

Kansas State University Libraries

New Prairie Press

Conference on Applied Statistics in Agriculture

1999 - 11th Annual Conference Proceedings

ANALYSIS OF GENOTYPE-BY-ENVIRONMENT INTERACTION WITH AMMI MODELS USING SAS PROC MIXED

Raul E. Macchiavelli

James S. Beaver

Follow this and additional works at: <https://newprairiepress.org/agstatconference>



Part of the [Agriculture Commons](#), and the [Applied Statistics Commons](#)



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License](#).

Recommended Citation

Macchiavelli, Raul E. and Beaver, James S. (1999). "ANALYSIS OF GENOTYPE-BY-ENVIRONMENT INTERACTION WITH AMMI MODELS USING SAS PROC MIXED," *Conference on Applied Statistics in Agriculture*. <https://doi.org/10.4148/2475-7772.1268>

This is brought to you for free and open access by the Conferences at New Prairie Press. It has been accepted for inclusion in Conference on Applied Statistics in Agriculture by an authorized administrator of New Prairie Press. For more information, please contact cads@k-state.edu.

ANALYSIS OF GENOTYPE-BY-ENVIRONMENT INTERACTION WITH AMMI MODELS USING SAS PROC MIXED

Raúl E. Macchiavelli and James S. Beaver

Dept. of Agronomy and Soils, Univ. of Puerto Rico, P.O. Box 9030, Mayagüez, PR 00681-9030

ABSTRACT

Genotype-by-environment (GE) interaction can be analyzed using different approaches. Among these, the additive main effects and multiplicative interaction model yields useful interpretations and can be applied successfully to plant breeding programs. In this paper we review fitting strategies for this model and show how to combine the capabilities of the Mixed and IML procedures in SAS to fit this model. This permits straightforward use of likelihood-based inference in standard and non standard situations like complex experimental designs. The proposed procedures were applied to data from red mottled bean variety trials conducted in the Dominican Republic and Puerto Rico in 9 environments with 30 lines (15 with indeterminate and 15 with determinate growth habit).

1. INTRODUCTION

Yield trials are usually repeated in different environments (typically a combination of locations and years). This permits general conclusions to be drawn, but also makes the interpretation of the interaction between genotype and environment (GE) interaction more difficult. Understanding this interaction is one of the aspects on which a successful plant breeding program depends (Gauch, 1988).

There are different approaches used to model GE interaction, among them the concurrence model, the regression on the mean models and the additive main effects and multiplicative interaction (AMMI) models (van Eeuwijk and Kroonenberg, 1998). Because it is the most versatile approach, the AMMI model has been used in a variety of situations (see for example Shafii *et al.*, 1992, and Aguiluz, 1998).

In this paper we review this model and show how to fit it using the Mixed and IML procedures in SAS. We apply the proposed methodology to red mottled bean (*Phaseolus vulgaris* L.) variety trials conducted in the Dominican Republic and Puerto Rico. In the Dominican Republic and Haiti, red mottled beans are grown by small-scale farmers both on hillsides and in valleys using diverse cultural practices. Hence, it is desirable to develop red mottled varieties which perform predictably over a wide range of environmental conditions. In this study 30 different lines were analyzed: 15 with indeterminate and 15 with determinate growth habit under 9 different environments (Beaver *et al.*, 1996).

2. METHODOLOGY

The general model used for a typical variety trial is:

$$Y_{ijk} = \mu + \alpha_i + \delta_{k(i)} + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk} \quad (1)$$

where μ , α_i , β_j , and $\alpha\beta_{ij}$ ($i=1, \dots, e; j=1, \dots, g$) are, respectively, the (fixed) general mean and the (fixed) effects of the environment, the genotype and the GE interaction. The term $\delta_{k(i)}$ is the (random) effect of the k th. block ($k=1, \dots, n$) in the i th. environment, δ i.i.d. $N(0, \sigma^2_{\text{block}})$, and the term ε_{ijk} is the (random) error term, ε_{ijk} i.i.d. $N(0, \sigma^2)$.

The AMMI model expresses each interaction term as a sum of multiplicative terms:

$$\alpha\beta_{ij} = \sum_{t=1}^m \lambda_t a_{ti} b_{tj} + \rho_{ij}$$

where λ_t , $\{a_{t1}, \dots, a_{te}\}$, $\{b_{t1}, \dots, b_{tg}\}$, are, respectively, the t th singular value, the t th column of the left-side matrix and the t th row of the right-side matrix of the singular value decomposition of E , the $e \times g$ matrix of interaction effects (i.e., the ij th. element of E is $\alpha\beta_{ij}$) (Mandel, 1971; Rao, 1973). The singular value decomposition has the property that the sum of squares of the interaction, SS_{GE} , equals $\sum_t \lambda_t^2$, and therefore the values λ_t^2 can be interpreted as the portion of the interaction sum of squares associated with the particular multiplicative component.

