

# Model Selection and Cross Validation in Additive Main Effect and Multiplicative Interaction Models

Carlos T. dos S. Dias\* and Wojtek J. Krzanowski

## ABSTRACT

The additive main effects and multiplicative interaction (AMMI) model has been proposed for the analysis of genotype–environmental data. For plant breeding, the recovery of pattern might be considered to be the principal objective of analysis. However, some problems still remain with the analysis, notably in selecting the number of multiplicative components in the model. Methods based on distributional assumptions do not have a sound methodological basis, while existing data-based approaches do not optimize the cross-validation process. This paper first summarizes the AMMI model and outlines the available methodology for selecting the number of multiplicative components to include in it. Then two new methods are described that are based on a full “leave-one-out” procedure optimizing the cross-validation process. Both methods are illustrated and compared on some unstructured multivariate data. Finally, their applications to analysis of genotype  $\times$  environment interaction (GEI) are demonstrated on experimental grain yield data. Conclusions of the study are that the “leave-one-out” procedure is preferable in practice to either distributional  $F$ -test or cross-validation randomization methods, and of the two “leave-one-out” procedures the Eastment-Krzanowski method exhibits the greater parsimony and stability.

MOST OF THE DATA collected in agricultural experiments are multivariate in nature because several attributes are measured on each of the individuals included in the experiments, i.e., genotypes, agronomic treatments, etc. Such data can be arranged in a matrix  $\mathbf{X}$ , where the  $(i,j)$ th element represents the value observed for the  $j$ th attribute measured on the  $i$ th individual (case) in the sample. Common multivariate techniques used to analyze such data include principal component analysis (PCA) if there is no a priori grouping of either individuals or variables; canonical variate or discriminant analysis if the individuals in the sample form a priori groups; canonical correlation analysis if the variables form a priori groups; and cluster analysis if some partitioning of the sample is sought.

In plant breeding, multi-environment trials (MET) are important for testing general and specific cultivar adaptation. A cultivar grown in different environments will frequently show significant fluctuation in yield performance relative to other cultivars. These changes are influenced by the different environmental conditions and are referred to as GEI. A typical example of a matrix  $\mathbf{X}$  arises in the analysis of MET, in which the rows of  $\mathbf{X}$  are the genotypes and the columns are the

environments where the genotypes are tested. Presence of GEI rules out simple interpretative models that have only additive main effects of genotypes and environments (Mandel 1971; Crossa, 1990; Kang and Magari, 1996). On the other hand, specific adaptations of genotypes to subsets of environments is a fundamental issue to be studied in plant breeding because one genotype may perform well under specific environmental conditions and may give a poor performance under other conditions.

Crossa et al. (2002) give a comprehensive review of the early approaches for analyzing GEI that include the conventional fixed two-way analysis of variance model, the linear regression approach, and the multiplicative models. The empirical mean response,  $\bar{y}_{ij}$ , of the  $i$ th genotype in the  $j$ th environment with  $n$  replicates in each of the  $i \times j$  cells is expressed as  $\bar{y}_{ij} = \mu + g_i + e_j + (ge)_{ij} + \bar{\epsilon}_{ij}$  where  $\mu$  is the grand mean across all genotypes and environments,  $g_i$  is the additive effect of the  $i$ th genotype,  $e_j$  is the additive effect of the  $j$ th environment,  $(ge)_{ij}$  is the GEI component for the  $i$ th genotype in the  $j$ th environment, and  $\bar{\epsilon}_{ij}$  is the error assumed to be NID  $(0, \sigma^2/n)$  (where  $\sigma^2$  is the within-environment error variance, assumed to be constant). This model is not parsimonious, because each GEI cell has its own interaction parameter, and uninformative, because the independent interaction parameters are complicated and difficult to interpret.

Yates and Cochran (1938) suggested treating the GEI term as being linearly related to the environmental effect, that is setting  $(ge)_{ij} = \xi_i e_j + d_{ij}$ , where  $\xi_i$  is the linear regression coefficient of the  $i$ th genotype on the environmental mean and  $d_{ij}$  is a deviation. This approach was later used by Finlay and Wilkinson (1963) and slightly modified by Eberhart and Russell (1966). Tukey (1949) proposed a test for the GEI using  $(ge)_{ij} = K g_i e_j$  (where  $K$  is a constant). Mandel (1961) generalized Tukey's model by letting  $(ge)_{ij} = \lambda \alpha_i e_j$  for genotypes or  $(ge)_{ij} = \lambda g_i \gamma_j$  for environments and thus obtaining a “bundle of straight lines” that may be tested for concurrence (i.e., whether the  $\alpha_i$  or the  $\gamma_j$  are all the same) or nonconcurrence.

Gollob (1968) and Mandel (1969, 1971) proposed a bilinear GEI term  $(ge)_{ij} = \sum_{k=1}^s \lambda_k \alpha_{ik} \gamma_{jk}$  in which  $\lambda_1 \geq$

**Abbreviations:** AMMI, additive main effects and multiplicative interaction model; COMM, completely multiplicative model; DF, degrees of freedom; GEI, genotype  $\times$  environment interaction; GREG, genotype regression model; IPCA, interaction principal component analysis; MET, multi-environment trials; NID, normally and independently distributed; PCA, principal components analysis; PRESS, predictive sum of squares; PRECORR, predictive correlation; RMSPD, root mean square predictive difference; SHMM, shifted multiplicative model; SREG, sites regression model; SS, sum of squares; SVD, singular value decomposition.

C.T. dos S. Dias, Dep. of Ciências Exatas, Univ. of São Paulo/ESALQ, Av. Pádua Dias 11, Cx.P.09, 13418-900, Piracicaba-SP, Brazil; W.J. Krzanowski, School of Mathematical Sciences, Laver Building, North Park Road, Exeter, EX4 4QE, UK. Received 8 Apr. 2002. \*Corresponding author (ctsdias@carpa.ciagri.usp.br).

$\lambda_2 \geq \dots \geq \lambda_s$  and  $\alpha_{ik}, \gamma_{jk}$  satisfy the ortho-normalization constraints  $\sum_i \alpha_{ik} \alpha_{ik'} = \sum_j \gamma_{jk} \gamma_{jk'} = 0$  for  $k \neq k'$  and  $\sum_i \alpha_{ik}^2 = \sum_j \gamma_{jk}^2 = 1$ . This leads to the linear-bilinear model  $\bar{y}_{ij} = \mu + g_i + e_j + \sum_{k=1}^s \lambda_k \alpha_{ik} \gamma_{jk} + \bar{\epsilon}_{ij}$ , which is a generalization of the regression on the mean model, with more flexibility for describing GEI because more than one genotypic and environmental dimension is considered. Zobel et al. (1988) and Gauch (1988) called this the Additive Main Effects and Multiplicative Interaction (AMMI) model.

A family of multiplicative models can then be generated by dropping the main effect of genotypes (Site Regression Model, SREG), the main effect of sites (Genotype Regression Model, GREG), or both main effects (Complete Multiplicative Model, COMM). Another multiplicative model, the Shifted Multiplicative Model (SHMM) (Seyedsadr and Cornelius, 1992) is useful for studying crossover GEI (Crossa et al., 2002.)

