

Kansas State University Libraries

New Prairie Press

Conference on Applied Statistics in Agriculture

1992 - 4th Annual Conference Proceedings

STATISTICAL ANALYSIS OF GENOTYPE-BY-ENVIRONMENT INTERACTION USING THE AMMI MODEL AND STABILITY ESTIMATES

Bahman Shafii

William J. Price

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

Shafii, Bahman and Price, William J. (1992). "STATISTICAL ANALYSIS OF GENOTYPE-BY-ENVIRONMENT INTERACTION USING THE AMMI MODEL AND STABILITY ESTIMATES," *Conference on Applied Statistics in Agriculture*. <https://doi.org/10.4148/2475-7772.1395>

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.

STATISTICAL ANALYSIS OF GENOTYPE-BY-ENVIRONMENT INTERACTION USING THE AMMI MODEL AND STABILITY ESTIMATES

Bahman Shafii and William J. Price
Statistical Programs
College of Agriculture
University of Idaho
Moscow, ID 83843

ABSTRACT

Understanding the implication of genotype-by-environment (GE) interaction structure is an important consideration in plant breeding programs. A significant GE interaction for a quantitative trait such as yield can seriously limit efforts in selecting superior genotypes for both new crop introduction and improved cultivar development. Traditional statistical analyses of yield trials provide little or no insight into the particular pattern or structure of the GE interaction. The Additive Main Effects and Multiplicative Interaction (AMMI) statistical model incorporates both additive and multiplicative components of the two-way data structure which can account more effectively for the underlying interaction patterns. Integrating results obtained from biplot graphic displays with those of the genotypic stability analysis enables clustering of genotypes based on similarity of response and the degree of stability in performance across diverse environments. The AMMI model is presented, and its usage in diagnosing the GE interaction structure is discussed. Tai's stability statistics are employed to determine the stability of genotypes tested. Empirical applications are demonstrated using data from a national winter rapeseed variety trial.

Key Words: genotype-by-environment interaction, biplot analysis, stability statistics, yield trials.

I. INTRODUCTION

The Genotype-by-Environment (GE) interaction structure is an important aspect in both plant breeding programs and the introduction of new crop commodities. GE interaction can occur when specified genotypes are grown across diverse environments (Zobel, 1990). Hence, selection of superior genotypes for both improved crop development and new crop introduction can be severely limited in the presence of a significant GE interaction. Statistically, the usefulness of subsequent analyses of means and inferences would be seriously reduced as a result of a significant interaction encountered in the analysis of two-way classification data.

The most commonly used statistical technique for analyzing two-way data structures is analysis of variance (ANOVA). While this technique is useful in identifying and testing sources of variability, it does not discern patterns of the underlying interaction. The additive nature of the ordinary ANOVA model allows adequate description of main effects, however, the interaction (residual from the additive model) is nonadditive and requires other techniques to identify interaction relationships.

Several regression techniques have been proposed which are based on some measure of the environment, usually an environmental index (Yates and Cochran,

1938; Finlay and Wilkinson, 1963; Eberhart and Russel, 1966). The use of the environmental index for regression can confound interaction and main effects. Furthermore, regression techniques are only effective when a particular pattern and model are assumed. Such approaches however, could be useful in calculating stability statistics (e.g. Tai, 1971).

Principle Components Analysis (PCA) provides a multiplicative model which can be used to diagnose two-way data classifications. This technique fails to identify significant main effects.

A hybrid statistical model which incorporates both the additive and multiplicative components of the two-way data structure is the Additive Main Effects and Multiplicative Interaction (AMMI) model. In this model, the additive portion of the variance is separated from the multiplicative variance (interaction) by ANOVA. PCA analysis is then applied to the interaction (residual) portion from the ANOVA to extract a new set of coordinate axes which account more effectively for the interaction patterns. Estimation of the PCA axes is accomplished according to the least squares principle (Bradu and Gabriel, 1978).

The purpose of this paper is to demonstrate the effectiveness of the AMMI statistical model (biplot analysis) in diagnosing the GE interaction patterns, and the usage of stability statistics in classifying genotypes based on similarity of response characteristics. Empirical results are illustrated with reference to a winter rapeseed variety trial.

