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
http://onlinelibrary.wiley.com/doi/10.1111/biom.12162/abstract

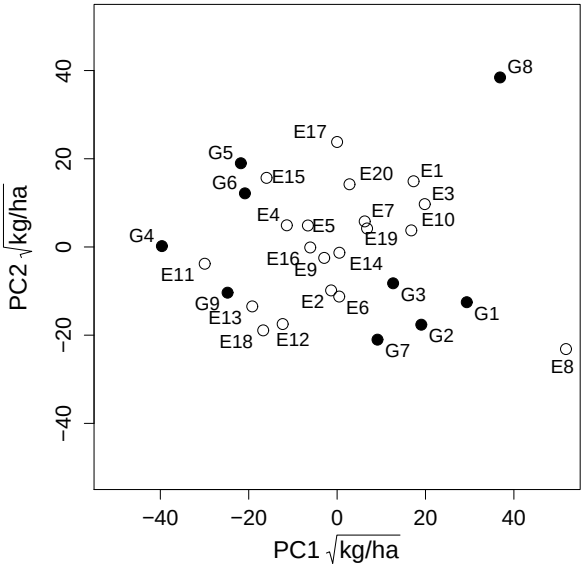

Figure 1. AMMI biplot for the CIMMYT maize data.

a dataset with nine maize genotypes (G1–G9) investigated in 20 environments (E1–E20). Points near the origin have small interaction effects, and points near each other have similar interaction effects (Gauch, 1992). Yang et al. (2009) discussed the validity of the biplot as a statistical method for analysis of genotype-by-environment interaction. One of their main concerns was the frequent lack of statistical hypothesis testing for determining the number of multiplicative terms. Figure 1 illustrates the first two multiplicative terms, but it is possible that more or fewer terms give a better description of the interaction. In practice, researchers would like to separate fixed genotype-by-environment interaction from random noise. This paper aims to develop a method for significance testing of multiplicative terms in GGE and AMMI models.

Testing for interaction in non-replicated two-way layouts goes back to Tukey (1949), who introduced the one-degree-of-freedom test for additivity, and Mandel (1961), who proposed row-specific regression on additive column effects. Yates and Cochran (1938) and Finlay and Wilkinson (1963) proposed genotype-specific regression of yield on site means. Such models can be fitted using nonlinear regression (Ng and Grunwald, 1997; Piepho, 1999). Mandel (1971) proposed the AMMI model. Johnson and Graybill (1972) derived a likelihood ratio test for the first multiplicateve term of the AMMI model. Based on their work, Marasinghe (1985) and Schott (1986) proposed a sequential testing procedure for all terms. This procedure tests the $(K+1)$th multiplicative term as if it were the first term in a problem with the numbers of rows and columns reduced by K. Cornelius et al. (1996) presented the approximately F-distributed JG/SM test statistic, which is built on the contributions by Johnson and Graybill (1972), Marashinge (1985) and Schott (1986). Cross-validation (Dias and Krzanowski, 2003, 2006) is another option for selecting the number of multiplicative 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 $K \geqslant 0$, the test statistic is the same as the likelihood ratio statistic that was derived by Yochmowitz and Cornell (1978) and advocated by Schott (1986). When $K \geqslant 1$, 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. By this means, simulation of the null distribution is made possible. Utilizing an approxmate 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 jth genotype in the ith environment ($i = 1, 2, \ldots, I; j = 1, 2, \ldots, J$). Let $\mathbf{Y}$ denote the $I \times J$ matrix of observations y_{ij}, , i.e. $\mathbf{Y} = \{y_{ij}\}$. GGE and AMMI models can be written in the form $\mathbf{Y} = \mathbf{A} + \mathbf{E}$, where $\mathbf{A}$ denotes an additive part and $\mathbf{E}$ denotes a matrix of interactions and residual errors. Let $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_I)^{\mathrm{T}}$ be a vector of environment main effects, and $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_J)^{\mathrm{T}}$ a vector of genotype main effects. Let $\mathbf{1}_I$ denote an I-vector of ones, $\mathbf{1}_J$ a J-vector of ones, and $\mathbf{J}_{I\times J} = \mathbf{1}_I\mathbf{1}_J^{\mathrm{T}}$ an $I \times J$ matrix of ones. Including an overall mean, μ, the additive part of the GGE model is

