STATISTICS

Singular-Value Partitioning in Biplot Analysis of Multienvironment Trial Data

Weikai Yan*

ABSTRACT

Multienvironment trials (MET) are conducted every year for all major crops throughout the world, and best use of the information contained in MET data for cultivar evaluation and recommendation has been an important issue in plant breeding and agricultural research. A genotype main effect plus genotype × environment interaction (GGE) biplot based on MET data allows visualizing (i) the which-won-where pattern of the MET, (ii) the interrelationship among test environments, and (iii) the ranking of genotypes based on both mean performance and stability. Correct visualization of these aspects, however, requires appropriate singular-value (SV) partitioning between the genotype and environment eigenvectors. This paper compares four SV scaling methods. Genotype-focused scaling partitions the entire SV to the genotype eigenvectors; environment-focused scaling partitions the entire SV to the environment eigenvectors; symmetrical scaling splits the SV symmetrically between the genotype and the environment eigenvectors; and equal-space scaling splits the SV such that genotype markers and environment markers take equal biplot space. It is recommended that the genotype-focused scaling be used in visualizing the interrelationship and comparison among genotypes and the environment-focused scaling be used in visualizing the interrelationship and comparison among environments. All scaling methods are equally valid in visualizing the which-won-where pattern of the MET data, but the symmetric scaling is preferred because it has all properties intermediate between the genotype- and the environment-focused scaling methods.

Regional performance trials are conducted every year for all major crops throughout the world with the purpose of identifying superior cultivars for the target region. The measured yield of each cultivar in each test environment is a mixture of environment main effect (E), genotype main effect (G), and genotype × environment interaction (GE). Typically, E explains most (up to 80% or higher) of the total yield variation, and G and GE are usually small. However, it is G and GE that are relevant to cultivar evaluation. Moreover, G and GE must be considered simultaneously when making cultivar selection decisions. For this reason, instead of trying to separate G and GE, Yan et al. (2000) deliberately put the two together and referred to the mixture as *GGE*. Yield data from regional performance trials, or more generally, multienvironment trials (MET), are usually quite large, and it is difficult to grasp the general pattern of the data without some kind of graphical presentation. The biplot technique (Gabriel, 1971) provides a powerful solution to this problem. A biplot that displays the GGE of a MET data, referred to as a *GGE biplot*, is an ideal tool for MET data analysis (Yan, 2001; Yan et al., 2000, 2001). A GGE biplot that sufficiently approximates the GGE of a MET data set allows, among other things, visualization of three important aspects: (i) the genotype × environment relations as represented by the which-won-where pattern, which facilitate megaenvironment investigation (Gauch and Zobel, 1997); (ii) the interrelationships among test environments, which facilitate identification of better environments for cultivar evaluation (Cooper et al., 1997) and of redundant environments that can be dismissed (Yan and Rajcan, 2002); and (iii) the interrelationships among genotypes, which facilitate comparison among genotypes and genotype ranking on both mean yield and stability (Yan et al., 2001). In all previous publications, it has been implicatively claimed that a single GGE biplot is sufficient for all these purposes (Yan, 2001; Yan et al., 2000, 2001). The purpose of this paper is to demonstrate that different GGE biplots are required to properly address different aspects.

THEORY

The Model for a GGE Biplot

A GGE biplot is constructed by first subjecting the GGE matrix, i.e., the environment-centered data, to singular-value (SV) decomposition. The GGE matrix is decomposed into three component matrices—the SV matrix (array), the genotype eigenvector matrix, and the environment eigenvector matrix—so that each element in the GGE matrix is recovered through

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

where

- Y_{ij} = the measured mean yield of genotype i (=1, 2,...n) in environment j (= 1, 2,...m)
- μ = the grand mean
- β_j = the main effect of environment j, ($\mu + \beta_j$) being the mean yield in environment j
- λ_l = the SV of lth principal component (PC), the square of which is the sum of squares explained by PCl (l = 1, 2,...k, with $k \leq \min(m, n)$ and $k = 2$ for a two-dimensional biplot)
- ξ_{il} = the eigenvector of genotype i for PC l
- η_{lj} = the eigenvector of environment j for PC l
- ε_{ij} = the residual associated with genotype i in environment j

Cereal Breeding and Biometrics, Univ. of Guelph, Guelph, ON, Canada N1G 2W1. Received 16 Nov. 2001. *Corresponding author (wyan@uoguelph.ca; wyan@ggebiplot.com).

Published in Agron. J. 94:990–996 (2002).

Abbreviations: AEC, average environment coordinates; G, genotype main effects; GE, genotype × environment interaction; GGE, genotype main effects plus genotype × environment interaction effects; MET, multienvironment trials; PC, principal component; SV, singular value.

