

COMPARISON BETWEEN AMMI, W-AMMI AND GGE METHODOLOGY IN THE CONTEXT OF SIMULATED DATA

Danilo Augusto SARTI¹
Carlos Tadeu dos Santos DIAS¹

- **ABSTRACT:** Genotype x environment interaction is a key issue in plant breeding and new cultivars development. The modelling of such interactions has huge importance to decisions in plant breeding and breeding program optimization. In this context AMMI, W-AMMI and GGE models claims to address such interactions. The present paper aims to check the behaviour of such models in face of data with well behaved parametric properties. The results shows that the three models are efficient to model GxE interactions.
- **KEYWORDS:** AMMI; WAMMI; GGE; genotype x environment.

1 Introduction

One constant interest in statistics, design and analysis of experiments is the study of main effects of factors and, in many cases, of the interactions among different levels of factors.

In the context of Multi Environmental Trials one of the main issues to be evaluated is the interaction between the factors genotype and environment.

Duarte and Venkovsky (1999), Yan *et al.* (2000), Hongyu *et al.* (2014), Hongyu *et al.* (2015), Rodrigues *et al.* (2016) show the role and importance of studying genotype x environment interactions and Sarti (2013) shows its economical importance for decisions in plant breeding. By its turn, Yan and Kang (2003) describes the role of models in cultivar recommendation.

Several models can be used to study interactions among factors and to evaluate genotype over environment interactions. This is the case of Additive main effects

¹Universidade de São Paulo - USP, LCE/ESALQ, CEP 13418900, Piracicaba, SP, Brazil. E-mail: danilo.estat@gmail.com; ctsdias@gmail.com

and multiplicative interaction - AMMI models Gauch (1992) and GGE model Yan and Kang (2003), as well as their variants, such as the weighted AMMI - WAMMI Hongyu *et al.* (2015) and the robust version of AMMI Rodrigues *et al.* (2016).

Every statistical model claims to answer a particular question. To check how good the models answer such questions we can use simulated data, with well established properties, such as the ones described in data section of this paper.

The scope of this study is to compare AMMI, WAMMI and GGE models in the context of simulated dataset called **simf2** described in the methodology obtained from Sarti (2019). The comparison of such statistical models can supply different parameters and information about the same key interest question and produce insight about genotype x environment interactions.

In addition to the comparison between the models, results show that AMMI, W-AMMI and GGE can be used in situations where the data are generated by a non normal but yet continuous distribution of probability such as logistic distribution.

2 Methodology

2.1 Data

In this paper we use a dataset called **simf2** obtained via simulation by Sarti (2019).

This data set is comprised by simulated values of yield ($ton * ha^{-1}$) for 10 maize hybrids genotypes in a 4 blocks experimental design conducted across five environments. The later represent the macro regions of maize production in Brazil. The data were simulated from distributions of probability fitted for data sets of real muti environmental maize trials in Brazil. Details about the fitting of such distributions and the simulations can be found on Sarti (2019).

The structure of **simf2** is presented in Table 1

Table 1 - General structure of **simf2** data set

block	environment	yeild $ton * ha^{-1}$	genotype
1	e1	9735.15	g1
2	e1	10483.14	g1
3	e1	11198.90	g1
4	e1	12430.01	g1
1	e2	7398.40	g2
2	e2	8146.39	g2
...
3	e5	9746.52	g10
4	e5	10977.63	g10

The statistical model for this block design is:

$$y_{ijh} = \mu + g_i + e_j + (ge)_{ij} + b_{h(j)} + \epsilon_{ijh} \quad (1)$$

Where:

- y_{ijh} is the response of i-th genotype in j-th environment for block h and $i = 1 \dots g, j = 1 \dots e$ and $h = 1 \dots r$
- μ is constant, frequently the general mean
- g_i is the effect of the genotype i
- e_j is the effect of the environment j
- $(ge)_{ij}$ is the interaction factor for genotype i in the environment j
- $b_{h(j)}$ is the effect of block h inside environment j.
- ϵ_{ijh} is the experimental error associated to genotype i, inside environment j in the block h, assumed

The distributions of probability from where Sarti (2019) simulated the dataset **simf2** are Logistic, Mixture of Normal and Normal. They were fitted from real data experiments in Brazil as shown in Sarti (2019).

This made possible an important result of this paper showing that AMMI, W-AMMI and GGE modelling can be applied in situations where the response variable comes from a non normal, but is still continuous model, like the logistic distribution.

The relationship between the region from Brazil used in the simulations to obtain **simf2**, and the respective model used by Sarti (2019) to describe the region where the experiments were conducted are shown in Table 2.

Given the economical importance of Maize in central Brazil, Sarti (2019) chooses 3 environments to represent it e1, e4 and e5.

For our purposes we will consider the general structure of **simf2** shown in Table 1 and more details about how the data set was obtained can be found in Sarti (2019).

The methodology used by Sarti (2019) in the simulation mentioned before enforced the relationships between genotypes and environments as stated in the Table 3.

In this table we see specific and non specific adaptation of genotypes to environments. For more details about it we suggested the reading of Sarti (2019).

Later in this article our interest will be the ability of AMMI, W-AMMI and GGE models to capture the relationships stated in Table 3.

Table 2 - Relationship among distributions from where simf2 dataset was simulated and Brazilian region they represented, as shown in Sarti (2019)

environment	related region considered	probabilistic model / parameters
e1,e4,e5	Center	logistic scale = 1106.83 kg location = 8392.22 kg
e2	North	normal mixture mean1 = 4640.68 kg mean2 = 8425.84 kg standard deviation1=1239.72 kg standard deviation2=1239.72 kg
e3	South	normal mean = 11562.492 kg standard deviation = 1620.763 kg

Table 3 - Specific relationships between genotype x environment

genotype	specific environment
g1	e1 (center1)
g2	e2 (north)
g3	e3 (south)
g4	e4 (center 2)
g5	e5 (center 3)
g6	e1 (center 1)
g7	e4 (center 2)
g8	e3 (south)
g9	e4 (center 2)
g10	e2 (north)

2.2 Analysis of variance

As a protocol for proceeding AMMI analysis we will use the recommendations presented by Gauch (2013) who suggests the following steps (i) analysis of variance, (ii) model diagnosis, (iii) mega-environment delineation, and (iv) agricultural recommendations.

The same reasoning will be applied to weighted AMMI and an analogous process to GGE analysis.

Analysis of variance is a statistical procedure to investigate the effect of two or more levels of treatments. Each experimental design has its own analysis of variance design.

In case of Multi Environmental Trials (MET) we conduct an analysis of

variance to check if the interaction term of the model is significant. If this is the case, we proceed interaction modelling techniques for the interaction term.

To present the reasoning about ANOVA we present first a single factor design as shown in Equation 2. Later in this section we expand the model to show how is the ANOVA for the block design regarding **simf2** data set.

$$y_{ij} = \mu + \tau_i + \epsilon_{ij} \quad (2)$$

Where:

- μ is the general mean
- τ_i is the effect of i-th treatment τ
- ϵ is the expected mean error associated with the measure $\epsilon N(o, \sigma^2)$
- $SQ = \sum_{i=1}^I \sum_{j=1}^J y_{ij}^2 - C$
- $C = \frac{(\sum_{i=1}^I \sum_{j=1}^J y_{ij})^2}{IJ}$
- $SQT = \sum_{i=1}^I y_{i.}^2 - C$
- $\sum_{i=1}^I y_{i.}$ is the total of treatment i
- $SQ_{res} = SQT_{total} - SQT_{treat}$

For which we are interested in testing the following hypotheses:

$$\begin{cases} H_0 \equiv \tau_1 = \tau_2 = \dots = \tau_i \equiv \mu_1 = \mu_2 = \dots = \mu_i \\ H_1 \equiv \tau_i \neq \tau_j \equiv \mu_i \neq \mu_j \quad \text{for at least one contrast} \end{cases} \quad (3)$$

The analysis of variance in this simple case decomposes the total variance in different parts, named between treatments and experimental error (inside). In general these two variations can be measured by sum of squares for each component.