$$\mathbf{A} = \mu\mathbf{J}_{I\times J} + \boldsymbol{\alpha}\mathbf{1}_J^{\mathrm{T}}, \tag{1}$$

and the additive part of the AMMI model is

$$\mathbf{A} = \mu\mathbf{J}_{I\times J} + \boldsymbol{\alpha}\mathbf{1}_J^{\mathrm{T}} + \mathbf{1}_I\boldsymbol{\beta}^{\mathrm{T}}. \tag{2}$$

In this paper, it will be assumed that $\mathbf{E}$ can be written

$$\mathbf{E} = \boldsymbol{\Theta}_{(\kappa)} + \mathbf{R} = \mathbf{U}_{(\kappa)}\boldsymbol{\Lambda}_{(\kappa)}\mathbf{V}_{(\kappa)}^{\mathrm{T}} + \mathbf{R}, \tag{3}$$

where $\boldsymbol{\Theta}_{(\kappa)}$ models interaction and $\mathbf{R} = \{r_{ij}\}$ is a matrix of independent $\mathrm{N}(0, \sigma^2)$ distributed errors r_{ij}. In (3), $\mathbf{U}_{(\kappa)}\boldsymbol{\Lambda}_{(\kappa)}\mathbf{V}_{(\kappa)}^{\mathrm{T}}$ is the singular value decomposition of $\boldsymbol{\Theta}_{(\kappa)}$. The subscript (κ) indicates that the rank of $\boldsymbol{\Theta}_{(\kappa)}$ is κ. Thus, the singular value decomposition of $\boldsymbol{\Theta}_{(\kappa)}$ comprises κ multiplicative terms that are not $\mathbf{0}$. It will be assumed that the diagonal elements of $\boldsymbol{\Lambda}_{(\kappa)}$ are sorted in decreasing order. The singular values will be denoted by $\lambda_1, \lambda_2, \ldots, \lambda_\kappa$, i.e. $\boldsymbol{\Lambda}_{(\kappa)} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_\kappa)$. The $I \times \kappa$ matrix of left-singular vectors is $\mathbf{U}_{(\kappa)} = (\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \ldots, \boldsymbol{\gamma}_\kappa)$, and the $J \times \kappa$ matrix of right-singular vectors is $\mathbf{V}_{(\kappa)} = (\boldsymbol{\delta}_1, \boldsymbol{\delta}_2, \ldots, \boldsymbol{\delta}_\kappa)$. In scalar form, GGE and AMMI models can be written $y_{ij} = \mu + \alpha_i + \gamma_{i1}\lambda_1\delta_{j1} + \gamma_{i2}\lambda_2\delta_{j2} + \ldots + \gamma_{i\kappa}\lambda_\kappa\delta_{j\kappa} + r_{ij}$ and $y_{ij} = \mu + \alpha_i + \beta_j + \gamma_{i1}\lambda_1\delta_{j1} + \gamma_{i2}\lambda_2\delta_{j2} + \ldots + \gamma_{i\kappa}\lambda_\kappa\delta_{j\kappa} + r_{ij}$, respectively.

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

Consider testing the null hypothesis $H_0 : \kappa = K$ against the alternative hypothesis $H_1 : \kappa > K$. Note that κ is the actual number of multiplicative terms, whereas K is the assumed number under the current null hypothesis. The ratio

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

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 $G = (bQ)/(a(1-Q))$, where a and b are defined as in Web 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 $\boldsymbol{\Theta}_{(\kappa)}$ in all B bootstrap simulations. Thus, $\boldsymbol{\Theta}_{(\kappa)}$ in (3) is estimated as

$$\hat{\Theta}_{(K)} = \begin{cases} 0 & \text{if } K = 0, \\ \sum_{K=1}^{K}\hat{\boldsymbol{\gamma}}_k\hat{\lambda}_k\hat{\boldsymbol{\delta}}_k^{\mathrm{T}} & \text{if } K > 0. \end{cases} \tag{5}$$

