Field Crops Research 257 (2020) 107929

Contents lists available at ScienceDirect

Field Crops Research

journal homepage: www.elsevier.com/locate/fcr

# Genotype-by-environment interaction and yield stability analysis of biomass sorghum hybrids using factor analytic models and environmental covariates

Isadora Cristina Martins Oliveira[a], José Henrique Soler Guilhen[a], Pedro César de Oliveira Ribeiro[b], Salvador Alejandro Gezan[c], Robert Eugene Schaffert[a], Maria Lucia Ferreira Simeone[a], Cynthia Maria Borges Damasceno[a], José Eustáquio de Souza Carneiro[d], Pedro Crescêncio Souza Carneiro[b], Rafael Augusto da Costa Parrella[a,*], Maria Marta Pastina[a,*]

[a] Embrapa Milho e Sorgo, 35701-970, Rod. MG 424, km 65, PO Box 151, Sete Lagoas, MG, Brazil
[b] Department of Biology, Federal University of Viçosa, 36570-000, Viçosa, MG, Brazil
[c] VSN International, Hemel Hempstead, United Kingdom
[d] Department of Plant Science, Federal University of Viçosa, 36570-000, Viçosa, MG, Brazil

ARTICLE INFO

*Keywords:*
*Sorghum bicolor* (L.) Moench
plant breeding
bioenergy
multi-environment-trials
mixed models

ABSTRACT

Biomass sorghum has emerged as an alternative crop for biofuel and bioelectricity production. Fresh biomass yield (FBY) is a quantitative trait highly correlated with the calorific power of energy sorghum cultivars, but also highly affected by the environment. The main goal of this study was to investigate the genotype-by-environment interaction (G × E) and the stability of sorghum hybrids evaluated for FBY across different locations and years, using factor analytic (FA) mixed models and environmental covariates. Pairwise genetic correlations between environments ranged from -0.21 to 0.99, indicating the existence of null to high G × E. The FA analysis unveiled that solely three factors explained more than 79% of the genetic variance, and that more than 60% of the environments were clustered in the first factor. Moderate correlations were found between some environmental covariates and the loadings of FA models for environments, suggesting the possible factors to explain the high G × E between environments clustered in a given factor. For example: precipitation, minimum temperature and speed wind were correlated to the environmental loadings of factor 1; minimum temperature, solar radiation and altitude to factor 2; and crop growth cycle to factor 3. The latent regression analysis was used to identify hybrids more responsive to a set of environments, as well as hybrids specifically adapted to a given environment. Finally, FA models can be successfully used to identify the main environmental factors affecting G × E, such as minimum temperature, precipitation, solar radiation, crop growth cycle and altitude.

## 1. Introduction

Biomass sorghum [*Sorghum bicolor* (L.) Moench] can be used as an alternative feedstock for bioelectricity (Demirbas et al., 2009), biogas (Mahmood et al., 2015) and second-generation ethanol production (Reis et al., 2016). For these purposes, one of the most important traits is dry biomass yield (DBY) (de Oliveira et al., 2018), which is highly correlated with the calorific power of sorghum cultivars and, consequently, with the potential to generate energy and heat during the burning process. Because of the large number of required processes for measuring DBY, fresh biomass yield (FBY) can be used as a proxy to perform indirect cultivar selections, due to the reported high genetic correlation between FBY and DBY (de Almeida et al., 2019). Besides the high FBY, other attributes are also attractive for using biomass sorghum as an energy crop in Brazil, such as drought and heat tolerance, water use efficiency, short crop growth cycle (160 to 180 days) (Brenton

*Abbreviations:* CV, coefficient of variation; FA, factor analytic; G × E, genotype × environment interaction; MET, multi-environment trials; FBY, fresh biomass yield; DBY, dry biomass yield; UN, unstructured; VCOV, variance-covariance matrix; VCU, value for cultivation and use

* Corresponding authors.
*E-mail addresses:* isadoracmo90@gmail.com (I.C.M. Oliveira), jhguilhen@gmail.com (J.H.S. Guilhen), pedroc.ribeiro14@gmail.com (P.C.d.O. Ribeiro), salvador.gezan@vsni.co.uk (S.A. Gezan), robert.schaffert@embrapa.br (R.E. Schaffert), marialucia.simeone@embrapa.br (M.L.F. Simeone), cynthia.damasceno@embrapa.br (C.M.B. Damasceno), jesc@ufv.br (J.E.d.S. Carneiro), carneiro@ufv.br (P.C.S. Carneiro), rafael.parrella@embrapa.br (R.A.d.C. Parrella), marta.pastina@embrapa.br (M.M. Pastina).

https://doi.org/10.1016/j.fcr.2020.107929
Received 16 June 2020; Received in revised form 3 August 2020; Accepted 9 August 2020

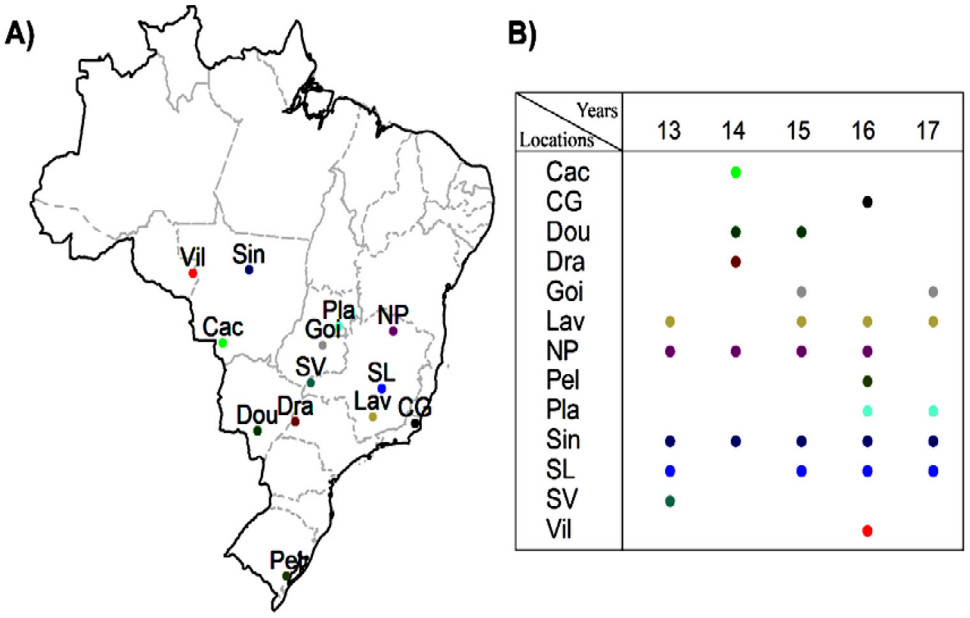


**Fig. 1.** (A) Geographic positions of the locations in which VCU trials were conducted; and (B) Locations evaluated in each year, from 2013 to 2017. Years were coded as 13 to 17. Locations were coded as: Cac, Cáceres-MT; CG, Campos dos Goytacazes-RJ; Dou, Dourados-MS; Dra, Dracena-SP; Goi, Goiânia-GO; Lav, Lavras-MG; NP, Nova Porteirinha-MG; Pel, Pelotas-RS; Pla, Planaltina-DF; SV, Santa Vitória-MG; SL, Sete Lagoas-MG; Sin, Sinop-MT; and Vil, Vilhena-RO.

et al., 2016; Parrella et al., 2011), and high adaptability to tropical and subtropical soil and climate conditions.

Annually, the Embrapa's sorghum breeding program performs late-stage trials, denominated value for cultivation and use (VCU) trials, in which the best biomass hybrids, previously selected in preliminary and/or late-stage breeding trials, are evaluated across different locations from distinct Brazilian regions. In these VCU trials, some commercial cultivars and/or parental lines are included as common checks, since different sets of hybrids are often tested across years and/or locations. The set of connected trials across multiple years and locations are called multi-environment trials (MET). The joint analysis of MET data can provide useful information about genotype-by-environment interaction (G × E), stability and adaptability of hybrids across distinct environments. This information is critical for the release of new cultivars showing yield stability across a set of environments, or that are specifically adapted to a given environment (Bornhofen et al., 2018; Burgueño et al., 2008; Dias et al., 2018b).

FBY is a quantitative trait highly affected by the environment. Therefore, the lack of information about G × E can lead to a reduction in the genetic gains, reinforcing the usefulness of MET studies (Quintero et al., 2018). Although late-stage trials are often balanced between locations within a harvest year, they are highly unbalanced between years, since low-performance hybrids are usually replaced by newly developed elite materials through the seasons of a breeding program. Another reason of unbalance in MET data sets is the occurrence of missing plots, due to biotic or abiotic stresses during the crop growth cycle. For these reasons, joint analysis of unbalanced experiments cannot be performed with the traditional statistical methods used to study G × E, adaptability and stability of hybrids across environments. In MET datasets, these studies require the use of more flexible approaches, such as linear mixed models (Kelly et al., 2007; Resende and Thompson, 2004; Smith et al., 2005).