II. METHODS

The basic linear model used is of the form:

$$y_{ij} = \mu + g_i + I_j + \delta_{ij} \quad (1)$$

where

- y_{ij} = observed response value of genotype i in environment j ,
- μ = grand mean,
- g_i = effect for genotype i , $i=1, \dots, k$,
- I_j = effect for environment j , $j=1, \dots, n$,
- δ_{ij} = residual

The residual component can be partitioned into the sum of multiplicative functions of i and j (Mandel, 1971). Thus,

$$\begin{aligned} y_{ij} &= \mu + g_i + I_j + (\lambda_1 v_{1i} v_{1j} + \lambda_2 v_{2i} v_{2j} + \dots) \\ &= \mu + g_i + I_j + \sum \lambda_m v_{mi} v_{mj} + \rho_{ij} \end{aligned} \quad (2)$$

Here, λ_m is the square root of the eigenvalue for PCA axis m of the matrix $E'E$ where E is the $k \times n$ matrix of interaction, δ_{ij} , from (1); v_{mi} and v_{mj} are the elements of the corresponding normalized eigenvectors of the matrices EE' and $E'E$, respectively; m is the number of axes retained, $m \leq \min(k, n-1)$; ρ_{ij} is a random error.

The AMMI model for replicated experiments takes the form :

$$y_{ijr} = \mu + g_i + I_j + b_{jr} + \sum \lambda_m v_{mi} v_{mj} + \rho_{ij} + \epsilon_{ijr} \quad (3)$$

where

- b_{jr} = effect of the r^{th} replicate in the j^{th} environment,

ϵ_{ijr} = error term .

Biplot analysis (Gabriel, 1971; Bradu and Gabriel, 1978, Zobel et. al., 1988) of the GE interaction allows for visual inspection and interpretation of the underlying structure and causes of interaction. This includes biplots of PCA interaction axes against one another, as well as genotype and environment means.

The corresponding regression approach leads to the following linear model:

$$y_{ijr} = \mu + g_i + l_j + b_{jr} + \beta_i l_j + \gamma_{ij} + e_{ijr} \quad (4)$$

where, for genotypic stability analysis, the interaction term, δ_{ij} (1), is partitioned into two components: the linear response to the environmental index, and the deviation from the linear response, i.e. $\delta_{ij} = \beta_i l_j + \gamma_{ij}$. The term e_{ijr} represents random residuals.

The general analysis of variance for both the AMMI and regression models is given in Table 1.

It can be shown that the least squares estimation of the β_i 's (regression coefficients) is equivalent to extracting the first principle component of the genotypic performances (Williams, 1952). Given the linear model in (2) and provided that the majority of the genotypic variation is accounted for by the specified regression equation, the following expectation holds:

$$\lambda_1^2 = \Sigma \beta_i^2 \Sigma l_j^2 \quad (\text{Williams, 1952; Perkins, 1972})$$

where λ_1^2 is the largest eigenvalue of the matrix $E'E$. Thus, the regression model may be considered as a special case of the AMMI model with only one important eigenvalue.

A regression based procedure for assessing genotypic stability and range of performance is outlined by Tai (1971). Tai's stability statistics are as follows:

$$\hat{\alpha}_i = \hat{\beta}_i(\text{MSL})/(\text{MSL}-\text{MSB}) \quad (5)$$

$$\hat{\lambda}_i = \text{MSD}_i/(\mathbf{k}-1)\text{MSE}/\mathbf{k}\mathbf{r} \quad (6)$$

where MSL, MSB, and MSE are the mean square location, block and error, respectively, and MSD_i is the mean square deviation from regression for the i^{th} genotype, α_i measures the linear response to the environmental component, while λ_i measures the deviation from the linear response in terms of the error variance (Tai, 1971). A genotype possessing ($\alpha = 0$, $\lambda = 1$) is considered to have average stability, whereas values of ($\alpha = -1$, $\lambda = 1$) indicate perfect stability. To assist in interpreting these statistics, a 100(1-a)% prediction interval for $\alpha_i=0$ and a corresponding confidence interval for $\lambda_i \geq 1$ may be constructed. Such a display enables classification of genotypes to different stability regions.

Statistical computations, estimation and graphics were carried out using procedures GLM and IML (SAS, 1990, 1990).

