

# Evaluation of the performance of sorghum genotypes using GGE biplot

Edmore Gasura<sup>1</sup>, Peter S. Setimela<sup>2,4</sup>, and Caleb M. Souta<sup>3</sup>

<sup>1</sup>Department of Crop Science, University of Zimbabwe, P.O. Box MP 167, Mt Pleasant, Harare, Zimbabwe (e-mail: egasura@agric.uz.ac.zw); <sup>2</sup>Maize Program, International Maize and Wheat Improvement Center, P.O. Box MP163, Mt Pleasant, Harare, Zimbabwe (e-mail: p.setimela@cgiar.org); and <sup>3</sup>Small grains breeding program, Rattray Arnold Research Station, Seed Co. Pvt Ltd, P.O. Box CH142, Chisipite, Harare, Zimbabwe (e-mail: calebsouta@yahoo.co.uk).

Received 9 April 2014, accepted 30 June 2015. Published on the web 13 July 2015.

Gasura, E., Setimela, P. S. and Souta, C. M. 2015. **Evaluation of the performance of sorghum genotypes using GGE biplot.** *Can. J. Plant Sci.* **95**: 1205–1214. In spite of sorghum's drought tolerance, it is largely affected by genotype  $\times$  environment interaction (GE), making it difficult and expensive to select and recommend new sorghum genotypes for different environments. The objectives of this study were to examine the nature of GE for sorghum grain yield, to identify superior sorghum genotypes for sorghum production environments and determine ideal testing locations for future breeding activities in Zimbabwe. The grain yield of 20 sorghum genotypes from Seed Co. Pvt. Ltd. were evaluated for 2 yr (2011/2012 and 2012/2013 cropping seasons) at five locations in different agro-ecological zones of Zimbabwe. Combined analyses of variance showed significant differences for genotypes ( $P < 0.01$ ), environments ( $P < 0.001$ ) and genotype  $\times$  location ( $P < 0.01$ ). Genotype  $\times$  environment variance component was seven times greater than that of genotypes. Genotype  $\times$  environment interaction was attributed to the variability in the predictable biotic and abiotic factors associated with the different locations. The genotype main effect plus GE biplot showed that the experimental sorghum genotypes W07, W09, W05, G06 and OP46 were high yielding and stable, and possessed other desirable agronomic traits. The most discriminating and representative location was Rattray Arnold Research Station.

**Key words:** Sorghum, genotype  $\times$  environment interaction, stability, ideal testing environment

Gasura, E., Setimela, P. S. et Souta, C. M. 2015. **Évaluation de la performance des génotypes de sorgho grâce au diagramme de double projection GGE.** *Can. J. Plant Sci.* **95**: 1205–1214. Bien qu'il résiste à la sécheresse, le sorgho est très affecté par les interactions génotype  $\times$  environnement (GE), ce qui rend la sélection et la recommandation de nouvelles variétés pour des milieux différents difficiles et onéreuses. L'étude devait établir la nature des interactions GE pour le rendement grainier du sorgho, et permettre l'identification de génotypes de qualité supérieure en vue de la culture du sorgho dans diverses conditions. Elle devait aussi servir à déterminer les meilleurs endroits où effectuer des essais dans le cadre des futurs travaux d'amélioration génétique réalisés au Zimbabwe. Les auteurs ont évalué le rendement grainier de 20 variétés de sorgho venant de la Seed Co. Pvt. Ltd. pendant deux ans (saisons végétatives de 2011/2012 et de 2012/2013), à cinq endroits représentant des zones agro-écologiques distinctes du Zimbabwe. L'analyse combinée de la variance révèle des écarts importants entre les génotypes ( $P < 0,01$ ), les milieux ( $P < 0,001$ ) et les interactions génotype  $\times$  emplacement ( $P < 0,01$ ). La variance de l'élément génotype  $\times$  environnement est sept fois plus importante que celle du génotype. On attribue les interactions génotype  $\times$  environnement à la variabilité des paramètres biotiques et abiotiques prévisibles associés aux divers endroits. L'effet principal du génotype associé à la double projection GE indique que les cultivars expérimentaux W07, W09, W05, G06 et OP46 donnent un rendement élevé et stable, tout en possédant d'autres caractères agronomiques intéressants. L'emplacement le plus représentatif et autorisant la meilleure discrimination est celui de la station de recherche de Rattray Arnold.

**Mots clés:** Sorgho, interaction génotype  $\times$  environnement, stabilité, milieu d'essai idéal

Sorghum is widely used for food, feed and beer brewing (Blum et al. 1997), especially in Africa. The research focus on small grains by the private and public sectors has been intensified in sub-Saharan Africa due to climate change threatening the production of major staple crops like maize, which are less drought tolerant than sorghum (Schlenker and Lobell 2010). In Zimbabwe, Seed Co. Pvt. Ltd., a private seed company, is developing sorghum

varieties ideal for food and/or brewing opaque beer. Generally, white-grained sorghums are desired for food, while red sorghums are ideal for malting because of the tannins, which make the beer bitter. Red sorghums are

**Abbreviations:** AMMI, additive main effect and multiplicative interaction biplot; ANOVA, analysis of variance; E, environment; G, genotype; GE, genotype  $\times$  environment interaction; GGE, genotype main effect plus genotype  $\times$  environment interaction; MET, multi-environmental trial; PC, principal component; SVP, singular value partitioning.

<sup>4</sup>Corresponding author (e-mail: p.setimela@cgiar.org).

also less susceptible to bird damage because they are less palatable and difficult to digest due to the presence of tannins and alcohol-soluble proteins. The majority of farmers tend to like high-yielding and early-maturing sorghums that are less susceptible to common diseases (*Exserohilum turcicum* and *Ramulispora sorghi*). Relatively shorter sorghums with minimal stem lodging are also desired because they can be manually harvested with less effort. Farmers normally would want all these traits incorporated in one variety. In most cases, this broad range of traits is highly correlated genetically, making their simultaneous improvement easy (Gasura et al. 2014). Furthermore, most of the traits are highly heritable and can be easily selected with accuracy during early generation testing. However, the expression of grain yield, a quantitative trait, is greatly influenced by the environment and genotype  $\times$  environment interaction (GE) (Bernardo 2002). Therefore, for breeding purposes it is ideal to classify the production environments into mega-environments to see which varieties are adapted to which mega-environment.

Setimela et al. (2005) grouped maize production environments in southern Africa into different mega-environments based on data from the International Maize and Wheat Improvement Center (CIMMYT) regional maize trials, and the maize-growing areas in Zimbabwe were divided into several mega-environments. Indeed, Zimbabwe has diverse agro-ecological zones that are classified into five natural regions based on their potential for crop production (Rukuni et al. 2006). Sorghum is grown in all these agro-ecological zones, which are highly variable in terms of soil characteristics, rainfall and temperature, among other factors (Nyamapfene 1991; Setimela et al. 2005; Rukuni et al. 2006). The sorghum breeding programs in Zimbabwe therefore target development of new varieties with superior agronomic performance suitable for the five agro-ecological zones.

