Adaptability and stability of cotton cultivars (Gossypium hirsutum L. race latifolium H.) using factor analytic model

J. J. Nuvunga[1], L. A. Oliveira1[2], C. P. Silva[1], A. K. A. Pamplona[1], A. Q. Silva[2], E. G. Moura[1], M. P. Maleia[3], M. Balestre[1]

[1]Departamento de Estatística, Universidade Federal de Lavras, Lavras, Minas Gerais, Brasil.

[2]Faculdade de Ciências Exatas e Tecnologia, Universidade Federal da Grande Dourados. Dourados, Mato Grosso do Sul, Brasil.

[3]Mozambique Agrarian Research Institute, Maputo, Mozambique.

Corresponding Author: M. Balestre

E-mail: marciobalestre@des.ufla.br

Genet. Mol. Res. 17 (1): gmr16039890

Received January 14, 2018

Accepted February 26, 2018

Published March 02, 2018

DOI http://dx.doi.org/10.4238/gmr16039890

ABSTRACT

To build up efficient strategies in plant breeding programs, it is requested a certain level of knowledge about the genotype–by-environments interaction (GEI) effects over the crop to be improved. One efficient way to gather this information is using linear mixed models using a parsimonious structure of GEI pattern such as factor analytic (FA) structure. In this work, we applied a multivariate analysis using the FA structure on a dataset composed of 11 cotton genotypes (Gossypium hirsutum) which were evaluated in seven environments in Mozambique, to identify stable and wide/specific adapted genotypes. Using the FA structure, it has been possible to select genotypes with specific and broad adaptability for the environmental network. The results indicated that FK37, Flash, BA525 and BA919 were the most productive genotypes, reaching the highest scores for first Factor. Nevertheless, the FK37 genotype presented Factor 2 close to zero classifying it as the most stable followed by Flash cultivar. The genotypes ISA205, QM301 and Albar SZ9314 were considered the less productive, although QM301 presented good stability. These findings suggest that the FK37 genotype might be recommended for Mozambique, since it had good yield and potential stability when compared to the most cultivated varieties, CA324 and Albar SZ9314.

KEY WORDS: Mixed model; Genotype-by-environment interaction; Multi-environmental trials; Multivariate analysis.

INTRODUCTION

The cotton crop (*Gossypium hirsutum*L. race*latifolium*H) is leading in the production of fibres and it is produced commercially worldwide, from temperate to tropical regions (Park et al., 2005; Naveed et al., 2007; Khadi et al., 2010). This crop is fifth in oil production and second in source of protein (Beltrão et al., 2010).

In general, 1 kg of cotton fiber may be obtained starting from 1.65 kg of seed, which contain roughly 21% oil and 23% protein (Wallace et al., 2008; Benzouba et al., 2010). The cotton occupies approximately 34 million hectares distributed across 60 countries around the world, in which Australia (2,000 $Kg.ha^{-1}$), Brazil (1,338 $Kg.ha^{-1}$), China (1,265 $Kg.ha^{-1}$), India (550 $Kg.ha^{-1}$), United States of America (985 $Kg.ha^{-1}$) and Pakistan (599 $Kg.ha^{-1}$) are the most important producers (Fengguo et al., 2007; Beltrão and Azevedo, 2008; Khadi et al., 2010; Acquaah, 2012). The *Gossypium* genus incorporates 50 different species dispersed worldwide and five of them are tetraploid and belong to the subgenus viz. *Karpas*.*Gossypium hirsutum* species is the most cultivated, representing about 96% of cotton world production (Brubaker and Wendel, 1994; Wendel and Cronn, 2003; Fang et al., 2013; Tyagi et al., 2014).

In Mozambique, cotton is one of the most important crops for national economy and contributes about 50 million dollars per year, being classified in the first two most important agricultural exports. Hence, it is a key source of income for more than 300,000 farmers in this country, mainly for small farmers which grades account 90% of the total Mozambique cotton production. In addition, since it is a crop for manufacture, cotton has in its production chain several sectors that employ and/or creating jobs from the farms until the ginning industry (Bias and Donavan, 2003).

One of main challenges of the cotton farmers is the yield variation caused by changes environmental conditions such as related to climate and soils quality that can interfere in crop performance. These factors may cause low heritability or genotypic repeatability which is very common in such quantitative traits. In this scenario, the expected marginal means obtained across several environmental conditions are requested in order to drop out the environmental nuisance factors. Therefore, the evaluation of genotypes in multi-environment trials (METs) network is very important in breeding programs for studying the stability and adaptability of genotypes, as well as its performance prediction in different environments.

