Multivariate Genotype and Genotype by Environment Interaction Biplot Analysis of Sugarcane Breeding Data Using R

Ouma Victor Otieno[*], Onyango Nelson Owuor

School of Mathematics, University of Nairobi (UoN), Nairobi, Kenya

Email address:

victor.otieno.oo@gmail.com (O. V. Otieno), onyango@uonbi.ac.ke (O. N. Owuor)

[*]Corresponding author

Abstract: Complexity of Genotype by environment interaction (GxEI) in sugarcane multi-environmental trial (MET) requires further evaluation for genotypes performance determination. Genotype and genotype by environment (GGE) is one of the many statistical techniques for evaluating the interaction with emphasis on genotypes. Many statistical analysis tools for GGE exists with usage depending on cost and knowhow. R open source analytical software ensures availability and the knowledge on the necessary packages is required thus the objective of the paper on utilization of GGE using R software in the evaluation of genotypes with presence GxEI. The application used secondary data of Kenyan Mtwapa series of 96 and 97 preliminary varietal trial stage 4 established under randomized complete block design (RCBD), consisting of 15 test genotypes and three controls in the environments of SONYsugar, Mumias and KibosF9 with the plant crop and ratoon crop cycles as seasons. The 2-way GEI data was handled using singular value decomposition (SVD) through the R package; GGEbiplot programmed scripts and graphical user interface (GUI) were used in ranking genotypes and environments, determining genotypes performance overall and in each environment, determining stabilities and adaptability of the genotypes and identifying mega trial environments. GGEbiplot unpacked the GEI through the principle components (PC) 1 and 2 that sufficiently explained 85.37% of the variations.

Keywords: Genotype by Environment Interaction, Genotype and Genotype by Environment, Singular Value Decomposition, R-software, GGEBiplot and Sugarcane

1. Introduction

Stages two-five of sugarcane varieties trials are multi-environmental trials (MET) that often result in genotype and environmental interaction (GEI). The interaction is the non-uniform change in the performance of the test genotypes given the change in test environment. Statistically significant GEI makes it difficult in recommending the best performing cultivars as it confounds their evaluation, thus the need to evaluate the GEI further for better understanding, interpretation and recommendation. Literature shows that a variety of regression based statistical methods had been used in evaluating the GEI, these includes Additive and multiplicative interaction (AMMI), Genotype and Genotype by Environment (GGE), Regression (SRER) among others. Multivariate GGEbiplot is a graphical extension of GGE and has been used in evaluating cultivars whenever GEI occurs in a number of trials for various crops. Examples includes; discriminating nitrogen use efficiency in popcorn lines in Brazil [1], determining yield stability of promising hybrid rice in Bangladesh [2], analyzing soybean multi-environment yield trial data in north western Ethiopia [3] and yield trials of rice produced in a temperate climate [4]. In other fields, biplot has been used in soya beans breeding [5], pathogen study [6], dialed cross table [7] and qualitative trait loci (QTL) by Environment [8].

Statistical computing made light processing of GEI using various statistical techniques and packages by many

scientists; Statistical analysis systems (SAS) in evaluating GEI [9], Minitab [10], Stata [11], SPSS [12] and various SAS products [13]. However, functionalities are often limited and results hard to obtain with the exception of STATISTICA [14]. Greater functionalities for biplot programs are provided by; XLS-Biplot [15], GGEBiplot [16] and BiPlot [17] and Genstat package [18]. Tailor made packages for multivariate visualization of data exists for specific disciplines such as ecology; brodgar [19, 20, 21] and PC-ORD [22, 23, 24]. The other alternative is programming using programming languages.

R is a free statistical programming language and environment capable of producing high-quality graphics; it is an open-source implementation of the S programming language, available for download for all the major platforms from the R project home page [25]. The R packages for Biplot analysis for the GxEI includes GGEBiplot [26], and GGE [27], that are better than the previously mentioned as they produce high quality biplot graphics. Many researchers have also use direct scripting in R [28].

Review of GGEBiplot analysis covered its principles and GxEI data handling. Biplot analysis solves genotypes stability problem in the presence of interaction by matching each genotype to specific environment through singular value decomposition (SVD) of the GxE matrix [16]. GGE implies G+GE, that is G and GE being the two sources of variation considered for genotype and test environment evaluation while the biplot expresses GE matrix graphically. Biplot analyses have mitigated the effect of GE by visualizing relationships among genotypes, environments and their interaction through SVD [8, 29-33].

Biplot analysis is applicable to a 2-way (G x E), 3 way (G x E x Trait (T)) and even 4 way (Year x E x G x T) interaction data. Biplot inner product property, considers matrices G_{mxr}, E_{rxn} and P_{mxn} of m row, r columns, r rows and n columns and m rows and n columns respectively that can be displayed as in equation 1

$$P_{ij} = x_i x'_j + y_i y'_j = \vec{g}_i \vec{e}_j = |g_i||e_j| \cos \theta_{ij} \quad (1)$$

Where (x_i, y_i) are the coordinates for genotype (row) i and (x'_j, y'_j) are coordinates for environments (column) j; $\vec{g}_i$ is the vector for genotype (row) i and $\vec{e}_j$ is the length for environment (column) j; $|g_i|$ is the vector length for row i and $|e_j|$ is the vector length for column j. θ_{ij} is the angle between the vectors of row i and column j.

