



This is an author produced version of a paper published in
Crop Science.

This paper has been peer-reviewed but may not include the final publisher
proof-corrections or pagination.

Citation for the published paper:

W. A. Malik, S. Hadasch, J. Forkman and H. P. Piepho. (2017)
Nonparametric Resampling Methods for Testing Multiplicative Terms in
AMMI and GGE Models for Multienvironment Trials. *Crop Science*.
Volume: 58, Number: 2, pp 752-761.
<http://dx.doi.org/10.2135/cropsci2017.10.0615>.

Access to the published version may require journal subscription.
Published with permission from: The American Crop Society.

Epsilon Open Archive <http://epsilon.slu.se>

Non-Parametric Resampling Methods for Testing Multiplicative Terms in AMMI and GGE Models for Multi-Environment Trials

W.A. Malik^{a*}, S. Hadasch^a, J. Forkman^b, H.P. Piepho^a

^a University of Hohenheim, Institute of Crop Science, Biostatistics Unit, Fruwirthstrasse 23, 70599 Stuttgart, Germany

^b Swedish University of Agricultural Sciences, Department of Crop Production Ecology, PO-Box 7043, 750 07 Uppsala, Sweden

* Corresponding Author:

Waqas Ahmed Malik

University of Hohenheim, Institute of Crop Science, Biostatistics Unit,
Fruwirthstrasse 23, 70599 Stuttgart, Germany

E-mail: w.malik@uni-hohenheim.de

Ph: +49 711 459 23021

Abstract:

The additive main effects and multiplicative interaction (AMMI) and genotype and genotype-environment interaction (GGE) models have been extensively used for the analysis of genotype-environment experiments in plant breeding and variety testing. Since their introduction, several tests have been proposed for testing the significance of the multiplicative terms, including a parametric bootstrap procedure. However, all of these tests are based on the assumptions of normality and homogeneity variance of the errors. In this paper, we propose tests based on non-parametric bootstrap and permutation methods. The proposed tests do not require any strong distributional assumptions. We also propose a test that can handle heterogeneity of variance between environments. The robustness of the proposed tests is compared with the robustness of other competing tests. The simulation study shows that the proposed tests always perform better than the parametric bootstrap method when the distributional assumptions of normality and homogeneity of variance are violated. The stratified permutation test can be recommended in case of heterogeneity of variance between environments.

Abbreviations: GEI, genotype \times environment interaction; AMMI, additive main effects and multiplicative interaction; GGE, genotype and genotype \times environment interaction; MET, multi-environment trial.

INTRODUCTION

Series of variety and plant breeding trials conducted at multiple environments, so-called multi-environment trials (MET), are the basis for the development and dissemination of new crop varieties. MET are important in order to test and identify genotypes for high and stable yield for general and specific adaptation. They are frequently analysed by linear models with effects for genotype, environment and their interaction. The pattern of genotype-environment interaction in MET data is usually of particular interest because it determines the adaptive pattern of a genotype's response to changing environments. Many models have been suggested for genotype-environment interaction effects in MET data (van Eeuwijk, 1995; Yang and Kang, 2002; Smith, et al., 2005). A very popular class of model has multiplicative terms for the genotype-environment interaction effects. The oldest model was proposed by Fisher and Mackenzie (1923) in the context of a factorial experiment with different types of manure and potato varieties which uses multiplicative terms for simple genotype-environment effects. The first proposal for modelling interaction in MET data was based on a regression of genotypes yields on an environment index computed from the mean of all genotypes at an environment. This model was proposed by Yates and Cochran (1938) and was popularized by Finlay and Wilkinson (1963). Williams (1952) and then Pike and Silverberg (1952) used the analysis of variance and singular value decomposition (SVD) to estimate multiplicative terms for interaction. Perhaps the most popular class of models, introduced by Mandel (1971), was labelled by Gauch (1988) as Additive Main Effects and Multiplicative Interactions (AMMI). The model was introduced in a plant breeding context by Kempton (1984) in connection with graphical displays of genotype-environment data, so-called biplots (Bradu and Gabriel, 1978; Kempton, 1984; Gauch, 1988; Yan and Kang, 2002; Laffont et al., 2013), which represent both genotypes and environments simultaneously and allow interactions to be studied graphically, and it used with both AMMI and GGE. The AMMI model has main effects for both genotypes and environments as well as multiplicative terms for the interaction. A further class of models that has a main effect only for environments, but no main effect for genotypes, are so-called Genotype and Genotype-Environment (GGE) models (Yan et al., 2000, Yan and Kang, 2002), also known as sites-regression (SREG) models (Cornelius et al., 1996).

AMMI and GGE models play an important role in the analysis of plant breeding and variety trials. Gauch (2013) provides a protocol for AMMI analysis of MET data. The merits of AMMI and GGE models have been discussed by Gauch (2006), Yan et al. (2007), Gauch et

al. (2008) and Gauch (2013). In AMMI models, the genotype-by-environment interaction is decomposed by a singular value decomposition (SVD) of the matrix of residuals after removing the genotype and environment main effects. The AMMI model can be written as

$$y_{ij} = \mu + \alpha_i + \beta_j + \sum_{k=1}^K \gamma_{ik} \lambda_k \delta_{jk} + r_{ij}, \quad (1)$$

where y_{ij} is the observed response (or mean yield) of the j -th genotype in the i -th environment ($i = 1, 2, \dots, I; j = 1, 2, \dots, J$), μ is the overall mean, α_i and β_j are the main effect of the i -th environment and j -th genotype, r_{ij} is the associated random residual error, $\lambda_k \geq 0$ ($k = 1, 2, \dots, K$) is the k -th singular value, and γ_{ik} and δ_{jk} are the scores associated with the i -th environment and j -th genotype, subject to the usual estimability constraints. The singular values λ_k are sorted in decreasing order. The sum of the multiplicative terms represents the interaction effect. Usually, the first few multiplicative terms show the systematic pattern and the remaining terms are just random noise. The GGE class of models is obtained by dropping the genotype main effect β_j from equation (1) and applying SVD to the data matrix centered on the environmental means.

The maximum possible number of multiplicative terms is $M = \min(I - 1, J - 1)$ for AMMI models, whereas it is $M = \min(I - 1, J)$ in GGE models. To maximize predictive accuracy, it is usually preferable to use a smaller number of terms (K) (Gauch, 1988). If too few terms are selected, predictions may be severely biased. If too many terms are selected, there is a danger of overfitting, entailing the risk of variance inflation. Thus, the challenge is to strike the right balance between bias and variance. Therefore, deciding how many multiplicative terms should be included in the model is a decisive step. Many procedures have been proposed for this purpose, including various significance tests (Gollob, 1968; Boik, 1993; Cornelius, 1993; Piepho, 1995; Forkman and Piepho, 2014) and cross-validation (Gauch, 1988; Dias and Krazanowski, 2003, 2006). The empirical type-I error rate and power of significance tests for unreplicated data has been studied in Forkman and Piepho (2014), and a comparison of cross-validation methods for replicated data can be found in Hadasch et al. (2017).