Although small grains like sorghum are regarded as resilient to various stresses (Schlenker and Lobell 2010), their yield performance is greatly influenced by GE (Bernardo 2002). Genotype  $\times$  environment interaction refers to the differential response of varieties grown in different environments (Finlay and Wilkinson 1963). Genotype  $\times$  environment interaction reduces the correlation between phenotype and genotype and thus reduces responses to selection during breeding (Yan and Kang 2002). Furthermore, GE complicates the recommendation of new varieties from multi-environment trial (MET) and identification of the best testing environment (Yan and Kang 2002). In this regard, several options for dealing with GE have been reported that largely involve reducing or exploiting the interaction (Yan and Kang 2002). Yan and Kang (2002) and Yan and Tinker (2005, 2006) highlighted some objectives of MET analysis that include identification of mega-environments to reduce negative GE, identification of ideal testing locations within mega-environments and identification of superior genotypes.

Superior genotypes should perform well in specific environments or could show yield stability across environments. Yield stability is a measure of the ability of a genotype to maintain relative performance across a wide range of environments (Finlay and Wilkinson 1963). An appropriate stable cultivar is capable of utilizing resources that are available in high-yielding environments, while maintaining above-average yield in other environments, a phenomenon known as the dynamic concept of stability (Yan and Kang 2002).

Several traditional methods of analysing GE have been reviewed (Freeman 1973; Westcott 1986). Some methods, such as analysis of variance, are good at detecting GE but cannot determine the pattern of the interactions. Regression-based methods use environmental scores, which have less to do with genotype plus GE (GGE) and thus explain only a small part of GGE. In the recent past, statistically effective methods, such as biplots based on principal component (PC) analysis, have been developed for GE analysis (Gauch 1993; Crossa et al. 2002; Yan and Kang 2002). Biplot analysis is a multivariate technique that graphically displays two-way data to permit visualization of the interrelationship (Gauch 1993, 2006; Yan and Tinker 2006). The biplot is a graphical presentation of a genotype  $\times$  environment two-way table (Gauch 1993, 2006, 2013). Approaches such as the additive main effect and multiplicative interaction biplot (AMMI) (Gauch 1993, 2006, 2013) and the genotype main effect plus the GE (GGE) biplot (Yan 2001; Yan and Tinker 2005, 2006) have been widely used. The GGE biplot can be subjected to different ways of singular value partitioning (SVP) (Yan and Tinker 2006). The biplot model that is fitted to residuals after the removal of the environmental main effect (environment-centered data) is called a GGE biplot (Crossa and Cornelius 1997; Yang et al. 2009). A GGE biplot is useful in evaluating the genotype main effects plus the GE (Yan and Tinker 2006). This approach has been commonly used in southern Africa for delineating maize production mega-environments and maize variety recommendations (Setimela et al. 2007; Setimela et al. 2010; Kamutando et al. 2013), but has not yet been fully exploited for sorghum breeding. The strengths of the GGE and AMMI methods have been debated unequivocally (Mandel and Gauch 1993; Gauch 2006, 2013; Laffont et al. 2007; Yan et al. 2007; Gauch et al. 2008; Yang et al. 2009). However, the GGE biplot mainly used in this study is capable of capturing much of the G plus GE variation and is also useful in understanding the test environments (Yan and Tinker 2006) as required for rationalizing the scarce resources available for future sorghum breeding programs. The objectives of this study were to examine the nature of the GE for sorghum MET, and use the GGE biplots to identify superior sorghum varieties for the sorghum production environments in Zimbabwe and better understand the sorghum-testing sites for future breeding activities.

## MATERIALS AND METHODS

### Sorghum Materials and Testing Environments

Twenty sorghum genotypes (Table 1) developed in Zimbabwe were evaluated for two seasons in the 2011/2012 and 2012/2013 cropping seasons (Table 2). Seventeen of these were experimental genotypes that were in the advanced stages of breeding. These experimental genotypes were bred for high yield and tolerance to two common diseases (*Ramulispora sorghi* disease and *Exserohilum turcicum* disease) and are suitable for either food or malting for brewing opaque beer. These sorghum genotypes were selected during the early stages of breeding based on these highly heritable traits. The other three genotypes (Sila sorghum, Smile sorghum and Macia sorghum), the commercial varieties grown in Zimbabwe, were obtained from Seed Co. Pvt. Ltd. The commercial variety Sila sorghum is known for its high yield ( $\approx 8 \text{ t ha}^{-1}$ ) under optimal conditions. The variety Smile sorghum is suitable for low-yielding areas, where it can maintain a guaranteed yield of about  $3 \text{ t ha}^{-1}$ . The variety Macia sorghum is an old but widely grown variety, mainly due to its general yield stability even in other countries in southern Africa.

### Experimental Design and Crop Management

At all locations, land was ploughed and disked to a fine tilth to allow ease of germination of small-seeded grain crops like sorghum. The 20 sorghum genotypes were planted in a  $5 \times 4$  rectangular lattice design replicated three times. A plot consisted of four rows of 6 m length and inter-row spacing of 0.75 m. The sorghum genotypes were planted at the rate of  $10 \text{ kg ha}^{-1}$  ( $100\,000 \text{ plants ha}^{-1}$ ).

**Table 1. Names and codes of the sorghum germplasm evaluated**

Entry number	Entry code	Entry name	Grain color
1	W05	SCSV (W) 7105	White
2	W10	SCSV (W) 07110	White
3	Sila	Sila	White
4	W09	SCSV (W) 07109	White
5	W18	SCSV (W) 07118	White
6	W14	SCSV (W) 07114	White
7	W02	SCSV (W) 07102	White
8	W06	SCSV (W) 7106	White
9	W07	SCSV (W) 07107	White
10	G06	SG 06/26	White
11	W19	SCSV (W) 07119	White
12	Macia	Macia <sup>z</sup>	White
13	W04	SCSV (W) 7104	White
14	W15	SCSV (W) 7115	White
15	W17	SCSV (W) 7117	White
16	OP11	SILA (KRC) OP11 (MZ) OP1	White
17	OP25	SILA (KRC) OP25 (MZ) OP1	White
18	OP46	OP46	White
19	OP03	SILA (KRC) OP3	White
20	Smile	SC Smile	Red

<sup>z</sup>Sorghum variety that was bred by the International Research Institute for the Semi-Arid Tropics (ICRISAT) and has been marketed by many companies including Seed Co. Pvt Ltd.