III. EMPIRICAL RESULTS

Data used in this study are from the U. S. National Winter Rapeseed Variety trials conducted during the years 1986, 1987, and 1988 at 12, 16, and 17 locations, respectively (Mahler, et. al. 1989). Six cultivars of Winter rapeseed [*Brassica napus* L. spp. *oleifera* (Metzg.) Sinsk. f. *biennis*] were included as controls: Bienvenu, Bridger, Cascade, Dwarf Essex, Glacier, and Jet Neuf. The experiment included four replications at each location. Locations were chosen to represent areas of existing or

potential rapeseed production as either an edible or industrial oil crop. The locations included: Georgia (GGA, TGA), Idaho (ID), Kansas (KS), Mississippi (MS), Montana (MT), New York (NY), North Carolina (NC), Oregon (OR), South Carolina (SC), Tennessee (TN), Texas (TX), Virginia (VA), and Washington (WA). The measured responses (of adaptation) were seed yield and oil content. The combined data set consisted of twenty seven location-years (environments) for yield, and twenty five environments for oil content. Only the seed yield analysis is considered here [see Shafii et.al. (1992) for further discussion].

Initial analysis of variance indicated a significant genotype by environment (GE) interaction for seed yield (Table 2). Environments were a much larger source of variability than genotypes. In accordance with the AMMI model, principle components analysis was used to decompose the GE interaction into six components (Table 3) of which the first two were deemed significant ($p=.0001$). These two axes accounted for over 80% of the variability (interaction sum of squares) while using less than half of the degrees of freedom (Table 2). Residual analysis revealed that the remaining PCA axes conformed to the expected pattern of random noise and were therefore assigned to the residual term. Overall, the AMMI model accounted for 98.3% of the total sum of squares.

Loadings of the PCA axes are good indicators of factors contributing to the variability. In the case of genotypes, the first PCA axis was dominated by cultivars Bienvenu and Bridger, suggesting that these cultivars exhibit a high degree of influence over the GE interaction. PCA axis 1 for environments, while not as clear in its loadings as genotypes, showed sizable scores for locations in the Pacific Northwest. The PCA scores of interaction could also be used to achieve more precise estimates of the true yield potential of both genotypes and specific environments (Shafii et. al. 1992).

AMMI biplots, as described earlier, are given in Figures 1 and 2. Figure 1 plots PCA axis 1 vs PCA axis 2 for both genotypes and environments. For the sake of clarity, only points of interest have been labeled. Since the GE component of the AMMI model is based on the product of PCA scores, it follows that cultivars or environments with small interactions (smaller scores) will appear close to the center of the axes. Cultivar Glacier exhibits this trait, while cultivars Bienvenu and Bridger are further from the center and thus, show strong interaction effects. The direction of cultivars and environments from the axes center also contains important information on the interaction. As an example, cultivar Bienvenu (BIENVENU) and environment Washington 1989 (WA89) appear opposite from each other indicating their contributions to the interaction was in opposing directions, e.g. they are negatively correlated. By contrast, cultivar Dwarf Essex (DWARF) and the environment Idaho 1989 (ID89) both have the same relative direction, so that both contribute positively to the interaction. Thus, the biplot can give information on relative stability, as well as suggesting trends of similar or dissimilar cultivars and environments.

The second biplot (Figure 2) is of PCA axis 1 vs mean yield of both genotypes and environments. This figure illustrates the wide discrepancy between the variability of environments and genotypes. Genotypes are represented as a narrow band at or around 2.5 Mg ha^{-1} , while environments cover a larger yield range. As before, the position of genotypes or environments is important with those appearing closer to the

vertical reference line at $PCA1=0$ indicating a greater relative stability. Grouping and trends are also important. For example, environments in the Pacific Northwest generally produce higher yields and tend to have PCA scores opposite in signs compared to those of the Southeast. The differential vernalization requirements of the six cultivars is thought to be responsible for the observed trend. Hence, these biplots are useful in determining genotypic stability and range of performance, as well as providing insight into the causes of the underlying interaction.

In addition to Tai's stability statistics, five other frequently cited stability measures were also computed from the data and are reported for the purpose of comparison (Table 4). These were: S_i^2 - the variance of genotypes across environments, W_i - ecovalence (Wricke, 1962), σ_i^2 - stability variance (Shukla, 1972), β_i - regression coefficient (Finlay and Wilkinson, 1963), and δ_i^2 - deviation from regression (Eberhart and Russel, 1966). All stability measures give identical rankings to the six genotypes with the exception of Finlay and Wilkinson's regression coefficient. This is not surprising given that S_i^2 , W_i , σ_i^2 , δ_i^2 , and λ_i are based on variability of the i^{th} genotype across environments, while β_i is based on the linear response of the i^{th} genotype to an environment. In spite of this difference, all six stability measures indicated the highest ranking cultivar (in terms of average stability) to be Glacier, and that the lowest rankings were consistently shared by cultivars Bienvenu and Bridger. This agrees with the results obtained from biplot analysis.