Forkman and Piepho (2014) proposed a significance test for the multiplicative terms of the AMMI or GGE model using a parametric bootstrap procedure assuming that the observed data follow a normal distribution with expected value given by the GGE or AMMI models. The null distribution is simulated using parameter estimates obtained from the observed data

under the null hypothesis. A large number of simulated random samples, i.e. bootstrap samples, are generated from the estimated model. Forkman and Piepho (2014) assume that residual errors are independently normally distributed with homogeneous variance. However, it is very likely that these assumptions do not hold for real data, especially when cultivar trials are performed in different locations with heterogeneous variance. Forkman and Piepho (2015) and Piepho (1995) showed that these types of test perform poorly when the assumptions of normality and homogeneity of variance are violated. In this case, the required conditions for a parametric test are not satisfied and a test that does not depend on parametric assumptions is desirable.

The aim of this article is to suggest robust procedures for testing the significance of the multiplicative terms. Specifically, we propose tests based on non-parametric bootstrap and permutation methods (Efron and Hastie, 2016; Good, 2005). Both methods are based on resampling of the residuals.

MATERIAL AND METHODS

To describe the method, it is convenient to rewrite model (1) in matrix form as:

$$\mathbf{Y} = \mathbf{1}_I \mu \mathbf{1}_J^T + \boldsymbol{\alpha} \mathbf{1}_J^T + \mathbf{1}_I \boldsymbol{\beta}^T + \mathbf{E} \quad (2)$$

where $\mathbf{Y} = \{y_{ij}\}$ is the $I \times J$ matrix of genotype-environment means y_{ij} . Let $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_I)^T$ be a vector of environment main effects, $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_J)^T$ a vector of genotype main effects and $\mathbf{1}_n$ is an n -vector of ones. The term μ is an overall mean. The $I \times J$ matrix \mathbf{E} is the sum of $\boldsymbol{\Theta}_{(\kappa)}$ and \mathbf{R} , i.e., $\mathbf{E} = \boldsymbol{\Theta}_{(\kappa)} + \mathbf{R}$, where $\boldsymbol{\Theta}_{(\kappa)}$ models the genotype-environment interaction and $\mathbf{R} = \{r_{ij}\}$ is an $I \times J$ matrix of random errors r_{ij} . In (1), the term $\sum_{k=1}^{\kappa} \gamma_{ik} \lambda_k \delta_{jk} = \boldsymbol{\Theta}_{(\kappa)}$ models the genotype-environment interaction, and it is based on a SVD of $\boldsymbol{\Theta}_{(\kappa)}$ which comprises κ multiplicative terms that are not zero. Also, $\sum_{k=1}^{\kappa} \gamma_{ik} \lambda_k \delta_{jk}$ can be written as $\mathbf{U}_{(\kappa)} \boldsymbol{\Lambda}_{(\kappa)} \mathbf{V}_{(\kappa)}^T$, where $\mathbf{U}_{(\kappa)}$ and $\mathbf{V}_{(\kappa)}$ are $I \times \kappa$ and $J \times \kappa$ matrices of the κ left and right singular vectors, respectively, and $\boldsymbol{\Lambda}_{(\kappa)}$ is a diagonal matrix that contains the κ singular values sorted in decreasing order.

The estimates of the overall mean (μ) and the main effects (α and β) are obtained by a simple two-way ANOVA of \mathbf{Y} . The resulting residual matrix $\hat{\mathbf{E}}$ is of rank M . Using singular value decomposition, $\hat{\mathbf{E}}$ can be written as $\hat{\mathbf{E}} = \hat{\mathbf{U}} \hat{\boldsymbol{\Lambda}} \hat{\mathbf{V}}^T$, where $\hat{\mathbf{U}}$ and $\hat{\mathbf{V}}$ are $I \times M$ and $J \times M$ matrices of the estimated left and right singular vectors, respectively, and $\hat{\boldsymbol{\Lambda}}$ is an $M \times M$ diagonal matrix of estimated singular values. If κ multiplicative terms are contained in the model then it is called an AMMI κ model.

The number of multiplicative terms appropriate for a given data set may be determined by a significance test of the null hypothesis that κ equals some specified value K against the alternative hypothesis that κ is larger than K . The test statistic T , suggested by Yochmowitz and Cornell (1978) and Schott (1986), was used by Forkman and Piepho (2014) in their parametric bootstrap methods, where $T = \hat{\lambda}_{K+1}^2 / \sum_{k=K+1}^M \hat{\lambda}_k^2$. In a parametric bootstrap method, the null distribution is simulated using parametric estimates under the null hypothesis (Efron and Tibshirani, 1993). A large number of samples are drawn from the fitted model using the estimates of model parameters. For each of these bootstrap samples, the test statistic T , say T_b , is calculated, where b indexes the b th bootstrap sample. The obtained distribution of T_b approximates the true distribution of T under the null hypothesis.

We here assume that the first K terms from the singular value decomposition represent the true fixed interaction $\Theta_{(K)}$ for all bootstrap samples and the random error \mathbf{R} follows an independent normal distribution with mean zero and constant variance $\sigma_{(K)}^2$ (Forkman and Piepho, 2014). The estimates of $\Theta_{(K)}$ and $\sigma_{(K)}^2$ are obtained as

$$\hat{\Theta}_{(K)} = \begin{cases} 0, & \text{if } K = 0, \\ \hat{\mathbf{U}}_{(K)} \hat{\Lambda}_{(K)} \hat{\mathbf{V}}_{(K)}^T & \text{if } K > 0. \end{cases} \quad (3)$$

and

$$\hat{\sigma}_{(K)}^2 = \frac{1}{v} \sum_{k=K+1}^M \hat{\lambda}_k^2, \quad 0 \leq K \leq M - 2, \quad (4)$$

where $v = (I - 1)(J - 1)$ for AMMI and $v = I(J - 1)$ for GGE analysis. The assumption of the parametric bootstrap method that random errors are normally distributed with constant variance, i.e. $\sigma_{(K)}^2$, is very strict, and it may be violated in agricultural experiments. In particular, there is likely to be heterogeneity of variance in multi-environment trials. Forkman and Piepho (2015) investigated the type-I error rate of the parametric bootstrap method when random errors are not normally distributed and heterogeneity of variance occurs. They found that the type-I error rate can be well above the nominal 5% rate when random error exhibits departure from normality and the error rate is also adversely affected by departures from homogeneity of variance. Therefore, it is desirable to have a test which is not based on such assumptions.

NON-PARAMETRIC BOOTSTRAP METHOD

Here we propose a non-parametric bootstrap procedure for computing the p -value for testing the K multiplicative term under the null hypothesis $H_0: \kappa = K$. Following Efron and Tibshirani (1993), the null distribution is simulated using parameter estimates under the null hypothesis. In first step, the fixed-effects parameters are estimated from the data and the residuals are computed. Then, a large number of samples, bootstrap samples, are drawn from the fitted model under the null hypothesis as follows. The first K terms from the SVD represent the true fixed interaction $\Theta_{(\kappa)}$ in all bootstrap samples, which is estimated using (3). The residual error of the b -th bootstrap sample, \mathbf{R}_b^B , is drawn randomly from the observed \mathbf{R} and added to the fixed interaction $\Theta_{(\kappa)}$ to obtain simulated data $\Theta_{(\kappa)} + \mathbf{R}_b^B$. We do not add main effects here because they are swept out before the SVD to estimate the interaction effect so their value is immaterial to the behavior of the test. Then, the test statistic T is computed to obtain the bootstrap test statistic T_b , which is compared to the observed test statistic T (based on actual data). The proportion of statistics T_b higher than the observed statistic T gives the p -value. Note that the parametric bootstrap methods are based on the assumption that residual errors, \mathbf{R} , are normally distributed with constant variance, whereas the non-parametric bootstrap method only assumes the exchangeability of residuals.