The estimation of the λ_s , a_s and b_s is done according to the procedure discussed by Mandel (1971). It consists in obtaining the residuals after fitting a main-effects-only model, and computing the singular value decomposition of the estimated interaction effects. Price and Shafii (1993) present a SAS IML program to do this. Note that for balanced data it is equivalent to compute the residuals from averages or from using all replicates and averaging the resulting residuals. It is not clear at present which of the two alternatives is preferable when the data are unbalanced.

For parsimony, m (the number of singular values retained in eq. 2) should be small, preferably much smaller than its maximum, $\min(e, g)$. Likelihood ratio tests have been proposed to determine the number of multiplicative terms to retain in (2) (Hegeman and Johnson, 1976; Milliken and Johnson, 1989). These tests can be computed directly using SAS Proc Mixed (SAS Institute, 1997) by fitting appropriate models under the null and alternative hypotheses, and letting the program compute the corresponding maximized likelihoods.

The hypotheses for testing whether it suffices to retain up to the m th multiplicative term are:

$$H_0: E(Y_{ijk}) = \mu + \alpha_i + \beta_j + \sum_{t=1}^m \lambda_t a_{ti} b_{tj}$$

$$H_a: E(Y_{ijk}) = \mu + \alpha_i + \beta_j + \alpha\beta_{ij}$$

Under the alternative hypothesis, the maximized likelihood is computed automatically by SAS Proc Mixed. In order to compute the maximized likelihood under H_0 , one can compute the residuals after fitting all fixed effects in the model:

$$r_{ijk} = Y_{ijk} - \left(\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \sum_{t=1}^m \hat{\lambda}_t \hat{\alpha}_{it} \hat{b}_{tj} \right).$$

Once these residuals are computed, they can be used as the dependent variable in a model with no fixed effects, no intercept, and the same random effects present in the original model (in the example, $\delta_{k(i)}$ and ε_{ijk}). If the data set is balanced, then the estimation procedure described above yields also the maximum likelihood estimators of the fixed effects, and hence the maximized likelihood can be obtained directly from the residuals r_{ijk} . To verify this, note that an individual term in the maximized log likelihood under the null hypothesis is:

$$\text{constant} - \frac{1}{2} \log(\widehat{\text{Var}}(Y_{ijk})) - \frac{1}{2} \frac{(Y_{ijk} - \hat{\mu}_{ijk})^2}{\widehat{\text{Var}}(Y_{ijk})}$$

The numerator of the last term of this equation is r_{ijk}^2 , and therefore the maximized likelihood can be computed directly for the proposed model using the residuals.

Once both maximized likelihoods are available, a likelihood ratio test statistic can be computed by taking the difference $-2 \log L_{H0} - (-2 \log L_{Ha})$. Since the null hypothesis is a proper subset of the alternative, this statistic has an asymptotic chi-square asymptotic null distribution. The degrees of freedom can be approximated using the formula suggested by Gollob (1968):

$$df = (e-1)(g-1) - \Sigma(g+e-1-2t) = (e-1)(g-1) - m(e+g-m-2).$$

Other approximations could be used (Mandel, 1971; Milliken and Johnson, 1989).

In summary, the following steps are necessary to fit the AMMI models in SAS:

1. Read data and fit a model with main effects but no interaction.
2. Obtain residuals and average them to obtain the E matrix (estimated GE interaction terms).
3. Obtain the singular value decomposition of E and generate data sets with the results (Shafii and Price, 1998).
4. Generate data sets containing the fitted values and residuals for models containing main effects and 1, 2, 3, etc. multiplicative interaction terms.
5. Fit models using the residuals obtained in 4 as dependent variables, no fixed effects, no intercept and the same random effects used in the model fitted in part 1. Compute the maximized likelihood for each of these models.
6. Fit a general model with main effects and interaction and compute its maximized likelihood.
7. Compute likelihood ratio test statistics for each of the models fitted in 5 and decide how many multiplicative terms to retain.

8. Compute the AMMI model means and prepare biplots to better interpret the results (Shafii and Price, 1998).

3. RED MOTTLED BEAN VARIETY TRIALS

From 1990 to 1992, several variety trials were carried out to evaluate the performance of breeding lines with determinate and indeterminate growth habit over a wide range of environmental conditions (Beaver *et al.*, 1996). A preliminary ANOVA shows large differences due to environment and genotype, and a large GE interaction ($F=4.14$, $p\text{-value}=0.0001$). Partition of the genotype effects and the GE interaction to account for the different growth habits also show highly significant effects and interaction (see Table 1). It also shows that a mayor contributor to the genotype and GE interaction effects is growth habit. The means show a wide range, from 0.4 to more than 3 ton/ha, with the determinate lines, in general, showing lower yields than the indeterminate ones (see Table 2).