To simulate residual error, random noise is added. Specifically, a random matrix $\mathbf{R}_b^B$ of independent $\mathrm{N}(0, \hat{\sigma}_{(K)}^2)$ distributed errors is added to $\hat{\boldsymbol{\Theta}}_{(K)}$. The superscript, B, indicates that $\mathbf{R}_b^B$ is a bootstrap version of $\mathbf{R}$, whereas the subscript, b, indicates that $\mathbf{R}_b^B$ is the bth bootstrap sample. The variance, $\hat{\sigma}_{(K)}^2$, which is an estimate of σ^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 σ^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 $\sum_{k=1}^{K}\lambda_k^2$ when $1 \leqslant K \leqslant M-2$. 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.

$$\sum_{k=1}^{K} \lambda_k^2 + v\sigma^2, \quad 1 \leqslant K \leqslant M-2, \tag{6}$$

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 $\sum_{k=1}^{M} \hat{\lambda}_k^2$. Hence, by substituting $\hat{\lambda}_k^2$ for λ_k^2 in (6), the estimating equation $\sum_{k=1}^{M} \hat{\lambda}_k^2 = \sum_{k=1}^{K} \hat{\lambda}_k^2 + v\hat{\sigma}_{(K)}^2$, $1 \leqslant K \leqslant M-2$, is obtained. Solving for $\hat{\sigma}_{(K)}^2$, the error variance σ^2 in (3) can be estimated as

$$\hat{\sigma}_{(K)}^2 = \frac{1}{v} \sum_{k=K+1}^{M} \hat{\lambda}_k^2, \quad 0 \leqslant K \leqslant M-2, \tag{7}$$