In general, GEI may be understood as the phenomenon where the genotypes present different responses across the environments-this scenario may alter the ranking of genotypes in the target environments. Therefore, it is an important issue for breeders and agronomists when the aim is to recommend cultivars for different locations and years or selecting for specific environments or for broad adaptability in mega-environments (Mortazavian et al., 2014). Several methods have been introduced to predict cultivar response in different scenarios (Crossa, 2012; Mortazavian et al., 2014). Among the statistical methods used for MET analysis, we can highlight the linear-bilinear fixed models such us the additive main effects and multiplicative interaction (AMMI) approach and genotype plus GEI (G+GE), the so called GGE biplot (Gauch, 1998; Cornelius and Seyedsadr, 1997; Naveed et al., 2006; Yan et al. 2000). In addition, mixed models approach based on factor analysis or FA structure has been used in MET analysis where the genotypes and GEI are taken as random effects, which genotype and interaction are random effects and environment is fixed (Smith et al., 2015; Stefanova and Buirchell, 2010; Nuvunga et al., 2015). The FA method has been gradually used for analysis in plant breeding (de los Campos and Gianola, 2007; Meyer, 2009; Raman et al., 2011; Zapata-Valenzuela, 2014; Cullis et al., 2014; Smith et al., 2015) and it offers advantages when compared to methods used in the traditional analysis.

Recently, approaches using Bayesian inference for the bilinear models such as AMMI-Bayesian has been applied in MET analysis to overcome some statistical problems related to biplot inferences (Crossa et al., 2011; Josse et al., 2014; Oliveira et al., 2015; Silva et al., 2015). Effective choice of methods for GEI analysis can provide accurate selection of the best varieties for specific or group of environments.

In this study, it was used a factors analysis (FA) to evaluate the response of cotton cultivars (stability and adaptability) in different environments distributed in Mozambique.

MATERIAL AND METHODS

Experimental Details

The experiments were installed in four locations at Namialo (District of Meconta, in Nampula Province), Namara (District of Balama, Cabo Delgado Province), Nhamatanda (District of Nhamatanda, Sofala Province) and Cuamba (District of Cuamba, Niassa Province) during the agricultural seasons 2011/12, 2012/13 and 2013/14, making seven different environments by combining local and years.

In Table 1 relevant characteristics are presented, referring to the different locations that were considered in this analysis.

Table 1. Description agroecologic of locals in evaluation of cotton cultivars productivity in Mozambique.

District	Meconta	Balama	Cuamba	Nhamatanda
Province	Nampula	Cabo Delgado	Niassa	Sofala
Region	North	North	North	Center
Location	38°N43' 9°W8'	13°S20' 38°E33'	34°S18' 36°E32'	19°C15' 34°W14'
Climate	Semi-humid	Semi-arid	Humid tropical	Humid Temp
Soil	Sandy	Alluvium	Loamy	Sedimentary
Precipitation	800-1000mm	1300-1500mm	800-1400mm	849mm
Temperature	24-26°C	20-25°C	24-26°C	17.8-32°C
Altitude	360-500m	200-500m	200-500m	300-1900m

The relationship between test locations and crop years resulted in seven different environments as shown in Table 2. Thus, our multi-environment dataset is composed of 11 varieties of genotypes (Flash, FK37, BA2018, BA320, BA919, BA525, QM301, ALBAR SZ9314, CA 324 and ISA 205) available in these environments.

Table 2. Environments code combining locals and years.

Local	Year	Environment
Cuamba	2011/12	E1
Nhamatanda	2011/12	E2
Cuamba	2012/13	E3
Nhamatanda	2011/12	E4
Meconta	2011/12	E5
Balama	2011/12	E6
Cuamba	2013/14	E7

The experimental design used in each location was a randomized complete block design with four replications. The plots consisted of five lines of 5.0 m long, with the two borders lines with a spacing of 0.70 m between rows and 0.20 cm between plants. The seeds were placed manually, being 4-10seeds per hole with approximately 4 cm deep. The first thinning was done 15 days after emergence leaving two plants per hole and the second thinning was done21 days after emergence leaving a plant per hole. Weeds were controlled manually when needed. The first pest control was done by spraying acetamiprid insecticide (222 $g.lt^{-1}$) in a dosage of 50 $ml.ha^{-1}$; then was sprayed Lambda-cyhalothrin (60 $g.L^{-1}$) fortnightly from the fourth week after emergence, at a dose of 250 $ml.ha^{-1}$. Insecticides were applied with micro-ulva (ULV). In the experiment were evaluated the cotton seed yield.

Linear Mixed Model

A general linear mixed model was used for the first-step analysis in the data from all seven environments. Considering *p* environments, *b* blocks and *m* genotypes, this dataset was subjected to a joint analysis:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Zu} + \mathbf{e} \qquad \text{(Equation 1)}$$

in which $\mathbf{y}(n \times 1)$ denotes the combined vector of data across environments, $\boldsymbol{\beta}(pb \times 1)$ are the fixed effects vector (blocks and environments) and $\mathbf{u}(mp \times 1)$ vector of genotypic effects within locations, $\mathbf{X}(n \times pb)$ and $\mathbf{Z}(n \times mp)$ are the associated design matrices, and $\mathbf{e}(n \times 1)$ denotes the random error vector (*m* is number of genotypes, *b* is number of environments and *p* is number of blocks). It was assumed that $\mathbf{e} \sim N(\mathbf{0}, \mathbf{R})$ and $\mathbf{u} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$, in which $\mathbf{R}$ is the diagonal residual matrix and $\boldsymbol{\Sigma}$ is the variance-covariance matrix of genetic effects (Nuvunga et al., 2015).