The algorithm for the nonparametric bootstrap method is given as follows:

- Step i. Compute $\hat{\mathbf{E}}$ and perform the SVD on $\hat{\mathbf{E}}$.
 - Step ii. Compute T using the singular values from Step i.
 - Step iii. Compute $\hat{\Theta}_{(K)}$ and $\hat{\Theta}_{(K)}^-$, where $\hat{\Theta}_{(K)}^- = \sum_{k=K+1}^M \hat{\gamma}_k \hat{\lambda}_k \hat{\delta}_k^T$ is an $I \times J$ matrix.
- For** $b = 1, 2, \dots, B$, **Do**:
- Step iv. Sample an $I \times J$ matrix \mathbf{R}_b^B from the elements of $\hat{\Theta}_{(K)}^-$ *with replacement*.
 - Step v. Compute $\mathbf{E}_b^B = \hat{\Theta}_{(K)} + \mathbf{R}_b^B$.
 - Step vi. Compute $\hat{\mathbf{E}}_b^B$ using

$$\hat{\mathbf{E}}_b^B = \mathbf{E}_b^B - \bar{\mathbf{E}}_{bi}^B - \bar{\mathbf{E}}_{b,j}^B + \bar{\mathbf{E}}_{b..}^B$$
 for AMMI analysis and,

$$\hat{\mathbf{E}}_b^B = \mathbf{E}_b^B - \bar{\mathbf{E}}_{bi}^B + \bar{\mathbf{E}}_{b..}^B$$
 for GGE analysis,
 where $\bar{\mathbf{E}}_{b..}^B$ is the general mean matrix, and $\bar{\mathbf{E}}_{bi}^B$ and $\bar{\mathbf{E}}_{b,j}^B$ are row (environment) and column (genotype) mean matrices from \mathbf{E}_b^B , respectively.
 - Step vii. Perform the SVD of $\hat{\mathbf{E}}_b^B$.
 - Step viii. Compute T_b in the same way as T using singular values from Step vii.

End For

Step ix. Estimate the p -value as the observed frequency of $T_b > T$.

It is often observed in multi-environment trials that the error variances are heterogeneous among environments. In the case of heterogeneity among environments, stratification of environments can be done in Step iv of the algorithm given above, i.e. the bootstrapping or permutation can be performed within each stratum (environment).

PERMUTATION-BASED METHOD

As discussed by Good (2004), a requirement of permutation tests is the exchangeability of data under the null hypothesis. In our case, the elements of errors E in Eq. (2) are exchangeable, but strictly speaking the elements of residuals \hat{E} are not exchangeable due to differences in the effects that need to be subtracted to obtain these residuals. Nevertheless, we may consider a permutation procedure, the expectation being that the elements of residuals \hat{E} approximately behave as the exchangeable elements of E , thus providing some robustness under violations of the normality and homogeneity of variance assumptions. The test proposed in the previous section can also be based on the permutation of residuals rather than bootstrapping. Note that if the residuals are sampled with replacement, we have a bootstrapping procedure, whereas if sampling is without replacement, we obtain a permutation procedure.

Therefore, we suggest the same test given in the previous section but now based on the permutation method. Thus, for the permutation test, the same algorithm can be used, however, the residual error in Step iv is obtained by taking a sample drawn at random without replacement.

Example

Description of Datasets

The non-parametric method outlined above is compared with the parametric bootstrap method using five real datasets from the literature on multi-environment trials.

- i. A New York soybean dataset with seven genotypes investigated in ten environments (Gauch, 1992; Zobel et al., 1988).
- ii. An international maize and wheat improvement center (CIMMYT) maize dataset with nine genotypes in 20 environments (Cornelius et al., 1996).
- iii. An Ontario winter wheat dataset with 18 genotypes in nine environments (Yan et al., 2007; Yan and Tinker, 2006).
- iv. A corn yield data with 20 genotypes and seven locations (Corsten and Denis, 1990).
- v. A pea dataset with five pea genotypes and four trails (Forkman and Piepho, 2014).

The first four datasets are available in R package “agridat” and the pea dataset is available in the supplementary section of Forkman and Piepho (2014). Table 1 gives an overview of AMMI analyses of these datasets. The sum of squares explained by a multiplicative term and percentage of sum of squares explained is given. The results for testing different multiplicative terms of the AMMI model using parametric and non-parametric methods are given in Table 2. The p -values were derived using 100,000 bootstrap samples. An R code function is available in the supplementary material to use the proposed method. The function requires less than a minute for a 300×100 data matrix with 100,000 bootstrap samples using a standard PC.

The first two multiplicative terms are identified as significant for the New York soybean data and the Ontario wheat data with both the parametric and non-parametric tests. For the CIMMYT maize data and the corn yield data, all methods agree that the AMMI model with one term is appropriate. However, no significant interaction is found in the Swedish pea data. The non-parametric bootstrap methods most often resulted in slightly higher p -values as compared to the parametric bootstrap method. However, the same numbers of multiplicative terms were identified as significant by the parametric and the non-parametric methods.

[Table 1 about here]

[Table 2 about here]

SIMULATION STUDY

To investigate the robustness of the simple parametric and non-parametric test to non-normality and heterogeneity of error terms, simulations were conducted using the *R* language environment, each with 100,000 simulations and $B = 5,000$ bootstrap runs. To make the simulation realistic, they were based on models fitted to the real data sets studied in this paper.

Suppose that y_{ij} is a response in a two-way table for a given data set. The simplest model included only an intercept and a random error. To test the multiplicative effects, we fitted the AMMIK model for a given value of K , that is

$$y_{ij} = \mu_{ij(K)} + r_{ij} = \mu + \alpha_i + \beta_j + \theta_{ij(K)} + r_{ij}$$

The overall mean was estimated as $\hat{\mu} = \bar{y}_{..}$, the environment effect as $\hat{\alpha}_i = \bar{y}_{i.} - \bar{y}_{..}$ and the genotype effect as $\hat{\beta}_j = \bar{y}_{.j} - \bar{y}_{..}$. The interaction $\theta_{ij(K)}$ was estimated from the first K components of the SVD of $\mathbf{E} = y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}$ using (3). The model assumes $r_{ij} \sim \text{dist}(0, \hat{\sigma}_{(K)}^2)$, where *dist* is the specified distribution. The residual variance $\hat{\sigma}_{(K)}^2$ was computed using (4). Type-I error rates were estimated through simulation by assuming that the actual model has $\kappa = K$ terms ($K = 0, 1, \dots, 4$), and the significance of the $(K + 1)$ th term was tested at the 5% significance level.