Any 2 way GxE interaction table can be analyzed graphically as long as it can be sufficiently approximated by a rank 2 matrix, thus matrix P_{mxn} is SVD into matrices G_{mxr}, D_{rxr} and E_{nxr} as in equation 2

$$P_{mxn} = G_{mxr} D_{rxr} E^T_{nxr} \quad r \le \min (m, n) \quad (2)$$

Where G_{mxr} is mxr matrix characterizing m genotypes, E_{nxr} is rxn matrix characterizing n environments and D_{rxr} is a diagonal matrix of r singular values. Matrix P_{mxn} is SVD into r principle components (PC), each having genotype vector (ε_{ii}), environment vector (η_{ij}) and singular values (λ_r) equation 3

$$P_{ij} = \sum_{i=1}^{r} \varepsilon_{ii} \lambda_r \eta_{jj} \ (\lambda_r \ge \lambda_{r+1}) \quad (3)$$

Where r- rank of GxEI 2 way table, it's the no. of PC necessary to fully represent P_{mxn}. $r \le \min (m, n)$. When $r < m$, there is a linear relationship among genotypes and that is equally true when m>n. When $r < n$, there are association among environments. Environments are independent iff r=n. (λ_r) is the singular value for PC$_i$, λ_r^2 is the eigen value of $P^T P$ or PP^T. G i^{th} columns are eigen vector of PP^T corresponding to λ_i^2 and the ith column of E are the eigen vector of $P^T P$ corresponding to the λ_i^2. The SVD requirement is that $G^T G$ =I_{rxr} =$E^T E$ where I_{rxr} – r x r identity matrix

The goodness of fit for the biplot is given by $(\lambda_1^2 + \lambda_2^2)/SS$ with SS being the sum of squares of the GxE interaction table. Given that $(\lambda_r > \lambda_{r+1})$, biplot PC1 and PC2 captures important patterns of P even when goodness of fit is poor. Excellent goodness of fit shows the strength of association among environments or among genotypes, poor goodness of fit is an indication of complex patterns or lack of discerning patterns.

The singular values are partitioned into genotypes and environment scores before biplot and then used to construct the approximated GxE two way data (equation 4)

$$P_{ij} = \sum_{i=1}^{r} \varepsilon_{ii} \lambda_r \eta_{ij} = \sum_{i=1}^{r} (\varepsilon_{ii} \lambda_r^f)(\lambda_r^{1-f} \eta_{ij}) \ f \ (0 \le f \le 1) \ (4)$$

f $(0 \le f \le 1)$ provides the ways of partitioning the single values; column metric preserving (f=0), row metric preserving (f=1) and symmetrical partitioning (f=1/2).

The dot product and correlations between two columns are provided and estimated as $\sum_{i=1}^{m} P_{ij} P_{ij'} = |e_j||e_{j'}| \cos \theta_{jj'}$ and $r_{ij} = \sum_{i=1}^{m} (P_{ij} - \bar{P}_j)(P_{ij'} - \bar{P}_{j'})/sqrt(\sum_{i=1}^{m} (P_{ij} - \bar{P}_j)^2 \sum_{i=1}^{m} (P_{ij'} - \bar{P}_{j'})^2$ respectively.

In the GGE biplot case the 2 way data is column centered and $\bar{P}_j = \bar{P}_{j'} = 0$ and thus $r_{ij} = \sum_{i=1}^{m} (P_{ij})(P_{ij'})/ sqrt(\sum_{i=1}^{m} (P_{ij})^2 \sum_{i=1}^{m} (P_{ij'})^2 = \sum_{i=1}^{m} P_{ij} P_{ij'} /|e_j||e_{j'}|$ and $\sum_{i=1}^{m} P_{ij} P_{ij'} /|e_j||e_{j'}|r_{ij}$ and $r_{ij} = \cos \alpha_{jj'}$ given that $\sum_{i=1}^{m} P_{ij} P_{ij'} = |g_j||e_j| \cos \theta_{ij}$ thus relationship of columns centered data becomes equality between cosine and correlation. The vector length of the column equals sqrt(m-1)*sd(column factor across the row); $S_j sqrt(m-1) = sqrt(\sum_{i=1}^{m} (P_{ij} - \bar{P}_j)^2 = sqrt \sum_{i=1}^{m} P_j^2 = |e_j|$

1.1. Data Centering Prior to SVD

Let the GxE 2 way table be Y. The value of each cell is the mixed effect of grand mean (µ), genotype main effect (α_i), environment main effect (β_j), genotype by environment interaction ($\emptyset_{ij}$) and a random effect (ε_{ij}) thus the model

$$Y_{ij} = \mu + \alpha_i + \beta_j + \emptyset_{ij} + \varepsilon_{ij} \quad (5)$$

The GE matrix P can be any part of Y_{ij} and could be modified to different models ignoring the random term thus

$$P_{ij} = Y_{ij} = \mu + \alpha_i + \beta_j + \emptyset_{ij} \quad (6)$$

$$P_{ij} = Y_{ij} - \mu = \alpha_i + \beta_j + \emptyset_{ij} \quad (7)$$

$$P_{ij} = Y_{ij} - \mu - \alpha_i = \beta_j + \emptyset_{ij} \qquad (8)$$

$$P_{ij} = Y_{ij} - \mu - \beta_j = \alpha_i + \emptyset_{ij} \qquad (9)$$

$$P_{ij} = Y_{ij} - \mu - \alpha_i - \beta_j = \emptyset_{ij} \qquad (10)$$

The biplot models 6, 7, 8, 9 and 10 are interpreted differently and depend on the research objective. Model 6 is the additive and multiplicative interactive model (AMMI) for visualizing the entire dataset, model 10 is for pure GE visualization, model 9 is the GGE (G+GE) model for genotype evaluation through GGE biplot analysis.

1.2. Data Scaling Prior to SVD

The GGE biplot model 9 is modified as

$$P_{ij} = (Y_{ij} - \mu - \beta_j)/S_i = (\alpha_i + \emptyset_{ij})/S_i \qquad (11)$$

Where S_i is the scaling factor for the column (environment or traits j) thus model 9 is the case where $S_i=1$. When S_j is the standard deviation (sd) for columns, the data is standardized giving all columns the same weight. When S_j is standard error (se), the heterogeneity among the environments is removed. The axes of the biplots are drawn to scale.

1.3. Biplot Analysis of GE in MET Data

Statistically significance GEI necessitates examination of changes in ranks of genotypes (crossover) in different environments resulting in different winners in different environments. Lack of recognizable GE patterns indicates presence of single mega environment and a model addressing random sources of variation is appropriate. GEI necessitates determining whether target environment should be divided into meaningful mega environments for exploiting /avoiding GE, determining the causes of GE interaction, the best test environments (representative and discriminating) and superior genotypes that are high and stable performers

The next section of the paper has the materials and methods enumerating the SVD of GEI for Environment and Genotype score and singular values and using them through the combination of GGEBiplot and gge R software packages in the production of biplots in the further evaluation of GEI in terms of the genotypes and environments adaptability and stability.