The Factor Analytic Structure

The genetic variance ($\boldsymbol{\Sigma}$) can be modelled through the FA structure, that considers the random effects of *m* genotypes in *p* environments as a linear function of latent variables $\mathbf{f_k}(m \times 1)$ with coefficients $\boldsymbol{\lambda}_k(p \times 1)$ known as loadings, for $k = 1, \ldots, t < p$, and a specific variance vector $\boldsymbol{\delta}$ $(mp \times 1)$. Then

$\boldsymbol{\Sigma} = (\boldsymbol{\Lambda}\boldsymbol{\Lambda}^{\cdot} + \boldsymbol{\Psi}) \otimes \mathbf{I}_m = \mathbf{FA}(k) \otimes \mathbf{I}_m$ in which $\boldsymbol{\Lambda} = \begin{pmatrix} \lambda_{11} & \cdots & \lambda_{1t} \\ \vdots & \ddots & \vdots \\ \lambda_{p1} & \cdots & \lambda_{pt} \end{pmatrix}$ is a matrix ($p \times t$) with the column containing the p environments delivery to the k^{th} latent factor and $\boldsymbol{\Psi}$ is the diagonal matrix of specific variances for the p environments.

The adopted approach (FA-SREG) was described by Meyer et al. (2009) and Nuvunga et al. (2015) based on the FA structure proposed by Smith et al. (2001, 2005), in which the genotypes effects are confounded with GEI, such as the linear-bilinear model of genotypes main effects and interaction GE (Burgueño et al., 2008; Stefanova and Buirchell, 2010). The factors analysis in the restricted maximum likelihood can be easily implemented with the covariance matrix $\boldsymbol{\Sigma}$ and diagonal $\mathbf{R}$ from step one since the unstructured G and diagonal V present very low dimensions. In this situation, the FA can be used as GEI pattern recognition method instead as dimensionality reduction one. It was assumed that $\mathbf{G}$ can be represented by an FA structure, such as $\boldsymbol{\Lambda}\boldsymbol{\Lambda}^{\cdot} + \boldsymbol{\Psi}$, and the BLUPs can be represented by common factors as $\mathbf{u} = \mathbf{Lf} + \boldsymbol{\delta}$, in which $\mathbf{L} = \boldsymbol{\Lambda} \otimes \mathbf{I}_m$. By implementing this transformation in linear mixed model (1) we have that:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}[\mathbf{Lf} + \boldsymbol{\delta}] + \mathbf{e} \qquad \text{(Equation 2)}$$

in which $\mathbf{f}$ ($mt \times 1$) is the factorial scores vector of missing information (BLUP's), $\boldsymbol{\delta}$ ($mp \times 1$) is a specific variance vector, $\mathbf{L}$ ($mp \times mt$) is the matrix of factorial loadings, and $\mathbf{X}$ and $\mathbf{Z}$ are the design matrix. Furthermore, it is assumed that $\mathbf{f} \sim N(\mathbf{0}, \mathbf{I})$, $\boldsymbol{\delta} \sim N(\mathbf{0}, \boldsymbol{\Psi})$ and $\mathbf{e} \sim N(\mathbf{0}, \mathbf{R})$.

Therefore, the reparametrized matrix solution to the mixed model equations, in which $\mathbf{W} = \mathbf{ZL}$, can be given as follows:

$$\begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{f}} \\ \hat{\boldsymbol{\delta}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}^{\cdot}\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}^{\cdot}\mathbf{R}^{-1}\mathbf{W} & \mathbf{X}^{\cdot}\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{W}^{\cdot}\mathbf{R}^{-1}\mathbf{X} & \mathbf{W}^{\cdot}\mathbf{R}^{-1}\mathbf{W} + \mathbf{I} & \mathbf{W}^{\cdot}\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}^{\cdot}\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}^{\cdot}\mathbf{R}^{-1}\mathbf{W} & \mathbf{Z}^{\cdot}\mathbf{R}^{-1}\mathbf{Z} + \boldsymbol{\Psi}^{-1} \otimes \mathbf{I}_m \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{X}^{\cdot}\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{W}^{\cdot}\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}^{\cdot}\mathbf{R}^{-1}\mathbf{y} \end{pmatrix} \qquad \text{(Equation 3)}$$

