Parametric Bootstrap Methods for Testing Multiplicative Terms in GGE and AMMI Models

Johannes Forkman
Department of Crop Production Ecology, Swedish University of Agricultural Sciences
PO Box 7043, 750 07 Uppsala, Sweden
email: johannes.forkman@slu.se

and

Hans-Peter Piepho
Institute of Crop Science, University of Hohenheim, 70 593 Stuttgart, Germany
email: hans-peter.piepho@uni-hohenheim.de

SUMMARY: The **genotype main effects and genotype-by-environment interaction effects** (GGE) model and the **additive main effects and multiplicative interaction** (AMMI) model are two common models for analysis of genotype-by-environment data. These models are frequently used by agronomists, plant breeders, geneticists and statisticians for analysis of multi-environment trials. In such trials, a set of genotypes, e.g. crop cultivars, are compared across a range of environments, e.g. locations. The GGE and AMMI models use singular value decomposition to partition genotype-by-environment interaction into an ordered sum of multiplicative terms. This article deals with the problem of testing the significance of these multiplicative terms in order to decide how many terms to retain in the final model. We propose parametric bootstrap methods for this problem. Models with fixed main effects, fixed multiplicative terms and random normally distributed errors are considered. Two methods are derived: a **full** and a **simple** parametric bootstrap method. These are compared with the alternatives of using approximate F-tests and cross-validation. In a simulation study based on four multi-environment trials, both bootstrap methods performed well with regard to Type I error rate and power. The simple parametric bootstrap method is particularly easy to use, since it only involves repeated sampling of standard normally distributed values. This method is recommended for selecting the number of multiplicative terms in GGE and AMMI models. The proposed methods can also be used for testing components in principal component analysis.

**Key words:** AMMI; Genotype-environment interaction; GGE; Multi-environment trials, Principal component analysis; Singular value decomposition.

1. Introduction

When differences between genotypes depend on environments, genotype-by-environment interaction is present. Genotype-by-environment interaction is studied in many branches of biology, not least in agriculture. In plant breeding and crop variety experimentation, cultivars or potential cultivars are commonly investigated at several environmentally different locations. For analysis of such data, two biadditive models (Denis and Gower, 1994) are especially common: the **genotype main effects and genotype-by-environment interaction effects** (GGE) model (Yan et al. 2000, Yan and Kang, 2002), and the **additive main effects and multiplicative interaction** (AMMI) model (Gauch 1988, 1992).

Both GGE and AMMI explore a matrix of genotype-by-environment means, using a combination of analysis of variance (ANOVA) and singular value decomposition. With GGE, singular value decomposition is performed on the matrix of residuals from a one-way ANOVA with fixed effects for environments. With AMMI, singular value decomposition is performed on residuals from a two-way ANOVA with fixed effects of genotypes and environments. Cornelius, Crossa and Seyedsadr (1996) called the GGE model a sites regression model (SREG). For a discussion about the relative merits of GGE and AMMI, see Gauch (2006), Yan et al. (2007) and Gauch, Piepho and Annicchiarico (2008). The present paper considers GGE and AMMI models with fixed main effects and normally distributed errors, although extensions have been made to mixed models (Piepho, 1997, 1998; Smith, Cullis and Thompson, 2001), other distributions (Tsujitani, 1992; van Eeuwijk, 1995) and three-way interactions (van Eeuwijk and Kroonenberg, 1998).

The result of the singular value decomposition is often presented in a biplot illustrating the first two multiplicative terms of the singular value decomposition. With GGE, such a biplot presents a rank-two approximation of the sum of genotype effects and genotype-by-environment interaction effects, which is a useful and popular tool for breeders (Yan and Tinker, 2006). With AMMI, genotype-by-environment interaction is studied separately from main effects of genotypes. Figure 1 is a biplot (see Section 3 for details) for an AMMI analysis of...

The definite version of this article is available at onlinelibrary.wiley.com
For replicated data, Gollob (1968) proposed an F-test for selecting the number of multiplicative terms. According to this method, the mean square of the multiplicative term is divided by the error mean square and compared with an F-distribution, similarly as with ANOVA. It is well-known that this method is too liberal (Cornelius et al., 1996). Cornelius, Seyedsadr and Crossa (1992), Cornelius (1993), and Piepho (1995) proposed and investigated various other F-tests for replicated data. In these tests, an error mean square is calculated from replicates within genotype-by-environment combinations. When there is only a single observation for each cell of the genotype-by-environment table, these methods do not apply.

The present paper considers using resampling for the problem of separating fixed genotype-by-environment interaction from random noise. Following Mandel (1971), it is suggested that the fixed interaction be modelled by the first $K$ terms of the singular value decomposition. Random noise is estimated as the remainder when $K$ is smaller than the maximum possible number of terms. The proposed methods for selecting $K$ use the parametric bootstrap technique (Efron and Tibshirani, 1993). With this approach, the distribution of the observed data is assumed to belong to a parametric family, and the expected value follows a statistical model, specifically the GGE or AMMI models. The model parameters are estimated from the observed data, and random samples are generated from the estimated model. Based on these so-called bootstrap samples, distributions of test statistics or other statistics can be approximated. In the present application, the distribution of a test statistic, $T$, for the significance of the $(K + 1)$th multiplicative term is approximated and used as a reference distribution for calculation of the $p$-value. When $K = 0$, this test statistic is the same as the likelihood ratio statistic used by Johnson and Graybill (1972), and when $K \geq 0$, the test statistic is the same as the likelihood ratio statistic that was derived by Yochmowitz and Corneli (1978) and advocated by Schott (1986). When $K \geq 1$, an exact test based on $T$ is not possible, since the distribution of $T$ depends upon unknown singular values (Schott, 1986). In the context of hypothesis testing, these singular values are nuisance parameters that are not specified by the null hypothesis. The null hypothesis is that the interaction be modelled by the first $K$ terms. For replicated data, Gollob (1968) proposed an F-test for selecting the number of multiplicative terms. According to this method, the mean square of the multiplicative term is divided by the error mean square and compared with an F-distribution, similarly as with ANOVA. It is well-known that this method is too liberal (Cornelius et al., 1996). Cornelius, Seyedsadr and Crossa (1992), Cornelius (1993), and Piepho (1995) proposed and investigated various other F-tests for replicated data. In these tests, an error mean square is calculated from replicates within genotype-by-environment combinations. When there is only a single observation for each cell of the genotype-by-environment table, these methods do not apply.

