Published March 22, 2018

REVIEW & INTERPRETATION

Biplots: Do Not Stretch Them!

Waqas Ahmed Malik* and Hans-Peter Piepho

ABSTRACT

Two-way tables of data, either observed or standardized in some way, are commonly analyzed by spectral decomposition or singular value decomposition, providing scores for both rows and columns of the two-way classification. Two of the most common examples in plant and crop research are sample × variable data (principal component analysis) and genotype × environment data, the latter either centered for environment only (genotype main effect plus genotype × environment interaction biplots) or doubly centered for both genotype and environment (genotype × environment interaction biplots based on the additive main effects and multiplicative interaction [AMMI] model). Results are often displayed by plotting the row scores, column scores, or both to visually study the structure of the data. Usually, arrows or lines are drawn from the origin to facilitate interpretation. Graphical features such as angles between arrows and distances between points, as well as graphical operations such as orthogonal projections, allow a number of useful interpretations. For the validity of such properties and operations, it is imperative that the two axes of a plot or biplot be equally scaled exactly (i.e., 1 cm on the vertical axis must represent the same number of units as 1 cm on the horizontal axis). Unfortunately, this important fact is often neglected by users when preparing such plots or integrating them into a text document for publication, rendering all of these features of a plot essentially meaningless. The purpose of the present note, therefore, is to highlight the importance of equal scaling using pertinent examples.

Univ. of Hohenheim, Institute of Crop Science, Biostatistics Unit, Fruwirthstrasse 23, 70599 Stuttgart, Germany. Received 28 Dec. 2017. Accepted 10 Feb. 2018. *Corresponding author (w.malik@uni-hohenheim.de). Assigned to Associate Editor Weikai Yan.

Abbreviations: AMMI, additive main effects and multiplicative interaction; CA, correspondence analysis; GE, genotype × environment interaction; GGE, genotype main effect + genotype × environment interaction; PCA, principal component analysis; PLRV, *Potato leafroll virus*; SVD, singular value decomposition.

Collecting data in a two-way matrix is very common in many scientific research areas, including agricultural science, ecology, psychology, medicine, business, and sociology. In the era of big data, reduction of dimensions and finding clusters is of major interest to data analysts. Singular value decomposition (SVD) is a popular method for dimension reduction of two-way data. The biplot introduced by Gabriel (1971) provides an efficient way to visualize a two-way matrix in a two-dimensional plane by plotting the first two scores obtained by SVD for rows, columns, or both. Numerous publications have made use of biplots in analyzing two-way data. Different variations of the biplot have been introduced (Gower et al., 2011; Greenacre, 2012); however, the underlying theory for the construction and interpretation of all biplots is the same.

In plant breeding and crop research, multienvironment trials are routinely conducted to compare several genotypes in multiple environments resulting in genotype × environment two-way data. Ecologists often assess species abundance over environmental gradients as species × environment two-way tables of frequencies or ordinal abundance scores. Similarly, sample × variable data are arising in many areas of agricultural research and also result in a two-way table of data. Principal component analysis (PCA) for sample × variable data, correspondence analysis

Published in Crop Sci. 58:1061–1069 (2018).
doi: 10.2135/cropsci2017.12.0747

(CA) for categorical abundance data and genotype main effect plus genotype × environment interaction (GGE) models (Yan and Kang, 2002), and additive main effect and multiplicative interaction (AMMI) models (Gauch, 1992) for multienvironment trial data are extensively used for analyzing two-way data in agricultural research. The results from these analyses are mostly visualized in the form of biplots (e.g., PCA biplots, CA biplots, GGE biplots, or genotype × environment interaction (GE) biplots based on AMMI analysis; Kempton, 1984; Gauch, 1992; Yan and Kang, 2002; Yang et al., 2009). Here, we will focus on GE biplots, but our main message equally applies to PCA, CA, and GGE biplots.

The interpretation of all biplots is based on three important geometric properties: the angles between row and column vectors, the length of vectors, and the distances between vectors. Moreover, inferences may be obtained from orthogonal projections. All these properties and operations need to be understood when interpreting these plots. Yan and Tinker (2006) lucidly discussed these principles. The graphical interpretation of biplots requires that the two axes are equally scaled exactly (i.e., 1 cm on the vertical axis must represent the same number of units as 1 cm on the horizontal axis). However, it has been observed in hundreds of agricultural research articles published in or submitted to peer-reviewed journals that authors, typesetters, or both often do not take care of this property. Some of the software packages developed exclusively for biplots take care of this crucial property (e.g., GGEbiplot; Yan and Kang, 2002), but often authors or typesetters unintentionally stretch figures to fit them within the page layout and, in doing so, render the plots essentially meaningless. Other times, when general-purpose plotting functions not specifically designed for biplots are used, the original graphics file may have unequally scaled axes by design, which is not a suitable format for biplots. The aim of this note, therefore, is to highlight the importance of equal scaling in biplots using pertinent examples for illustration.

BIPLOT GEOMETRY

Consider a set of multienvironment trials where g genotypes are tested in each of e environments with r replications. The mean response of individual genotypes averaged over r replications within each environment can be computed and used to fill a $g \times e$ matrix, denoted here as $\mathbf{P}_{g,e}$ (or $\mathbf{P}$ for brevity). The entries of that matrix could be the individual genotype-environment means themselves, or the means could be environment centered, in case a GGE biplot is to be produced, or doubly centered by genotypes and environments, in the case of a GE biplot from AMMI analysis. Below, we will focus on the latter, but all our statements apply equally (mutatis mutandis) to other biplots. Data standardization or data scaling prior to SVD is usually performed on two-way tables when columns represent different variables measured in different units. This happens mostly in sample × variable data where standardization is crucial to give each variable the same weight. Such standardization is not usually crucial, however, for genotype × environment data, where a single trait is measured in different environments, unless heterogeneity of variance between environments is very high (Yan and Tinker, 2006).