To test hypothesis H_0 the test F is used given our parametric context under the assumptions of normality.

In this case, $MSTret/MSRes$ has a distribution F on (i-j) and I(J-1) degrees of freedom under iid and normality assumptions.

In some situations, when the same experiments are repeated in some design inside several environments, a joint analysis can be made. In our case the main interest is to evaluate if the interaction between genotype and environment are statistically significant.

The basic model of ANOVA for a block experiment conducted in several environments is in Equation 4 followed by its ANOVA table.

$$y_{ijh} = \mu + g_i + e_j + (ge)_{ij} + b_{h(j)} + \epsilon_{ijh} \quad (4)$$

Table 4 - Analysis of Variance table for Equation 4

Source of Variation	Df	Sq	Mean Sq
blocks inside environments	e(r-1)	SQ_{bie}	MS_{Bie}
Genotypes	(g-1)	SQ_g	MS_g
Environments	(e-1)	SQ_e	MS_e
GxE Interaction	(g-1)(e-1)	SQ_{gxe}	MS_{gxe}
Residuals	e(g-1)(r-1)	SQ_{res}	MS_{gxe}
Total	(ger-1)	SQ_{res}	

Equation 4 describes the joint analysis of variance for detecting genotype by environment interactions.

Where:

- y_{ijh} is the response of i-th genotype in j-th environment for block h and $i = 1...g, j = 1...e$ and $h = 1...r$
- μ is constant, frequently the general mean
- g_i is the effect of the genotype i
- e_j is the effect of the environment j
- $(ge)_{ij}$ is the interaction factor for genotype i in the environment j
- $b_{h(j)}$ is the effect of block h inside environment j.
- ϵ_{ijh} is the experimental error associated to genotype i, inside environment j in the block h, assumed

In this context of joint analysis mean analysis require estimatives of mean residual variance $QM_{EM} = \frac{\sum_j SQ_{Resj}}{\sum_{j=1} GF_{Res}}$ and $j = 1, 2...e$

2.3 AMMI model

According to Dias (2005), Hongyu (2015), Hongyu *et al.* (2015), Rodrigues *et al.* (2016), Sarti (2013) and Sarti (2019) in data set containing g genotypes and e environments we have g x e interaction components to be estimated. Each effect is unfolded in its relative components.

Hongyu (2015) citing Gauch (1992) relates that the mean of a variable for a given genotype in an environment are result of additive effects (additive main effects). The interaction effects, by its turn, are a result of a multiplication of contributions from the genotypes and from the environments (multiplicative interaction effects).

Hongyu (2015) states that given a set of g genotypes tested inside e environments with r repetitions can be summarized as in Equation 5 that represents a matrix of means of genotypes inside each environment.

$$\mathbf{Y}_{gxe} = \begin{pmatrix} \bar{y}_{1,1} & \bar{y}_{1,2} & \cdots & \bar{y}_{1,e} \\ \bar{y}_{2,1} & \bar{y}_{2,2} & \cdots & \bar{y}_{2,e} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{y}_{g,1} & \bar{y}_{g,2} & \cdots & \bar{y}_{g,e} \end{pmatrix} \quad (5)$$

Hongyu (2015) states that in cases when interactions are statistically significant the SQ_{gxe} should be decomposed to avoid additional residual in some sum of squares.

Gollob (1968), Mandel (1969) suggest the decomposition of the interaction GxE term by the bilinear term shown in Equation 6.

$$(ge)_{ij} = \sum_{k=1}^n \lambda_k \gamma_{ik} \alpha_{jk} \quad (6)$$

In equation 6 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$, e γ_{ik}, α_{jk} satisfy the orthonormalization coefficient $\sum_i \gamma_{ik} \gamma'_{ik} = \sum_j \alpha_{jk} \alpha'_{jk} = 0$ para $k \neq k'$ and $\sum_i \gamma_{ik}^2 = \sum_j \alpha_{jk}^2 = 1$.

On the other hand, the statistical equation underlying AMMI models is shown in Equations 7, 8 and 9, based in Duarte and Vencovsky (1999), Dias (2005), Hongyu (2015) and Sarti (2019).

$$y_{ijr} = \mu + g_i + e_j + \sum_{k=1}^n \lambda_k \gamma_{ik} \alpha_{jk} + \rho_{ij} + b_{r(j)} + \epsilon_{ijr} \quad (7)$$

Where

- y_{ijr} is value for i -th genotype inside environment j in r -th block and $i = 1, 2, \dots, g, j = 1, 2, \dots, e$ and $r = 1, 2, \dots, h$;
- μ is a constant that corresponds to the general means
- g_i is the effect of genotype i
- e_j is the effect of the environment j
- λ_k is the square root of eigen value k from matrix $(\mathbf{GE})(\mathbf{GE})^t$ or $(\mathbf{GE})^t(\mathbf{GE})$ with $k = 1, 2, \dots, p$ and $p = \min\{(g-1)(e-1)\}$
- γ_{ik} is the i -th element of column vector γ_k associated to λ_k
- α_{jk} is the j -th element associated to the vector row α_k associated to λ_k
- ρ_{ij} is the additional residual value that contains the two multiplicative effects terms not included in the model

- $b_{r(j)}$ is the effect of block r inside environment j
- ϵ_{ijk} is the experimental error associated to genotype i environment j and block r assumed iid normally distributed with $(\mu, \sigma^2/r)$ and r the number of blocks

$$Y_{ijk} = \overbrace{\mu + g_i + e_j}^{\text{additive}} + \overbrace{\sum_{k=1}^n \lambda_k \gamma_{ik} \alpha_{jk}}^{\text{multiplicative}} + \epsilon_{ijk} \quad (8)$$

$$Y_{ijk} = \overbrace{\mu + g_i + e_j}^{\text{additive}} + \overbrace{\sum_{k=1}^n \lambda_k \gamma_{ik} \alpha_{jk} \sum_{k=n+1}^p \lambda_k \gamma_{ik} \alpha_{jk}}^{\text{multiplicative}} + \epsilon_{ijk} \quad (9)$$

Hongyu (2015) citing Gauch (1992), Dias (2005) and Sarti (2019) state that the term $(ge)_{ij}$ is represent by a summation of p terms, each of the λ_k expressed in terms of the same unit of Y_{ij} by an effect of genotype (γ_{ik}) and an environmental factor (α_{jk}), both of them being adimensional, or $\sum_{k=1}^n \lambda_k \gamma_{ik} \alpha_{jk}$. The term λ_k describes the information relative to the interaction in the k-plot, therefore, the summation of p plots comprises $(SQGEI) = \sum_{k=1}^n \lambda_k^2$ and the effects γ_{ik} and α_{jk} represent the weights of genotypes i and environment j in this specific part of the interaction.

Hongyu (2015) citing Gauch (2013) states that AMMI models do not try to recover all $SQGEI$ but just the part of it most determined by genotype and environment (rows and columns of matrix GE).

Hongyu (2015) quoting Duarte e Venkovsky (1999) states that AMMI models separate the pattern $\sum_{k=1}^n \lambda_k \gamma_{ik} \alpha_{jk}$ from the noise $\sum_{k=n+1}^p \lambda_k \gamma_{ik} \alpha_{jk}$.

Hongyu (2015) states that each element from the **GE** matrix can be given by the relation stated in Equation 10. For details about the single value decomposition of matrix GE we recommend Hongyu (2015) and Sarti (2019).

$$(\hat{ge})_{ij} = Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..} \quad (10)$$

Hongyu (2015) citing Gauch (2013), Duarte and Venkovsky (1999) and Dias (2005) state that AMMI models explain the sum of squares of the interaction by an approximation of rank n to the matrix GE, in general with n much less than p, this aims to explain the GEI by a reduced number of axes, resulting in an informative model without losing too many degrees of freedom.