To generate a biplot that can be used in visual analysis of MET data, the SVs have to be partitioned into the genotype and environment eigenvectors so that Eq. [1] can be written in the form of

$$Y_{ij} - \mu - \beta_j = \sum_{l=1}^{k} g_{il} e_{lj} + \varepsilon_{ij} \quad [2]$$

where g_{il} and e_{lj} are called PC l scores for genotype i and environment j, respectively. In a biplot, genotype i is displayed as a point defined by all g_{il} values, and environment j is displayed as a point defined by all e_{lj} values (l =1 and 2 for a two-dimensional biplot). Singular-value partitioning is implemented by

$$g_{il} = \lambda_l^{f_l} \xi_{il} \text{ and } e_{lj} = \lambda_l^{1-f_l} \eta_{lj} \quad [3]$$

where f_l is the partition factor for PC l. Theoretically, f_l can be anything between 0 and 1 although 0.5 is so far most commonly used. Therefore, there are numerous ways to construct a GGE biplot, leading to numerous GGE biplots of different shapes. The influence of different partitioning factors on the interpretation of a GGE biplot has rarely been documented, except in DeLacy et al. (1996). This paper compares four special SV partition methods in GGE biplot construction for their suitability in visualizing the three aforementioned aspects.

Environment-Focused Scaling

It is referred to as environment-focused scaling if $f_l = 0$, i.e., if the SV is completely partitioned into the environment eigenvectors so that $g_{il} = \xi_{il}$ and $e_{lj} = \lambda_l \eta_{lj}$. In this scaling, the environmental scores are in the original unit of yield (e.g., t ha^{-1}), and the genotype scores are normalized (unitless). Because all of the SV is partitioned into the environment scores, the range of the environment scores is likely many times greater than that of the genotypes, and when directly plotted, the genotypes are likely to be crowded in the biplot. To generate a biplot in which the ranges of the genotypes and the environments are comparable, the genotype scores for both axes can be multiplied by an arbitrary number. Multiplying both axes of the genotype scores with a *positive* number is equivalent to multiplying such a number to each element of the environment-centered data matrix and will not alter the genotype × environment pattern of the data. Properties of the environment-focused scaling were discussed by DeLacy et al. (1996) under the term "principal component scaling."

By partitioning all SV to the environment scores, the relative importance of PC1 and PC2 is fully reflected by the locations of the environment markers in the GGE biplot. Therefore, a GGE biplot based on environment-focused scaling is most suitable for visualizing the interrelationship among the environments but not for that of the genotypes.

Genotype-Focused Scaling

It is referred to as genotype-focused scaling when $f_l = 1$, i.e., when the SV is partitioned entirely into the genotype eigenvectors so that $g_{il} = \lambda_l \xi_{il}$ and $e_{lj} = \eta_{lj}$. In this scaling, the unit of the genotype scores (g_{il}) is the original unit of yield, and the environmental scores (e_{lj}) are unitless. Because all of the SV is partitioned into the genotype scores, the range of the genotype scores are likely to be many times greater than that of the environment scores. As a result, the environments in the biplot are likely to be crowded relative to the genotypes. For a genotype × environment table, genotype-focused scaling is the default scaling method of the SAS procedure PRINCOM, as demonstrated in Yan and Hunt (2002). To generate a biplot in which the ranges of the genotypes and the environments are comparable, the environment scores of both axes can be multiplied by an arbitrary factor.

By partitioning all SV to the genotype scores, the relative importance of PC1 and PC2 is fully reflected by the locations of the genotypes in the GGE biplot. Therefore, a GGE biplot based on genotype-focused scaling is suitable for evaluating the genotypes but not the environments.

Symmetric Scaling

It is called symmetrical scaling when f_l takes the value of 0.5 so that $g_{il} = \lambda_l^{0.5} \xi_{il}$ and $e_{lj} = \lambda_l^{0.5} \eta_{lj}$. This type of scaling has the unique property that genotype scores and environmental scores have the same unit for both PC1 and PC2, which is the square root of the original unit [e.g., $(t/ha)^{0.5}$]. This property makes it possible to visualize the relative magnitude of genotype variation and environment variation for both PC1 and PC2. This is the scaling method used in AMMI analysis (Gauch, 1988) and some GGE biplot analysis (Yan et al., 2000). It is intermediate between the environment-focused scaling and the genotype-focused scaling in all aspects.

Equal-Space Scaling

The equal-space scaling method was first proposed by Dr. Paul L. Cornelius (University of Kentucky) and reported in Yan et al. (2001). It is devised so that the biplot space taken by genotypes is equal to that by environments. This is achieved by assigning the SV partition factor to:

$$f_l = 0.5\left\{1 + \frac{\ln\left(\frac{\max(\eta_{lj}) - \min(\eta_{lj})}{\max(\xi_{il}) - \min(\xi_{il})}\right)}{\ln\lambda_l}\right\} \quad [4]$$

In this scaling, the unit of the genotype scores and that of the environments are usually different; the unit of PC1 and that of PC2 are also different. The meaning of this scaling in terms of genotype and environment evaluation is not defined. This problem was not realized when Yan et al. (2001) was prepared. Equal-space scaling is equivalent to the symmetric scaling only when $\max(\eta_{lj}) - \min(\eta_{lj}) = \max(\xi_{il}) - \min(\xi_{il})$ for both PC1 and PC2. Recently, Dr. Paul L. Cornelius (personal communication, 2002) has proposed two additional scaling methods, equal maximum vector length scaling and equal maximum ordinate length scaling. The discussion on equal-space scaling in this paper should also apply to these two scaling methods.