Biplot analysis starts by decomposing the $\mathbf{P}$ matrix into a product of three matrices, $\mathbf{U}$, $\mathbf{\Lambda}$, and $\mathbf{V}$, using SVD:

$$\mathbf{P}_{g,e} = \mathbf{U}_{g,s}\mathbf{\Lambda}_{s,s}\mathbf{V}^{\mathbf{T}}_{e,s} \qquad [1]$$

where $\mathbf{\Lambda}_{s,s}$ is a diagonal matrix containing s singular values, ordered from largest to smallest, where s is the rank of the matrix $\mathbf{P}_{g,e}$ with $s \leq \min(g-1, e-1)$. The matrices $\mathbf{U}_{g,s}$ and $\mathbf{V}_{e,s}$ are orthogonal matrices with columns known as left and right singular vectors of $\mathbf{P}_{g,e}$, respectively. The Eq. [1] can be rewritten as

$$\mathbf{P}_{g,e} = \left(\mathbf{U}_{g,s}\mathbf{\Lambda}^{\alpha}_{s,s}\right)\left(\mathbf{\Lambda}^{1-\alpha}_{s,s}\mathbf{V}^{\mathbf{T}}_{e,s}\right) = \mathbf{G}_{g,s}\mathbf{H}^{\mathbf{T}}_{e,s} \qquad [2]$$

where α is a scalar that, in principle, can take on any value on the real line but typically is chosen to lie between 0 and 1. The scalar α is a factor that partitions the singular values into genotype and environment scores. We may refer to the choice of α as singular value partitioning.

If we let $\mathbf{G}_{g,2}$ and $\mathbf{H}_{e,2}$ be the submatrices formed by the first two columns of $\mathbf{G}_{g,s}$ and $\mathbf{H}_{e,s}$, respectively, then $\mathbf{P}_{g,e} \approx \mathbf{G}_{g,2}\mathbf{H}^{\mathbf{T}}_{e,2}$ is a rank 2 approximation of $\mathbf{P}_{g,e}$. This, in fact, is the closest rank 2 approximation to $\mathbf{P}$ in a least-squares sense. In a biplot, the rows of the $g \times 2$ matrix $\mathbf{G}_{g,2}$ are plotted as points, which correspond to g genotypes. The rows of the $e \times 2$ matrix $\mathbf{H}_{e,2}$ are plotted as vectors, which correspond to e environments. Any approximating biplot of $\mathbf{P}$ (or the exact biplot of $\mathbf{P}$, in case it is a matrix of rank 2 [$s = 2$]) allows several approximations, which can be expressed mathematically (see Appendix; Gabriel, 1971) or verbally, as will be outlined in the section below.

The singular value partitioning (choice of α) determines the scaling of the points and vectors in the biplot. The interpretation of the biplot is based on the choice of α, and this choice depends on the underlying research question. The conventional choices of α are 0, 1, and 1/2. The effects and implications of the choice of α will be discussed in the next section.

BIPLOT INTERPRETATION

The interpretation of biplots relies on geometrical properties and operations, and the underlying principles of these geometrical properties and operations are the same for all biplots, regardless of the assumed model and type of data

preprocessing used. A brief verbal summary of geometrical properties is given below, and the mathematical underpinnings are given in the Appendix:

1. The cosine of the angle between the vectors of two environment (genotype) vectors approximates the correlation between the corresponding environments (genotypes) if $\alpha = 0$ ($\alpha = 1$).

2. The length of an environment vector is approximately proportional to the square root of the variance of the corresponding environment if the data are environment centered and $\alpha = 0$, whereas for representing the variance of genotypes, the data should be genotype centered and $\alpha = 1$ should be used. Incidentally, in a GE biplot with $\alpha = 1$, the length of a genotype vector corresponds to the square root of Wricke's (1962) ecovalence for the stability of the genotype.

3. The genotype points can be projected perpendicularly onto the environment vectors, the projection being proportional to the inner product of genotype points and the environment vector, which in turn gives an approximation of the response of a genotype in that environment. This interpretation holds for any choice of α.

4. The distance between genotype points is a two-dimensional approximation of the Euclidean distance between two genotypes if $\alpha = 1$ is used. Similarly, the distances between arrowheads of environment vectors are two-dimensional approximations of the Euclidean distances between environments, if $\alpha = 0$ is used.

Biplots based on different models (e.g., AMMI, GGE) have different interpretations. For example, the meaning of the correlation depends on the model used. Thus, the correlation of two genotypes in a GE biplot is the correlation of interaction effects for these two genotypes, whereas the correlation of two genotypes in a GGE biplot is in terms of genotype performances.

In the next section, these key properties will be illustrated graphically, and the detrimental effects of stretching the graph will be discussed, hopefully convincing the reader that equal scaling of both biplot axes is indeed indispensable.

ILLUSTRATION WITH PERFECTLY RANK-2 MATRIX

A three-by-three toy dataset **P** is given in Table 1. After doubly centering the data for AMMI analysis, the resulting matrix of genotype × environment interaction (**C**) effects is of rank 2. Thus, any element of the **C** matrix can be represented exactly, based on a SVD in Eq. [1], as

Table 1. A toy dataset.

	Environment			
Genotype	E1	E2	E3	Total
G1	12.1	22.1	28.2	62.4
G2	9.1	22.1	13.2	44.4
G3	16.3	14.3	17.4	48.0
Total	37.5	58.5	58.8	154.8

the inner product of the two vectors corresponding to its rows and to its columns.