According to Gauch (2013), Duarte and Venkovsky (1999) and Dias (2005), Hongyu *et al.* (2015) and Hongyu (2015) show that procedures can be used to determine the number of axes to be retained in order to explain the patterns related to the interaction. The authors state that one method is to quantify the number of degrees of freedom related to SQGEI related to each member o AMMI family.

Thus we can get the mean square related to each plot. After that an F-test can be performed to evaluate the significance of each component relate to the mean square error.

Hongyu (2015) states that a stop rule to select which AMMI model to be used ($AMMI_0, AMMI_1, \dots, AMMI_n$) is based in a F test for the axes of interaction.

Hongyu (2015) quoting Cornelius (1993) states the statistic described in Equation 11 which is considered a robust method. Under the hypothesis that there are no more than n multiplicative terms for the interaction it follows a F distribution with parameters $(g - 1 - n)(e - 1 - n)$ degrees of freedom for residuals. In this case Hongyu (2015) states that this can be a good test to get the n+1 first components of the interaction.

$$F_r = \frac{SQ_{GEI} - \sum_{k=1}^n \lambda_k^2}{f_2 QM_{Res}} \quad (11)$$

where

- λ_k is the k-th eigen value of the matrix $(GE)(GE)^t$ and QM_{res} is the mean square of the residuals

2.4 Weighted AMMI model

Dias (2005), Hongyu (2015) and Sarti (2019) state that in cases when a double entry table of data containing genotypes and environments has heterogenous variances among environments the values of table must have different weights for its related residual squares in estimation processes. The same author quoting Robertson (2014) states that when the variance and covariance structure for the error differ significative among the environments this should be considered in AMMI analysis.

In this paper we use the generalization proposed in Hongyu *et al.* (2014) , Hongyu *et al.* (2015) and Hongyu (2015) named as weighted AMMI and stated in Equation 12.

$$y_{ijr}/s_j^2 = \mu + g_i + e_j + \sum_{k=1}^n \lambda_k \gamma_{ik} \alpha_{jk} + \rho_{ij} + b_{r(j)} + \epsilon_{ijr} \quad (12)$$

where

- y_{ijr} = is value for i-th genotype inside environment j in r-th block and $i = 1, 2, \dots, g$, $j = 1, 2, \dots, e$ and $r = 1, 2, \dots, h$;
- s_j^2 is the variance of experimental error of environment j or means square residual for the related environment
- μ is a constant that corresponds to the general means

- g_i is the effect of genotype i
- e_j is the effect of the environment j
- λ_k is the square root of eigen value k from matrix $(\mathbf{GE})(\mathbf{GE})^t$ or $(\mathbf{GE})^t(\mathbf{GE})$ with $k = 1, 2, \dots, p$ and $p = \min\{(g - 1)(e - 1)\}$
- γ_{ik} is the i-th element of column vector γ_k associated to λ_k
- α_{jk} is the j-th element associated to the vector row α_k associated to λ_k
- ρ_{ij} is the additional residual value that contains the two multiplicative effects terms not included in the model
- $b_{r(j)}$ is the effect of block r inside environment j ϵ_{ijk} is the experimental error associated to genotype i environment j and block r assumed iid normally distributed with $(\mu, \sigma^2/r)$ and r the number of blocks

2.5 Genotype main effects + genotype environment interaction - GGE methods

Sites regression models

Sites regression is a fundamental concept originated by the kind of multi environmental trials we address in this paper. The reasoning behind this methods is to explore the GEI interaction via the fundamental relationship expressed in Equation 13 and its fundamental consequence: $P - E = G + GE$ Hongyu (2015) quoting Yan and Kang (2003).

$$p = G + E + GE \quad (13)$$

Hongyu (2015) and Hongyu *et al.* (2014) , Hongyu *et al.* (2015) argues that Sites regression methods are based in biplot methods. In this sense, Hongyu (2015) argues that the methods of sites regression is similar to AMMI modelling, the difference being that in Sites regression the principal effects of genotypes are considered together with GEI effect meanwhile in AMMI models they are considered additive effects.

Sites regression can be mathematically posed as in Equation 14

$$\bar{Y}_{ij} = \mu + G_i + E_j + (GE)_{ij} \quad \bar{Y}_{ij} - \mu - E_j = G_i + (GE)_{ij} \quad (14)$$

Where:

- \bar{Y}_{ij} is the general mean of i-th genotype inside the j-th environment
- μ is the general mean

- G_i is the effect of i=th genotype
- E_j is the effect of environment j
- GE_{ij} is the effect of the genotype x environment interaction

The same Equation can be stated as:

$$\bar{Y}_{ij} - \mu - E_j = g_{i1}e_{1j} + g_{i2}e_{2j} + \epsilon_{ij} \quad (15)$$

These relationships can be rewritten and be used to describe a single value decomposition as shown in Hongyu (2015) quoting Yan and Kang (2003) and Sarti (2019).

$$\bar{Y}_{ij} - \mu - E_j = \lambda_1\mu_{i1}v_{j1} + \lambda_2\mu_{i2}v_{j2} + \epsilon_{ij} \quad (16)$$

Where:

- λ_1 λ_2 are the greatest eigenvalues of first and second principal components
- μ_{i1} and μ_{i2} are the eigen vectors of genotype i for Principal components 1 and 2 and v_{j1} and v_{j2} the eigen vectors of environment j to Principal Components 1 and 2.

GGE Biplot

Hongyu (2015) quoting Yan and Kang (2003), Yan and Thinker (2006), Yan (2010) and Yan (2011) states that the GGE biplot model is based in a graphical representation of a matrix of data. These biplots models are built in the first and second principal components that use Site Regression modelling. Same authors stat that in conditions when the first principal components is highly correlated to main effects of genotypes it can be viewed as a proportion of yield given by genotype effect. Hongyu (2015) and Hongyu *et al.* (2014) , Hongyu *et al.* (2015) argue that the second principal component represent the GEI part of the variable. Hongyu (2015) and Hongyu *et al.* (2014) , Hongyu *et al.* (2015) state also that three main aspects can be obtained via gge biplot Yan and Kang (2003):

- Grouping of environments with similar behaviours showing the genotypes with greater potential and its identification inside each subgroup of environments
- Inter relationship between environments indicating the best environment to test the cultivars e indicating which environment is less favorable
- Inter relationship among genotypes making easier their comparisons and ranking

Hongyu (2015) states quoting Yan and Kang (2003) and Yan et.al. (2000) that when different genotypes are adapted to different groups of environments and the variation between groups is greater than inside groups we have the formation of a mega environment. Hongyu *et al.* (2014) , Hongyu *et al.* (2015) states quoting Trethowan, *et al.* (2001) also that this definition of mega environments can be viewed as an important decision framework for plant breeders.

Hongyu (2015) states that in mega environments studies the mean of the graphic is not related to the general mean but to the environment mean instead. Hongyu (2015), Hongyu *et al.* (2014) , Hongyu *et al.* (2015) and Yan and Kang (2003) state that the cosine of an angle between two environments can be seen as the correlation among them.

Hongyu (2015), Hongyu *et al.* (2014) , Hongyu *et al.* (2015) quoting Yan (2001), Yan (2002) , Yan and Kang (2003) state that GGE biplot is built by means of the simple dispersion of g_{i1} and g_{i2} for genotypes and e_{1j} and e_{2j} for environments. Hongyu (2015) state yet that the Equation 15is environment centered given origin to one main factors to be considered in GGE analysis: Evaluations should consider only one variable by MET. example given: grains weight

To deal with this issue Yan and Kang (2003) proposed the correction presented in Equation 17

$$\frac{Y_{ij} - \bar{\mu} - E_j}{d_j} = g_{i1}e_{1j} + g_{i2}e_{2j}\epsilon_{ij} \quad (17)$$

Where

- $d_j = \sqrt{\frac{1}{g-1} \sum_{i=1}^g (Y_{ij} - \bar{Y}_{.j})^2}$ is considered the standard phenotypic deviation for environment j
- \bar{Y}_{ij} is the response of genotype i and environment j
- $\bar{Y}_{.j}$ is the mean of environment j
- g is the number of tested genotypes

Hongyu (2015) states that for MET with just one characteristic the last correction assume that environments have the same importance, thus, do not allow to detect differences among environments.