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 $\hat{\mathbf{\Theta}}_{(K)}$ 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 $\mathbf{R}_b^B$ of independent $\mathrm{N}(0, \hat{\sigma}_{(K)}^2)$ distributed errors
   ii. Compute $\mathbf{E}_b^B = \hat{\mathbf{\Theta}}_{(K)} + \mathbf{R}_b^B$
   iii. Let e_{bij}^B denote the element in the ith row and jth column of $\mathbf{E}_b^B$. Let $\bar{e}_{bi\cdot}^B$ denote row (i.e. environment) means, $\bar{e}_{b\cdot j}^B$ column (i.e. genotype) means, and $\bar{e}_{b\cdot\cdot}^B$ the general mean of $\mathbf{E}_b^B$. For GGE analysis, let $\hat{\mathbf{E}}_b^B = \{e_{bij}^B - \bar{e}_{bi\cdot}^B\}$. For AMMI analysis, let $\hat{\mathbf{E}}_b^B = \{e_{bij}^B - \bar{e}_{bi\cdot}^B - \bar{e}_{b\cdot j}^B + \bar{e}_{b\cdot\cdot}^B\}$
3. Subject $\hat{\mathbf{E}}_b^B$ 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 $\mathbf{L}_J$ be a $J \times (J-1)$ matrix such that $\mathbf{L}_J\mathbf{L}_J^{\mathrm{T}} = \mathbf{I}_J - (1/J)\mathbf{J}_{J\times J}$. Then $\hat{\mathbf{E}}$ equals $\mathbf{Y}\mathbf{L}_J\mathbf{L}_J^{\mathrm{T}}$ and $\mathbf{L}_I\mathbf{L}_I^{\mathrm{T}}\mathbf{Y}\mathbf{L}_J\mathbf{L}_J^{\mathrm{T}}$ in GGE and AMMI analysis, respectively. For the AMMI model with $J \leqslant I$, Johnson and Graybill (1972) showed that $\hat{\mathbf{E}}^{\mathrm{T}}\hat{\mathbf{E}}/\sigma^2$ is noncentral Wishart distributed with $I-1$ degrees of freedom, scale matrix $\mathbf{I}$, and noncentrality matrix $\mathbf{L}_J^T\mathbf{\Theta}_{(\kappa)}^{\mathrm{T}}\mathbf{\Theta}_{(\kappa)}\mathbf{L}_J/\sigma^2$, i.e. $\hat{\mathbf{E}}^{\mathrm{T}}\hat{\mathbf{E}}/\sigma^2$ is $\mathrm{W}_{J-1}(I-1, \mathbf{I}, \mathbf{L}_J^{\mathrm{T}}\Theta_{(\kappa)}^{\mathrm{T}}\mathbf{\Theta}_{(\kappa)}\mathbf{L}_J/\sigma^2)$. When $J > I$, $\hat{\mathbf{E}}\hat{\mathbf{E}}^{\mathrm{T}}/\sigma^2$ is $\mathrm{W}_{I-1}(J-1, \mathbf{I}, \mathbf{L}_I^{\mathrm{T}}\mathbf{\Theta}_{(\kappa)}\mathbf{\Theta}_{(\kappa)}^{\mathrm{T}}\mathbf{L}_I/\sigma^2)$. The positive eigenvalues of $\hat{\mathbf{E}}^{\mathrm{T}}\hat{\mathbf{E}}/\sigma^2$ are the same as those of $\hat{\mathbf{E}}\hat{\mathbf{E}}^{\mathrm{T}}/\sigma^2$ and equal $\hat{\lambda}_1^2/\sigma^2$, $\hat{\lambda}_2^2/\sigma^2$, ..., $\hat{\lambda}_M^2/\sigma^2$. Provided that H_0 is correct and $\lambda_1^2/\sigma^2, \lambda_2^2/\sigma^2, \ldots, \lambda_M^2/\sigma^2$ 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 $\mathrm{W}_{J-1-K}(I-1-K, \mathbf{I})$ (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 $\mathrm{W}_{J-1-K}(I-1-K, \mathbf{I})$ 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 $\hat{\mathbf{E}}_b^B$ of independent $\mathrm{N}(0,1)$ distributed errors, where $D = I-1$ in AMMI analysis and $D = I$ in GGE analysis.
   ii. Subject $\hat{\mathbf{E}}_b^B$ 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 Kth interaction term is $I+J-2K$ and $I+J-2K-1$ in GGE and AMMI analysis, respectively. Consequently, the residual error degrees of freedom, $\mathrm{DF}_{Residual}$, 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 $\hat{\mathbf{E}}^{(,-j)}$ the result of deleting the jth column of $\hat{\mathbf{E}}$ and after this subtracting row means. Denote by $\hat{\mathbf{E}}^{(-i,)}$ the result of deleting the ith row of $\hat{\mathbf{E}}$ and after this, in case of AMMI-analysis, subtracting column means. Let $\hat{\mathbf{U}}^{(-i,)}$ and $\hat{\mathbf{U}}^{(,-j)}$ denote the matrix of left-singular vectors, $\hat{\mathbf{\Lambda}}^{(-i,)}$ and $\hat{\mathbf{\Lambda}}^{(,-j)}$ the matrix of singular values, and $\hat{\mathbf{V}}^{(-i,)}$ and $\hat{\mathbf{V}}^{(,-j)}$ the matrix of right-singular vectors, of $\hat{\mathbf{E}}^{(-i,)}$ and $\hat{\mathbf{E}}^{(,-j)}$, respectively. Write $\hat{\mathbf{U}}^{(-i,)} = \{\hat{u}_{st}^{(-i,)}\}$, $\hat{\mathbf{U}}^{(,-j)} = \{\hat{u}_{st}^{(,-j)}\}$, $\hat{\mathbf{V}}^{(-i,)} = \{\hat{v}_{st}^{(-i,)}\}$ and $\hat{\mathbf{V}}^{(,-j)} = \{\hat{v}_{st}^{(,-j)}\}$. Denote by $\hat{\lambda}_k^{(-i,)}$ and $\hat{\lambda}_k^{(,-j)}$ the kth 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, $\hat{\lambda}^2_{K+1}$, and proportions (%) of the additive model error sum of squares, $\sum_{k=1}^{M} \hat{\lambda}^2_k$. 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