In this way, it was possible to generate repeated data sets which mimic the patterns of the original data and have a specified number of interaction components. Also, by varying the distribution of error, it was possible to assess the robustness of tests when the error term belonged to different distributions. The simulations are made using nine different cases for drawing random errors r_{ij} . The same distributions which were used by Forkman and Piepho (2015) were used in these simulations. A description of these distributions and error variance used in the simulation is given in Table 3. The distributions were scaled so that the mean is $\hat{\mu}_{ij(K)}$ and the variance equals to $\hat{\sigma}_{(K)}^2$, and in the cases with heterogeneous variance among environments $\hat{\sigma}_{(K)}^2 = \sum_i \hat{\sigma}_{i(K)}^2 / I$, where $\hat{\sigma}_{i(K)}^2$ is the error variance in the i -th environment. In the case of heterogeneity among environments, the non-parametric bootstrapping or permutation was performed within each stratum.

To investigate the effect of the data dimension on the type-I error rate of the non-parametric bootstrap method, simulations were also performed based on datasets of increased dimensions. The data dimension was increased by combining the AMMI0 model fitted to the

Soya bean data matrix two times, three times, four times and five times, simultaneously in both row and column directions. In this way, the numbers of environments and genotypes were increased. Finally, the error term from the exponential distributions (Table 3) is added.

[Table 3 about here]

RESULTS AND DISCUSSION

The comparison of the simple parametric bootstrap method with the non-parametric bootstrap and permutation methods is done in terms of the observed frequency of type-I errors. The results are shown in Figures 1, 2 and 3. Each plot represents results using a specified distribution for the error terms. The three bars in a group represent three methods, and the K -th group ($K = 0, 1, 2, 3, 4$) shows the assumed AMMIK model.

The results of the simulation of the different distributions (Table 3) with homogenous error variance are presented in Figure 1. The results for normally distributed errors are given in Figure 1(a), for the lognormal in Figure 1(b), gamma in Figure 1(c), uniform in Figure 1(d) and exponential in Figure 1(e). For all non-normal distributions, the type-I error rates of the non-parametric bootstrap and the permutation methods are nearer to the nominal 5% as compared to the parametric bootstrap method. The same pattern of results is obtained for all three data sets. When the assumptions of normality and homogeneity of variance are satisfied, the non-parametric bootstrap and the permutation methods maintain the type-I error rates close to the nominal 0.05 significance level, as does the parametric bootstrap method. For gamma distributed error terms (Figure 1c), the non-parametric bootstrap and permutation methods gave much lower and more correct type-I error rates than the parametric method. The exponential distribution is an extreme case that is not plausible in crop variety trials. For this distribution, the parametric bootstrap method shows a very high type-I error rate while the non-parametric bootstrap and permutation methods show much lower, but still high type-I error rates (Figure 1e). For all three methods, the type-I error rate decreases with increasing number K of multiplicative terms. The parametric bootstrap method produces a too low type-I error rate for uniformly distributed errors (Figure 1d); however, the non-parametric bootstrap and permutation methods give error rates closer to the nominal 5% rate. In all these cases, the non-parametric bootstrap and permutation methods perform similarly.

[Figure 1 about here]

The results with heterogeneous error variance are presented in Figure 2. For the normal distribution with heterogeneous variance (i.e. homogenous CV), the type-I error rates of the non-parametric bootstrap and permutation methods are much lower than that of the parametric bootstrap method (Figure 2a), although they are higher than 0.05 but decrease with the number of multiplicative terms. The simulated lognormal distributions are slightly skewed (Forkman and Piepho, 2015), therefore, all three methods maintain the type-I error

rate close to 5% when the assumption of homogeneity is fulfilled (Figure 1b). However, when the error term is simulated with heterogeneous variance (Figure 2b), the type-I error rate is much higher than 0.05 for all three methods, but the type-I error rate of the non-parametric bootstrap and permutation methods is almost half of the error rate for the parametric method.

The results for normally distributed errors with heterogeneous variance among environments are shown in Figure 2(c). The type-I error rate is also quite high for all three methods (although the non-parametric bootstrap and permutation methods do much better than the parametric bootstrap method). To overcome this problem, each environment is considered as a stratum and non-parametric bootstrapping and permutation are done within each stratum separately. The results given in Figure 2(d) indicate that the stratified non-parametric bootstrap method is conservative, whereas the permutation method performs better than the non-parametric bootstrap method.

[Figure 2 about here]

The case of non-normally distributed errors with heterogeneous variance among environments is considered in Figure 3. The non-parametric bootstrap and permutation method performed much better, with regard to type-I error, than the parametric bootstrap method. Figure 3(c) shows that these tests perform much better using stratification of environments.

[Figure 3 about here]

To investigate the effect of data dimension on the type-I error rate of the non-parametric bootstrap method, the simulation is performed by increasing the data dimension. The results from different data sizes with error terms from the exponential distribution and testing AMMI0 models are given in Figure 4. From this figure, it is visible that the type-I error rate gets closer to 5% as the size of the data matrix increases.

For testing the main effects of environment and genotypes, the observed type-I error rates of these methods are close to the nominal 5% significance level, which was already confirmed by Forkman and Piepho (2015).

[Figure 4 about here]

CONCLUSION

The replication of treatments is a basic principle of experimental design which permits calculating an error mean square. In this paper, however, we consider unreplicated data. But yet, because of the structured form of interaction in AMMI models, residuals, which form the basis of our resampling procedure, can be computed even though we have no replication at the genotype-by-environment level. It should be emphasized, however, that wherever feasible, replication is preferable because it is expected to allow more robust inferences than are afforded by unreplicated data.

In this paper, we were concerned with statistical tests for finding the number of multiplicative terms in AMMI and GGE models for analyzing unreplicated genotype-environment data. The introduced tests based on the non-parametric bootstrap and permutation methods are straightforward. No strong assumptions are needed. The results of the simulation study indicate that these methods have better performance than the parametric bootstrap method with regard to the type-I error rate. When the distribution of the observations is non-normal or the data is heteroscedastic, the parametric bootstrap method cannot maintain the nominal type-I error rate. However, the non-parametric bootstrap and permutation method performed much better, especially when used with stratification by environment. In conclusion, the non-parametric bootstrap and permutation methods always perform better or at least as good as the parametric bootstrap method. The stratified permutation test is the recommended test when observations are heteroscedastic or not normally distributed. While these non-parametric tests provide robustness to violations of the usual assumptions, control of the type I error rate is not exact, so there is room for further improvement and the development of further improved procedures is highly desirable. Also, the investigation of our proposed methods has been limited to the balanced case. It would be worthwhile to check their performance under a range of scenarios with unbalanced data.

Lastly, the choice of statistical analysis should be governed by the agricultural research purpose followed by statistical recommendations. As pointed out by Gauch (2006), the research purposes are the masters whereas statistical methods are the servants. Gauch (2013) identified three criteria for model diagnosis in AMMI literature, i.e., statistical significance, agricultural interpretability, and predicative accuracy. There are different possible criteria for

different agricultural purposes even for a given dataset. The statistical significance test tries to optimize predicative accuracy. This is the main research purpose we are addressing. Multiplicative terms that are not significant are automatically not relevant for any research purpose, however, specific research purposes such as delineating mega-environment may call for fewer terms than those that are statistically significant. Therefore, it is important to consider the research purpose and the properties of the dataset when diagnosing the multiplicative terms to retain.

ACKNOWLEDGMENT

This work was supported by DFG grant PI 377/1-17.