Regardless of the SV partitioning method, the genotype × environment matrix represented by Eq. [1] is not altered. Therefore, all possible scaling methods should reveal the same which-won-where pattern.

MATERIALS AND METHODS

The yield data from 1998 winter wheat (*Triticum aestivum* L.) performance trials are used in this investigation, which tested 33 cultivars in eight environments (Table 1). This data set was used previously in Yan et al. (2001) in comparing two types of GGE biplots. The analysis could be conducted using statistical packages such as SAS (SAS Inst., 1996), as described in detail in Yan and Hunt (2002). However, this is a tedious and laborious process. All analyses were done using *GGEbiplot*, which is a Windows application that fully automates biplot analysis (Yan, 2001). A demo version of the program is available at www.ggebiplot.com (verified 7 July 2002).

Table 1. Yield data of 33 winter wheat genotypes in nine locations (data from 1998 Ontario Winter Wheat Performance Trials).

Genotypes	Locations							
	EA	HN	ID	NN	OA	RN	WE	WK
	kg ha^{-1}							
2510	5.7	4.7	4.9	6.0	4.8	5.8	4.8	4.9
2526	6.3	6.1	5.4	6.5	5.1	6.6	5.3†	6.0
2533	6.6	6.3	4.7	6.5	5.4	5.9	4.5	4.8
2540	7.1	6.5	5.6	7.0	6.0	6.2	5.0	5.6
2557	7.6	6.1	5.5	6.5	5.1	6.5	4.6	6.2
2560	7.6	6.1	5.6	6.9	5.4	7.1	5.0	6.4
2737	6.7	5.9	5.2	6.4	5.1	6.1	5.2	5.9
Ari	6.3	5.7	4.4	6.2	5.8	5.0	4.2	5.8
Ash	5.9	5.3	4.8	6.4	4.6	5.9	4.4	5.8
Car	5.7	5.5	4.5	5.6	5.8	4.8	4.1	5.9
Dex	7.0	5.8	4.8	6.5	5.8	5.6	4.3	6.1
Dlt	5.5	4.5	4.3	4.5	4.7	3.1	2.6	4.1
Ess	7.0	5.8	5.3	6.3	6.1	5.8	4.7	6.1
Fre	6.8	5.6	4.9	6.2	4.7	5.6	4.2	6.1
Fun	6.5	5.1	4.4	5.8	4.6	5.5	4.4	5.5
Han	5.9	5.3	4.4	5.7	4.8	5.2	4.1	5.0
Har	7.0	5.8	4.6	6.2	4.9	5.2	4.6	5.7
Hur	6.6	6.0	4.9	6.3	5.4	5.1	4.5	6.2
Kar	7.0	5.7	4.8	6.5	5.5	5.3	4.2	6.4
Ljh95	5.6	5.0	4.8	5.8	4.2	5.7	4.8	5.7
Mac	6.6	6.8	5.2	6.6	6.0	5.4	4.6	6.8
Mar	6.6	5.4	4.5	6.2	5.3	5.0	3.8	5.6
Men	7.6	6.1	5.1	6.8	5.8	5.6	4.8	6.2
Mon	6.1	4.7	4.1	5.5	4.0	4.8	4.2	5.0
Mor	7.1	5.9	4.4	6.0	5.2	4.7	3.9	5.6
Mou	7.4	6.1	5.3	6.7	6.0	6.0	5.2	6.7
Mwh	5.8	5.5	4.8	5.8	4.3	5.8	3.7	5.8
Pat	6.7	5.8	4.9	6.2	5.0	5.7	4.5	5.8
Ron	6.6	5.9	4.5	6.3	5.7	5.4	4.1	6.4
S93	5.2	4.6	3.7	4.7	5.2	3.8	2.6	4.6
Sup	7.1	6.1	4.6	6.7	5.7	4.9	4.1	6.3
Maximum	7.6	6.8	5.6	7.0	6.1	7.1	5.3	6.8

† Underlined values are the maximum yields at each test location.

RESULTS AND DISCUSSION

The first two PCs explained 87% of the total GGE variation (Fig. 1–3), suggesting that a biplot of PC1 and PC2 adequately approximates the environment-centered data. Different scaling methods are compared for each of the three aspects: the which-won-where pattern, which is a summary of the genotype × environment relations, the genotype ranking based on mean and stability of the genotypes, and the interrelationships among the environments.

The Which-Won-Where Pattern

Visualization of the which-won-where pattern of MET data is important for studying the possible existence of different megaenvironments in a region (Gauch and Zobel, 1997; Yan et al., 2000, 2001). The polygon view of a GGE biplot explicitly displays the which-won-where pattern, and hence is a succinct summary of the GE pattern of a MET data set (Fig. 1). The polygon is formed by connecting the markers of the genotypes that are further away from the biplot origin such that all other genotypes are contained in the polygon. The rays in Fig. 1 are lines that are perpendicular to the sides of the polygon or their extensions. Take Fig. 1A as an example. Ray 1 is perpendicular to the side that connects cultivars Dlt and Zor (the Dlt-Zor side); Ray 2 is perpendicular to side Zor-Sup; similarly, Ray 3 is perpendicular to side Sup-Mac, Ray 4 to side Mac-Mou, Ray 5 to side Mou-2560, Ray 6 to side 2560-2526, Ray 7 to side 2526-Ljh95, Ray 8 to side Ljh95-Mon, and Ray 9 to side Mon-Dlt. Side Ljh95-Mon and side Mon-Dlt almost completely coincide.

