Biplot Analysis of Genotype × Environment Interaction: Proceed with Caution

Rong-Cai Yang,* Jose Crossa, Paul L. Cornelius, and Juan Burgueño

ABSTRACT

Biplot analysis has been used for studying genotype × environment interaction (GE) or any two-way table. Its descriptive and visualization capabilities along with the availability of user-friendly software have enabled plant scientists to examine any two-way data by a click on a computer button. Despite widespread use, the validity and limitations of biplot analysis have not been completely examined. Here we identify and briefly discuss six key issues surrounding overutilization or abuse of biplot analysis. We question (i) whether the retention of the first two multiplicative terms in the biplot analyses is adequate; (ii) whether the biplot can be more than a simple descriptive technique; (iii) how realistic a "which-won-where" pattern is identified from a biplot; (iv) what if genotypes and/or environments are random effects; (v) how relevant biplot analysis is to the understanding of the nature and causes of interaction; and (vi) how much the biplot analysis can contribute to detection of crossover interaction. We stress the need for use of confidence regions for individual genotype and environment scores in biplots to make critical decisions on genotype selection or cultivar recommendation based on a statistical test. We conclude that the biplot analysis is simply a visually descriptive statistical tool and researchers should proceed with caution if using biplot analysis beyond this simple function.

R.-C. Yang, Agriculture Research Division, Alberta Agriculture and Rural Development, #307 , 700-113 St., Edmonton, AB, T6H 5T6, Canada; and Dep. of Agricultural, Food and Nutritional Science, Univ. of Alberta, Edmonton, AB, T6G 2P5, Canada; J. Crossa and J. Burgueño, Biometrics and Statistics Unit, Crop Informatics Lab. (CRIL), International Maize and Wheat Improvement Center (CIMMYT), Apdo. Postal 6-641, Mexico, D.F., Mexico; P.L. Cornelius, Dep. of Plant and Soil Sciences and Dep. of Statistics, Univ. of Kentucky, Lexington, KY 40546-03121. Received 19 Nov. 2008. *Corresponding author (rong-cai.yang@ualberta.ca).

Abbreviations: AMMI, additive main effects and multiplicative interaction; ANOVA, analysis of variance; BLUP, best linear unbiased prediction; CI, confidence interval; COI, crossover interaction; FA(2), factor analytic model with the first two latent factors; GE, genotype × environment interaction; GGE, genotype main effects and genotype × environment interaction; GL, genotype × location; GLBM, general linear-bilinear model; MET, multi-environment trials; PC, principal component; PCA, principal components analysis; SHMM, shifted multiplicative model; SREG, site regression model; SVD, singular value decomposition.

A large amount of literature on the use of biplot analysis for studying genotype × environment interaction (GE) has recently appeared. Such biplot analyses are based on one or another of several different linear-bilinear models all of which conform to the framework of the general linear-bilinear model (GLBM) (e.g., Cornelius and Seyedsadr, 1997). The additive main effects and multiplicative interaction (AMMI) model and the genotype main effects and genotype × environment interaction effects (GGE) model (fitted to residuals after removal of environment main effects) have been the two most commonly used models for the biplot analysis. The GGE model is sometimes called the Sites Regression (SREG) model

Published in Crop Sci. 49:1564–1576 (2009).
doi: 10.2135/cropsci2008.11.0665
© Crop Science Society of America
677 S. Segoe Rd., Madison, WI 53711 USA

All rights reserved. No part of this periodical may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Permission for printing and for reprinting the material contained herein has been obtained by the publisher.

(Crossa and Cornelius, 1997). Several recent reviews have exhaustively compared and contrasted AMMI and GGE with respect to their suitability for GE analysis (Gauch, 2006; Yan and Tinker, 2006; Yan et al., 2007; Gauch et al., 2008). Basic principles and methodology as well as applications have been described in books on AMMI by Gauch (1992) and on GGE by Yan and Kang (2002).

While there is no denial that the AMMI or GGE biplot analysis is very useful for quick visualization and exploration of patterns inherent in the complex GE two-way table, we have developed a concern with the utility and interpretations of such biplot analyses beyond their functionality and capability. Because of widespread distribution of user-friendly software for AMMI and GGE analyses, it is worrisome if any inadvertent use of the software may lead to dubious results and conclusions. Recent reviews (Gauch, 2006; Yan and Tinker, 2006; Yan et al., 2007; Gauch et al., 2008) focused on what each of the AMMI and GGE biplot analyses can do and on which analysis is better than the other in handling different situations. In this paper, we will instead examine the key issues regarding the limitations of both AMMI and GGE biplot analyses as descriptive and visualization tools and discuss the consequences of ignoring them these limitations. These issues are critical because they are inherently related to the validity and scope of the functionalities and capabilities claimed by proponents of the AMMI or GGE biplot analysis.