However, one aspect that has not yet been fully resolved concerns the determination of the number of multiplicative components to be retained in the model to adequately explain the pattern in the interaction. Some proposals have been put forward by, among others, Gollob (1968), Mandel (1971), Gauch and Zobel (1988), Cornelius (1993), and Piepho (1994, 1995). All take into consideration the proportion of the variance accumulated by the components (Duarte and Vencovsky, 1999), and the more recent ones focus on cross validation as a predictive data-based methodology. However, some problems still remain, notably in optimizing the cross-validation process.

In this paper, we first summarize the AMMI model and analysis for genotype-environment data, and sketch out the available methodology for selecting the number of multiplicative components in the model. We then describe two methods based on a full leave-one-out procedure that optimizes the cross-validation process. Both methods are illustrated on some unstructured multivariate data. Their application to analysis of GEI is then exemplified on some experimental data, and a comparison of all available methods is made on data from five multi-environment cultivar trials.

## MATERIAL AND METHODS

### The AMMI Model

Suppose that a set of  $g$  genotypes has been tested experimentally in  $e$  environments. The mean of each combination of genotype and environment, obtained from  $n$  replications of an experiment (a balanced set of data), can be represented by the following array

$$\mathbf{Y}_{(g \times e)} = \begin{bmatrix} Y_{11} & Y_{12} & \dots & Y_{1e} \\ Y_{21} & Y_{22} & \dots & Y_{2e} \\ \dots & \dots & \dots & \dots \\ Y_{g1} & Y_{g2} & \dots & Y_{ge} \end{bmatrix}$$

The AMMI model postulates additive components for the main effects of genotypes ( $g_i$ ) and environments ( $e_j$ ) and multiplicative components for the effect of the interaction ( $ge$ )<sub>ij</sub>. Thus, the mean response of genotype  $i$  in an environment  $j$  is modeled by:

$$Y_{ij} = \mu + g_i + e_j + \sum_{k=1}^m \lambda_k \alpha_{ik} \gamma_{jk} + \rho_{ij} + \epsilon_{ij}$$

in which  $(ge)_{ij}$  is represented by:

$$\sum_{k=1}^m \lambda_k \alpha_{ik} \gamma_{jk} + \rho_{ij},$$

under the restrictions:

$$\sum_i g_i = \sum_j e_j = \sum_i (ge)_{ij} = \sum_j (ge)_{ij} = 0.$$

Estimates of the overall mean ( $\mu$ ) and the main effects ( $g_i$  and  $e_j$ ) are obtained from a simple two-way ANOVA of the array of means  $\mathbf{Y}_{(g \times e)} = [Y_{ij}]$ . The residuals from this array then constitute the array of interactions:

$$\mathbf{GE}_{(g \times e)} = [(\hat{ge})_{ij}]$$

and the multiplicative interaction terms are estimated from the singular value decomposition (SVD) of this array. Thus,  $\lambda_k$  is estimated by the  $k$ th singular value of  $\mathbf{GE}$ ,  $\alpha_{ik}$  is estimated by the  $i$ th element of the left singular vector  $\alpha_{k(g \times 1)}$ , and  $\gamma_{jk}$  is estimated by the  $j$ th element of the right singular vector  $\gamma_{k(1 \times e)}$  associated with  $\lambda_k$  (Good, 1969; Mandel, 1971; Piepho, 1995). Correspondences between SVD and PCA are as follows:  $\lambda_k$  is the  $k$ th singular value or the square root of the  $k$ th largest eigenvalue of the arrays  $(\mathbf{GE})(\mathbf{GE})^T$  and  $(\mathbf{GE})^T(\mathbf{GE})$ , which have equal nonnull eigenvalues;  $\alpha_{ik}$  is the  $i$ th element of the eigenvector of  $(\mathbf{GE})(\mathbf{GE})^T$  associated with  $\lambda_k^2$ ;  $\gamma_{jk}$  is the  $j$ th element of the eigenvector of  $(\mathbf{GE})^T(\mathbf{GE})$  associated with  $\lambda_k^2$ .

The GEI in this model is thus expressed as a sum of components, each multiplied by  $\lambda_k$ , for a genotypic effect ( $\alpha_{ik}$ ) and an environment effect ( $\gamma_{jk}$ ). The term  $\lambda_k$  gives the proportion of the variance due to the GEI interaction in the  $k$ th component. The effects  $\alpha_{ik}$  and  $\gamma_{jk}$  represent weights for genotype  $i$  and environment  $j$ , in that component of the interaction ( $\lambda_k^2$ ). The rank of  $\mathbf{GE}$  is  $s = \min\{g - 1, e - 1\}$ , so the index  $k$  in the sum of multiplicative components can run from 1 to  $s$ . Use of all  $s$  components regains all the variation:  $SS(\text{GEI}) = \sum_{k=1}^s \lambda_k^2$  and the model is saturated so it produces an exact fit to the data, with no residual error term against which to test effects (except in the situation when an independent error is estimated). When  $m < s$ , the model is said to be truncated. However, for AMMI, one does not try to recoup the whole  $SS(\text{GEI})$  but only the components most strongly determined by genotypes and environments. Consequently, the index is generally set to run to  $m < s$ , so the estimates are obtained from the first  $m$  terms of the SVD of the  $\mathbf{GE}$  array (Good, 1969; Gabriel, 1978). This is a least-squares analysis that leaves an additional residual denoted by  $\rho_{ij}$ . Thus, the interaction of genotype  $i$  with environment  $j$  is described by  $\sum_{k=1}^m \lambda_k \alpha_{ik} \gamma_{jk}$ , discarding the noise given by  $\sum_{k=m+1}^s \lambda_k \alpha_{ik} \gamma_{jk}$ . Here, as in PCA, the components account, successively, for decreasing proportions of the variation present in the  $\mathbf{GE}$  array ( $\lambda_1^2 \geq \lambda_2^2 \geq \dots \geq \lambda_s^2$ ). Therefore, the AMMI method is seen as a procedure capable of separating signal and noise in the analysis of the GEI (Weber et al., 1996).

### Determining the Optimal Number of Multiplicative Terms in the AMMI Model

The main objective is the prediction of the true trait response in the cell of the two-way table of genotypes and environments. To achieve this, a truncated AMMI model should be used and thus criteria for determining the number of components needed to explain the pattern in the GEI term have been the objects of some research (Gollob, 1968; Mandel, 1971; Gauch and Zobel, 1988; Piepho, 1994, 1995; Cornelius, 1993; Cornelius et al., 1996).

Two basic approaches have evolved to determine the opti-

mal number of multiplicative terms to be retained in the GEI component. One approach uses a cross-validation method in which the data are randomly split into modeling data and validation data. AMMI is fitted to the modeling data and the mean squared errors of prediction (expressed as the root mean squared predictive difference, RMSPD) are determined from the validation data. The main criticism of this approach is that the best predictive model computed from a subset of data may not be the best model when all data are considered (Cornelius and Crossa, 1999); if crossvalidation is used in MET data, the data must be adjusted for replicate differences within environments (Cornelius and Crossa, 1999). The other approach for determining the best predictive truncated model is to use tests of hypotheses about the  $k$ th component,  $H_{0k}$ :  $\lambda_k = 0$ , using the complete data set (and not a subset like in the cross-validation approach). These tests are based on the sequential sum of squares explained by the multiplicative terms.