The present paper considers using resampling for the problem of separating fixed genotype-by-environment interaction from random noise. Following Mandel (1971), it is suggested that the fixed interaction be modelled by the first $K$ terms of the singular value decomposition. Random noise is estimated as the remainder when $K$ is smaller than the maximum possible number of terms. The proposed methods for selecting $K$ use the parametric bootstrap technique (Efron and Tibshirani, 1993). With this approach, the distribution of the observed data is assumed to belong to a parametric family, and the expected value follows a statistical model, specifically the GGE or AMMI models. The model parameters are estimated from the observed data, and random samples are generated from the estimated model. Based on these so-called bootstrap samples, distributions of test statistics or other statistics can be approximated. In the present application, the distribution of a test statistic, $T$, for the significance of the $(K + 1)$th multiplicative term is approximated and used as a reference distribution for calculation of the $p$-value. When $K = 0$, this test statistic is the same as the likelihood ratio statistic used by Johnson and Graybill (1972), and when $K \geq 0$, the test statistic is the same as the likelihood ratio statistic that was derived by Yochmowitz and Corneli (1978) and advocated by Schott (1986). When $K \geq 1$, an exact test based on $T$ is not possible, since the distribution of $T$ depends upon unknown singular values (Schott, 1986). In the context of hypothesis testing, these singular values are nuisance parameters that are not specified by the null hypothesis. The full parametric bootstrap method of the present paper resolves this problem by substituting estimates for unknown parameters. Utilizing an approximate result for distributions of eigenvalues (Muirhead, 1978), a simplified version of the full parametric bootstrap method is derived. This simple parametric bootstrap method is particularly easy to apply, because it uses standard normally distributed values as bootstrap samples. Thus, with this method it is not necessary to estimate the parameters; it suffices to assume that errors are normally distributed. The proposed bootstrap methods for GGE and AMMI can with small adjustments also be used for the the completely multiplicative model (COMM), and the genotypes regression model (GREG) (Cornelius et al. 1996). Furthermore, the methods may be used for testing components in principal component analysis (PCA).
2. Methods

2.1 Statistical Model

Assume that \( J \) genotypes have been investigated in \( I \) environments. Let \( y_{ij} \) denote the observed mean yield of the \( j \)th genotype in the \( i \)th environment (\( i = 1, 2, \ldots, I; j = 1, 2, \ldots, J \)). Let \( Y \) denote the \( I \times J \) matrix of observations and \( E \) denotes a matrix of interactions and residual errors. Let \( \alpha = (\alpha_1, \alpha_2, \ldots, \alpha_J)^T \) be a vector of environment main effects, and \( \beta = (\beta_1, \beta_2, \ldots, \beta_I)^T \) a vector of genotype main effects. Let \( \mathbf{I} \) denote an \( I \)-vector of ones, \( \mathbf{J} \) a \( J \)-vector of ones, and \( \mathbf{I} \times \mathbf{J} = \mathbf{I} \mathbf{J}^T \) an \( I \times J \) matrix of ones. Including an overall mean, \( \mu \), the additive part of the GGE model is

\[
\mathbf{A} = \mu \mathbf{I} \times \mathbf{J} + \alpha \mathbf{1}_I \mathbf{1}_J^T, \tag{1}
\]

and the additive part of the AMMI model is

\[
\mathbf{A} = \mu \mathbf{I} \times \mathbf{J} + \alpha \mathbf{1}_I \mathbf{1}_J^T + \beta \mathbf{1}_J \mathbf{1}_I^T. \tag{2}
\]

In this paper, it will be assumed that \( \mathbf{E} \) can be written \( \mathbf{E} = \Theta(\kappa) + \mathbf{R} = \mathbf{U}(\kappa) \mathbf{V}(\kappa)^T + \mathbf{R} \), \( \Theta(\kappa) \) models interaction and \( \mathbf{R} = \{r_{ij}\} \) is a matrix of independent \( \mathbf{N}(0, \sigma^2) \) distributed errors \( r_{ij} \). In (3), \( \mathbf{U}(\kappa) \mathbf{V}(\kappa)^T \) is the singular value decomposition of \( \Theta(\kappa) \). The subscript \((\kappa)\) indicates that the rank of \( \Theta(\kappa) \) is \( \kappa \). Thus, the singular value decomposition of \( \Theta(\kappa) \) comprises \( \kappa \) multiplicative terms that are not \( 0 \). It will be assumed that the diagonal elements of \( \mathbf{U}(\kappa) \) are sorted in decreasing order. The singular values will be denoted by \( \lambda_1, \lambda_2, \ldots, \lambda_k \), i.e. \( \lambda_k = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_k) \). The \( I \times \kappa \) matrix of left-singular vectors is \( \mathbf{U}(\kappa) = (\gamma_1, \gamma_2, \ldots, \gamma_k) \) and the \( \kappa \times J \) matrix of right-singular vectors is \( \mathbf{V}(\kappa) = (\delta_1, \delta_2, \ldots, \delta_k) \).