Compound D basal fertilizer (7–14–7, NPK) was applied at the rate of  $200 \text{ kg ha}^{-1}$  at planting followed by top dressing with ammonium nitrate (34.5% N) at flowering stage (68–76 d after planting). Weeds were controlled using both hand hoeing and herbicides (Gramoxone<sup>®</sup> and Atrazine<sup>®</sup> at the rate of  $3 \text{ L ha}^{-1}$ ). Stalk borers were controlled using the insecticide Thionex<sup>®</sup> at the rate of  $1 \text{ L ha}^{-1}$ .

### Data Collection

Data collected on sorghum included: *Exserohilum turcicum* disease scores, *Ramulispora sorghi* disease scores, plant height, days to physiological maturity, percent stem lodging, percent bird damage, and grain yield. *Exserohilum turcicum* disease and *Ramulispora sorghi* disease were scored on a scale of 1 to 5 where 1 is for a plot with healthy plants and 5 for a plot severely diseased. Plant height was measured in centimeters from the base of the plant up to the point where the flag leaf is attached. Days to physiological maturity were recorded from the date of planting until the formation of the black layer on the sorghum grain. Stem lodging was recorded as the proportion of the plants in a plot that showed a stem inclination of  $\geq 45^\circ$  and was expressed as a percentage. Bird damage was recorded as the proportion of plants in a plot with grains eaten by birds and was expressed as a percentage. All plots were carefully hand harvested to minimize shattering and threshed separately. Grain yield was measured in grams per plot and then converted to tonnes per hectare with adjustments to 12.5% moisture content.

### Data Analyses

Genotype  $\times$  location  $\times$  year analysis of variance (ANOVA) was performed on grain yield and related traits using a mixed model (where genotypes and locations were fixed while years, all the interactions involving year, replications, blocks and error were random) in Genstat software version 13 (Genstat 2010). The following model for the combined ANOVA was used:

$$Y_{ijkm(l)} = \mu + r_1(pt)_{jk} + b_m(ptr)_{jkl} + g_i + p_j + t_k + (gp)_{ij} + (gt)_{ik} + (pt)_{jk} + (gpt)_{ijk} + e_{ijkm(l)}$$

where  $Y_{ijkm(l)}$  is the yield of the  $i$ th genotype in the  $j$ th location and the  $k$ th year in the  $m$ th block within the  $l$ th replication,  $\mu$  is the grand mean,  $r_1(pt)_{jk}$  is the effect of the  $l$ th replication within locations and years,  $b_m(ptr)_{jkl}$  is the effect of the  $m$ th block within the  $l$ th replication that is also within locations and years,  $g_i$ ,  $p_j$  and  $t_k$  are the main effects of the genotype, locations and years,  $(gp)_{ij}$ ,  $(gt)_{ik}$ ,  $(pt)_{jk}$  are the first order interactions and  $(gpt)_{ijk}$  is the second-order interaction, and finally  $e_{ijkm(l)}$  is the pooled error term. The terms  $i = 1, 2, 3 \dots 20$ ;  $j = 1, 2, 3, 4, 5$ ;  $k = 1, 2$ ;  $l = 1, 2, 3$  and  $m = 1, 2, 3, 4, 5$ .

The appropriate  $F$ -test for a mixed model that involves fixed genotypes and locations and random years was

carried out as described by McIntosh (1983) and recently by Moore and Dixon (2014). The assumption used for the combined experiments is that the effect of random interactions sum to zero across each level of a fixed factor (Moore and Dixon 2014). In brief, the mean squares for genotypes, genotypes  $\times$  locations, genotypes  $\times$  years and genotypes  $\times$  locations  $\times$  years were tested against the pooled error mean square, while locations, years and location  $\times$  years were tested against the mean square of replications within locations and years (McIntosh 1983). The variance components attributed to locations ( $\delta^2_l$ ), years ( $\delta^2_y$ ), locations  $\times$  years ( $\delta^2_{ly}$ ), genotypes ( $\delta^2_g$ ), genotypes  $\times$  location ( $\delta^2_{gl}$ ), genotypes  $\times$  years ( $\delta^2_{gy}$ ), genotypes  $\times$  locations  $\times$  years ( $\delta^2_{gly}$ ) and random error ( $\delta^2_{error}$ ) were estimated by solving the equations formed by equating the mean squares to their respective expected mean squares (McIntosh 1983; Moore and Dixon 2014). The variance component due to environments (location  $\times$  year combinations) was estimated by summing up  $\delta^2_l$ ,  $\delta^2_y$  and  $\delta^2_{ly}$ , while the variance component due to genotype  $\times$  environment ( $\delta^2_{ge}$ ) was obtained by summing up  $\delta^2_{gl}$ ,  $\delta^2_{gy}$  and  $\delta^2_{gly}$ . The broad sense coefficients of genetic determination (broad sense heritability based on fixed genotypes) on a single plot basis, single environment basis and across environments basis were estimated as  $\delta^2_g/(\delta^2_g + \delta^2_{gl} + \delta^2_{gy} + \delta^2_{gly} + \delta^2_{error})$ ;  $\delta^2_g/(\delta^2_g + \delta^2_{gl} + \delta^2_{gy} + \delta^2_{gly} + \delta^2_{error}/nr)$  and  $\delta^2_g/(\delta^2_g + \delta^2_{gl}/nl + \delta^2_{gy}/ny + \delta^2_{gly}/nly + \delta^2_{error}/nylr)$ , respectively, where nr = number of replications, nl = number of locations, ny = number of years, nly = number of location  $\times$  year combinations and nylr is the number of years  $\times$  location  $\times$  replications.

In order to determine the number of principal components to retain during GGE biplot analyses, a post-dictive evaluation was done for model fitting using Gollob's (1968) *F*-test (Zobel et al. 1988; Dias and Krzanowski 2003; Gauch 2013). Gollob's (1968) *F*-test (Zobel et al. 1988; Gauch 2013) showed that the two principal components of the biplot were significant and thus could explain much of the variation (67%) in the two-way data. Therefore, a GGE-2 biplot analysis (Yan and Tinker 2006) was conducted using Genstat Software version 13 (Genstat 2010). The GGE biplot model was described by Yan et al. (2000), Yan and Hunt (2001) and Yan (2002) as:

$$Y_{ij} - \mu - \beta_j = \sum_{l=1}^k \lambda_l \xi_{il} \eta_{jl} + \varepsilon_{ij}$$

where  $Y_{ij}$  is the mean yield of the  $i$ th genotype in the  $j$ th environment,  $\mu$  is the grand mean,  $\beta_j$  is the main effect of the environment  $j$ ,  $\lambda_l$  is the singular value of the  $l$ th principal component and  $k = 2$  in this case,  $\xi_{il}$  is the eigen vector of the genotype  $i$  for PC  $l$ ,  $\eta_{jl}$  is the eigen vector of environment  $j$  for PC  $l$ ,  $\varepsilon_{ij}$  is the residual associated with genotype  $i$  in the environment  $j$ .