We will now briefly review these two approaches. It may be noted that shrinkage estimators of multiplicative models have been shown recently to be good predictors of cultivar performance in environments (Cornelius and Crossa, 1999), but these estimators require df estimates that the authors find problematic. Moreover, other classes of estimators can be better than shrinkage estimators (see Venter and Steel, 1993). So we do not consider them any further.

### Tests of Significance of Multiplicative Terms

The sequential sum of squares of the AMMI model for the  $k$ th component,  $S_k$ , is given by  $n \lambda_k^2$  for  $k = 1, 2, \dots, \text{rank}(\mathbf{GE})$  (where  $\mathbf{GE} = \bar{y}_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}$ ). As in PCA, all of the test criteria involve, at least indirectly, the ratio of the accumulated sum of squares for the first  $m$  components to the total SS(GEI), i.e.,  $\sum_{k=1}^m \lambda_k^2 / \text{SS}(\text{GEI})$ .

One of the usual procedures consists of determining the degrees of freedom associated with a particular component of SS(GEI) for each member of the family of AMMI models. This enables mean squares to be computed for each component, together with an error mean square. Since we have an orthogonal partition of the interaction sum of squares, the ratio of the mean square of any interaction component to the error mean square is then assumed to follow an  $F$  distribution with the corresponding degrees of freedom. This implicitly assumes a normal distribution for the original response variable, and enables individual interaction components to be subjected to significance tests. However, validity of the  $F$  distribution in these circumstances is subject to considerable doubt. The eigenvalues  $\lambda_k^2$  of the matrix  $(\mathbf{GE})(\mathbf{GE})^T$  (or  $(\mathbf{GE})^T(\mathbf{GE})$ ) are distributed as eigenvalues of a Wishart matrix but do not have a chi-square distribution. Since the  $S_k$  are not independent random variables following a chi-square distribution, an  $F$  test does not hold. Nonetheless, selection of the optimal model is often based on  $F$  tests for the successive terms of the interaction, the number of included terms corresponding to the number of significant components. The Gollob (1968) approximate  $F$  test assumes that  $n \hat{\lambda}_k^2 / \sigma^2$  is distributed as chi-square and so obviously does not hold. Computer simulations done by Cornelius (1993) showed that Gollob tests at the 0.05 level are very liberal with Type I error rate of 66% for testing  $H_{01}: \lambda_1 = 0$ . The  $F$ -approximation tests  $F_{\text{GHI}}$ ,  $F_{\text{GH2}}$  (Cornelius et al., 1992; Cornelius et al., 1993), effectively control Type I error rates, and are generally more parsimonious than the Gollob test. However, these tests are conservative for testing multiplicative terms for which the previous term is small. Simulation and iteration tests have greater power than the  $F_{\text{GHI}}$  and  $F_{\text{GH2}}$  tests with good control of Type I error rates. The

residual AMMI, collected in the last term of SS(GEI), can also be tested to confirm its nonsignificance.

Turning to the question of degrees of freedom, Gauch and Zobel (1996) mention some methods for attributing degrees of freedom to components of an AMMI model; those of Gollob (1968) and Mandel (1971) are particularly popular. However, the authors warn that, unfortunately, there are disagreements between these methods. Choosing one requires both theoretical and practical considerations. The approach of Gollob (1968) is very easily applied, since the number of degrees of freedom for component  $m$  of the interaction is simply defined to be  $\text{DF}(\text{IPCAm}) = g + e - 1 - 2m$ , whereas most other approaches require extensive simulations before they can be used.

For instance, Mandel (1971) defines the number of degrees of freedom for component  $k$  to be  $\text{DF}(\text{IPCk}) = E[\lambda_k^2] / \sigma^2$ , where  $\sigma^2$  is the population variance. However, simulations then have to be conducted to evaluate the number of degrees of freedom in particular cases. Mandel gives some tables derived from such simulations for a limited set of conditions, whereas Krzanowski (1979) gives some exact versions. These tables, however, are not exhaustive and this reduces the practical utility of the method. By contrast with Gollob (1968), Mandel's (1971) system generally results in a nonlinear decrease in the degrees of freedom for the successive interaction terms, which can still be fractions.

For some years, the degrees of freedom have been obtained by Mandel's (1971) proposal, which was considered exact and therefore correct. However, this proposal has received much criticism recently (e.g., Gauch, 1992), and it is now felt to be less appropriate than the approach of Gollob (1968). The reason for this criticism centers on the assumptions made by Mandel in his simulations, that the matrix contains only noise and not signals, whereas the presence of signal affects the component patterns substantially.

Gauch (1992) discusses the question of obtaining the degrees of freedom for the multiplicative components of an AMMI model. He concludes that rigorous simulations seem unnecessary or impractical, and generally recommends the use of Gollob's system when one is using an  $F$ -test approach, bearing in mind that the procedure is an intuitive guide. In cases where there seems to be a clear division between the large components determining the systematic part and small noise components, the author suggests that assigning equal degrees of freedom  $\{\text{DF}(\text{IPCk}) = [(g - 1)(e - 1)]/e\}$  is especially useful for early components because normally there will be little interest in partitioning the noise components. In addition, definitive questions of research require the exact assigning of the degrees of freedom to each multiplicative term. Therefore by Gollob's system, the full joint analysis of variance (computed from means) has the structure as shown in Table 1.

Piepho (1995) investigated the robustness (to the assumptions of homogeneity and normality of the errors) of some alternative tests to select an AMMI model. He comments that  $F$  tests applied in accordance with Gollob's (1968) criterion are liberal, in that they select too many multiplicative terms. Of the four methods he studied, including that of Gollob (1968), the test proposed by Cornelius et al. (1992) was the most robust. The author thus recommends that preliminary evaluations should be conducted to verify the validity of the assumptions if one of the other tests is to be used.

