominance effects and D_2^* is the variance of homozygous dominance effects. The covariances of half-sib families are

$$\begin{aligned}\mathrm{Cov}(y_{iik}, y_{ijk'}) &= \sigma_n^2 \\ &= V_A + D_1\end{aligned}$$

for F_1 organisms and their inbred parents, and

$$\begin{aligned}\mathrm{Cov}(y_{ijk}, y_{ij'k'}) &= \sigma_n^2 \\ &= \tfrac{1}{2}V_A\end{aligned}$$

for F_1 organisms with one common parent. A fundamental assumption for the bio-model of diallel crosses is that all the genetic effects are independent random variables. This implies that nuclear additive and nuclear dominance effects are not correlated and D_1 is approximately zero. It is approximately true for

$$\begin{aligned}\mathrm{Cov}(y_{iik}, y_{iik'}) &= 4\sigma_n^2 + \sigma_t^2 \\ &\approx 2V_A + D_2^*\end{aligned}$$

and

$$\begin{aligned}\mathrm{Cov}(y_{iik}, y_{ijk'}) &= 4\sigma_n^2 \\ &\approx \tfrac{1}{2}V_A.\end{aligned}$$

For self-pollinated plants with no serious inbreeding depression, it may be acceptable to assume $t_{ij} \sim (0, \sigma_t^2)$ for $i \leqslant j$ and

$$\mathrm{Var}(t_{ii}) = \mathrm{Var}(t_{ij}) = \sigma_t^2.$$

Therefore the variance of homozygous dominance effects D_2^* is approximately equal to the dominance variance V_D. When diallel mating is conducted from completely inbred parents, direct genetic variances (or covariances) can be estimated by $V_A = 2\sigma_n^2$, and $V_D = \sigma_t^2$ (or $C_A = 2\sigma_{n/n}$, and $C_D = \sigma_{t/t}$) if parents are not included, or approximately by $V_A \approx 2\sigma_n^2$ and $V_D \approx \sigma_t^2$ (or $C_A \approx 2\sigma_{n/n}$, and $C_D \approx \sigma_{t/t}$) if parents are included. One advantage of including parents in the bio-model is that heterosis can be measured by the predicted interaction effects $\sum_i \hat{t}_{ii}$. A negative sum implies that positive heterosis may be important for the trait studied.

Griffing's diallel model (Griffing, 1956) is one of the most popular models for diallel analyses. Among four methods of Griffing's model, methods 1 and 3 are the diallel crosses which involve reciprocal F_1 organisms and so can give maternal and paternal effect estimates. Griffing's model for genetic entry from line $i \times$ line j is

$$y_{ij} = \mu + g_i + g_j + s_{ij} + r_{ij} + \text{error}.$$

The reciprocal effects r_{ij} can then be analysed as the extranuclear effects. The variances of general combining ability, σ_g^2, specific combing ability, σ_s^2, and reciprocal effects, σ_r^2, can be estimated by methods 1 and 3 of Griffing's model. The variance components

https://doi.org/10.1017/S0016672300034200 Published online by Cambridge University Press

in the bio-model for diallel crosses have an equivalence to these: $\sigma_n^2 \sim \sigma_g^2, \sigma_t^2 \sim \sigma_s^2, (\sigma_m^2 + \sigma_p^2)/2 \sim \sigma_r^2$. By Griffing's model, only an average variance for maternal and paternal effects can be estimated as the reciprocal variance σ_r^2. The causes of extranuclear effects are not distinguishable for maternal and paternal effects. When both maternal and paternal effects are present, the bio-model is superior to Griffing's model for unbiased estimation of variance and covariance components for nuclear, maternal and paternal effects.

In this study the unbalanced mating design has an experimental size of 168 (56 genotypes and three replications) which is the same size as an eight-parent balanced design. Monte Carlo simulations for the balanced design (simulation results are not presented in this paper) showed no considerable differences of bias, MSE and power value for estimating variance and covariance components between balanced and unbalanced mating designs. It is indicated that MINQUE(1) is equally efficient for estimating variance and covariance components for both balanced and unbalanced diallel crosses with the same experimental sizes. By using partial diallel crosses, the number of parents sampled at random from the reference population can be increased but not the experiment size.

In breeding practice, parental genetic merits are sometimes of more concern to the breeder. There is no way to estimate separately the genetic effects of the bio-model by any side conditions (Cockerham & Weir, 1977), although these random genetic effects are predictable by the BLUP procedure (Henderson, 1963). Parents as a random sample from a reference population can then be evaluated by predicted genetic merits. Random genetic effects have most often been predicted by using estimates of variance components. 'BLUP' predictors are unbiased for means ($\hat{\mathbf{l}}'\hat{\mathbf{e}}_{u(\hat{\theta})} \approx 0$) but under-estimated for variances ($\hat{\mathbf{e}}'_{u(\hat{\theta})}\hat{\mathbf{e}}_{u(\hat{\theta})} < df_u\hat{\sigma}_u^2$) (Searle *et al.*, 1992). We have shown that an adjusted unbiased prediction (AUP) by the MINQUE(1) method can give predictors with unbiased means and variances.