BACKGROUND THEORY OF BIPLOT ANALYSIS

Before embarking on such a description and discussion, we will first provide a brief overview of theory and principles underlying the biplot analyses. Consider a set of multi-environment trials (MET) where g genotypes are tested in each of e environments each with r replications. The phenotypic values of individual genotypes are averaged across r replications within each environment, resulting in the $g \times e$ GE cell means arranged as a two-way table. Such a two-way table may be analyzed through the joint use of analysis of variance (ANOVA) and singular value decomposition (SVD) which is also known as principal component analysis (PCA). The analysis is performed using a GLBM (e.g., Cornelius and Seyedsadr, 1997; Cornelius et al., 2001), the general form of which can be written as

$$y_{ij} = \sum_{h=1}^{m} \beta_h x_{hij} + \sum_{k=1}^{t} \lambda_k \alpha_{ik} \gamma_{jk} + \varepsilon_{ij} \qquad [1]$$

where y_{ij} is the mean of the ith genotype in the jth environment; the x_{hij}'s are known constants and the β_h parameters (regression coefficients) are for the linear terms; λ_k's ($\lambda_1 \geq \lambda_2 \geq \ldots \lambda_t$) are scaling constants (singular values) that allow the imposition of orthonormality constraints on the singular vectors for genotypes, [$\alpha_k = (\alpha_{1k}, \ldots, \alpha_{gk})$] and for environments, [$\gamma_k = (\gamma_{1k}, \ldots, \gamma_{ek})$], such that

$$\sum_i \alpha_{ik}^2 = \sum_j \gamma_{jk}^2 = 1 \text{ and } \sum_i \alpha_{ik}\alpha_{ik'} = \sum_j \gamma_{jk}\gamma_{jk'} = 0$$

for $k \neq k'$. The α_{ik} and γ_{jk} for $k = 1, 2, 3 \ldots$ are called "primary," "secondary," "tertiary,"... etc. effects of genotypes and environments, respectively; ε_{ij} is the residual error assumed to be NID $(0, \sigma^2/r)$ with σ^2 being the pooled within-environment error variance.

AMMI and GGE are just the two special cases of the GLBM. The AMMI model is expressed as

$$y_{ij} = \mu + \tau_i + \delta_j + \sum_{k=1}^{t} \lambda_k \alpha_{ik} \gamma_{jk} + \varepsilon_{ij} \qquad [2]$$

where μ is the overall mean, τ_i is the effect of the ith genotype, d_j is the effect of the jth environment and the multiplicative terms in the sum are the same as defined in Eq. [1]. The GGE model is expressed as

$$y_{ij} = \mu + \delta_j + \sum_{k=1}^{t} \lambda_k \alpha_{ik} \gamma_{jk} + \varepsilon_{ij} . \qquad [3]$$

The maximum number of multiplicative terms in the sum is $t = \min(g - 1, e - 1)$ for the full AMMI model and $t = \min(g, e - 1)$ for the full GGE model. In the AMMI model, only the GE interaction is modeled by the bilinear terms, whereas in the GGE or SREG model, the bilinear terms model the main effects of genotypes (G) plus the GE interaction. To present results of fitting Eq. [1] in a biplot, the singular value λ_k is often absorbed by the vectors of genotypic and environmental scores, that is, $\alpha_{ik}^{\star} = \lambda_k^f \alpha_{ik}$ and $\gamma_{jk}^{\star} = \lambda_k^{1-f} \gamma_{jk}$, with $0 \leq f \leq 1$. Obviously, the full models in Eq. [2] and [3] are unparsimonious because all t multiplicative terms in the sum are retained. In most biplot applications, a reduced model with only the first two multiplicative terms being retained is often used, i.e.,

$$y_{ij}^{\star} = y_{ij} - \mu - \tau_i - \delta_j \approx \alpha_{i1}^{\star}\gamma_{j1}^{\star} + \alpha_{i2}^{\star}\gamma_{j2}^{\star} + \varepsilon_{ij} \qquad [4]$$