The Cornelius test statistic with  $m$  multiplicative terms in the model is as follows:

$$F_{R,m} = (\text{SS}(\text{GEI}) - \sum_{k=1}^m \lambda_k^2) / (f_2 \text{MS}(\text{Error mean}));$$



**Table 1. Full joint analysis of variance computed from averages using Gollob and Cornelius's system.**

Source of variation	Degrees of freedom	Sum of squares Gollob	DF† Cornelius	Sum of squares Cornelius
Genotypes (G)	$g - 1$	SS(G)	—	—
Environment (E)	$e - 1$	SS(E)	—	—
Interaction (GEI)	$(g - 1)(e - 1)$	SS(GEI)	—	—
IPCA 1	$g + e - 1 - (2 \times 1)$	$\lambda_1^2$	$(g - 1 - 1)(e - 1 - 1)$	$\sum_{k=2}^g \lambda_k^2$
IPCA 2	$g + e - 1 - (2 \times 2)$	$\lambda_2^2$	$(g - 1 - 2)(e - 1 - 2)$	$\sum_{k=3}^g \lambda_k^2$
IPCA 3	$g + e - 1 - (2 \times 3)$	$\lambda_3^2$	$(g - 1 - 3)(e - 1 - 3)$	$\sum_{k=4}^g \lambda_k^2$
IPCA s	$g + e - 1 - (2 \times s)$	$\lambda_s^2$	—	—
Error mean/n	$ge(n - 1)$	SS(Error mean)	—	—
Total	$gen - 1$	SS(Total)	—	—

† Degrees of freedom.

$$\text{with } f_2 = (g - 1 - m)(e - 1 - m).$$

This is the  $F_R$  test of Cornelius et al. (1992) that may turn out to be liberal as compared with  $F_{GH}$ ,  $F_{GH}$ , or simulation iteration tests. Under the null hypothesis that no more than  $m$  terms determine the interaction, the numerator (i.e., the residual SS(GEI) for the fitted AMMI model) is, approximately, a chi-square variable (Piepho, 1995) so the test statistic has an  $F$  distribution with  $f_2$  and Error mean square degrees of freedom.

Thus, a significant result for the test suggests that at least one more multiplicative term must be added to the  $m$  already included. It can therefore be seen as a test of significance of the first  $m + 1$  terms of the interaction (similar to the test of lack of fit in linear regression). When  $m = 0$ , i.e., when no multiplicative term is included, the test is just equivalent to the  $F$  test for global GEI in the joint ANOVA. It is an exact test. One also notices that the number of degrees of freedom of the numerator of  $F_R$  is equal to the degrees of freedom for the whole interaction minus the degrees of freedom attributed by Gollob (1968) for the  $m$  first terms. It is concluded, therefore, that the application of  $F_R$  is equivalent to the test of residual AMMI for GEI, as suggested.

### Predictive Assessment Using Cross Validation

Gauch and Zobel (1988) comment that evaluations such as those above by means of distributional assumptions via the  $F$  test can be termed "postdictive," in that they search for a model to explain a great part of the variation in the observed data (with high coefficient of determination). Thus, they argue, such methods are not efficient for selecting parsimonious models and are liable to include noise. By contrast, "predictive" criteria of evaluation capitalize on the ability of a model to form predictions with data not included in the analysis, simulating future responses not yet measured, so it would be preferable to base the model choice on such criteria.

To make predictions, in general, it is necessary to use computationally intensive statistical procedures. The less the model choice or assessment of performance of a predictor is based on distributional assumptions, the more general is the result. Thus, methods that are essentially data-based and free of theoretical distributions will have the greatest generality. Such methods involve resampling the given data set, using techniques such as the jackknife, the bootstrap and cross validation. Gauch (1988) introduced the name "predictive evaluation" when it is based on cross validation (Stone, 1974; Wold, 1978), and this is the principle underlying his proposal for selection of number of components in AMMI models.

In his method, the replications for each combination of genotypes and environments are randomly divided into two subgroups: (i) data for the fit of the AMMI model and (ii) data for validation. The responses are predicted for a family of AMMI models (i.e., for different values of  $m$ ) and these are compared with the respective validation data, calculating the differences between these values. Then, the sum of squares

of these differences is obtained and the result is divided by the number of predicted responses. This method was developed further by Crossa et al. (1991). The authors call the square root of this result the mean predictive difference (RMSPD) and suggest that the procedure be repeated about 10 times, getting an average of the results for each member of the family of models. A small value of RMSPD indicates predictive success of the model, so the best model is the one with smallest RMSPD. The chosen model is then used to analyze the data of all the  $n$  replications, jointly, in a definitive analysis.

Further modifications have been proposed in recent years. Piepho (1994) suggests obtaining the average value of RMSPD for 1000 different randomizations, instead of the 10 suggested by Crossa et al. (1991). The author considers a modification of the completely random partition of the data (modeling and validation) when the experiment is blocked. In this case, he recommends drawing entire blocks from the experiment and not making components for each combination of genotype and environment. Thus, the original block structure is preserved. However, despite the logical coherence of this type of proposal, studies confirming its effectiveness are still not available. Gauch and Zobel (1996) suggest that the validation data set should always be just one observation for each treatment. This is because that it is most likely, from  $n - 1$  data points, to find a model that is closest to the analysis of the full set of  $n$  data points. We thus take up this idea in the present contribution, and describe two methods that optimize the cross-validation process by validating the fit of the model on each data point in turn and then combining these validations into a single overall measure of fit.

Cornelius and Crossa (1999) used cross validation for comparing performance of shrinkage estimators, truncated multiplicative models and best linear unbiased predictor (BLUP) by computing the RMSPD as the square root of the mean squared difference between the predictive value and their corresponding validation data on replication adjusted data of five MET. The authors used a stopping rule for the number of crossvalidations that consisted in calculating the pooled mean square predictive difference on the  $m$ th execution of the loop (PMSPD<sub>m</sub>). The crossvalidation was terminated if the maximum absolute value of PMSPD<sub>m</sub> - PMSPD<sub>m-1</sub> was less than 0.01. The maximum number of cross validations required was 64 and the minimum was 39.

### Leave-One-Out Methods

We now propose two methods based on a full leave-one-out procedure that optimizes the cross-validation process. In the following, we assume that we wish to predict the elements  $x_{ij}$  of the matrix  $\mathbf{X}$  by means of the model:  $x_{ij} = \sum_{k=1}^m d_k u_{ik} v_{jk} + \varepsilon_{ij}$ . The methods are those outlined by Krzanowski (1987) and Gabriel (2002) respectively, in which we predict the value  $\hat{x}_{ij}^m$  of  $x_{ij}$  ( $i = 1, \dots, g; j = 1, \dots, e$ ) for each possible choice of  $m$

(the number of components), and measure the discrepancy between actual and predicted values as

$$\text{PRESS}(m) = \sum_{i=1}^g \sum_{j=1}^e (\hat{x}_{ij}^m - x_{ij})^2.$$

However, to avoid bias, the data point  $x_{ij}$  must not be used in the calculation of  $\hat{x}_{ij}^m$  for each  $i$  and  $j$ . Hence, appeal to some form of cross validation is indicated, and the two approaches differ in the way that they handle this. Both, however, assume that the SVD of  $\mathbf{X}$  can be written as  $\mathbf{X} = \mathbf{UDV}^T$ .