The which-won-where scatter biplot (for mega-environment delineation), genotype comparison biplot (for comparing genotypes based on mean yield and

stability) and location comparison biplot (for identifying the most discriminating and representative locations) were generated using the appropriate SVP methods (Yan 2002). In the scatter biplot, the polygon view displaying the which-won-where pattern was formed by connecting the genotype markers furthest away from the biplot origin such that the polygon contained all other genotypes (Yan 2002). The polygon was then dissected by straight lines perpendicular to the polygon sides and running from the biplot origin. Visualization of the mean and stability of genotypes using a genotype comparison biplot was achieved by representing an average environment by an arrow. A line that passes through the biplot origin to the average environment (average genotype axis) was drawn followed by a perpendicular line that passes through the biplot origin. For the analyses of test location, the environment vectors were drawn from the biplot origin to the markers of the environment. The average environment (AE) was represented with a small arrow and a line from the biplot origin to the AE (average environment axis) was drawn followed by a perpendicular line that passes through the biplot origin.

For each genotype, the grand means of all traits across environments that include yield, percent bird damage, percent stem lodging, *Exserohilum turcicum* disease score, *Ramulispora sorghi* disease score, plant height and days to physiological maturity were calculated. This formed a two-way table of genotypes  $\times$  traits means. The cross-environment mean data were scaled (standardized) by dividing each trait mean value with the within-trait standard deviation, as outlined by Yan and Tinker (2006). The standardization helps to remove the different units found among different traits (Yan et al. 2000; Yan 2001). The resultant data were subjected to biplot analysis using the trait focused SVP method and the data were trait-centered. The vectors were drawn to connect the specific traits from the biplot origin. The across-environment means of the seven traits studied were subjected to Pearson's phenotypic correlation coefficient analyses using Genstat software 13th ed. (Genstat 2010). The absence of statistical assessments of the biplots suggests that all the biplot outputs should be interpreted with caution as they are descriptive tools with limited capacity for hypothesis testing (Yang et al. 2009).

## RESULTS

### Combined ANOVA, Variance Components and Heritability Estimates

The combined ANOVA for sorghum grain yield showed significant differences for locations ( $P < 0.001$ ), years ( $P < 0.001$ ), location  $\times$  year interaction ( $P < 0.01$ ), genotypes ( $P < 0.01$ ), and the genotypes  $\times$  locations interaction ( $P < 0.001$ ) (Table 3). However, the genotype  $\times$  years and genotypes  $\times$  locations  $\times$  years interactions were non-significant (Table 3) and also contributed much less to

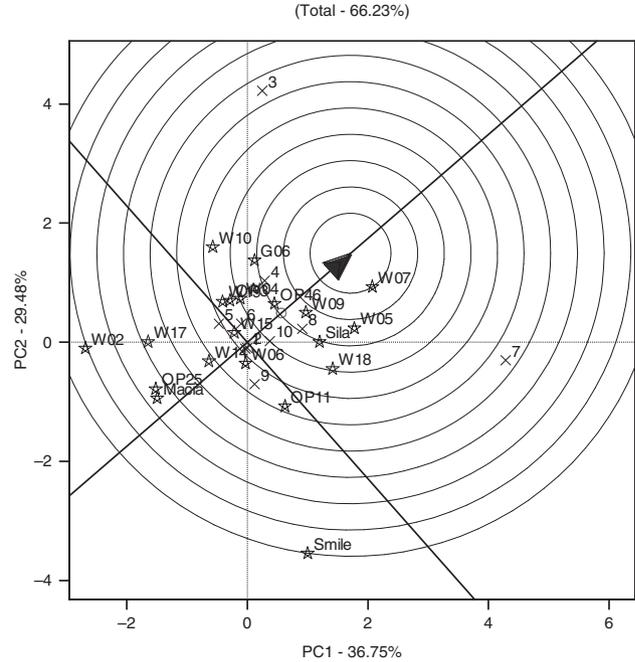
the variance component due to genotype  $\times$  environment interaction (Table 3).

The variance component due to environments was the largest at 9.37, followed by the  $\delta^2$ error at 0.78, then  $\delta^2$ ge at 0.2 and last  $\delta^2$ g at 0.03 (Table 3). The  $\delta^2$ y had the highest contribution to the environmental variance followed by  $\delta^2$ l (Table 3) and much less came from  $\delta^2$ ly. Importantly,  $\delta^2$ ge was seven (7) times larger than  $\delta^2$ g (Table 3). The broad sense coefficients of genetic determination estimates (equivalent to broad sense heritability of fixed genotypes) on a single plot basis, single environment basis and across environments basis were 2.8, 5.8 and 31.8%, respectively.

**Genotype Evaluation Based on GGE Biplots**

Model diagnosis (fitting) showed that the first two PCs were significant. PC1 and PC2 explained 36.8 and 29.5% of the G + GE sum of squares, respectively, while 33.7% was attributed to noise. This justifies the use of GGE-2 biplots to explain the G plus GE. The which-won-where biplot showed different winning genotypes in different environments (Fig. 1). However, genotype W07 won in most of the environments. The mega-environments in Fig. 1 were all overlapping. The GGE biplot analysis of individual years revealed that the crossover interaction observed in Fig. 1 was not repeatable in individual years (results not shown).

The genotype comparison biplot ranked genotypes according to their mean yield and stability as W07 >



**Fig. 2.** A genotype comparison biplot showing the best sorghum genotypes based on mean performance and stability across 10 environments (five locations and 2 yr). To avoid congesting the graph, the environments are numbered 1–10, while genotypes are shown by their codes. The biplot was produced based on genotype focused SVP, no scaling, no transformation and the data were environment centered.

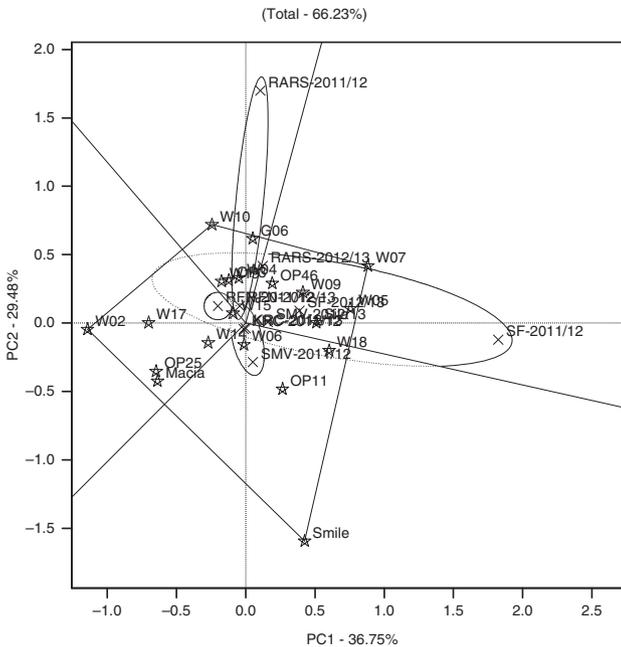
W09 = W05 > OP46 = G06 = Sila sorghum followed by the rest of the experimental genotypes and the check varieties, Macia sorghum and Smile sorghum (Fig. 2).