Within the linear mixed models' framework, different variance and covariance (VCOV) structures can be considered for random effects, allowing for complex modeling of genetic correlations across environments, in addition to heterogeneity of variances between environments. The unstructured (UN) is the most complex VCOV structure, which considers heterogeneity of variances across environments and all possible specific covariances between pairs of environments. However, due to the high number of parameters (i.e. variance-covariance components) to estimate when several environments are analyzed simultaneously, fitting an UN structures is often a complex task (Smith et al., 2005), which may lead to unstable model fits, with high standard errors for the variance components, and/or frequently to a lack of model convergence.

Multiplicative factor analytic (FA) structures have been proposed as a more parsimonious approach, allowing the estimation of a smaller number of parameters compared to UN structures (Kelly et al., 2007; Piepho, 1997, 1998; Smith et al., 2001b). In addition, graphical tools, such as latent regression plots (Smith et al., 2015; Thompson et al., 2003) and heatmaps of estimated genetic correlation matrices across environments (Cullis et al., 2014; Smith et al., 2015), can be explored to infer about G × E, adaptability and stability of hybrids. In addition, the environmental loadings of FA models can also be correlated to environmental covariates, such as geographic, soil and climatic covariates, in order to observe trends in the genotype performance across environments according to environmental conditions (Sae-Lim et al., 2014). In this context, the main goals of this study were: i) to understand G × E interaction, and adaptability and stability of Embrapa's biomass sorghum hybrids across the measured environments; ii) to select stable hybrids across a set of environments or specifically adapted hybrids to a given environment; and iii) to identify the main environmental covariates that explain G × E interaction.

## 2. Material and Methods

### 2.1. Genetic Material and Experimental Design

A total of 55 sorghum genotypes were evaluated, corresponding to 51 experimental hybrids and four checks, i.e., two commercial biomass sorghum hybrids and two commercial forage sorghum hybrids. These 55 genotypes were evaluated across 29 VCU trials, comprising 13 different locations from nine Brazilian states, geographically distributed as shown in Fig. 1, over five years (2013 to 2017). Some of the 55 hybrids were evaluated in all 29 VCU trials, while others were evaluated only across locations on a given year. Hybrids were obtained from crossing five R-lines (restorer) to 26 A-lines (cytoplasmic male-sterile), belonging to the breeding program. Trials were named according to the initials of each location followed by the last two digits of its respective year.

For each year between five and eight trials were established. Table 1 shows the year and location of each trial evaluated in this study. The number of evaluated trials and genotypes per year varied. In addition, on average each location was evaluated twice, ranging from one to five

**Table 1**
Embrapa's trials evaluated over 13 locations and five years.

| Location* | Years | Location | Average· |
|---|---|---|---|
| Cáceres – MT | 2014 | Cac.14 | 71.25 |
| C. dos Goytacazes - RJ | 2016 | CG.16 | 79.12 |
| Dourados – MS | 2014 | Dou.14 | 85.16 |
| Dourados – MS | 2015 | Dou.15 | 102.97 |
| Dracena – SP | 2014 | Dra.14 | 53.51 |
| Goiânia – GO | 2015 | Goi.15 | 73.06 |
| Goiânia – GO | 2017 | Goi.17 | 65.51 |
| Lavras – MG | 2013 | Lav.13 | 74.93 |
| Lavras – MG | 2015 | Lav.15 | 68.11 |
| Lavras – MG | 2016 | Lav.16 | 64.54 |
| Lavras – MG | 2017 | Lav.17 | 50.73 |
| Nova Porteirinha – MG | 2013 | NP.13 | 74.45 |
| Nova Porteirinha – MG | 2014 | NP.14 | 70.50 |
| Nova Porteirinha – MG | 2015 | NP.15 | 102.23 |
| Nova Porteirinha – MG | 2016 | NP.16 | 93.79 |
| Pelotas – RS | 2016 | Pel.16 | 52.25 |
| Planaltina – DF | 2016 | Pla.16 | 60.91 |
| Planaltina – DF | 2017 | Pla.17 | 54.50 |
| Sinop – MT | 2013 | Sin.13 | 82.06 |
| Sinop – MT | 2014 | Sin.14 | 68.70 |
| Sinop – MT | 2015 | Sin.15 | 74.16 |
| Sinop – MT | 2016 | Sin.16 | 71.29 |
| Sinop – MT | 2017 | Sin.17 | 70.99 |
| Sete Lagoas – MG | 2013 | SL.13 | 64.12 |
| Sete Lagoas – MG | 2015 | SL.15 | 66.03 |
| Sete Lagoas – MG | 2016 | SL.16 | 75.09 |
| Sete Lagoas – MG | 2017 | SL.17 | 81.09 |
| Santa Vitória – MG | 2013 | SV.13 | 60.69 |
| Vilhena – RO | 2016 | Vil.16 | 72.38 |

* DF: Distrito Federal; GO: Goiás; MG: Minas Gerais; MT: Mato Grosso; RJ: Rio de Janeiro; RS: Rio Grande do Sul; RO: Rondônia; SP: São Paulo; · Phenotypic average of each evaluated location.

years of evaluation.

Each trial was arranged as a lattice design with three replicates. Plots consisted of two rows of 5 m spaced by 0.7 m, consisting on an initial density of 140,000 plants $ha^{-1}$. The number of genotypes evaluated in each year of VCU trials ranged from 16 (2013) to 36 (2014), with an average of 25 per trial. Seven to 21 genotypes, with an average of 12, were common across years (Table 2). Two to four checks were added to each trial. Nevertheless, BRS716 and BRS655, biomass and forage Embrapa's commercial hybrids, respectively, were included as common checks across all VCU trials. Although the MET dataset used in this study is highly unbalanced across years, it is balanced across locations within the same year.

Fertilizer management weed and pest control, and other agricultural practices were performed as recommended for sorghum cultivation in each Brazilian region. In all VCU trials, FBY (t $ha^{-1}$) was evaluated at grain physiological maturity. At harvest, all plants in a given plot, i.e. two rows of 5 m, were cut at ground level by hand and weighed using a digital suspension scale. Subsequently, the weights were converted into tons per hectare.

**Table 2**
Number of genotypes within each year (diagonal), and the number of common genotypes between years (upper diagonal) of the VCU trials and the percentage of common genotypes between the years (lower diagonal).

| Year | 2013 | 2014 | 2015 | 2016 | 2017 |
|---|---|---|---|---|---|
| 2013 | 16(5*) | 12 | 7 | 6 | 6 |
| 2014 | 75% | 36(5*) | 17 | 16 | 14 |
| 2015 | 44% | 46% | 25(6*) | 12 | 11 |
| 2016 | 38% | 43% | 48% | 25(8*) | 21 |
| 2017 | 38% | 38% | 44% | 84% | 25(5*) |

* number of locations evaluated each year.

### 2.2. Environmental covariates

In order to identify potential climatic covariates that affect biomass sorghum productivity, maximum, mean and minimum temperature (°C), solar radiation (MJ $m^{-2}$ $day^{-1}$), precipitation (mm $day^{-1}$), relative humidity (%), wind speed (m $s^{-1}$) and altitude (m) were collected for the whole crop growth cycle, i.e. from planting to harvest of each trial. The climatic covariates were obtained from the database of the National Aeronautics and Space Administration Prediction of Worldwide Energy Resource (NASA POWER) project (https://power.larc.nasa.gov/data-access-viewer/) according to the coordinates of each location and the crop growth cycle duration of each trial.