AMMI biplots were compared to a plot of distribution of Tai's stability statistics (Figure 3). There is strong agreement between the two graphic displays. Cultivar Glacier, which was close to the axes center in Figure 1, is in the region defined as average stability ($\alpha=0$, $\lambda=1$). Cultivars Jet Neuf and Cascade while having acceptable α levels have larger λ values, and are therefore regarded as stable with lower reliability. These cultivars were a modest distance from the axes center in Figure 1. Bridger and Bienvenu, both of which had large PCA scores, appear well outside the stability region in Figure 3. Therefore, the AMMI model and stability estimates both concur on the stability of genotypes. While Tai's statistics give an easily interpretable report on the stability and reliability of genotypes, the AMMI model provides information on the stability and trends of environments, as well as the correlation between a genotype and environment.

IV. CONCLUDING REMARKS

The AMMI model provides a useful tool in diagnosing GE interaction patterns and improving the accuracy of response estimates. It enables clustering of genotypes based on similarity of response characteristics and identifying potential trends across environments. Tai's stability statistics may be used, along with biplot analysis, to evaluate relative stability, reliability and ranking of genotypes in regional trials. The suggested strategy could extract more information from the GE interaction, thereby aiding researchers in identifying specific cultivars with competitive yields across diverse environments.

ACKNOWLEDGEMENTS

Contribution from the College of Agriculture, University of Idaho, Idaho Agricultural Experiment Station paper number 9302.

REFERENCES

- Bradu, D. and K.R. Gabriel. 1978. The biplot as a diagnostic tool for models of two-way tables. *Technometrics*. 20: 47-68.
- Eberhart, S.A. and W.A. Russel. 1966. Stability parameters for comparing varieties. *Crop Sci.* 6: 36-40.
- Finlay, K.W. and G.N. Wilkinson. 1963. An analysis of adaptation in a plant breeding programme. *Aust. J. Agri. Res.* 14: 742-754.
- Gabriel, K.R. 1971. The biplot graphic display of matrices with application to principle component analysis. *Biometrika*. 58: 453-467.
- Gauch, H.G. 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 analytic and analysis of variance. *Psychometrika*. 33: 73-115.
- Mahler, K.A., D.L. Auld, N.L. Christensen, W.T. Fike, H. Fribourg, J.E. Hairston, A.N. Hang, J. Hill, R. Karow, G.G. McBee, P.L. Raymer, and D.E. Starner. 1989. National winter rapeseed variety trial 1988-89. *Univ. of Idaho Misc. Ser, No.* 130.
- Mandel, J. 1971. A new analysis of variance model for non-additive data. *Technometrics*. 13: 1-18.
- Perkins, J. M. 1972. The principle component analysis of genotype-environmental interactions and physical measures of the environment. *Heredity* 29:51-70.
- SAS Institute Inc. 1990. *SAS/IML User's Guide, Version 5, Version 6 Edition*, Cary, NC.
- SAS Institute Inc. 1990. *SAS/STAT User's Guide, Version 6, Fourth Edition, Volume 2*, Cary, NC.
- Shafii, B., K.A. Mahler, W.J. Price, and D.L. Auld. 1992. Genotype by environment interaction effects on winter rapeseed yield and oil content. *Crop Sci.* 32: 922-927.

- Shukla, G.K. 1972. Some aspects of partitioning genotype-environmental components of variability. *Heredity* 29:237-245.
- Tai, G.C.C. 1971. Genotypic stability analysis and its application to potato regional trials. *Crop Sci.* 11: 184-190.
- Williams, E.J. 1952. The interpretation of interactions in factorial experiments. *Biometrika*, 39, 65-81.
- Wricke, G. 1962. Über eine methode zur erfassung der ökologischen streubreite in feldversuchen. *Z. Pflanzenzuecht.* 47:92-96.
- Yates, F., and W. G. Cochran. 1938. The analysis of groups of experiments. *J. Afric. Sci., Camb.*, 28, 556-580.
- Zobel, R.W. 1990. A powerful statistical model for understanding genotype-by-environment interaction. p. 126-140. In M.S. Kang (Ed), *Proc. Genotype-by-environment interaction and plant breeding.* Louisiana State University, Baton Rouge, LA.
- Zobel, R.W., M.J. Wright, and H.G. Gauch. 1988. Statistical analysis of a yield trial. *Agron. J.* 80: 388-393.