2. Materials and Methods

2.1. Study Environment: Kibos, Sony and Mumias

Data: The test clones used in the study were; 96_153, 96_173, 96_50, 97_100, 97_102, 97_131, 97_204, 97_215, 97_221, 97_224, 97_226, 97_264, 97_317, 97_321, 97_399 and 4 controls; CO421, CO617, N14 and CO945 of Mtwapa series established in 96 and 97. CO617 and CO945 were removed from the analysis as they didn't contribute the interaction as they were planted in single sites

2.2. Statistical Method: Multivariate GGE Biplot

Multivariate GGE biplot analyses are based on the mean yield phenotypic measurement of the test cultivars. The model was as shown in equation 12

$$Y_{ij} - \mu = G_i + E_j + GE_{ij} \ (i = 1,2,\cdots,17; j = 1,\cdots,3) \ (12)$$

Where Y_{ij} represents the phenotypic average of genotype i in environment j, μ is the grand mean, G_i is the fixed effect of genotype i, E_j is the fixed effect of environment j, GE_{ij} is the random effect of the interaction between genotype i and environment j.

Equation 13 indicates the principle of biplot. It referred to as the inner-product property of the biplot allowing estimation of elements of matrix P and visualizing them, ranking the rows relative to any column, ranking the columns relative to any row, comparing any two rows relative to individual columns, identifying the rows with largest (or smallest) values for each column, or vice versa [29].

$$GE_{ij} = G_iG_j' + E_iE_j' = \vec{g}\vec{e} = |g_i||e_j|\cos\theta_{ij} \ (i = 1,2,\cdots,17; j = 1,\cdots,3) \qquad (13)$$

Where (G_i, E) are the coordinates for genotype i and (G_j', E_j') are coordinates for environments j; $\vec{g}_i$ is the vector for genotype i and $\vec{e}_j$ is the length for environment j; $|g_i|$ is the vector length for genotype i and $|e_j|$ is the vector length for environment j. θ_{ij} is the angle between the vectors of row i and column j.

SVD

GGE biplot model keeps the genotype effect (G) and genotype x environment (GE) interaction effects together as multiplicative terms as expressed by equation 14

$$Y_{ij} - \mu - \beta_j = g_{1i}e_{1j} + g_{i2}e_{2j} + \varepsilon_{ij} \ (i = 1,2,\cdots,17; j = 1,\cdots,3) \qquad (14)$$

where Y_{ij} is the yield performance expected of genotype i in environment j; μ is the overall mean; $β_j$ is the main effect of environment j; g_{1i} and e_{1j} are the main scores for the ith genotype in the jth environment, respectively; g_{i2} and e_{2j} are the secondary scores for the ith genotype in the jth environment, respectively; and ε_{ij} is the residual not explained by either of the effects ("noise").

Thus, construction of the biplot in the GGE model is accomplished by the simple dispersion of g1i and gi2 for genotypes and e_{1j} and e_{2j} for environments, via singular value decomposition (SVD), in accordance with the equation 15

$$Y_{ij} - \mu - \beta_j = \lambda_1\xi_{i1}\eta_{1j} + \lambda_2\xi_{i2}\eta_{2j} + \varepsilon_{ij} \ (i = 1,2,\cdots,17; j = 1,\cdots,3) \qquad (15)$$

where λ_1 and λ_2 are the largest eigenvalues of the first and second principal components; PC1 and PC2, respectively; ξ_{i1} and ξ_{i2} are the eigenvalues of the ith genotype for PC1 and PC2, respectively; and η_{1j} and η_{2j} are the eigenvalues of the jth environment for PC1 and PC2, respectively

2.3. R GGEbiplotGUI and GGEbiplot Scripting

The GGE biplot analysis was performed with the aid of R software [25]. It includes mixed model analysis the for the phenotypic yield data of the genotypes. The workings are as follows;

```
setwd() # setting working directory
model <- AMMI(Locality, Genotype, Rep, MKMGHA, console=FALSE)model$ANOVA
install.packages("GGEBiplotGUI", dependencies=TRUE) #Installing GGEBiplotGUI package
library(GGEBiplotGUI) #Calling the library
head(gge) #View top 6 rows of data
GGEBiplot(Data=gge) # GGE biplot analysis
```

R codes for the analysis

```
#Biplot analysis in R using GGEBiplot analysis for GxE
Setwd()
dtge<-read.csv("G:/GGE                     Biplot analysis/MS97&98SeriesGxE.csv", header=TRUE)
View(dtge)
library(agricolae)
library(GGEBiplotGUI)
attach(dtge)
dtge2 <- dtge[,-1]
rownames(dtge2) <- dtge[,1]
GGEBiplot(Data=dtge2)
svd(dtge2)
as.matrix(svd(dtge2)$d)
dtge2<-as.matrix(dtge2)
prcomp(dtge2)  #The get the principle components standards deviation and variences
prcomp(dtge2)$sd # standard deviation
(prcomp(dtge2)$sd)^2 # variances
######################################
dtge<-read.csv("G:/GGE                     Biplot analysis/MS96&97series.csv", header=TRUE)
View(dtge)
library(GGEBiplotGUI)
library(gge)#loading the library gge
#install. packages("GGEBiplots", dependencies=TRUE)# Installing the package GGEBiplots and
#dependent packages;'farver', 'tweenr', 'ggforce'
library(GGEBiplots)#Loading the library (GGEBiplot)
#Creating the two way GxE table of means
meantab<-stattable           (dtge$VARIETY,           dtge$Site, dtge$TCH, FUN=mean, na. rm=TRUE)
View(meantab)# Viewing the GxE table of means
GGE1<-GGEModel(meantab) # GGE modeling of the data "meantab"for the GxE data
CompareGens(GGE1,"N14","96_50")#Comparing        two genotypes at a time
DiscRep(GGE1)#Getting  the  discriminativeness  and Representativeness
EnvRelationship(GGE1)  #Examining  the  environments relations
#Examining the Environments
#mfrow3d(1,3)
```

```
ExamineEnv(GGE1,"Mumias","Sony")#Examining the Environment "Mumias"
ExamineEnv(GGE1,"Sony")#Examining  the  Environment "Sony"
ExamineEnv(GGE1,"KibosF9")#Examining the Environment "KibosF9"
# Examining the Genotypes
ExamineGen(GGE1,"96_153");
ExamineGen(GGE1,"96_173");
ExamineGen(GGE1,"96_50");
ExamineGen(GGE1,"97_100");
ExamineGen(GGE1,"97_102");
ExamineGen(GGE1,"97_131");
ExamineGen(GGE1,"97_204");
ExamineGen(GGE1,"97_215");
ExamineGen(GGE1,"97_221");
ExamineGen(GGE1,"97_224");
ExamineGen(GGE1,"97_226");
ExamineGen(GGE1,"97_264");
ExamineGen(GGE1,"97_317");
ExamineGen(GGE1,"97_321");
ExamineGen(GGE1,"97_399");
ExamineGen(GGE1,"CO421");
ExamineGen(GGE1,"N14");
#General GGE plot
GGEPlot(GGE1)
# Specifying the outcome plot outlook
GGEPlot(GGE1, type = 4, d1 = 1, d2 = 2, selectedE = NA, selectedG = NA, selectedG1 = NA, selectedG2 = NA, colGen = "forestgreen", colEnv = "blue", colSegment = "red", colHull = "black", sizeGen = 4, sizeEnv = 4, largeSize = 4.5, axis_expand = 1.2, axislabels = TRUE, axes = TRUE, limits = TRUE, titles = TRUE, footnote = TRUE)
GGEPlot(GGE1, type = 4, d1 = 1, d2 = 2, selectedE = NA, selectedG = NA, selectedG1 = NA, selectedG2 = NA, colGen = "forestgreen", colEnv = "blue", colSegment = "red", colHull = "black", sizeGen = 4, sizeEnv = 4, largeSize = 4.5, axis_expand = 1.2, axislabels = TRUE, axes = TRUE, limits = TRUE, titles = FALSE, footnote = FALSE)
#Using the gge package
GGE2<-gge(as. matrix(meantab))
GGEPlot(GGE2) #ploting
####################
MeanStability(GGE1, footnote = FALSE) # Getting the means versus stability
RankEnv(GGE1, footnote = FALSE)# Ranking the Environments
RankGen(GGE1, footnote = FALSE)# Ranking the Genotypes
GGEPlot(GGEModel(meantab)) #GGE plot
WhichWon(GGE1) # which won where plot
```