In scalar form, GGE and AMMI models can be written \( y_{ij} = \mu + \alpha_i + \gamma_{ij} \lambda_1 \delta_1 + \gamma_{ij} \lambda_2 \delta_2 + \ldots + \gamma_{ij} \lambda_k \delta_k + r_{ij} \) and \( y_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij} \lambda_1 \delta_1 + \gamma_{ij} \lambda_2 \delta_2 + \ldots + \gamma_{ij} \lambda_k \delta_k + r_{ij} \), respectively.

The general mean, \( \mu \), can be estimated as the average \( \hat{\mu} = \frac{1}{I} \sum_{i=1}^I \frac{1}{J} \sum_{j=1}^J y_{ij} / (IJ) \), and the row and column effects as \( \hat{\alpha}_i = \frac{1}{J} \sum_{j=1}^J y_{ij} / J - \hat{\mu} \) and \( \hat{\beta}_j = \frac{1}{I} \sum_{i=1}^I y_{ij} / I - \hat{\mu} \), respectively. Define \( \hat{\mathbf{E}} \) as the least squares estimator of \( \mathbf{E} \) in (1) or (2), where \( \hat{\mathbf{E}} = \mathbf{Y} - \mathbf{A} \). Let \( \mathbf{M} \) denote the rank of \( \hat{\mathbf{E}} \). Generally, \( \mathbf{M} = \min(I, J - 1) \) in GGE analysis, whereas \( \mathbf{M} = \min(I - 1, J - 1) \) in AMMI analysis. Through singular value decomposition, \( \hat{\mathbf{E}} \) can be written as \( \hat{\mathbf{E}} = \hat{\mathbf{U}} \hat{\mathbf{A}} \hat{\mathbf{V}}^T \), where \( \hat{\mathbf{U}} = (\hat{\gamma}_1, \hat{\gamma}_2, \ldots, \hat{\gamma}_\mathbf{M}) \) is an \( I \times \mathbf{M} \) matrix of estimated left-singular vectors, \( \hat{\mathbf{A}} = \text{diag}(\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_\mathbf{M}) \) is an \( \mathbf{M} \times \mathbf{M} \) diagonal matrix of estimated singular values sorted from largest to smallest, and \( \hat{\mathbf{V}} = (\hat{\delta}_1, \hat{\delta}_2, \ldots, \hat{\delta}_\mathbf{M}) \) is a \( \mathbf{M} \times J \) matrix of estimated right-singular vectors.

Considering the test for the null hypothesis \( H_0 : \kappa = K \) against the alternative hypothesis \( H_1 : \kappa > K \). Note that \( \kappa \) is the actual number of multiplicative terms, whereas \( K \) is the assumed number under the current null hypothesis. The ratio

\[
T = \frac{\hat{\lambda}_{K+1}^2}{\sum_{k=K+1}^\mathbf{M} \hat{\lambda}_k^2} \tag{4}
\]

can be used as a test statistic for \( H_0 \). Large values of (4) present evidence against \( H_0 \). For \( K = 0 \), Johnson and Graybill (1972) provided simulation-derived critical values for selected rows and columns.

2.2 JG/SM-test

Yochmowitz and Cornell (1978) and Schott (1986) proposed using the test statistic (4) for the problem of selecting the number of multiplicative terms. In (4), the squared singular values are distributed as eigenvalues of a Wishart matrix (Johnson and Graybill, 1972). Cornelius et al. (1996) showed, based on Johnson and Graybill (1972), Marasinghe (1985) and Schott (1986), that (4) can be transformed into an approximately \( F \)-distributed statistic \( G = (AQ)/(A(1 - Q)) \), where \( a \) and \( b \) are defined as in Appendix A, and \( Q = ((M - K)T - 1)/(M - K - 1) \). The values \( a \) and \( b \) are functions of expressions that approximate the first two moments, estimated through Monte Carlo simulation, of eigenvalues of Wishart matrices (Liu and Cornelius, 2001).

2.3 Full parametric bootstrap method

We propose parametric bootstrapping for computation of the \( p \)-value. According to this approach (Efron and Tibshirani, 1993), the null distribution is simulated using parameter estimates under the null hypothesis. First, the model parameters, including the variance, are estimated, and then a large number, \( B \), of samples are drawn from the fitted model using the estimates. For each sample, the test statistic, \( T \), is calculated. The obtained distribution of simulated test statistics approximates the true sampling distribution of \( T \) under \( H_0 \) and can be used for estimation of the \( p \)-value.

In this application, the first \( K \) terms from the singular value decomposition are taken to represent the true fixed interaction \( \Theta(\kappa) \) in all \( B \) bootstrap simulations. Thus, \( \Theta(\kappa) \) in (3) is estimated as

\[
\hat{\Theta}(\kappa) = \left\{ \begin{array}{ll}
0 & \text{if } K = 0, \\
\sum_{k=1}^\kappa \hat{\gamma}_k \hat{\delta}_k \hat{\lambda}_k^T & \text{if } K > 0.
\end{array} \right. \tag{5}
\]

To simulate residual error, random noise is added. Specifically, a random matrix \( \mathbf{R}^B \) of independent \( \mathbf{N}(0, \sigma^2 \hat{\delta}(\kappa)^2) \) distributed errors is added to \( \hat{\Theta}(\kappa) \). The superscript, \( B \), indicates that \( \mathbf{R}^B \) is a bootstrap version of \( \mathbf{R} \), whereas the subscript, \( b \), indicates that \( \mathbf{R}_b^B \) is the \( b \)th bootstrap sample. The variance, \( \hat{\delta}^2(\kappa) \), which is an estimate of \( \sigma^2 \), can be derived as follows.