The AMMI model is

$$\mathbf{P} = \mathbf{1}_I \mu \mathbf{1}_J^{\mathbf{T}} + \mathbf{a}\mathbf{1}_J^{\mathbf{T}} + \mathbf{1}_I \mathbf{b}^{\mathbf{T}} + \mathbf{C} \qquad [3]$$

where $\mathbf{a} = (a_1, a_2, \ldots, a_I)^{\mathbf{T}}$ and is a vector of environment main effects, $\mathbf{b} = (b_1, b_2, \ldots, b_J)^{\mathbf{T}}$ and is a vector of genotype main effects, and $\mathbf{1}_n$ is an n-vector of ones. The term μ is an overall mean. The matrix **C** is the genotype × environment interaction effect (for simplicity, we have omitted a residual error term). Thus, the interaction matrix **C** can be given as

$$\mathbf{C} = \mathbf{P} - \mathbf{1}_I \mu \mathbf{1}_J^{\mathbf{T}} - \mathbf{a}\mathbf{1}_J^{\mathbf{T}} - \mathbf{1}_I \mathbf{b}^{\mathbf{T}} \qquad [4]$$

or

$$\begin{bmatrix} -4 & -1 & 5 \\ -1 & 5 & -4 \\ 5 & -4 & -1 \end{bmatrix} = \begin{bmatrix} 12.1 & 22.1 & 28.2 \\ 9.1 & 22.1 & 13.2 \\ 16.3 & 14.3 & 17.4 \end{bmatrix} - \begin{bmatrix} 17.2 & 17.2 & 17.2 \\ 17.2 & 17.2 & 17.2 \\ 17.2 & 17.2 & 17.2 \end{bmatrix}$$
$$- \begin{bmatrix} -4.7 & 2.3 & 2.4 \\ -4.7 & 2.3 & 2.4 \\ -4.7 & 2.3 & 2.4 \end{bmatrix} - \begin{bmatrix} 3.6 & 3.6 & 3.6 \\ -2.4 & -2.4 & -2.4 \\ -1.2 & -1.2 & -1.2 \end{bmatrix}$$

The **C** matrix can be subjected to a SVD as shown in Eq. [1], yielding **U**, **V**, and **Λ** matrices. It should be noted that, in this small example, both singular values are the same. In larger datasets, the singular values form a declining series.

$$\mathbf{U} = \begin{bmatrix} -0.617 & 0.535 \\ -0.154 & -0.802 \\ 0.772 & 0.267 \end{bmatrix}$$

$$\mathbf{\Lambda} = \begin{bmatrix} 7.94 & 0 \\ 0 & 7.94 \end{bmatrix}$$

$$\mathbf{V} = \begin{bmatrix} 0.816 & 0 \\ -0.408 & -0.707 \\ -0.408 & 0.707 \end{bmatrix}$$

Biplots of the **C** matrix using Eq. [2] with three different singular value partitionings ($\alpha = 0$, 1, and 1/2) are given in Fig. 1. The genotype and environmental scores are represented as vectors and will be illustrated briefly

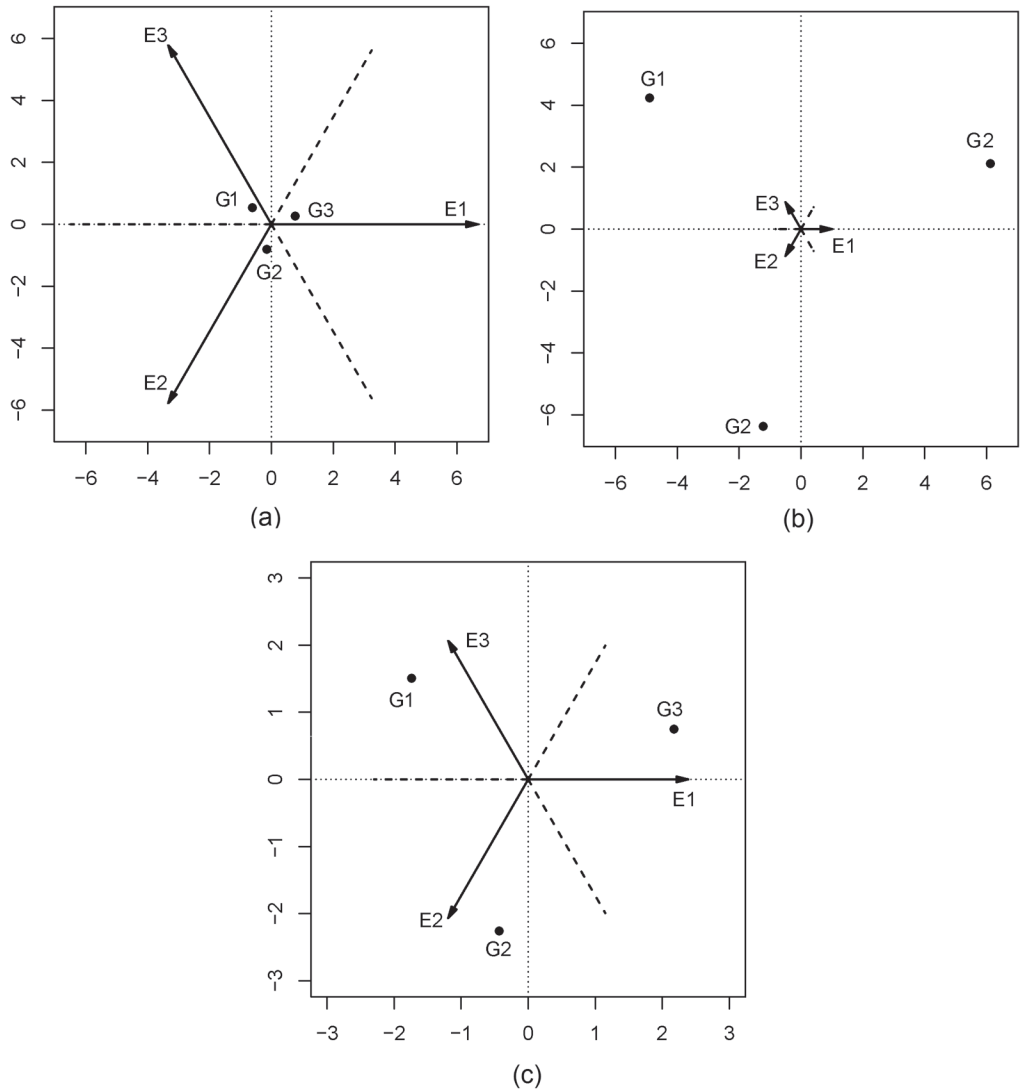

Fig. 1. Biplots of the genotype × environment interaction (**C**) matrix with different singular value partitioning: (a) $\alpha = 0$, (b) $\alpha = 1$, (c) $\alpha = 1/2$. The **C** matrix has a rank of 2. Both axes are equally scaled.

