Chapter 16

Mixed Linear Model Approaches for Quantitative Genetic Models

Jixiang Wu
Jun Zhu
Johnie N. Jenkins

Purpose

Computer software for estimating variance and covariance components, correlations, and predicting genetic effects.

Software Description

We describe a suite of genetic software that employs mixed linear model approaches. The various components relate to three categories, viz, genetic models for diallel crosses (Table 16.1), seed traits (Table 16.2), and developmental traits (Table 16.3). It can also be used to analyze regional agronomic trials.

This software has several features:

1. Handles complicated genetic models for agronomic traits, seed traits, and developmental traits
2. Analyzes unbalanced data
3. Utilizes jackknifing techniques to test the significance of each genetic parameter
4. Provides some important references containing results
5. Fast computation

System Requirements

Hardware: PC 486 or above; 16MB RAM, or more; 10 MB or more available hard disk space

Operating System: Microsoft Windows 95/98, Microsoft Windows NT 3.5 or above.

Installing

The software suite is available upon request to the author or from the Web site: <http://msa.ars.usda.gov/ms/msstate/csrl/jenkins.htm>.

All related files are compressed into RUNWIN32.EXE, a self-extracting file. To install the software,

1. Copy RUNWIN32.ZIP to a subdirectory (e.g. *c:\WIN32*) on hard disk using Windows Explorer or Windows NT Explorer.
2. Double click RUNWIN32.ZIP and all files will be extracted into the current subdirectory WinZip 6.3 extraction software.

Tasks Performed by the Software

This software performs analyses on agronomic traits for diallel cross models; seed models; developmental models; and regional trials.

A. Programs for Diallel Cross Models

Table 16.1 shows programs for diallel crosses.

This software can be used to estimate genetic variance components and genetic covariance components and to predict genetic effects and heterosis for AD, ADM, and ADAA models.

B. Programs for Seed Models

Table 16.2 includes programs for seed models. These programs can also be used to estimate genetic variance components, genetic covariance components, and to predict genetic effects of diploid seed and triploid seed models.

TABLE 16.1. Diallel Crosses

Genetic Models	Jackknife by cell	Jackknife by block
AD[1]	GENAD	GENAD
	GENVAR1C	GENVAR1R
	GENCOV1C	GENCOV1R
	GENHET1C	GENHET1R
ADM[2]	GENADM	GENADM
	GENVAR1C	GENVAR1R
ADAA[3]	GENADAA	GENADAA
	GENVAR1C	GENVAR1R
	GENCOV1C	GENCOV1R

[1]Additive-dominance models
[2]Additive-dominance maternal models
[3]Additive-dominance additive × additive epistasis models

TABLE 16.2. Seed Models

Genetic models	Jackknife by cell	Jackknife by block
Diploid	GENDIPLD	GENDIPLD
	GENVAR0C	GENVAR0R
	GENCOV0C	GENCOV0R
	GENHET0C	GENHET0R
Triploid	GENTRIPL	GENTRIPL
	GENVAR0C	GENVAR0R
	GENCOV0C	GENCOV0R
	GENHET0C	GENHET0R

C. Programs for Developmental Genetic Models

Table 16.3 includes programs for developmental traits. These programs can be used to create conditional data files, to estimate conditional genetic variance components, and to predict conditional genetic effects for AD, ADM, ADAA for diploid or triploid seed models. Some programs have appeared in Tables 16.1 and 16.2.