### 2.3. Phenotypic Data Analyses

#### 2.3.1. Single-trial Analyses

Linear mixed models were fitted using the statistical package ASReml-R v.3 (Butler et al., 2009), available for the R software (R Core Team, 2018). Phenotypic data quality, experimental accuracy, coefficient of variation and generalized heritability were calculated by single-trial analyses using the following linear mixed model:

$$\boldsymbol{y} = \mu 1_{\boldsymbol{n}} + \boldsymbol{Xr} + \boldsymbol{Z}_1\boldsymbol{b}.\,\boldsymbol{r} + \boldsymbol{Z}_2\boldsymbol{g} + \boldsymbol{e}$$

where $\boldsymbol{y}$ is the vector ($n \times 1$) of phenotypic values, in which $n$ is the number of observations; $\mu$ is the overall mean; $\boldsymbol{r}$ is the vector ($j \times 1$) of fixed effects of $j$ replicates; $\boldsymbol{g}$ is the vector ($i \times 1$) of the random effects of $i$ genotypes, with $\boldsymbol{g} \sim N(0, \sigma_g^2\boldsymbol{I_i})$, in which $\sigma_g^2$ is the total genetic variance; $\boldsymbol{b}.\,\boldsymbol{r}$ is the vector ($jk \times 1$) of random $k$ block effects within $j$ replicates, with $\boldsymbol{b}.\,\boldsymbol{r} \sim N(0, \sigma_b^2\boldsymbol{I_{jk}})$, in which $\sigma_b^2$ is the variance of blocks; and $\boldsymbol{e}$ is the vector of residual effects, with $\boldsymbol{e} \sim N(0, \sigma_e^2\boldsymbol{I_n})$, in which $\sigma_e^2$ is the residual variance. $\boldsymbol{X}$, $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$ represent the incidence matrices for their respective effects of replicates, genotypes and blocks within replicates, $1_{\boldsymbol{n}}$ is a vector of ones, and $\boldsymbol{I_i}$, $\boldsymbol{I_{jk}}$ and $\boldsymbol{I_n}$ are identity matrices with their corresponding orders.

Diagnostic plots were used to detect outliers and to verify residuals assumptions using the fitted models from each trial. Generalized heritability ($H^2$) (Cullis et al., 2006), and experimental accuracy ($A$ c) (Mrode, 2014), were estimated for each trial using the following equations: $H^2 = 1 - [P\bar{E}V/(2 \times \hat{\sigma}_g^2)]$ and $Ac = \sqrt{1 - (P\bar{E}V/\hat{\sigma}_g^2)}$; in which $P\bar{E}V$ is the average of the prediction error variance, and $\hat{\sigma}_g^2$ is the estimated genetic variance. The coefficient of variation was calculated using the formula $CV\% = (\hat{\sigma}_e/\hat{\mu}) \times 100$, in which $\hat{\sigma}_e$ is the estimated residual standard deviation, and $\hat{\mu}$ is the overall mean of each trial.

#### 2.3.2. Joint Analysis of Multiple Trials

The combined analysis of all trials was carried out in two stages. First, the adjusted means of genotypes ($\boldsymbol{y}_{adj}$) and the environmental residuals were obtained by the single-trial analyses (i.e., by environment), but considering the effect of the genotypes as a fixed effect. At this stage, the means of genotypes were corrected for the other experimental design related effects.

In the second stage, linear mixed models were fitted across environments, using the adjusted means of genotypes from the first stage and associated weights:

$$\boldsymbol{y}_{adj} = \mu 1_{\boldsymbol{n}} + \boldsymbol{Xs} + \boldsymbol{Zg}.\,\boldsymbol{s} + \boldsymbol{e}$$

where $y_{adj}$ is the vector ($il \times 1$) of adjusted means, from the first stage, of $i$ treatments (genotypes) in each environment $l$; $\mu$ is the overall mean; $\boldsymbol{s}$ is the vector ($l \times 1$) of fixed effects of environments; $\boldsymbol{g}.\,\boldsymbol{s}$ is the vector ($il \times 1$) of random genetic effects of $i$ genotypes within $l$ environments, with $\boldsymbol{g}.\,\boldsymbol{s} \sim N(0,\ \boldsymbol{G} \otimes \boldsymbol{P})$; and $\boldsymbol{e}$ is the vector ($il \times 1$) of residual effects, with $\boldsymbol{e} \sim \boldsymbol{N}(0, \boldsymbol{\Sigma})$. Here, $\boldsymbol{G}$ is the ($l \times l$) genetic VCOV matrix for the effect of genotypes within environments, $\boldsymbol{P}$ is the ($i \times i$) pedigree-based relationship matrix estimated by the Henderson's recursive method (Amadeu et al. 2016), and $\boldsymbol{\Sigma}$ is a ($l \times l$) diagonal matrix, in which its elements are given by the inverse of the residual variance of the

adjusted means of genotypes in each environment (Smith et al., 2001a). A factor analytic structure of order $k$ ($FA_k$), in which $k$ is the number of multiplicative components, was considered to model the $\boldsymbol{G}$ matrix. $\boldsymbol{X}$ and $\boldsymbol{Z}$ are the incidence matrices for their respective effects; $1_n$ is a vector of ones; and $\boldsymbol{I_{il}}$ is identity matrices with their corresponding orders.

The overall percentage ($\bar{v}$) of the genetic variance explained by the $k$ factors from the FA structure was calculated using:

$$\bar{v} = 100 \times tr(\boldsymbol{\Lambda\Lambda^T})/\, tr(\boldsymbol{\Lambda\Lambda^T} + \boldsymbol{\psi})$$

where $\boldsymbol{\Lambda}$ is the matrix ($l \times k$) of factor loadings, $\{\lambda_{lk}\}$, in which $\boldsymbol{\lambda_{lk}}$ is the $k$ factor loading ($k = 1,2, \ldots, K$) for the environment $1$; $\boldsymbol{\psi}$ it is a diagonal matrix ($l \times 1$) with the specific variances for each environment; and $tr$ is the trace function. FA models of different orders can be compared based on the overall percentage of the genetic variance explained by the factors in the model, or on their AIC (Akaike Information Criterion) (Bozdogan, 1987) and BIC (Bayesian Information Criterion) (Schwarz, 1978) values. In this study, models from the first ($FA_1$) to sixth orders ($FA_6$) were compared. Here, the best FA model was selected based on the AIC values.

The VCOV matrix for the effect of genotypes within environments, defined by the $FA_k$ model is:

$$\boldsymbol{G} = (\boldsymbol{\Lambda\Lambda}^T + \boldsymbol{\psi}) \otimes \boldsymbol{I_l}$$

where $\boldsymbol{\Lambda}$ is the matrix $l \times k$ of factor loadings $\{\lambda_{lk}\}$, in which $\lambda_{lk}$ is the $k^{\text{th}}$ factor loading of the ($k = 1, 2, \ldots, k$) for the environment $l$; $\boldsymbol{\psi}$ is a diagonal matrix ($l \times l$) with specific variances for each environment; $\boldsymbol{I_l}$ is an identity ($l \times l$) matrix. The genetic correlations between pairs of environments ($\boldsymbol{\rho_{ll'}}$) was calculated via the FA model using the terms of the above $\boldsymbol{G}$ matrix as $\rho_{ll'} = COV_{ll'}/\sqrt{\sigma^2_{ll}\sigma^2_{ll'}}$, where $COV_{ll'}$ is the genetic covariance among trials $l$ and $l'$; and $\sigma^2_{ll}$ and $\sigma^2_{l'l'}$ are the genetic variances for the trials $l$ and $l'$, respectively. The genetic correlations between environments were used to infer about the presence or the absence of G × E. Thus, a high correlation between two environments correspond to a low G × E, i.e. the genotypes have similar genetic responses in both environments.

After estimating the variance components and solving the equation of mixed models, the factor scores for genotypes ($\widetilde{\boldsymbol{f}}$), i.e. the genotype effects for each factor, and the factor loadings for environments ($\widetilde{\boldsymbol{\delta}}$), i.e. the environmental effects for each factor, were obtained as described by Resende and Thompson (2004).

#### 2.3.3. Yield Stability Analysis Across Environments

Latent regression plots were built for a selected group of 10 genotypes that expressed the best yield performance in the joint analysis. These plots can be used to investigate yield adaptability and stability of genotypes across different environments (Smith and Cullis, 2018). The predicted breeding values reflect the genotype responses to a factor loading of a given environment. According to Cullis et al. (2010), for a meaningful interpretation, environmental factor loadings must be rotated to a principal components solution, maximizing the proportion of the genetic covariance accounted by for the first rotated factor loading, while the second rotated factor loading accounts for the next largest proportion and is orthogonal to the first factor, and so on. For this, the rotation of the factors was performed via Varimax (Kaiser, 1958).

#### 2.3.4. Correlations with Environmental Covariates

Pearson's correlation coefficients were estimated between each environmental covariate and the factor loadings for environments extracted from the FA analysis. This information is useful to understand the effect that each covariate has over the genotype performance across environments, allowing to identify the most likely factors affecting G × E in a given set of environments.

