Published May 2007
Published online May 31, 2007

RESEARCH

Numerical and Graphical Measures to Facilitate the Interpretation of GGE Biplots

Jean-Louis Laffont, Mohamed Hanafi, and Kevin Wright*

ABSTRACT

The genotype + genotype-by-environment (GGE) biplot technique has been widely used in the recent years for the analysis of multienvironment trials, as is evident by the large number of articles published where there is a reference to this technique. One question often raised by the users of this technique is how much of genotype and/or genotype-by-environment variability is captured by the GGE biplot axes. This article provides an answer to this question by establishing a link between the partitioning of the total sum of squares (TSS) of the genotype-by-environment-centered matrix provided by singular value decomposition and the partitioning of this TSS provided by the analysis of variance technique. An artificial dataset is used to illustrate this link, which is visualized through a mosaic plot. This new GGE biplot interpretation tool is found to be useful and is discussed in contrast with other interpretation tools.

J.-L. Laffont, Pioneer Génétique, Chemin de l'Enseigure, 31840 Aussonne, France; M. Hanafi, ENITIAA-INRA, Unité de Sensométrie et Chimiométrie, Rue de la Géraudière, BP 82225, 44322 Nantes Cedex 03, France; K. Wright, Pioneer Hi-Bred Int., 7300 NW 62nd Ave., Johnston, IA 50156. Received 29 Aug. 2006. *Corresponding author (kevin.d.wright@pioneer.com).

Abbreviations: AEC, average environment coordinate; G, genotype; GE, genotype-by-environment; GGE, genotype + genotype-by-environment; SSG, sums of squares due to genotypes; SSGE, sums of squares due to genotype-by-environment; SVD, singular value decomposition; TSS, total sum of squares.

Multienvironment trials are widely conducted to identify superior genotypes (G) in a target population of environments. The data generated from these trials can be rearranged in a two-way genotype-by-environment (GE) table where each cell is the result of one genotype in one environment for a particular trait (e.g., yield), this result being either a raw data point or being estimated when replications are available (simple means or more sophisticated estimates from linear models). When analyzing such tables, it is important to correctly assess both the G effects and the GE effects, and this can be challenging for large tables.

Gabriel (1971) developed the biplot that simultaneously displays the rows and the columns of a matrix. Multiple versions of this graphical multivariate technique have found wide applicability and acceptance by plant breeders for analysis of GE tables (Gauch, 2006; Yan and Tinker, 2006). One particular version is the genotype + genotype-by-environment (GGE) biplot, which simultaneously plots information about G and GE interactions. The term "GGE biplot" comes from Yan et al. (2000), but the

Published in Crop Sci. 47:990–996 (2007).
doi: 10.2135/cropsci2006.08.0549

677 S. Segoe Rd., Madison, WI 53711 USA

technique has appeared previously in sources such as Kempton (1984) and Cooper and DeLacy (1994).

A biplot can be constructed with the use of singular value decomposition (SVD) (Kroonenberg, 1997). Let $\mathbf{Y}$ be an n by p matrix of rank $r \le \min\{n, p\}$. The SVD of $\mathbf{Y}$ is given by $\mathbf{Y} = \mathbf{U\Lambda V'}$, where $\mathbf{\Lambda}$ (n by n) is a diagonal matrix with singular values $\lambda_1 \ge \lambda_2 \ge \ldots \ge \lambda_r > \lambda_{r+1} = \ldots = \lambda_n = 0$, and $\mathbf{U}$ (n by n) and $\mathbf{V}$ (p by n) are orthogonal matrices. Let $\mathbf{G} = \mathbf{U\Lambda}^{\alpha}$ and let $\mathbf{H} = \mathbf{V\Lambda}^{1-\alpha}$ for $0 \le \alpha \le 1$. A two-dimensional biplot uses the first two columns of $\mathbf{G}$ for the genotype coordinates and the first two columns of $\mathbf{H}$ for the environment coordinates. The value of the parameter α can give different types of biplots with different interpretation properties (Gabriel, 1971; Jackson, 1991, p. 199–202). Yan and Kang (2003, p. 42–49) discussed the particular properties of the GGE biplot when considering different values for α (which they call f_j). For this article, $\alpha = 0$ will be used, which preserves the Euclidean distances between the columns (environments) of $\mathbf{Y}$. This is called "environment-focused" scaling in the terminology of Yan and Kang (2003). When the columns of $\mathbf{Y}$ are centered (and possibly standardized), the SVD of $\mathbf{Y}$ produces a principal component analysis of the covariance matrix of $\mathbf{Y}$.