TABLE 16.3. Developmental Traits

Genetic Models	Jackknife by cells	Jackknife by block
AD	GENAD	GENAD
	GENCOND1	GENCOND1
	GENVAR1C	GENVAR1R
ADM	GENADM	GENADM
	GENCOND1	GENCOND1
	GENVAR1C	GENVAR1R
ADAA	GENADAA	GENADAA
	GENCOND1	GENCOND1
	GENVAR1C	GENVAR1R
Diploid	GENDIPLD	GENDIPLD
	GENCOND0	GENCOND0
	GANVAR0C	GENVAR0R
Triploid	GENTRIPL	GENTRIPL
	GENCOND0	GENCOND0
	GENVAR0C	GENVAR0R

Use of the Software Package

A. Diallel Model Analysis

Step 1. Build a data file. The arrangement of data is shown in file dial1.txt. on the Web page. The first five columns in this file represent environment (e.g., year or location), female, male, generation, and block. In column 1, enter environment number (1...e); in column 2 and 3, enter female and male number, respectively (1...p); and in column 5 enter block number (1...b). The data identifiers should be consecutive positive integers, each beginning with 1. The generation codes for column 4 are 0 for parents, 1 for F_1, and 2 for F_2. Enter data in columns 6 to *n*.

Step 2. Create an information matrix based on genetic models. Run GENAD for AD model, GENADM for ADM model, and GENADE for ADAA model. For example, when GENAD is chosen, prompts will automatically appear on screen as follows:

Input name of your data file: (e.g., dial 1.txt)
Do you have block effects within location (or environment)? Y/N

When running GENADM, you will see an extra prompt:

Do you analyze triploid endosperm? Y/N

Two files will be automatically created, e.g., *dial1.dat, dial1.mat,* where, *dial1.mat* contains matrix information, and *dial1.dat* contains the data of traits to be analyzed.

Step 3. Estimate the variance components and predict genetic effects. For example, when GENVAR1C or GENVAR1R is selected on the screen you will see the following prompts:

Input name of your data file: (input the data file name as given in Step 2).
What kind of parents did you use? Input 1 for inbred or 0 for outbred.
Choose prediction method. Do you want to use LUP or AUP? Input L for LUP and O for AUP.
Input coefficients for each parent: 1 for first group, –1 for second group, 0 for others.
Input sampling number for the jackknife procedure if running GENVAR1C.

The results are automatically stored in a file named, for example, *dial1.var.*

Step 4. Estimate covariance components and correlation coefficients. After finishing step 3, you can run GENCOV1C or GENCOV1R, and you will see the following prompts:

Input name of your data file: (the name given in Step 2);
What kind of parents did you use? Input 1 for inbred or 0 for outbred;
Input sampling number for the jackknife procedure: if running GENCOV1C.

The results are automatically stored in a file named, for example, *dial1.cor.*

Step 5. Predict heterosis. After running step 2, you can run GENHET1C or GENHET1R. Follow the prompts that automatically appear on the screen after you have chosen to run either model:

Input name of your data file: (input the name used in Step 2);
What kind of parents did you use? Input 1 for inbred or 0 for outbred;

Input sampling number for the jackknife procedure: if running GENHET1C.

The results are automatically stored in a file named, for example, *dial1.pre.* Note: During the process, other temporary files such as *matrix.var, matrix.uq2, matrix.uq3, matrix.uq4, matrix.uq5,* or *matdjc.var* will be created. The user should delete these files after finishing all analyses.

B. Seed Model Analysis

Step 1. Build a data file. The arrangement of data is shown in file *ctseed.txt.* The first five columns in this data file represent environment (e.g., year, location), female, male, generation, and block. In column 1, enter environment number (1 . . . e), in column 2, female number, in column 3, male number (1 . . . p), and in column 5, block number (1. . . b). The data identifiers should be consecutive integers, each beginning with 1. The generation codes for column 4 are 0 for parent, 1 for F_1, 2 for F_2, 3 for $BC1 = (F_1 \times P_1)$, 4 for $BC2 = (F_1 \times P_2)$, 5 for $RBC1 = (P_1 \times F_1)$, and 6 for $RBC2 = (P_2 \times P_2)$. Enter data in columns 6 to *n.*

Step 2. Construct an information matrix based on genetic models; GENDIPLD for diploid seed model and GENTRIPL for triploid seed model. When GENDIPLD is run, the following prompts appear on the screen:

Input name of your data file: (for example, enter ctseed.txt)
Do you have block effect within location? Y/N

Note: Two files will be automatically created, *ctseed.dat, ctseed.mat,* where *ctseed.mat* contains matrix information, and *ctseed.dat* contains data on traits to be analyzed.