The standard cross-validation procedure is to subdivide  $\mathbf{X}$  into a number of groups, delete each group in turn from the data, evaluate the parameters of the predictor from the remaining data, and predict the deleted values (Wold, 1976, 1978). Krzanowski (1987) argued that the most precise prediction results when each deleted group is as small as possible, and in the present instance that means a single element of  $\mathbf{X}$ . Denote by  $\mathbf{X}^{(-i)}$  the result of deleting the  $i$ th row of  $\mathbf{X}$  and mean-centering the columns. Denote by  $\mathbf{X}_{(-j)}$  the result of deleting the  $j$ th column of  $\mathbf{X}$  and mean centering the columns, following the scheme given by Eastment and Krzanowski (1982). Then we can write

$$\mathbf{X}^{(-i)} = \bar{\mathbf{U}} \bar{\mathbf{D}} \bar{\mathbf{V}}^T \text{ with } \bar{\mathbf{U}} = (\bar{u}_{st}), \bar{\mathbf{V}} = (\bar{v}_{st}), \text{ and } \bar{\mathbf{D}} = \text{diag}(\bar{d}_1, \dots, \bar{d}_p),$$

and

$$\mathbf{X}_{(-j)} = \tilde{\mathbf{U}} \tilde{\mathbf{D}} \tilde{\mathbf{V}}^T \text{ with } \tilde{\mathbf{U}} = (\tilde{u}_{st}), \tilde{\mathbf{V}} = (\tilde{v}_{st}), \text{ and } \tilde{\mathbf{D}} = \text{diag}(\tilde{d}_1, \dots, \tilde{d}_{(p-1)}).$$

Now consider the predictor

$$\hat{x}_{ij}^m = \sum_{t=1}^m (\tilde{u}_{it} \sqrt{\tilde{d}_t}) (\bar{v}_{tj} \sqrt{\bar{d}_t})$$

Each element on the right-hand side of this equation is obtained from the SVD of  $\mathbf{X}$ , mean-centered after omitting either the  $i$ th row or the  $j$ th column. Thus, the value  $x_{ij}$  has nowhere been used in calculating the prediction, and maximum use has been made of the other elements of  $\mathbf{X}$ . The calculations here are exact, so there is no problem with convergence as opposed to expectation maximization approaches that have also been applied to AMMI, but are not guaranteed to converge.

Gabriel (2002), on the other hand, takes a mixture of regression and lower-rank approximation of a matrix as the basis for his prediction. The algorithm for cross-validation of lower rank approximations proposed by the author is as follows:

For given (GEI) matrix  $\mathbf{X}$ , use the partition

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_{11} & \mathbf{x}_1^T \\ \mathbf{x}_{.1} & \mathbf{X}_{.11} \end{bmatrix}$$

and approximate the submatrix  $\mathbf{X}_{.11}$  by its rank  $m$  fit using the SVD

$$\mathbf{X}_{.11} = \sum_{k=1}^m \mathbf{u}_{(k)} \mathbf{d}_k \mathbf{v}_{(k)}^T = \mathbf{UDV}^T,$$

where  $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_m]$ ,  $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_m]$ , and  $\mathbf{D} = \text{diag}(d_1, \dots, d_m)$ . Then predict  $x_{11}$  by

$$\hat{x}_{11} = \mathbf{x}_1^T \mathbf{VD}^{-1} \mathbf{U}^T \mathbf{x}_{.1}$$

and obtain the cross-validation residual  $e_{11} = x_{11} - \hat{x}_{11}$ .

Similarly obtain the cross-validation fitted values  $\hat{x}_{ij}$  and residuals  $e_{ij} = x_{ij} - \hat{x}_{ij}$  for all other elements  $x_{ij}$ ,  $i = 1, \dots, g$ ;  $j = 1, \dots, m$ ;  $(i, j) \neq (1, 1)$ . Each will require a different partition of  $\mathbf{X}$ .

These residuals and fitted values can be summarized by

$\text{PRESS}(m) = \frac{1}{ge} \sum_{i=1}^g \sum_{j=1}^e e_{ij}^2$  and  $\text{PRECORR}(m) = \text{Corr}(x_{ij}, \hat{x}_{ij} | \forall i, j)$  respectively.

With either method, choice of  $m$  can be based on some suitable function of

$$\text{PRESS}_-(m) = \frac{1}{ge} \sum_{i=1}^g \sum_{j=1}^e (\hat{x}_{ij}^m - x_{ij})^2.$$

However, the features of this statistic differ for the two methods. Gabriel's approach yields values that first decrease and then (usually) increase with  $m$ . He therefore suggests that the optimum value of  $m$  is the one that yields the minimum of the function. The Eastment-Krzanowski approach produces (generally) a set of values that is monotonically decreasing with  $m$ . They therefore argue for the use of

$$W_m = \frac{\text{PRESS}_-(m-1) - \text{PRESS}_-(m)}{D_m} \div \frac{\text{PRESS}_-(m)}{D_r},$$

where  $D_m$  is the number of degrees of freedom required to fit the  $m$ th component and  $D_r$  is the number of degrees of freedom remaining after fitting the  $m$ th component. Consideration of the number of parameters to be estimated together with all the constraints on the eigenvectors at each stage, shows that  $D_m = g + e - 2m$ .  $D_r$  can be obtained by successive subtraction, given  $(g-1)e$  degrees of freedom in the mean-centered matrix  $\mathbf{X}$ , i.e.,  $D_1 = (g-1)e$  and  $D_r = D_{r-1} - [g + e - (m-1)2]$ ,  $r = 2, 3, \dots, (g-1)$ , (Wold, 1978).  $W_m$  represents the increase in predictive information supplied by the  $m$ th component, divided by the mean predictive information in each of the remaining components. Thus, "important" components should yield values of  $W_m$  greater than unity. Basing choice of  $m$  on  $W_m$  in this way can thus be seen as a natural counterpart to the selection of a best set of orthogonal regressor variables in multiple regression analysis.

On a computational level, the best accuracy of prediction seems to be achieved when the entries  $(x_{ij})$  in different columns of  $\mathbf{X}$  are comparable in size and there is relatively little variation among the  $d_i$ . The most stable procedure is thus one in which the mean  $\bar{x}_j$  and standard deviation  $s_j$  of column  $j$  ( $j = 1, \dots, e$ ) are first found from the values present in that column. Existing entries  $x_{ij}$  of  $\mathbf{X}$  are then standardized to  $\hat{x}_{ij} = (x_{ij} - \bar{x}_j)/s_j$ , estimates are found by applying  $\hat{x}_{ij} = \mathbf{x}_i^T \mathbf{VD}^{-1} \mathbf{U}^T \mathbf{x}_{.j}$  to the standardized data, and then the final values are obtained from  $\hat{x}_{ij} = \bar{x}_j + s_j \hat{x}_{ij}$ .

Turning to the case of genotype-environment data, it would appear that  $\mathbf{X}$  should be the array of interactions previously denoted  $\mathbf{GE}$ . However, since we are merely looking for the appropriate number of multiplicative terms in the model, and any additive constants can be absorbed into the  $\epsilon_{ij}$  component of the model, we can apply the leave-one-out procedure directly to the data matrix  $\mathbf{Y}$ . Indeed, this may often be preferable given the small values taken by most elements of  $\mathbf{GE}$ .