3. Results and Discussions

The results and discussions are presented in four sections: section one represents the results of analysis of variance and testing the GEI significance hypothesis, section two is the

SVD of GEI providing genotypes and environment scores singular values, variations explained and information ratio, section three on genotype performance and stability, relationships amongst test environments, which genotypes won-where identifying the best genotypes for each environment; ranking genotypes and environments, discriminativeness and representativeness and stability and adaptability.

3.1. GGEBiplot ANOVA

ANOVA (table 1) indicates the from the combined analysis

sugarcane cultivars yields were affected by genotypes (9.38%), environment (42.42%) crop class-CC (6.93%) and GE interaction (8.09%). G, E and GEI were highly significant (P<0.0001, P<.0001, P<.0001 and P=0.0006 respectively) for sugarcane yield. The highly significant G × E effects suggest that genotypes be tested for adaptation to specific environments [34-35]. GEI effects show that genotypes performed differently to the variations in environmental conditions of sugarzones and qualified multi-location trials.

Table 1. ANOVA.

Source	DF	SS	Mean Square	F-Value	Pr>F	Explained SS%
VARIETY	16	55659.57	3478.7229	5.05	<.0001	9.38
Site	2	251671.8	125835.8753	182.63	<.0001	42.42
REP(Site)	6	15093	2515.4992	3.65	0.0018	2.54
CC	2	41119.65	20559.824	29.84	<.0001	6.93
Site*VARIETY	32	48015.29	1500.4777	2.18	0.0006	8.09
Site*VARIETY*CC	49	45308.72	924.6678	1.34	0.0832	7.64
Error	198	136424.9	689.0145			22.99
Corrected Total	305	593292.8				

3.2. Singular Value Decomposition (SVD) of Genotype and Environment Interaction

SVD of the GEI provides the environmental scores, genotype scores and the singular values as shown in tables 2, 3 and 4

Table 2. Environmental scores.

	[,1]	[,2]	[,3]
[1,]	-0.42425	0.10430	0.89952
[2,]	-0.55028	0.75921	-0.34756
[3,]	-0.71917	-0.64244	-0.26470

Environments; 1-KibosF9, 2-Mumias and 3-Sonysugar.

Table 3. Genotypes scores.

Genotypes	[,1]	[,2]	[,3]	genotypes	[,1]	[,2]	[,3]
[96_153,]	-0.21389	0.222204	0.228274	[97_224,]	-0.25347	0.188533	0.457249
[96_173,]	-0.24563	0.115254	-0.1171	[97_226,]	-0.26859	-0.03879	0.29493
[96_50,]	-0.22438	-0.15053	-0.04905	[97_264,]	-0.2273	-0.29002	-0.29496
[97_100,]	-0.21374	0.103862	-0.16661	[97_317,]	-0.27351	-0.37906	0.04373
[97_102,]	-0.22387	0.178234	-0.34435	[97_321,]	-0.23048	0.023301	-0.08715
[97_131,]	-0.29254	-0.01848	0.022013	[97_399,]	-0.19884	0.312403	0.239912
[97_204,]	-0.23006	-0.03426	0.03436	[CO421,]	-0.26641	0.04307	-0.19717
[97_215,]	-0.23208	0.526887	-0.13988	[N14,]	-0.26478	-0.4451	0.367712
[97_221,]	-0.24247	-0.17335	-0.37416				

Singular value, explained variation, cumulative proportion, and variances for the principal components (PCs) of the GGE biplot analysis.

Table 4. Singular values, variations and variances.