Meyer (2009) and Nuvunga et al. (2015) reported the solutions for the fixed and random effects to be similar as follow:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^{\cdot}\mathbf{R}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\cdot}\mathbf{R}^{-1}(\mathbf{y} - \mathbf{Wf} - \mathbf{Z}\boldsymbol{\delta}) \qquad \text{(Equation 4)}$$

$$\hat{\mathbf{f}} = (\mathbf{W}^{\cdot}\mathbf{R}^{-1}\mathbf{W} + \mathbf{I})^{-1}\mathbf{W}^{\cdot}\mathbf{R}^{-1}(\mathbf{y} - \mathbf{Xb} - \mathbf{Z}\boldsymbol{\delta}) \qquad \text{(Equation 5)}$$

$$\hat{\boldsymbol{\delta}} = (\mathbf{Z}^{\cdot}\mathbf{R}^{-1}\mathbf{Z} + \boldsymbol{\Psi}^{-1} \otimes \mathbf{I})^{-1}\mathbf{Z}^{\cdot}\mathbf{R}^{-1}(\mathbf{y} - \mathbf{Xb} - \mathbf{Wf}) \qquad \text{(Equation 6)}$$

Loadings rotation

When $k > 1$, the loadings matrix ($\boldsymbol{\Lambda}$) is not unique. This non-uniqueness requires some restriction through the fitting of the FA(k) models because the variance models are termed as non-identifiable (Smith et al., 2002; Cullis et al., 2010).

To ensure the identifiability in **Λ**, one set of constraints must be imposed on the factor analytic parameters i.e. **Λ** must be a lower triangular matrix (Smith et al., 2001; Meyer, 2009; Cullis et al., 2010). This is purely for computational simplicity and these restrictions rarely have any basis or biological interpretation. In the context of MET data, a representation of the main components of the loadings (and thus the scores) is usually more significant (Cullis et al., 2010).

After fitting the model, the required rotation is as follows:

$$\mathbf{\Lambda}^{*} = \mathbf{\Lambda B} \qquad \text{(Equation 7)}$$

in which **B** is the orthogonal matrix, obtained from the singular value decomposition of **Λ**, whose columns are the eigenvectors of $\mathbf{\Lambda}^{\cdot}\,\mathbf{\Lambda}$, More details can be found in Meyer (2009) and Cullis et al. (2010).

Then, $\mathbf{\Lambda}^{*} = \left[\boldsymbol{\lambda}_1^{*}\ \boldsymbol{\lambda}_2^{*}\ \ldots\ \boldsymbol{\lambda}_k^{*}\ \right]$ means that the first vector (latent variable) accounts for the largest covariance in the FA(*k*) model, the second vector is orthogonal to the first and it accounts for the next largest amount and so on.

In the following, we assume that the loadings correspond to the rotated versions.

RESULTS

The lower triangular unstructured (UN) genotypic (co)variance matrix and diagonal one for residual are shown in Table 3. Based on the Hartley test, there is enough evidence to reject the hypothesis of homogeneity for genetic variance and environmental variance.

It suggests the existence of genetic and residual heterogeneity of variances.

Table 3. Genotypic and residual (in red) (co)variance matrix for cotton dataset evaluated in of cotton seed yield at seven different sites.

E nv.*	E1	E2	E3	E4	E5	E6	E7
1	0.02$^{(0.17)}$						
2	0.03	0.11$^{(0.27)}$					
3	0.01	0.01	0.03$^{(0.17)}$				
4	0.01	0.03	0.00	0.02$^{(0.05)}$			
5	0.00	0.00	0.00	0.00	0.01$^{(0.29)}$		
6	0.01	0.01	0.01	0.00	0.00	0.01$^{(0.11)}$	
7	0.00	-0.02	0.00	-0.01	0.00	0.00	0.02$^{(0.04)}$

*Env. = Environment

It is notably also the presence null co-variances among environments (these co-variances represent the genotypic variance plus the variance of interaction between local peers).

The pairs of environments with zero covariance can be interpreted as those that present non-correlated genotypic responses contributing largely for GEI.

The residual variance was greater for environments 2 and 5, suggesting the influence of those environments in the magnitude of these variances (Table 3).

The FA-SREG2 model explained the 90% of overall genotypic variation. In the Table 4 are the factor loadings rotated by Varimax method. It is observed that the first factor explained almost 80% of the total variation and the second one, 10%. The specific variances were low for all considered environments. The high values of common variance show that the two factors explained a large percentage of each environment variance.

Table 4. Estimated loadings (correlation scale) for the balanced data fit using the FA-SREG2 model.

Environment	Factor 1	Factor 2	Common variance	Specific variance
E1	0.0801	-0.0781	0.99	0.01
E2	0.3280	0.0042	1.00	0.00
E3	0.0291	-0.1078	0.99	0.01
E4	0.0714	0.0332	0.99	0.01
E5	0.0015	0.0213	0.99	0.01
E6	0.0611	-0.0501	0.99	0.01
E7	-0.0372	-0.0449	0.99	0.01
Variance Total	80%	10%		

The Figure 1 shows the factor scores graph of the adjusted FA-SREG2 model. Similarly, to GGEbiplot, the first axis (Factor 1) provides information about the adaptability (the linear correlation of scores related to factor 1 with the marginal BLUPS was 0.90) and the second axis (Factor 2) describes the contribution of the environment or genotype for GEI.