Table 3 shows the percentage of variability associated with each of the singular values. The last 3 columns show the results of the approximte likelihood ratio tests for testing the goodness of fit of a model retaining the first m multiplicative terms ($m=1, 2, 3, 4$). From this table it is clear that the first three multiplicative terms explain most of the GE interaction, and that additional terms do not improve the fit.

Table 1. ANOVA table for fixed effects in red mottled bean variety trial

Source of Variation	F	p-value
Environment	42.92	.0001
Growth habit	134.45	.0001
Genotype (Growth habit)	3.40	.0001
Envir. X Growth habit	42.65	.0001
Envir. X Genot. (Growth habit)	2.76	.0001

The biplots are presented in Figures 1-4. They were constructed using the program provided by Price and Shafii (1993). Plots involving the first component clearly indicate that the most important aspects of the GE interaction can be explained by the growth habit: essentially all the genotypes with indeterminate growth (numbers 16-30) have positive scores for the first component, while most of the genotypes with determinate growth have negative scores. Because of the multiplicative effects used in the model, a given environment will manifest opposite effects for determinate and indeterminate lines. Environments 6-9 (all of the Puerto Rico planting dates except January 1990) will tend to have a positive correlation with indeterminate lines, while environments 1-5 (Dominican Republic locations and Puerto Rico's January 1990 planting date) seem positively correlated with the lines with determinate growth habit.

Table 2. Observed mean yields for the red mottled bean variety trials (ton/ha). Lines 1-15 have determinate growth habit and lines 16-30 have indeterminate growth habit. Environments 1-4 are in the Dominican Republic (San Cristóbal Dec. 90, Vallejuelo Sep. 91, Manoguayabo Dec. 91 and San Juan de la Maguana Dec. 91 respectively). Environments 5-9 are in Isabela, Puerto Rico (Jan. 90, Oct. 90, Jan. 91, Oct. 91 and Jan. 92 respectively).

Line	Environment									Mean
	1	2	3	4	5	6	7	8	9	
1 (8738-7B)	1.13	1.23	1.81	1.56	1.11	1.16	1.44	0.38	1.38	1.24
2 (8738-12B)	0.88	1.52	1.68	1.23	1.26	1.18	1.74	0.35	1.55	1.27
3 (8738-5B)	1.10	1.40	1.34	1.29	0.97	1.24	1.77	1.47	2.30	1.43
4 (Indiana Roja)	0.72	1.10	1.35	0.75	1.26	1.07	1.60	2.26	1.54	1.30
5 (8738-3B)	1.29	1.94	1.75	1.21	1.13	1.61	1.68	1.03	1.92	1.51
6 (Pompadour F)	0.61	1.52	1.79	1.02	1.13	1.39	1.25	0.35	1.19	1.14
7 (8738-4B)	1.17	1.31	1.55	1.33	1.09	1.13	1.36	1.34	1.74	1.34
8 (Pompad. Checa)	1.00	1.06	1.80	1.27	1.09	1.32	0.98	0.74	1.08	1.15
9 (8738-8B)	0.83	1.60	1.93	1.10	1.12	1.85	1.13	0.56	1.43	1.28
10 (Pompadour N)	1.05	1.60	1.84	1.35	1.02	0.95	1.69	0.52	1.57	1.29
11 (8738-1B)	1.11	1.13	1.88	1.08	1.05	0.89	1.51	1.21	1.51	1.26
12 (8911-55)	1.03	1.31	1.21	1.25	0.90	2.41	1.25	0.28	1.08	1.19
13 (8857-67)	1.28	1.15	1.15	1.24	0.75	2.32	2.10	0.92	2.20	1.46
14 (PC50)	1.25	1.58	2.53	2.08	1.09	1.88	1.05	0.47	1.38	1.48
15 (Jose Beta)	1.04	1.60	1.83	1.75	1.15	1.92	1.81	0.63	1.25	1.44
16 (8856-12)	1.42	1.40	1.06	0.76	0.93	2.92	1.84	1.99	1.55	1.54
17 (8856-61)	1.02	1.02	0.77	0.87	0.90	2.10	2.44	2.05	2.60	1.53
18 (8953-12)	0.64	0.79	0.87	0.79	1.01	2.28	1.72	2.13	2.31	1.39
19 (8953-1)	1.03	1.25	1.13	1.15	1.17	3.08	2.19	1.54	1.81	1.59
20 (8953-2)	1.04	1.48	1.62	1.48	1.11	1.70	1.95	2.05	2.42	1.65
21 (8953-3)	0.82	1.09	1.10	1.31	0.66	2.24	1.80	2.24	2.37	1.52
22 (8953-6)	0.78	1.15	1.01	1.46	0.84	3.15	2.20	1.65	2.11	1.59
23 (8953-9)	1.18	1.44	1.41	1.49	0.84	2.31	2.37	2.83	2.17	1.78
24 (8953-10)	1.45	1.60	1.66	1.77	1.14	1.67	2.37	1.80	2.17	1.74
25 (8738-7B)	1.58	0.88	1.56	1.50	0.57	1.75	2.33	1.26	2.00	1.49
26 (Pompadour G)	1.12	1.21	1.49	1.98	1.05	2.19	1.95	2.78	2.71	1.83
27 (Pompadour V)	1.29	1.16	1.26	0.92	1.19	1.70	2.30	1.93	1.82	1.51
28 (Pompadour K)	1.03	1.38	1.35	1.44	1.38	2.31	1.95	1.48	1.83	1.57
29 (Pompadour T)	1.63	0.98	1.23	3.33	1.28	2.04	1.91	1.33	2.57	1.81
30 (8858-28)	1.30	1.43	1.13	1.45	1.02	3.10	2.20	2.80	2.87	1.92
Mean	1.09	1.31	1.47	1.37	1.04	1.90	1.80	1.41	1.88	1.47