Consider an \( I \)-by-\( J \) two-way layout with additive main effects of rows and columns, additive fixed effects of interactions, and independent additive errors with variance \( \sigma^2 \). When there is no replication within cells, a two-way ANOVA without interaction is commonly fitted. In this case, the expected residual sum of squares equals the unknown interaction sum of squares plus \( (I - 1)(J - 1) \sigma^2 \) (Searle, Casella and McCulloch, 1992). Similarly, as relevant for GGE analysis, when a one-way ANOVA with row effects is performed, the expected residual sum of squares equals the unknown sum of squares due to column and interaction effects plus \( I(J - 1) \sigma^2 \). Under \( H_0 \), the sum of squares of true interaction effects is 0 when \( K = 0 \) and \( \sum_{k=1}^\kappa \lambda_k^2 \) when \( 1 \leq K \leq M - 2 \). We propose that the observed error sum of squares from the additive model, i.e. the model without multiplicative terms, be equated to the
expected error sum of squares under $H_0$, i.e.,
\[ \sum_{k=1}^{K} \hat{\lambda}_b^2 + v\sigma^2, \quad 1 \leq K \leq M - 2, \tag{6} \]
where $v = I(J - 1)$ for GGE, and $v = (I - 1)(J - 1)$ for AMMI. When fitting the additive model, the error sum of squares always equals $\sum_{k=1}^{K} \hat{\lambda}_b^2$. Hence, by substituting $\hat{\lambda}_b^2$ for $\lambda_b^2$ in (6), the estimating equation $\sum_{k=1}^{K} \hat{\lambda}_b^2 = \sum_{k=1}^{K} \lambda_b^2 + v\sigma^2(k)$, $1 \leq K \leq M - 2$, is obtained. Solving for $\hat{\sigma}_k^2$, the error variance $\sigma^2$ in (3) can be estimated as
\[ \hat{\sigma}_k^2 = \frac{1}{v} \sum_{k=K+1}^{M} \hat{\lambda}_b^2, \quad 0 \leq K \leq M - 2, \tag{7} \]
where $v$ is defined as in (6).

In summary, to test $H_0$ using the full parametric bootstrap procedure for GGE and AMMI analysis, we suggest the following:

1. Compute $\Theta_{(K)}$ as specified by (5) and $\hat{\sigma}_k^2$ as specified by (7).
2. For $b = 1, 2, \ldots, B$, where $B$ is large, do the following:
   i. Sample an $I \times J$ matrix $R_b^B$ of independent $N(0, \hat{\sigma}_k^2)$ distributed errors
   ii. Compute $E_b^B = \Theta_{(K)} + R_b^B$
   iii. Let $\hat{e}_{ibi}$ denote the element in the $i$th row and $j$th column of $E_b^B$. Let $\bar{e}_{ij}$ denote row (i.e. environment) means, $\bar{e}_{bj}$ column (i.e. genotype) means, and $\bar{e}_{ib}$ the general mean of $E_b^B$. For GGE analysis, let $\bar{e}_{ib} = \{\bar{e}_{ib} - \bar{e}_{ij} - \bar{e}_{bj} + \bar{e}_{ib}\}$. For AMMI analysis, let $\bar{e}_{ib} = \{\bar{e}_{ib} - \bar{e}_{ij} - \bar{e}_{bj} + \bar{e}_{ib}\}$.
3. Subject $E_b^B$ to singular value decomposition and use the obtained singular values for calculation of bootstrap samples $T_b$ according to the right hand side of (4).
4. Estimate the $p$-value as the observed frequency of $T_b$ larger than $T$ computed from the data.

2.4 Simple parametric bootstrap method

An approximate version of the full parametric bootstrap method can be derived as follows. Let $L_j$ be a $J \times (J - 1)$ matrix such that $L_j L_j^T = I_j - (1/J)1_{J \times J}$. Then $E$ equals $\sum_{j} Y_i L_j^T L_j Y_j + L_j^T Y_i L_j$, where $\sum_{j}$ is the GGE and AMMI analysis, respectively. For the AMMI model with $J \leq I$, Johnson and Graybill (1972) showed that $E^T E / \sigma^2$ is noncentral Wishart distributed with $I - 1$ degrees of freedom, scale matrix $I$, and noncentrality matrix $L_j^T \Theta_{(K)} \Theta_{(K)} L_j / \sigma^2$, i.e. $E^T E / \sigma^2$ is $W_{J-1}(I - 1, L_j^T \Theta_{(K)} \Theta_{(K)} L_j / \sigma^2)$. When $I > J$, $E^T E / \sigma^2$ is $W_{I-1}(J - 1, L_j^T \Theta_{(K)} \Theta_{(K)} L_j / \sigma^2)$. The positive eigenvalues of $E^T E / \sigma^2$ are the same as those of $EE^T / \sigma^2$ and equal $\lambda_1^2 / \sigma^2, \lambda_2^2 / \sigma^2, \ldots, \lambda_J^2 / \sigma^2$. Provided that $H_0$ is correct and $\lambda_1^2 / \sigma^2, \lambda_2^2 / \sigma^2, \ldots, \lambda_J^2 / \sigma^2$ are large, the asymptotic joint distribution of these eigenvalues can be approximated by the joint distribution of the eigenvalues of a central Wishart matrix distributed as $W_{J-1,K}(I - 1 - K, I)$ (Muirhead, 1978; Marasinghe, 1985). As a result, the distribution of $T$ in (4) may be approximated by the distribution of the ratio of the first eigenvalue to the sum of all eigenvalues of a central Wishart matrix. Since the joint distribution of the eigenvalues of a $W_{J-1,K}(I - 1 - K, I)$ distributed matrix is equal to the joint distribution of the squared singular values of a random $(I - 1 - K) \times (J - 1 - K)$ matrix of independent standard normal values, the following simple parametric bootstrap method can be proposed:

1. For $b = 1, 2, \ldots, B$, where $B$ is large, do the following:
   i. Sample a $(D - K) \times (J - 1 - K)$ matrix $E_b^B$ of independent $N(0, \hat{\sigma}_k^2)$ distributed errors, where $D = I - 1$ in AMMI analysis and $D = J$ in GGE analysis.
   ii. Subject $E_b^B$ to singular value decomposition and use the obtained singular values for calculation of bootstrap samples $T_b$ according to the right hand side of (4), here using $K = 0$ (i.e. compute $T_b$ as the ratio of the first squared singular value to the sum of all squared singular values).

2. Estimate the $p$-value as the observed frequency of $T_b$ larger than $T$ computed from the data.

Web Appendix C and D provide R and SAS code, respectively, for the simple parametric bootstrap method.

2.5 Sequential $F$-test

Gollob (1968) and Wold (1978) defined degrees of freedom associated with parameters in multiplicative models as the number of parameters minus the number of constraints. With this definition, the number of degrees of freedom, $DF_K$, needed to estimate the $K$th interaction term is $I + J - 2K$ and $I + J - 2K - 1$ in GGE and AMMI analysis, respectively. Consequently, the residual error degrees of freedom, $DF_{Residual}$, is $I(J - 1) - \sum_{K=1}^{K}(I + J - 2k) = (I - K)(J - 1 - K)$ and $(I - 1)(J - 1) - \sum_{K=1}^{K}(I + J - 2k - 1) = (I - 1 - K)(J - 1 - K)$ in GGE and AMMI analysis, respectively. Gollob (1968) proposed an $F$-test with denominator mean squared error calculated from replicates within genotype-by-environment combinations. This test was not applicable to the examples of the following section, since these examples used means or non-replicated data. For this reason, sequential $F$-test statistics were instead calculated as ratios between the mean square due to fixed interaction, i.e. the multiplicative terms, and residual mean square.

2.6 Cross-validation method

Dias and Krzanowski (2003, 2006) proposed cross-validation for choosing the number of multiplicative interaction terms. Based on their work, the present paper uses the following method. Denote by $E_s^{(-i)}$ the result of deleting the $j$th column of $E$ and after this subtracting row means. Denote by $E_s^{(-i)}$ the result of deleting the $i$th row of $E$ and after this, in case of AMMI-analysis, subtracting column means. Let $U_i^{(-i)}$ and $S_i^{(-i)}$ denote the matrix of left-singular vectors, $\Lambda_i^{(-i)}$ and $\Lambda_i^{(-i)}$ the matrix of singular values, and $V_i^{(-i)}$ and $V_i^{(-i)}$ the matrix of right-singular vectors, of $E_s^{(-i)}$ and $E_s^{(-i)}$, respectively. Write $U_i^{(-i)} = \{u_{si}^{(-i)}\}$, $U_i^{(-i)} = \{u_{si}^{(-i)}\}$, $V_i^{(-i)} = \{v_{sj}^{(-i)}\}$, and $V_i^{(-i)} = \{v_{sj}^{(-i)}\}$. Denote by $\lambda_i^{(-i)}$ and $\lambda_i^{(-i)}$ the $k$th diagonal element of the diagonal
Table 1
AMMI analyses of the New York soybean, CIMMYT maize, Ontario wheat and Swedish pea datasets. Sum of squares, $\hat{\lambda}_{K+1}^2$, and proportions (%) of the additive model error sum of squares, $\sum_{k=1}^M \hat{\lambda}_k^2$. Test statistics for the cross-validation method ($W$), the sequential $F$-test ($F$), the JG/SM-test ($G$) and the parametric bootstrap tests ($T$). Probability values for the sequential $F$-test, the JG/SM-test, and the full and simple parametric bootstrap tests. Bootstrap test $p$-values were computed using $B = 100,000$ bootstrap samples. Under $H_0$, the model has $K$ terms.