The total sum of the squared singular values is equal to the total sum of squares (TSS) variability in $\mathbf{Y}$. Any squared singular value divided by TSS will give the portion of the variability in $\mathbf{Y}$ accounted for by the corresponding biplot axis. This provides one partition of TSS. Another TSS partition is the one usually summarized in an ANOVA table, TSS = SSG + SSGE, where SSG is the sum of squares attributable to G and SSGE is the sum of squares attributable to GE. (Due to the columnar centering of the matrix, the sum of squares for environments [SSE] is 0.) Establishing a link between these two partitions of TSS will allow for an interpretation of the biplot axes in terms of a portion of G and GE variability accounted for by the axes. This article establishes this link and provides additional interpretation tools of the GGE biplot by using this link.

MATERIALS AND METHODS

Three important matrix identities will be presented for reference.

First, for an n by p matrix $\mathbf{A}$, the trace of $\mathbf{AA'}$ (that is, the sum of the diagonal elements of $\mathbf{AA'}$) is equal to the total (uncorrected) sum of squares in $\mathbf{A}$ (Searle, 1982, p. 46):

$$\mathrm{tr}(\mathbf{AA}') = \sum_{ij} a_{ij}^2 \qquad [1]$$

Second, for any square matrix $\mathbf{B}$ and an orthogonal matrix $\mathbf{U}$ with columns u_k that satisfies $\mathbf{U'U} = \mathbf{I}$,

$$\mathrm{tr}(\mathbf{B}) = \sum_{k=1}^{n} u_k' \mathbf{B} u_k \qquad [2]$$

This is easily proven by the invariance of the trace operator under orthogonal transformation (Flowers, 2000, p. 251).

Third, let $\mathbf{Y}$ be an environment-centered matrix with n rows (corresponding to genotypes) and p columns (corresponding to environments). Because the matrix is environment-centered, the environment (column) means are zero. The matrix $\mathbf{Y}$ can be (but is not required to be) environment-standardized so that the environment (column) variances are 1. In the derivations that follow it will be assumed that $\mathbf{Y}$ is *not* environment-standardized—this will facilitate an intuitive interpretation of the rows of $\mathbf{Y}$ as genotype means. Now consider a partitioning of the matrix $\mathbf{Y}$ as $\mathbf{Y} = \mathbf{Y}_G + \mathbf{Y}_{GE}$, where $\mathbf{Y}_G$ is a matrix the same size as $\mathbf{Y}$, with each column representing the row-means from $\mathbf{Y}$,

$$\mathbf{Y}_G = \begin{bmatrix} \bar{y}_{1\cdot} & \bar{y}_{1\cdot} & \cdots & \bar{y}_{1\cdot} \\ \bar{y}_{2\cdot} & \bar{y}_{2\cdot} & \cdots & \bar{y}_{2\cdot} \\ \cdots & \cdots & \cdots & \cdots \\ \bar{y}_{n\cdot} & \bar{y}_{n\cdot} & \cdots & \bar{y}_{n\cdot} \end{bmatrix}$$

and

$$\mathbf{Y}_{GE} = \mathbf{Y} - \mathbf{Y}_G = \begin{bmatrix} y_{11} - \bar{y}_{1\cdot} & y_{12} - \bar{y}_{1\cdot} & \cdots & y_{1p} - \bar{y}_{1\cdot} \\ y_{21} - \bar{y}_{2\cdot} & y_{22} - \bar{y}_{2\cdot} & \cdots & y_{2p} - \bar{y}_{2\cdot} \\ \cdots & \cdots & \cdots & \cdots \\ y_{n1} - \bar{y}_{n\cdot} & y_{n2} - \bar{y}_{n\cdot} & \cdots & y_{np} - \bar{y}_{n\cdot} \end{bmatrix}$$

(The dot notation in the subscript indicates a summation over that index). The matrices $\mathbf{Y}$, $\mathbf{Y}_G$, and $\mathbf{Y}_{GE}$ are related through the identity,

$$\mathbf{YY}' = \mathbf{Y}_G\mathbf{Y}_G' + \mathbf{Y}_{GE}\mathbf{Y}_{GE}' \qquad [3]$$

A proof of this identity is provided in Appendix A.

Numerical Measures to Facilitate the Interpretation of GGE Biplot Axes

The identities stated above will now be used in partitioning the sums of squares in $\mathbf{Y}$. First, define TSS to be the total sum of squares in $\mathbf{Y}$, SSG to be the total sum of squares in $\mathbf{Y}_G$, and SSGE to be the total sum of squares in $\mathbf{Y}_{GE}$. Using property [1] and applying the trace operator to [3], the total sum of squares in $\mathbf{Y}$ can be partitioned into a sum of squares that depends on $\mathbf{Y}$ only through the genotype means and a sum of squares that depends on $\mathbf{Y}$ only through deviations (across environments) from the genotype means. In short form, TSS = SSG + SSGE.