Term	Sum of squares		Test statistic				Probability value			
									Bootstrap	
$K+1$	$\hat{\lambda}^2_{K+1}$	%	W	F	G	T	F-test	JG/SM	Full	Simple
New York soybean data										
1	8 189 065	82.4	6.81	13.42	8.81	0.824	0.000	0.000	0.000	0.000
2	1 170 288	11.8	1.28	4.77	2.86	0.671	0.000	0.003	0.006	0.005
3	254 964	2.6	-0.11	1.44	0.57	0.445	0.239	0.862	0.864	0.865
4	200 449	2.0	0.13	2.13	1.01	0.631	0.130	0.484	0.466	0.470
5	107 532	1.1	0.15	7.26	4.27	0.916	0.038	0.109	0.095	0.096
CIMMYT maize data										
1	35 078 698	56.2	1.82	6.22	4.56	0.562	0.000	0.000	0.000	0.000
2	9 426 035	15.1	0.15	2.24	1.29	0.345	0.003	0.156	0.154	0.156
3	6 515 627	10.4	0.00	2.08	1.17	0.364	0.010	0.278	0.264	0.272
4	5 383 947	8.6	0.47	2.68	1.74	0.472	0.002	0.041	0.047	0.046
5	3 091 503	5.0	0.41	2.47	1.59	0.514	0.008	0.106	0.108	0.111
Ontario wheat data										
1	9.616	48.2	0.64	4.34	2.92	0.482	0.000	0.000	0.000	0.000
2	4.652	23.3	1.56	3.35	2.15	0.450	0.000	0.001	0.003	0.003
3	1.933	9.7	0.09	1.80	0.91	0.340	0.037	0.594	0.577	0.580
4	1.249	6.3	-0.02	1.44	0.61	0.333	0.151	0.896	0.905	0.905
5	1.083	5.4	0.20	1.72	0.86	0.433	0.088	0.618	0.606	0.610
Swedish pea data										
1	69 259	73.7	0.47	2.80	1.30	0.737	0.118	0.366	0.363	0.362
2	20 720	22.0	0.23	2.59	1.05	0.838	0.297	0.554	0.542	0.544

matrix $\hat{\mathbf{\Lambda}}^{(-i,)}$ and $\hat{\mathbf{\Lambda}}^{(,-j)}$, respectively. Define $\hat{e}_{ij}$ through $\hat{\mathbf{E}} = \{\hat{e}_{ij}\}$ and let $\hat{e}^C_{ij(K)} = \sum_{k=1}^{K} \hat{u}^{(,-j)}_{ik} \sqrt{\hat{\lambda}^{(,-j)}_k \hat{\lambda}^{(-i,)}_k} \hat{v}^{(-i,)}_{jk}$, where C indicates that $\hat{e}^C_{ij(K)}$ is a cross-validation prediction of $\hat{e}_{ij}$. Following Eastment and Krzanowski (1982), when calculating $\hat{e}^C_{ij(K)}$, the sign of $\hat{u}^{(,-j)}_{ik} \sqrt{\hat{\lambda}^{(,-j)}_k \hat{\lambda}^{(-i,)}_k} \hat{v}^{(-i,)}_{jk}$ was set equal to the sign of $\hat{\boldsymbol{\gamma}}_k \hat{\lambda}_k \hat{\boldsymbol{\delta}}^{\mathrm{T}}_k$. This is a cross-validation method, because due to deletions of rows and columns before singular value decompositions, $\hat{e}_{ij}$ is not used in $\hat{e}^C_{ij(K)}$. With $\hat{e}^C_{ij(K)}$ as prediction of $\hat{e}_{ij}$, the predicted residual sum of squares is $\mathrm{PRESS}_K = \frac{1}{IJ} \sum_{i=1}^{I} \sum_{j=1}^{J} (\hat{e}^C_{ij(K)} - \hat{e}_{ij})^2$, where $\hat{e}^C_{ij(0)} = 0$. Eastment and Krzanowski (1982), proposed using $W = ((\mathrm{PRESS}_{K-1} - \mathrm{PRESS}_K)/\mathrm{DF}_K)/(\mathrm{PRESS}_K/\mathrm{DF}_{Residual})$ as a decision rule for model selection. According to their suggestion, the optimum value for K is the largest value of K at which W is greater than 1.