before investigating the effect of distortions by stretching or compressing an axis. The labels E1 through E3 are used to denote the three environment vectors, displayed as arrows, and G1 through G3 denote the genotype vectors, each of which starts at the origin, as do the environmental vectors. The genotype vectors are represented only by dots placed at the terminal ends of the vectors. The biplot with singular value partitioning $\alpha = 0$ given in Fig. 1a gives the so-called environment view (lengths of vectors and angles between them, as well as distances of environments, can be interpreted), whereas the genotype view given in Fig. 1b uses the singular value partitioning $\alpha = 1$, which provides the so-called genotype view (Yan and Kang, 2002). The distance between environments is the Euclidean distance of vectors in Fig. 1a. The distance between environments E1 and E3 is 11.2, which is the Euclidean distance between vectors E1 and E3. Note that this Euclidean distance is defined in terms of the interaction effects. Thus, if two environments have a short distance, their interaction profiles are similar. Similarly, the Euclidean distance between genotypes G1 and G3 is 11.2 in Fig. 1b. Genotypes with a short Euclidean distance have similar interaction profiles.

In Fig. 1a, the cosine of the angle θ between two environment vectors represents the correlation of interaction effects between two environments with an angle of 0° indicating a correlation of +1, an angle of 90° (or 270°) a correlation of 0, and an angle of 180° a correlation of −1. For example, the cosine of the angle between environments E1 and E3 [$\cos(\theta_{E1,E3})$] gives the correlation of interaction effects between environments E1 and E3. From Fig. 1a, it is evident that the angle between E1 and E3 is >90°, which represents a negative correlation. Similarly, the cosine of the angle between two genotype points in Fig. 1b represents the correlation of the two genotypes.

The squared length of a genotype vector in Fig. 1b is the approximation of the sum of squares of interaction effects of a genotype, which is Wricke's (1962) ecovalence

of the genotype. Therefore, genotype vectors having the same length corresponds to genotypes having the same ecovalence.

The **G** and **H** matrices using the symmetric singular value partitioning (i.e., $\alpha = 1/2$) are

$$\mathbf{G} = \begin{bmatrix} -1.739 & 1.506 \\ -0.435 & -2.259 \\ 2.174 & 0.753 \end{bmatrix}$$

$$\mathbf{H} = \begin{bmatrix} 2.300 & 0 \\ -1.150 & -1.992 \\ -1.150 & 1.992 \end{bmatrix}$$

Using these matrices, the interaction of a genotype with an environment can directly be derived from a biplot drawn in Fig. 2. For example, the inner product of $\overrightarrow{\mathbf{OG3}}$ and $\overrightarrow{\mathbf{OE1}}$ gives the interaction of genotype G3 with environment E1. The inner product of genotype G3 and environment E1 can be computed from the vector coordinates of G3 (2.174, 0.753) and E1 (2.30, 0) as (2.174 × 2.30) + (0.753 × 0) = 5, which is exactly equal to the interaction effect for G3 in E1, due to perfect fit.

The interaction effect of G3 in E1 may also be derived from Fig. 2a using the vector geometry of $\overrightarrow{\mathbf{OG3}}$ and $\overrightarrow{\mathbf{OE1}}$:

$$\overrightarrow{\mathbf{OG3}} \times \overrightarrow{\mathbf{OE1}} = \left|\overrightarrow{\mathbf{OG3}}\right|\left|\overrightarrow{\mathbf{OE1}}\right|\cos(\theta) = \left|\overrightarrow{\mathbf{OG3}'}\right|\left|\overrightarrow{\mathbf{OE1}}\right| \quad [5]$$

where $\overrightarrow{\mathbf{OG3}'}$ is the orthogonal projection of $\overrightarrow{\mathbf{OG3}}$ on $\overrightarrow{\mathbf{OE1}}$. In Fig. 2a, the approximate angle between G3 and E1 is 19°. The lengths of $\overrightarrow{\mathbf{OG3}}$ on $\overrightarrow{\mathbf{OE1}}$ are $\left|\overrightarrow{\mathbf{OG3}}\right| = \sqrt{(2.174^2 + 0.753^2)} = 2.3$, and $\left|\overrightarrow{\mathbf{OE1}}\right| = \sqrt{(2.30^2 + 0^2)} = 2.3$, respectively. Thus, the inner product of G3 and E1 is

$$\overrightarrow{\mathbf{OG3}} \times \overrightarrow{\mathbf{OE1}} = 2.3 \times 2.3 \times \cos(19°) = 5.00$$

which is the interaction effect of this genotype × environment combination. The sign of the interaction is determined by the angle subtended by the genotype on the environment vector. Specifically, an angle of <90° or >270° between a genotype vector and an environment vector indicates that the genotype has a positive interaction effect at that environment. A negative interaction is indicated if the angle is between 90° and 270°. For example, the angle between genotype G3 and environment E1 is <45°, which shows a positive interaction between them. However, the angle between G3 and E2 is >90° which shows a negative interaction between them.