## 3. Results

### 3.1. Data Description and Single-trial Analyses

The average FBY across trials ranged from 50.73 (Lav.17) to 102.97 t ha$^{-1}$ (Dou.15), with an overall mean of 71.87 t ha$^{-1}$ (Table 1). Genetic variances ranged from 43.06 (SV.13) to 524.54 (Sin.13) across trials and differed significantly from zero ($p < 0.05$), based on the Likelihood ratio test (LRT). Considerable heterogeneity of residual variances ($\hat{\sigma}^2_e$) was observed, with values ranging from 16.61 (Pla.17) to 217.43 (Sin.13) across trials. Generalized heritability values were high, ranging from 0.69 (Pel.16) to 0.95 (Pla.16 e Pla.17), while experimental accuracy values ranged from medium (0.61, Pel.16) to very high (0.95, Pla.16 e Pla.17). By contrast, the coefficients of variation (CV%) ranged from 7.45 (Pla.17) to 17.97% (Sin.13), showing high experimental precision. Additional information of the individual analyses per trial are presented in Supplementary Table S1.

### 3.2. Joint Analysis of Multiple Trials

The results from the MET analysis based on a factor analytic structure, together with the total number of parameters (NP) and the values of AIC and REML log-likelihood (logREML) of the fitted models, considering distinct VCOV structures for the estimated **G** matrix, are presented in Table 3. Here, the lowest AIC value was observed for $FA_{(3)}$, with a total of 99 parameters.

Genetic correlations between environments (Fig. 2) ranged from -0.21 (Goi.17 and Pel.16) to 0.99 (Lav.13 and Lav.15; Lav.13 and Pel.16; Lav.15 and Pel.16), indicating the existence of null to high G × E between environments. Goiânia in the year 2017 (Goi.17) was the environment that presented the highest G × E with correlations close to zero or negative with other locations (Fig. 2), such as CG.16 (-0.07), NP.15 (-0.11), Pel.16 (-0.21), Pla.16 (-0.01) and Sin.15 (-0.03). The pair Dou.15 and Sin.14 (-0.02) also presented a negative correlation close to zero. More than 95% of the genetic correlations between environments were positive. In this study, genetic correlations above 0.60 were considered high, indicating the occurrence of low G × E, i.e. the genotypes exhibited similar FBY between environments. By contrast, pairs of environments showing low correlation (bellow 0.30), suggest the occurrence of high G × E, i.e. the performance of genotypes changed across environments. Additionally, the genetic correlation between environments can be used to define mega-environments.

The three factors of the $FA_{(3)}$ model jointly explained 79.03% of the observed genetic variance. The first factor of this structure captured 35.42% of the genetic variability. The environments Cac.14, CG.16, Dou.14, Dra.14, Lav.13, Lav.15, NP.15, NP.16, Pel.16, Sin.13, Sin.14, Sin.15, Sin.16, Sin.17, SL.13, SL.15 and Vil.16 showed the highest loadings for this first factor (identified marked in bold on Table 4). The second factor explained 24.07% of the genetic variability, representing the highest loadings for Dou.15, Lav.16, Lav.17, NP.13, NP.14, Pla.16, Pla.17 and SL.16. And the third factor explained 19.55% of the genetic variability, and the environments Goi.15, Goi.17, SL.17 and SV.13 were

**Table 3**
Total number of parameters (NP), Akaike Information Criterion (AIC) and REML log-likelihood (logREML) of the variance and covariance (VCOV) models examined for the estimated **G** matrix in the combined analysis of environments.

| Structure | NP | AIC | logREML |
|---|---|---|---|
| $FA_{(1)}$ | 58 | 3956.56 | −1920.28 |
| $FA_{(2)}$ | 81 | 3918.99 | −1878.50 |
| **$FA_{(3)}$** | **99** | **3910.49** | **-1856.25** |
| $FA_{(4)}$ | 121 | 3918.62 | −1838.31 |
| $FA_{(5)}$ | 144 | 3937.89 | −1824.95 |
| $FA_{(6)}$ | 165 | 3945.53 | −1807.77 |

$FA_{(k)}$: factor analytic model of order k.

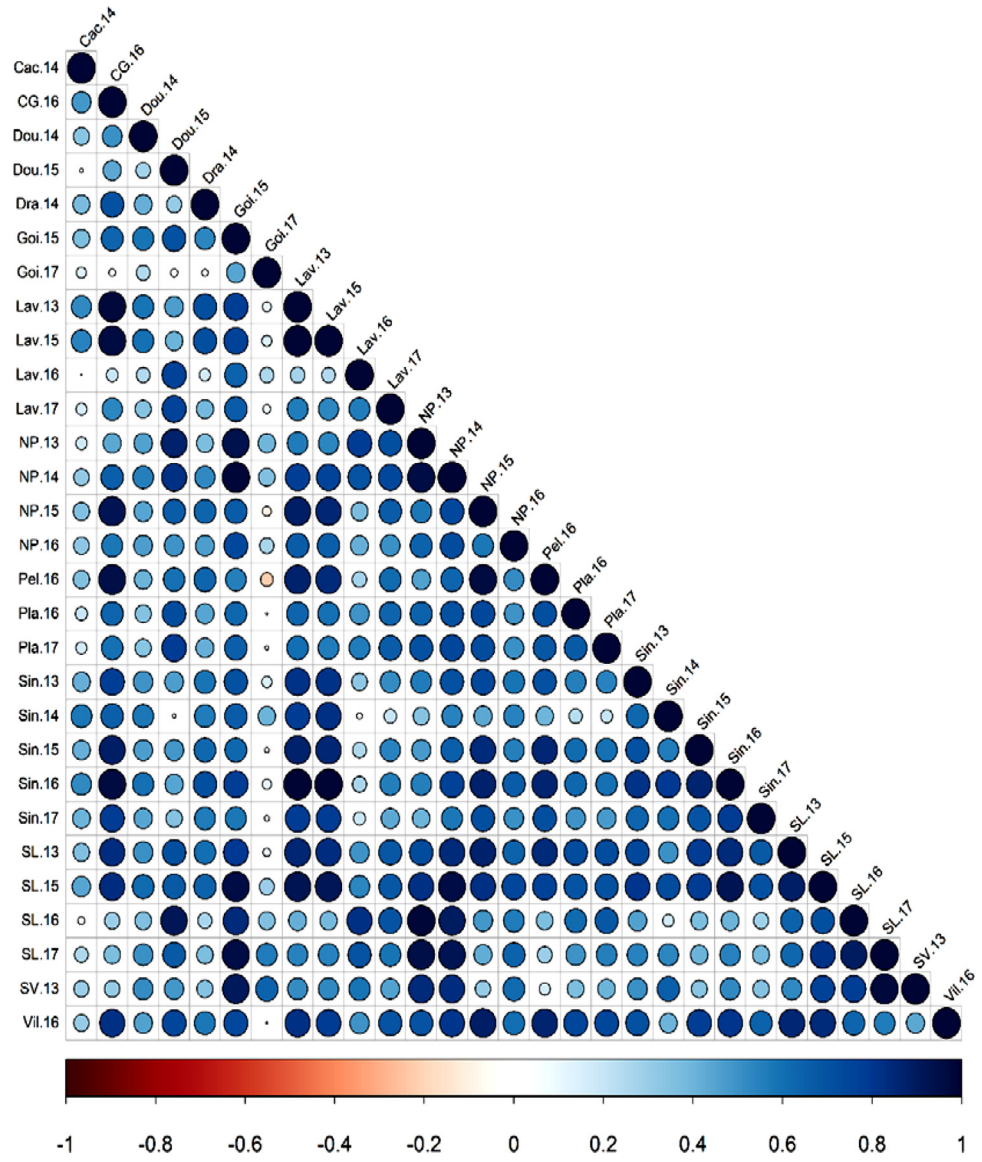


**Fig. 2.** Estimated pairwise genetic correlations for FBY between 13 environments: Cac: Cáceres-MT; CG: Campos dos Goytacazes-RJ; Dou: Dourados-MS; Dra: Dracena-SP; Goi: Goiânia-GO; Lav: Lavras-MG; NP: Nova Porteirinha-MG; Pel: Pelotas-RS; Pla: Planaltina-DF; SV: Santa Vitória-MG; SL: Sete Lagoas-MG; Sin: Sinop-MT; Vil: Vilhena-RO. Evaluated in the years 2013 to 2017, which were coded as 13 to 17. The size and color of the circles are related to the magnitude and direction of the genetic correlations between environments, respectively.

represented by this factor.

The location Sinop was mostly described by the first factor for all experimental years. A similar tendency was noted for Planaltina and Goiânia, that had all years allocated in the second and third factors, respectively. A slight tendency was also observed to group locations evaluated in 2014 in the first factor, which represented four of the five locations tested in this year.

