Comparing Bayesian and Frequentist Approaches for GGE Bi-plot Analysis in Multi-Environment Trials in Sorghum

Siraj Osman Omer1* and Murari Singh2

1Experimental Design and Analysis Unit, Agricultural Research Corporation (ARC), Sudan
2International Center for Agricultural Research in the Dry Areas (ICARDA), Amman, Jordan

*Corresponding author: Siraj Osman Omer, Experimental Design and Analysis Unit Agricultural Research Corporation (ARC), P.O. Box 126, Wad Medani, Sudan, Tel: + 249908246491; E-mail: sirajstat@yahoo.com

Received date: August 17, 2017; Accepted date: November 22, 2017; Published date: November 30, 2017

Copyright: © 2017 Omer SO, Singh M. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.


Abstract

The GGE stands for genotype main effect (G) added to genotype by environment (GE) interaction. GGE-biplot a graphical tool was applied to study adaptation of 18 sorghum genotypes to the six environments using data from a multi-environment trials (MET) conducted in randomized complete block designs at two locations during 2009-2012 in Sudan. Analysis of variance was used to test the significance of GE interactions, estimates of variance components and predicted means were obtained using restricted maximum likelihood REML method. A number of priors for the variance components of the model were considered for Bayesian analysis using R2WinBUGS software. The best set of priors was selected using the deviance information criterion (DIC). Thus, the predicted estimates of GE means using REML method for frequentist approach and posterior estimates for the Bayesian approach were used for the graphical presentation of the genotypes and the environments. In frequentist approach, the first two principal components accounted for 64% of variation in total GGE interactions where the individual two principal components accounted for PC1=43% and PC2=23% respectively. The Bayesian approach accounted for 89% variation in the total GGE interaction with PC1=58% and PC2=31% respectively. The Bayesian GGE biplot analysis explained much larger proportion of variation in GGE interaction in comparison with frequentist approach, and thus resulted in a more powerful inference on the adaptation of genotypes to the environments considered.

Keywords: GGE biplot; Multi-environment trials; Bayesian approach

Introduction

Multi-environment trials (MET) are used to identify superior crop genotypes for a target environment in plant breeding programs [1,2]. Selecting the best genotypes, with stable and high yield across a number of environments presents statistical issues in the presence of significant genotype x environment interaction (GEI) due to changes in the magnitude of the genotypic response across the environments [3]. GGE biplot has been shown to be very useful for analyzing MET datasets and identifying adaptable genotypes with high yield performance in several studies [4-6]. The GGE biplot analysis exhibits the aspects of genotype stability and adaptability, when a high proportion of the sum of squares of G+GEI could be retained in first two principal components [7]. The polygon view of the biplot is the sound way to visualize the interaction patterns between genotypes and environments [8]. Yan and Kang (2003) have shown the presence or absence of crossover GE interaction which is helpful in identifying different mega environments. Recently, the extensive usefulness of GGE-biplot has been elucidated for analyzing data from multi-environment trials in wheat [9]. These aspects make GGE biplot a most comprehensive tool in plant breeding. GGE biplot is for: 1) carrying out mega-environment analysis (see example, [10-12]: 2) genotypes evaluation (the mean performance and stability) and 3) environments’ evaluation (the power to discriminate among genotypes in target environment). A two-way table of GE data may be analyzed through the joint use of analysis of variance (ANOVA) and singular value decomposition (SVD) in term of principal component analysis (PCA) [13].

In an ongoing crop improvement program, priors information is available for distribution of variance components, for example, for genotypes and genotype x environment integration. These can be used to improve the results on the predication of GE means. The commonly used frequentist approaches also not make use of such information. The analysis of multi-environment trials with a view to estimate genotypes stability can be carried out using a Bayesian approach. Edwards and [14] modeled heterogeneity using exponential of an additive model followed by assigning suitable priors to the variance components of the effects terms of the model. Crossa et al. have studied some practical and theoretical aspect of Bayesian stability in the context of additive main effects and multiplicative interaction (AMMI) and posterior means for genotype x environment interaction component using Gibbs sampler and applied to data on maize [15]. Applications of the Bayesian approach to the AMMI model and GGE have been presented by Josse et al. [16]
and da Silva et al. [17]. Using Bayesian analysis, Omer et al. conducted a similar study on genotype × environment interactions and the GGE-biplot assessment of balanced classifications with missing values [18]. Bayesian analysis of GGE biplot models and its implications for the interpretation of the biplots have been discussed in de Oliveira et al. [19]. The purpose of this study is to compare Bayesian and frequentist approaches for GGE biplot of sorghum genotypes yields in six environments in Sudan. The first section presents a frequentist approach, the commonly used GGE biplots. The second section is Bayesian GGE biplot analysis obtained from posterior estimates of genotype × environment interaction of predicted means of grain yield.