A second transformation can be obtained via

$$\bar{Y}_{ij}/s_j - \mu - E_j = g_{i1}e_{1j} + g_{i2}e_{2j} + \epsilon_{ij} \quad (18)$$

Where:

- $s_j = \sqrt{\frac{1}{m(r-1)} \sum_{i=1}^m \sum_{k=1}^y (Y_{ijk} - \bar{Y}_{.ij})^2}$

- \bar{Y}_{ij} is the mean of genotype i and environment j in repetition k inside environment j .

Single values decomposition in GGE biplots

Hongyu (2015), Yan and Kang (2003) and Yan *et al.* (2001) state that single value decomposition is very important in biplot analysis returning the principal components providing measures of the variability captured by each of these component.

$$P_{ge} = U_{g,r} S_{r,r} V_{e,t}^T [r \leq \min(g, e)] \quad P_{ij} = \sum_{k=1}^r u_{ik} \lambda_k v_{kj} \quad (19)$$

Where

- P_{ge} is a interactions matrix (G+GEi)
- $U_{g,r}$ is a matrix containing g rows and r columns which characterizes genotypes
- $V_{e,t}$ has r rows and e columns describes environments
- $S_{r,r}$ is the diagonal matrix containing r single values.
- Single value decomposition decomposes the matrix in r principal components each of them containing genotype vector u_i , vector environment v_j and single values λ decomposed in P_{ij} expression
- $\lambda_k \geq \lambda_{k+1}$,
- r is the number of principal components to represent P $r \leq \min(g, e)$. When $r < g$ there are association between genotypes and if $g < r$ we have association between environments.

Hongyu (2015) quoting Yan and Kang (2003), states that principal components cannot be directly computed before making the partition of of single values in genotypes and environments. This partition is described in Equation 19.

For a biplot (dimension=2) Hongyu (2015) state the Equation 20.

$$P_{ij} = (\lambda_1^s u_{i1}) \lambda_1^{1-s} v_{j1} + (\lambda_2^s u_{i2}) \lambda_2^{1-s} v_{j2} + \epsilon_{ij} \quad (20)$$

Where

- $s=[0;0.5;1]$ is the factor of partition of the single values
- $\lambda_1^s u_{i1}$ is used as abscissa and $\lambda_2^s u_{i2}$ for genotypes
- $\lambda_1^{1-s} v_{j1}$ as abscissa $\lambda_2^{1-s} v_{j2}$ as ordinate for environments

Hongyu (2015) quoting Yan and Kang (2003) informs that the factor s in previous equations is used to resize scores of rows and columns for a better visual interpretation of the biplot. In this sense if $s=1$ the single value partition is centered in genotypes, if $s=0$ the allocation is made completely to environment scores if the value is $s=0.5$ squared roots of λ_k will be allocated for both genotype and environment scores.

Hongyu (2015) states that partition centered in genotypes or in environments allow evaluation of genotypes and environments of test.

Hongyu (2015) , Yan and Kang (2003) and Dias (2005) state that the approximation of any element of matrix P of rank 2 can be estimated visually by inner product of genotype and environment and the cossin of the angle (inner product property of biplot).

The interpretations of the graphic respects the following rules Hongyu (2015), Hongyu *et al.* (2015) and Yan and Thinker (2006).

- if the angle of a genotype and a vector of environment is $< 90^\circ$ the genotype is greater than mean
- if the angle is $> 90^\circ$ then is less than average
- if the angle is close to 90° then the genotype is closer to the mean

Hongyu (2015) and Yan and Kang (2003) declare that the arrow that passes through the origin of biplot is the mean average environment (MAE). The arrow points to a greater average performance of genotypes. The line that is perpendicular to MAE indicates less stability over environments. In this graph PVS=1 and the single values are totally directed to genotypic scores Hongyu (2015) quoting Yan (2011).

Hongyu (2015) and Hongyu *et al.* (2015) quoting Yan (2011) state that when data are centered in environments the length of the vector of an environment is proportional to the standard deviation of genotypes means, this value being equal to phenotypic variance σ_p in a test environment and this can be used as a value of discrimination of the environment.

Hongyu (2015) and Hongyu *et al.* (2015) quoting Yan (2011) also argues that environments with shorter vectors are less discriminative, in other words, all genotypes have a tendency of performing similarly and thus little or no information about genotypic differences can be relevant in this environment, therefore, this environment should not be used as tester.

2.6 Information criteria

Hongyu (2015), Hongyu *et al.* (2015) quoting Yan and Thinker (2006) state an information criteria to evaluate the accuracy of a biplot in describing a two entry data table. Given a double entry data table containing g genotypes and e environments, the number of principal components to represent such table is $k =$

$\min(e, g-1)$. Same authors state that if there is no correlation among environments all the k PCS should be completely independent and the proportion of total variance explained by each principal component is $1/k$ Hongyu (2015).

When some correlation among environments is present the proportion of variance explained by the first principal components should be greater than $1/k$, and the variance explained by other pcs should be less than $1/k$ Hongyu (2015), Hongyu *et al.* (2015) and Yan and Thinker (2006).

This information criteria can be measure as in Equation 21:

$$\frac{\text{varibility captured by } PC_i}{\text{total variability}} \times k \quad (21)$$

Where

- $k = \min(e, g - 1)$

If the information criteria is $IR > 1$ we have some association among environments, if it is approximately 1 there are no patterns and when IR is less than 1 no information is present.

2.7 Computational Implementation

The analysis which results are presented later were made using R programming language (R CORE TEAM, 2019).

Ammi analysis were conducted by means of the R package *agricolae* Mendiburu (2015), specifically using the function AMMI. The help session of the package about AMMI analysis is a good source for users interested in details about AMMI models. Weighted AMMI was implement by usage of a function programmed by the author based on the one present in *agricolae* .

GGE Biplot methodology was implemented by means of GGEbiplots package for R Dumble. All the graphics used in GGE biplots result session were obtained with functions from this package.

3 Results

Here are presented the results according to the protocol mentioned in the methodology section.

3.1 AMMI analysis of simulated dataset

Analysis of variance

The ANOVA table for dataset simf2, containing 10 genotypes simulated in 5 environments is shown in Table 5.

From it we can deduce that environments, genotypes and repetition inside environments and the interaction were all statistically significant justifying the

posterior analysis of interaction. The coefficient of variation for the experiment was calculated as 4,74%.

The table of means of genotypes across environments is presented in Table 7 and in Table 9. Data show that the sum of squares of the interaction is 246221662.25 which comprises 21.8% if the total data variability.

We can check that g1 was the highest mean production genotype followed by g2, g4, g5 and g6. Regarding environments e3 and e5 are the ones with greater mean of yield.

Another way of seeing the GEI interaction is in Figure 1.

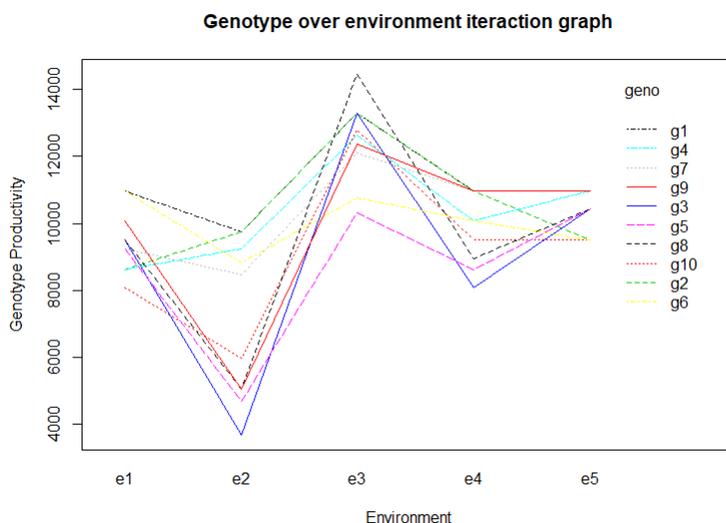


Figure 1 - Interaction Plot for simf2 showing the productivity of genotypes $ton * ha^{-1}$.