Using Eq. [5], the interaction effect of G3 in E1 can directly be derived from Fig. 2a using the orthogonal projection of genotype $\overrightarrow{\mathbf{OG3}}$ on $\overrightarrow{\mathbf{OE1}}$. Equation [5] shows that interaction is proportional to the length of projection $\overrightarrow{\mathbf{OG3}'}$, which is also true for $\overrightarrow{\mathbf{OG1}'}$ and $\overrightarrow{\mathbf{OG2}'}$. Thus, the interactions of two genotypes with the same environment can be assessed by comparing the length and its direction relative to the origin of their projections onto that environment. The orthogonal projections of the three genotypes onto environment E1 are drawn in Fig. 2a. Up to a factor of proportionality, the interaction of all genotypes in environment E1 can be read from this graph. The length of the $\overrightarrow{\mathbf{OG3}'}$ is larger than that of $\overrightarrow{\mathbf{OG1}'}$ and $\overrightarrow{\mathbf{OG2}'}$ in environment E1, showing that G3 has higher interaction with E1 than G1 and G2. The direction of $\overrightarrow{\mathbf{OG1}'}$ and $\overrightarrow{\mathbf{OG2}'}$ is in the opposite direction of $\overrightarrow{\mathbf{OE1}}$, which shows that G1 and G2 have negative interaction with environment E1. The length of the $\overrightarrow{\mathbf{OG1}'}$ is larger than that of $\overrightarrow{\mathbf{OG2}'}$, which shows that G1 has a larger negative interaction effect than G2 with environment E1.

The orthogonal projections of genotype G3 onto the three environments are also drawn in Fig. 2b. The length of the projection vector and its direction relative to the

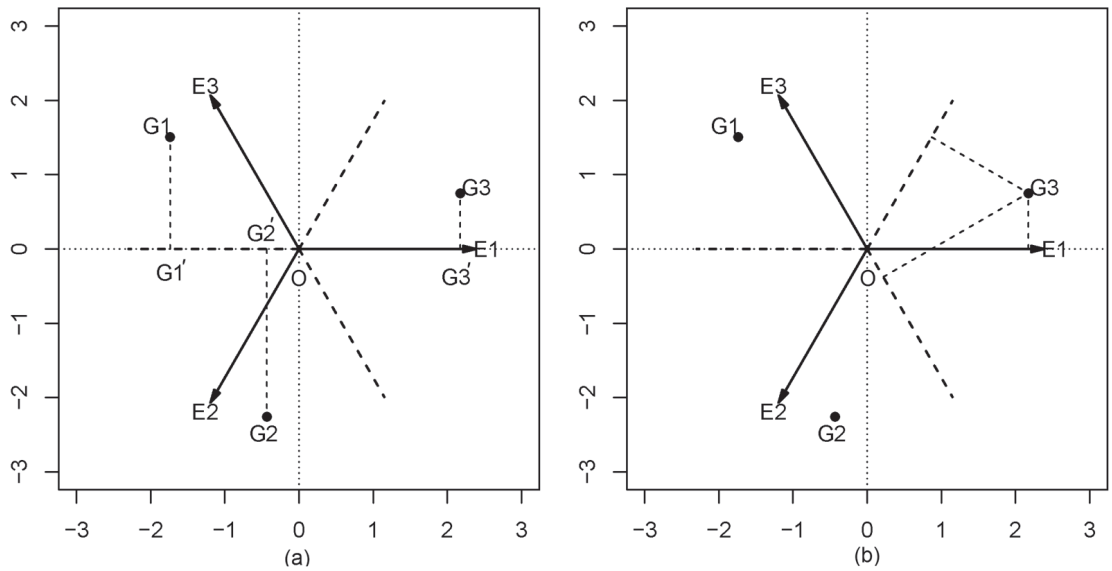

Fig. 2. Biplots of the genotype × environment interaction (**C**) matrix with symmetric singular value partitioning ($\alpha = 1/2$): (a) projections of all genotypes on environment E1, (b) the projections of genotype G3 on all environments. The axes are equally scaled.

origin of the vector determine the relative size and sign of the interaction effect of G3 in the three environments, respectively. The projections of G3 onto E2 and E3 are in the opposite direction of the vectors E2 and E3, respectively, which reflects the negative interaction of G3 with these environments.

It is important to reiterate that all interpretations of a biplot are based on the assumption that both axes are drawn to scale (Yan and Tinker, 2006). The axes of biplots given in Fig. 1 and 2 are drawn to scale exactly equally. By contrast, the biplots shown in Fig. 3 are two examples where axes are drawn without taking care of the axis scales. The plots were drawn using R (R Core Team, 2017) with default settings without explicitly defining them to be of equal scale. The biplot shown in Fig. 3a is drawn with plot height and width being equal, but axes are not drawn with equal relative scaling. Here, one unit on the horizontal axis is not equal in length to one unit on the vertical axis. By comparison, both axes of the biplot shown in Fig. 3b were drawn with relative scaling, but later the plot was stretched to fit within the page layout. The biplots shown in Fig. 3 distort all geometrical properties of genotypic and environmental vectors in biplots shown in Fig. 1 and 2. All the angles between different environments and genotypic vectors in Fig. 3a and 3b are changed compared with those in Fig. 1 and 2. For example, the angle between environments E2 and E3 in Fig. 2a is obtuse, but it is turned into an acute angle in Fig. 3. The projections of genotypes onto different environments and distances between them are stretched out, which eventually leads to a faulty interpretation. The lengths of vectors are also stretched out. The lengths of genotype vectors are therefore not correctly representing the sum of squares of interactions (i.e., the ecovalence).