### 3.3. Yield Stability Analysis Across Environments

The latent regression plots were built for 10 genotypes that were evaluated in at least 50% of the locations. These were selected according to the highest overall predicted means based on the $FA_{(3)}$ model, such as proposed by Smith and Cullis (2018). First, latent regressions plots were built for the first factor, regressing the predicted breeding values on the rotated environmental loadings of factor 1 (Fig. 3.A). Then, for the second factor, the predicted breeding values were adjusted for the first factor and then regressed on the rotated environmental loadings of factor 2 (Fig. 3.B), and for the third factor, the predicted breeding values were adjusted for the second factor and then regressed on the rotated environmental loadings of factor 3 (Fig. 3.C). In each latent regression plot, the solid circles and hollow circles correspond to the predicted breeding values of genotypes in tested and untested locations, respectively. The overall predicted means obtained by the $FA_{(3)}$ structure ranged from 37.71 t $ha^{-1}$ (T_03, which corresponds to the check BRS655) to 84.38 t $ha^{-1}$ (H_30, hybrid), with an overall mean of 72.00 t $ha^{-1}$. Additionally, the overall predicted means for the 10 highest and 10 smallest yielding genotypes, and their respective factor scores from the $FA_{(3)}$ structure, are presented in the supplementary Tables S2 and S3, respectively.

In the latent regression plots for the factors 1, 2 and 3 (Fig. 3), the slope of the regression line ($\beta_1$) corresponds to the genotype score of the respective factor. Thus, genotypes showing high and positive slopes are more responsive to environmental improvements, i.e. exhibited higher predicted breeding values in environments having higher factor loadings, meaning that they were more adapted to these environments. For example, among all environments, NP.15 (15.40), Lav.15 (13.45), Sin.13 (13.00), Sin.16 (12.26), Sin.17 (11.99) and Sin.15 (11.60)

**Table 4**
Environmental loadings for the factor analytic model of order three, i.e. three factors [$FA_{(3)}$].

| Environments | Factors | | |
|---|---|---|---|
| | Factor 1 | Factor 2 | Factor 3 |
| Cac.14 | **4.95** | −0.90 | −2.32 |
| CG.16 | **7.71** | 2.00 | 0.17 |
| Dou.14 | **8.83** | 3.10 | −6.81 |
| Dou.15 | 1.87 | **9.44** | −0.37 |
| Dra.14 | **7.21** | 1.75 | −1.39 |
| Goi.15 | 4.40 | 5.13 | **-5.32** |
| Goi.17 | −0.75 | 0.65 | **-7.91** |
| Lav.13 | **8.70** | 2.56 | −1.97 |
| Lav.15 | **13.45** | 3.07 | −3.67 |
| Lav.16 | 0.05 | **9.23** | −3.53 |
| Lav.17 | 4.02 | **7.32** | −0.97 |
| NP.13 | 3.04 | **10.40** | −6.37 |
| NP.14 | 4.34 | **6.23** | −4.11 |
| NP.15 | **15.40** | 9.80 | 2.10 |
| NP.16 | **7.96** | 6.23 | −6.04 |
| Pel.16 | **5.23** | 2.71 | 1.49 |
| Pla.16 | 5.75 | **7.51** | 0.21 |
| Pla.17 | 4.40 | **7.33** | −0.11 |
| Sin.13 | **13.00** | 5.78 | −4.29 |
| Sin.14 | **8.26** | −2.11 | −7.10 |
| Sin.15 | **11.60** | 4.56 | −0.17 |
| Sin.16 | **12.27** | 3.31 | −3.02 |
| Sin.17 | **11.99** | 3.15 | −1.69 |
| SL.13 | **8.23** | 6.74 | −1.55 |
| SL.15 | **8.89** | 6.36 | −5.18 |
| SL.16 | 1.16 | **12.18** | −5.85 |
| SL.17 | 2.96 | 7.82 | **-9.50** |
| SV.13 | 1.58 | 2.91 | **-6.13** |
| Vil.16 | **6.70** | 6.11 | −0.24 |

presented the highest loadings for the first factor (Table 4 and Fig. 3). Thus, the hybrids H_07 ($\beta_1$ = 2.10) and H_09 ($\beta_1$ = 1.72) can be identified as the more responsive genotypes for these environments, because they present a higher angular coefficient. On the other hand, slopes close to zero are observed for genotypes with yield stability across the set of environments, i.e. even with an improvement of the environment the genotype presents yield stability. For example, the hybrids H_26 and H_28 did not respond to the environmental changes, i.e. showed stable performance across environments.

For the second factor, SL.16 (12.18), NP.13 (10.43), NP.15 (9.80), Dou.15 (9.43) and Lav.16 (9.25) showed the highest environmental loadings (Table 4 and Fig. 3). The latent regression plots on the second factor highlighted the hybrids H_07 ($\beta_1$ = 1.92), H_26 ($\beta_1$ = 2.10), H_29 ($\beta_1$ = 2.41) and H_38 ($\beta_1$ = 2.39) as more responsive genotypes to these environments.

For the third factor, the environments SL.17 (-9.50), Goi.17 (-7.91) and Sin.14 (-7.10) showed the highest loadings, but with negative values (Table 4 and Fig. 3). The latent regression plots highlighted the hybrid H_26 ($\beta_1$ = -2.76), H_28 ($\beta_1$ = -2.96), and T_04 ($\beta_1$ = -1.87) as the genotypes of higher score, but in an opposite sense, i.e. a decrease in genotype performance with an increase in the environmental loadings. Only hybrid H_07 ($\beta_1$ = 0.40) showed a positive value for this factor, but in a low degree, indicating that it is the most stable genotype across environments clustered in the factor 3. Additionally, Figs. 3B and 3C also show changes from downward to upward direction for the genotypes H_26, H_28 and T_04 indicating a positive response of these hybrids to the environmental loadings of the second factor, and from upward to downward showing a negative response of these hybrids to third factor environmental loadings.

Table 5 presents the 10 hybrids exhibiting the highest overall predicted means, which were evaluated in at least 50% of the locations. The hybrid H_30 showed the highest predicted breeding value among all the evaluated genotypes for all environments. Among the most productive materials, the hybrid H_07 can be highlighted as a highly adapted genotype across environments represented by the first and the second factors, and as a stable genotype across the environments represented by the third factor.

### 3.4. Correlation with environmental covariates

In Table 6, the Pearson's correlation coefficients between the environmental loadings and the environmental covariates are presented for the three factors of the $FA_{(3)}$ structure. This information is useful to understand the effect of each environmental covariate in the genotype's performance across environments, allowing to identify the most likely factors affecting G × E in a given set of environments. Based on these correlations, we can identify that the covariates of precipitation, minimum temperature and wind speed affected the genotypes performance in the environments clustered by factor 1 (highlighted in bold in Table 6). On the other hand, the main covariates affecting FBY in the environments clustered by factor 2 were altitude, minimum temperature and solar radiation. Finally, the environments clustered by the third factor were mostly affected by crop growth cycle duration, expressed here as the number of days from planting to harvesting.

## 4. Discussion

The Embrapa's biomass sorghum breeding program constantly seeks for more productivity, adaptability and stability for a wide range of environments. In this sense, multi-environment trials (MET) are annually executed to evaluate yield performance of genotypes in different environments, distributed across Brazil, to cover distinct regions and edaphoclimatic conditions of interest. The release of new cultivars only occurs when they present specific features that prove their desirable performance for a given geographic region, highlighting the importance of MET studies in breeding programs.

Factor analytic (FA) structures fitted within the linear mixed model framework have emerged as a flexible and robust approach for modeling genetic variance-covariance matrices, being more parsimonious for MET analyses than unstructured models (Smith et al., 2001a). Additionally, mixed models show great flexibility to deal with unbalanced data, which is extremely important for breeding programs since low-performance genotypes are often replaced by newly developed materials over the years. G × E studies have been widely implemented using FA structures to understand the adaptability and stability of genotypes across environments (Dias et al., 2018a; Li et al., 2017; Peixouto et al., 2016), and also to define mega-environments in plant breeding programs (Monteverde et al., 2018; Smith et al., 2015; Smith and Cullis, 2018). However, few studies have used FA models to infer about the main factors affecting G × E in crop species, as done here. Additionally, to our knowledge, this is the first study that explores the correlations between environmental covariates and FA loadings to investigate the potential factors affecting G × E for FBY in biomass sorghum. This information can help breeders to recommend genotypes for a given set of environments based on historical series of climatic data, as well as to optimize the number of environments tested in late-stage trials, prioritizing environments with different edaphoclimatic conditions and, consequently, expected genotypes response.