Cornelius et al. (1993) compared results of cross validation with those obtained after computing the PRESS statistics in multiplicative models in a complete MET data. The data splitting involved three replicates for modeling and one replicate for validation. They computed the RMSPD of PRESS by adjusting the value of PRESS as  $[\text{PRESS}/ge + 3s^2/4]^{1/2}$ , where  $g$  and  $e$  denote the number of genotypes and sites in the MET and  $s^2$  is the pooled within-site error variance. The term in  $s^2$  is an adjustment for the difference in variance of the validation data on cell means, to make results comparable to the RMSPD from 3-1 data splitting. Results on an MET with nine genotypes and twenty sites showed that PRESS is more sensitive to

**Table 2. RMSPD from 3-1 crossvalidation and adjusted RMSPD(PRESS) for models fitted to a multi-environment trial.**

Terms	Model form†				
	AMMI	GREG	SREG	COMM	SHMM
Data splitting					
0	980	—	—	—	—
1	915	954	908	962	947
2	934	907	935	911	906
3	951	926	947	930	924
4	955	946	949	951	944
5	963	957	967	957	959
Adjusted RMSPD(PRESS)‡					
0	970	—	—	—	—
1	956	939	892	942	925
2	2 725	912	9 155	935	886
3	30 708	994	7 071	1 557	925
4	19 030	1 071	14 670	2 682	2 246
5	30 540	3 251	8 165	8 688	5 094

† AMMI: Additive main effect and multiplicative interaction model; GREG: Genotype regression model; SREG: Sites regression model; COMM: Completely multiplicative model; SHMM: Shifted multiplicative model.

‡ RMSPD: Root mean square predictive difference; PRESS: Predictive sum of squares.

overfitting than is data splitting. Table 2 shows that PRESS differentiates more clearly the model forms than does data splitting. For some model forms (SHMM and SREG), the model with smallest PRESS predicted data in a deleted cell better than they were predicted by three replicates of data with all cells present. On the other hand, many overfitted models gave very unreliable prediction of a deleted cell.

## RESULTS

### Illustrative Data Sets

Krzanowski (1988) considered a simple multivariate data set from Kendall (1980, p. 20), relating to 20 samples of soil and five variables: percentages of sand content, silt content, and clay content; organic matter; and pH. The author called attention to the fact that the three percentages added to 100 so that applying any regression-based technique to the raw data would incur multicollinearity problems, but the singular-value approach could be applied directly without any computational drawbacks.

Jeffers (1967) described two detailed multivariate case studies, one of which concerned 19 variables measured on each of 40 winged adelgids that had been caught in a light trap. Of the 19 variables, 14 are length or width measurements, four are counts, and one (anal fold) is a presence/absence variable score 0 or 1.

In Table 3, we show a comparison between the Eastment–Krzanowski and Gabriel methods for the soil data. In this case, when the data matrix was standard-

ized, the rank of the resulting matrix was four and all submatrices associated with columns four and five had a singular matrix **D**. Thus we used a Moore–Penrose generalized inverse instead of the ordinary inverse when implementing Gabriel's method. We can see that the PRESS(*m*) values are much lower with the Eastment–Krzanowski method than with the Gabriel method up as far as *m* = 4.

This shows that the Gabriel criterion is sensitive to the appearance of a singular matrix **D**. On the other hand, the use of the Eastment–Krzanowski criterion *W* suggests two components in the model, whereas the Gabriel approach indicates that all four are needed. In this case, both methods yield similar PRECORR values for all components.

In Table 4, we show a comparison between both approaches using the Jeffers data. Now, with standardized matrix **X**, no singular submatrices appeared. We can see from PRESS(*m*) that the Gabriel values are lower until the third component. From the *W* statistic, we can see that the Eastment–Krzanowski method suggests that four components should be retained in the model, whereas the Gabriel criterion suggests two.

### Genotype × Environment Examples

Vargas and Crossa (2000) present a complete data set from a wheat (*Triticum aestivum* L.) variety trial with eight genotypes tested during six years (1990–1995) in Cd. Obregón, Mexico. In each year, the genotypes were arranged in a complete block design with three replicates. The eight genotypes correspond to a historical series of cultivars released from 1960 to 1980. We divided the original data by 1000, and analyzed the mean grain yields (kg ha<sup>-1</sup>). Results of analysis of variance incorporating both the Gollob and Cornelius *F* tests are shown in Table 5.

This table shows that genotypes, years, and GEI are highly significant (*P* < 0.01) and account for 39.29, 45.20, and 15.51% of the treatment sum of squares, respectively. At the 1% significance level, both the Gollob and the Cornelius *F* tests indicate that the first two interaction components (IPCA1 and IPCA2) should be included in the model.

The result of a full cross validation across 1000 randomizations of the data can be seen in the first three columns of Table 6. For each randomization, one of the three observations at each treatment combination was randomly selected and used to create the interaction matrix, whereas the other two replicates were averaged to form the validation data. Also shown in the remaining columns of Table 6 are the Eastment–Krzanowski and

**Table 3. Data on twenty samples of soil and five variables (from Kendall, 1980, p. 20, based on Krzanowski, 1988).**

Rank <i>m</i>	Eastment-Krzanowski			Gabriel		
	PRESS <sub><i>m</i></sub> †	PRECORR	W	PRESS <sub><i>m</i></sub>	PRECORR	W
1	4.36	0.9963	27.78	8.08	0.9932	13.60
2	2.23	0.9981	2.14	7.45	0.9937	0.18
3	2.14	0.9982	0.05	5.60	0.9952	0.45
4	2.13	0.9982	0.00	0.21	0.9998	10.20

† PRESS: Predictive sum of squares; PRECORR: Predictive correlation; W: Eastment–Krzanowski criterion.

**Table 4. Data on forty winged aphids and nineteen variables (from Jeffers, 1967, based on Krzanowski, 1987).**

Rank m	Eastment–Krzanowski			Gabriel		
	PRESS <sub>m</sub> †	PRECORR	W	PRESS <sub>m</sub>	PRECORR	W
1	0.4500	0.9799	29.04	0.4240	0.9810	31.56
2	0.3391	0.9849	3.71	0.2883	0.9871	5.34
3	0.3389	0.9849	0.00	0.2934	0.9869	−0.18
4	0.2865	0.9874	1.85	0.2957	0.9868	−0.07
5	0.2823	0.9876	0.14	0.3031	0.9864	−0.23
6	0.2815	0.9876	0.02	0.3096	0.9862	−0.18
7	0.2760	0.9878	0.16	0.3117	0.9861	−0.05
8	0.2723	0.9880	0.10	0.3239	0.9855	−0.28
9	0.2679	0.9882	0.11	0.3668	0.9836	−0.80
10	0.2677	0.9882	0.00	0.3589	0.9839	0.13
11	0.2666	0.9883	0.02	0.3687	0.9835	−0.14
12	0.2651	0.9883	0.02	0.4222	0.9812	−0.59
13	0.2640	0.9884	0.01	0.4842	0.9786	−0.50
14	0.2622	0.9885	0.02	0.5039	0.9776	−0.12
15	0.2616	0.9885	0.00	0.4986	0.9778	0.02
16	0.2610	0.9885	0.00	0.5004	0.9778	−0.00
17	0.2604	0.9885	0.00	0.5443	0.9759	−0.03
18	0.2601	0.9886	−0.00	0.5778	0.9744	0.03

† PRESS: Predictive sum of squares; PRECORR: Predictive correlation; W: Eastment–Krzanowski criterion.

Gabriel leave-one-out results, as obtained from the single matrix of averages across the three replicates. The full cross validation yields minimum RMSPD at four components, although the RMSPD values for three, four, and five components are very similar in size and any could be chosen to represent the optimum number of components for the model. However, both the Eastment–Krzanowski and the Gabriel methods suggest that one component should be retained in the model.