REFERENCES

- Boik, R.J. 1993. A comparison of three invariant tests of additivity in two-way classifications with no replications. *Comp. Stat. Data Anal.* 15:411–424.
- Bradu, D., and Gabriel, K. 1978. The biplot as a diagnostic tool for models of two-way tables. *Technometrics* 20:47–68.
- Corsten, L.C.A., and Denis, J.B. 1990. Structuring interaction in two-way tables by clustering. *Biometrics* 46:207–215.
- Cornelius, P.L. 1993. Statistical tests and retention of terms in the additive main effects and multiplicative interaction model for cultivar trials. *Crop Sci.* 33:1186–1193.
- Cornelius, P.L., Crossa, J., and Seyedsadr, M.S. 1996. Statistical tests and estimators of multiplicative models for genotype-by-environment interaction. In M. S. Kang and H. G. Gauch (ed.) *Genotype-by-Environment Interaction*, CRC Press, Boca Raton, FL.
- Dias, C.T. dos S., and Krzanowski, W.J. 2006. Choosing components in the additive main effect and multiplicative interaction models. *Sci. Agri.* 63:169–175.
- Dias, C.T. dos S., and Krzanowski, W.J. 2003. Model selection and cross validation in additive main effect and multiplicative interaction models. *Crop Sci.* 43:865–873.
- Efron, B. 1979. Bootstrap methods: Another look at the jackknife. *Annals Statist.* 7:1–26.
- Efron, B., and Hasti, T. 2016. *Computer Age Statistical Inference: Algorithms, Evidence, and Data Science*. Cambridge University Press, Cambridge.
- Efron, B., and Tibshirani, R. 1993. *An Introduction to the Bootstrap*. Chapman & Hall/CRC, Boca Raton, FL.
- Finlay, K.W., and Wilkinson, G. N. 1963. Analysis of adaptation in a plant-breeding programme. *Australian J. Agri. Res.* 14:742–754.
- Fisher, R., and Mackenzie, W. 1923. Studies in crop variation. II. The manurial response of different potato varieties. *J. Agri. Sci.* 13:311–320.
- Forkman, J., and Piepho, H.-P. 2015. Robustness of the simple parametric bootstrap method for the additive main effects and multiplicative interaction (AMMI) model. *Biuletyn Oceny Odmian* 34:11–18.
- Forkman, J., and Piepho, H.-P. 2014. Parametric bootstrap methods for testing multiplicative terms in GGE and AMMI models. *Biometrics* 70:639–647.
- Gauch, H.G. 1992. *Statistical Analysis of Regional Yield Trials: AMMI Analysis of Factorial Designs*. Elsevier, Amsterdam.

- 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 Sci.* 46:1488–1500.
- Gauch, H.G. 2013. A Simple protocol for AMMI analysis of yield trials. *Crop Sci.* 53:1860–1869.
- Gauch, H.G., Piepho, H.-P., and Annicchiarico, P. 2008. Statistical analysis of yield trials by AMMI and GGE: Further considerations. *Crop Sci.* 48:866–889.
- Good, P. I. 2005. *Permutation, Parametric and Bootstrap Tests of Hypotheses*. Springer, New York.
- Hadasch S., Forkman, J., and Piepho, H-P. 2017. Cross-validation in AMMI and GGE models: a comparison of methods. *Crop Sci.* 57:264–274
- Kempton, R.A. (1984). The use of biplots in interpreting variety by environment interactions. *J. Agri. Sci.* 103:123–135.
- Laffont, J.L., Wright, K., and Hanafi, M. 2013. Genotype plus genotype-by-block of environments biplots. *Crop Sci.* 53: 2332–41.
- Mandel, J. 1971. A new analysis of variance model for nonadditive data. *Technometrics* 13:1–18.
- Pesarin, F. 2001. *Multivariate Permutation Tests: With Applications in Biostatistics*. Wiley Chichester.
- Piepho, H.P. 1995. Robustness of statistical tests for multiplicative terms in the additive main effects and multiplicative interaction model for cultivar trials. *Theor. Appl. Genet.* 90:438–443.
- Pike, E.W., and Silverberg, T.R. 1952. Designing mechanical computers. *Machine Design* 24:131–137, 159–163.
- Schott, J.R. 1986. A note on the critical-values used in stepwise tests for multiplicative components of interaction. *Comm. Stat. Theor. Meth.* 15:1561–1570.
- Smith A.B, Cullis, B.R., and Thompson, R. 2005. The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. *J. Agri. Sci.* 143:449–462.
- Van Eeuwijk, F.A. 1995. Linear and bilinear models for the analysis of multi-environment trials: I. An inventory of models. *Euphytica* 84:1–7.
- Williams, E.J. 1952. The interpretation of interactions in factorial experiments. *Biometrika* 39:65–81.

- Yan, W., Hunt, L.A., Sheng, Q., and Szlavnic, Z. 2000. Cultivar evaluation and mega-environment investigation based on the GGE biplot. *Crop Sci.* 40:597–605.
- Yan, W., and Kang, M.S. 2002. *GGE Biplot Analysis: A Graphical Tool for Breeders, Geneticists, and Agronomists*. CRC Press, Boca Raton, FL.
- 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 Sci.* 47:643–655.
- Yates, F., and Cochran, W. G. 1938. The analysis of groups of experiments. *J. Agri. Sci.* 28:556–580.
- Yochmowitz, M. G., and Cornell, R. G. 1978. Stepwise tests for multiplicative components of interaction. *Technometrics* 20: 79–84.
- Zobel, R.W., Wright, M.J., and Gauch, H.G. 1998. Statistical analysis of a yield trial. *Agron. J.* 80:388–393.

Table 1: Sums of squares ($\hat{\lambda}^2$) in AMMI analyses of real datasets.

Singular value	New York Soybean data		CIMMYT maize data		Ontario wheat data		Corn yield data		Swedish pea data	
	$\hat{\lambda}^2$	%	$\hat{\lambda}^2$	%	$\hat{\lambda}^2$	%	$\hat{\lambda}^2$	%	$\hat{\lambda}^2$	%
1	8,189,065	82.4	35,078,698	56.2	9.616	48.2	880.13	41.74	69,259	73.7
2	1,170,288	11.4	9,426,035	15.1	4.652	23.3	545.25	25.86	20,720	22.0
3	254,964	2.6	6,515,627	10.4	1.933	9.7	259.61	12.31	3,997	4.3
4	200,449	2.0	5,383,947	8.6	1.249	6.3	231.01	10.96	–	–
5	107,532	1.1	3,091,503	5.0	1.083	5.4	133.05	6.31	–	–

Table 2: AMMI analyses of the New York soybean, CIMMYT maize, Ontario wheat, Swedish pea and Corn yield datasets using the simple parametric bootstrap method, the non-parametric bootstrap method and the permutation-based method.