Smith et al. (2015) proposed the use of latent regression plots to study yield stability and adaptability of genotypes across environments. In their approach, the predicted breeding values of genotypes are regressed on the environmental factor loadings of the FA model. In the present study, factor analysis and latent regression plots allowed identifying genotypes adapted to different environments, such as the hybrid H_48 that presented better performance according to the environmental improvements, i.e. higher adaptability (1.75, 1.41 and 1.21 for factor 1, 2 and 3, respectively, data not shown). However, the predicted breeding values of this genotype was close to the average genotype performance within environments. By contrast, the hybrid

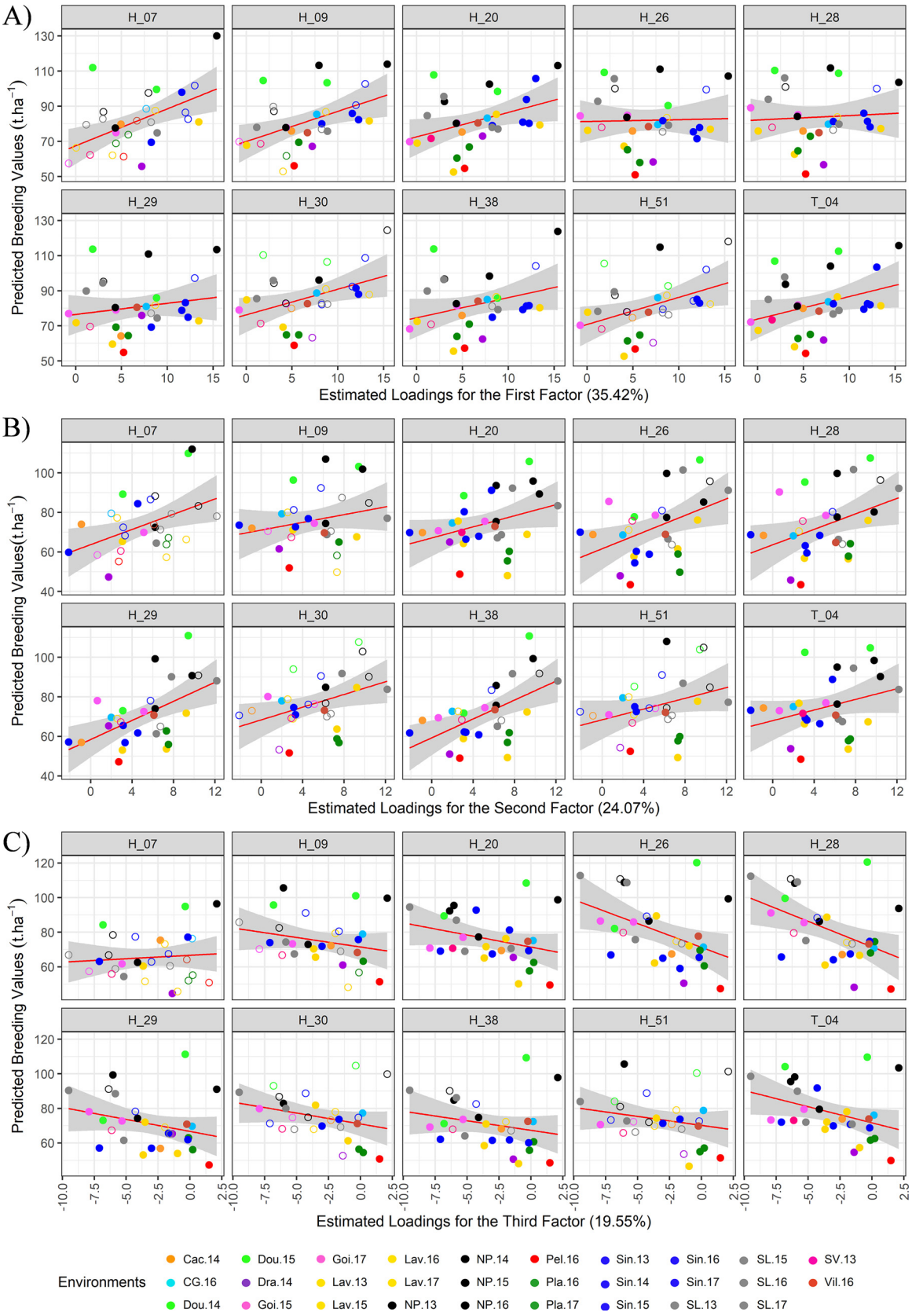


**Fig. 3.** Latent regression plots for the best 10 genotypes grown in at least 50% of the locations using: A) First factor; B) Second factor; and C) Third factor. The solid circles correspond to predicted breeding values of genotypes in tested locations, and the hollow circles to predicted breeding values of genotypes in untested locations. The solid red line and the gray shade correspond to the latent regression line and to the confidence interval of 95%, respectively. The values within parenthesis in the label of *x* axis correspond to the proportion of the genetic variance explained by each factor.

**Table 5**
Rank of the best 10 genotypes evaluated in at least 50% of the environments, based on their overall predicted means (Means); the slope (genotype score) of the latent regression lines for the first (Factor 1), the second (Factor 2) and the third (Factor 3) factors.

| Hybrids | Means | Factor 1 | Factor 2 | Factor 3 |
|---|---|---|---|---|
| H_07 | 80.32 | 2.10 | 1.92 | 0.40 |
| H_09 | 80.19 | 1.72 | 0.98 | −1.10 |
| H_20 | 80.59 | 1.39 | 1.40 | −1.40 |
| H_26 | 80.72 | 0.10 | 2.10 | −2.76 |
| H_28 | 82.38 | 0.25 | 1.96 | −2.96 |
| H_29 | 79.57 | 0.63 | 2.41 | −1.39 |
| H_30 | 84.38 | 1.46 | 1.61 | −1.26 |
| H_38 | 80.95 | 1.15 | 2.39 | −1.10 |
| H_51 | 79.93 | 1.54 | 1.36 | −1.03 |
| T_04 | 81.11 | 1.28 | 1.34 | −1.87 |

**Table 6**
Pearson's correlation coefficient between the environmental covariates and the environmental loadings extracted from the factor analysis. The highest correlations between environmental covariates and the environmental loadings of $FA_{(3)}$ model are highlighted in bold.

| Covariate | Factor 1 | Factor 2 | Factor 3 |
|---|---|---|---|
| Precipitation | **0.362** | 0.094 | −0.009 |
| Maximum Temperature | 0.083 | 0.044 | −0.012 |
| Minimum Temperature | **0.353** | **-0.391** | 0.199 |
| Average Temperature | 0.252 | −0.193 | 0.102 |
| Solar Radiation | 0.024 | **0.305** | −0.107 |
| Crop's Growth Cycle | 0.130 | 0.069 | **-0.221** |
| Humidity | 0.175 | −0.279 | 0.076 |
| Wind Speed | **-0.390** | 0.250 | 0.051 |
| Altitude | −0.237 | **0.454** | −0.254 |

H_07 presented a high and positive response to the environments clustered in the first and the second factors, with a slight stability over the environments clustered in the third factor, besides presenting one of the highest predicted means for all environments. In breeding programs, high-yielding stable hybrids are very desired. Additionally, latent regression plots also allowed selecting, among the best 10 genotypes, those specifically adapted to a given set of environments. Moreover, it was also possible to select genotypes showing yield stability across some environments, such as the hybrids H_26 and H_28 to the environments allocated in the first factor, and H_65 to all other environments.

Moreover, based on the latent regression plots, the hybrids H_20, H_30, and T_04 were the most responsive high-yielding genotypes to the environmental loadings for all factors. Interestingly, the T_04 is an Embrapa's biomass sorghum commercial hybrid, BRS716, released in 2014. This hybrid is recommended for cultivation in a wide range of Brazilian regions, mainly in the Southeast and Central-West regions, being able to reach FBY of up to 150 t $ha^{-1}$, much higher than the Brazilian overall average of 70 t $ha^{-1}$ for biomass sorghum.

Overall low productivities were observed for the checks T_01 and T_02 (Table S3), which correspond to cultivars commercially released in Brazil by two distinct private companies as biomass and forage hybrids, respectively. Another of Embrapa's check, the hybrid T_03 (BRS 655), also presented an overall low FBY. However, BRS 655 is a photoperiod insensitive forage sorghum cultivar, explaining this low FBY (approximately 55.00 t $ha^{-1}$). The above-mentioned genotypes exhibited the three lowest overall predicted means (Table S3). Moreover, these hybrids did not show adaptability to any of the environments, except T_03 that showed adaptability to the environments clustered by the third factor.