Step 3. Estimate variance components and predict genetic effects. For example, for GENVAR0C or GENVAR0R, the following prompts will appear on the screen:

Input name of your data file: (name given in Step 2)
Choose prediction method. Do you want to use LUP or AUP? For LUP, input L, for AUP input O.
Input coefficients for each parent: 1 for first group, –1 for second group, and 0 for others;

Input sampling number for the jackknife procedure: if running GENVAR0C.

The results are automatically stored in the *ctseed.var* file.

Step 4. Estimate covariance components and correlation coefficients. For example, run GENCOV0C or GENCOV0R. When running this program, on-screen prompts include:

Input name of your data file: (name given in Step 2)
Input sampling number for the jackknife procedure: if running GENCOV0C.

The results are automatically stored in the *ctseed.cov* file. Note: The user should delete temporary files created during the execution of the program, after finishing all analyses.

C. Developmental Genetic Model Analysis

Step 1. Construct the file. The file format is the same as for the diallel and seed models.

Step 2. Convert traits to conditional traits.

(a) Construct information matrix based on genetic models.

For example, for AD model run GENAD and follow the on-screen prompts.

Input name of your data file: filename.txt
Do you have blocks within location? Y/N

(b) Run GENCOND1 or GENCOND0, where GENCOND1 is for AD, ADM, and ADAA models; and GENCOND0 is for diploid and triploid seed models.

Step 3. Now run steps 2 through 5 from A (diallel models) or B (seed models), the only difference being the change in the name of input file from *filename.txt* to *filename.doc* (the latter is a conditional data file).

D. Crop Regional Trial Analysis

Software included: GENTEST, GENETESTM, and GENTESTW. These programs can be used to estimate variance components, compare

the significance of differences among varieties, and to evaluate the stability of each variety.

Step 1. Build the data file. The arrangement of data is shown in the file *msbean.txt*. The four columns represent variety, year, location, and replication. In the first column, enter variety number (check variety should be the highest number; this is important if you choose to transform data relative to the check). In the second column, enter year number. In the third column, location number; and in the fourth column enter replication number. The data identifiers should be consecutive positive integers beginning with 1.

Step 2. Construct an information matrix based on chosen genetic models.

For example, run GENTEST and follow these on-screen prompts:
Input name of your data file:

Step 3. Estimate stability for a single trait.

For example, run GENTESTM and follow the on-screen prompts:
Input name of your data file: (from Step 1).
Do you want to transform data relative to check genotype? Y/N
How many linear contrasts do you want?
Input coefficients for each variety: 1 for first group, –1 for second group, 0 for others.

The results are automatically stored in the *region.var* file. The results include variance components, linear contrasts among different genotypes, and stability of each genotype for each trait.

Step 4. Estimate stability for multiple traits.

For example, run GENTESTW and follow on-screen prompts:
Input name of your data file: (from Step 1).
Input weight or values for each trait (sum of these weights = 1.0).
How many linear contrasts do you want?
Input coefficients for each variety: 1 for first group, –1 for second group, 0 for others.

The results are automatically stored in the *region.cov* file. These results include variance and covariance components and stability of each genotype for multiple traits.

The following references may help the reader to understand the use of software packages and Internet sites.

Atchley, W.R. and Zhu, J. (1997). Developmental quantitative genetics, conditional epigenetic variability and growth in mice. *Genetics* 147:765-776.

Cockerham, C.C. (1980). Random and fixed effects in plant genetics. *Theoretical and Applied Genetics* 56:119-131.

Cockerham, C.C. and Weir, B.S. (1977). Quadratic analysis of reciprocal crosses. *Biometrics* 33:187-203.