Term $K + 1$	Test statistic T	P -value		
		Parametric	Non-parametric	Permutation
New York soybean data				
1	0.824	0.0000	0.0000	0.0000
2	0.671	0.0050	0.0221	0.0294
3	0.445	0.8650	0.8743	0.8753
4	0.631	0.4700	0.4792	0.4809
5	0.916	0.0960	0.0998	0.0995
CIMMYT maize data				
1	0.562	0.0000	0.0000	0.0000
2	0.345	0.1560	0.1876	0.1885
3	0.364	0.2720	0.3246	0.3277
4	0.472	0.0460	0.0654	0.0677
5	0.514	0.1110	0.1211	0.1211
Ontario wheat data				
1	0.482	0.0000	0.0000	0.0000
2	0.450	0.0030	0.0025	0.0025
3	0.340	0.5800	0.6659	0.6779
4	0.333	0.9050	0.9034	0.9057
5	0.433	0.6100	0.6439	0.6484
Corn yield data				
1	0.417	0.0244	0.0354	0.0363
2	0.444	0.0693	0.0572	0.0579
3	0.380	0.8134	0.8172	0.8217
4	0.546	0.2902	0.3151	0.3185
5	0.691	0.3309	0.3673	0.3694
Swedish pea data				
1	0.737	0.3620	0.3027	0.2922
2	0.838	0.5440	0.5338	0.5356

Table 3: Distributions used in the simulation study.

Distribution	Test of multiplicative effects
1. Normal distribution with homogeneous variance	$y_{ij} \sim N(\hat{\mu}_{ij(K)}, \hat{\sigma}_{(K)}^2)$
2. Normal distribution with heterogeneous variance (homogeneous coefficient of variation)	$y_{ij} \sim N(\hat{\mu}_{ij(K)}, \gamma_{(K)}^2 \hat{\mu}_{ij(K)}^2)$ $\gamma_{(K)} = \hat{\sigma}_{(K)} / \hat{\mu}$
3. Normal distribution with heterogeneous variance among environments	$y_{ij} \sim N(\hat{\mu}_{ij(K)}, \hat{\sigma}_{i(K)}^2),$
4. Lognormal distribution with homogeneous variance	$y_{ij} \sim \ln N(m_{(K)}, s_{(K)}^2),$ $m_{(K)} = \log(\hat{\mu}_{ij(K)}) - s_{(K)}^2/2,$ $s_{(K)}^2 = \log(\hat{\sigma}_{(K)}^2 / \hat{\mu}_{ij(K)} + 1)$
5. Lognormal distribution with heterogeneous variance (homogeneous coefficient of variation)	$y_{ij} \sim \ln N(m_{(K)}, s_{(K)}^2),$ $m_{(K)} = \log \hat{\mu}_{ij(K)} - s_{(K)}^2/2,$ $s_{(K)}^2 = \log(\hat{\sigma}_{(K)}^2 / \hat{\mu}^2 + 1)$
6. Gamma distribution with shape parameter $\alpha = 4$	$y_{ij} \sim \hat{\mu}_{ij(K)} + X - 2\hat{\sigma}_{(K)},$ $X \sim \text{gamma}(4, \hat{\sigma}_{(K)}/2)$
7. Exponential distribution (i.e. Gamma distribution with shape parameter $\alpha = 1$)	$y_{ij} \sim \hat{\mu}_{ij(K)} + X - \hat{\sigma}_{(K)},$ $X \sim \text{gamma}(1, \hat{\sigma}_{(K)})$
8. Uniform distribution	$y_{ij} \sim \hat{\mu}_{ij(K)} + R,$ $R \sim \text{Uniform}\left(-\sqrt{3\hat{\sigma}_{(K)}^2}, \sqrt{3\hat{\sigma}_{(K)}^2}\right)$
9. Gamma distribution ($\alpha = 4$) with heterogeneous variance among environments	$y_{ij} \sim \hat{\mu}_{ij(K)} + X - 2\hat{\sigma}_{i(K)},$ $X \sim \text{gamma}(4, \hat{\sigma}_{i(K)}/2)$

FIGURE CAPTION

Figure 1: Observed type-I error rates of the parametric, non-parametric and permutation bootstrap methods for three datasets in columns and distributions in rows. The distribution of errors are (a) normal, (b) lognormal, (c) gamma ($\alpha = 4$), (d) uniform, and (e) exponential. All these distribution assumes homogenous error variance.

Figure 2: Observed type-I error rates of the parametric, non-parametric and permutation bootstrap methods for three datasets in columns and distributions in rows. The distribution of errors is normal for (a, c, d) and lognormal for (b). The heterogeneous error variance was used for cases (a) and (b), and heterogeneous error variance among environments are used for cases (c) and (d). In case (d), the bootstrapping and permutation is done within each environment.

Figure 3: Observed type-I error rates of the parametric, non-parametric and permutation bootstrap methods for three datasets in columns and distributions in rows. The gamma distribution with shape parameter ($\alpha = 4$) and heterogeneous error variance among environments are used for both cases. In case (b), the bootstrapping and permutation is done within each environment.

Figure 4: Observed type-I error rates by using non-parametric bootstrap methods for different sizes of the data matrix with exponentially distributed errors with homogeneous variance. The left most bar represents the error rate of original Soya bean data and the right most bar shows error rate of data that was increased by combining the original data matrix 5 times in both environment and genotype direction.