Table 1a. The general analysis of variance for the AMMI model.

| Source | df |
|----------------|---------------------------------|
| Environment | $e-1$ |
| Blocks | $e(b-1)$ |
| Genotype | $g-1$ |
| GE Interaction | $(e-1)(g-1)$ |
| † PCA 1 | $g + e - 3$ |
| PCA 2 | $g + e - 5$ |
| . | |
| . | |
| PCA m | $g + e - 1 - 2m$ |
| Residual | $(g-1)(e-1) - m(g + e - m - 2)$ |
| Exp. Error | $e(g-1)(b-1)$ |

† The degrees of freedom for axis m are determined by:
 $df_m = g + e - 1 - 2m$, where g and e and the number of
genotypes and environments, respectively (Gollob, 1968).

Table 1b. The general analysis of variance for the regression model.

| Source | df |
|----------------|---------------|
| Environment | $e-1$ |
| Blocks | $e(b-1)$ |
| Genotype | $g-1$ |
| GE Interaction | $(e-1)(g-1)$ |
| Regression | $g-1$ |
| Residual | $(g-1)(e-2)$ |
| Exp. Error | $e(g-1)(b-1)$ |

Table 2. Analysis of variance of AMMI model for seed yield (Mg ha^{-1}) of six rapeseed cultivars grown in 27 environments during 1986-87, 1987-88 and 1988-89.

| Source | df | Sum of squares | Mean square | F value | P > F |
|-----------|-----|----------------|-------------|---------|-------|
| Model | 242 | 1458.21 | 6.03 | 25.67 | .0001 |
| Environ | 26 | 1312.94 | 50.50 | 219.57 | .0001 |
| Blocks | 81 | 18.59 | 0.23 | 0.96 | .5783 |
| Genotype | 5 | 5.09 | 1.02 | 4.25 | .0009 |
| GE Inter. | 130 | 121.58 | 0.94 | 3.92 | .0001 |
| PCA 1 | 30 | 63.29 | 2.11 | 8.79 | .0001 |
| PCA 2 | 28 | 33.99 | 1.21 | 5.04 | .0001 |
| Residual | 72 | 24.30 | 0.34 | 1.42 | .0205 |
| Exp Error | 405 | 95.56 | 0.24 | | |
| C. Total | 647 | 1553.77 | | | |

Table 3. Eigenvalues and associated variance proportions for yield of six rapeseed cultivars grown in 27 environments during 1986-87, 1987-88, and 1988-89.

| PCA | Eigenvalues | Variance Proportion (%) | |
|-----|-------------|-------------------------|------------|
| | | Component | Cumulative |
| 1 | 63.29 | 52.1 | 52.1 |
| 2 | 33.99 | 28.0 | 80.1 |
| 3 | 10.95 | 9.0 | 89.1 |
| 4 | 7.95 | 6.5 | 95.6 |
| 5 | 5.40 | 4.4 | 100.0 |
| 6 | 0.00 | 0.0 | 100.0 |

Table 4. Various stability measures and their rankings of mean yield (Mg ha⁻¹) for six rapeseed cultivars grown in 27 environments during 1986-87, 1987-88, and 1988-89.

| Variety | $\underline{S_i^2}$ | $\underline{W_i}$ | $\underline{\sigma_i^2}$ | $\underline{\beta_i}$ | $\underline{\delta_i^2}$ | $\underline{\lambda_i}$ |
|----------|---------------------|-------------------|--------------------------|-----------------------|--------------------------|-------------------------|
| Bienvenu | 1.78 (6) | 8.73 (6) | 1.63 (6) | 0.139 (5) | 0.295 (6) | 5.90 (6) |
| Bridger | 1.59 (5) | 7.88 (5) | 1.14 (5) | -0.203 (6) | 0.217 (5) | 4.33 (5) |
| Cascade | 0.80 (4) | 4.50 (4) | 0.84 (4) | -0.043 (3) | 0.169 (4) | 3.38 (4) |
| Dwarf E. | 0.54 (2) | 3.37 (2) | 0.47 (2) | 0.097 (4) | 0.110 (2) | 2.19 (2) |
| Glacier | 0.17 (1) | 1.75 (1) | 0.21 (1) | -0.005 (1) | 0.067 (1) | 1.35 (1) |
| Jet N. | 0.73 (3) | 4.16 (3) | 0.78 (3) | 0.015 (2) | 0.160 (3) | 3.19 (3) |

Figure 1. Biplot of principle components analysis (PCA) axis 2 vs axis 1 for yield (Mg ha^{-1}) for six rapeseed cultivars grown in 27 environments during 1986-87, 1987-88, and 1988-89.