PC	Singular Values	Explained%	Cumulative%	Variances
1	988.0726	65.48	65.48	707.16
2	66.51622	19.89	85.37	214.82
3	54.04387	14.63	100.00	158.02

Principal components one and two accounted for 85.37% of the total variation hence GGEbiplot method explaining a bigger proportion of the sum of G+GE. 14.63% could be attributed to PC3 or noise. PC with an IR value more than one shows that it accounts for more

information and variation unlike that with lesser than one as indicated by PC1 and thus making biplot analysis appropriate in explaining patterns related to GE interaction

3.3. Genotype Performance and Stability

Productive performance: genotypes allocated on the vertex of the polygon are farther from the origin than the other genotypes within the sector bounded by them and the most responsive. They show better or worse performance in some or in all of the environments and therefore they can be used to identify possible mega-environments. Genotypes located within the polygon are the least responsive to the stimuli of the environments
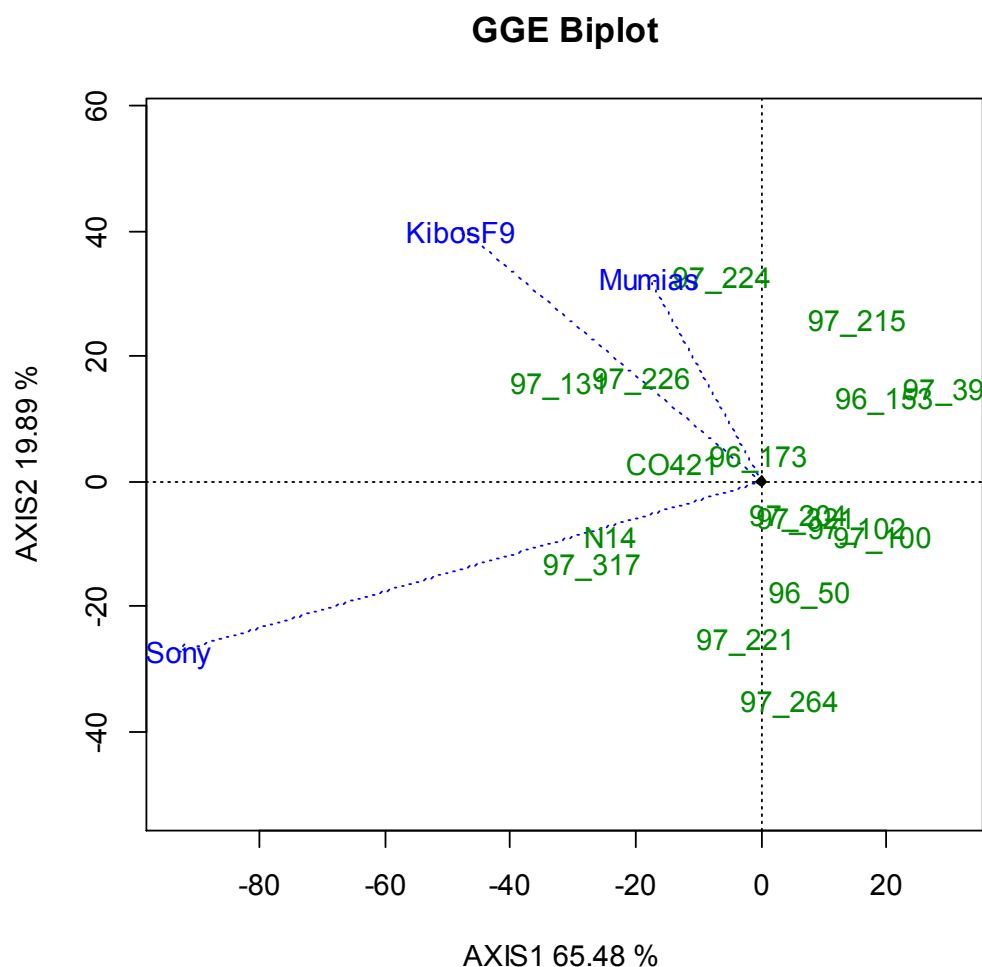
GGE Biplot

Figure 1. General GGE biplot.

3.4. Relationships Amongst Test Environments

Interrelationships among the three test environments are shown figure 2. Lines connecting the biplot origin and environments markers are environment vectors while the angle between the vectors of two environments provides the correlation coefficient between them; the cosine of the angle between the vectors of two environments estimates correlation coefficient between them [36-37]. Based on the cosine the environments were grouped into two; one being KibosF9 and Mumias and two being Sonysugar as shown by the distance between those environments. Nonetheless the angles between all the three environments are acute hence all were positively correlated hence the GE was moderately low and no crossover. The distance between environments measure their dissimilarity in discriminating genotypes and helps in grouping environments. Similarity (covariance) between two environments is determined by both length of their vector and cosine of angle between them. Closer association between KibosF9 and Mumias shows that the same information about genotypes could be obtained from either of the two and hence reduction in testing costs.
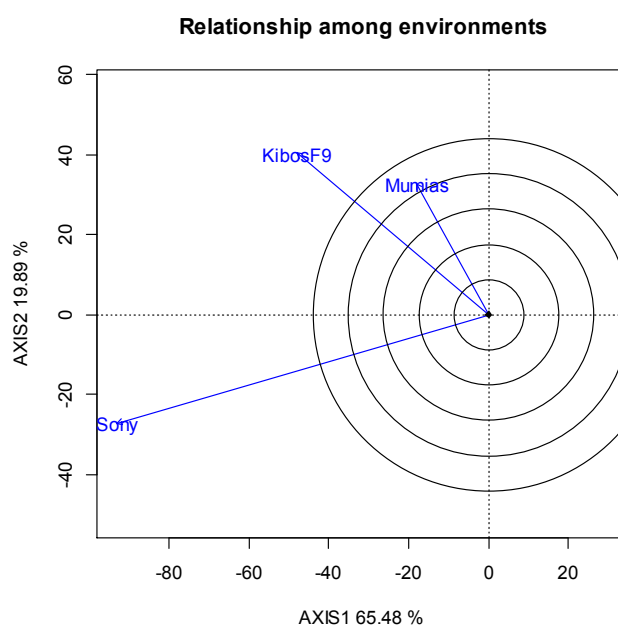
Relationship among environments

Figure 2. Relationships among environments.