Table 3. Percentage of variability and approximate goodness of fit tests for multiplicative components for red mottled beans variety trials.

Multiplicative Components	λ^2	Variance proportion (%)		Goodness of fit tests		
		Component	Cumulative	χ^2	d.f.	p-value
1	101.72	53.91	53.91	368.83	196	0.0000
2	34.54	18.32	72.22	237.46	162	0.0001
3	23.97	12.71	84.93	135.48	130	0.3532
4	12.19	6.46	91.39	79.56	100	0.9345
5	6.28	3.33	94.72			
6	4.54	2.40	97.12			
7	3.15	1.67	98.79			
8	2.29	1.21	100.00			

Lines having scores close to 0 for the first 3 components can be interpreted as more stable (presenting more predictable yields in most of the environments). The lines 3, 7, 13, 20, 24, 25 and 28 seem most stable, since their scores for the first component are smaller than 0.1 (in absolute value), and their scores for the second and third components are smaller than 0.3 in absolute value. Furthermore, line 24 combines a high mean yield with predictable performance.

Figure 4 graphs the first component versus the mean yield (ton/ha). From this graph the growth habit effect can be noted: indeterminate lines tend to have higher yields than determinate lines. The genotypes which are closer to the reference line at 0 indicate a greater relative stability. Separate analysis for indeterminate and determinate lines show a similar interaction structure, and 2-3 multiplicative components would need to be retained in each case.

4. SUMMARY

In this paper we reviewed the additive main effects and multiplicative interaction model and discussed fitting strategies. We present an algorithm to fit this model and compute goodness of fit ratio tests to decide how many multiplicative components are important. This algorithm combines the capabilities of SAS PROC MIXED and IML. We present an example applying the proposed methodology to data from red mottled bean variety trials conducted in the Dominican Republic and Puerto Rico in 9 environments with 30 lines.

ACKNOWLEDGMENTS

We thank Mónica Balzarini and a referee for suggestions which greatly improved this paper.

REFERENCES

- Aguiluz, A. (1998). Evaluacion de híbridos de maíz (*Zea mays* L.) de grano blanco y amarillo en ambientes de Centroamérica, Panamá y el Caribe en 1996. *Agronomía Mesoamericana* 9, 28-37.
- Beaver, J.S., E. Arnaud-Santana and D.P. Coyne (1996). Yield stability of determinate and indeterminate red mottled beans. *Journal of Agriculture of the Univ. of Puerto Rico* 80, 187-189.
- Gauch, Hugh G, Jr. (1988). Model selection and validation for yield trials with interaction. *Biometrics* 44, 705-715.
- Gollob, H.F. (1968). A statistical model which combines features of factor analytis and analysis of variance. *Psychometrika* 33, 73-115.
- Hegeman, V. and D.E. Johnson (1976). On analyzing two-way AOV data with interaction. *Technom.* 18, 273-281.
- Mandel, J. (1971). A new analysis of variance model for non-additive data. *Technom.* 13, 1-18.
- Milliken, G.A. and D.E. Johnson (1989). *Analysis of Messy Data Volume 2: Nonreplicated Experiments*. New York: Van Nostrand-Reinhold.
- Price, W.J. and B. Shafii (1993). The use of biplots in diagnosing interaction patterns of two-way classification data. *Proceedings of the Eighteenth Annual SAS Users Group International Conference*, Cary (NC): SAS Institute, Inc.
- Rao, C.R. (1973). *Linear Statistical Inference and its Applications*. 2nd. ed., New York: Wiley.
- SAS Institute, Inc. (1997). *SAS/STAT Software: Changes and Enhancements through Release 6.12*. Cary, NC: SAS Institute, Inc.
- Shafii, B., K.A. Mahler, W.J. Price and D.L. Auld (1992). Genotype x environment interaction effects on winter rapeseed yield and oil content. *Crop Science* 32, 922-927.
- Shafii, B. and W.J. Price (1998). Analysis of genotype-by-environment interaction using the additive main effects and multiplicative interaction model and stability estimates. *Journal of Agricultural, Biological and Environmental Statistics* 3, 335-345.
- van Eeuwijk, F.A. and P.M. Kroonenberg (1998). Multiplicative models for interaction in three-way ANOVA, with applications to plant breeding. *Biometrics* 54, 1315-1333.