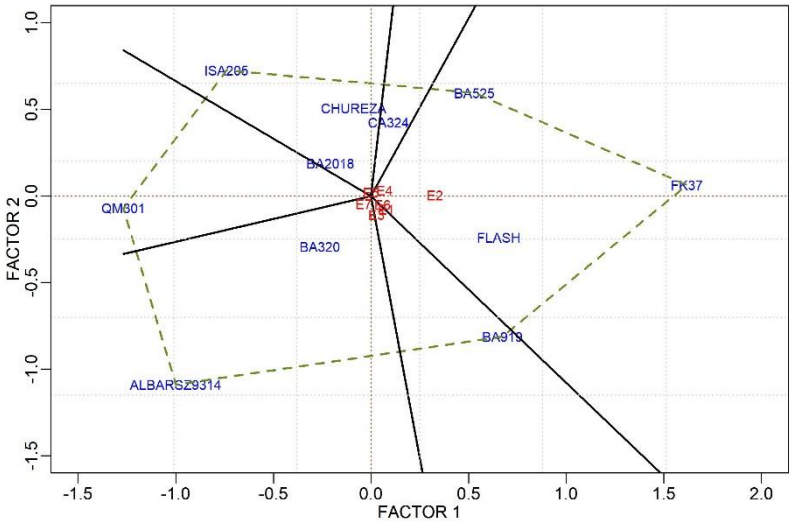

Figure 2. Biplot representation for the first two scores and factorial loads separately for yields of cottonseed (t / ha). Figure 2a - biplot of the scores (genotypes). Figure 2b-factorial loads (environments).

Genotypes whose scores have values greater than zero for Factor 1 are the most productive, while scores near the biplot origin for Factor 2 indicate genotypes that contribute poorly or do not contribute to the GEI; in other words, they are not explained by the environmental loadings and therefore stable. In this scenario, the FK37 genotype was considered the most adapted and stable and it can be widely recommended for all sites in this study. The genotypes that are most distant from the biplot origin for the second score contributed more for the interaction (or specific response) i.e.: Albar SZ9314, BA919, BA525, ISA205.

Figure 2 shows the biplots related to the scores and factorial loadings. In this figure, it is observed four environmental subsets {4, 5}, {1, 3, 6}, {7} and {2}, with respect to the GEI pattern (Figure 2b).

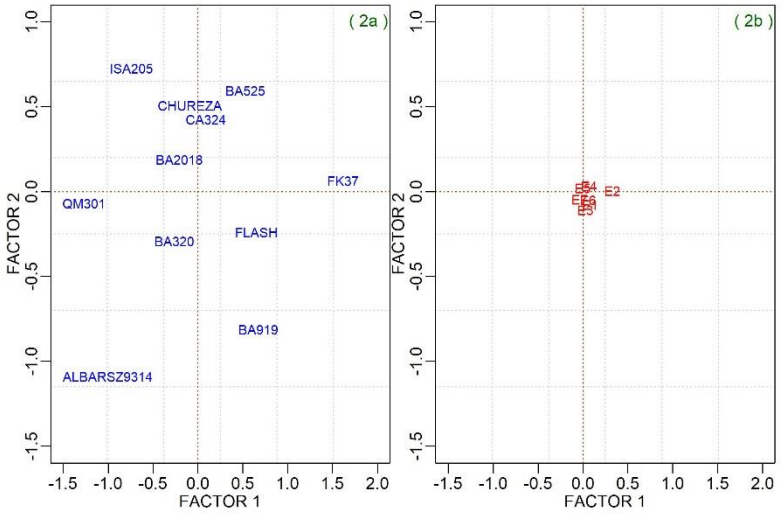

Figure 1. Biplot representation for the first two scores and factorial loads (simultaneous representation) for yields of cotton seed (t / ha), considering the FA-SREG2 model.

The joint response of genotypes and environments can be examined when in Figure 1. The simultaneous representation of genotypes and environments facilitates interpretation; however, it should be noted that environmental and genotypic scores are not on the same scale. Genotypic and environmental scores located in the same quadrants represent specific positive combinations or adaptability between genotypes and environments.

DISCUSSION

Regarding to the application and interpretation of FA models, there are two issues of primary interest: the estimate of the overall performance of genotypes in all environments and the evaluation of the genotypes stability and adaptability (Stefanova and Buirchell, 2010).

The FA-SREG model can be understood as the linear regression of genotype and GEI in latent environmental covariates (environmental loading up) in which each genotype has a separate slope (genotypic scores, **f**), but a common intercept (main effects of genotypes are indistinguishable from GEI). The genotypes slopes measure the sensitivity of genotypes to hypothetical environmental factors represented by loading of each environment (Burgueno et al., 2008).

In the FA biplot the first score may represent the general adaptability whether it is correlated with the marginal BLUPs, while the second one may be interpreted as those related to GEI, that is, it evaluates the genotypic stability. Thus, positive values for the first axis correspond to genotypes with high yield and genotypes presenting the second axis values near to biplot origin are those presenting good stability.