Table 5 - Anova for Simulated data set simf2

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
ENV	4	625528118.89	156382029.72	15.40	3.441e-05 ***
REP(ENV)	15	152357920.20	10157194.68	45.57	< 2.2e - 16 ***
GEN	9	74161716.42	8240190.71	36.97	< 2.2e - 16 ***
ENV:GEN	36	246221662.25	6839490.62	30.69	< 2.2e - 16 ***
Residuals	135	30089061.71	222881.94		

Table 6 - Means of AMMI model

ENV	GEN	Prod (ton/ha)
e1	g1	10961.80
e1	g10	8088.94
e1	g2	8625.06
e1	g3	9509.42
e1	g4	8625.06
e1	g5	9230.74
e1	g6	10961.80
e1	g7	9230.74
e1	g8	9509.42
e1	g9	10072.42
e2	g1	9751.12
e2	g10	5965.96
e2	g2	9751.12
e2	g3	3681.58
e2	g4	9234.70
e2	g5	4687.14
e2	g6	8825.58
e2	g7	8472.29
e2	g8	5040.42
e2	g9	5040.42
e3	g1	13295.11
e3	g10	12784.19
e3	g2	13295.11
e3	g3	13295.11

Table 7 - Means of AMMI models part2

ENV	GEN	Prod (ton/ha)
e3	g4	12619.96
e3	g5	10308.60
e3	g6	10759.10
e3	g7	12085.10
e3	g8	14456.07
e3	g9	12352.35
e4	g1	10961.80
e4	g10	9509.42
e4	g2	9509.42
e4	g3	10434.40
e4	g4	10961.80
e4	g5	10434.40
e4	g6	9509.42
e4	g7	10961.80
e4	g8	10434.40
e4	g9	10961.80
e5	g1	10961.80
e5	g10	9509.42
e5	g2	9509.42
e5	g3	10434.40
e5	g4	10961.80
e5	g5	10434.40
e5	g6	9509.42
e5	g7	10961.80
e5	g8	10434.40
e5	g9	10961.80

Table 8 - Coordinates of Simf2 AMMI biplot

	type	Prod	PC1	PC2	PC3	PC4
g1	GEN	11186.33	21.13	-2.96	11.63	-0.00
g10	GEN	9171.59	-5.29	19.32	-1.20	0.00
g2	GEN	10138.03	33.72	25.90	5.36	0.00
g3	GEN	9470.98	-40.07	5.62	7.24	-0.00
g4	GEN	10480.66	21.75	8.05	-25.83	0.00
g5	GEN	9019.06	-18.41	-30.47	-14.96	0.00
g6	GEN	9913.07	28.09	-29.57	26.12	0.00
g7	GEN	10342.35	14.13	-5.31	-19.17	-0.00
g8	GEN	9974.94	-29.04	22.69	12.35	0.00
g9	GEN	9877.76	-26.00	-13.27	-1.53	0.00
e1	ENV	9481.54	-11.26	-32.38	33.84	0.00
e2	ENV	7045.03	71.95	4.37	-2.56	0.00
e3	ENV	12525.07	-22.13	48.85	13.18	0.00
e4	ENV	10367.87	-19.28	-10.42	-22.23	0.00
e5	ENV	10367.87	-19.28	-10.42	-22.23	0.00

Table 9 - Genotype x Environment means of genotypes from data simf2

genotype	e1	e2	e3	e4	e5
g1	10961.80	9751.12	13295.11	10961.80	10961.80
g2	8625.06	9751.12	13295.11	9509.42	8625.06
g3	9509.42	3681.58	13295.11	10434.40	9509.42
g4	8625.06	9234.70	12619.96	10961.80	8625.06
g5	9230.74	4687.14	10308.60	10434.40	9230.74
g6	10961.80	8825.58	10759.10	9509.42	10961.80
g7	9230.74	8472.29	12085.10	10961.80	9230.74
g8	9509.42	5040.42	14456.07	10434.40	9509.42
g9	10072.42	5040.42	12352.35	10961.80	10072.42
g10	8088.94	5965.96	12784.19	9509.42	8088.94

Considering the matrix stated in Equation 5 we proceed the interaction analysis by the related decomposition in single values $GE = USV^t$. The genotypes means are shown in Table 9. This matrix will have rank $p = \min(9, 4)$ so the interaction can be decomposed in 4 principal components.

The decomposition of such matrices is presented in Matrices S, U and V.

U matrix contains the singular vector to the left of the interaction, S is a diagonal matrix containing singular values and V a matrix which columns owns singular values to the right of the interaction.

$$S = \begin{pmatrix} 6355.64 & 0 & 0 & 0 & 0 \\ 0 & 3625.09 & 0 & 0 & 0 \\ 0 & 0 & 2147.54 & 0 & 0 \\ 0 & 0 & 0 & 1646.21 & 0 \\ 0 & 0 & 0 & 0 & 0.00 \end{pmatrix}$$

$$U = \begin{pmatrix} 0.24 & -0.04 & 0.33 & 0.02 \\ -0.07 & 0.30 & -0.17 & -0.22 \\ 0.39 & 0.42 & -0.02 & -0.33 \\ -0.52 & 0.08 & 0.29 & 0.19 \\ 0.28 & 0.21 & -0.19 & 0.60 \\ -0.18 & -0.48 & -0.19 & 0.36 \\ 0.33 & -0.51 & 0.50 & -0.22 \\ 0.21 & -0.06 & -0.40 & 0.11 \\ -0.40 & 0.35 & 0.31 & -0.02 \\ -0.28 & -0.27 & -0.46 & -0.51 \end{pmatrix}$$

$$V = \begin{pmatrix} -0.23 & -0.60 & 0.59 & -0.19 \\ 0.83 & 0.14 & 0.20 & 0.24 \\ -0.40 & 0.76 & 0.25 & -0.11 \\ 0.11 & -0.10 & -0.61 & -0.64 \\ -0.31 & -0.19 & -0.43 & 0.70 \end{pmatrix}$$

Table 10 - Proportion of variance measured by the principal components

	percent	acum	Df	Sum.Sq	Mean.Sq	F.value
PC1	66.40	66.40	12.00	161576760.49	13464730.04	60.41
PC2	21.60	88.00	10.00	52565081.27	5256508.13	23.58
PC3	7.60	95.50	8.00	18447774.50	2305971.81	10.35
PC4	4.50	100.00	6.00	10840092.47	1806682.08	8.11

The decomposition of sum of squares of interaction in principal components is shown in Table 10. The sum of squares of each component comprises the total sum of squares of interaction. In this case all the principal components are significant. In same table we can see that the percentage of variation captured by each principal component are 66,40; 21,60;7,60 and 4,50 respectively. We can check that PC1 and PC2 summarizes 88% of total variability of the interaction.

In next stage we proceed the graphical analysis of AMMI modelling, the coordinates of such graphics are in table 8.

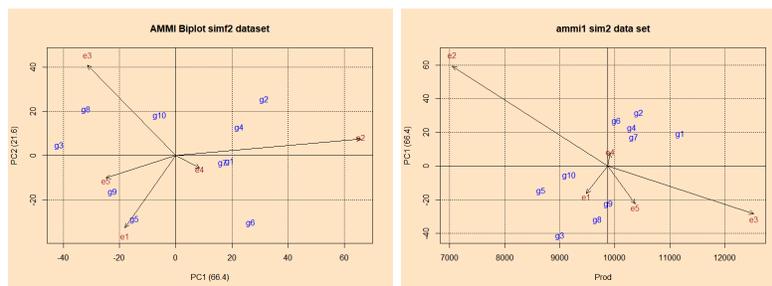


Figure 2 - AMMI biplots PC1 x PC2 and PC1 x yeild for simf2 datasets.