**Test Location Evaluation Based on GGE Biplots**

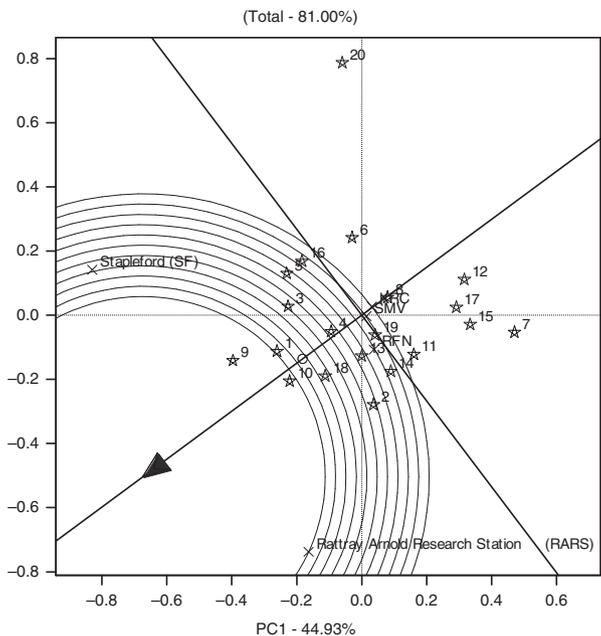
The locations at Stapleford (SF) and Rattray Arnold Research Station (RARS) had the longest vectors from the biplot origin. The angle between these locations is almost a right angle (Fig. 3). However, locations KRC, RF and SMV had very short vectors. Locations KRC and SMV had small angles between them (Fig. 3). RARS and SF were more discriminating, while other locations were not discriminating at all (Fig. 3). The RARS location was found closer to the average environment followed by SF (Fig. 3).

**Trait and Genotype  $\times$  Trait Evaluation Based on the GGE Biplot and Correlation Analysis**

The genotype means across environments are shown in Table 4. The variance components for genotypes were lower than for GE for all traits except plant height (Table 4). However, broad sense heritability (repeatability) based on genotype means across environments was high as it was above 65% for all traits except grain yield and percent stem lodging (Table 4). The genotype  $\times$  trait biplot showed that some traits were correlated (Fig. 4).

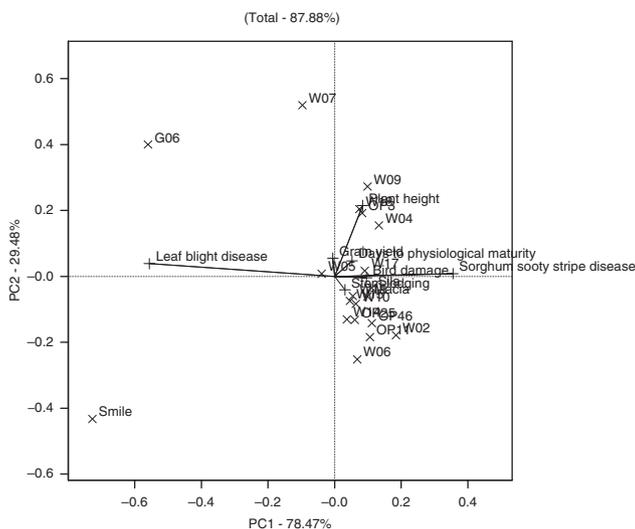


**Fig. 1.** The which-won-where and mega-environment delineation biplot for the 20 sorghum genotypes evaluated in five locations for 2 yr. The biplot was produced based on symmetric focused SVP, no scaling, no transformation and the data were environment centered.



**Fig. 3.** The ideal testing location for sorghum among the locations used in evaluations. The biplot was produced based location focused SVP, no scaling, no transformation and the data were location centered. To avoid congesting the graph, the genotypes are numbered 1–20, while locations are shown by their codes.

This was also supported by the Pearson’s correlation analysis that showed significant negative correlation between *Exserohilum turcicum* and sorghum stripe disease ( $r = -0.925$ ;  $P < 0.001$ ). Percentage bird damage



**Fig. 4.** Genotype × trait biplot showing genotypes with a given set of traits. Data were standardized with the within-trait standard deviation. The biplot was produced based on trait focused SVP, no scaling, no transformation and the data were trait centered.

was negatively correlated with *Exserohilum turcicum* ( $r = -0.66$ ,  $P < 0.01$ ), but positively correlated with *Ramulispora sorghi* disease ( $r = 0.625$ ,  $P < 0.01$ ). However, other traits were not correlated.

### DISCUSSION

#### Variance Components, Nature of the GE and Heritability Estimates

The presence of significant genotype main effect and GE suggest differential responses of the genotypes and the need to identify high-yielding and stable genotypes across the test environments. The large variance component due to environment alone justifies the need to use the GGE biplot, which captures much of the variation due to G plus GE as a fraction of the total variation ( $G + E + GE$ ) (Yan et al. 2007). The large variance component due to locations and years shows that the locations used in this study were diverse across years. Indeed, Zimbabwe is sub-divided into five natural agro-ecological zones based on crop production potential (Nyamapfene 1991; Rukuni et al. 2006). In these agro-ecological zones there is huge variability in predictable factors (soil characteristics) and unpredictable factors (temperature and rainfall) (Table 2).

The presence of GE that is seven times larger than the genotypic main effects indicates substantial differences in genotypic responses across locations. The large GE and error variance components observed in this study present challenges in breeding and selecting for adapted genotypes. Large GE and error terms affect the repeatability of the experiment (Annicchiarico et al. 2000) thus confounding the efforts in variety breeding, selection and recommendations (Yan and Kang 2002). The large GE and error variance components might have resulted in low heritability (2.8%) on a single plot basis. However, heritability across environments increased (31.8%) as the number of locations and years increased. This shows that large GE and error variance components increase the evaluation cost due to the large numbers of replications, locations and years needed to improve heritability, and hence the selection efficiency.