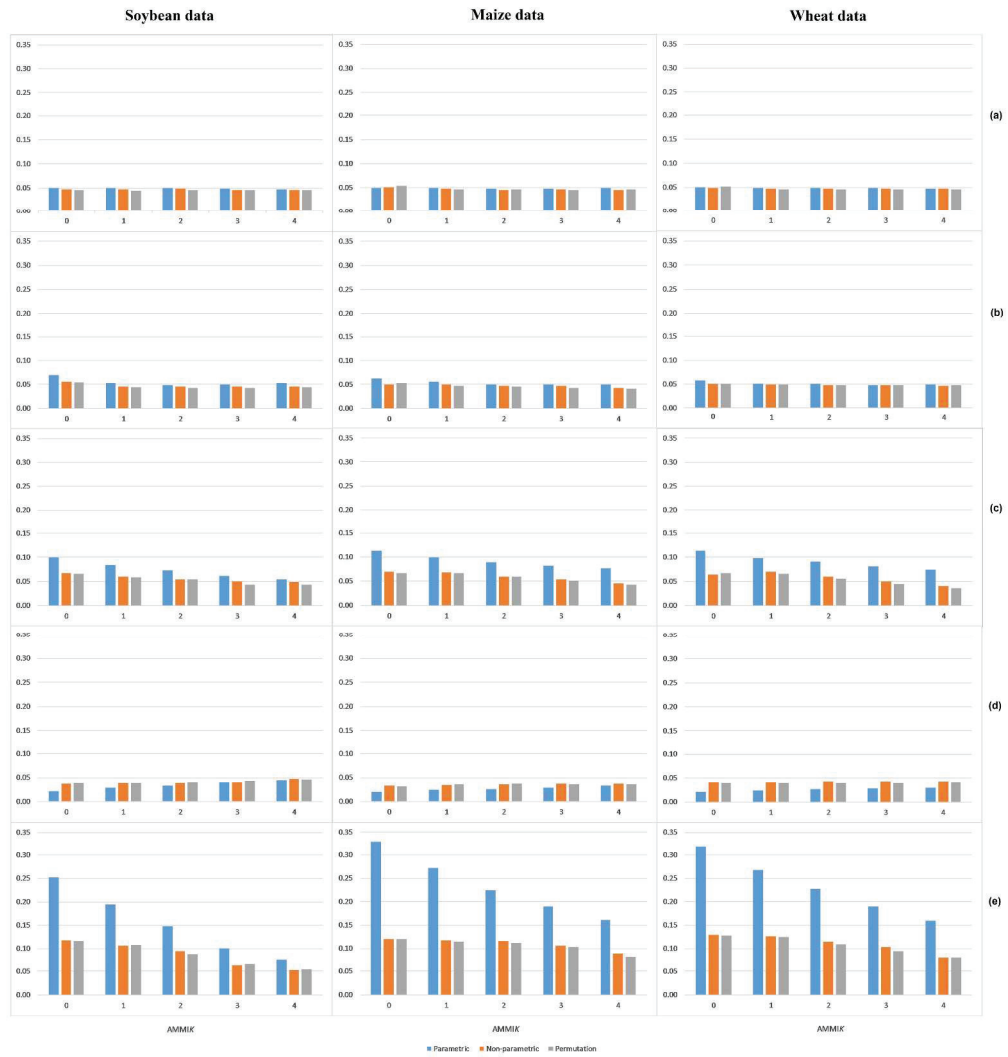


Figure 1: Observed type-I error rates of the parametric, non-parametric and permutation bootstrap methods for three datasets in columns and distributions in rows. The distribution of errors are (a) normal, (b) lognormal, (c) gamma ($\alpha=4$), (d) uniform, and (e) exponential. All these distribution assumes homogenous error variance.

417x444mm (300 x 300 DPI)

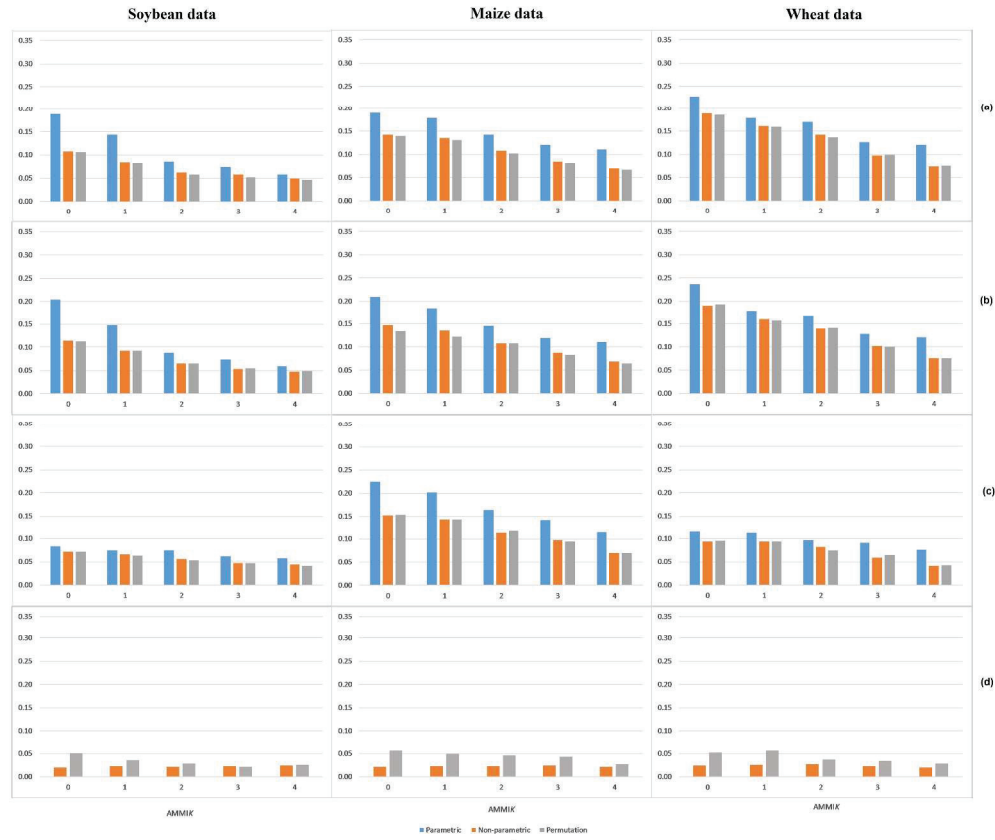


Figure 2: Observed type-I error rates of the parametric, non-parametric and permutation bootstrap methods for three datasets in columns and distributions in rows. The distribution of errors are normal for (a, c, d) and lognormal for (b). The heterogeneous error variance was used for cases (a) and (b), and heterogeneous error variance among environments are used for cases (c) and (d). In case (d), the bootstrapping and permutation is done within each environment.

334x281mm (300 x 300 DPI)

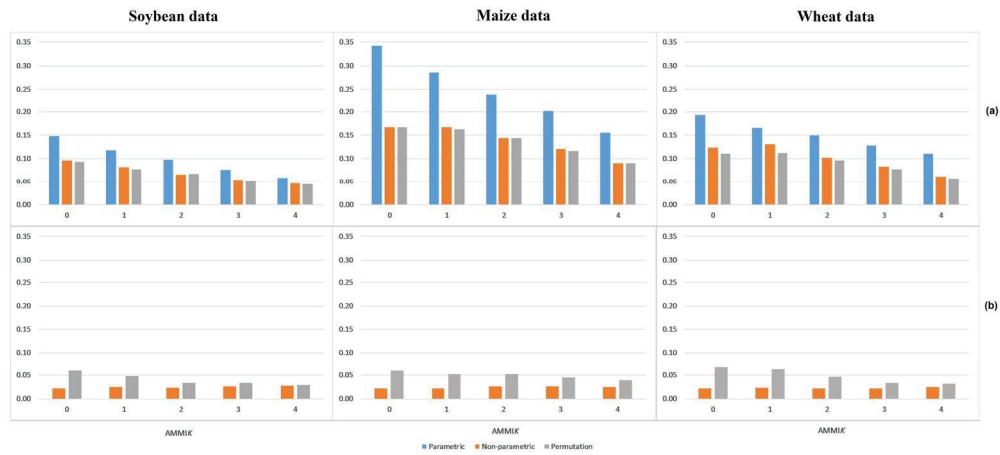


Figure 3: Observed type-I error rates of the parametric, non-parametric and permutation bootstrap methods for three datasets in columns and distributions in rows. The gamma distribution with shape parameter ($\alpha=4$) and heterogeneous error variance among environments are used for both cases. In case (b), the bootstrapping and permutation is done within each environment.