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 $\hat{\boldsymbol{\gamma}}_1\sqrt{\hat{\lambda}_1}$ and $\hat{\boldsymbol{\delta}}_1\sqrt{\hat{\lambda}_1}$, whereas axis 2 (PC2) displays the values of $\hat{\boldsymbol{\gamma}}_2\sqrt{\hat{\lambda}_2}$ and $\hat{\boldsymbol{\delta}}_2\sqrt{\hat{\lambda}_2}$. 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 $\mathrm{N}(0,1)$ distributed values. The p-values (Table 1) correspond to the areas on the right hand side of the dashed lines in Figure 2.

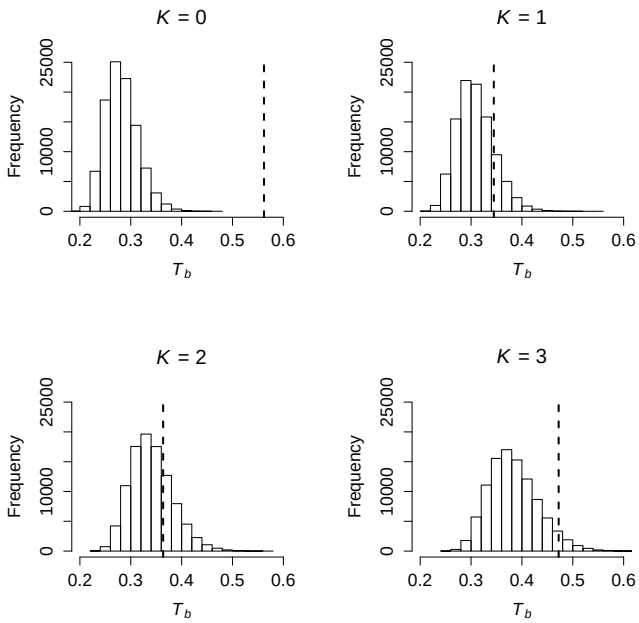

Figure 2. Bootstrap distribution of T_b in the AMMI analysis of the CIMMYT maize data using the simple parametric bootstrap method. The dashed lines indicate the location of the observed test statistic (4) on the T_b axis.

4. Simulation studies

For each dataset, $\hat{\boldsymbol{\Theta}}_{(\kappa)}$ was calculated according to (5). Repeatedly, matrices of random $\mathrm{N}(0, \hat{\sigma}^2_{(\kappa)})$ distributed values were generated and added to $\hat{\boldsymbol{\Theta}}_{(\kappa)}$, with $\hat{\sigma}^2_{(\kappa)}$ 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 method 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 0.95 tolerance interval for probability 0.05 can be written $\pm 1.96\sqrt{0.05(1-0.05)/100\,000} = \pm 0.00135$. 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, $\hat{\boldsymbol{\Theta}}_{(K)}$ was computed and random $\mathrm{N}(0, \hat{\sigma}^2_{(K)})$ terms were repeatedly added. For examination of power, the actual model was assumed to have $\kappa > K$ terms. When $\kappa = K + 1$, then $\hat{\boldsymbol{\Theta}}_{(K+1)}$ was computed and random $\mathrm{N}(0, \hat{\sigma}^2_{(K+1)})$ 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

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

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

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

reason not be used. The cross-validation method was very conservative, but this method was not designed to control Type I error.

