

AN APPROACH OF ELIMINATING SYSTEMATIC ERRORS IN GENETIC ANALYSIS

JIXIANG WU, HAIMING XU and JUN ZHU

*Agronomy Department, Zhejiang Agricultural University,
Hangzhou 310029, China*

New general genetic models including row and column effects and the corresponding analysis approaches were proposed. The results of Monte Carlo simulations indicated that the variance components could be estimated unbiasedly by the models containing row and column effects under both cases with and without systematic errors. A worked example will be given for eliminating systematic errors and estimating the genetic variance components.

1 Introduction

One of the primary aims of most agricultural field experiments is to obtain an unbiased and efficient estimation. To achieve this aim, it is important to control field variation that is possibly due to experimental management, fertility trends and other environmental factors. In general, blocking is employed to control field variation by arranging plots in appropriate ways. These methods include incomplete block design which were widely used².

When a large number of varieties are to be compared in a field trial, a latic square containing the whole plots is impossible because of requirement of too many replications. Recently, there has been an increased interest in "neighbor" or "spatial" methods for analyzing field experiments, where an attempt is made to estimate and remove the effects of the association of neighboring plots from the treatment contrasts. Examples of this work include^{1,4-6,10}. These methods were based on regular variation in the field. However, there were some limitations in use of these methods under general assumption.

The consideration of the present research was that the genetic materials could be grown in random or in sequence, and the same check set regularly. A general genetic model including row and column effects and the corresponding analysis methods were proposed. Monte Carlo simulations were conducted to confirm the unbiasedness of the estimates obtained by the model with and without row and column effects. For illustration, a worked example is presented that has been analyzed following AD model including systematic error.

2. General Genetic Models

If there does not exist the genotype by environment interaction, then the phenotypic performance of k th genotype in a plot can be expressed as the following linear model:

$$y = \mu + R + C + G + e, \quad (1)$$

With genetic experiments in multiple environments, the average phenotypic performance of genetic population in one environment can be expressed by the following,

$$y = \mu + E + R + C + G + GE + e, \quad (2)$$

where μ = population mean, fixed effect, E = environment effect, fixed effect, R = row effect within environment, fixed effect, C = column effect within environment, fixed effect, G = total genotype effect, random effect, GE = genotype \times environment interaction effect, random effect, and e = random error effect.

Cockerham's³ general genetic model could be extended by including genetic components of G and GE interaction¹².

If these models do not include systematic effects, the row and column effects in model (1) or model (2) can be dropped.

3. Analysis Methodology

The general genetic model can be expressed by a matrix form of mixed linear model:

$$y = Xb + \sum_{u=1}^m U_u e_u = Xb + Ue,$$

where, y is a $(n \times 1)$ vector of phenotypic mean for all entries; b is the vector of fixed effects, X is the known incidence matrix with coefficients 1 or 0 related with the fixed effects; e_u is the vector of random effects, $e_u \sim (0, \sigma_u^2 I)$; U_u is the known coefficient matrix related with to the random vector e_u .

If random effects are not correlated, the random vector y has a multivariate distribution with mean Xb and a variance-covariance matrix V

$$y \sim (Xb, V = \sum_{u=1}^m \sigma_u^2 U_u U_u^T)$$

where U_u^T is the transpose of U_u , $U_m = I$ is an identity matrix.

Variance components in the general model could be estimated by mixed linear model approaches such as restricted maximum likelihood (REML) method⁸

Table 1: The simulation results of two models without row and column effects

Parameter	True value	Model(1)			Model(2)		
		Bias	CE	Power	Bias	CE	Power
V_G	50	0.51	0.30	0.99	0.70	0.26	1.00
V_e	10	-0.07	0.26	0.97	-0.08	0.21	1.00
V_G	50	1.12	0.36	0.88	1.04	0.29	0.99
V_e	25	-0.25	0.25	0.98	-0.23	0.19	1.00
V_G	50	1.80	0.45	0.56	1.60	0.36	0.85
V_e	50	-0.64	0.24	0.99	-0.49	0.19	1.00

Note: Model(1) is the model containing row and column effects, Model(2) is the model not containing row and column effects.

or minimum norm quadratic unbiased estimation (MINQUE)⁹. Prediction of genetic effects in the mixed linear model could be obtained by linear unbiased prediction (LUP)¹³ or adjusted unbiased prediction (AUP)¹¹.

4. Simulation Results

Monte Carlo simulations were performed for an experimental design with 5 rows and 22 columns with two replications by the general model and adjusted model. Checks were assigned for every 5 entries. The unbiasedness and efficiency were compared between two cases with and without systematic errors. For each case, 500 simulations were run to obtain sample means of estimates $\hat{\sigma}^2$, bias = $\hat{\sigma}^2 - \sigma^2$, mean square error ($MSE = Bias^2 + var(\hat{\sigma}^2)$)¹¹, coefficient of efficiency $CE = \frac{\sqrt{MSE}}{|\theta| + |Bias|}$ and power value.