Eisen, E.J., Bohren, B.B., and McKean, H.E. (1966). Sex-linked and maternal effects in the diallel cross. *Australian Journal of Biological Science* 19:1061-1071.

Fisher, R.A. (1925). *Statistical Methods for Research Workers,* First Edition. Oliver & Boyd, Edinburgh and London.

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

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

Hartley, H.D. and Rao, J.N.K. (1967). Maximum-likelihood estimation for the mixed analysis of variance model. *Biometrika* 54:93-108.

Henderson, C.R. (1963). Selection index and expected genetic advance. In Hanson, W.D. and Robinson, H.F. (Eds.), *Statistical Genetics and Plant Breeding* (pp. 141-163). Washington, DC: National Academy of Science, National Research Council.

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

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

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

Rao, C.R. and Kleffe, J. (1980). Estimation of variance components. In Krishnaiah, P.R. (Ed.), *Handbook of Statistics,* Vol. 1 (pp. 1-40). North-Holland, New York.

Searle, S.R., Casella, G., and McCulloch, C.E. (1992). *Variance Components.* John Wiley and Sons, New York.

Shi, C.H., Zhu, J., Zeng, R.C., and Chen, G.L. (1997). Genetic and heterosis analysis for cooking quality traits of indica rice in different environments. *Theoretical and Applied Genetics* 95:294-300.

Wu, J.X., Zhu, J., Ji, D.F., and Xu, F.H. (1995). Genetic analysis for heterosis of fiber traits in Upland cotton (Chinese). *Acta Gossypii Sinica* 7(4):217-222.

Yan, J.Q., Zhu, J., He, C.X., Benmoussa, M., and Wu, P. (1998). Molecular dissection of developmental behavior of plant height in rice (*Oryza sativa* L.). *Genetics* 150:1257-1265.

Yan, X.F., Xu, S.Y., Xu, Y.H., and Zhu, J. (1998). Genetic investigation of contributions of embryo and endosperm genes to malt kolbach index, alpha-amylase activity and wort nitrogen content in barley. *Theoretical and Applied Genetics* 96(5):709-715.

Zeng, Z.-B. (1994). Precision mapping of quantitative trait loci. *Genetics* 136:1457-1468.

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

Zhu, J. (1993a). Methods of predicting genotype value and heterosis for offspring of hybrids (Chinese). *Journal of Biomathematics* 8(1):32-44.

Zhu, J. (1993b). Mixed model approaches for estimating covariances between two traits with unequal design matrices (Chinese). *Journal of Biomathematics* 8(3):24-30.

Zhu, J. (1994). General genetic models and new analysis methods for quantitative traits (Chinese). *Journal of Zhejiang Agricultural University* 20(6):551-559.

Zhu, J. (1996). Analysis methods for seed models with genotype × environment interactions (Chinese). *Acta Genetica Sinica* 23(1):56-68.

Zhu, J. (1998). Mixed model approaches of mapping genes for complex quantitative traits. In Wang, L.Z. and Dai, J.R. (Eds.), *Proceedings of Genetics and Crop Breeding of China* (pp. 19-20). Chinese Agricultural Science and Technology Publication House, Beijing.

Zhu, J., Wang, G.J., and Zhang, R.C. (1997). Genetic analysis on gene effects and GE interaction effects for kernel nutrient quality traits of Upland cotton (Chinese). *Journal of Biomathematics* 12(2):111-120.

Zhu, J. and Weir, B.S. (1994a). Analysis of cytoplasmic and maternal effects: I. A genetic model for diploid plant seeds and animals. *Theoretical and Applied Genetics* 89:153-159.

Zhu, J. and Weir, B.S. (1994b). Analysis of cytoplasmic and maternal effects: II. Genetic models for triploid endosperm. *Theoretical and Applied Genetics* 89:160-166.

Zhu, J. and Weir, B.S. (1996). Diallel analysis for sex-linked and maternal effects. *Theoretical and Applied Genetics* 92(1):1-9.