Results in Table 3 show the existence of heterogeneity of genotypic and residual variances. A direct implication of this fact would be that the use of analysis of variance (ANOVA)based on least square for GEI and multivariate methods such as AMMI and GGEbiplot (Gauch et al., 2006; Yan et al., 2000; Yan et al., 2007), would not be recommended due the simplification of GEI structure. This claim can also be extended to mixed models or Bayesian methods that simplify the GEI structure using the homogeneity of variance assumption.

Maleia et al. (2010), for instance, evaluated the stability and adaptability of cotton cultivars in Mozambique employing methodologies of Annicchiarico (1992) and of Burrows and Toler (1998). Exploiting the Toler and Burrows (1998) methodology, they identified cotton cultivars presenting specific adaptability for high and low quality environments and others showing wide adaptability for the trail network. In other hand, using the Annicchiarico method, they found cultivars that presented high phenotypic stability.

The genotypes evaluated here were analyzed by Maleia et al. (2017) using fixed AMMI methods found the FK 37, BA 919 e Flash as the most productive genotype and they identified BA2018 e BA320 as more stable genotypes. Although the AMMI method is recognized for allowing the simultaneous study of stability and adaptability, in a single approach, separating pattern of noise, it has the inherent limitations of fixed effects models, such as difficulty in dealing with heteroscedastic data set.

Moiana et al. (2014), on the other hand, using REML/BLUP methodologies (mixed models) to evaluate adaptability and harmonic means to evaluate stability founds genotypes presenting stability and adaptability for Mozambique. Interestingly, the methodology used by these authors is for ranking of the predictions of genotypes and do not evaluate separately the stability and adaptability since the harmonic means of genotypic values is correlated with marginal BLUPs.

However, these methodologies assume the homogeneity of variances between the tested environments which it not requested in mixed models with FA structure. Burgueño et al. (2008) and Cullis et al. (2010) recognized the importance of separating the two GEI causes: the heterogeneity of genetic variances between environments (i.e., the interaction due to scale) and heterogeneity of correlation between environments (Table 3). The last one is often considered more important to MET in plant breeding, since it has impact on the genotypic classification and the selection. The method used in this study allows detailed analysis in this direction, as well as several advantages when compared to traditional methods of analysis as the AMMI fixed effects model and even conventional mixed models such as those used in Maleia et al. (2010), Moiana et al. (2014) and Naveed et al. (2006).

The model used in this study was explained and interpreted the relation of FA-SREG2 and SREG2 models (Burgueno et al 2008. The biplot interpretation of the of FA models is very similar to the GGEbiplot model presented in Yan et al. (2007), CROSSA et al. (2010) and Stefanova and Buirchell (2010). In GGEBiplot, genotypes are evaluated for adaptability from the first principal component scores and in the FA-SREG model this relationship can be obtained through the first factorial score because negative loadings of environments are not observed. In addition, and there was a high correlation between the first factor scores for genotypes and its marginal BLUPs (0.90). The stability of a genotype on GGE analysis, in turn, can be described through the second principal component. A similar interpretation can be given to factor score 2 in the FA-SREG analysis. Smith et al. (2002), in other hand) explained the relation between the FA-AMMI2 and AMMI2 models when the genotype are marginalized from GEI matrix.

In this study, from 11 evaluated genotypes, four are checks- CA324, Albar SZ9314, Chureza and ISA205– being the first two the most grown in Mozambique and the ISA205 is no longer produced. The remaining seven genotypes are genotypes candidates and, therefore, deserve special attention in the trials. The results in Figure 1 of the factorial scores for the FA-SREG2 model shows that the genotypes FK37, Flash, BA525 and BA919 are the most productive reaching the highest scores for Factor 1. Nevertheless, the FK37 genotype with Factor 2 close to zero was the most stable followed by Flash genotype. The genotypes ISA 205, QM301 and Albar SZ9314 are the less productive, although QM301 presents good stability.

As already mentioned, Maleia et al. (2017) analyzed these same genotypes from the conventional AMMI model. In the analysis of adaptability, the results agree with those obtained from the FA analysis. However, with regard to the interaction study, the results diverged between the two approaches. This discrepancy may be partly explained by the fact that we analyzed the MET set in seven environments, while Maleia et al. (2017) evaluated only six of them. In addition, we assume different variances between the test environments, while the AMMI approach has the homogeneity of variances as fundamental assumption.

CONCLUSION

These results suggest that the FK37 genotype should be recommended for Mozambique, since has a good yield potential and good stability when compared with CA324 and Albar SZ 9314. The FA-SREG model permitted selecting genotypes with specific and broad adaptation in Mozambique.

REFERENCES

Acquaah G (2012). Principles of Plant Genetics and Breeding, (2nd edn), Wiley-Blackwell, Oxford, 758 p.

Annicchiarico P (1992). Cultivar adaptation and recommendation from alfalfa trials in Northern Italy. *Journal of Genetics and Breeding*, 46: 269-278.