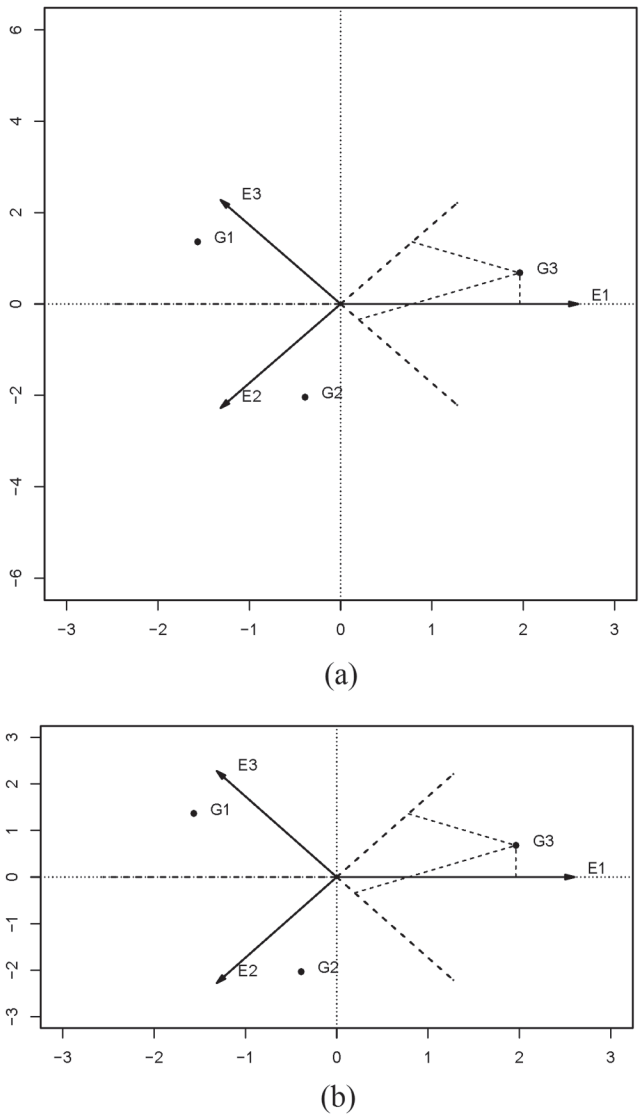

Fig. 3. Biplots of the genotype × environment interaction (**C**) matrix with symmetric singular value partitioning (α = 1/2): (a) the biplot drawn with axes drawn on different scales, (b) the axes of biplots drawn on an equal scale, but the plot was stretched later on.

A REAL DATASET

Six trials were conducted in Peru to evaluate the development of *Potato leafroll virus* (PLRV)-resistant potato cultivars at the International Potato Center, Lima, Peru. The data from 28 genotypes were analyzed to determine the yield gain and resistance from PLRV. The data is available in R package "agricolae" (de Mendiburu, 2017).

The yield data will be used to show the GE biplot based on an AMMI model fitted to this data. The two-dimensional biplot of the first two components is given in Fig. 4 (R code to generate this plot is given in the supplemental material). The biplot is drawn with α = 0 (i.e., environment

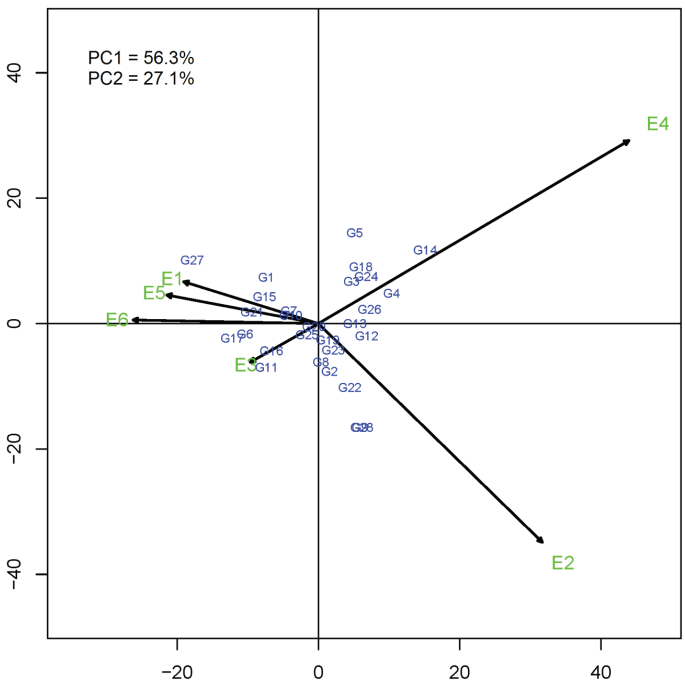

Fig. 4. Biplot of the *Potato leafroll virus* dataset from fitting an additive main effects and multiplicative interaction (AMMI) model using α = 0 (environment view) with equally scaled axes. PC1 and PC2 are Principal Components 1 and 2.

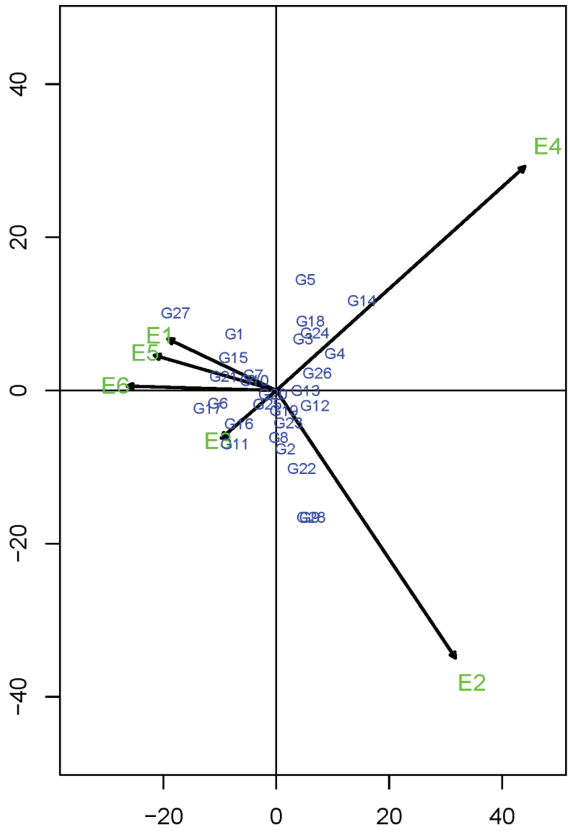

Fig. 5. Biplot shown in Fig. 4 when stretched vertically.

view) with the length of genotype vectors multiplied by 40 ($\mathbf{G}_{g,2} = 40\mathbf{U}_{g,2}$) so that the vectors for genotypes are on a scale commensurate with those for the environments (Digby and Kempton, 1987, p. 64). The proportion of the sum of squares of singular values of these two components to the total sum of squares of singular values $\left[\left(\lambda_1^2+\lambda_2^2\right)\Big/\sum_{i=1}^{s}\lambda_i^2\right]$ represents the variation explained by them. Here, the first two dimensions represent 83.4% of the variation in the genotype × environment interaction matrix. A third component accounts for 9.4%. The axes of this biplot are equally scaled.