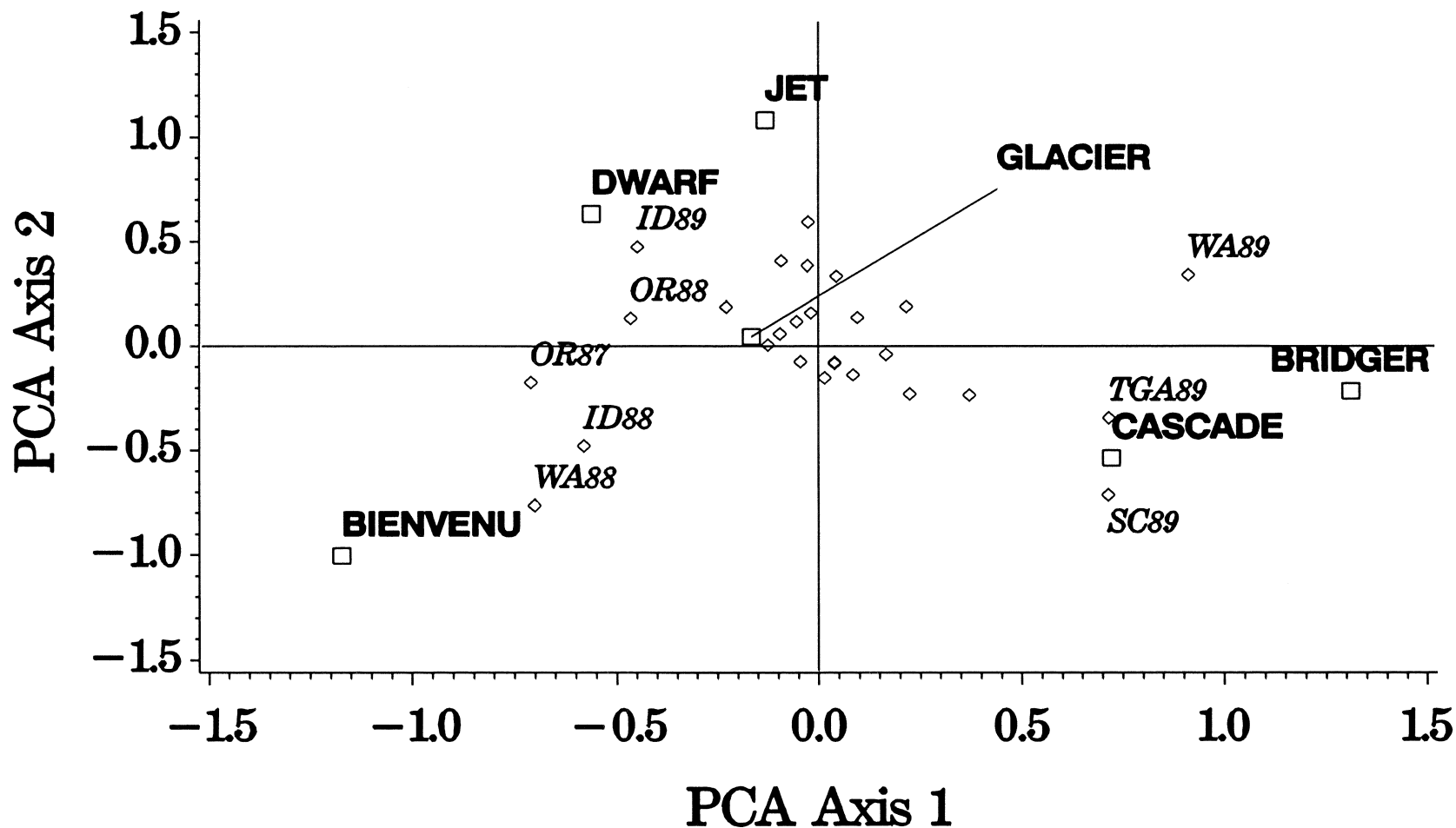


Figure 2. Biplot of principle components analysis (PCA) axis 1 vs mean yield (Mg ha⁻¹) for six rapeseed cultivars grown in 27 environments during 1986-87, 1987-88, and 1988-89.