The main graphical representations of AMMI modelling of dataset simf 2 are shown in Figure 2 that shows on the left the representation of PC1 x PC2 biplot and on the right we have the picture that contrasts PC1 component and yield of genotypes.

Via such figures it is possible to analyze the variability dispersion of genotypes and several important characteristics about environments.

The variability of main effects of genotype and environments is displaced in horizontal axis of Figure 2b PC1xPC2 the multiplicative ones related to GEI is represented in vertical axis.

The interpretation of PC1xPC2 model form AMMI allows evaluation of stability of genotypes over environments by evaluation of magnitude of genotypes and environments scores in the interaction axis. Hongyu (2015), Hongyu *et al.* (2015) quoting Gauch (2013) state that lower scores imply in more stable accros all environments genotypes. Visually these genotypes will be closer to the origin of x, y coordinates.

Genotypes closer from the environment vector are adapted in such environments.

In the case of dataset simf2 the genotypes more close to the center are g10, g1 and g4 therefore are the most stable across environments from the set considered.

By the other hand the graphic shown in Figure 2 b allow to contrast PC1 and yield of grains. By this graph we can see that g1, g2, g4, g6 and g7 are higher than mean genotypes. They should be recommended to be produced but considering their stability to specific environments. In case of genotype g10 in despite of being lower than average it has great stability over environments and its improvement should be strongly recommended via breeding methods, ex given: usage of sister lines of its parental inbred lines. Sister Lines are inbred lines extremely close to each other from genetic pedigree point of view. Such sister liner can be crossed between themselves and then the resulting line crossed with the line which will then form the hybrid genotype of maize.

The environment which behaviour was the better considering these genotypes was the south.

Table 3 shows the specific adaptation of each genotype to environments.

We also note given the directions of environment vectors that e2 and e3 are divergent (exactly as it should be once one represents north and south regions from Brazil as stated in the data description).

All these properties shown by AMMI analysis reflected the properties induced in simulations of dataset simf2 described in the methodology section, showing that AMMI modelling is indeed efficient in capturing such aspects of genotype x environment interaction.

Weighted AMMI analysis of simulated dataset

The analysis of Variance for weighted AMMI analysis is the same as ANOVA for AMMI and is shown in Table 11. The GEI interactions is present and significant.

The use of weighted AMMI is motivated by situations when we face heterogeneous environments, this heterogeneity is expressed by the heterocedasticity among environments. In this situation environment error variance is given by the means squares of residuals.

The means and genotypes for weighted AMMI are in Tables 13 and 14. The values are in a different scale from the original data because of the weighting produced.

Table 11 - Weighted AMMI ANOVA

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
ENV	4	617225742.25	154306435.56	15.19	3.441e-05***
REP(ENV)	15	152357920.23	10157194.68	45.57	< 2.2e - 16 ***
GEN	9	104580085.15	11620009.46	52.14	< 2.2e - 16 ***
ENV:GEN	36	243429708.74	6761936.35	30.34	< 2.2e - 16 ***
Residuals	135	30089061.71	222881.94		

Table 12 - Means of environments and genotypes Weighted AMMI

	ENV	GEN	WEIGHTED MEAN (TON HA)
1	e1	g1	0.011941
2	e1	g10	0.00881
3	e1	g2	0.00939
4	e1	g3	0.01035
5	e1	g4	0.0093
6	e1	g5	0.01005
7	e1	g6	0.01194
8	e1	g7	0.01005
9	e1	g8	0.01035
10	e1	g9	0.01097
11	e2	g1	0.00173
12	e2	g10	0.00106
13	e2	g2	0.00173
14	e2	g3	0.00065
15	e2	g4	0.00164
16	e2	g5	0.00083
17	e2	g6	0.00157
18	e2	g7	0.00150
19	e2	g8	0.00089
20	e2	g9	0.00089
21	e3	g1	0.00865
22	e3	g10	0.00831
23	e3	g2	0.00865
24	e3	g3	0.0086
25	e3	g4	0.0082
26	e3	g5	0.0067
27	e3	g6	0.0070
28	e3	g7	0.0078
29	e3	g8	0.0094
30	e3	g9	0.0080

Table 13 - Means of environments and genotypes Weighted AMMI Part 2

	ENV	GEN	WEIGHTED MEAN (TON HA)
31	e4	g1	0.0271
32	e4	g10	0.02354
33	e4	g2	0.023
34	e4	g3	0.0258
35	e4	g4	0.02714
36	e4	g5	0.02583
37	e4	g6	0.02354
38	e4	g7	0.02714
39	e4	g8	0.02583
40	e4	g9	0.02714
41	e5	g1	0.02714
42	e5	g10	0.02354
43	e5	g2	0.02
44	e5	g3	0.02583
45	e5	g4	0.0271
46	e5	g5	0.0258
47	e5	g6	0.02354
48	e5	g7	0.02714
49	e5	g8	0.02583
50	e5	g9	0.02

Table 14 - Table of genotypes means weithed AMMI

	type	Y	PC1	PC2	PC3	PC4
g1	GEN	0.01	0.01	-0.01	0.00	-0.01
g10	GEN	0.01	-0.02	0.02	-0.00	0.00
g2	GEN	0.01	-0.03	0.02	-0.01	-0.00
g3	GEN	0.01	0.01	0.00	0.03	0.00
g4	GEN	0.01	0.02	0.02	-0.01	0.01
g5	GEN	0.01	0.01	-0.01	-0.00	0.02
g6	GEN	0.01	-0.03	-0.04	0.00	0.00
g7	GEN	0.01	0.02	0.00	-0.02	-0.00
g8	GEN	0.01	0.00	0.01	0.03	-0.01
g9	GEN	0.01	0.02	-0.01	-0.01	-0.02
e1	ENV	0.01	-0.01	-0.04	0.02	-0.00
e2	ENV	0.00	-0.02	0.00	-0.01	0.03
e3	ENV	0.01	-0.01	0.03	0.03	-0.01
e4	ENV	0.01	-0.02	0.00	-0.04	-0.02
e5	ENV	0.03	0.06	0.00	-0.00	0.00

The single value decomposition for GE weighted AMMI is shown in the following matrices S, U and V.

$$S = \begin{pmatrix} 0.00413 & 0 & 0 & 0 & 0 \\ 0 & 0.00304 & 0 & 0 & 0 \\ 0 & 0 & 0.00247 & 0 & 0 \\ 0 & 0 & 0 & 0.00112 & 0 \\ 0 & 0 & 0 & 0 & 0.0000 \end{pmatrix}$$

$$U = \begin{pmatrix} 0.15 & -0.26 & 0.04 & -0.27 \\ -0.38 & 0.38 & -0.02 & 0.06 \\ -0.54 & 0.32 & -0.25 & -0.12 \\ 0.16 & 0.08 & 0.59 & 0.11 \\ 0.30 & 0.28 & -0.26 & 0.40 \\ 0.21 & -0.24 & -0.03 & 0.54 \\ -0.48 & -0.67 & 0.04 & 0.15 \\ 0.26 & 0.05 & -0.44 & -0.02 \\ 0.06 & 0.24 & 0.53 & -0.20 \\ 0.27 & -0.17 & -0.21 & -0.62 \end{pmatrix}$$

$$V = \begin{pmatrix} -0.13 & -0.77 & 0.41 & -0.14 \\ -0.29 & 0.08 & -0.16 & 0.83 \\ -0.19 & 0.63 & 0.55 & -0.26 \\ -0.28 & 0.01 & -0.71 & -0.47 \\ 0.89 & 0.05 & -0.10 & 0.05 \end{pmatrix}$$

The w-AMMI biplots are shown in Figure 3.

The understanding about the specific adaptation of each genotypes to the different environments is the same as the conclusions made by AMMI models. The main results from W-AMMI for simf2 dataset are pretty much the same got by AMMI analysis. Thus w-AMMI captured all the properties of the simulated dataset being therefore a good recommendation of model to study GEI.