<table>
<thead>
<tr>
<th>Term</th>
<th>Sum of squares</th>
<th>Test statistic</th>
<th>Probability value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$W$</td>
<td>$F$</td>
</tr>
<tr>
<td>$K+1$</td>
<td>$\hat{\lambda}_{K+1}^2$</td>
<td>%</td>
<td></td>
</tr>
<tr>
<td>New York soybean data</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>8 189 065</td>
<td>82.4</td>
<td>6.81</td>
</tr>
<tr>
<td>2</td>
<td>1 170 288</td>
<td>11.8</td>
<td>1.28</td>
</tr>
<tr>
<td>3</td>
<td>254 904</td>
<td>2.6</td>
<td>-0.11</td>
</tr>
<tr>
<td>4</td>
<td>200 449</td>
<td>2.0</td>
<td>0.13</td>
</tr>
<tr>
<td>5</td>
<td>107 532</td>
<td>1.1</td>
<td>0.15</td>
</tr>
<tr>
<td>CIMMYT maize data</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>35 078 698</td>
<td>56.2</td>
<td>1.82</td>
</tr>
<tr>
<td>2</td>
<td>9 426 035</td>
<td>15.1</td>
<td>0.15</td>
</tr>
<tr>
<td>3</td>
<td>6 515 627</td>
<td>10.4</td>
<td>0.00</td>
</tr>
<tr>
<td>4</td>
<td>5 383 947</td>
<td>8.6</td>
<td>0.47</td>
</tr>
<tr>
<td>5</td>
<td>3 091 503</td>
<td>5.0</td>
<td>0.41</td>
</tr>
<tr>
<td>Ontario wheat data</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>9 616</td>
<td>48.2</td>
<td>0.64</td>
</tr>
<tr>
<td>2</td>
<td>4 652</td>
<td>23.3</td>
<td>1.56</td>
</tr>
<tr>
<td>3</td>
<td>1 933</td>
<td>9.7</td>
<td>0.09</td>
</tr>
<tr>
<td>4</td>
<td>1 249</td>
<td>6.3</td>
<td>-0.02</td>
</tr>
<tr>
<td>5</td>
<td>1 083</td>
<td>5.4</td>
<td>0.20</td>
</tr>
<tr>
<td>Swedish pea data</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>69 259</td>
<td>73.7</td>
<td>0.47</td>
</tr>
<tr>
<td>2</td>
<td>20 720</td>
<td>22.0</td>
<td>0.23</td>
</tr>
</tbody>
</table>

3. Examples
In this section, the full and simple parametric bootstrap methods are compared with the cross-validation method, the sequential $F$-test and the JG/SM-test, using three datasets from the GGE and AMMI literature and a small dataset of mean yields from four pea trials:

1. Gauch (1992, 2006) studied a dataset with seven soybean genotypes in ten New York environments. The first multiplicative term of the AMMI model sorts genotypes with respect to time required for maturation, and environments with regard to length of season (Gauch, 1992). All trials comprised four replicates. We used the table of genotype-by-environment means.

2. The international maize and wheat improvement center (CIMMYT) trial EVT16B investigated nine maize cultivars at 20 international sites (Cornelius et al., 1996).

3. Yan and Tinker (2006), Yan et al. (2007) and Yang et al. (2009) studied the Ontario winter wheat data of mean yields from 18 cultivars at nine locations. Using this example, Gauch et al. (2008) discussed interpretations depending on the choices of model (GGE or AMMI) and the number of multiplicative terms.

4. To investigate small-sample properties, a dataset with five pea genotypes in four trials was analyzed (Web Table 1). The trials were performed at locations in southern Sweden during the three-year period 2004–2006.
The two first multiplicative terms of the AMMI analysis of the CIMMYT maize data are illustrated in Figure 1. Principal component axis 1 (PC1) displays the values of $\sqrt{\lambda_1}$ and $\delta_1$, whereas axis 2 (PC2) displays the values of $\sqrt{\lambda_2}$ and $\delta_2$. GGE and AMMI analysis can be used to determine which genotypes are performing similarly (e.g., G5 and G6) or dissimilarly (e.g., G4 and G8) in varying environments, and to classify environments into groups of environments that are similar with regard to performance of genotypes.

Table 1 and Web Table 2 present results of AMMI and GGE analyses, respectively. The tables show tests of the first five multiplicative terms, except for the Swedish pea dataset, for which only two terms were tested. The $p$-value of the parametric bootstrap test was derived using $B = 100,000$ bootstrap samples.

The full and simple parametric bootstrap methods resulted in very similar $p$-values. The parametric bootstrap tests and the $JG/SM$-tests gave similar, but not identical, $p$-values. The sequential $F$-test was generally much more liberal. In some cases, the cross-validation $W$-statistic was smaller than 1 although the parametric bootstrap and $JG/SM$-tests indicated significant effects at level 0.05. Following the Eastment and Krzanowski (1982) decision rule, the cross-validation method agreed with the parametric bootstrap and $JG/SM$-tests.

Using the bootstrap tests, an AMMI model with two terms was appropriate for the New York soybean and the Ontario winter wheat datasets. For the small Swedish pea dataset, there were no significant patterns in the interaction.

For the CIMMYT maize dataset, AMMI PC1 and PC2 captures 56.2% and 15.1% of the genotype-by-environment interaction, respectively (Table 1). The parametric bootstrap methods indicate that the first term (PC1) is significant, but the second term (PC2) is not. Since $H_0 : \kappa = 1$ cannot be rejected, the interaction pattern displayed by the vertical axis in Figure 1 is not larger than one would expect by chance. For example, the difference between genotypes G1 and G8 should not be overemphasized. In this case, a biplot that illustrates genotype and environment means on the horizontal axis and PC1 on the vertical axis (Gauch, 1992) is appropriate. We propose that terms are tested sequentially until a nonsignificant result is obtained (see the Discussion). With this decision rule, the final model for the maize data contains a single multiplicative term although Table 1 shows that the fourth term is significant assuming a null model with three terms.

Figure 2 shows the empirical distribution of $T_b$ for the four first terms of the AMMI analysis of the CIMMYT maize data using the simple parametric bootstrap method. Notice that $T_b$ is not dependent on the data. Each $T_b$ is computed from a matrix of independent $N(0,1)$ distributed values. The $p$-values (Table 1) correspond to the areas on the right hand side of the dashed lines in Figure 2.