3.5. Genotype Best Performance with Reference to Environment (Which Won Where)

GGE polygon (figure 3) views 15 sugarcane genotypes and two controls tested in the three environments. It identifies winning genotypes in different environments by visualizing GEI [29] in MET and help in estimating possible existence of different mega environments [5, 8, 38]. The vertex genotypes were 97_264, 97_317, 97_131, 97_224, 97_215, 97_39 and 97_100 having the largest distance from the origin. They are the best or poorest in some or all environments given they are farthest from the origin of biplot [29], and thus were more responsive to environmental change and are considered as specially adapted genotypes. The biplot is divided into seven sections by the seven rays with genotypes falling in all sections while the environments falling in three areas. The 97_131 and 97_224 cultivars are in vertices of the polygon in which the KibosF9 and Mumias environments are contained; thus, they are the most productive in those two environments in terms of yield. Cultivar 97_317 was in the vertex that had Sonysugar environment thus it's the best performer for yield in Sonysugar. The other genotypes on the vertices of the polygon not contain any of three environments are unfavorable in the three test environments and showed low responsiveness and yields. Test environments fell in to two sectors while five of the sectors in the polygon had no test environment. The best performing genotype in yield was 97_399 being the furthest to the right
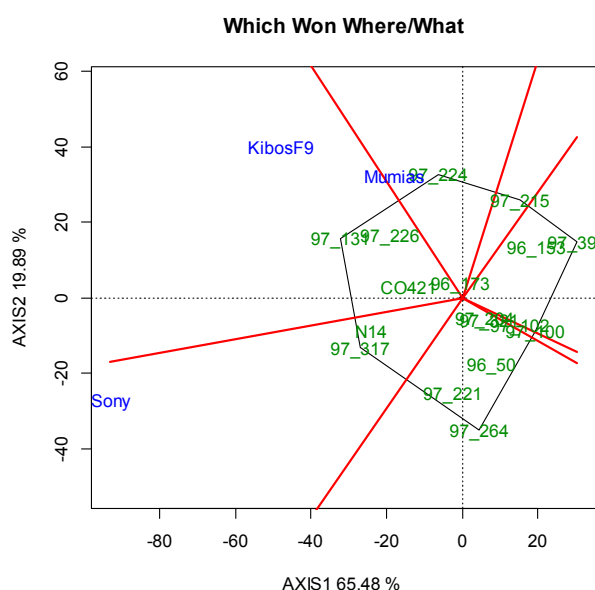

Figure 3. Genotypes adaptability and performance to environment.

3.6. Ranking the Genotypes and Environments

Figure 4 ranks 15 genotypes and two controls by mean yield and stability. Line passing through biplot origin from lower right to upper left is the average environment axis (AEA) as defined by the first two PCs of the environments scores. The furthest from the arrow is the genotype with the highest yield
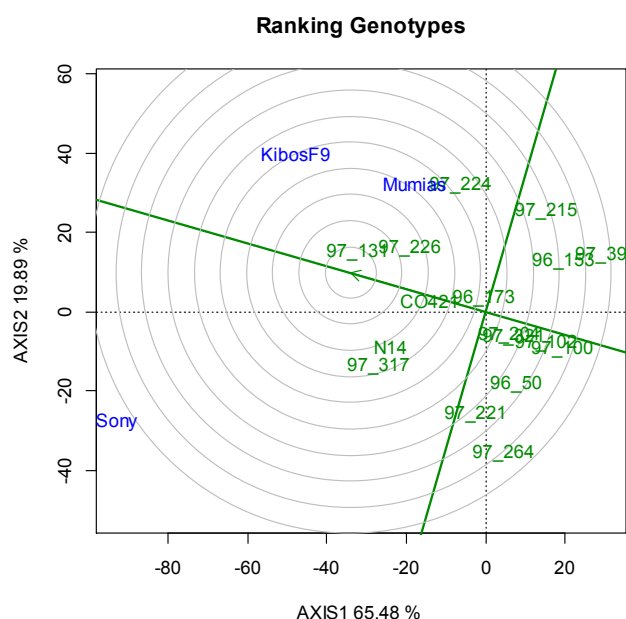

Figure 4. Genotypes ranks.

Ranking genotypes in a particular environment; the environment axis through the origin and that specific environment, genotypes closer to the environment along the axis are high yielding and vice versa. Thus 97_317 and N14 are ranked best on Sony, 97_224 in Mumias, 97_131, 97_226 and 97_224 in Kibos. Genotype axis through the biplot origin and that genotype, along that axis are the rankings of the environments. Genotypes located closer to the ideal genotype are more desirable than others. Figure 5 ranks the test environments, KibosF9 and Sony allocated to the third and fourth and second concentric circles are the closest to ideal in terms of yields and phenotypic stability. Kibos was ranked 1st at was closer to the middle of the concentric cycles followed by Sony and Mumias.
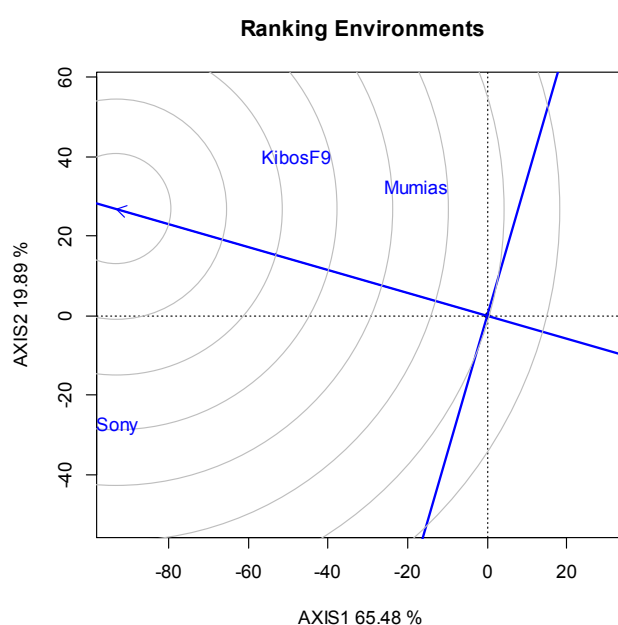

Figure 5. Environments ranking.

3.7. Discriminativeness and Representativeness

The circle indicated by the arrow represents the average environment. If the angle formed between the test environment and the line passing through the average environment is small, it means that this test environment is representative, and the larger the vector for each environment, the greater the discrimination capability [8]. Thus Sony was best at discriminating the genotypes followed by KibosF9 and Mumias in that order.