To obtain a broader comparison of methods, we turn to the data sets in Cornelius and Crossa (1999) who describe five multienvironment international cultivar trials, all in randomized block designs. Trial 1 was a wheat trial with 19 durum wheat cultivars, one bread wheat cultivar, and 34 sites. Trials 2 to 5 were maize (*Zea mays* L.) trials with numbers of cultivars and sites equal to (16,24), (9,20), (18,30), and (8,59), respectively. For each trial, the number of multiplicative terms was obtained from a range of methods and the results are given in Table 7.

## DISCUSSION

The three distinct approaches to selection of number of multiplicative interaction components have yielded different results on the data of Vargas and Crossa (2000).

Distributional  $F$  tests indicate two components as optimum; randomization cross validation suggests three or four, whereas leave-one-out methods indicate just one important component. So how do we assess these differences?

The first point to make is that the  $F$ -test methods all rely heavily on distributional assumptions (normality of data and validity of  $F$  distributions for mean squares), which may not be appropriate in many cases. Also, it is documented that the different  $F$  tests can come up with conflicting recommendations on a particular data set (Duarte and Vencovsky, 1999), while Piepho (1995) has noted that some of the tests select too many interaction components. This feature can be seen clearly in the comparisons of Table 7 also. So, in general, it seems that a data-based cross-validation method should be more appropriate.

Turning then to the full cross-validation randomization approach, the weakness here is that a large portion of the data must be set aside for the validation set. This means that the model is fitted to only a relatively small part of the data. For example, in the analysis reported in Table 6, the fit was to just one observation at each genotype–environment combination, while the assessment was on the mean of the other two replicates. Be-

**Table 5. Additive main effects and multiplicative interaction analysis of the Vargas and Crossa (2000) data, up to the first five interaction principal component analysis (IPCA).**

Source of variation	Sum of squares	DF†	F <sub>Gollob</sub>	Sum of squares	DF <sub>Cornelius</sub>	F <sub>Cornelius</sub>
Block	0.2001	2	0.63	—	—	—
Treatment	108.8393	47	14.65**	—	—	—
Genotypes (G)	42.7587	7	38.65**	—	—	—
Years (E)	49.1997	5	62.27**	—	—	—
Interaction (GEI)	16.8809	35	3.05**	—	—	—
IPCA 1	7.2428	11	4.16**	9.6379	24	2.54**
IPCA 2	5.4232	9	3.81**	4.2147	15	1.78*
IPCA 3	2.9696	7	2.68*	1.2451	8	0.98
IPCA 4	1.1906	5	1.50	0.0545	3	0.12
IPCA 5	0.0545	3	0.11	—	—	—
Error	14.8543	94	—	—	—	—
Correct Total	123.8939	143	—	—	—	—

\* The 0.05 probability level.

\*\* Significant at the 0.01 probability level.

† DF: degrees of freedom.



**Table 6. Cross-validation data analysis and leave-one-out method on the Vargas and Crossa (2000) data.**

Rank m	Randomisation Cross-validation		Eastment–Krzanowski		Gabriel	
	RMSPD†	PRECORR	PRESS_m	W	PRESS_m	W
0	0.5040	0.8436	—	—	—	—
1	0.5149	0.8386	0.1861	2.8587	0.1886	2.7882
2	0.4968	0.8521	0.1989	–0.1029	0.2020	–0.1057
3	0.4830	0.8617	0.1721	0.1167	0.2610	–0.1695
4	0.4776	0.8655	0.1615	–0.0218	0.3543	0.0877
5	0.4812	0.8635	0.1394	–0.3171	0.5285	0.6592

† RMSPD: Root mean square predictive difference; PRECORR: Predictive correlation; PRESS: Predictive sum of squares; W: Eastment–Krzanowski criterion.

tween-replicate variation may generally be very high, which inflates assessment error sums of squares and has probably contributed to the high number of components selected by this method.

By contrast, the leave-one-out methods make the most efficient use of the data and result in the most parsimonious model (AMMI 1) for the example of Tables 5 and 6. This model has 23 df (5 for years plus 7 for genotypes plus 11 for interaction PCA component 1) and is twice as parsimonious as AMMI 5 (in the sense that AMMI 5 contains twice as many degrees of freedom as AMMI 1). Thus, we conclude that a final model may be constructed by applying AMMI 1 to all the data (i.e., all three replications). The first interaction component recovers 43% of the GEI SS in only 31.4% of the interaction df (Table 5). The higher interaction components are judged by predictive assessment to be just noise for the purpose of yield prediction, and thus may be pooled with the residual.

The feature of parsimony is illustrated most clearly by the results of the Eastment–Krzanowski method in Table 7. Trials 1 through 4 are complex trials in which there is clear evidence of GEI, whereas Trial 5 is much simpler and probably free of interaction. From a practitioner's point of view, capturing the essence of any interaction in relatively few components is an attraction as these components can be interpreted clearly, whereas fitting many components may create problems of interpretation. Most of the methods shown in Table 7 exhibit quite a large variability in the number of components selected, with large numbers in some data sets for each method. Such large numbers are undesirable in practice. At the other extreme, PRESS provides a maximum of

one component for several complex trials in which the interaction structure is evidently not so straightforward, and several trials with no components including one (Trial 5) in which there is clear evidence of interaction. By contrast, the Eastment–Krzanowski method provides a stable pattern of relatively low and hence interpretable numbers of components.

In summary, therefore, distributional *F* tests are often based on questionable assumptions while full cross-validation randomization methods remove too much of the available data for validation purposes and hence lead to less reliable fitted models. Use of leave-one-out method is therefore recommended in general. Of the two such methods investigated here, the Eastment–Krzanowski method, has shown the greater parsimony and stability of fitted model.

## ACKNOWLEDGMENTS

The authors thank Dr. José Crossa for his very generous contributions that significantly improved the draft, and Dr. João Batista Duarte for his constructive criticism. This research was financially supported by FAPESP proc. 00/12292-1.