When the GE is substantial it should not be ignored; the causes must be identified (Yan and Kang 2002) and the GE should be addressed (Bernardo 2002). Most of the GE variation in this study was due to predictable factors (location-intrinsic factors) rather than unpredictable factors (years). In this study, the GE could be attributed to differences in soil types, rainfall patterns, and temperatures (Table 2), as well as various diseases found in specific locations. Under such circumstances, predictable factors can be managed, but modifying the environment to suit the crop will further constrain the resource-poor farmers, who are already struggling to obtain inputs. The cheaper option is to develop sorghum varieties adapted to the target environments. However, because the locations have no clearly defined boundaries and farmers tend to influence each other in the choice

**Table 2. Description of the sorghum-testing locations in Zimbabwe**

Location code <sup>a</sup>	Location name	Natural region	Altitude (masl)	Average rainfall (mm)	Soil type
RARS	Ratray Arnold Research Station	2a	1341	750–1000	Red Clay
SF	Stapleford Research Centre	2a	1492	750–1000	Red Clay
RFN	Raffingora (On-farm)	2b	1200	750–800	Sandy Loam
SMV	Panmure Research Station	2b	881	650–800	Red Clay
KRC	Kadoma Research Centre	3	1149	650–800	Red Clay

<sup>a</sup>In the biplots, location codes with the suffix 2011/12 refer to the first year while location codes with the suffix 2012/13 refer to the second year.

of variety that is grown, the development of varieties with broad adaptation is strongly supported, rather than location-specific varieties. The absence of a genotype  $\times$  year interaction is expected in sorghum because the major variation across years is rainfall, and sorghum is known to be drought tolerant due to its capacity to remobilize stem reserves (Blum et al. 1997). Indeed, Yan and Rajcan (2003) noted that a single year was good enough for predicting the performance of soybean cultivars. Therefore, the non-significance of genotype  $\times$  year and genotype  $\times$  location  $\times$  year interactions makes it easier to recommend sorghum cultivars without extensive testing across years.

#### Evaluation of Sorghum Genotypes across Environments

The GGE of the 20 sorghum genotypes evaluated in 10 environments was best predicted by the first two PCAs based on Gollob's *F*-test (Zobel et al. 1988). Therefore, a biplot with two PCAs was used to describe the GGE. Yan (2000) reported that the first two PCs captured the most useful variation in a biplot. When the GE is larger than the genotype main effect, ignoring the interaction is not recommended (Yan and Kang 2002). In our study, the GGE biplots explained 67% of the total GGE sum of squares, while G plus GE explained over 5% of the total variance (G + E + GE). Yan and Tinker (2005) highlighted such complex situations as being common in MET data. However, irrespective of such complex situations most patterns in the two-way data are revealed by the first two PCs (Yan and Tinker 2006). Based on this reasoning, the number of PCs used in our study can be considered as adequate, particularly when coupled with Gollob's *F*-test, which also suggested two PCs.

The success of different genotypes in different environments shows the existence of crossover genotype  $\times$  environment interactions (Yan and Tinker 2006). The mega-environments in Fig. 1 are overlapping, suggesting the absence of a clear pattern of GEs. The GE is of concern if it is of the crossover type (Baker 1988; Yan and Kang 2002) because it reduces heritability and compromises cultivar recommendations (Bernardo 2002). The existence of crossover GE suggests that cultivar evaluation and recommendation based on any single location is unreliable because there is differential performance of varieties across locations (Yan and Kang 2002). However, the non-repeatable crossover interaction across

years suggests no need to delineate mega-environments (Yan and Kang 2002; Yan and Tinker 2005, 2006) especially based on the data from these relatively few and fixed locations (Table 2).

The existence of crossover interactions suggests the need to reduce or exploit GE. When the crossover genotype  $\times$  environment interaction is non-repeatable over years, Yan and Kang (2002) suggested that there is no need to delineate mega-environments, but that cultivar evaluation should be based on mean performance and stability. The genotype comparison biplot with genotype focused SVP shows that the most stable and high-yielding genotype is W07, followed by W09 and W05 (Fig. 2). These three genotypes yielded higher and were more stable than the commercial variety, Sila sorghum. Based on their stability and high yield, these three varieties were outstanding in the majority of the environments (Fig. 1). Such varieties will display stability and outstanding yield performance in high yield environments, which is highly desirable (Finlay and Wilkinson 1963). Genotypes G06 and OP46 were comparable in yield performance to the commercial variety Sila sorghum. These two genotypes, together with the top three genotypes, should be selected for advanced variety trials and commercialization. Other commercial varieties did not perform well (Fig. 2). Sila sorghum was ranked fourth in terms of mean performance and stability, while Smile sorghum had high yield

**Table 3. Mean squares and variance components for the grain yield ( $\text{t ha}^{-1}$ ) of 20 sorghum genotypes grown across locations and years**

Source	Degrees of freedom	Mean square	Variance components
Locations	4	359.4148***	2.88
Years	1	1588.8578***	5.25
Locations $\times$ Years	4	88.0249**	1.24
Location $\times$ Year (Replications)	20	13.3835***	0.61
Location $\times$ Year (Replications, Blocks)	120	1.2833***	0.03
Genotypes	19	1.6244**	0.03
Genotypes $\times$ Locations	76	1.5479***	0.13
Genotypes $\times$ Years	19	0.8586NS	0.01
Genotypes $\times$ Locations $\times$ Years	76	0.9660NS	0.06
Error	260	0.7778	0.78
Total	599	7.0797	

\*\*, \*\*\* Significant at the 1% and 0.1% probability levels, respectively; NS, non-significant.

**Table 4. Mean values and related genetic parameters of seven traits studied in 20 sorghum genotypes evaluated across 10 environments**

Genotype code	Grain yield (t ha <sup>-1</sup> )	Days to 50% physiological maturity <sup>z</sup>	Plant height (cm)	Stem lodging (%) <sup>y</sup>	Bird damage (%) <sup>y</sup>	<i>Exserohilum turcicum</i> disease score	<i>Ramulispora sorghi</i> disease score
G06	3.96	122.50	155.50	1.21	7.23	2.73	1.41
Macia	3.06	123.60	148.80	2.05	20.24	1.27	3.19
OP11	3.87	123.40	141.60	2.08	14.10	1.21	3.31
OP25	3.48	123.10	145.00	2.71	17.23	1.41	3.32
OP3	3.74	125.90	152.00	2.08	12.39	1.43	3.49
OP46	3.98	124.10	142.10	1.94	12.63	1.17	3.25
Sila	3.83	123.20	148.70	2.81	14.41	1.24	3.32
Smile	3.35	123.80	120.80	1.89	6.46	3.22	1.51
W02	3.29	123.80	144.40	2.43	11.79	1.07	3.61
W04	3.76	125.70	151.20	1.63	14.66	1.27	3.51
W05	4.04	124.20	144.70	2.23	16.96	1.70	3.17
W06	3.55	123.80	139.40	2.28	13.91	1.29	3.18
W07	4.05	125.20	161.00	1.55	12.36	1.92	3.12
W09	3.66	124.90	157.00	1.81	18.29	1.37	3.36
W10	3.70	123.90	145.80	2.59	18.81	1.33	3.07
W14	3.23	126.60	139.40	1.33	15.27	1.44	3.19
W15	3.75	124.10	145.30	3.05	12.80	1.43	3.28
W17	3.54	125.20	147.00	1.14	14.64	1.24	3.13
W18	3.83	125.40	152.80	1.99	11.81	1.42	3.41
W19	3.56	124.80	143.90	1.16	13.50	1.33	3.11
Grand mean	3.66	124.36	146.32	2.00	13.97	1.52	3.10
<i>F</i> probability	NS	0.042	<0.001	NS	0.033	<0.001	<0.001
Least significant difference (5%)	—	1.43	6.91	—	6.15	0.59	0.31
Coefficient of variation (%)	24.10	1.63	8.02	81.23	57.66	20.76	16.81
$\delta^2_g$	0.03	1.14	47.90	0.00	14.50	0.17	0.21
$\delta^2_{ge}$	0.20	1.93	26.40	0.40	11.05	0.34	0.31
$\delta^2_{error}$	0.78	4.09	137.70	2.63	64.94	0.10	0.27
Repeatability based on single plot basis	2.80	0.16	0.23	0.00	0.16	0.28	0.27
Repeatability on individual environment basis	5.80	0.26	0.40	0.00	0.31	0.32	0.34
Repeatability on across environments basis	31.80	0.67	0.87	0.00	0.69	0.82	0.84