The average environment is at the point where the arrow is and the line being average environment axis (AEA). A test environment that has a smaller angle with the AEA is more representative of other test environments. Thus, KibosF9 was most representative as compared to Mumias and Sony. None of the zones was best in both discrimitiveness and representativeness as none felt on the AEA line. Discriminating but non-representative test environments such as SONY and Mumias are good for selecting specifically adapted genotypes if they can divided into mega-environments. Discriminating but non-representative test environments like Sony could be useful for culling unstable genotypes given that it was a single mega-environment. Mumias and Kibos could have been a mega environment given that they were close with an acute angle between them as shown in figure 6
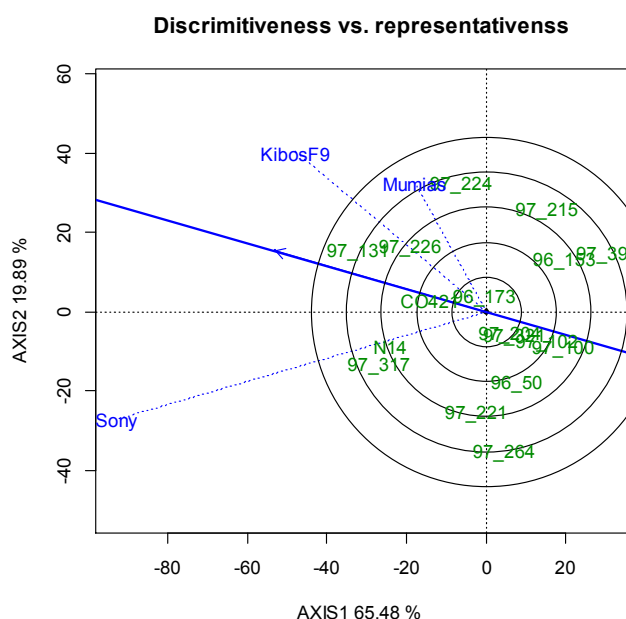
Discrimitiveness vs. representativenss

Figure 6. Genotypes and environments discriminativeness and representativeness.

3.8. Genotypes Mean and Stability

Genotypes yields and stability evaluation are based on the average environment coordination (AEC). An ideal environment is given by the mean score of principal components 1 and 2 (PC1 and PC2) for all test environments. Genotypes that showed higher yield than the overall mean were 97_224, 97_226, 97_131, 97_317 and N14. Genotypes

96_173, 97_221 and CO421 yields were similar to the overall mean yield. The others genotypes showed a lower yield than the overall mean (Figure 7).
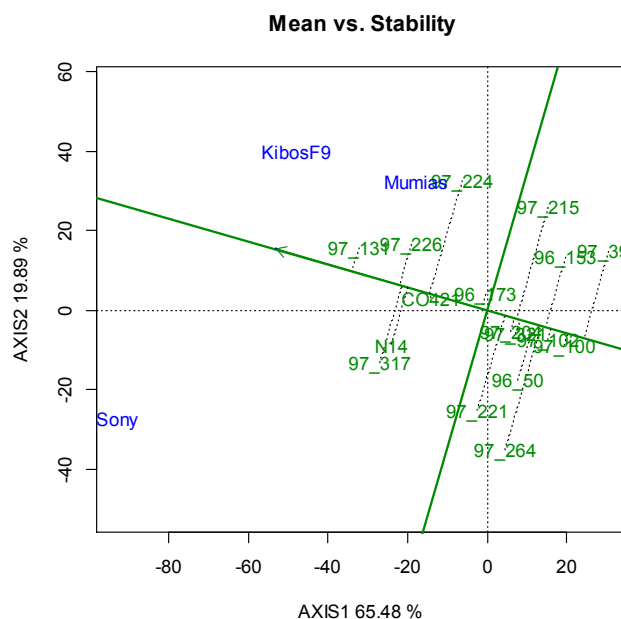
Mean vs. Stability

Figure 7. Mean and stability.

Comparing all the genotypes, the genotype metric preserved is used with the Euclidian distance between them measuring their overall dissimilarity. Genotype vector gives its contribution to the G+GE, thus the longest vector the large contribution to either G or GE or both (G+GE). Angle between genotype vector and AEA partitions the vector length into G and GE components. Right angle between them implies contribution to GE only, obtuse

angle implies contribution to mainly G giving lower than average performance and an acute angle implies contribution is mainly to G leading to higher than average mean performance. Angle between genotypes indicates similarity in response to the environment; an acute angle indicate they respond similarly to an environment, an obtuse angle implies they perform inversely (differently) to the environment; if one performs well then the other performs poorly. The right angle implies they perform independently to the environment.

4. Conclusion

GGEbiplot analyses simplified the unknown content of the GE interaction and enhance the evaluation of the test genotypes. 85.37% of the GEI patterns were accounted for the PC1 (65.48%) and PC2 (18.89%). The three trial environments were grouped into two mega trial environments; one being KibosF9 and Mumias and the second being Sonysugar. The genotypes 97_131 and 97_224 in the vertices of the polygon were most productive KibosF9 and Mumias environments while 97_317 was in Sonysugar environment. Genotypes 97_317 and N14 are ranked best on Sonysugar, 97_224 in Mumias

and 97_131, 97_226 and 97_224 in KibosF9. Genotypes that showed higher yield than the overall mean were 97_224, 97_226, 97_131, 97_317 and N14. Sony was best at discriminating the genotypes followed by KibosF9 and Mumias in that order. Genotypes 96_173, 97_221 and CO421 yields were similar to the overall mean yield. The others genotypes showed a lower yield than the overall mean. GGEbiplot as a as tool for evaluating GxEI through visualization is very good and accurate, it is easy to understand and show stable genotypes in the face of GxEI. The biplot itself allows flexibility on whether one desires an additive and interactive models or just interactive model. The graphics need to be accompanied with statistical test to make it more powerful. Combined with other statistical estimation methods it strongly complements visual aspect.

Acknowledgements

References

[1]   Santos, A. D., Amaral Júnior, A. T. D., Kurosawa, R. D. N. F., Gerhardt, I. F. S., & Fritsche Neto, R. (2017). GGE Biplot projection in discriminating the efficiency of popcorn lines to use nitrogen. *Ciência e Agrotecnologia*, *41* (1), 22-31.

[2]   Akter, A., Hasan, M. J., Kulsum, U., Rahman, M. H., Khatun, M., & Islam, M. R. (2015). GGE biplot analysis for yield stability in multi-environment trials of promising hybrid rice (Oryza sativa L.). *Bangladesh Rice Journal*, *19* (1), 1-8.