Using the SVD of $\mathbf{Y}$, $\mathbf{Y} = \mathbf{U\Lambda V'}$, and by the orthogonality of $\mathbf{U}$ and $\mathbf{V}$, $\mathbf{\Lambda} = \mathbf{U'YV}$ and $\mathbf{\Lambda}^2 = \mathbf{U'YY'U}$. Let TSS_k be the kth diagonal element of $\mathbf{\Lambda}^2$, $\mathrm{TSS}_k = u_k'\,\mathbf{YY}'\,u_k$, where u_k is the kth column of $\mathbf{U}$ (the kth principal component of $\mathbf{YY'}$). The quantity TSS_k represents the portion of TSS that is accounted for by the kth principal component of $\mathbf{Y}$, and $\mathrm{TSS}_k/\mathrm{TSS}$ is the proportion of total variation accounted for by the kth principal component (and biplot axis). This result is well known; see Kroonenberg (1997). Using Eq. [3], TSS_k can be further partitioned as $\mathrm{TSS}_k = u_k'\mathbf{Y}_G\mathbf{Y}_G'u_k + u_k'\mathbf{Y}_{GE}\mathbf{Y}_{GE}'u_k$. For notational convenience, this will be written as $\mathrm{TSS}_k = \mathrm{SSG}_k + \mathrm{SSGE}_k$, where $\mathrm{SSG}_k = u_k'\mathbf{Y}_G\mathbf{Y}_G'u_k$ and $\mathrm{SSGE}_k = u_k'\mathbf{Y}_{GE}\mathbf{Y}_{GE}'u_k$. By identities [1] and [2], $\mathrm{SSG} = \sum_{k=1}^{n} u_k'\mathbf{Y}_G\mathbf{Y}_G'u_k = \sum_{k=1}^{n}\mathrm{SSG}_k$, so that SSG_k represents the portion of SSG that is accounted for by the kth principal component, and $\mathrm{SSG}_k/\mathrm{SSG}$ is the proportion of genotypic variation accounted for by the kth principal component. In a similar manner, $\mathrm{SSGE} = \sum_{k=1}^{n} u_k'\mathbf{Y}_{GE}\mathbf{Y}_{GE}'u_k = \sum_{k=1}^{n}\mathrm{SSGE}_k$. The

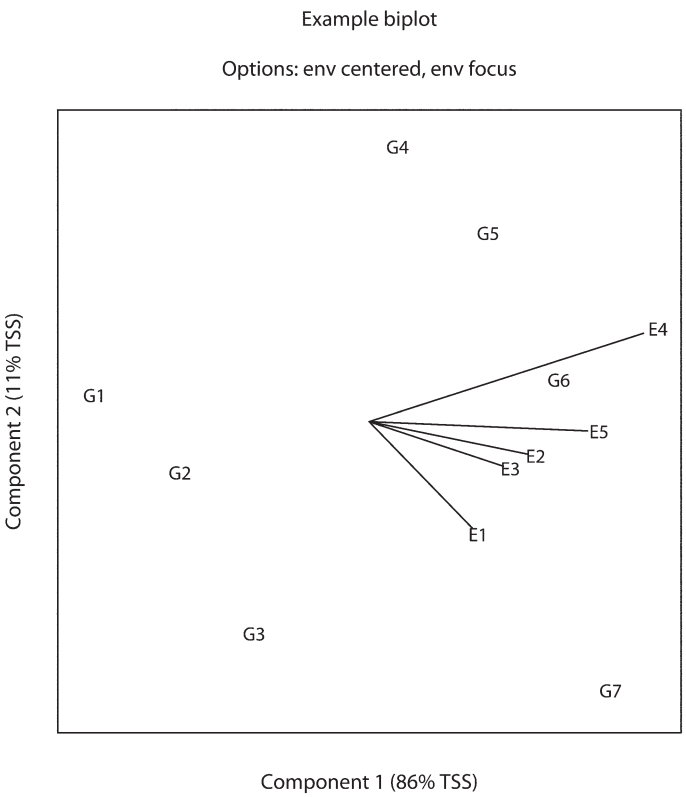

Figure 1. An environment-centered genotype + genotype-by-environment biplot of the data in Table 1. Environments are shown as vectors, and genotypes are shown as individual labels. TSS, total sum of squares.

term $SSGE_k$ represents the portion of SSGE that is accounted for by the kth principal component axis, and $SSGE_k/SSGE$ is the proportion of total GE variation accounted for by the kth principal component.

While previous authors have considered one-way partitionings of TSS (DeLacy et al., 1996; Kroonenberg, 1997), this article introduces a simultaneous partitioning of the total variation into (i) portions for the principal component axes of the biplot and (ii) portions representing variability for genotypes and for variability due to differences among GE interaction means.