Figure 1. Second vs. First Multiplicative Components. (Large numbers represent environments and small numbers represent genotypes.)

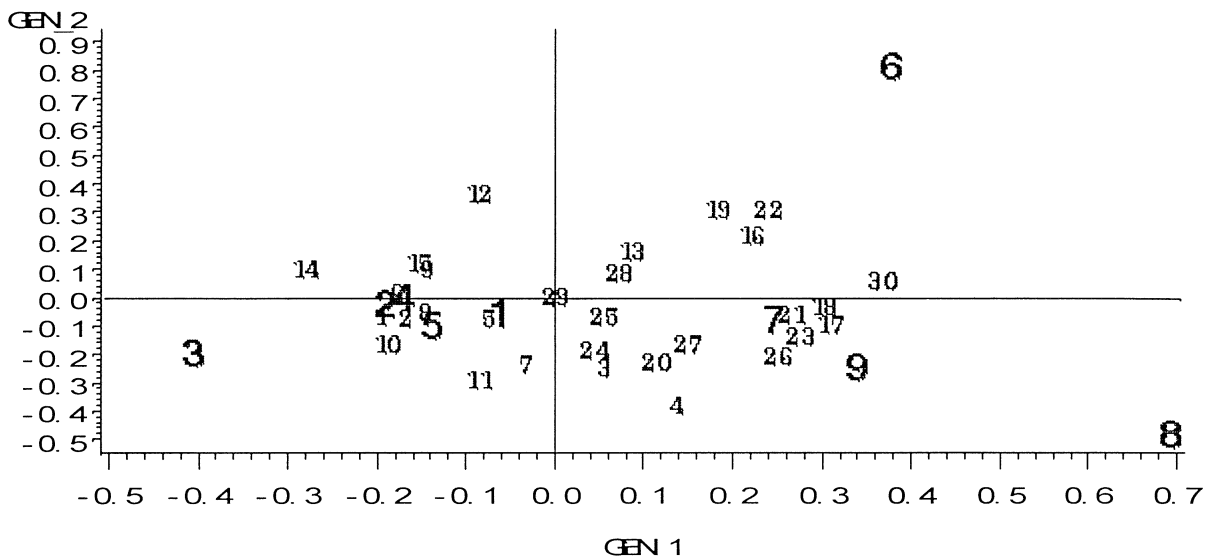


Figure 2. Third vs. First Multiplicative Components. (Large numbers represent environments and small numbers represent genotypes.)