These nine rays divide the biplot into nine sectors, and the environments fall into three of them (Fig. 1A). An interesting feature of this view of a GGE biplot is that the vertex cultivar *for* each sector had the highest yield in all environments that fall in the sector (Yan et al., 2000). Thus, four environments—RN, WE, ID, and NN—fall into the sector delineated by Rays 5 and 6, and the vertex cultivar for this sector is 2560, suggesting that the highest-yielding cultivar for these four environments in 1998 was 2560. Similarly, three environments—WK, HN, and EA—fall into the sector delineated by Rays 4 and 5, and the vertex cultivar for this sector is Mou, suggesting that the highest-yielding cultivar for these three environments in 1998 was Mou. A single environment, OA, falls into the sector delineated by Rays 3 and 4, and the vertex cultivar for this sector is Mac, suggesting that Mac was the highest-yielding cultivar for OA in 1998.

Although the biplot based on genotype-focused scaling (Fig. 1B) is quite different in shape from Fig. 1A, it displays the same genotype × environment relations. First, the biplot is divided into the same number of sectors. Second, the same groups of environments are displayed. Third, the same vertex cultivars can be identified although it is less obvious than in Fig. 1A. The vertex cultivar for the sector between Rays 5 and 6 is 2560. The vertex cultivar between Rays 4 and 5 is Mou rather than Mac as explained below. Ray 5 is perpendicular to side Mou-2560. It separates environments OA, EA, HN, and WK, where Mou is nominally better than 2560, from environments RN, WE, ID, and NN where 2560 is nominally better than Mou. Ray 4 is perpendicular to side Mac-Mou. It separates environment OA, where Mac is nominally better than Mou, from all other environments where Mou is nominally better than Mac. Thus, Mou is nominally the best cultivar for environments EA, HN, and WK. For the same reasoning, Mac, rather than Sup, is the vertex for the sector where OA resides. The use of a perpendicular line to a polygon side as a comparison facility was first proposed in Yan et al. (2000) and more fully described in Yan and Kang (2002). Needless to say, biplots based on symmetric scaling (Fig. 1C) and equal-space scaling (Fig. 1D) display the same which-won-where pattern. In Fig. 1D, cultivars Ljh95, Mon, and Dlt aligned on a straight line; thus, only eight rays are displayed.

The above statements on the which-won-where pattern based on the biplots can be largely, though not entirely, validated from the original data (Table 1) because only 87%, rather than 100%, of the GGE are explained by the biplots. Arguably, however, 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. For example, the biplots indicate that cultivar Mou was the highest yielding in EA, HN, and WK (Fig. 1), whereas Mac was actually the highest yielding in HN and WK (Table 1). This is partially because Mac was, on the whole, poorer than