## REFERENCES

- Cornelius, P.L. 1993. Statistical tests and retention of terms in the additive main effects and multiplicative interaction model for cultivar trials. *Crop Sci.* 33:1186–1193.
- Cornelius, P.L., and J. Crossa. 1999. Prediction assessment of shrinkage estimators of multiplicative model for multi-environment cultivar trials. *Crop Sci.* 39:998–1009.
- Cornelius, P.L., J. Crossa, and M.S. Seyedsadr. 1993. Tests and estimators of multiplicative models for variety trials. p. 156–166. *In* Proceedings of Annual Kansas State University Conference on Applied Statistics in Agriculture, 5th., Manhattan, KS. 25–27 Apr. 1993. Dep. of Statistics, Kansas State Univ., Manhattan, KS.
- Cornelius, P.L., J. Crossa, and M.S. Seyedsadr. 1996. Statistical tests and estimators of multiplicative models for genotype-by-environment interaction. p. 199–234. *In* M.S. Kang and H.G. Gauch (ed.) *Genotype-by-environment interaction*. CRC Press, Boca Raton, FL.
- Cornelius, P.L., M. Seyedsadr, and J. Crossa. 1992. Using the shifted multiplicative model to search for “separability” in crop cultivar trials. *Theor. Appl. Genet.* 84:161–172.
- Crossa, J. 1990. Statistical analyses of multilocation trials. *Adv. Agron.* 44:55–85.
- Crossa, J., P.L. Cornelius, and W. Yan. 2002. Biplot of linear-bilinear models for studying crossover genotype  $\times$  environment interaction. *Crop Sci.* 42:619–633.
- Crossa, J., P.N. Fox, W.H. Pfeifer, S. Rajaram, and H.G. Gauch. 1991. AMMI adjustment for statistical analysis of an international wheat yield trial. *Theor. Appl. Genet.* 81:27–37.
- Duarte, J.B., and R. Vencovsky. 1999. Interação genótipos  $\times$  ambientes—uma introdução à análise “AMMI”. Ribeirão Preto, S.P.
- Eastment, H.T., and W.J. Krzanowski. 1982. Cross-validatory choice of the number of components from a principal component analysis. *Technometrics* 24:73–77.

**Table 7. Number of AMMI multiplicative terms in five data sets that are statistically significant for various tests, and using the PRESS, crossvalidation, Eastment–Krzanowski, and the Gabriel criteria.**

Test†	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5
JG	4	1	1	4	0
AL	4	4	4	7	0
F <sub>GHI</sub>	5	5	2	7	2
F <sub>R</sub>	5	6	2	8	3
PRESS	1	1	1	0	0
Crossvalidation	4	8	1	10	0
Eastment–Krzanowski	2	2	2	1	1
Gabriel	5	6	7	4	1

† JG = Seyedsadr–Cornelius/Johnson–Graybill/Schot–t–Marasinghe test.  
 AL = Anderson–Lawley test of equality of the last  $p - k + 1$  principal components (Jackson, 1991, Section 4.4.1)  
 F<sub>GHI</sub> = approximate sequential tests against the pure error based on the Goodman–Haberman theorem (Cornelius, 1993).  
 F<sub>R</sub> = test of the residual mean square.



- Eberhart, S.A., and W.A. Russell. 1966. Stability parameters for comparing varieties. *Crop Sci.* 6:36–40.
- Finlay, K.W., and G.N. Wilkinson. 1963. The analysis of adaptation in a plant-breeding programme. *Austr. J. Agric. Res.* 14:742–754.
- Gabriel, K.R. 1978. Least squares approximation of matrices by additive and multiplicative models. *J. Roy. Stat. Soc. Series B* 40:186–196.
- Gabriel, K.R. 2002. Le biplot—outil d'exploration de données multidimensionnelles. *Journal de la Societe Francaise de Statistique* 143 (to appear).
- Gauch, H.G. 1988. Model selection and validation for yield trials with interaction. *Biometrics* 44:705–715.
- Gauch, H.G. 1992. Statistical analysis of regional yield trials; AMMI analysis of factorial designs. Elsevier Science, New York.
- Gauch, H.G., and R.W. Zobel. 1988. Predictive and postdictive success of statistical analysis of yield trials. *Theor. Appl. Genet.* 76:1–10.
- Gauch, H.G., and R.W. Zobel. 1996. AMMI analysis of yield trials. p. 85–122. *In* M.S. Kang and H.G. Gauch (ed.) *Genotype by environment interaction*. CRC Press, Boca Raton, FL.
- Gollob, H.F. 1968. A statistical model which combines features of factor analytic and analysis of variance techniques. *Psychometrika* 33(1):73–115.
- Good, I.J. 1969. Some applications of the singular decomposition of a matrix. *Technometrics* 11(4):823–831.
- Jackson, J.E. 1991. A user's guide to principal components. Wiley and Sons, New York.
- Jeffers, J.N.R. 1967. Two case studies in the application of principal component analysis. *Appl. Stat.* 16:225–236.
- Kang, M.S., and R. Magari. 1996. New developments in selecting for phenotypic stability in crop breeding. p. 1–14. *In* M.S. Kang and H.G. Gauch (ed.) *Genotype by environment interaction*. CRC Press, Boca Raton, FL.
- Kendall, M.G. 1980. *Multivariate analysis* (2nd ed.). Charles Griffin & Co., London.
- Krzanowski, W.J. 1979. Some exact percentage points of a statistic useful in analysis of variance and principal component analysis. *Technometrics* 21:261–263.
- Krzanowski, W.J. 1987. Cross-validation in principal component analysis. *Biometrics* 43:575–584.
- Krzanowski, W.J. 1988. Missing value imputation in multivariate data using the singular value decomposition of a matrix. *Listy Biometryczne-Biometrical Letters* XXV (1,2):31–39.
- Mandel, J. 1961. Non-additivity in two-way analysis of variance. *J. Am. Statist. Assoc.* 56:878–888.
- Mandel, J. 1969. The partitioning of interaction in analysis of variance. *J. Res. Int. Bur. Stand. Sect. B* 73:309–328.
- Mandel, J. 1971. A new analysis of variance model for non-additive data. *Technometrics* 13(1):1–18.
- Piepho, H.P. 1994. Best linear unbiased prediction (BLUP) for regional yield trials: a comparison to additive main effects and multiplicative interaction (AMMI) analysis. *Theor. Appl. Genet.* 89:647–654.
- Piepho, H.P. 1995. Robustness of statistical test for multiplicative terms in additive main effects and multiplicative interaction model for cultivar trial. *Theor. Appl. Genet.* 90:438–443.
- Seyedsadr, M., and P.L. Cornelius. 1992. Shifted multiplicative models for nonadditive two-way tables. *Commun. Stat. B Simul. Comp.* 21:807–832.
- Stone, M. 1974. Cross-validated choice and assessment of statistical predictions (with Discussion). *J. Roy. Stat. Soc. Series B* 36:111–148.
- Tukey, J.W. 1949. One degree of freedom for non-additivity. *Biometrics* 5:232–242.
- Vargas, M.V., and J. Crossa. 2000. The AMMI analysis and graphing the biplot. CIMMYT, INT., Mexico.
- Venter, J.H., and S.J. Steel. 1993. Simultaneous selection and estimation for the some zeros family of normal models. *J. Statist. Computation Simulation*. 45:129–146.
- Weber, W.E., G. Wricke, and T. Westermann. 1996. Selection of genotypes and predictions performance and analysing genotype-by-environment interactions. p. 353–371. *In* M.S. Kang and H.G. Gauch (ed.) *Genotype by environment interaction*. CRC Press, Boca Raton, FL.
- Wold, S. 1976. Pattern recognition by means of disjoint principal component models. *Pattern Recognition* 8:127–139.
- Wold, S. 1978. Cross-validated estimation of the number of components in factor and principal component models. *Technometrics* 20:397–405.
- Yates, F., and W.G. Cochran. 1938. The analysis of groups of experiments. *J. Agric. Sci.* 28:556–580.
- Zobel, R.W., M.J. Wright, and H.G. Gauch, Jr. 1988. Statistical analysis of a yield trial. *Agron. J.* 80:388–393.