One additional useful measure of interest is the absolute value of the correlation between the genotype means and each axis of the biplot. Let $m = (1/p)\mathbf{Y1}_p$ be the vector of genotype means, where $\mathbf{1}_p$ is a vector of 1's of length p. Note that $\mathbf{Y}_G = m\mathbf{1}_p'$, and so SSG = $\text{tr}(\mathbf{Y}_G\mathbf{Y}_G') = \text{tr}(m\mathbf{1}_p'\mathbf{1}_p m') = pm'm$. Since the vectors u_k and m are centered (with mean 0), the squared correlation of u_k and m is

$$\text{corr}^2(u_k, m) = \frac{u_k' m m' u_k}{u_k' u_k m' m} \quad [4]$$

which is the same as

$$\frac{SSG_k}{SSG} = \frac{u_k'\mathbf{Y}_G\mathbf{Y}_G'u_k}{pm'm} = \frac{u_k' m\mathbf{1}_p'\mathbf{1}_p m' u_k}{pm'm} = \frac{u_k' m m' u_k}{u_k' u_k m' m}$$

Therefore the absolute value of the correlation between the genotype means and axis k of the biplot is $\sqrt{SSG_k/SSG}$.

Table 1. Example data showing simulated yields for seven genotypes (G1 to G7) in each of five environments (E1 to E5).

	E1	E2	E3	E4	E5	Mean
G1	50	67	90	98	120	85
G2	55	71	93	102	129	90
G3	65	76	95	105	134	95
G4	50	80	102	130	138	100
G5	60	82	97	135	151	105
G6	65	89	106	137	153	110
G7	75	95	117	133	155	115
Mean	60	80	100	120	140	100

A Graphical Measure to Facilitate the Interpretation of GGE Biplot Axes

The two-way partitioning of the sums of squares calculated above can be presented in a table; however, a graphical representation provides for faster and easier interpretation of the partitioning. Hartigan and Kleiner (1981) introduced mosaic plots for the purpose of displaying data from a contingency table. In a mosaic plot, the counts in a contingency table are represented by rectangles whose area is proportional to the cell frequency. The method can be generalized to any two-way table such as the sums of squares considered here. The mosaic plot will be covered fully in the example that follows.

RESULTS

Table 1 shows simulated data for the yield of seven genotypes in each of five environments. The genotypes and environments in this table are ordered according to their respective means. The deviation of each cell from the marginal effects is assumed to represent a GE effect (and not error). Some interaction is present; for example, in environment E4, the highest yield is from genotype G6, whereas in all other environments, the highest yield is from genotype G7. In environment E1, genotype G7 is much higher yielding than the other genotypes. An environment-focused GGE biplot of this table is shown in Fig. 1.

Kroonenberg (1997) provides concise rules for interpretation of the biplot. The rules will not be completely restated here, but two interpretations of Fig. 1 will be given. To visually interpret the GE interaction in Table 1, note that projecting the genotype points orthogonally onto the E4 environment vector shows that genotype G6 is the highest yielding in this environment. Similarly, the orthogonal projection of the genotype points onto the E1 vector reveals that G7 is much higher yielding than other genotypes in environment E1.

The coordinates of the genotypes are given by the first two column-vectors of **U**:

u_1 = [–0.568 –0.391 –0.238 0.059 0.246 0.392 0.500]′

u_2 = [0.055 –0.105 –0.439 0.569 0.390 0.086 –0.557]′

The vector of genotype means is [85 90 95 100 105 110 115]′, and the absolute values of the correlations between the genotype means and u_1 and u_2 are 0.99 and 0.11, respectively. The first axis of the biplot is highly correlated with the genotype means and can be informally described as a

"genotype axis." The vector u_1 can, if needed, be scaled by –1 so that increasing genotype means are always positively correlated with the first axis.

Applying the two-way partitioning of the TSS to the data in Table 1 gives the results shown in Table 2 and illustrated by the mosaic plot in Fig. 2.

The table and mosaic plot are interpreted in the following manner.

The entire square of the mosaic plot represents 100% of the TSS.

1. The dark-colored areas of the mosaic plot show the total variation due to genotypic effects. In this example 78.1% of the total variation is explained by the differences between genotype means.
2. The light-colored areas of the mosaic plot correspond to the (remaining) variation due to genotype-by-environment effects—21.9% of the total variation.

The columns of the mosaic plot correspond to the principal component axes.

1. The first column represents 86.2% of the total area in the square, and the first principal component accounts for 86.2% of the TSS.
2. The second column/principal component captures 10.9% of the TSS.
3. The first two principal components (used to construct the biplot in Fig. 1) together account for 97.1% of the TSS, (3451.0 + 48.2)/3500 = 99.9% of the SSG, and (410.0+440.8)/980 = 86.8% of the SSGE.