4. Simulation studies

For each dataset, $\Theta_{(K)}$ was calculated according to (5). Repeatedly, matrices of random $N(0, \hat{\sigma}_{(K)}^2)$ distributed values were generated and added to $\Theta_{(K)}$, with $\hat{\sigma}_{(K)}^2$ calculated from the original data using (7). For the New York soybean data, the CIMMYT maize data and the Ontario winter wheat data, $H_0$ was tested for $K = 0, 1, \ldots, 4$. For the Swedish pea dataset, $H_0$ was tested for $K = 0, 1$. For each scenario, 100,000 random matrices were generated. For each of these datasets, the sequential $F$ statistic, the cross-validation $W$ statistic, the $JG/SM$-test $G$ statistic and the parametric bootstrap $T$ statistic were calculated.

With this large number of simulated datasets an approximate 0.05 tolerance interval for probability 0.05 can be written as $\pm 1.96 \sqrt{0.05 (1 - 0.05) / 100,000} = \pm 0.00135$. In the inner simulation loops, the $p$-values for the parametric bootstrap methods were determined based on $B = 1000$ bootstrap samples. For the cross-validation method, the Eastment and Krzanowski (1982) decision rule was used.

For examination of Type I error rates, the actual model was assumed to have $\kappa = K$ terms. Thus, $\Theta_{(K)}$ was computed and random $N(0, \hat{\sigma}_{(K)}^2)$ terms were repeatedly added. For examination of power, the actual model was assumed to have $\kappa > K$ terms. When $\kappa = K + 1$, then $\Theta_{(K+1)}$ was computed and random $N(0, \hat{\sigma}_{(K+1)}^2)$ terms were repeatedly added. The null hypothesis was always $H_0 : \kappa = K$.

The parametric bootstrap methods showed Type I error rates that were very close to the nominal significance level 0.05 (Table 2 and Web Table 3 for AMMI and GGE, respectively). The $JG/SM$-test usually produced significant results slightly too often. With this test, Type I error rates were less favorable for the New York soybean and Swedish pea data than for the other datasets. This indicates that the $JG/SM$-test might perform less well for small datasets than for large. The sequential $F$-test was found to be much too liberal in declaring multiplicative terms significant, and should for this
Table 2

Type I error rates in AMMI analyses using cross-validation, the sequential $F$-test, the $JG/SM$-test, and the full and simple parametric bootstrap tests. Nominal significance level 0.05. Under $H_0$, the model has $K$ terms.

<table>
<thead>
<tr>
<th>Term</th>
<th>$K+1$ Bootstrap</th>
<th>Cross-validation</th>
<th>$F$-test</th>
<th>$JG/SM$</th>
<th>Full</th>
<th>Simple</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>New York soybean data</td>
<td>0.021 0.529 0.057 0.049 0.050</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 0.018 0.417 0.056 0.049 0.052</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 0.000 0.309 0.058 0.051 0.049</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>4 0.000 0.209 0.059 0.048 0.050</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>5 0.000 0.124 0.040 0.051 0.050</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CIMMYT maize data</td>
<td>1 0.004 0.880 0.054 0.050 0.049</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 0.006 0.808 0.055 0.050 0.050</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 0.003 0.706 0.054 0.049 0.049</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>4 0.000 0.580 0.054 0.049 0.049</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>5 0.000 0.444 0.054 0.049 0.050</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ontario wheat data</td>
<td>1 0.005 0.854 0.055 0.050 0.049</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 0.011 0.776 0.054 0.050 0.049</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 0.002 0.672 0.055 0.050 0.050</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>4 0.000 0.551 0.055 0.049 0.049</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>5 0.000 0.416 0.054 0.049 0.048</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Swedish pea data</td>
<td>1 0.068 0.168 0.064 0.050 0.050</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 0.000 0.098 0.017 0.050 0.049</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3

Estimated power in AMMI analyses when using the $JG/SM$-test and the full and simple parametric bootstrap tests. Nominal significance level 0.05. Under $H_0$, the model has $K$ terms. The actual model has $\kappa = K+1$ terms.

<table>
<thead>
<tr>
<th>Term</th>
<th>Bootstrap</th>
<th>$K+1$</th>
<th>$JG/SM$</th>
<th>Full</th>
<th>Simple</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>New York soybean data</td>
<td>1 1.000 1.000 1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 1.000 1.000 1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 0.825 0.803 0.804</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>4 0.984 0.978 0.978</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>5 1.000 1.000 1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CIMMYT maize data</td>
<td>1 1.000 1.000 1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 0.992 0.991 0.991</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 0.996 0.996 0.995</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>4 1.000 1.000 1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>5 1.000 1.000 1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ontario wheat data</td>
<td>1 1.000 1.000 1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 1.000 1.000 1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 0.980 0.979 0.977</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>4 0.971 0.967 0.967</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>5 0.999 0.999 0.999</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Swedish pea data</td>
<td>1 0.563 0.500 0.499</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 0.139 0.346 0.342</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5. Discussion

The problem of testing multiplicative terms in GGE and AMMI analysis is complicated, because estimated squared singular values are not chi-square distributed (Schott, 1986). The $F$-distribution has been proposed as reference distribution for various approximate test statistics based on ratios of mean squares (Cornelius et al., 1996), and the question has been how to calculate numbers of degrees of freedom (Mandel, 1971; Gauch, 1992). Using resampling methods, the degrees-of-freedom problem of identifying the actual reference distribution can be circumvented. The reference distribution is simulated, which enables approximate inference. This paper proposed parametric bootstrap methods for testing multiplicative terms in GGE and AMMI models. The results of the simulation study indicated that these methods can be used to select the number of multiplicative terms to be retained in the model.