and

$$y_{ij}^{\star} = y_{ij} - \mu - \delta_j \approx \alpha_{i1}^{\star}\gamma_{j1}^{\star} + \alpha_{i2}^{\star}\gamma_{j2}^{\star} + \varepsilon_{ij} \qquad [5]$$

for AMMI and GGE, respectively, where the approximation indicated in Eq. [4] and Eq. [5] reflects the constraint that the third singular value and all subsequent singular values are zero, i.e., $\sum_{k=3}^{t} \alpha_{ik}^{\star}\gamma_{jk}^{\star} = 0$. If such approximation is inadequate, then more multiplicative terms should be included. In general, the number of PCs for the AMMI and GGE models can be indicated by attaching that number as a suffix to the model name. For example, three commonly used models for the biplot analysis are AMMI1 (the AMMI model with one PC), GGE2 (the GGE model with two PCs) and AMMI2 (the AMMI model with two PCs).

The sampling variance of $\overline{y}_{ij}^{\star}$ is,

$$Var(y_{ij}^{\star}) = \frac{\sigma^2}{\gamma}\left[Var(\hat{\alpha}_{i1}^{\star}\hat{\gamma}_{j1}^{\star}) + Var(\hat{\alpha}_{i2}^{\star}\hat{\gamma}_{j2}^{\star}) + 2Cov(\hat{\alpha}_{i1}^{\star}\hat{\gamma}_{j1}^{\star}, \hat{\alpha}_{i2}^{\star}\hat{\gamma}_{j2}^{\star})\right] \qquad [6]$$

where $Var(\hat{\alpha}^{\star}_{i1}\hat{\gamma}^{\star}_{j1})$, $\mathrm{Var}(\hat{\alpha}^{\star}_{i2}\hat{\gamma}^{\star}_{j2})$ and $2Cov(\hat{\alpha}^{\star}_{i1}\hat{\gamma}^{\star}_{j1}, \hat{\alpha}^{\star}_{i2}\hat{\gamma}^{\star}_{j2})$ are given in Denis and Gower (1994, 1996) and Denis and Pazman (1999). These authors have constructed asymptotic confidence regions for genotypic and environmental scores. While the above parametric approach to constructing confidence regions for genotypic and environmental scores certainly helps to make more reliable decisions on genotype selection and cultivar recommendation, it is not easily implemented for models with more than two bilinear terms, and furthermore requires restrictive assumptions such as asymptotic normality. It is also unclear how the confidence regions constructed under the strictly fixed-effects model can be extended under a mixed-effects model. Here we advocate the use of bootstrapping, a non-parametric resampling technique (Efron, 1982; Yang et al., 1996; Timmerman et al., 2007), for constructing confidence regions for genotypic and environmental scores. Bootstrapping operates by drawing random samples of the same size as the original sample from that sample with replacement and these bootstrap samples are used to construct empirical distributions of estimated genotypic and environmental scores. This non-parametric approach is more flexible, requires no distributional assumption concerning the estimates, and can be used for both fixed- and mixed-effects models. More details will be given later in the context of discussion of needs for statistical testing for these scores in a particular example.

BIPLOT TECHNIQUES

Rank-Two Biplots

A biplot is a scatter plot that graphically displays a point or score for each genotype and each environment (Gabriel, 1971; Kempton, 1984). Recent reviews (Yan and Tinker, 2006; Gauch, 2006; Yan et al., 2007; Gauch et al., 2008) have described in detail two kinds of biplots for the analysis of a two-way GE matrix from the MET data. The first kind includes AMMI2, GGE2, or any other rank-two biplots. A GGE2 biplot graphs scores of environments and genotypes in the first two PCs from SVD of the deviations of cell means from additive environment means with PC1 for its abscissa and PC2 for its ordinate. An AMMI2 biplot graphs scores of environments and genotypes in the first two PCs from SVD of the empirical interactions (i.e., deviations of cell means from additive main effects of genotypes and environments). As described above, GGE applies SVD to the environment-centered two-way data containing G and GE but AMMI applies SVD to the doubly-centered two-way data containing GE only.