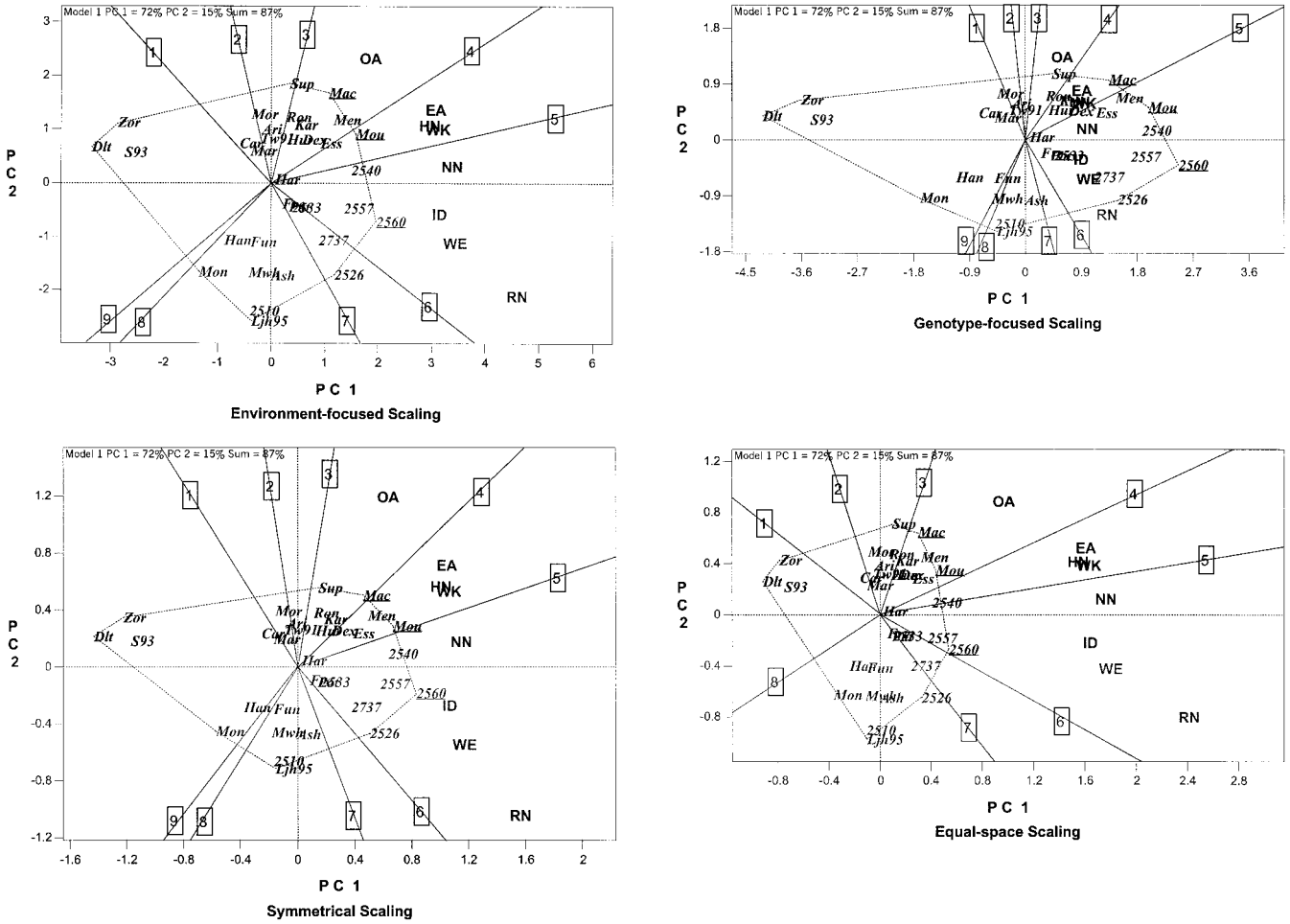

Fig. 1. Polygon view of the GGE biplot: (A) environment-focused scaling, (B) genotype-focused scaling, (C) symmetrical scaling, and (D) equal-space scaling. Genotypes are in title case, and environments are in upper case. PC, principal component.

Mou in environments ID, RN, and WE where 2560 was the highest-yielding genotype.

Mean Yield and Stability of Genotypes

Visualization of the mean performance and stability of genotypes is always an important issue in cultivar evaluation. It was pointed out that if PC1 of a GGE biplot approximates the genotype main effects (i.e., mean performance), PC2 must approximate the GE effects associated with each genotype, which is a measure of instability (Yan et al., 2000). However, this condition is not always met. To deal with possible exceptions, an alternative GGE biplot was devised (Yan et al., 2001), which forces the abscissa to present the genotype main effect and is, therefore, more interpretable in terms mean performance and stability. This is, however, at the expense of explaining slightly smaller GGE variation. The merits of the two types of GGE biplots are combined to some extent by introducing an average environment coordination (AEC) (Yan, 2001; Yan and Hunt, 2002). This is implemented as follows. First, an average environment is defined by the average PC1 and PC2 scores of all environments, represented by a small circle (Fig. 2). A line is then drawn to pass through this average environment and the biplot origin; this line is called the average environment axis and serves as the abscissa of the AEC. The ordinate of the AEC is the line that passes through the origin and is perpendicular to the AEC abscissa (Fig. 2). Unlike the AEC abscissa, which has one direction, with the arrow pointing to greater genotype main effect, the AEC ordinate is indicated by double arrows, either direction away from the biplot origin indicates greater GE effect and reduced stability. As a rule, the genotype projections onto the AEC abscissa are good approximations of the genotype main effects. For our case, the correlation between the projections and the genotype main effects was 0.982.

An ideal cultivar should have the highest mean performance and be absolutely stable (i.e., perform the best in all environments). Such an ideal cultivar is defined by having the greatest vector length of the high-yielding genotypes and with zero GE, as represented by the dot with an arrow pointing to it (Fig. 2). Although such an ideal cultivar may not exist in reality, it can be used as a reference for cultivar evaluation. A genotype is more desirable if it is located closer to the ideal cultivar. Thus, using the ideal cultivar as the center, concentric circles were drawn to help visualize the distance between each genotype and the ideal cultivar (Fig. 2).

Because the units of both PC1 and PC2 for the geno-

types are the original unit of yield in the genotype-focused scaling (Fig. 2B), the units of the AEC abscissa (mean yield) and ordinate (stability) should also be the original unit of yield. The unit of the distance between genotypes and the ideal cultivar, in turn, is the original unit of yield as well. Therefore, the ranking based on the genotype-focused scaling assumes that stability and mean yield are equally important.