177x80mm (300 x 300 DPI)

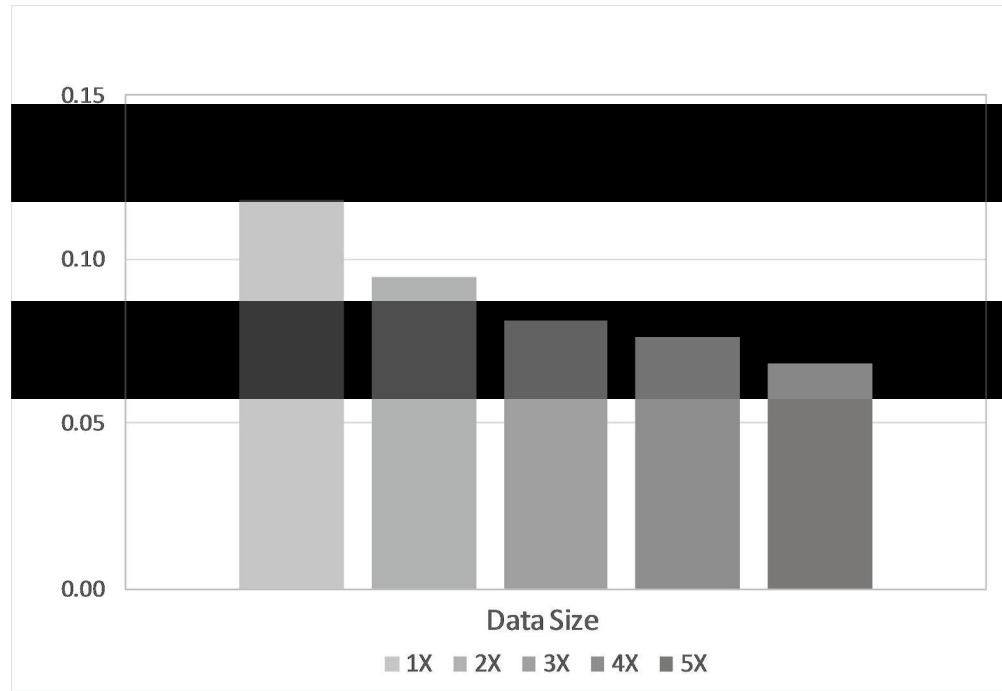


Figure 4: Observed type-I error rates by using non-parametric bootstrap methods for different sizes of data matrix with exponentially distributed errors with homogenous variance. The left most bar represents the error rate of original Soya bean data and the right most bar shows error rate of data that was increased by combining the original data matrix 5 times in both environment and genotype direction.

```
#####
##### Function for Non-parametric bootstrapping #####
#####
```

```
## Columns are genotypes
## Rows are Environments
```

```
require(reshape2)
require(agridat)
require(e1071)
```

```
test.npb <- function(Y, n.boot=1000, K = 2)
```

```
{
```

```
  I <- nrow(Y)
  J <- ncol(Y)
  M <- min(I-1, J-1)
```

```
  E <- sweep(Y, 1, rowMeans(Y))
  E <- sweep(E, 2, colMeans(Y))
  E <- E + mean(Y)
```

```
  E.svd <- svd(E)
  U <- E.svd$u[,1:M]
  V <- E.svd$v[,1:M]
  D <- diag(E.svd$d[1:M])
  lam <- E.svd$d[1:M]
```

```
  t.obs <- lam[K+1]^2/sum(lam[(K+1):M]^2)
```

```
  t.boot <- rep(NA, n.boot)
```

```
  if (K > 0){
```

```
    U.K <- U[,1:K]
    D.K <- diag((lam[1:K]), nrow=K, ncol=K)
    V.K <- V[,1:K]
    theta.K <- U.K %*% D.K %*% t(V.K)
```

```
  }else
```

```
    theta.K <- matrix(0, nrow= I, ncol=J)
```

```
  U.B <- U[, (K+1):M]
```

```
  D.B <- diag((lam[(K+1):M]), nrow=(M-K), ncol=(M-K))
```

```
  V.B <- V[, (K+1):M]
```

```
  R.B <- U.B %*% D.B %*% t(V.B)
```

```
  for(bb in 1:n.boot){
```

```
    R.b <- matrix(sample(R.B, I*J, replace=T), nrow=I, ncol=J)
    E.b <- theta.K + R.b
```

```

E.bb <- sweep(E.b, 1, rowMeans(E.b))
E.bb <- sweep(E.bb, 2, colMeans(E.b))
E.bb <- E.bb + mean(E.b)

lam.b <- svd(E.bb)$d
t.boot[bb] <- lam.b[K+1]^2/sum(lam.b[(K+1):M]^2)
}

pvalue <- colMeans(t.boot > matrix(rep(t.obs, n.boot), nrow=n.boot, byrow=TRUE))

cat("Test statistics of observed data:", t.obs, "\n")
cat("p-value of AMMI", K, "using Non-parametric bootstrap:", pvalue)

}

#####
#### L C A Corsten and J B Denis, (1990) Structuring Interaction in
#### Two-Way Tables By Clustering, Biometrics, 46, 207--215. Table 1.
#####

data(corsten.interaction)
dat <- corsten.interaction
corsten <- acast(dat, gen~loc, value.var='yield')

test.npb(corsten, n.boot=100000, K=1)

test.pc(corsten, n.boot=100000, K=0)

#####

```



```
#####
#####          Permutation Method          #####
#####
```

```
## Columns are genotypes
## Rows are Environments
```

```
require(reshape2)
require(agridat)
require(e1071)
```

```
test.ppb <- function(Y, n.boot=1000, K = 2)
```

```
{
```

```
  I <- nrow(Y)
  J <- ncol(Y)
  M <- min(I-1, J-1)
```

```
  E <- sweep(Y, 1, rowMeans(Y))
  E <- sweep(E, 2, colMeans(Y))
  E <- E + mean(Y)
```

```
  E.svd <- svd(E)
  U <- E.svd$u[,1:M]
  V <- E.svd$v[,1:M]
  D <- diag(E.svd$d[1:M])
  lam <- E.svd$d[1:M]
```

```
  t.obs <- lam[K+1]^2/sum(lam[(K+1):M]^2)
```

```
  t.boot <- rep(NA, n.boot)
```

```
  if (K > 0){
```

```
    U.K <- U[,1:K]
    D.K <- diag((lam[1:K]), nrow=K, ncol=K)
    V.K <- V[,1:K]
    theta.K <- U.K %*% D.K %*% t(V.K)
```

```
  }else
```

```
    theta.K <- matrix(0, nrow= I, ncol=J)
```

```
  U.B <- U[, (K+1):M]
```

```
  D.B <- diag((lam[(K+1):M]), nrow=(M-K), ncol=(M-K))
```

```
  V.B <- V[, (K+1):M]
```

```
  R.B <- U.B %*% D.B %*% t(V.B)
```

```
  for(bb in 1:n.boot){
```

```
    R.b <- matrix(sample(R.B, I*J, replace=F), nrow=I, ncol=J)
    E.b <- theta.K + R.b
```

```

E.bb <- sweep(E.b, 1, rowMeans(E.b))
E.bb <- sweep(E.bb, 2, colMeans(E.b))
E.bb <- E.bb + mean(E.b)

lam.b  <- svd(E.bb)$d
t.boot[bb] <- lam.b[K+1]^2/sum(lam.b[(K+1):M]^2)
}

pvalue <- colMeans(t.boot > matrix(rep(t.obs, n.boot), nrow=n.boot, byrow=TRUE))

cat("Test statistics of observed data:", t.obs, "\n")
cat("p-value of AMMI", K, "using Permutation based bootstrap:", pvalue)

}

#####
#### L C A Corsten and J B Denis, (1990) Structuring Interaction in
#### Two-Way Tables By Clustering, Biometrics, 46, 207--215. Table 1.

data(corsten.interaction)
dat <- corsten.interaction
corsten <- acast(dat, gen~loc, value.var='yield')

#####

test.ppb(corsten, n.boot=100000, K=1)

test.ppb(corsten, n.boot=100000, K=0)

```