The interpretation of GGE2 and AMMI2 biplots is similar. Briefly, the genotypic and environmental scores are represented as vectors in a two-dimensional space. The genotypic and environmental vectors are drawn from the origin (0,0) to the end points determined by their scores. An angle of less than 90° or larger than 270° between a genotypic vector and an environmental vector indicates that the genotype has a positive response at that environment. A negative genotypic response is indicated if the angle is between 90° and 270°. The cosine of the angle between two environments (or genotypes) approximates the phenotypic correlation of the two environments (or genotypes) with an angle of zero indicating a correlation of +1, an angle of 90° (or −90°) a correlation of 0, and an angle of 180° a correlation of −1. A full description of the interpretation of the biplots of multiplicative models is given in Gower and Hand (1996).

Beginning with Yan et al. (2000), GGE2 biplots have often been used to identify "which-wins-where" patterns. Specifically, lines are drawn to connect the markers of the furthest genotypes in the biplot such that they are the corners (i.e., vertices) of an irregular polygon and, for each side of the polygon, drawing a line segment perpendicular to that side of the polygon so as to pass through (or, more commonly, to stop at) the origin. These line segments subdivide the polygon into sectors involving different subsets of environments and genotypes. The genotype, the marker for which is at the corner of one sector, is the best performer in the environments included in that sector. Environments, the markers for which are located far away from the origin, discriminate the genotypes more than those near the origin. Recently there has been an ongoing debate on merits and demerits of AMMI2 vs. GGE2 biplots for genotype and environment identifications (Yan et al., 2007; Gauch et al., 2008).

Rank-One Biplots

Rank-one biplots are the second kind of biplots. One of the commonly used biplots in this category is an AMMI1-based scatter plot where main genotypic and environmental effects are provided for the abscissas and PC1 scores from SVD of the empirical interactions (i.e., deviations of cell means from additive main effects of genotypes and environments) are the ordinates (Zobel et al., 1988). The AMMI1 biplot enables a simultaneous view of the mean performance and the stability of the genotypes. Similarly, in the GGE1 (or SREG1) model, predicted responses ($\hat{\gamma}_{ij} = \hat{\mu} + \hat{\delta}_j + \hat{\alpha}^{\star}_{i1}\hat{\gamma}_{j1}$) are plotted against the environmental scores ($\hat{\gamma}_{j1}$) as shown in Crossa and Cornelius (1997). The plotted points for a given genotype do not necessarily fall on a straight line, but are connected with line segments to form an overlaid set of broken-line graphs, one for each genotype. These broken line graphs help assess whether crossover GE interactions (COIs) are present. The absence of COIs (i.e., the broken lines will not cross one another within the region of plotted points) is indicated if the $\hat{\gamma}_{j1}$ are all of like sign, but the presence of COIs (the broken lines will cross over at one or more points) is not indicated if the $\hat{\gamma}_{j1}$ are not of like sign.

The rank-one biplot analysis is very similar to the classic joint-regression analysis or its variants where the observed performance is fitted to a linear regression of the environmental mean so that the plots of predicted responses vs. environmental means are all straight lines (Yates and Cochran, 1938; Finlay and Wilkinson, 1963; Eberhart and Russell, 1966; Perkins and Jinks, 1968). Significant heterogeneity of slopes of individual lines indicates the presence of GE, but it does not tell whether or not the GE interactions involve COIs. In other words, individual lines may not be paralleled (i.e., different slopes) but they may not be crossed-over to each other either.

ISSUES OF BIPLOT ANALYSIS

Here we identify and briefly discuss six key issues concerning the use of biplot analysis as a descriptive statistical tool. The list of issues is by no means exhaustive and is probably biased toward our own experiences and work on the subject over the years. Nevertheless, it is our belief that these issues critically affect the validity of such analysis, but they are generally ignored by the current biplot literature. We now provide a brief discussion on each of them in the hope that this discussion will raise the awareness of readers of this and other agricultural journals where the biplot analysis articles are frequently published with respect to these issues.

Issue 1: Is Rank-Two Approximation Adequate?

Most applications of the biplot analysis have focused on the use of rank-two approximation, that is, any biplot is constructed using the scores derived from the first two bilinear terms (the first two PCs) to approximate the information content of the two-way GE table. The first two PCs of the two-way GE table are the largest contributors to the total variability in the table; however, unless they can completely explain the total variability, the question of whether this rank-two approximation is adequate would naturally arise.