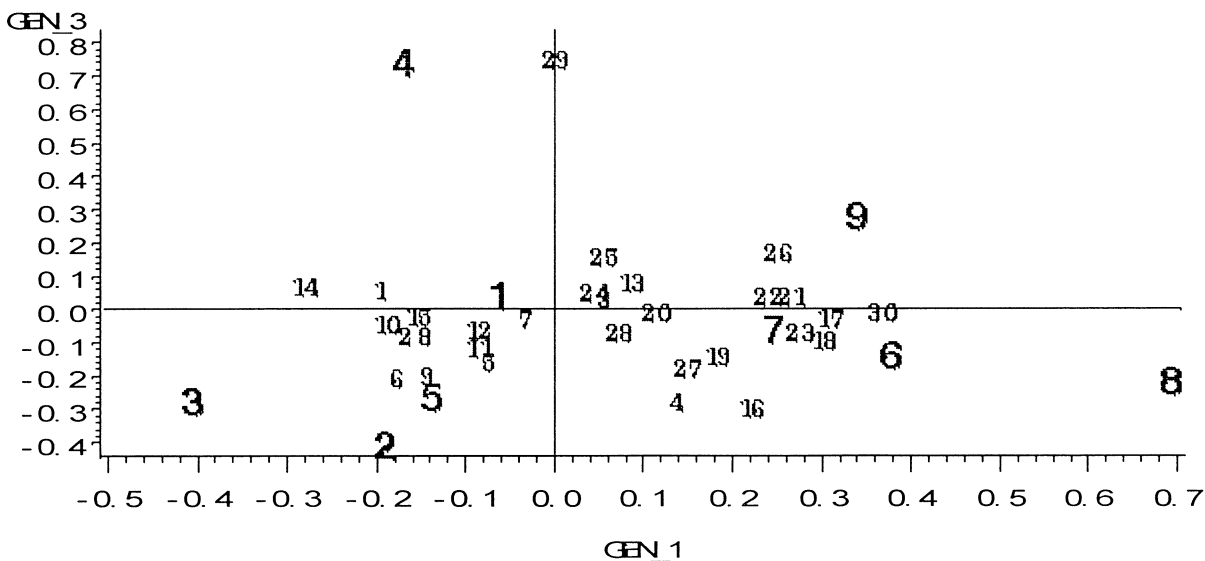


Figure 3. Third vs. Second Multiplicative Components. (Large numbers represent environments and small numbers represent genotypes.)

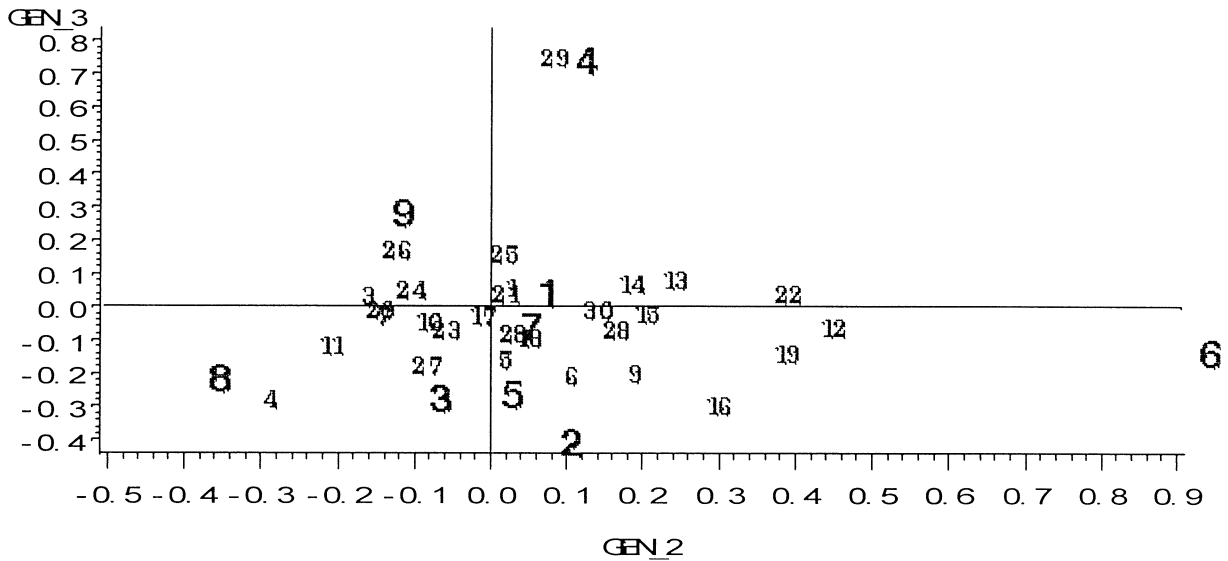
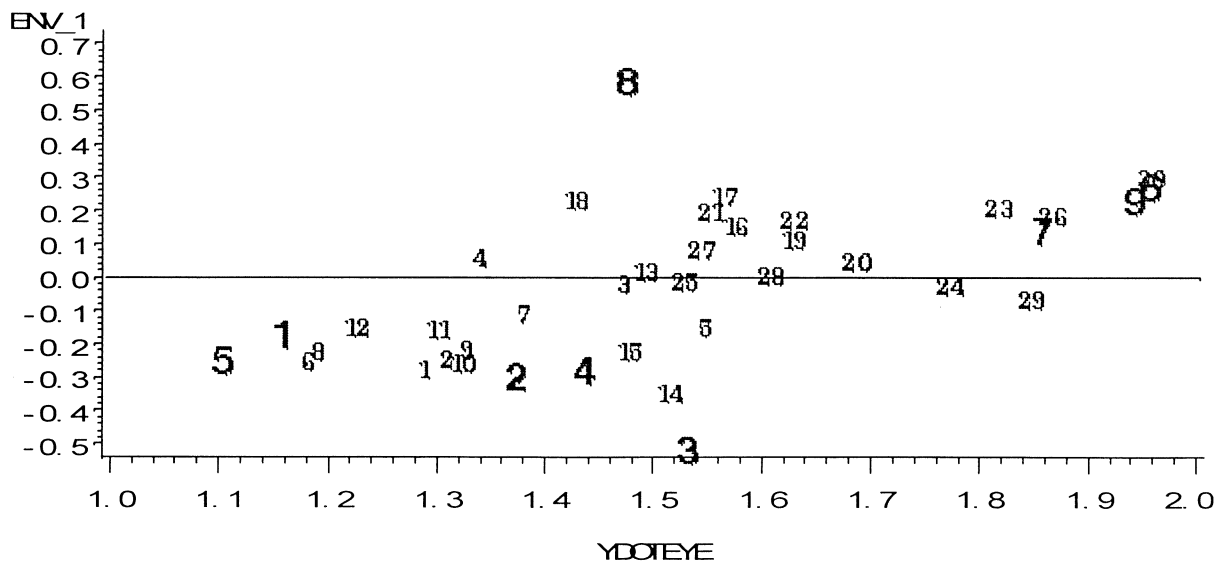