Although factor analysis allowed for an efficient identification of hybrids exhibiting adaptability to a given environment, as well as hybrids showing yield stability across a set of environments, no clear patterns were observed to cluster locations in mega-environments. However, high genetic correlations were observed among some environments, suggesting that the evaluated hybrids exhibited very similar genetic responses across them. For example, Campos dos Goytacazes (CG), Dracena (Dra), Nova Porteirinha (NP), Pelotas (Pel), Santa Vitória (SV) and Vilhena (Vil) exhibited higher pairwise genetic correlations to most of the other environments, mainly to the environments clustered in the same factor. These locations can be dropped from future evaluations, optimizing the resources to evaluate other untested (or poorly tested) geographic regions in Brazil.

Furthermore, the FA structures considered in the present study can be easily extended to incorporate genomic relationship matrices, which are used to predict the performance of untested genotypes across environments via MET genomic selection (MET-GS) models. This is an interesting approach to accelerate the genetic gains in biomass sorghum breeding programs in the future. Several studies have been focused on the prediction of untested genotypes, such as Burgueño et al. (2012), Dias et al. (2018b) and Dias et al. (2020). These authors presented interesting results about the accuracy of MET-GS models to predict untested genotypes across all environments, or genotypes tested in some locations but not in others, based on the genetic relationship among tested and untested genotypes. MET-GS models provided better accuracies when predicting the performance of genotypes within correlated environments.

Latent regression plots allowed studying genotype performance across different environments, showing their responses to varying environmental covariates. According to Smith et al. (2015), environmental loadings are difficult to interpret, being more consistent when environmental covariates are included with the FA structure. Thus, the FA structure considered in the present study were extended to include some edaphoclimatic covariates, and showed that minimum temperature, solar radiation, precipitation, wind speed, crop growth cycle duration and altitude were affecting the genotypes response within environments clustered by different factors. Habyarimana (2004) and Olson et al. (2012), report that the accumulation of biomass in sorghum is directly influenced by the availability of water, crop growth cycle duration, temperatures and photoperiod, which is an indication of the importance of these environmental covariates for the biomass sorghum.

In this study, the correlations between some environmental covariates and the environmental loadings extracted from the factor analysis were used to explain the possible factors affecting G × E, such as proposed by Sae-Lim et al. (2014). The correlations between the environmental loadings and the environmental covariates were considered low (0.01) to moderate (0.45). In general, the minimum temperature was the environmental covariate that explained the most genetic variability of FBY across environments. It is also important to highlight that minimum temperature presented high correlations to factors 1 and 2, but in opposite directions, affecting the adaptability and stability of genotypes across the environments clustered by these factors. These results are in agreement with the findings of previous studies in sorghum. For example, House (1985) and Bantilan et al. (2004) showed that low temperatures reduce the growth rate in sorghum, increasing the crop growth cycle duration and, consequently, decreasing biomass yield. Moreover, Reddy and Patil (2015) reported an increase in tillering when sorghum genotypes are exposed to low temperatures, which is valued by breeders since these are directly linked to the increased biomass production (Murray et al., 2008). On the other hand, the maximum temperature presented low correlations with all factors, showing less influence on the genotypes performance across environments. Additionally, precipitation was the covariate that affected the most the classification of environments by the first factor, showing a low correlation with the other factors. The same was observed for solar radiation in the second factor, and for crop growth cycle duration in the third factor.

These results suggest that genotypes adapted to the majority of environments, such as H_07, presented a high genetic response to

environments clustered by factors 1 and 2 (i.e. 25 environments) and, consequently, were more affected by changes in minimum temperature. By contrast, H_26 and H_28 presented high scores for factors 2 and 3, suggesting that they were more responsive to changes in minimum temperature, solar radiation, altitude, and crop growth cycle duration. In addition to the minimum temperature, genotypes more adapted to the environments clustered by factor 1 should also be more sensitive to changes in other covariates also expressing moderate correlation to the environmental loadings of factor 1, such as, precipitation and wind speed. In this sense, genotypes adapted to a given set of environments are more affected by variations in the environmental covariates showing moderate to high correlations to factors exhibiting high loadings for these environments.

Therefore, this study shows that the edaphoclimatic covariates can be used successfully to investigate the environmental factors that affect the genotypes response across distinct environments, helping breeders to identify the most suitable hybrids according to the environmental features, as well as stable hybrids to some adverse environmental conditions.

## 5. Conclusion

Despite the presence of high G × E, high genetic correlations were found between some environments, suggesting that some locations can be dropped and, in order to optimize resources, replaced by new sites representing geographic regions not yet tested in late-stage trials. The hybrids H_07, H_20, H_26 and H_29 stood out for their high FBY and for the best response to the environmental improvements, showing adaptability across the evaluated environments, and stability to some specific environments. The use of environmental covariates, such as minimum temperature, altitude and solar radiation, can help breeders in the study and selection of genotypes showing high adaptability and stability across environments, considering the information of current environmental features, without the need of extensive field evaluations.

## Financing

This work was supported by BNDES (Brazilian National Bank for Economic and Social Development), CNPq (National Science and Technology Development Council), CAPES (National Council for the Improvement of Higher Education), Fapemig (Research Support Foundation of Minas Gerais) and Embrapa (Brazilian Agricultural Research Corporation).

## CRediT authorship contribution statement

**Isadora Cristina Martins Oliveira:** Conceptualization, Methodology, Software, Writing - original draft, Writing - review & editing. **José Henrique Soler Guilhen:** Conceptualization, Methodology, Software, Writing - review & editing. **Pedro César de Oliveira Ribeiro:** Validation, Writing - review & editing. **Salvador Alejandro Gezan:** Software, Writing - review & editing. **Robert Eugene Schaffert:** Investigation, Writing - review & editing. **Maria Lucia Ferreira Simeone:** Validation. **Cynthia Maria Borges Damasceno:** Validation. **José Eustáquio de Souza Carneiro:** Writing - review & editing. **Pedro Crescêncio Souza Carneiro:** Conceptualization, Writing - review & editing. **Rafael Augusto da Costa Parrella:** Investigation, Project administration. **Maria Marta Pastina:** Conceptualization, Methodology, Software, Writing - review & editing.

## Declaration of Competing Interest

The authors report no declarations of interest.

## Acknowledgments

The authors thank all scientists, research assistants, undergraduate and graduate students associated with the experimental design, management and data collection at Embrapa Maize and Sorghum, and all the graduate students of the Federal University of Viçosa, who directly or indirectly collaborated in carrying out this study.

## Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.fcr.2020.107929.

## References

Amadeu, R.R., Cellon, C., Olmstead, J.W., Garcia, A.A., Resende, M.F., Muñoz, P.R., 2016. AGHmatrix: R package to construct relationship matrices for autotetraploid and diploid species: A blueberry example. Plant Genome. 9. https://doi.org/10.3835/plantgenome2016.01.0009.

Bantilan, M.C.S., Gowda, C.L.L., Reddy, B.V.S., Obilana, A.B., Evenson, R.E., 2004. Sorghum genetic enhancement: research process, dissemination and impacts. International Crops Research Institute for the Semi-Arid Tropics. Andhra Pradesh, India.

Bornhofen, E., Todeschini, M.H., Stoco, M.G., Madureira, A., Marchioro, V.S., Storck, L., Benin, G., 2018. Wheat yield improvements in Brazil: Roles of genetics and environment. Crop Sci. 58, 1082–1093. https://doi.org/10.2135/cropsci2017.06.0358.

Bozdogan, H., 1987. Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. Psychometrika. 52, 345–370.

Brenton, Z.W., Cooper, E.A., Myers, M.T., Boyles, R.E., Shakoor, N., Zielinski, K.J., Rauh, B.L., Bridges, W.C., Morris, G.P., Kresovich, S., 2016. A genomic resource for the development, improvement, and exploitation of sorghum for bioenergy. Genetics. 204, 21–33. https://doi.org/10.1534/genetics.115.183947.

Burgueño, J., Crossa, J., Cornelius, P.L., Yang, R.-C., 2008. Using factor analytic models for joining environments and genotypes without crossover genotype x environment interaction. Crop Sci. 48, 1291–1305. https://doi.org/10.2135/cropsci2007.11.0632.

Burgueño, J., de los Campos, G., Weigel, K., Crossa, J., 2012. Genomic prediction of breeding values when modeling genotype x environment interaction using pedigree and dense molecular markers. Crop Sci. 52, 707–719. https://doi.org/10.2135/cropsci2011.06.0299.

Butler, D.G., Cullis, B.R., Gilmour, A.R., Gogel, B.J., 2009. ASReml-R reference manual. State Queensland, Dep. Prim. Ind. Fish., Brisbane.

Cullis, B.R., Smith, A.B., Coombes, N.E., 2006. On the design of early generation variety trials with correlated data. J. Agric. Biol. Environ. Stat. 11, 381–393. https://doi.org/10.1198/108571106X154443.