Yan and Tinker (2006) considered this issue and reviewed several strategies to determine an appropriate number of bilinear terms to be retained. These strategies include judgment from the size of singular values (or their squares) associated with individual PCs and use of cross-validation methods. These authors further suggested that if three or more PCs are required to capture an adequate amount of the total variability in the GE table, either 3D plots simultaneously involving three PCs or multiple 2D biplots can be used. There are two adverse effects if these suggestions are followed. First, inclusion of more than two PCs increases the number of mega-environments recognized (Gauch et al., 2008). As an example, Gauch et al. (2008) considered the Ontario winter wheat (*Triticum aestivum* L.) data (Table 1; Yan et al. 2007). The AMMI1 model identifies two mega-environments that are the same as those of GGE2 identification by Yan et al. (2007). Moreover, AMMI2 to AMMI8 partitions the nine locations into three to six mega-environments, depending on model choice. This immediately raises the question of how many mega-environments really exist for a given MET data set. Second, despite the possibility of including more than two PCs for multiple biplots or high-dimensional plots, practical applications of biplot analysis have rarely gone beyond the first two PCs. The converse side of the issue is that if the first two PCs are used uncritically in a GGE biplot for delineating mega-environments when there is lack of any GE pattern, then a breeding region/zone may be subdivided into several mega-environments, when, in fact, no subdivision is warranted. In this regard, we agree with Gauch et al.'s (2008) warning that such a "mistake increases effort and cost while it decreases accuracy and benefit."

When the first two PCs capture only a small percentage of the total variability but are nevertheless used for a biplot, the patterns identified may be inaccurate or unreliable. To the best of our knowledge, the agricultural biplot literature provides no guidance concerning how much of the total variability accounted for by the first two PCs is considered adequate. To illustrate why such guidance is important, we take the relevant results from Navabi et al. (2006) for the SREG (GGE) analysis of yield data of 472 regional hard-red spring wheat cultivar trials across Alberta from 1981 to 2002 (Table 2). As shown in Table 2, the first two PCs capture a wide range of percentages of the total variability due to the sum of genotype (G) and genotype × location (GL) effects (G + GL), from 45% for the trials in 1993 to 83% in 2002. With almost two-fold difference in the contributions of PC1 and PC2 over years, the biplots based on these PCs are certainly not equally informative from year to year. The results in the last column of Table 2 present yet another but related problem: of three effects, location (L), genotype (G), and genotype × location (GL) interaction, the combined (G + GL) effect is much smaller than the L effect and the (G + GL) contribution can be as little as 2% to the total variability. Since AMMI models involve only the GL interaction effect, the contribution of this interaction effect to the total variability is even smaller. If a small amount of (G + GL) variation is left after removal of the L effect, there may be insufficient resolution for genotype evaluation and identification. Thus, guidance is also needed to determine how large the combined (G + GL) effect should be, relative to the total variability (L + G + GL), to ensure a meaningful biplot on the basis of PC1 and PC2.

Different statistical tests including the usual Gollob's *F*-test (Zobel et al., 1988) or modified *F*-tests known as the F_{GH} test and F_R test (Cornelius et al., 1996) are devised to determine how many PCs should be retained for a

desired precision (to be discussed below) but are insensitive to the question of how much of the total variability the first two PCs should capture before claiming the usefulness of biplots. It is clear from Table 2 that PC1 and PC2 are significant across all years despite (i) the wide range of percentages of total variability that they capture and (ii) a small percentage of the total variability (L + G + GL) that is accounted for by the combined (G + GL) effect in some years (only 2% in 1985). Thus, before some definite theoretical criteria can be developed, some empirical thresholds may be helpful for the practical evaluation of biplots. For example, if the long-term average as shown in Table 2 is used to provide a rule of thumb, then we may recommend that the first two PCs should account for > 60% of the (G + GL) variability and the combined (G + GL) effect should account for >10% of the (L + G + GL) variability before claiming the usefulness of biplots. More research is certainly needed to substantiate such a claim.

When a rank-two approximation is confirmed to be inadequate, AMMI2 and GGE2 biplots would not be very useful for delineating mega-environments and evaluating genotypes. The focus should then be shifted to identify appropriate higher-order AMMI or GGE models that allows for gaining more accuracy by separating a parsimonious and signal-rich model from a noise-rich residual. The cell means in a two-way table that are predicted by an AMMI or GGE model should be more accurate than the simple arithmetic cell means over actual replications because the AMMI or GGE model consider the entire data set to be relevant in predicting future performance by fitting a multiplicative model. The actual MET data comprise a mixture of signal and noise with predictive accuracy increasing with the first few PCs being included, but declining with more PCs being added, a phenomenon known as "Ockham's hill" (Gauch, 2006; Gauch et al., 2008). The best and usually parsimonious model strikes a balance to avoid the models that underfit signal or overfit noise. Criteria other than predictive success for determining the number of significant multiplicative terms in AMMI, GGE (SREG), and other linear-bilinear models include Gollob's *F*-test (Zobel et al., 1988) or the F_{GH} test and F_R test (Cornelius et al., 1996), Akaike information criterion (AIC), or Bayesian information criterion (BIC) (Casanoves et al., 2005), and various schemes for cross validation and model choice (e.g., Gauch, 1988; Cornelius and Crossa, 1999; Dias and Krzanowski, 2003).