[3]   Atnaf, M., Kidane, S., Abadi, S., & Fisha, Z. (2013). GGE biplots to analyze soybean multi-environment yield trial data in north Western Ethiopia. *Journal of Plant Breeding and Crop Science*, *5* (12), 245-254.

[4]   Donoso-Ñanculao, G., Paredes, M., Becerra, V., Arrepol, C., & Balzarini, M. (2016). GGE biplot analysis of multi-environment yield trials of rice produced in a temperate climate. *Chilean journal of agricultural research*, *76* (2), 152-157.

[5]   Yan, W., & Rajcan, I. (2002). Biplot analysis of test sites and trait relations of soybean in Ontario. *Crop Science*, *42* (1), 11-20.

[6]   Yan, W., & Falk, D. E. (2002). Biplot analysis of host-by-pathogen data. *Plant Disease*, *86* (12), 1396-1401.

[7]   Yan, W., & Hunt, L. A. (2002). 19 Biplot Analysis of Multi-environment Trial Data.

[8]   Yan, W., & Tinker, N. A. (2005). An integrated biplot analysis system for displaying, interpreting, and exploring genotype× environment interaction. *Crop Science*, *45* (3), 1004-1016.

[9]   Dia, M., Wehner, T. C., & Arellano, C. (2016). Analysis of genotype× environment interaction (G× E) using SAS programming. *Agronomy Journal*, *108* (5), 1838-1852.

[10]   Minitab, I. (2014). MINITAB release 17: statistical software for windows. *Minitab Inc, USA*.

[11]   StataCorp, L. P. (2007). Stata data analysis and statistical Software. *Special Edition Release*, *10*, 733.

[12]   SPSS, I. (2017). SPSS 17.0. *Chicago, Ill, SPSS*.

[13]   SAS Institute Inc 2009

[14]   StatSoft, P. L. (2009). STATISTICA®, version 9.0.

[15]   Udina, F. (2005). Interactive biplot construction. *Journal of Statistical Software*, *13* (5), 1-16.

[16]   Yan W, Kang MS (2006). GGEbiplot, Version 5. URL http://www.ggebiplot.com/.

[17]   Lipkovich, I., & Smith, E. P. (2002). Biplot and singular value decomposition macros for Excel. *Journal of statistical software*, *7* (5), 1-15.

[18]   VSN International Ltd (2017). Genstat for Windows 19th Edition. VSN International Ltd, Hemel Hempstead, UK. URL http://www.genstat.co.uk/.

[19]   Highland Statistics Ltd. (2009). Brodgar: Software Package for Data Exploration, Univariate Analysis, Multivariate Analysis and Time Series Analysis.

[20]   Lepš, J., & Šmilauer, P. (2003). *Multivariate analysis of ecological data using CANOCO*. Cambridge university press.

[21]   Ter Braak, C. J., & Smilauer, P. (2002). *CANOCO reference manual and CanoDraw for Windows user's guide: software for canonical community ordination (version 4.5)*. www. canoco. com.

[22]   McCune, B., & Mefford, M. J. (1999). *PC-ORD: multivariate analysis of ecological data; Version 4 for Windows; [User's Guide]*. MjM software design.

[23]   Grandin, U. (2006). PC-ORD version 5: A user-friendly toolbox for ecologists. *Journal of Vegetation Science*, *17* (6), 843-844.

[24]   La Grange, A., le Roux, N., & Gardner-Lubbe, S. (2009). BiplotGUI: interactive biplots in R. *Journal of Statistical Software*, *30* (12), 1-37.

[25]   R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

[26]   Frutos, E., Galindo, M. P., Leiva, V. (2014). An interactive biplot implementation in R for modeling genotype-by-environment interaction. Stoch. Environ. Res. Risk Assess. 28:1629-1641

[27]   Kevin Wright and Jean-Louis Laffont (2018). gge: Genotype Plus Genotype-by-Environment Biplots. R package version 1.4. https://CRAN.R-project.org/package=gge

[28]   Dia, M., Wehner, T. C., & Arellano, C. (2017). RGxE: an r program for genotype x environment interaction analysis. *American Journal of Plant Sciences*, *8* (07), 1672.

[29]   Yan, W., & Kang, M. S. (2002). *GGE biplot analysis: A graphical tool for breeders, geneticists, and agronomists*. CRC press.

[30] Kang, M. S. (1997). Using genotype-by-environment interaction for crop cultivar development. In *Advances in agronomy* (Vol. 62, pp. 199-252). Academic Press.

[31] Gauch Jr, H. G. (1992). *Statistical analysis of regional yield trials: AMMI analysis of factorial designs*. Elsevier Science Publishers.

[32] Yan, W., Hunt, L. A., Sheng, Q., & Szlavnics, Z. (2000). Cultivar evaluation and mega-environment investigation based on the GGE biplot. *Crop Science*, *40* (3), 597-605.

[33] Yan, W. (2001). GGEbiplot—a Windows application for graphical analysis of multienvironment trial data and other types of two-way data. Agronomy journal, 93 (5), 1111-1118.

[34] Aina, O. O., Dixon, A. G. O., Paul, I., & Akinrinde, E. A. (2009). G× E interaction effects on yield and yield components of cassava (landraces and improved) genotypes in the savanna regions of Nigeria. *African Journal of Biotechnology*, *8* (19).

[35] Xu, F. F., TANG, F. F., SHAO, Y. F., CHEN, Y. L., Chuan, T., & BAO, J. S. (2014). Genotype× environment interactions for agronomic traits of rice revealed by association mapping. *Rice Science*, *21* (3), 133-141.

[36] Kroonenberg, P. M. (1995). Introduction to biplots for G× E tables. Res. Rep. 51. Dep. of Mathematics, Univ. of Queensland, QLD, Australia. *Introduction to biplots for G× E tables. Res. Rep. 51. Dep. of Mathematics, Univ. of Queensland, QLD, Australia.*

[37] Yan, W. (2002). Singular-value partitioning in biplot analysis of multienvironment trial data. *Agronomy Journal*, *94* (5), 990-996.

[38] Gauch, H. G., & Zobel, R. W. (1996). AMMI analysis in yield trials. KANG, MS, GAUCH, HG (Ed) Genotype by environment interaction.