Beltrão NEM, Azevedo DMP (2008). O agronegócio do algodão no Brasil. 2ND edI. rev. ampl. Brasília: EMBRAPA, 13 p.

Beltrão NEM, Vale LS, Marques LF, Cardoso GD, et al. (2010). O cultivo do algodão orgânico no semiárido Brasileiro. *Revista Verde* 5: 8-13.

Benbouza, H, Lacape JM, Jacquemin JM, Courtois B, (2010). Introgression of the low gossypol seed and high-gossypol plant trait in upland cotton: analysis of [(Gossypium hirsutum L. X G. raimondii)^2 x G. sturtianum] trispecific hybrid and selected derivatives using mapped SSRs. *Mol. Breed*. 25: 273-286. https://doi.org/10.1007/s11032-009-9331-6

Bias C, Donovan C (2003). Gaps and opportunities for agricultural sector development in Mozambique. *Research Report* No 54E, Ministry of Agriculture and Rural Development, Mozambique.

Brubaker CL, Wendel JF (1994). Re-evaluating the origin of domesticated cotton (Gossypium hirsutum L.; Malvaceae) using nuclear restriction fragment length polymorphism (RFLPs). *American J. of Botany*. 81: 1309-1326. https://doi.org/10.2307/2445407

Burgueño J, Crossa J, Cornelius PL, Yang RC (2008). Using factor analytic models for joining environments and genotypes without crossover genotype × environment interaction. *Crop Sci*. 48: 1291-1305. https://doi.org/10.2135/cropsci2007.11.0632

Cornelius PL, Seyedsadr MS (1997). Estimation of general linear-bilinear models for two-way tables. *Journal of Statistical Computation and Simulation* 58: 287-322. https://doi.org/10.1080/00949659708811837

Crossa J (2012). From genotype × environment interaction to gene × environment interaction. *Current Genomics* 13: 225-244. https://doi.org/10.2174/138920212800543066

Crossa J, de Los Campos G, Pérez P, Gianola D, et al. (2010). Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics* 186: 713-724. https://doi.org/10.1534/genetics.110.118521

Crossa J, Perez-Elizalde S, Jarquin D, Cotes JM, et al. (2011). Bayesian Estimation of the Additive Main Effects and Multiplicative Interaction Model. *Crop Sci*. 51: 1458-1469. https://doi.org/10.2135/cropsci2010.06.0343

Cullis BR, Smith AB, Beeck C, Cowling WA (2010). Analysis of yield and oil from a series of canola breeding trials. Part II: Exploring variety by environment interaction using factor analysis. *Genome* 53: 1002-1016. https://doi.org/10.1139/g10-080

Cullis BR, Jefferson P, Thompson R, Smith AB (2014). Factor analytic and reduced animal models for the investigation of additive genotype-by-environment interaction in outcrossing plant species with application to a Pinus radiata breeding programme. *Theor Appl Genet.* 127: 2193-2210. https://doi.org/10.1007/s00122-014-2373-0

Da Silva CP, de Oliveira LA, Nuvunga JJ, Pamplona AKA (2015). A Bayesian Shrinkage Approach for AMMI Models. *PLoSOne* 10:e0131414. https://doi.org/10.1371/journal.pone.0131414

De los Campos G, Gianola D (2007). Factor analysis models for structuring covariance matrices of additive genetic effects: a Bayesian implementation. *Genetics Selection Evolution* 39: 481-494. https://doi.org/10.1051/gse:20070016

De Oliveira LA, da Silva CP, Nuvunga JJ, da Silva, et al. (2015). Credible intervals for scores in the AMMI with random effects for genotype. *Crop Sci*. 55: 465-476. https://doi.org/10.2135/cropsci2014.05.0369

Fang L, Choi SH, Baek JS, Liu C, et al. (2013). Control of angiogenesis by AIBP-mediated cholesterol efflux. *Nature*. 498: 118-122. https://doi.org/10.1038/nature12166

Fengguo S, Xiangyun Z, Junlan L, Baosheng G, et al. (2007). Training course on cotton breeding and management technology. *Shijiazhuan*. 231.

Gallant AR (1987) Non-linear statistics models. New York: *John Wiley*. 610.

Gauch HG (2006) Statistical analysis of yield trials by AMMI and GGE. *Crop Sci.* 46: 1488-1500.

Iam. Instituto de Algodão de Moçambique. 2015. Variedades de sementes do algodão. http://www.iam.gov.mz/index.php?option=com_content&view=article&id=102&Itemid=83.

Josse J, van Eeuwijk FA, Piepho HP, Denis JB (2014). Another look at Bayesian analysis of AMMI models for genotype-environment data. *J. Agric. Biol. Environ. Stat.* 19: 240-257. https://doi.org/10.1007/s13253-014-0168-z

Khadi BM, Santhy V, Yadav MS (2010). Cotton: an introduction. Biotechnol. *Agric. For*. 65:1-14.