Issue 2: Can Biplots Be More than a Simple Descriptive Graphic Tool?

The original intent of a biplot was to reduce the data dimensionality and to enable the data analysts and researchers to have a quick look at relationships among genotypes, among environments, or interactions between genotypes and environments. Thus, biplots were simply a descriptive graphic tool for a quick view. However, many recent applications of biplot analysis have gone beyond the boundary of its limited functionality. As repeatedly claimed in recent review papers (Yan and Tinker, 2006; Gauch, 2006; Yan et al., 2007; Gauch et al., 2008), AMMI or GGE biplot analysis allows for mega-environment delineation, genotypic evaluation, and test environment evaluation. All of these claims are based solely on simple visualization of biplots without any statistical hypothesis testing. This raises a serious concern about the credibility of these claims.

Given that any two-way GE table is a 'sample' data set, the singular values and scores (singular vectors) for genotypes and environments used in the biplot are the point estimates of the corresponding parameter values. The asymptotic variances and covariances of these point estimates along with elliptic confidence regions are given in Denis and Gower (1994, 1996) and Denis and Pazman (1999). As illustrated from the analysis of a rye-grass trial by Denis and Gower (1996), confidence ellipses for each and every genotype and environment must be constructed and imposed on the biplot before a definite conclusion can be reached concerning mega-environment delineation, genotype evaluation, and test environment evaluation.

As already pointed out earlier, it remains unclear how the parametric approach by Denis and others to the construction of confidence regions can be implemented under different linear-bilinear models and its statistical inference relies on restrictive assumptions (e.g., asymptotic normality of individual genotypic and environmental scores). Here we illustrate through the analysis of an example how the non-parametric bootstrapping technique can be used to construct confidence regions on the basis of the empirical distributions of estimated parameters. Once again, we use the Ontario winter wheat example which consists of the yield data of 18 winter wheat genotypes (G1 to G18) tested at nine Ontario locations (E1 to E9). As in the usual GGE biplot analysis, we calculate the deviations of cell means for all 162 (18 × 9) genotype–location combinations from location means (Table 1). The strategies for drawing bootstrap samples for genotype-focused and environment-focused GGE biplots are different. For the genotype-focused GGE biplot ($f = 1$), the genotypic and environmental scores are $\alpha^{\star}_{ik} = \lambda_k \alpha_{ik}$ and $\gamma^{\star}_{jk} = \gamma_{jk}$, respectively. On the other hand, for the environment-focused GGE biplot ($f = 0$), the genotypic and environmental scores are $\alpha^{\star}_{ik} = \alpha_{ik}$ and $\gamma^{\star}_{jk} = \lambda_k \gamma_{jk}$, respectively. A direct application of bootstrapping would require that each bootstrap sample is drawn at random with replacement from the 162 GE cell means. Since the resampling with replacement means that some of the original 162 values will not appear in a bootstrap sample, whereas some others may appear many times, the two-way GE data from the bootstrap sample is obviously unbalanced. Because SVD needs to be done on a balanced data set, we propose

Table 1. Deviations of mean yield (Mg ha^{-1}) of 18 winter wheat cultivar (G1 to G18) tested at nine environments (E1 to E9) from the corresponding environmental means. Genotype and environment scores corresponding to the first and second principal components (PC1 and PC2) and their 95% confidence limits generated by bootstrapping.