Considering power, Table 3 and Web Table 4, for AMMI and GGE, respectively, confirm the close agreement between the parametric bootstrap method and the JG/SM-test. Web Tables 5–10 include estimated power for the methods that did not control the Type I error rate, i.e. the cross validation method and the sequential F-test. The JG/SM-test rejected H_0 more often than the parametric bootstrap test, as expected considering the conclusion from Table 2 and Web Table 3 that the JG/SM-test tends to be slightly too liberal. When $K+1=\kappa$, the parametric bootstrap tests showed very high power (Tables 3 and Web Table 4). When $K+1<\kappa$, all tests encountered some cases with poor power, including the very liberal sequential F-test (Web Tables 5–10).

With the exception the AMMI analysis of the small Swedish pea dataset, the first multiplicative term was clearly significant in all analyses (Table 1 and Web Table 2). An additional simulation study was performed in order to investigate the performance of the two bootstrap methods when the first term is less distinct (Web Appendix B). The full and the simple parametric bootstrap methods performed similarly.

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.

The JG/SM-test was proposed together with a sequential testing procedure (Schott, 1986), under which terms are tested sequentially, beginning with $K = 0$ and continuing

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 Kth 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 probabilities 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 α_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 α_0. Provided that α_0 is close to the nominal significance level α, 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 $\mathbf{X}$ be a data matrix with I rows (observations) and J columns (variables). Let $\bar{\mathbf{X}}$ denote the matrix when each column in $\mathbf{X}$ has been standardized to have zero means, and $\mathbf{Z}$ the matrix when each column has been standardized to have zero means and unit standard deviations. PCA uses the sample covariance matrix of either $\bar{\mathbf{X}}$ or $\mathbf{Z}$. The variances of the computed principal components are proportional to the squared singular values of $\bar{\mathbf{X}}$ or $\mathbf{Z}$, depending on whether $\bar{\mathbf{X}}$ or $\mathbf{Z}$ 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 methods could be detected, we specifically advocate the use of the simple parametric bootstrap method, which is easier to program 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 Mathematical & Statistical Psychology* **31**, 144–152.

Cornelius, P. L. (1993). Statistical tests and retention of terms in the additive main effects and multiplicative interaction model for cultivar trials. *Crop Science* **33**, 1186–1193.

Cornelius, P. L. (1980). Functions approximating Mandel's tables for the means and standard deviations of the first three roots of a Wishart matrix. *Technometrics* **22**, 613–616.

Cornelius, P. L., Crossa, J., and Seyedsadr, M. S. (1996). Statistical tests and estimators of multiplicative models for genotype-by-environment interaction. In *Genotype-by-environment interaction*, M. S. Kang and H. G. Gauch (eds.), 199–231. Boca Raton: CRC Press.

Denis, J.-B., and Gower, J. C. (1994). Biadditive models. *Biometrics* **50**, 310–311.

Dias, C. T. D., and Krzanowski, W. J. (2003). Model selection and cross validation in additive main effect and multiplicative interaction models. *Crop Science* **43**, 865–873.

Dias, C. T. D., and Krzanowski, W. J. (2006). Choosing components 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 principal component analysis. *Technometrics* **24**, 73–77.

Efron, B., and Tibshirani, R. (1993). *An introduction to the bootstrap*. Boca Raton: Chapman & Hall/CRC.

Finlay, K. W., and Wilkinson, G. N. (1963). Analysis of adaptation in a plant-breeding programme. *Australian Journal of Agricultural Research* **14**, 742–754.

Gauch, H. G. (1992). *Statistical analysis of regional yield trials: AMMI analysis of factorial designs*. Amsterdam: Elsevier.

Gauch, H. G. (1988). Model selection and validation for yield trials with interaction. *Biometrics* **44**, 705–715.

Gauch, H. G. (2006). Statistical analysis of yield trials by AMMI and GGE. *Crop Science* **46**, 1488–1500.

Gauch, H. G., Piepho, H. P., Annicchiarico, P. (2008). Statistical analysis of yield trials by AMMI and GGE: Further considerations. *Crop Science* **48**, 866–889.