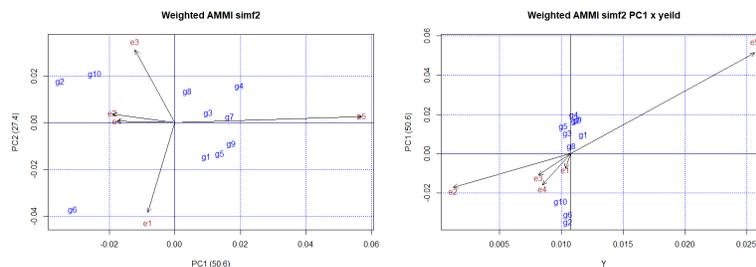


Figure 3 - AMMI biplots PC1 x PC2 and PC1 x yield for simf2 datasets. The graph on the left describes the variability captured by first and second principal components and the the picture on the left describes the variability got by PC1 versus the yield of each genotype.

GGE analysis of simulated dataset

In this section we will cover the four objectives of GGE methodology: investigation of mega environments, genotype evaluation, test environments and comprehension of GEI.

GGE Biplot:

The GGE biplot for simulated data set is present in Figure 4 that can be used to:

- genotype classification in any environment
- classify environments regarding any given genotype performance

By example, g1 and g6 have obtuse angles considering environment e2 and e4 therefore they had less than average productivity in environments.

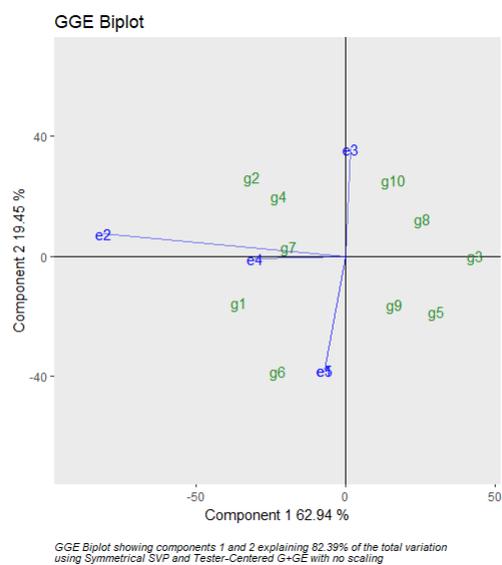


Figure 4 - GGE Biplot for simf2. In the x-axis, scores for the first principal component are shown, and in the y-axis, scores for PC2 are shown. The graphic explains 81.58% of the total variability. Genotypes are labeled g1 to g10, and environments are labeled e1 to e5.

Mega environment analysis

Hongyu (2015) and Hongyu *et al.* (2015) state that a mega environment is a group of environments or sub regions where a set of genotypes have similar behaviour and adaptability. The biplot "Which won where" is an important tool to make this analysis.

This graph for simf2 data is shown in Figure 5. As we can see there is an irregular polygon and lines that arise from the origin of the biplot, intercepting each side of the polygons perpendicularly. The vertices of such polygons are the genotypes markers which the distant is the greatest from origin. Thus, all the genotypes are inside the polygon Hongyu (2015), Hongyu *et al.* (2015) quoting Yan (2011).

According to Yan (2000) a line that crosses perpendicularly one side of the polygon represent an hypothetical environment and the two genotypes that define this side of polygon represent the same productivity. The lines that irradiate from origin divide the biplot in sectors and for each sector there is a genotype at the vertex, and such genotype showed the best behaviour for environments in this sector Hongyu (2015), Yan (2000), Yan and Kang (2003) and Yan (2011).

Following the reasoning from last paragraph we can see from Figure 5 that g1, g2, g10, g3, g5 and g6 represent the vertices of the polygon. Regarding environments, e2 and e4 can be grouped, e3 is a single environment stratus and the same happens with e5.

The GGE biplot which won where what is shown in Figure 5.

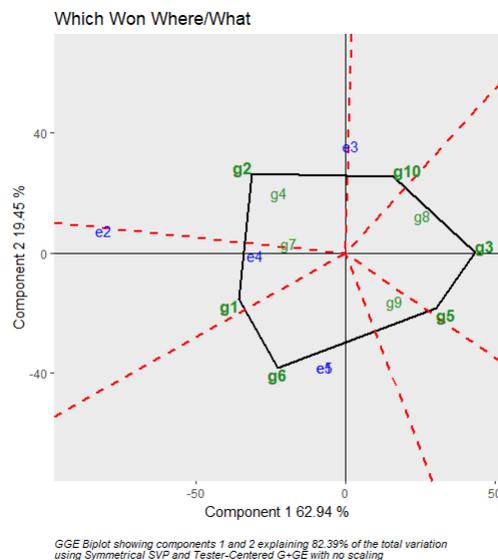


Figure 5 - Which won where what biplot for simf2.

Mean x stability

The GGE biplot shown the means x stability is present in 6.

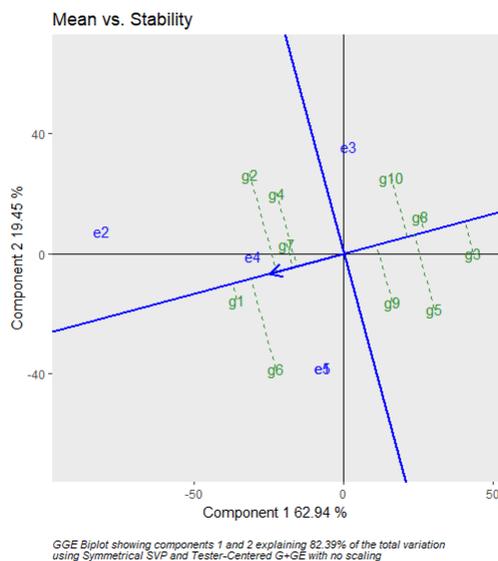


Figure 6 - Means x stability gge biplot for simf2.

Hongyu (2015), as well as Yan and Kang (2003), state that an ideal genotype should have high mean performance and produce at a range of several environments. The GGE biplot that shows discrimination and representative is in Figure 7 is a good tool to evaluate these aspects of genotypes.

Figure 6, according with the rules described in methodology section, allows the ranking of genotypes as $g1 > g6 > g2 > g7 > g4 > g9 > g10 > g5 > g8 > g3$.

Mean x stability

Hongyu (2015) states that the Discriminative and representative gge biplot allow the discrimination of genotypes and representation of target environments.

In the case of simf 2 all vectors of environments were long implies that are good environments to discriminate the genotypes. Environments that have small angles with MAE are better representatives of the target environment.

Regarding environment stability and performance the relationship among genotypes and environments via GGE analysis was the same as in Table 3 showing the capacity of GGE analysis in detecting the properties of the data simulated with specific properties.

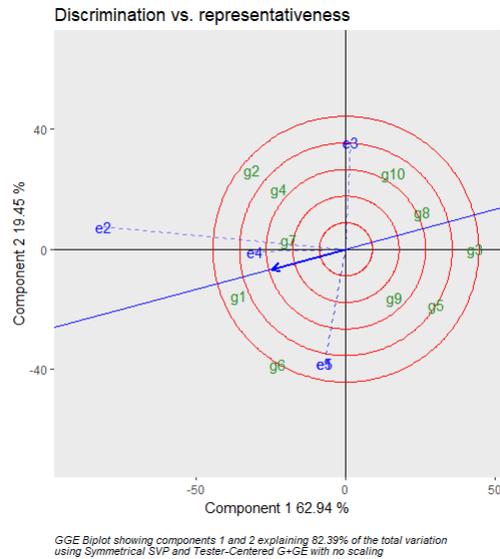


Figure 7 - Discrimination representative of ggbiplot for simf2.

3.2 Comparison between AMMI, w-AMMI and GGE biplots

Table 14 describes the single values for GGE, AMMI and WAMMI models and its respective principal components.

Using Table 14 was obtained Table 15 which is a correlation matrix. Such table describes the correlation between the explained variability obtained by means of principal components in GGE, AMMI and WAMMI models.