When the same biplot is stretched vertically as shown in Fig. 5 or stretched horizontally as shown in Fig. 6, all geometrical representations of genotypic and environmental vectors are distorted. The angles between different environments and genotypes are changed. The angle between environment E2 and E4 was ∼90° in Fig. 4 but it becomes >90° in Fig. 5 and <90° in Fig. 6. Thus, environments E2 and E4 are not correlated in Fig. 4, but they seem negatively correlated in Fig. 5 and positively correlated in Fig. 6. The lengths of environment vectors and projections of genotypes on environments are also changed in Fig. 5 and 6. These geometrical changes make a meaningful interpretation of the plot impossible.

CONCLUSION

The biplot is a widely used graphical tool for analyzing multienvironment trial data. However, valid graphical interpretation is based on several geometric foundations. The use of such a graphical display is permissible only if the axes are equally scaled. An incorrect interpretation can easily occur without this condition. This article provides some insights into the interpretation of biplots, highlighting the importance of equal scaling of both biplot axes.

In conclusion, we suggest that data analysts always draw biplots with an equal relative scale on both axes, thus ensuring that one unit scale on the horizontal axis is equal to one unit scale on the vertical axis. This can be achieved in R by specifying the aspect ratio to equal one. Care should also be taken by authors to preserve equal scaling when exporting the plot to an image file for publishing. Similarly, technical editors and publishers should make sure no distortion is introduced by stretching to fit the journal's page layout at typesetting.

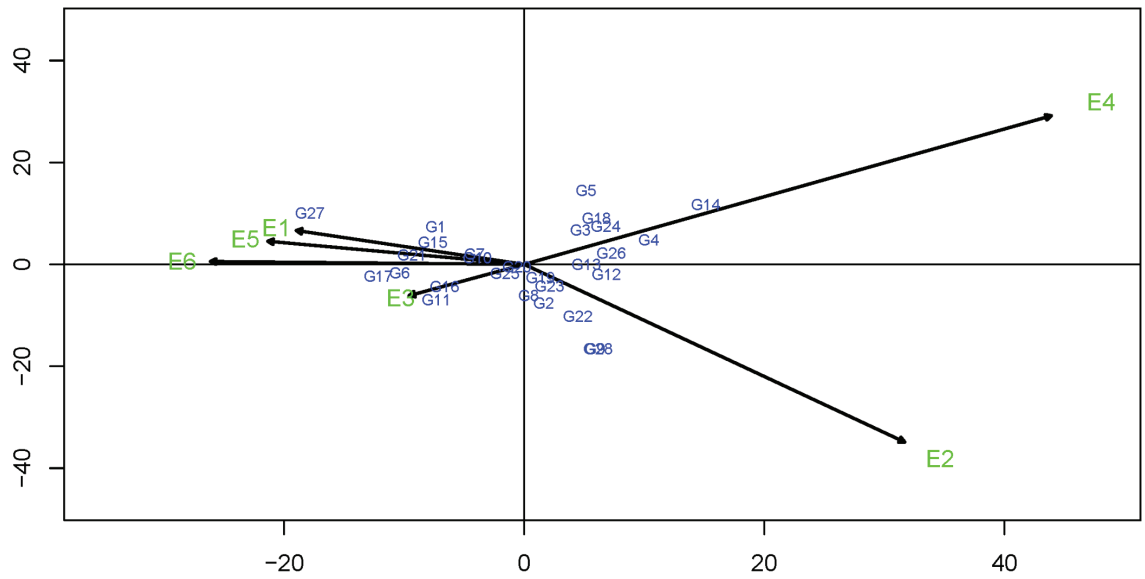

Fig. 6. Biplot shown in Fig. 4 when stretched horizontally.

Conflict of Interest

The authors declare that there is no conflict of interest.

Supplemental Material Available

Supplemental material for this article is available online.

Acknowledgments

We are thankful to the Associate Editor and two anonymous reviewers for their useful comments and suggestions. This work was supported by German Research Foundation (DFG) Grant PI 377/1-17.

References

de Mendiburu, F. 2017. agricolae: Statistical procedures for agricultural research. R Found. Stat. Comput. https://CRAN.R-project.org/package=agricolae (accessed 28 Feb. 2018).

Digby, P.G.N., and R.A. Kempton. 1987. Multivariate analysis of ecological communities. Springer, the Netherlands.

Gabriel, K.R. 1971. Biplot display of multivariate matrices with application to principal component analysis. Biometrika 58:453–467. doi:10.1093/biomet/58.3.453

Gauch, H.G. 1992. Statistical analysis of regional yield trials: AMMI analysis of factorial designs. Elsevier, Amsterdam.

Gower, J.C., S. Lubbe, and N. le Roux. 2011. Understanding biplots. Wiley, New York. doi:10.1002/9780470973196

Greenacre, M.J. 2012. Biplots: The joy of singular value decomposition. Wiley Interdiscip. Rev. Comput. Stat. 4:399–406. doi:10.1002/wics.1200

Kempton, R. 1984. The use of biplots in interpreting variety by environment interactions. J. Agric. Sci. 103:123–135. doi:10.1017/S0021859600043392

R Core Team. 2017. R: A language and environment for statistical computing. R Found Stat. Comput., Vienna.