Materials and Methods

Experimental data set

Eighteen genotypes of sorghum were evaluated in randomized complete block design (RCBD) during three growing seasons, 2009/10 to 2011/12, at two different locations (North Gedarif and South Gedarif) in Sudan. Data on grain yield in kg/ha was recorded for analysis.

Statistical analysis

Analysis of variance (ANOVA) and REML methods were applied on the combined data using Genstat software [20] to obtain the frequentist estimates of variance components and predicted means while R2WinBUGS was used to obtain posterior means under Bayesian approach. GGE biplot graphs were drawn based on predicted means of under frequentist and Bayesian approach. Therefore, the Bayesian posterior means of GE two ways table based on priors for the SDCs, priors were obtained by using data on sorghum yield (kg ha⁻¹) from three similar experiments conducted to evaluate 18 genotypes in RCBDs with four replications during 2006/07-2008/09 at Rahd station in Sudan. A priors information and the WinBUGS and R codes are available [21], the number of iterations was set at one 50,000, the number of chains was set at three, and the last 5,000 simulated values of the parameters were taken for evaluating the posterior distributions.

The Bayesian approach uses prior information which was considered in terms of distributions for variance or standard deviation components for effects of blocks within environments, genotypes, environments and GEI and experimental error variances σ²ε, assumed to be homogeneous across environments. The a priori distributions of the variance components in terms of the scale parameters, in the GEI model in the application were taken from past data for half-normal distributions. The prior of half-normal distribution was used for the various standard deviation components of the data model. Using the best a priori distributions, the a posteriori expected values of predicted GE means were obtained (details are aimed for presentation in a separate publication and are not included here). Details of the deviance information criterion (DIC), a Bayesian counterpart of the Akaika information criterion (AIC) for model selection, and selection of the best priors data see [22,23].

The Bi-plots model

Yan and Kang (2003) observed phenotypic variation (P) of genotypes across environments is made up of environment variation (E), genotype variation (G) and genotype-by-environment (GE) interaction variation Yan [24,25]. This can be written as P=E+G+GE, usually E is the dominant source of variation separate for genotype, so environmental means are removed and analysis concentrates on the genotype variation and genotype-by-environment interaction [26]. The sum of these two terms can be approximated as first two principal components to obtain GGE bi-plot using Genstat software. The basic model for a GGE bi-plot is given as

\[ Y_{ij} = \mu + b_i + a_j + d_{ij} \]  \hspace{1cm} (1)

Where \( Y_{ij} \) is the estimated yield of genotype i in environment j, \( \mu \) is the grand mean of all observations, \( a_j \) is the main effect of genotype i, \( b_i \) is the main effect of environment j and \( d_{ij} \) is the interaction between genotype i and environment j. Instead of trying to separate G and GE, a GGE biplot model accounts G and GE together and expresses their joint contribution G+GE into two multiplicative terms [27]. Thus, the GGE bi-plot model can be rewritten as

\[ Y_{ij} = \mu + b_i + \gamma_1 P_{i1} + \gamma_2 P_{i2} + d_{ij} + \epsilon_{ij} \]  \hspace{1cm} (2)

where \( \gamma_1 \) and \( \gamma_2 \) are the singular values (SV) for the first and second principal component (PC1 and PC2), respectively, \( P_{i1} \) and \( P_{i2} \) are elements of eigenvectors of genotype i for PC1 and PC2, respectively, \( d_{ij} \) and \( \epsilon_{ij} \) are elements of eigen vectors of environment j for PC1 and PC2, respectively, \( \epsilon_{ij} \) is the residual associated with genotype i in environment j. PC1 and PC2 eigenvectors cannot be plotted to uniquely construct a meaningful bi-plot before the singular values are partitioned into the genotype and environment eigenvectors. Singular-value partitioning is implemented by

\[ \theta_{ij} = \gamma_1 f_i P_{i1} + \delta_{ij} = \gamma_1^{1-f_i} \delta_{ij} \]  \hspace{1cm} (3)

Where, \( f_i \) is the partition factor for PC1. Theoretically, \( f_i = 1 \) is most commonly used and is interpreted as environment focused. To generate the GGE bi-plot, the equation [1] is presented as:

\[ Y_{ij} = \mu + b_i + \theta_{1j} P_{i1} + \theta_{2j} P_{i2} + \epsilon_{ij} \]  \hspace{1cm} (4)

In a bi-plot, genotype i is displayed as a point defined by all \( \theta_{ij} \) values \((i=1,2)\) on PC1 (x-axis) and PC2 (y-axis), and environment j is displayed as a point defined by all \( \theta_{ij} \) values [28].