According to Bradley (1978), the empirical level should not deviate from the nominal level by more than 10 %. Under this rule, the Type I error rate should, at nominal level 0.05, not exceed 0.055. The parametric bootstrap methods fulfil this requirement, but the $JG/SM$-test does in general not.
with \( K = 1, 2, \ldots, M - 2 \) as long as \( H_0 \) is rejected. The 
\((K + 1)\)th term should not be tested unless the \( K \)th term is 
significant. Through this procedure, the requirement that the 
first \( K \) singular values be large is expected to be fulfilled. 
Since the simple parametric bootstrap method makes use of 
the same approximation, a sequential testing procedure can 
be recommended for this method as well. Equivalently, if all 
multiplicative terms are tested and the results compiled in a 
table similar to Table 1, a forward selection procedure may be 
applied when deciding on which terms to retain in the final 
model. 

If models are tested sequentially until a non-significant 
result is obtained, then computed \( p \)-values are not strictly 
correct provided they have not taken into account the proba-
bilities that the tests are performed at all. A Type I error in 
the test of the \((K + 1)\)th term can occur only if the first 
\( K \) terms were declared significant. The probability of this 
event is smaller than 1. As a consequence, the probabilities 
of Type I error that were estimated through simulation are 
upper bounds for family-wise error rates. If 
\( \alpha_0 \) is the probability of 
Type I error when testing the \((K + 1)\)th term, then 
the probability of selecting more than \( K \) terms is smaller 
than \( \alpha_0 \). Provided that \( \alpha_0 \) is close to the nominal significance 
level \( \alpha \), the sequential testing procedure gives protection 
against overfitting. To maintain the family-wise error rate, 
we recommend that the number of terms is increased only as 
long as \( H_0 \) is rejected. 

The parametric bootstrap methods of the present paper 
can be used for the problem of selecting principal components 
in PCA. Let \( X \) be a data matrix with \( I \) rows (observations) and 
\( J \) columns (variables). Let \( \hat{X} \) denote the matrix when 
each column in \( X \) has been standardized to have zero means, 
and \( Z \) the matrix when each column has been standardized 
to have zero means and unit standard deviations. PCA uses 
the sample covariance matrix of either \( X \) or \( Z \). The variances 
of the computed principal components are proportional to 
the squared singular values of \( X \) or \( Z \), depending on whether 
\( X \) or \( Z \) was used for the PCA. Thus, large singular values 
indicate important principal components, and the proposed 
parametric bootstrap methods can be applied to test their 
significance. Performance of the proposed methods in PCA 
deserves further study. 

Due to the demonstrated fine performance with regard to 
power and probability of Type I errors, we recommend the 
parametric bootstrap methods for testing multiplicative terms 
in GGE and AMMI analyses. Since no important differences 
in performance between the two parametric bootstrap meth-
ods could be detected, we specifically advocate the use of the 
simple parametric bootstrap method, which is easier to pro-
gram and computationally more effective. With this method, 
it is advisable to determine the number of multiplicative terms 
to retain in the model through a forward-selection procedure.

**Supplementary Materials**

Web Appendices A–D and the Web Tables referenced in 
Sections 2–4 are available with this paper at the Biometrics 
website on Wiley Online Library.

**References**

Bradley, J. V. (1978). Robustness? British Journal of Mathema-
tical & Statistical Psychology 31, 144–152.

in the additive main effects and multiplicative interaction 

tables for the means and standard deviations of the first 
three roots of a Wishart matrix. *Technometrics* 22, 613– 
616.

Statistical tests and estimators of multiplicative models 
for genotype-by-environment interaction. In *Genotype-
by-environment interaction*, M. S. Kang and H. G. Gauch 

*Biometrics* 50, 310–311.

and cross validation in additive main effect and multi-
plicative interaction models. *Crop Science* 43, 865–873.

ponents in the additive main effect and multiplicative 
interaction (AMMI) models. *Scientia Agricola* 63, 169– 
175.

Eastment, H. T., and Krzanowski, W. J. (1982). Cross-
validatory choice of the number of components from a 

bootstrap*. Boca Raton: Chapman & Hall/CRC.

adaptation in a plant-breeding programme. Australian 
*Journal of Agricultural Research* 14, 742–754.

trials: AMMI analysis of factorial designs*. Amsterdam: 
Elsevier.


AMMI and GGE. *Crop Science* 46, 1488–1500.

analysis of yield trials by AMMI and GGE: Further 

Gollob, H. F. (1968). A statistical model which combines 
features of factor analytic and analysis of variance tech-

a two-way model with interaction and no replication. 
*Journal of the American Statistical Association* 67, 862– 
868.

approximations for the means and standard deviations of 
the characteristic roots of a Wishart matrix. *Communica-
tions in Statistics – Simulation and Computation* 30, 
963–989.

Mandel, J. (1961). Non-additivity in two-way analysis of 
variance. *Journal of the American Statistical Association* 
56, 878–888.

Mandel, J. (1971). A new analysis of variance model for non-
additive data. *Technometrics* 13, 1–18.

Marasinghe, M. G. (1985). Asymptotic tests and Monte-Carlo


Received May 2013. Revised February 2014. Accepted February 2014.