Wricke, G. 1962. On a method of understanding the biological diversity in field research. (In German.) Z. Pflanzenzuecht. 47:92–146.

Yan, W., and M.S. Kang. 2002. GGE biplot analysis: A graphical tool for breeders, geneticists, and agronomists. CRC Press, Boca Raton, FL. doi:10.1201/9781420040371

Yan, W., and N.A. Tinker. 2006. Biplot analysis of multi-environment trial data: Principles and applications. Can. J. Plant Sci. 86:623–645. doi:10.4141/P05-169

Yang, R.C., J. Crossa, P.L. Cornelius, and J. Burgueño. 2009. Biplot analysis of genotype × environment interaction: Proceed with caution. Crop Sci. 49:1564–1576. doi:10.2135/cropsci2008.11.0665

APPENDIX

1. Let $\overrightarrow{\mathbf{og}}$ be a vector of two elements g_1 and g_2 (Fig. A1a). The length of the vector is indicated by $\left|\overrightarrow{\mathbf{og}}\right|$ and calculated as

$$\left|\overrightarrow{\mathbf{og}}\right| = \sqrt{g_1^2 + g_2^2}$$

2. The inner product of two vectors $\overrightarrow{\mathbf{og}}$ and $\overrightarrow{\mathbf{oh}}$ (Fig. A1b) can be defined as:

$$\overrightarrow{\mathbf{og}} \times \overrightarrow{\mathbf{oh}} = g_1 h_1 + g_2 h_2 = \left|\overrightarrow{\mathbf{og}}\right|\left|\overrightarrow{\mathbf{oh}}\right|\cos(\theta)$$

From this, we can obtain:

$$\theta = \cos^{-1}\left(\frac{\overrightarrow{\mathbf{og}} \times \overrightarrow{\mathbf{oh}}}{\left|\overrightarrow{\mathbf{og}}\right|\left|\overrightarrow{\mathbf{oh}}\right|}\right)$$

where θ is the angle between the vectors $\overrightarrow{\mathbf{og}}$ and $\overrightarrow{\mathbf{oh}}$.

3. The vectors $\overrightarrow{\mathbf{og}}$ and $\overrightarrow{\mathbf{oh}}$ are orthogonal if θ = 90° (Fig. A1c).

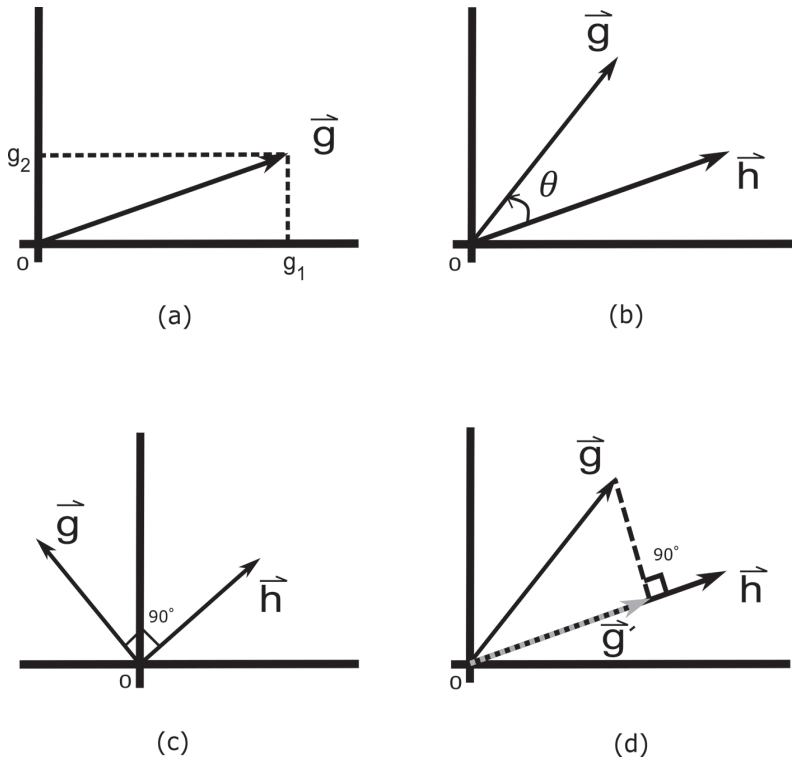

Fig. A1. Some basic vector geometry.

4. The projection $\overrightarrow{\mathbf{og}'}$ of $\overrightarrow{\mathbf{og}}$ on $\overrightarrow{\mathbf{oh}}$ is a vector collinear with $\overrightarrow{\mathbf{oh}}$ that can be found by dropping a perpendicular line from the tip of $\overrightarrow{\mathbf{og}}$ onto $\overrightarrow{\mathbf{oh}}$ (Fig. A1d).

5. The matrix $\mathbf{P}_{g,e}$ may be approximated in a two-dimensional subspace using SVD as $\mathbf{P}_{g,e} \approx \mathbf{G}_{g,2}\mathbf{H}_{e,2}^{\mathbf{T}}$, whose elements are given by

$$p_{ij} = g_{i1}h_{j1} + g_{i2}h_{j2}$$

For example, the response of Genotype 1 (row) in Environment 3 (column) is approximated in two dimensions by

$$p_{13} = g_{11}h_{31} + g_{12}h_{32}$$

The coordinates for Genotype 1 are $\overrightarrow{\mathbf{og}} = \left(g_{11}, g_{12}\right)$ and for Environment 3 are $\overrightarrow{\mathbf{oh}} = \left(h_{31}, h_{32}\right)$ (Fig. A1d):

$$p_{13} = \overrightarrow{\mathbf{oh}} \times \overrightarrow{\mathbf{og}'} = \overrightarrow{\mathbf{oh}} \times \overrightarrow{\mathbf{og}} \times \cos\left(\theta\right)$$