Maleia MP, Vidigal Filho PS, Kvitschal MV, Gonçalves-Vidigal MC (2010). Stability and adaptability of commercial cotton cultivars (Gossypium hirsutum L. race latifolium H.*)* in Mozambique. *Afr. J. Agric. Res.* 5: 539-550. https://doi.org/10.21475/ajcs.17.11.04.pne60

Meyer K (2009). Factor-analytic models for genotype x environment type problems and structured covariance matrices. *Genet Sel Evol*. 41: 21. https://doi.org/10.1186/1297-9686-41-21

Moiana LD, Vidigal Filho PD, Gonçalves-Vidigal MC, Maleia MP (2014). Application of mixed models for the assessment genotype and environment interactions in cotton (Gossypium hirsutum) cultivars in Mozambique. *Afr. J. Biotechnology* 13:1985-1991. https://doi.org/10.5897/ajb2013.12926

Mortazavian, SMM, Nikkhah HR, Hassani FA, Sharif-al-Hosseini M, M. (2014). GGE biplot and AMMI analysis of yield performance of barley genotypes across different environments in Iran. *Journal of agricultural science and Technology*.16:609-622,

Naveed M, Nadeem M, Islam M. (2007). AMMI analysis of some upland cotton genotypes for yields stability in different milieus. *World J. Agric. Sci.* 3: 39-43.

Naveed M, Mukhtar N, Farooq J, Ilyas M, Islam NU (2006). Evaluation of some new strains of Gossypium hirsutum L. for yield stability across environments. *J. Agric. Soc. Sci.* 2: 17-19.

Nuvunga JJ, de Oliveira LA, Pamplona AKA, da Silva CP (2015). Factor analysis using mixed models of multi-environment trials with different levels of unbalancing. *Genet Mol Res*. 14: 14262-14278.

Park YH, Alabady MS, Ulloa M, Sickler B (2005). Genetic mapping of new cotton fiber loci using EST-derived microsatellites in an interspecific recombinant inbred line cotton population. *Molecular Genetics and Genomics*. 274: 428-431. https://doi.org/10.1007/s00438-005-0037-0

Raman A, Ladha JK, Kumar V, Sharma S, 2011. Stability analysis of farmer participatory trials for conservation agriculture using mixed models. Field Crops Research 121:450-459. https://doi.org/10.1016/j.fcr.2011.02.001

Smith AB, Cullis BR, Thompson R (2001). Analysing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. *Biometrics*. 57: 1138-1147. https://doi.org/10.1111/j.0006-341x.2001.01138.x

Smith AB, Cullis BR, Luckett D, Hollamby G, et al. (2002). Exploring variety-environment data using random effects AMMI models with adjustments for spatial field trend. Part II: Applications. In Quantitative Genetics, Genomics and Plant Breeding (Ed. M. Kang). 337-352.https://doi.org/10.1079/9780851996011.0337

Smith AB, Cullis BR, Thompson R (2005). The analysis of crop cultivar breeding and evaluation trials: An overview of current mixed model approaches. *J. Agric. Science* 143: 1-14.

Smith AB, Ganesalingam A, Kuchel H, Cullis BR (2015). Factor analytic mixed models for the provision of grower information from national crop variety testing programs. *Theor Appl. Genet.* 128: 55-72.

Stefanova K, Buirchell B (2010). Multiplicative Mixed Models for Genetic Gain Assessment in Lupin Breeding. *Crop Sci*. 50:880-891.

Tyagi P, Gore MA, Bowman DT, Campbell BT, et al. (2014). Genetic diversity and population structure in the US Upland cotton (Gossypium hirsutum l.). *Theoretical and Applied Genetics*. 127: 283-295.

Toler JE, Burrows, PM (1998). Genotypic performance over environmental arrays: a non-linear grouping protocol. *Journal of Applied Statistics* 25: 131-143. https://doi.org/10.1080/02664769823368

Wallace TP, Bowman D, Campbell BT, Chee P, et al. (2008). Status of the USA cotton germplasm collection and crop vulnerability. *Genet. Res. Crop Evol*. 56: 507-532. https://doi.org/10.1007/s10722-008-9382-2

Wendel JF, Cronn RC (2003). *Polyploidy and the evolutionary history of cotton. In: Advances in Agronomy* 78: 139-186.

Yan W, Hunt LA, Sheng Q, Szlavnics Z (2000). Cultivar evaluation and mega-environment investigation based on the GGE biplot. *Crop Sci.* 40:597-605. https://doi.org/10.2135/cropsci2000.403597x

Yan W, Kang MS, Ma B, Woods S, et al. (2007). GGE biplot vs. AMMI analysis of genotype-by-environment data. *Crop Sci.* 47:643-655.

Zapata-Valenzuela J (2012). Uso de La Estructura de Factor Analítico para Aumentar La Heredabilidad de Ensayos Clonales de Pinus taeda L. *Chilean Journal of Agricultural Research*. 72:309-315.