Simulation results for bias, CE and power value were summarized in Tables 1 and 2 for different variance components. The estimates of genotypic and residual variances were unbiased ($\frac{|Bias|}{\sigma} < 0.05$) for both models for the case without systematic errors. The power value of V_G for the models containing row and column effects was slightly less than that for the reduced model. When systematic errors do exist, the estimates for two variance components were identical as those for the case without systematic errors for the adjusted model. However, two variance components were overestimated by the model not containing row and column effects. Therefore the adjusted models and their analysis methods should be suggested for estimating the genetic variance components if there exist the systematic errors or if we do not know whether systematic errors exist or not.

Table 2: The simulation results of two models with row and column effects

Parameter	True value	Model(1)			Model(2)		
		Bias	CE	Power	Bias	CE	Power
V_G	50	0.51	0.30	0.99	3.77	0.31	0.99
V_e	10	-0.07	0.26	0.97	23.72	0.72	1.00
V_G	50	1.12	0.36	0.88	4.12	0.35	0.91
V_e	25	-0.25	0.25	0.98	23.49	0.51	1.00
V_G	50	1.80	0.45	0.56	4.94	0.41	0.70
V_e	50	-0.64	0.24	0.99	23.13	0.36	1.00

Note: Model(1) is the model containing row and column effects, Model(2) is the model not containing row and column effects.

5 Worked Example

The data of the 10 parents of the Upland cotton and their 20 F_1 s and F_2 s planted with three replications in 1992 and 1993 were analyzed with the additive-dominance model (AD model). The AD model and its GE interaction which included row and column effect could be expressed by the following matrix form:

$$\begin{aligned}
 \mathbf{y} &= \mathbf{1}\mu + \mathbf{X}_E\mathbf{b}_E + \mathbf{X}_R\mathbf{b}_R + \mathbf{X}_C\mathbf{b}_C \\
 &\quad + \mathbf{U}_{AE}\mathbf{e}_A + \mathbf{U}_{DE}\mathbf{e}_D + \mathbf{U}_{AE}\mathbf{e}_{AE} + \mathbf{U}_{DE}\mathbf{e}_{DE} + \mathbf{U}_{BE}\mathbf{b}_B + \mathbf{e} \quad (3) \\
 &= \mathbf{X}\mathbf{b} + \sum_{u=1}^6 \mathbf{U}_u\mathbf{e}_u.
 \end{aligned}$$

The row and column effects could be deleted in model (3) if the systematic errors were not included

The variance components for lint yield and lint percentage were estimated for these two models. The estimates and their standard errors could be estimated by removing one block each time with the jackknife techniques (Miller, 1974).

The results in Table 3 indicated that the estimates of genetic variance components for lint yield and lint percentage were different for the two models. The proportion of residual variance for lint yield in full model decreased by about 13% and that for lint percentage decreased by about 5%. This implied that the variance components were estimated biasedly for the model not containing row and column effects if the trait was largely influenced by systematic errors. The results also suggested that the systematic errors could be eliminated by the method presented in this paper if the treatments were arranged randomly

Table 3: Estimates of variance components for lint yield and lint percentage of Upland cotton by reduced model and full model.

Var. Comp.	Lint yield		Lint percentage	
	Model (a)	Model (b)	Model (a)	Model (b)
V_A	1.302+	28.934**	5.252**	5.074**
V_D	83.634*	28.738**	0.579*	0.253+
V_{AE}	0.840*	21.295**	0.263*	0.474+
V_{DE}	72.367*	13.404+	0.527*	0.647**
V_e	119.280**	40.624**	2.175**	1.553**

Note: +, *, **: significance at 0.10, 0.05 and 0.01 level, respectively. Model (a), (b): AD model not containing row and column effects, containing row and column effects, respectively.

with several replications.

Acknowledgments

This research was supported in part by the China Natural Science Foundation 39470377.

References

1. Besag J. and R. Kempton R., *Biometrics* **42**, 231 (1986).
2. Cochran W.G. and G.M. Cox., *Experimental Designs*, John Wiley and Sons(1957).
3. Cockerham C.C., *Theor. Appl. Genet.* **56**, 119 (1980).
4. Cullis B.R. and A.C. Gleeson, *Biometrics* **47**, 1449 (1991).
5. Federer W.T. and K.E. Basford, *Biometrics* **47**, 1461 (1991).
6. Gleeson A.C. and B.R. Cullis, *Biometrics* **43**, 277 (1987).
7. Miller R.G., *Biometrika* **61**, 1 (1974).
8. Patterson H.D. and R. Thompson, *Biometrika* **58**, 545 (1971).
9. Rao C.R., *Journal of Multivariate Analysis* **1**, 257 (1971).
10. William E.R., *Biometrika* **73**(2), 279 (1986).
11. Zhu J., *Journal of Biomathematics* **8**(1), 32 (1993).
12. Zhu J., *Journal of Zhejiang Agricultural University* **20**(6), 551 (1994).
13. Zhu J. and Weir B.S., *Theor. Appl. Genet.* **89**, 153 (1994).