Gollob, H. F. (1968). A statistical model which combines features of factor analytic and analysis of variance techniques. *Psychometrika* **33**, 73–115.

Johnson, D. E., and Graybill, F. A. (1972). An analysis of a two-way model with interaction and no replication. *Journal of the American Statistical Association* **67**, 862–868.

Liu, G., and Cornelius, P. L. (2001). Simulations and derived approximations for the means and standard deviations of the characteristic roots of a Wishart matrix. *Communications 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

studies associated with the multiplicative interaction-model. *Communications in Statistics – Theory and Methods* **14**, 2219–2231.

Muirhead, R. J. (1978). Latent roots and matrix variates: a review of some asymptotic results. *Annals of Statistics* **6**, 5–33.

Ng, M. P., and Grunwald, G. K. (1997). Nonlinear regression analysis of the joint-regression model. *Biometrics* **53**, 1366–1372.

Piepho, H. P. (1995). Robustness of statistical tests for multiplicative terms in the additive main effects and multiplicative interaction model for cultivar trials. *Theoretical and Applied Genetics* **90**, 438–443.

Piepho, H. P. (1997). Analyzing genotype-environment data by mixed models with multiplicative terms. *Biometrics* **53**, 761–766.

Piepho, H. P. (1998). Empirical best linear unbiased prediction in cultivar trials using factor-analytic variance-covariance structures. *Theoretical and Applied Genetics* **97**, 195–201.

Piepho, H. P. (1999). Fitting a regression model for genotype-by-environment data on heading dates in grasses by methods for nonlinear mixed models. *Biometrics* **55**, 1120–1128.

Schott, J. R. (1986). A note on the critical-values used in stepwise tests for multiplicative components of interaction. *Communications in Statistics – Theory and Methods* **15**, 1561–1570.

Searle, S. R, Casella, G., and McCulloch, C. E. (1992). *Variance components*. Hoboken: Wiley.

Smith, A., Cullis, B., and Thompson, R. (2001). Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. *Biometrics* **57**, 1138–1147.

Tsujitani, M. (1992). A note on the additive and multiplicative models in two-way contingency tables. *Biometrics* **48**, 267–269.

Tukey, J. W. (1949). One degree of freedom for non-additivity. *Biometrics* **5**, 232–242.

van Eeuwijk, F. A. (1995). Multiplicative interaction in generalized linear models. *Biometrics* **51**, 1017–1032.

van Eeuwijk, F. A., and Kroonenberg, P. M. (1998). Multiplicative models for interaction in three-way ANOVA, with applications to plant breeding. *Biometrics* **54**, 1315–1333.

Wold, S. (1978). Cross-validatory estimation of the number of components in factor and principal components models. *Technometrics* **20**, 397–405.

Yan, W., Hunt, L. A., Sheng, Q., and Szlavnics, Z. (2000). Cultivar evaluation and mega-environment investigation based on the GGE biplot. *Crop Science* **40**, 597–605.

Yan, W., and Kang, M. S. (2002). *GGE biplot analysis: a graphical tool for breeders, geneticists, and agronomists*. Boca Raton: CRC Press.

Yan, W., Kang, M. S., Ma, B. L., Woods, S., and Cornelius, P. L. (2007). GGE biplot vs. AMMI analysis of genotype-by-environment data. *Crop Science* **47**, 643–655.

Yan, W., and Tinker, N. A. (2006). Biplot analysis of multi-environment trial data: Principles and applications. *Canadian Journal of Plant Science* **86**, 623–645.

Yang, R. C., Crossa, J., Cornelius, P. L., and Burgueno, J. (2009). Biplot analysis of genotype x environment interaction: proceed with caution. *Crop Science* **49**, 1564–1576.

Yates, F., and Cochran, W. G. (1938). The analysis of groups of experiments. *Journal of Agricultural Science* **28**, 556–580.

Yochmowitz, M. G., and Cornell, R. G. (1978). Stepwise tests for multiplicative components of interaction. *Technometrics* **20**, 79–84.

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