Cullis, B.R., Smith, A.B., Beeck, C.P., Cowling, W.A., 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, B.R., Jefferson, P., Thompson, R., Smith, A.B., 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.

de Almeida, L.G.F., Parrella, R.A.C., Simeone, M.L.F., Ribeiro, P.C.O., Santos, A.S., da Costa, A.S.V., Guimarães, A.G., Schaffert, R.E., 2019. Composition and growth of sorghum biomass genotypes for ethanol production. Biomass and Bioenergy. 122, 343–348.

de Oliveira, A.A., Pastina, M.M., Parrella, R.A.C., Noda, R.W., Simeone, M.L.F., Schaffert, R.E., de Magalhães, J.V., Damasceno, C.M.B., Margarido, G.R.A., et al., 2018. Genomic prediction applied to high-biomass sorghum for bioenergy production. Mol. Breed. 38, 49.

Demirbas, M.F., Balat, M., Balat, H., 2009. Potential contribution of biomass to the sustainable energy development. Energy Convers. Manag. 50, 1746–1760. https://doi.org/10.1016/j.enconman.2009.03.013.

Dias, K.O.D.G., Gezan, S.A., Guimarães, C.T., Parentoni, S.N., Guimarães, P.E., de, O., Carneiro, N.P., Portugal, A.F., Bastos, E.A., Cardoso, M.J., Anoni, C., de, O., de Magalhães, J.V., de Souza, J.C., Guimarães, L.J.M., Pastina, M.M., 2018a. Estimating genotype × environment interaction for and genetic correlations among drought tolerance traits in maize via factor analytic multiplicative mixed models. Crop Sci. 58, 72–83. https://doi.org/10.2135/cropsci2016.07.0566.

Dias, K.O.D.G., Gezan, S.A., Guimarães, C.T., Nazarian, A., e Silva, L.D.C., Parentoni, S.N., Guimarães, P.E., de, O., Anoni, C.O., Pádua, J.M.V., Pinto, M.O., et al., 2018b. Improving accuracies of genomic predictions for drought tolerance in maize by joint modeling of additive and dominance effects in multi-environment trials. Heredity 121, 24–37.

Dias, K.O.G., Piepho, H.P., Guimarães, L.J.M., Guimarães, P.E.O., Parentoni, S.N., Pinto, M.O., Noda, R.W., Guimarães, J.V., Magalhães, C.T., Garcia, A.A.F., Pastina, M.M., 2020. Novel strategies for genomic prediction of untested single-cross maize hybrids using unbalanced historical data. Theor. Appl. Genet. 133, 443–455.

Habyarimana, E., Laureti, D., De Ninno, M., Lorenzoni, C., 2004. Performances of biomass sorghum [*Sorghum bicolor* (L.) Moench] under different water regimes in Mediterranean region. Ind. Crops Prod. 20, 23–28. https://doi.org/10.1016/j.

indcrop.2003.12.019.
House, L.R., 1985. A guide to sorghum breeding, second ed. International Crops Research Institute for the Semi-Arid Tropics, India.
Kaiser, H.F., 1958. The varimax criterion for analytic rotation in factor analysis. Psychometrika. 23, 187–200.
Kelly, A.M., Smith, A.B., Eccleston, J.A., Cullis, B.R., 2007. The accuracy of varietal selection using factor analytic models for multi-environment plant breeding trials. Crop Sci. 47, 1063–1070. https://doi.org/10.2135/cropsci2006.08.0540.
Li, Y., Suontama, M., Burdon, R.D., Dungey, H.S., 2017. Genotype by environment interactions in forest tree breeding: Review of methodology and perspectives on research and application. Tree Genet. Genomes. 13–60 doi: I 10.1007/s11295-017-1144-x.
Mahmood, A., Hussain, A., Shahzad, A.N., Honermeier, B., 2015. Biomass and biogas yielding potential of sorghum as affected by planting density, sowing time and cultivar. Pak J Bot. 47, 2401–2408.
Monteverde, E., Rosas, J.E., Blanco, P., de Vida, F.P., Bonnecarrère, V., Quero, G., Gutierrez, L., McCouch, S., 2018. Multienvironment models increase prediction accuracy of complex traits in advanced breeding lines of rice. Crop Sci. 58, 1519–1530. https://doi.org/10.2135/cropsci2017.09.0564.
Mrode, R.A., 2014. Linear models for the prediction of animal breeding values, third ed. CABI, Oxfordshire, UK.
Murray, S.C., Rooney, W.L., Mitchell, S.E., Sharma, A., Klein, P.E., Mullet, J.E., Kresovich, S., 2008. Genetic improvement of sorghum as a biofuel feedstock: II. QTL for stem and leaf structural carbohydrates. Crop Sci. 48, 2180–2193. https://doi.org/10.2135/cropsci2008.01.0068.
NASA POWER, Available at: https://power.larc.nasa.gov/data-access-viewer/ (accessed November 2019).
Olson, S.N., Ritter, K., Rooney, W., Kemanian, A., McCarl, B.A., Zhang, Y., Hall, S., Packer, D., Mullet, J., 2012. High biomass yield energy sorghum: developing a genetic model for C4 grass bioenergy crops. Biofuels, Bioprod. Biorefining. 6, 640–655. https://doi.org/10.1002/bbb.1357.
Parrella, R.A., da, C., Schaffert, R.E., May, A., Emygdio, B., Portugual, A.F., Damasceno, C.M.B., 2011. Desempenho agronômico de híbridos de sorgo biomassa. Embrapa Milho e Sorgo-Boletim Pesqui. e Desenvolv.
Peixouto, L.S., Nunes, J.A.R., Furtado, D.F., 2016. Factor analysis applied to the G+ GE matrix via REML/BLUP for multi-environment data. Crop Breed. Appl. Biotechnol. 16, 1–6.
Piepho, H.P., 1997. Analyzing genotype-environment data by mixed models with multiplicative terms. Biometrics. 53, 761–766.
Piepho, H.P., 1998. Empirical best linear unbiased prediction in cultivar trials using factor-analytic variance-covariance structures. Theor. Appl. Genet. 97, 195–201.
Quintero, A., Molero, G., Reynolds, M.P., Calderini, D.F., 2018. Trade-off between grain weight and grain number in wheat depends on GxE interaction: A case study of an elite CIMMYT panel (CIMCOG). Eur. J. Agron. 92, 17–29.
R Core Team, 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria ISBN 3-900051-07-0. http://www.R-project.org/.
Reddy, S., Patil, J.V., 2015. Genetic enhancement of rabi sorghum: Adapting the Indian Durras. Academic Press, San Diego, USA.
Reis, A.L.S., Damilano, E.D., Menezes, R.S.C., de Morais Jr, M.A., 2016. Second-generation ethanol from sugarcane and sweet sorghum bagasses using the yeast Dekkera bruxellensis. Ind. Crops Prod. 92, 255–262.
Resende, M.D.V., Thompson, R., 2004. Factor analytic multiplicative mixed models in the analysis of multiple experiments. Rev. Mat. Estat. 22, 31–52.
Sae-Lim, P., Komen, H., Kause, A., Mulder, H.A., 2014. Identifying environmental variables explaining genotype-by-environment interaction for body weight of rainbow trout (*Onchorynchus mykiss*): reaction norm and factor analytic models. Genet. Sel. Evol. 46, 16. https://doi.org/10.1186/1297-9686-46-16.
Schwarz, G., 1978. Estimating the dimension of a model. Annals of Statistics. 6, 461–464.
Smith, A., Cullis, B., Gilmour, A., 2001a. The analysis of crop variety evaluation data in Australia. Aust. New Zeal. J. Stat. 43, 129–145.
Smith, A., Cullis, B., Thompson, R., 2001b. Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. Biometrics 57, 1138–1147.
Smith, A.B., Cullis, B.R., Thompson, R., 2005. The analysis of crop cultivar breeding and evaluation trials: An overview of current mixed model approaches. J. Agric. Sci. 143, 449–462. https://doi.org/10.1017/S0021859605005587.
Smith, A.B., Ganesalingam, A., Kuchel, H., Cullis, B.R., 2015. Factor analytic mixed models for the provision of grower information from national crop variety testing programs. Theor. Appl. Genet. 128, 55–72. https://doi.org/10.1007/s00122-014-2412-x.
Smith, A.B., Cullis, B.R., 2018. Plant breeding selection tools built on factor analytic mixed models for multi-environment trial data. Euphytica. 214, 143.
Thompson, R., Cullis, B., Smith, A., Gilmour, A., 2003. A sparse implementation of the average information algorithm for factor analytic and reduced rank variance models. Aust. New Zeal. J. Stat. 45, 445–459.