<sup>z</sup>Trait recorded in six environments.<sup>y</sup>Trait recorded in five environments.

NS, non-significant; negative genotypic variance component for stem lodging was regarded as zero.

but relatively poor stability and Macia sorghum had low yield but was relatively stable. The top five experimental genotypes are potential candidates for release. Varieties such as W07, which performed well in specific environments, could be targeted to those environments to maximize grain yield. Interestingly, the same genotype was the most stable across all environments. This sorghum variety can be recommended for release in Zimbabwe and also for use in the sorghum breeding program.

The genotypes W05 and W09 have a similar mean yield across environments as shown in Fig. 2. Therefore, it is logical to release the better of the two.

#### *Correlation among Traits and Genotypes Associated with Particular Traits*

Interestingly, the high-yielding and stable genotypes were also correlated with other desirable traits (Fig. 4, Table 4). The high heritability values (>65%) for most traits (Table 4) make the genotype  $\times$  trait biplot more relevant in examining the relationship between genotypes and traits across environments (Fig. 4). The best genotype, W07, has desirable traits including high yield, earliness, less bird damage, and less stem lodging, as well as reduced attack from diseases such as *Ramulispora sorghi* and *Exserohilum turcicum*. Because genotype W07 possessed most of the desirable traits, this reinforces the need for the selection of this genotype for release. Genotypes W09 and W07 have similar characteristics, except that it takes relatively longer days for W07 to reach physiological maturity (Table 4). Genotypes that have delayed physiological maturity are not attractive to farmers, but they are much less prone to bird damage. They are therefore suitable for large-scale commercial production so that bird scaring can be reduced. Genotypes W05 and OP46 have average attributes of all the traits studied. This makes them ideal genotypes for release. Genotype G06 has moderate resistance to sorghum *Exserohilum turcicum* disease. However, it has all the other desirable traits, making it an ideal candidate for release.

The presence of significant correlation among traits (Fig. 4) is desirable because it can facilitate indirect selection (Gasura et al. 2014). In general, plants that take long days to mature are taller among the cereals. In maize, most early genotypes are shorter than late-maturing ones (Gasura et al. 2013). Greater plant height results in more leaves and thus could explain the greater yield in these relatively tall varieties. The correlation of bird damage to physiological maturity could be attributed to the fact that birds attack early crops heavily because there will be no other crops around. Because quaila birds common to southern Africa flock in large numbers, their activities also lead to stem lodging, and this results in close relationships among earliness (days to physiological maturity), bird damage and stem lodging.

The white sorghum genotypes are prone to bird damage due to their palatability and lack of tannins. The high tannin levels and the large proportion of prolamine

(an alcohol-soluble protein) reduce the palatability and digestibility of the brown sorghum, “Smile sorghum”, making it less preferred as food. Nonetheless, given its benefits in malting, it is worth continuing to grow it. Our results show that all the commercial check varieties are of inferior yield and stability compared with the newly developed materials. However, the new varieties W09 and W07 seem to be slightly taller than the original varieties, while W05 is still within the average range of heights. However, none of the potential varieties exceeded 160 cm tall, suggesting that they are still in a range in which the majority of farmers, especially women carrying babies on their backs, can still harvest the crop.

#### *Evaluation of the Test Environments*

The presence of GE in sorghum justifies undertaking METs during cultivar breeding and selection, but this will greatly increase the cost of evaluation. Therefore, environmental analysis will help to better understand the testing environments and possibly help reduce the cost of genotype evaluations (Yan and Kang 2002). For example, the vector length of the biplot approximates the standard deviation within each location, a measure of the discriminating ability of the location (Yan and Tinker 2006). The RARS and SF locations, which had the longest vectors from the biplot origin (Fig. 4), were the most discriminating sites. Therefore these two sites can be used jointly as discriminating locations during early generation testing. Discriminating and representative environments are best for selecting superior genotypes while culling inferior ones. The right angle between the locations RARS and SF (Fig. 3) shows that there is no correlation between these two locations ( $r = \text{Cos } 90 = 0$ ) (Yan and Kang 2002). Such lack of correlation is a crucial parameter in selecting locations for future evaluation. Locations that are highly correlated are redundant in METs because these locations do not add much information about cultivar performance, but rather increase the cost of evaluation (Yan and Kang 2002).

In this study, the RARS location is the ideal testing environment for sorghum because it is closer to the average environment (Fig. 3). Early generation testing at RARS will reduce breeding costs by reducing the number of test varieties to be evaluated in multi-location trials as well as inherently reducing the number of test locations. Because Seed Co. Pvt. Ltd. has its research station at RARS, using this location as an ideal testing environment has an added advantage of proximity to essential experimental resources.

#### **CONCLUSIONS**

There is a large effect of the genotype  $\times$  environment interaction on sorghum grain yield, which was seven times larger than the effect of genotypes. We recommend W07, W09, W05, OP46 and G06 as the most stable and high-yielding varieties that could be commercialized in Zimbabwe as better substitutes for the old varieties. These varieties also displayed the least damage

from *Ramulispora sorghi* and *Exserohilum turcicum*, have less stem lodging and are relatively early in maturity. Rattray Arnold Research Station (RARS) is an ideal test location that can be used to select superior genotypes during sorghum breeding and testing in order to reduce operational costs.

### ACKNOWLEDGEMENTS

We are grateful to Seed-Co. Pvt. Ltd. for financial support and the International Research Institute for the Semi-Arid Tropics (ICRISAT) for the provision of some breeding materials used in the experiments.

**Annicchiarico, P., Pecetti, L., Boggini, G. and Doust, M. A. 2000.** Repeatability of large-scale germplasm evaluation results in durum wheat. *Crop Sci.* **40**: 1810–1814.