The rows of the mosaic plot (nested inside each column) divide each axis into a portion due to SSG and a portion due to SSGE.

1. The first principal component axis has a very large contribution from SSG. In such a situation, the first axis can be considered a genotype axis, and the distances between the perpendicular projections of the genotype points onto the first axis are highly correlated with the distances between the genotype means. In particular, the absolute value of the correlation is $\sqrt{SSG_1/SSG}$.
2. The second component axis sum of squares has a much higher contribution from GE effects than G effects. It might be tempting to label this a genotype-by-environment axis, but the sum of squares due to GE along axis 1 ($SSGE_1$ = 410.0) is nearly as large as along axis 2 ($SSGE_2$ = 440.8), so a better description for axis 2 might be a "nongenotype axis."

DISCUSSION

The partitioning of the total sums of squares presented here (and the corresponding graphical display) has proven to be a useful tool for the interpretation of biplots. We have generally found that the first principal component axis captures more of the TSS due to genotypes, relative to the GE effects, than does the second axis, the second axis captures more than the third axis, and so forth. This trend is not always monotonic, however, and suggests that useful information may be extracted from higher-order principal components. Further, when the first principal component axis is highly correlated with G means, it is good practice to examine a biplot constructed with the second and third principal components to more fully explore GE effects.

Table 2. A simultaneous partitioning of the sums of squares (for the data in Table 1) along principal-component axes and genotype/genotype-by-environment classification.†

Axis	SSG	SSGE	Total (TSS)	Percent of TSS
1	3451.0	410.0	3861.0	86.2
2	48.2	440.8	489.0	10.9
3	0.8	123.9	124.7	2.8
4	0.0	4.8	4.8	0.1
5	0.0	0.5	0.5	0.0
Total	3500	980	4480	100.0
Percent of TSS	78.1	21.9	100.0	

†SSG, sums of squares due to genotypes; SSGE, sums of squares due to genotype-by-environment; TSS, total sum of squares.

The two-dimensional biplot approximation of a matrix **Y** is safely interpreted only when the biplot representation captures an acceptable portion of the total variation of **Y**. Unless the proportion of total variability explained by the biplot is "sufficiently large, the interpretation of the biplot is suspect" (Jackson, 1991, p. 199). Somewhat more explicitly, Bartkowiak and Szustalewicz (1995, p. 177) examine a biplot capturing 70% of the total variation and "conclude therefore that the representation in the shown biplot is not very satisfactory." In another study, Yan and Rajcan (2002, p. 15) refer

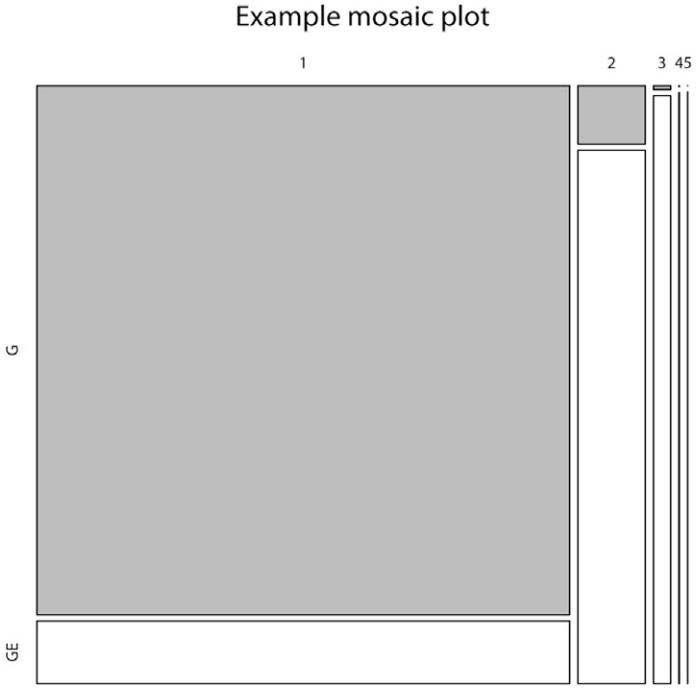

Figure 2. Mosaic plot visualizing the two-way partitioning of the total sums of squares into genotype (G) and genotype-by-environment (GE) components along each principal component axis.