In contrast, the genotype scores in a biplot based on environment-focused scaling (Fig. 2A) take no account of the relative importance of PC1 and PC2. The ratio of the PC1 score to the PC2 score (PC1/PC2 ratio) is reduced by a factor of λ_1/λ_2 compared with the genotype-focused scaling. If the AEC abscissa happens to coincide with the PC1 axis, the mean/stability ratio is also reduced, i.e., the stability is overemphasized, by a factor of λ_1/λ_2. On the other extreme, if the AEC abscissa happens to coincide with the PC2 axis, the mean/stability ratio is enlarged, i.e., the mean performance is overemphasized, by a factor of λ_1/λ_2. Obviously, numerous

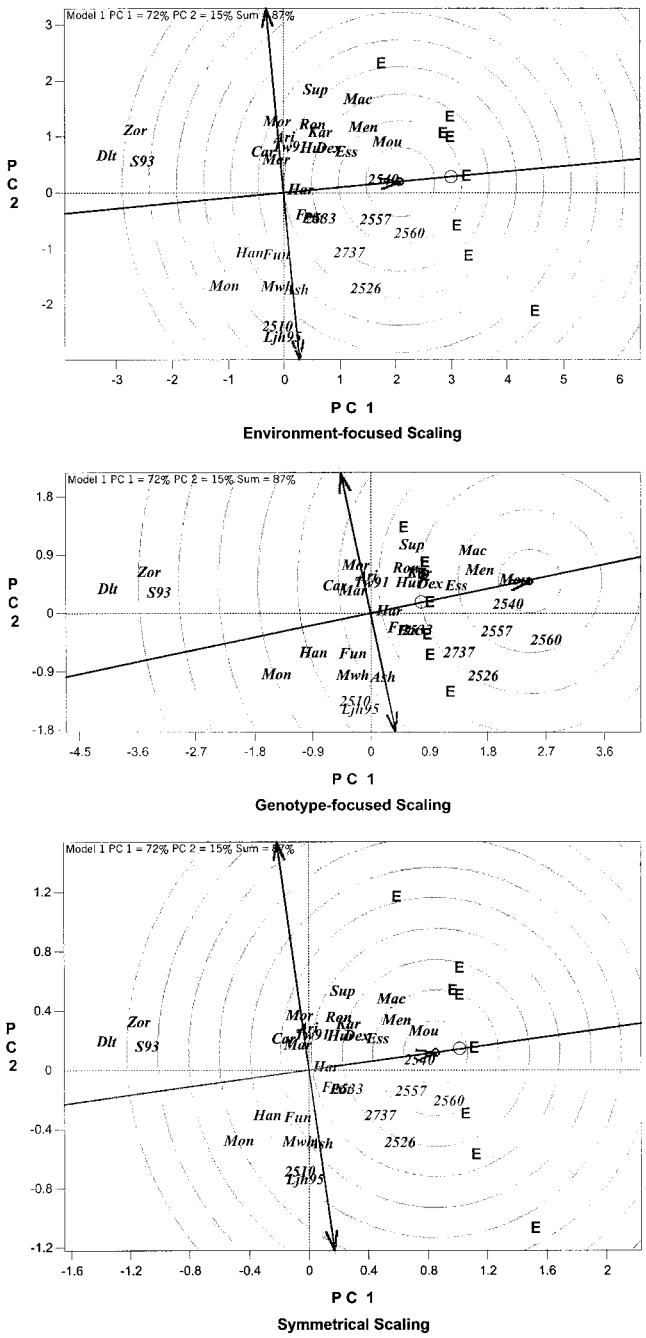

Fig. 2. Average environment coordination (AEC) view of the GGE biplot: (A) environment-focused scaling, (B) genotype-focused scaling, and (C) symmetrical scaling. Genotypes are in title case, and environments are represented by E. The concentric circles are used to compare all genotypes with the ideal cultivar, which is represented by the concentric center. The single isolated circle represents the mean environment. PC, principal component.