Figure 4. First Multiplicative Component Scores vs. Mean Yields. (Large numbers represent environments and small numbers represent genotypes.)



APPENDIX

SAS program for analyzing the data

```

%LET ENV_N = 9;  /*** Set the number of environments ***/
%LET GEN_N = 30; /*** Set the number of genotypes   ***/
%LET REP_N = 4;  /*** Set the number of replicates  ***/

%LET VAR = yield; /*** Set the name of the response variable here   ***/
%LET ENV = env;   /*** Set the name of the environment variable here ***/
%LET GEN = gen;   /*** Set the name of the genotype variable here   ***/
%LET REP = repet; /*** Set the name of the replication variable here ***/

/***** READ IN DATA *****/

data yld;
  set data.yield;

/***** RUN MIXED FOR INTERACTION/RESIDUALS *****/

proc mixed noitprint noclprint data=yld;
  class &GEN &ENV &REP;
  model &VAR = &ENV &GEN / p;
  random &REP(&ENV); /* FOR THE MODEL, PUT EVERYTHING BUT THE INTERACTION*/
  make 'Predicted' out=resid1 noprint;
  id &GEN &ENV &REP;

/***** COMPUTE AVERAGE OF RESIDUALS TO OBTAIN THE E MATRIX *****/

proc sort data=resid1;
  by &ENV &GEN ;
proc means mean noprint;
  by &ENV &GEN ;
  var _resid_;
  output out=resid2 mean=res;

/***** START IML FOR SINGULAR VALUE DECOMPOSITION *****/

proc iml;
  use resid2 var{&ENV &GEN res};
  read all var{res} into r;
  e=shape(r,&ENV_N);
  call svd(envvec,val1,genvec,e);
  val=val1#val1*&REP_N;
  propeig= val/val[+];
  propacum=propeig;
  do i=2 to nrow(val);
    propacum[i]=propacum[i-1]+propeig[i];
  end;
  print 'Proportions of each component', val propeig propacum;

```

Applied Statistics in Agriculture

```

/***** OUTPUT E, G, & G*E VECTORS TO SAS DATA SET FOR PLOTTING ****/

    create vece from envvec (|colname=label|);
    append from envvec;
    create vecg from genvec (|colname=labg|);
    append from genvec;

/**** CREATE DATASET FOR COMPUTING PREDICTED VALUES IN MIXED ****/

    matcomb = envvec` || genvec` || val1;
    label1="env1": "env&ENV_N" ;
    label2="gen1": "gen&GEN_N";
    label3={"lambda"};
    label=label1||label2||label3;
    create matrix from matcomb (|colname=label|);
    append from matcomb;
quit;

/**** PREPARE DATASETS WITH PREDICTED VALUES USING 0, 1, 2, etc. PCA ****/

proc mixed noitprint noclprint data=yld method=ML;
    class &ENV &GEN &REP;
    model &VAR = &ENV &GEN /pm;
    random &REP(&ENV) ;
    make 'PredMeans' out=pre0 noprint;
    make 'Fitting' out=pca0 noprint;
    id &ENV &GEN &REP;

data pre0; set pre0; check=1;

data matrix1; set matrix;
    if _n_=1; check=1;

data pred1; merge pre0 matrix1; by check;
    array environ {&ENV_N} env1-env&ENV_N;
    array genotyp {&GEN_N} gen1-gen&GEN_N;
    keep &VAR &ENV &GEN &REP _PRED_ _RESID_ check;
    _PRED_ = _PRED_+lambda*environ(&ENV)*genotyp(&GEN);
    _RESID_ = &VAR - _PRED_;
data matrix1; set matrix;
    if _n_=2; check=1;

data pred2; merge pred1 matrix1; by check;
    array environ {&ENV_N} env1-env&ENV_N;
    array genotyp {&GEN_N} gen1-gen&GEN_N;
    keep &VAR &ENV &GEN &REP _PRED_ _RESID_ check;
    _PRED_ = _PRED_+lambda*environ(&ENV)*genotyp(&GEN);
    _RESID_ = &VAR - _PRED_;

data matrix1; set matrix;
    if _n_=3; check=1;