	Environments															
Genotype	E1	E2	E3	E4	E5	E6	E7	E8	E9	Mean	PC1	Lower limit	Upper limit	PC2	Lower limit	Upper limit
G1	0.10	−0.29	−0.29	−0.41	0.26	−0.61	0.11	−0.32	−0.23	4.00	−0.31	−0.98	0.39	−0.74	−1.07	0.57
G2	0.06	0.33	−0.23	0.02	0.02	0.09	0.72	0.03	0.04	4.31	0.38	−0.32	0.99	−0.37	−0.73	0.60
G3	0.31	0.14	−0.04	−0.03	0.39	−0.03	0.49	−0.46	−0.28	4.24	0.42	−0.44	1.06	−0.69	−1.00	0.74
G4	0.37	0.31	0.24	0.41	0.54	0.28	−0.01	0.53	0.55	4.54	0.96	0.24	1.54	0.53	−0.29	0.85
G5	0.03	0.16	0.37	0.36	0.09	0.36	0.91	−0.26	−0.07	4.40	0.72	−0.02	1.33	−0.39	−0.79	0.74
G6	0.82	0.04	−0.15	0.28	0.90	−0.01	−0.25	−0.09	−0.12	4.34	0.71	0.02	1.31	−0.13	−1.05	0.89
G7	−0.98	−0.26	−0.40	−0.33	−0.34	−0.79	−0.08	−0.30	−0.87	3.70	−1.34	−1.91	−0.55	−0.85	−1.52	0.42
G8	0.49	0.22	1.29	0.46	−0.14	0.77	−0.07	0.70	0.67	4.67	1.14	0.21	1.78	1.33	−1.04	1.63
G9	0.68	0.30	0.37	−0.05	0.28	−0.20	0.74	0.15	−0.04	4.43	0.89	0.16	1.48	−0.38	−0.73	0.61
G10	0.84	0.22	0.46	0.27	0.26	0.29	−0.34	0.09	0.40	4.46	0.83	0.17	1.42	0.65	−0.61	0.99
G11	−0.07	0.09	−0.38	−0.07	0.46	0.19	0.62	−0.22	0.25	4.28	0.39	−0.43	1.00	−0.55	−0.89	0.67
G12	−1.21	−1.40	−0.75	−1.14	−1.45	−0.80	−0.86	−0.29	−0.80	3.22	−3.01	−3.34	−2.29	−0.30	−1.35	0.50
G13	−0.26	−0.56	−0.84	0.23	−1.12	0.09	−1.64	0.60	−0.01	3.80	−1.64	−2.45	−0.87	1.45	−1.72	1.76
G14	−1.02	−0.59	−0.72	−0.71	−1.05	0.03	−0.96	−0.44	−0.34	3.54	−2.10	−2.41	−1.40	0.18	−1.07	0.66
G15	0.02	0.26	0.52	0.10	0.51	0.08	−0.31	−0.15	0.03	4.30	0.45	−0.02	1.01	0.10	−0.39	0.43
G16	0.58	0.26	−0.19	0.41	0.38	0.27	0.06	−0.06	0.13	4.39	0.68	0.06	1.25	0.09	−0.50	0.56
G17	−0.57	0.53	0.24	−0.14	−0.91	0.24	0.08	0.50	0.48	4.24	−0.24	−0.97	0.48	0.72	−1.20	1.40
G18	−0.12	0.21	0.47	0.42	0.96	−0.23	0.77	0.00	0.21	4.48	1.08	0.18	1.66	−0.63	−0.97	0.87
Mean	4.36	4.44	3.14	3.49	5.68	5.06	4.24	4.36	2.9	4.19						
PC1	0.43	0.34	0.34	0.30	0.49	0.20	0.38	0.08	0.24							
Lower limit	0.21	0.13	0.20	0.01	0.17	−0.02	−0.01	−0.21	0.01							
Upper limit	0.58	0.41	0.52	0.37	0.62	0.38	0.67	0.29	0.39							
PC2	0.14	0.05	0.19	0.21	−0.36	0.37	−0.57	0.43	0.35							
Lower limit	−0.38	−0.14	−0.32	−0.09	−0.61	−0.25	−0.78	−0.35	−0.24							
Upper limit	0.54	0.37	0.55	0.41	0.62	0.45	0.77	0.52	0.48							

an alternative sampling strategy. Rather than randomizing individual cell means in the two-way table, we randomize only either columns or rows (but not both), keeping rows or columns unchanged. Thus, for the genotype-focused GGE biplot, each bootstrap sample consists of 18 random draws (with replacement) from 18 genotypes while keeping nine locations unchanged. This resampling process is repeated 10,000 times to obtain 10,000 bootstrap samples. The estimates of genotypic and environmental scores from the singular value decomposition are computed directly from the original data and from each of 10,000 bootstrap samples. From the bootstrap distribution, we found an approximate 95% confidence interval (CI) for each score examined (Table 1). The CIs for the genotypic scores are generally much larger than those for the environmental scores. This is likely a fair reflection of much larger environment-to-environment variation than genotype-to-genotype variation (the environmental effect accounts for 73.2% of the total variability whereas the genotypic effect accounts for only 13.7%).