**Bernardo, R. 2002.** Breeding for quantitative traits in plants. Stemma Press, St. Paul, MN. 369 pp.

**Blum, A., Golan, G., Mayer, J. and Sinmena, B. 1997.** The effect of the dwarfing genes on sorghum grain-filling from remobilized stem reserves under stress. *Field Crops Res.* **52**: 43–54.

**Crossa, J. and Cornelius, P. L. 1997.** Site regression and shifted multiplicative model clustering of cultivar trials sites under heterogeneity of error variances. *Crop Sci.* **37**: 406–415.

**Crossa, J., Cornelius, P. L. and Yan, W. 2002.** Biplots of linear-bilinear models for studying crossover genotype  $\times$  environment interaction. *Crop Sci.* **42**: 619–633.

**Dias, C. T., dos, S. and Krzanowski, W. J. 2003.** Model selection and cross validation in additive main effect and multiplicative interaction models. *Crop Sci.* **43**: 865–873.

**Finlay, K. W. and Wilkinson, G. N. 1963.** The analysis of adaptation in a plant breeding programme. *Austin J. Agric. Res.* **14**: 742–754.

**Freeman, G. H. 1973.** Statistical methods for the analysis of genotype–environment interactions. *Heredity* **31**: 339–354.

**Gasura, E., Setimela, P. S., Edema, R., Gibson, P. T., Okori, P. and Tarekegne, A. 2013.** Exploiting grain-filling rate and effective grain-filling duration to improve grain yield of early-maturing maize. *Crop Sci.* **53**: 1–9.

**Gasura, E., Setimela, P. S., Tarekegne, A., Icishahayo, D., Edema, R., Gibson, P. T. and Okori, P. 2014.** Variability of grain-filling traits in early maturing CIMMYT tropical maize inbred lines. *Crop Sci.* **54**: 530–536.

**Gauch H. G., Jr. 1993.** Prediction, parsimony and noise. *Am. Sci.* **81**: 468–472.

**Gauch H. G., Jr. 2006.** Statistical analysis of yield trials by AMMI and GGE. *Crop Sci.* **46**: 1488–1500.

**Gauch H. G., Jr. 2013.** A simple protocol for AMMI analysis of yield trials. *Crop Sci.* **53**: 1860–1869.

**Gauch H. G., Jr., Piepho, H. and Annicchiarico, P. 2008.** Statistical analysis of yield trials by AMMI and GGE: Further considerations. *Crop Sci.* **48**: 866–889.

**GenStat. 2010.** GenStat 13th ed. VSN International Ltd., Hemel Hempstead, UK.

**Gollob, H. F. 1968.** A statistical model which combines features of factor analytic and analysis of variance techniques. *Psychometrika* **33**: 73–115.

**Kamutando, C. N., Muungani, D., Masvodza, D. R. and Gasura, E. 2013.** Exploiting genotype  $\times$  environment interaction in maize breeding in Zimbabwe. *Afr. J. Agric. Res.* **8** (29): 4058–4066.

**Laffont, J. L., Hanafi, M. and Wright, K. 2007.** Numerical and graphical measures to facilitate the interpretation of GGE biplots. *Crop Sci.* **47**: 990–996.

**Mandel, J. and Gauch, H. G., Jr. 1993.** Non-additivity and biplots. *Biometrics* **49** (3): 952–954.

**McIntosh, M. S. 1983.** Analysis of combined experiments. *Agron. J.* **75**: 153–155.

**Moore, K. J. and Dixon, P. M. 2014.** Analysis of combined experiments revisited. *Agron. J.* **106**: 1–9.

**Nyamapfene, K. 1991.** Soils of Zimbabwe. Nehanda Publishers. Harare, Zimbabwe.

**Rukuni, M., Tawonezwi, P., Eicher, C., Munyuki-Hungwi, M. and Matondi, P. 2006.** Zimbabwe's agricultural revolution. University of Zimbabwe publications, Sable Press, Harare, Zimbabwe.

**Schlenker, W. and Lobell, D. B. 2010.** Robust negative impacts of climate change on African agriculture. *Environ. Res. Lett.* **5**: 1–8.

**Setimela, P., Chitalu, Z., Jonazi, J., Mambo, A., Hodson, D. and Banziger, M. 2005.** Environmental classification of maize-testing locations in the SADC region and its implication for collaborative maize breeding strategies in the subcontinent. *Euphytica* **145**: 123–132.

**Setimela, P. S., Crossa, J. and Banziger, M. 2010.** Targeting of early to intermediate maize hybrids for yield performance and yield stability using SREG model. *S. Afr. J. Plant Soil* **27** (3): 207–214.

**Setimela, P. S., Vivek, B., Banziger, M., Crossa, J. and Maiden, F. 2007.** Evaluation of early to medium maturity open pollinated maize varieties in SADC region using GGE biplot based on the SREG model. *Field Crops Res.* **103**: 161–169.

**Westcott, B. 1986.** Some methods of analysing genotype–environment interactions. *Heredity* **56**: 243–253.

**Yan, W. 2001.** GGEbiplot-A windows application for graphical analysis of multi-environment trial data and other types of two-way data. *Agron. J.* **93**: 1111–1118.

**Yan, W. 2002.** Singular-value partitioning in biplot analysis of multi-environment trial data. *Agron. J.* **94**: 990–996.

**Yan, W. and Hunt, L. A. 2001.** Interpretation of genotype  $\times$  environment interaction for winter wheat yield in Ontario. *Crop Sci.* **41**: 19–25.

**Yan, W. and Kang, M. S. 2002.** GGE biplot analysis: A graphical tool for breeders, geneticists, and agronomists. CRC Press, London, UK.

**Yan, W. and Rajcan, I. 2003.** Prediction of cultivar performance based on single-versus multiple-year tests in soybean. *Crop Sci.* **43**: 549–555.

**Yan, W. and Tinker, N. A. 2005.** An integrated biplot analysis system for displaying, interpreting, and exploring genotype  $\times$  environment interaction. *Crop Sci.* **45**: 1004–1016.

**Yan, W. and Tinker, N. A. 2006.** Biplot analysis of multi-environment trial data: Principles and applications. *Can. J. Plant Sci.* **86**: 623–645.

**Yang, R., Crossa, J., Cornelius, P. L. and Burgueno, J. 2009.** Biplot analysis of genotype  $\times$  environment interaction: Proceed with caution. *Crop Sci.* **49**: 1564–1576.

**Yan, W., Hunt, L. A., Sheng, Q. and Szlavnics, Z. 2000.** Cultivar evaluation and mega-environment investigation based on the GGE biplot. *Crop Sci.* **40**: 597–605.

**Yan, W., Kang, M. S., Ma, B., Woods, S. and Cornelius, P. L. 2007.** GGE biplot vs. AMMI analysis of genotype-by-environment data. *Crop Sci.* **47**: 641–653.

**Zobel, R. W., Wright, M. J. and Gauch, H. G. 1988.** Statistical analysis of a yield trial. *Agron. J.* **80**: 388–393.