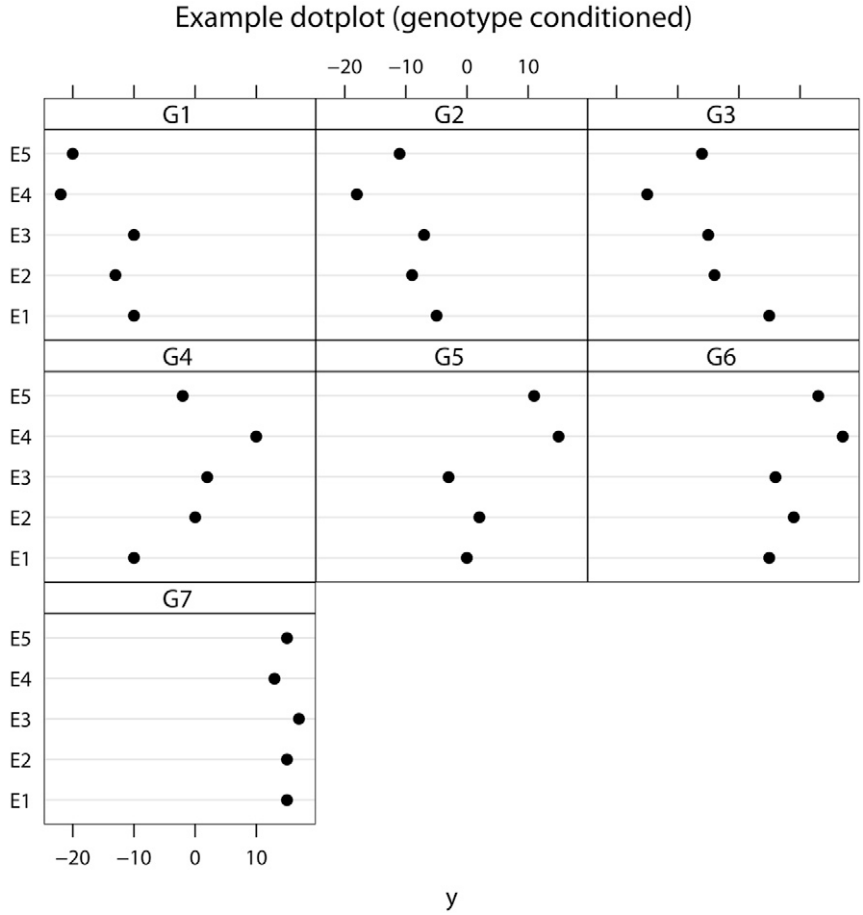

Figure 3. Dot plot of yield in each (centered) environment (E1 to E5), conditioned on genotypes (G1 to G7) (i.e., each panel represents a unique genotype).

to 52% as a "relatively low proportion" of variation of the standardized data. According to Kroonenberg (1997), moderate or low proportions of explained variability may indicate that the main structure of the table is represented in the biplot, but additional structure may lie in higher dimensions. Biplots that capture even less variation may still provide useful information, with the recognition that interpretations of the biplot might need to be verified through other methods.

Increasing the number of principal components used to display the biplot increases the percentage of total variation explained, but at the cost of losing the simplicity of a two-dimensional view of the data. Peres-Neto et al. (2005) evaluated many rules for identifying the number of nontrivial principal components.

To explore additional structure in the data, an alternative graphic such as a dot plot (Cleveland, 1993) can be useful. Figure 3 represents the example data of Table 1 as a dot plot in which each panel represents a different genotype, each row corresponds to a different environment, and the (environment-centered) yield is shown by the position of the dot along a horizontal line. In this dot plot, it is immediately apparent that genotype G7 has the highest (overall) yield. While a similar interpretation can be inferred from the biplot, the dot plot provides a clearer visualization in this case. The dot plot also provides a clear visualization of the stability of genotype G7. Assessing stability of the genotypes in a biplot, however, is not best accomplished using an environment-focused scaling based GGE biplot, as pointed out by Yan and Kang (2003). Instead, they recommend a *genotype*-focused scaling based GGE biplot ($\alpha = 1$) with a display of average environment coordinate (AEC) for that purpose.

A biplot shows the "big picture" of GE interactions and sometimes displays the data more accurately than the raw data (Yan et al., 2000), but does not always reveal all important structure, including possible outliers. Yan (2002, p. 992) suggests that "the pattern displayed by the biplots may be more robust than the individual data points in the raw data because the biplot is based on all data points." However, serious noise is not unusual in plant breeding (van Eeuwijk and Keizer, 1995), and Hawkins et al. (2001, p. 3) note that the SVD is a least-squares method that is "highly susceptible to outliers," so that the accuracy of a biplot can be adversely affected by outliers or other data errors. For example, Cleveland (1993) used dot plots to uncover an error in a 2-yr multilocation experiment in which one location's data had the years switched. A biplot of the same data does not reveal the error.

Consequently, we recommend that biplots be accompanied by dot plots. Dot plots and biplots can each answer questions that are more difficult or impossible for the other to answer. One advantage of dot plots is that they require nearly no explanation, while biplots require some education for proper interpretation. One disadvantage of dot plots is that when there are "many" (perhaps 40 or more) environments and/or genotypes, there may be too much data to fit comfortably on a single page.