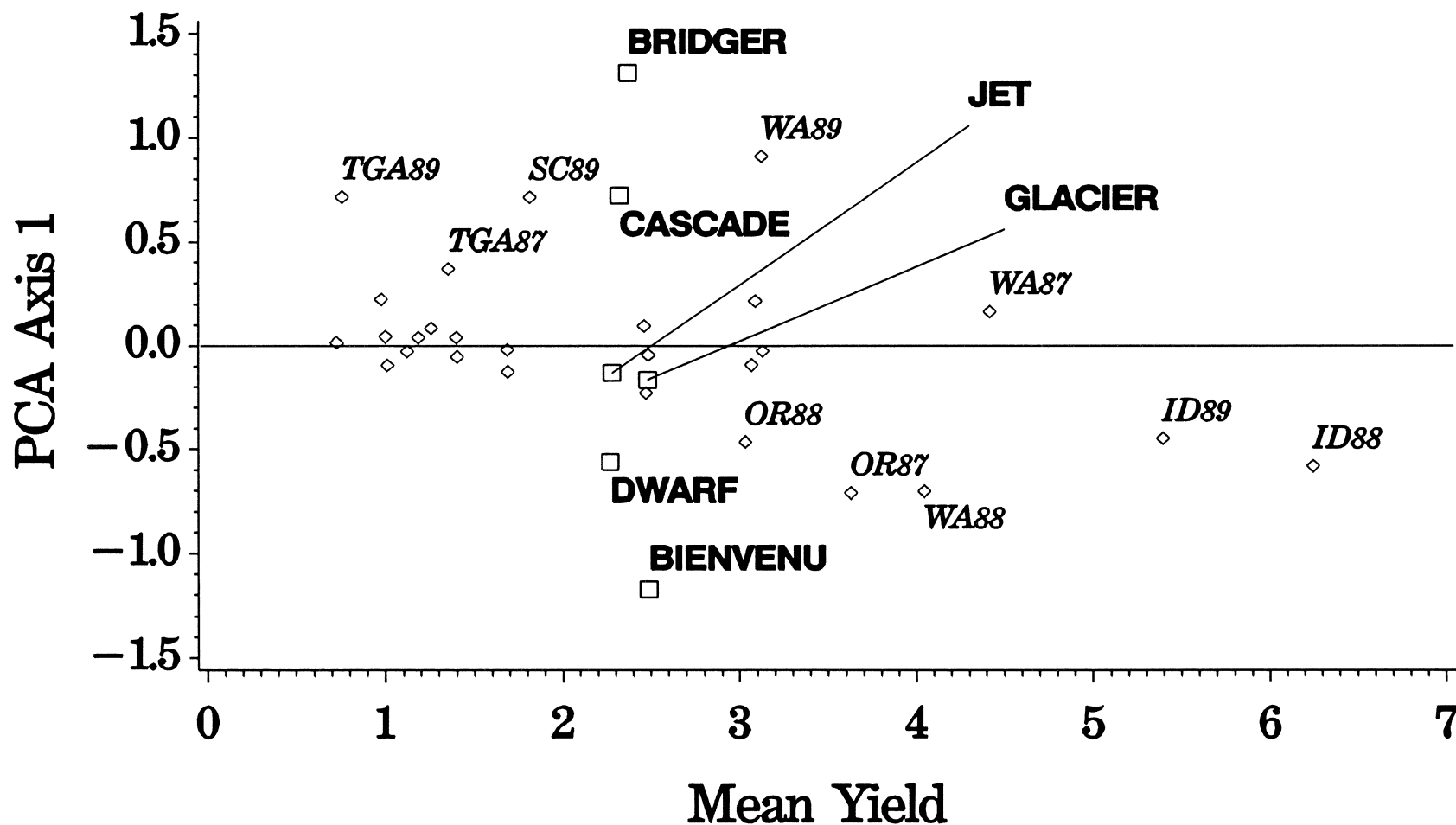


Figure 3. Distribution of Tai's stability statistics of yield (Mg ha^{-1}) for six rapeseed cultivars grown in 27 environments during 1986-87, 1987-88, and 1988-89. Prediction intervals computed at $\alpha = .05$.