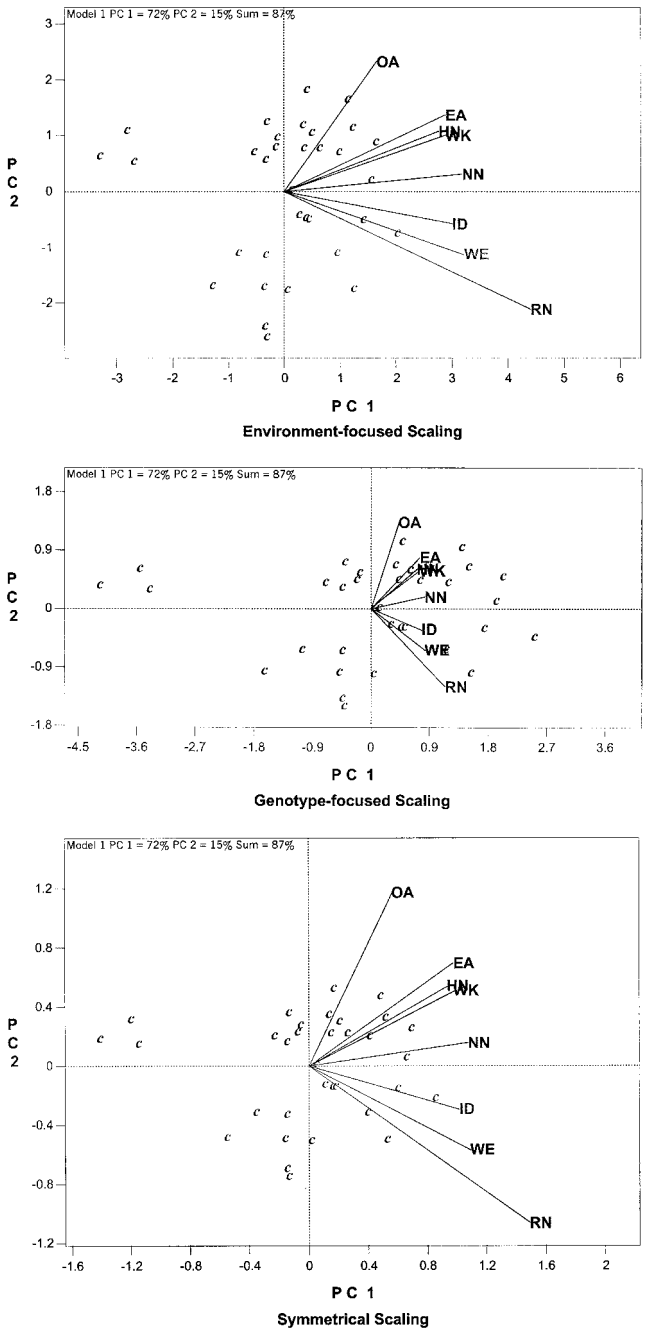

Fig. 3. Vector view of the GGE biplot: (A) environment-focused scaling, (B) genotype-focused scaling, and (C) symmetrical scaling. Genotypes are represented by *c*, and environments are in upper case.

possibilities exist between these two extremes since the rotation angle of the AEC relative to the original coordinates can be anything between 0 and 90°. The stability is overemphasized if the angle is <45° and underemphasized if the angle is >45°. Consequently, the relative importance of mean vs. stability—hence, the meaning of the distance between a genotype and the ideal cultivar in a biplot based on environment-focused scaling—is not defined.

Similar discussion applies to the symmetrical scaling, though to a lesser extent. The units of PC1 and PC2 for the genotypes, the units of the AEC axes for the genotypes, and the unit of the distance between a genotype and the ideal cultivar are all in square root of the original unit. Relative to genotype-focused scaling, the symmetric scaling tends to put more weight on PC2 vs. PC1 by a factor of $\sqrt{\lambda_1/\lambda_2}$. As for the environment-focused scaling, depending on the angle of rotation of the AEC relative to the original coordinates, the mean/stability ratio in the symmetric scaling may be over- (rotation angle >45°) or underemphasized (rotation angle <45°) relative to the genotype-focused scaling. The relative importance of mean vs. stability—hence, the meaning of the distance between a genotype and the ideal cultivar in a biplot based on symmetric scaling—is also undefined.

It is important to know that different scaling methods put different weights on mean vs. stability. Consequently, the choice of scaling methods may influence the ranking of the genotypes based on mean performance and stability. For example, based on the genotype-focused scaling (Fig. 2B), cultivar Mou was the most desirable. It was more desirable than 2560 even though the latter had the highest mean yield. Genotypes '2540', 2560, '2557', 'Men', and Mac seemed to be equally desirable although their yields differed in individual environments. Genotypes Dlt, Zor, and 'S93' were the least desirable because they had the lowest mean yield. In contrast, based on the environment-focused scaling (Fig. 2A), 2540 is identified as the most desirable because this scaling method puts more weight to stability relative to mean yield (because the rotation angle is <45°) and because 2540 was more stable than the other high-yielding genotypes. For the same reason, cultivars 2526 and Sup, which had relatively large GE were put to the fourth and fifth layers from the concentric center compared with Fig. 2B where they were in the third and fourth layers, respectively. The ranking of genotypes in the symmetrical scaling (Fig. 2C) is intermediate between the genotype-focused scaling and the environment-focused scaling.

The units of the axes of a biplot based on the equal-space scaling are variable, depending on the data. The genotype ranking in such a biplot, therefore, has no clear interpretations although its AEC still indicates the mean and stability of the genotypes (biplot not shown) as other scaling methods do.

Interrelationship among Environments

The correlation coefficients among the eight test environments are presented in Table 2. It contains 28 correlation coefficients. The number of correlation coefficients increases quickly to an unmanageable level as more environments are involved. For example, if there were 20 environments, this table would have 190 correlation coefficients. Admirably, the vector view of a GGE biplot (Fig. 3) provides a succinct summary of the interrelationships among the environments. The lines that connect the biplot origin and the markers of the environments are called environment vectors. The angle between the vectors of two environments is related to the correlation coefficient between them. The accuracy of a biplot in displaying the interrelationships among the environments, however, has much to do with the SV scaling method. When the biplot adequately approximates the environment-centered data, and when the environment-focused scaling is used (Fig. 3A), the cosine of the angle between the vectors of two environments approximates the correlation coefficient between them (Kroonenburg, 1995). To verify, all environments should be positively correlated because all angles among them are smaller than 90°. Sure enough, there are no negative numbers in Table 2. The angle between environments OA and RN is only slightly smaller than 90°; therefore, the correlation between them should be close to 0. In Table 2, it was 0.181. The loose association of OA with ID and WE (Table 2) was also well reflected in Fig. 3A. There were inconsistencies, however. For example, Fig. 3A suggests that HN and WK are the most closely correlated environments, but the largest correlation coefficient was actually between RN and ID (Table 2). Some inconsistencies are expected because the biplot did not explain 100% of the GGE variation.