```

```

data pred3; merge pred2 matrix1; by check;
  array environ {&ENV_N} env1-env&ENV_N;
  array genotyp {&GEN_N} gen1-gen&GEN_N;
  keep &VAR &ENV &GEN &REP _PRED_ _RESID_ check;
  _PRED_ = _PRED_+lambda*environ(&ENV)*genotyp(&GEN);
  _RESID_ = &VAR - _PRED_;

data matrix1; set matrix;
  if _n_=4; check=1;

data pred4; merge pred3 matrix1; by check;
  array environ {&ENV_N} env1-env&ENV_N;
  array genotyp {&GEN_N} gen1-gen&GEN_N;
  keep &VAR &ENV &GEN &REP _PRED_ _RESID_ check;
  _PRED_ = _PRED_+lambda*environ(&ENV)*genotyp(&GEN);
  _RESID_ = &VAR - _PRED_;

proc mixed noitprint noclprint data=pred1 method=ml;
  class &ENV &GEN &REP;
  model _RESID_ = /noit;
  random &REP(&ENV);
  make 'Fitting' out=pca1 noprint;

proc mixed noitprint noclprint data=pred2 method=ml;
  class &ENV &GEN &REP;
  model _RESID_ = /noit;
  random &REP(&ENV);
  make 'Fitting' out=pca2 noprint;

proc mixed noitprint noclprint data=pred3 method=ml;
  class &ENV &GEN &REP;
  model _RESID_ = /noit;
  random &REP(&ENV);
  make 'Fitting' out=pca3 noprint;

proc mixed noitprint noclprint data=pred4 method=ml;
  class &ENV &GEN &REP;
  model _RESID_ = /noit;
  random &REP(&ENV);
  make 'Fitting' out=pca4 noprint;

proc mixed noitprint noclprint data=yld method=ml;
  class &ENV &GEN &REP;
  model &VAR = &ENV|&GEN;
  random &REP(&ENV);
  make 'Fitting' out=pca11 noprint;

data pca0; set pca0;
  if descr='-2 Log Likelihood';
  pca0=value; keep pca0;
data pca1; set pca1;
  if descr='-2 Log Likelihood';
  pca1=value; keep pca1;

```

Applied Statistics in Agriculture

```

data pca2; set pca2;
  if descr='-2 Log Likelihood';
  pca2=value; keep pca2;
data pca3; set pca3;
  if descr='-2 Log Likelihood';
  pca3=value; keep pca3;
data pca4; set pca4;
  if descr='-2 Log Likelihood';
  pca4=value; keep pca4;
data pcall; set pcall;
  if descr='-2 Log Likelihood';
  pcall=value; keep pcall;

data likratio; merge pca0 pca1 pca2 pca3 pca4 pcall;
  pca=0; chisq=pca0-pcall; df=(&ENV_N-1)*(&GEN_N-1);
  p_value=1-probchi(chisq,df); output;
  pca=1; chisq=pca1-pcall; df=(&ENV_N-1)*(&GEN_N-1)-(&ENV_N+&GEN_N-3);
  p_value=1-probchi(chisq,df); output;
  pca=2; chisq=pca2-pcall; df=(&ENV_N-1)*(&GEN_N-1)-(2*&ENV_N+2*&GEN_N-8);
  p_value=1-probchi(chisq,df); output;
  pca=3; chisq=pca3-pcall; df=(&ENV_N-1)*(&GEN_N-1)-(3*&ENV_N+3*&GEN_N-15);
  p_value=1-probchi(chisq,df); output;
  pca=4; chisq=pca4-pcall; df=(&ENV_N-1)*(&GEN_N-1)-(4*&ENV_N+4*&GEN_N-24);
  p_value=1-probchi(chisq,df); output;
proc print data=likratio;
  var pca chisq df p_value;
  title 'Goodness of fit LRTs for adding PCA terms';

```