The two dot plots of greatest utility for comparing to a biplot are (i) yield of each genotype, with separate panels for each environment, and (ii) yield in each environment, with separate panels for each genotype. The former can provide information about the variance within each environment, while the latter is more useful for examining variance across environments. Biplots can also provide information about the within-environment variance, as the lengths of the environment vectors (in an environment-centered biplot) are approximately proportional to the standard deviations of the genotypes in each environment (Kroonenberg, 1997), whereas the AEC for a genotype-focused biplot roughly displays the variance across environments (Yan and Kang, 2003).

CONCLUSIONS

This article establishes a link between the partition of the TSS by the biplot axes and the partition by SSG and SSGE. This link quantifies the amount of G and GE variation accounted for by the biplot axes. The graphical display of the partitioning in the mosaic plot enables a quick interpretation. The methods in this article have been applied to GGE biplots but could be useful for other types of biplots as well.

APPENDIX A

First note that

$$\mathbf{Y}_{GE}\mathbf{1}_p = \begin{bmatrix} y_{11}-\overline{y}_{1\cdot} & y_{12}-\overline{y}_{1\cdot} & \cdots & y_{1p}-\overline{y}_{1\cdot} \\ y_{21}-\overline{y}_{2\cdot} & y_{22}-\overline{y}_{2\cdot} & \cdots & y_{2p}-\overline{y}_{2\cdot} \\ \cdots & \cdots & \cdots & \cdots \\ y_{n1}-\overline{y}_{n\cdot} & y_{n2}-\overline{y}_{n\cdot} & \cdots & y_{np}-\overline{y}_{n\cdot} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} y_{1\cdot}-p\overline{y}_{1\cdot} \\ y_{2\cdot}-p\overline{y}_{2\cdot} \\ \vdots \\ y_{n\cdot}-p\overline{y}_{n\cdot} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

and so

$$\mathbf{Y}_{GE}\mathbf{Y}_G' = \mathbf{Y}_{GE}\mathbf{1}_p \begin{bmatrix} \overline{y}_{1\cdot} \\ \overline{y}_{2\cdot} \\ \vdots \\ \overline{y}_{n\cdot} \end{bmatrix}' = [0]$$

and (by transposing)

$$\mathbf{Y}_G\mathbf{Y}_{GE}' = [0]$$

From $\mathbf{Y} = \mathbf{Y}_G + \mathbf{Y}_{GE}$, the cross-product is $\mathbf{YY}' = \mathbf{Y}_G\mathbf{Y}_G' + \mathbf{Y}_{GE}\mathbf{Y}_{GE}' + \mathbf{Y}_G\mathbf{Y}_{GE}' + \mathbf{Y}_{GE}\mathbf{Y}_G'$, which reduces to $\mathbf{YY}' = \mathbf{Y}_G\mathbf{Y}_G' + \mathbf{Y}_{GE}\mathbf{Y}_{GE}'$.

APPENDIX B

Software for the construction of (generic) biplots is available for a variety of software programs including Excel (Lipkovich and Smith, 2002), SAS (Friendly, 1991), Xlisp-Stat (Udina, 2005), and the stand-alone GGE biplot (Yan and Kang, 2003). The authors of this article have used both S-Plus (http://www.insightful.com; verified 4 Mar. 2007) and R (http://www.r-project.org; verified 4 Mar. 2007) and began with the biplot function by Venables and Ripley (2002), which was customized to produce the biplot given in this article. Since R is open source and freely available, the interested reader may find the following example useful. This R (version 2.3.1) code can be used to generate an environment-focused GGE biplot and a mosaic plot.

```
# Let Y be a G × E matrix. Center the environments
Y = data.frame(E1 = c(50, 55, 65, 50, 60, 65, 75),
E2 = c(67, 71, 76, 80, 82, 89, 95),
E3 = c(90, 93, 95, 102, 97, 106, 117),
E4 = c(98, 102, 105, 130, 135, 137, 133),
E5 = c(120, 129, 134, 138, 151, 153, 155))
rownames(Y) = c("G1", "G2", "G3", "G4", "G5",
   "G6", "G7")
Y = scale(Y, center = TRUE, scale = FALSE)
YG = matrix(rowMeans(Y)) %*% rep(1,ncol(Y))
YGE = Y - YG
Ysvd = svd(Y) # Singular value decomposition
G = U = Ysvd$u
H = Ysvd$v %*% diag(Ysvd$d)
# Formula: SSGk = diag(U´ YG YG´ U)
SSGk = diag(crossprod(crossprod(YG,U)))
# Formula: SSGEk = diag(U´ YGE YGE´ U)
SSGEk = diag(crossprod(crossprod(YGE,U)))
# Proportion of SS for each axis
TSS = sum(Y^2)
AxisProp = round(100*(SSGk + SSGEk)/TSS,0)
# First, the environment-focused GGE biplot
biplot(G[,1:2], H[,1:2], xlabs = rownames(Y),ylabs =
   colnames(Y),
   xlab = paste("Axis 1: ",AxisProp[1],"%"),
   ylab = paste("Axis 2: ",AxisProp[2],"%"))
# Then the mosaic plot
mosaicdata = data.frame(G = SSGk, GE = SSGEk)
mosaicplot(mosaicdata, main = "", off = c(5,1),
cex.axis = 1, col = c("dimgray", "snow2"))
```