Standard errors of estimated variances and covariances or of predicted random effects in the bio-model can be obtained directly by some approximate formulae (Kackar & Harville, 1984; Searle *et al.*, 1992). In the present paper we suggest using the jack-knife method for estimating standard errors. Statistical tests for the hypothesis of no variation tend to be more powerful when using standard errors obtained by the jack-knife procedure (Zhu, 1989). Standard errors for functions of estimated variances and covariances (e.g. heritabilities and correlation coefficients) or of predicted random effects (e.g. heterosis) are also obtainable by the jack-knife procedure.

In summary, the bio-model with mixed model approaches allows the analysis of data from unbalanced designs and the separate estimation of variance and covariance components for maternal and paternal effects. We have evaluated three estimation procedures for the components of the bio-model and have given a methodology for covariance component estimation and for prediction of genetic effects. The methods of diallel analysis suggested in this paper are also applicable to balanced and unbalanced data of other diallel models with independent random genetic effects.

We thank two anonymous reviewers for useful comments and suggestions on the earlier version of the manuscript. This research was supported in part by NIH grant GM32518.

References

Cockerham, C. C. (1963). Estimation of genetic variances. In *Statistical Genetics and Plant Breeding* (ed. W. D. Hanson & H. F. Robinson), publication 982, pp. 53–94. Washington, DC: National Academy of Sciences/National Research Council.

Cockerham, C. C. & Weir, B. S. (1977). Quadratic analyses of reciprocal crosses. *Biometrics* **33**, 187–203.

Cockerham, C. C. & Weir, B. S. (1984). Covariances of relatives stemming from a population undergoing mixed self and random mating. *Biometrics* **40**, 157–164.

Comstock, R. E. & Robinson, H. F. (1952). Estimation of average dominance of genes. In *Heterosis* (ed. J. W. Gowan), pp. 494–516. Ames, Iowa: Iowa State University Press.

Corbeil, R. R. & Searle, S. R. (1976). Restricted maximum likelihood (REML) estimation of variance components in the mixed model. *Technometrics* **18**, 31–38.

Efron, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans*. Philadelphia: Society for Industrial and Applied Mathematics.

Gardner, C. O. & Eberhart, S. A. (1966). Analysis and interpretation of the variety cross diallel and related populations. *Biometrics* **22**, 439–452.

Giesbrecht, F. G. (1985). MIXMOD: a SAS procedure for analyzing mixed models. Mimeo series no. 1659. Raleigh, NC: Institute of Statistics, North Carolina State University.

Griffing, B. (1956). Concept of general and specific combining ability in relation to diallel crossing systems. *Australian Journal of Biological Sciences* **9**, 463–493.

Hallauer, A. R. & Miranda, J. B. (1981). *Quantitative Genetics in Maize Breeding*. Ames, Iowa: Iowa State University Press.

Hayman, B. I. (1954). The theory and analysis of diallel crosses. *Genetics* **39**, 789–809.

Henderson, C. R. (1963). Selection index and expected genetic advance. In *Statistical Genetics and Plant Breeding* (ed. W. D. Hanson & H. F. Robinson), publication 982, pp. 141–163. Washington, DC: National Academy of Sciences/National Research Council.

Henderson, C. R. (1979). Using estimates in predictions of breeding values under a selection model. In *Variance Components and Animal Breeding* (ed. L. D. Van Vleck & S. R. Searle), pp. 217–227. Ithaca, NY: Cornell University.

Kackar, R. N. & Harville, D. A. (1984). Approximations for standard errors of estimators of fixed and random effects in mixed linear models. *Journal of the American Statistical Association* **79**, 853–862.

Kinderman, A. J. & Monahan, J. F. (1977). Computer generation of random variables using the ratio of normal deviates. *Association for Computing Machinery Transactions on Mathematical Software* **3**, 257–260.

Matzinger, D. F. & Kempthorne, O. (1956). The modified diallel table with partial inbreeding and interactions with environment. *Genetics* **41**, 822–833.

Miller, R. G. (1974). The jackknife: a review. *Biometrika* **61**, 1–15.

Patterson, H. D. & Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika* **58**, 545–554.

Rao, C. R. (1970). Estimation of heteroscedastic variances in linear models. *Journal of the American Statistical Association* **65**, 161–172.

Rao, C. R. (1971). Estimation of variance and covariance components: MINQUE theory. *Journal of Multivariate Analysis* **1**, 257–275.

Rao, C. R. (1972). Estimation of variance and covariance components in linear models. *Journal of the American Statistical Association* **67**, 112–115.

Rao, C. R. & Kleffe, J. (1980). Estimation of variance components. In *Handbook of Statistics*, vol. 1 (ed. P. R. Krishnaiah), pp. 1–40. New York: North-Holland.

Searle, S. R., Casella, G. & McCulloch, C. E. (1992). *Variance Components*. New York: Wiley.

Yates, F. (1947). Analysis of data from all possible reciprocal crosses between a set of parental lines. *Heredity* **1**, 287–301.

Zhu, J. (1989). Estimation of genetic variance components in the general mixed model. PhD dissertation, North Carolina State University, Raleigh, NC.