In Table 15 we can check that the 3 models are highly correlated with correlations near to 1. In other words the three models describe in an efficient way the variability contained in simf2 data set.

Table 15 - Single Values from PCS of GGE, AMMI and W-AMMI models

PC	GGE biplot	AMMI	W AMMI
1	7156.89	6355.64	0.00413
2	4231.38	3625.09	0.00304
3	3516.99	2147.54	0.00247
4	1797.54	1646.20	0.00112
5	0.00	0.00	0.00

This information is confirmed by the Table 16 that shows the Information Criteria for Each model and respective principal components. For all the models AMMI and WAMMI we have $PC1 > 1$.

Table 16 - Correlation between the PCS for GGE, AMMI and W-AMMI models

	GGE	AMMI	W AMMI
GGE	1.00	0.98	0.99
AMMI	0.98	1.00	0.95
W-AMMI	0.99	0.95	1.00

Comparing the models the PC1 of AMMI model was the greater being the one that most explained the variability of the data set followed by GGE and finally by WAMMI.

Hongyu (2015) and Hongyu *et al.* (2015) quoting (GAUCH, 2006) state that AMMI model have higher accuracy in visualizing patterns in experiments when compared to GGE biplots.

Hongyu (2015) and Hongyu *et al.* (2015) quoting (GAUCH, 2006) state that one critic to be made to GGE is that this methodology does not separate the G effect from GEI, but Yan (2011) declares that if the breeder interest in selection of genotype is not based in separation of G and GEI factors GGE biplot can be used.

Table 17 - Information criteria for AMMI, W AMMI and GGE models

PC	AMMI	IR	W-AMMI	IR	GGE	IR
1	66.4	3.32	50.6	2.53	62.94	3.147
2	21.6	1.08	27.4	1.37	19.45	0.9725
3	7.6	0.038	18.2	0.91	13.25	0.6625
4	4.5	0.0225	3.8	0.19	4.36	0.218

A general remark, however, is that, even capturing the same amount of variability and general characteristics of environments and genotypes, GGE and AMMI should be used carefully. This is because in GGE the terms of main effects of genotypes and interaction are summed together G+GEI. This doesn't happen when we use AMMI and WAMMI models. In breeding programs this information should be considered during selection of genotypes via GGE and AMMI methods.

Lastly, but not least important, we can say that AMMI modelling can be used to model of data generated by non normal continuous distribution such as logistic distribution. This result is possible once such distribution was used to simulated the data sets.

As recommendation for futures studies we suggested the comparison of the same models with another important version of AMMI called ROBUST AMMI described in Rodrigues *et.al.* (2016), as well as use of simulated data with different values for the coefficient of variation for the experiments.

Acknowledgments

Authors would like to thank reviewers and editors for their comments and suggestions.

SARTI, D. A.; DIAS, C. T. S.; Comparação entre AMMI, WAMMI e GGE no contexto de dados simulados. *Rev. Bras. Biom.*, Lavras, v.38, n.3, p.290-323, 2020.

- **RESUMO:** A interação genótipo x ambiente é uma questão fundamental no melhoramento de plantas e no desenvolvimento de novas cultivares. A modelagem de tais interações tem enorme importância nas decisões de melhoramento de plantas e otimização de programas de melhoramento. Nesse contexto, os modelos AMMI, W-AMMI e GGE permitem o estudo dessas interações. O presente trabalho tem como objetivo verificar o comportamento de tais modelos na presença de dados simulados com propriedades paramétricas bem conhecidas. Os resultados mostram que os três modelos são eficientes para modelar interações GxE.
- **PALAVRAS-CHAVE:** AMMI; WAMMI; GGE; genótipo x ambiente

References

CORNELIUS, P. L. Statistical tests and retention of terms in the additive main effects and multiplicative interaction model for cultivar trials *Crop Science*, v.33, p.1186-1193, 1993.

DIAS, C. T. S. *Metodos para escolha de componentes em modelo de efeito principal aditivo e interação multiplicativa (AMMI)*, 2005. 73p. Thesis (Ph.D.) - Universidade de São Paulo, Piracicaba, 2005.

DUARTE, J.; VENCOSKY, R. Interação genótipo ambiente: uma introdução análise AMMI. *SBG*, v1, p.1-60, 1999.

FRIESTSCHE-NETO, R.; MIRANDA, G. V.; DELIMA, R. O.; SOUZA, H. N. Factor analysis and sreg gge biplot for the genotype environment interaction stratification in maize. *Ciência Rural*, v.14, p.1043-1048, 2010.

GAUCH, H. *Statistical analysis of regional yield trials: ammi analysis of factorial designs*. 1.ed. London: Elsevier, 1992. 200p.

GAUCH, H. Statistical analysis of yield trials by ammi and gge. *Crop Science*, v.46, p.1488-1500, 2006.

GAUCH, H. A simple protocol for ammi analysis. *Crop Science*, v.53, p.1860-1869, 2013.

HONGYU, K. *Comparação do GGE biplot ponderado e AMMI ponderado com outros modelos de interação genótipo ambiente*, 2015. 100p. Thesis (Ph.D.) - Universidade de São Paulo, Piracicaba, 2015.

- HONGYU, K.; GARCIA PENA, M.; ARAUJO, L.; DIAS, C. Statistical analysis of yield trials by ammi analysis of genotype x environment interaction. *Biometrical Letters*, v.51, p.89-102, 2014.
- HONGYU, K.; SILVA, F. L., OLIVEIRA, A., SARTI, D. A., DIAS, C. T. S. Comparação entre os modelos ammi e gge biplot para dados de ensaios multi-ambientes. *Revista Brasileira de Biometria*, v.3, p.139-155, 2015.
- MENDIBURU, F. D. *Agricolae: Statistical Procedures for Agricultural Research*. R Package Version 1.2-3. 2015. <http://CRAN.R-project.org/package=agricolae>
- R CORE TEAM. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2019. URL <http://www.R-project.org/>
- ROBERTSON, A. A weighted ammi algorithm to study genotype-by-environment interaction and qtl-by-environment interaction. *Crop Science*. v.54, p.1555-1570, 2014.
- RODRIGUES, P. C; MONTEIRO, A.; LOURENÇO V. M. A robust AMMI model for the analysis of genotype-by-environment data. *Bionformatics*. v.32, p.58-66, 2016.
- SARTI, D. A. *Uncertainty management through decision analysis: applications to production optimization and uncertain demands*, 2013. 100p. Thesis (MSc) - Universidade de São Paulo, Piracicaba, 2013.
- SARTI, D. A. *The statistical paradigm: probabilistic and multivariate analysis applied through computational simulation in the interaction between genotype x environment*, 2013. 100p. Thesis (MSc) - Universidade de São Paulo, Piracicaba, 2013.
- YAN, W. Biplot analysis of multi environment trial data: principles and applications. *Canadian Journal of Plant Sciences*. v.86, p.623-645, 2006.
- YAN, W.; HOLLAND, J. B. A heritability-adjusted gge biplot for test environment evaluation. *Crop Science*. v.171, p.355-369, 2010.
- YAN, W.; HUNT, L. A.; SHENG, Q.; Szlavnic, Z. Cultivar Evaluation and Mega-Environment Investigation Based on the GGE Biplot. *Crop Science*. v.40, p.597-604, 2000.
- YAN, W.; KANG, M. *GGE Biplot Analysis: A graphical tool for breeders, geneticists and agronomists*. London, 2003. 400p.
- YAN, W; TINKER, A. Biplot analysis of multienvironment trial data: principles and applications. *Canadian Journal of Plant Sciences*. v.86, p.623-645, 2006.
- YAN, W. Singular-value partition for biplot analysis of multi-environment trial data. *Agronomy Journal*. v.94, p.990-996, 2002.

YAN, W. Gge biplot vs. ammi graphs for genotype-by-environment data analysis.
Indian Society of Agricultural Statistics. v.65, p.181-193, 2011.

Received on 18.05.2019.

Approved after revised on 05.05.2020.