Bayesian approach for evaluating genotype and environment interaction

From Bayesian perspective, model in equation (1) can be rewritten as

\[ Y_{ij} \sim N(\mu + b_i + a_j + d_{ij}, \sigma^2) \]  \hspace{1cm} \text{for} \ (i=1,2)

Where, the priors of half-normal distributions, the a posteriori expected values of predicted GE means were obtained (details are aimed for presentation in a separate publication and are not included here). Details of the deviance information criterion (DIC), a Bayesian counterpart of the Akaika information criterion (AIC) for model selection, and selection of the best priors data see [22,23].
The variance components of various effects and interactions in equation (1) will be assumed to be random variables having distributions, called the a the priori distribution, with known parameters. The Bayesian methodology for evaluating genotype and environment interaction (GEI) was presented by Omer at al. on an unbalanced dataset of sorghum yields. In this context, frequentist and Bayesian GEI data model will be used for investigating GGE bi-plot analysis when the dataset is balanced for genotypes and environment classifications.

**Results and Discussion**

The estimates of variance component, Table 1 indicated that the GEI was significant (P<0.01) under both the approaches. Variation due to G was significant under Bayesian and frequentist approach. Bayesian approach, compared with the frequentist approach, gave the better differentiation of genotypes as assessed in term of variance component and standard error.

**Table 1**: Summary of the estimates of variance components from combined analysis of grain yield in the evaluation of 18 sorghum genotypes in 6 environments under frequentist and Bayesian approach.

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>Degree of freedom</th>
<th>Frequentist approach</th>
<th>Bayesian approach</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Component of variance</td>
<td>Standard error</td>
</tr>
<tr>
<td>Env(Rep)</td>
<td>1152</td>
<td>848</td>
<td>1169**</td>
</tr>
<tr>
<td>Genotypes</td>
<td>17</td>
<td>1169**</td>
<td>2095</td>
</tr>
<tr>
<td>Environment</td>
<td>5</td>
<td>149318**</td>
<td>95586</td>
</tr>
<tr>
<td>GE interaction</td>
<td>85</td>
<td>21342**</td>
<td>4249</td>
</tr>
<tr>
<td>Error</td>
<td>306</td>
<td>24659**</td>
<td>1994</td>
</tr>
</tbody>
</table>

**Predicted values of the genotype and environment interaction**

The a posteriori means of the GE predicted values is shown in Table 2 (Bayesian approach). Table 3 gives the predicted means of GE under frequentist (REML method). Genotypes are denoted as G1, G2,..., G18 and environment as E1, E2,...,E6. Table 2 shows that G10 was the best genotype in environments E1, E4 and E5 with yields 306, 596 and 151 kg/ha respectively. While G13 was the best genotype in environment E2 (yield=491.8 kg/ha), G18 was best in E3 (yield=1010 kg/ha) and G15 in E6 (yield=1636 kg/ha).

**Table 2**: Bayesian mean predicted of grain yield (kg/ha) of 18 sorghum genotypes (G1 to G18) across the sex environments E1 to E6 comprising two locations and three years.

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Environments</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1</td>
<td>E2</td>
<td>E3</td>
</tr>
<tr>
<td>G1</td>
<td>129</td>
<td>253</td>
</tr>
<tr>
<td>G2</td>
<td>175</td>
<td>327</td>
</tr>
<tr>
<td>G3</td>
<td>154</td>
<td>231</td>
</tr>
<tr>
<td>G4</td>
<td>173</td>
<td>165</td>
</tr>
<tr>
<td>G5</td>
<td>153</td>
<td>150</td>
</tr>
<tr>
<td>G6</td>
<td>155</td>
<td>194</td>
</tr>
</tbody>
</table>

**Table 3**: Frequentist mean predicted of grain yield (kg/ha) of 18 sorghum genotypes (G1 to G18) across the sex environments E1 to E6 comprising two locations and three years.

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Environments</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1</td>
<td>E2</td>
<td>E3</td>
</tr>
<tr>
<td>G1</td>
<td>129</td>
<td>253</td>
</tr>
<tr>
<td>G2</td>
<td>175</td>
<td>327</td>
</tr>
<tr>
<td>G3</td>
<td>154</td>
<td>231</td>
</tr>
<tr>
<td>G4</td>
<td>173</td>
<td>165</td>
</tr>
<tr>
<td>G5</td>
<td>153</td>
<td>150</td>
</tr>
<tr>
<td>G6</td>
<td>155</td>
<td>194</td>
</tr>
</tbody>
</table>
This article is available from: http://www.imedpub.com/european-journal-of-experimental-biology/

Table 3 shows that the G10 yielded highest in environments E1, E4 and E5 with 259, 622.1 and 356.3 kg/ha, respectively. G13 was the best genotype in environment E2 (yield=574.4 kg/ha), G18 in environment E3 (yield=951.4 kg/ha) and G15 in environment E6 (yield=1627.5 kg/ha). Genotypes G10, G13, G15 and G18 gave the highest yield under both the approaches in an environment were found best. The posterior means were estimated with a higher precision compared to the frequentist approach (Tables 2 and 3). For instance, average standard error genotype in environment E1, mean in 38.66 kg/ha under Bayesian approach and 70.03 kg/ha under frequentist approach.