Acknowledgments

The authors would like to thank our colleagues Mark Cooper, Steve Dofing, Steve Langton, Roger Lafosse, Carlos Löffler, and Deanne Wright for discussions and reviews of early versions of this article. We also thank the referees and editors for helpful comments and corrections.

References

Bartkowiak, A., and A. Szustalewicz. 1995. The augmented biplot and some examples of its use. Machine Graphics & Vision 4:161–186. Available at http://www.ii.uni.wroc.pl/~aba/biaug.pdf (verified 4 Mar. 2007).

Cleveland, W.S. 1993. Visualizing data. Hobart Press, Summit, NJ.

Cooper, M., and I.H. DeLacy. 1994. Relationships among analytical methods used to study genotypic variation and genotype-by-environment interaction in plant breeding multi-environment experiments. Theor. Appl. Genet. 88:561–572.

DeLacy, I.H., M. Cooper, and K.E. Basford. 1996. Relationships among analytical methods used to study genotype by environment interaction and evaluation of their impact on response to selection. p. 51–84. *In* M.S. Kang and H.G. Gauch (ed.) Genotype-by-environment interaction. CRC Press, Boca Raton, FL.

Friendly, M. 1991. SAS system for statistical graphics. BBU Press, Cary, NC.

Flowers, B.H. 2000. An introduction to numerical methods in C++. Oxford Univ. Press, Oxford, U.K.

Gabriel, K.R. 1971. The biplot graphical display of matrices with application to principal component analysis. Biometrika 58:453–467.

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

Hartigan, J.A., and B. Kleiner. 1981. Mosaics for contingency tables. p. 268–273. *In* Computer science and statistics: Proc. 13th Symp. on the Interface. Springer-Verlag, NY.

Hawkins, D.M., L. Liu, and S.S. Young. 2001. Robust singular value decomposition. Technical report 122, National Institute of Statistical Sciences. Available at http://www.niss.org/technicalreports/tr122.pdf (verified 4 Mar. 2007). National Institute of Statistical Sciences, Research Triangle Park, NC.

Jackson, J.E. 1991. A user's guide to principal components. John Wiley & Sons, New York.

Kempton, R.A. 1984. The use of biplot in interpreting variety by environment interactions. J. Agric. Sci. (Cambridge) 103:123–135.

Kroonenberg, P.M. 1997. Introduction to biplots for G×E tables. Research report no. 51, Centre for Statistics, The University of Queensland. Available at http://three-mode.leidenuniv.nl/document/biplot.pdf (verified 4 Mar. 2007). The University of Queensland, Brisbane, Australia

Lipkovich, I., and E. Smith. 2002. Biplot and singular value decomposition macros for Excel. J. Statist. Software 7(5). Available at http://www.jstatsoft.org/v07/i05/ (verified 4 Mar. 2007).

Peres-Neto, P.R., D.A. Jackson, and K.M. Somers. 2005. How many principal components? Stopping rules for determining the number of non-trivial axes revisited. Computat. Statist. Data Anal. 49:974–997.

Searle, S.R. 1982. Matrix algebra useful for statistics. John Wiley & Sons, New York.

Udina, F. 2005. Interactive biplot construction. J. Statist. Software 13(5). Available at http://www.jstatsoft.org/v13/i10/ (verified 4 Mar. 2007).

van Eeuwijk, F.A., and L.C.P. Keizer. 1995. On the use of diagnostic biplots in model screening for genotype by environment tables. Stat. Comput. 5:141–153.

Venables, W.N., and B.D. Ripley. 2002. Modern applied statistics with S. Software available at http://www.stats.ox.ac.uk/pub/MASS4/ (verified 4 Mar. 2007). Springer, New York.

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

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

Yan, W., and M.S. Kang. 2003. GGE biplot analysis: A graphical tool for breeders, geneticists, and agronomists. CRC Press, Boca Raton, FL.

Yan, W., and I. Rajcan. 2002. Biplot analysis of test sites and trait relations of soybean in Ontario. Crop Sci. 42:11–20.

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