A biplot based on the environment-focused scaling (Fig. 3A) correctly displays the interrelationships among environment because the environment scores reflect the relative importance of PC1 and PC2. The genotype-focused scaling cannot correctly display the correlation coefficients among environments (Fig. 3B). The most obvious example is the obtuse angle between OA and RN, which suggests a negative correlation between

Table 2. Correlation coefficients among test environments.

Locations	EA	HN	ID	NN	OA	RN	WE	WK
EA	1	0.728	0.641	0.781	0.535	0.529	0.553	0.664
HN	0.728	1	0.750	0.855	0.653	0.614	0.631	0.735
ID	0.641	0.75	1	0.873	0.395	0.928	0.877	0.719
NN	0.781	0.855	0.873	1	0.550	0.807	0.810	0.772
OA	0.535	0.653	0.395	0.550	1	0.181	0.264	0.543
RN	0.529	0.614	0.928	0.807	0.181	1	0.867	0.583
WE	0.553	0.631	0.877	0.810	0.264	0.867	1	0.611
WK	0.664	0.735	0.719	0.772	0.543	0.583	0.611	1

them. Figure 3B also suggests that there was no association between RN and EA (and also HN and WK), which is not true (Table 2). Thus, the genotype-focused scaling cannot be used to reliably visualize the interrelationship among environments. Its accuracy in displaying the correlation among environments, relative to that of the environment-focused scaling, decreases as $(\lambda_1 - \lambda_2)$ increases. The biplot based on symmetrical scaling (Fig. 3C) was in between the two extreme scaling methods. Its accuracy in displaying the interrelationship among environments also decreases as $(\lambda_1 - \lambda_2)$ increases.

CONCLUSIONS

The GGE biplots of MET data allow visualizing the interrelationship among genotypes (including the ranking of cultivars based on both mean performance and stability), interrelationship among environments, and interaction between genotypes and environments (including the which-won-where pattern). Depending on the SV partitioning or scaling methods, there are numerous ways to construct a GGE biplot. Four special methods were examined in this paper: genotype-focused scaling, environment-focused scaling, symmetric scaling, and equal-space scaling.

It was demonstrated that all scaling methods are equally valid in displaying the which-won-where pattern of the MET data although one method may be better than others in terms of clarity, which depends on the data. However, the scaling method influences the ranking of genotypes based on mean performance and stability and the display of the interrelationship among environments. Cultivar ranking based on the genotype-focused scaling assumes that mean performance and stability are equally important; the environment-focused scaling tends to put more or less weight on stability, and the symmetric scaling is intermediate between genotype- and environment-focused scaling methods in all aspects. The equal-space scaling does not have a defined interpretation for its ranking of genotypes. For visualization of the interrelationship among environments, the environment-focused scaling should be used. The genotype-focused scaling, and the symmetrical scaling to a lesser extent, can distort the interrelationship among environments because the relative magnitude of PC1 and PC2 is not taken into account or only partially taken into account.

It is recommended that the genotype-focused scaling should be used in visualizing the interrelationship and comparing among genotypes and the environment-focused scaling be used in visualizing the interrelationship and comparing among environments. All scaling methods are equally valid in visualizing the which-won-where pattern of the MET data, but the symmetric scaling is preferred because it has all properties intermediate between the genotype- and environment-focused scaling methods. These understandings have been incorporated in the GGEbiplot software (Yan and Kang, 2002; www.ggebiplot.com). The equal-space scaling method is not recommended for either ranking the genotypes or visualizing the interrelationship among environments although it is equally valid as other scaling methods in displaying the which-won-where patterns.

REFERENCES

Cooper, M., R.E. Stucker, I.H. DeLacy, and B.D. Harch. 1997. Wheat breeding nurseries, target environments, and indirect selection for grain yield. Crop Sci. 37:1168–1176.

DeLacy, I.H., K.E. Basford, M. Cooper, J.K. Bull, and C.G. McLaren. 1996. Analysis of multi-environment trials—a historical perspective. p. 39–124. *In* M. Cooper and G.L. Hammer (ed.) Plant adaptation and crop improvement. CAB Int., Wallingford, Oxon, UK.

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

Gauch, H.G. 1988. Model selection and validation for yield trials with interaction. Biometrics. 44:705–715.

Gauch, H.G., and R.W. Zobel. 1997. Identifying mega-environments and targeting genotypes. Crop Sci. 37:311–326.

Kroonenberg, P.M. 1995. Introduction to biplots for G×E tables. Res. Rep. 51. Dep. of Mathematics, Univ. of Queensland, Brisbane, Australia.

SAS Institute. 1996. SAS/STAT user's guide. Version 6. 4th ed. SAS Inst., Cary, NC.

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., P.L. Cornelius, J. Crossa, and L.A. Hunt. 2001. Two types of GGE biplots for analyzing multi-environment trial data. Crop Sci. 41:656–663.

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

Yan, W., and L.A. Hunt. 2002. Biplot analysis of diallel data. Crop Sci. 42:21–30.

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

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