GGE bi-plot analysis

The partitioning of GGE interaction under Bayesian approach (Table 2) in two principal components showed that PC1 and PC2 accounted for 67.80% and 13.84% respectively, and thus explaining a total of 81.64% variation. While for the Frequentist estimates, the corresponding values were 47.29% for (PC1) and 24.62% for (PC2) and 71.84% for total variation. The total percentage variance explained by the two component representation was more effective in Bayesian case compared to frequentist approach. Thus the predicted values of GE means under Bayesian method assemble a clearer pattern between genotypes and environment in smaller number of components, compared to frequentist approach. The bi-plots have been shown by polygon views in Figure 1 (Bayesian approach) and Figure 2 (Frequentist approach). The comparison views of biplots have been shown in Figure 3 (Bayesian approach).

It has been pointed out that PC1 of a GGE bi-plot approximates the genotype main effects (mean performance) and PC2 approximate the GEI effects associated with each genotype, which is a measure of instability. In frequentist, the genotypes with the best response to particular environments to identify specifically adapted genotypes, we grouped G14 and G15 had the highest yielding performance in environments E6, and the G17 and G18 performed well in the environments E4 and E3 whereas G2, G5, G8 and G9 were poor in E1, E2 and E5 environments. For instance, the G15 had the highest yielding performance in environments E6 and G18 well performed in the environments E3, whereas G2, G5, G8 and G9 were poor in E1, E2 and E5 environments with low yield performance.
Comparison GGE bi-plot is used to evaluate the genotypes relative to an ideal genotype. This genotype has large PC1 scores (high mean yield) and small (absolute) PC2 scores (high stability). In Bayesian approach, genotypes G13, G15 and following to G10 and G17 were more desirable than other durum genotypes Polygon view (Figure 1). A genotype is more favorable if it is closer to the ideal genotype position. Therefore, in frequentist approach, genotypes G17 and following to G10, G13, G2 and G15 were more desirable than other genotypes using Polygon view (Figure 2). Using the comparison plot, Bayesian approach highlighted that genotype G15 and following to G14, G17 and G12 were more desirable than the other genotypes. While poor genotypes were G4, G3, G1 and G16 ordinarily (Figure 3). In frequentist approach, genotypes G15 and following to G14, G17 and G12 were more desirable than other genotypes and poor genotypes were G4, G3 and G1, ordinarily (Figure 4). In this way one can compare the two approaches for GGE biplot analysis because it allows visual interpretation of GE interaction. In both approaches, it seems that GGE biplot methodology is a proper tool for identifying high yielding genotypes as the most stable ones. Bayesian GGE bi-plots can be used as a new approach for visualizing GGE bi-plot on the statistical model of principal component analysis (PCA). The difference between the two approaches in GGE bi-plot analysis is based on differences in the predicted GE means. The Bayesian approach of GGE bi-plot showed higher variability accounted for by the total PCA compared to the frequentist approach. GGE bi-plot analysis may be used for quick visual evaluation and comparison based on the two approaches. These issues are critical because they are inherently related to the validity and scope of the functionalities and capabilities claimed by proponents of GGE bi-plot analysis. The Bayesian codes can be obtained from the first author.

### Conclusion

GGE bi-plot methodology, as has been shown to be useful in the analysis of MET dataset, with a view to have a graphical assessment of relationships among genotypes and test environments for a combined response in terms of GEI and genotypic performance (GGEI). The Bayesian approach integrates the prior information available from already conducted trials with that of the current dataset and provides a more realistic and wider coverage for statistical inference. Predicted GE means were obtained using Bayesian and frequentist approaches. Bayesian GGE bi-plot analysis can be easily used to draw conclusions and make right decisions.

### Acknowledgments

The authors are grateful to Mr. Mohammed Hamz Mohammed, Cereal Research Center, Agricultural Research Corporation (ARC), Wad Medani, Sudan for providing the data used in the illustration. First author is grateful to ICARDA and Arab Fund for Economic and Social Development (AFESD) for granting a fellowship for carrying out the research study.

### References


This article is available from: http://www.imedpub.com/european-journal-of-experimental-biology/