Yan and Tinker (2006), Yan et al. (2007), and Gauch et al. (2008) have all used the same data set as given in Table 1 to provide detailed descriptions of mega-environment delineation, genotype evaluation and test-environment evaluation based on GGE or AMMI biplots. Of many results on patterns of relationships and differences among genotypes and environments claimed by these studies, two have repeatedly been identified as successful uses of biplots. The first result is that "the nine test environments fell into two apparent groups: E7 and E5 formed one group, and the remaining environments formed another" (Yan and Tinker, 2006, p. 630). The second result is that "G18 had higher yield in E5 and E7 whereas G8 had higher yield in other environments. This is a clear example of a "cross-over" interaction" (Yan and Tinker, 2006, p. 634). Let us now examine these two specific results in light of CIs for individual genotypic and environmental scores corresponding to PC1 and PC2 developed by bootstrapping in Table 1. It is visually evident from Fig. 1 [and also from Figure 2 of Yan and Tinker (2006) and Fig. 1 of Yan et al. (2007)] that there is little difference among environments on the PC1 axis and that the difference between "the two apparent groups" as stated in the first claim must be due to the difference between the group on the PC2

axis. However, the CIs of individual environmental scores on either PC1 or PC2 axis all overlapped, suggesting that group one (E5 and E7) would not be statistically different from group two (the remaining seven environments). On the second result, once again, it is visually clear from Fig. 1 [and Figure 11 of Yan and Tinker (2006)] that the difference between the two genotypes (G8 and G18) is virtually the difference on the PC2 axis, but the CIs for PC2 scores of these two genotypes overlapped, (–1.04, 1.63) for G8 and (–0.97, 0.87) for G18. Therefore, both results are based merely on visual observations but are not supported by statistical tests. While the development of 2D confidence ellipses of individual genotypic and environmental scores imposed on the biplots would generally facilitate the statistical inference, it is not particularly necessary here because the environmental and genotypic differences involved in evaluating the two results are essentially one dimensional (on the PC2 axis).

The above discussion on the use of CIs for evaluating visually observed claims serves to emphasize that the use of uncertainty measures for genotypic and environmental scores may lead to the different results and conclusions that are drawn from the point estimates of individual scores based on the biplots. Thus, it is our opinion that acceptance of any future submissions to *Crop Science* or other agricultural science journals should be conditional on whether or not appropriate statistical assessment of the biplots is provided. At a minimum, when the uncertainty around genotypic and environmental scores is not measured, the authors need to acknowledge explicitly in their manuscripts that the biplot results should be interpreted with caution.

Issue 3: How Realistic Is a "Which-Won-Where" Pattern from a Biplot?

The proudest contribution that the biplot (particularly GGE-based biplot) literature has claimed is its ability to identify the which-won-where pattern through an extended use of the inner-product property of the biplot. In addition to the usual scatter plot of genotypic and environmental scores in a biplot, such identification requires two more steps: (i) a polygon is first drawn to connect the scores of the genotypes that are furthest from the origin (0,0) with the scores of all remaining genotypes lying within the polygon; (ii) a perpendicular line to each side of the polygon is drawn from the origin. With these two additional steps, the following new interpretation arises. The perpendicular lines to the polygon sides divide the biplot into sectors, each having its own winning cultivar, viz., the marker for which is at the polygon vertex formed by those two polygon sides. The marker for the winning cultivar for a sector is positioned within its winning sector.

The concept of such biplot-based identification is definitely very appealing due to its simplicity and straightforwardness. However, the validity of the interpretation depends critically on the key assumption that the genotypic and environmental scores for polygon construction are the "true" values with no error. There are at least two sources of error with the genotypic and environmental scores. First, as mentioned above, the GE two-way table is a sample data set and all genotypic and environmental scores are simply point estimates with sampling errors. In other words, the so-called wining cultivar identified in a particular sector may not differ significantly from the adjacent, non-wining cultivars if the confidence regions are imposed. Thus any which-won-where pattern based on initial inspection of biplots is simply a curious visual observation only and must be subject to subsequent statistical tests before being definitely recommended for practical utility. Second, the year-to-year

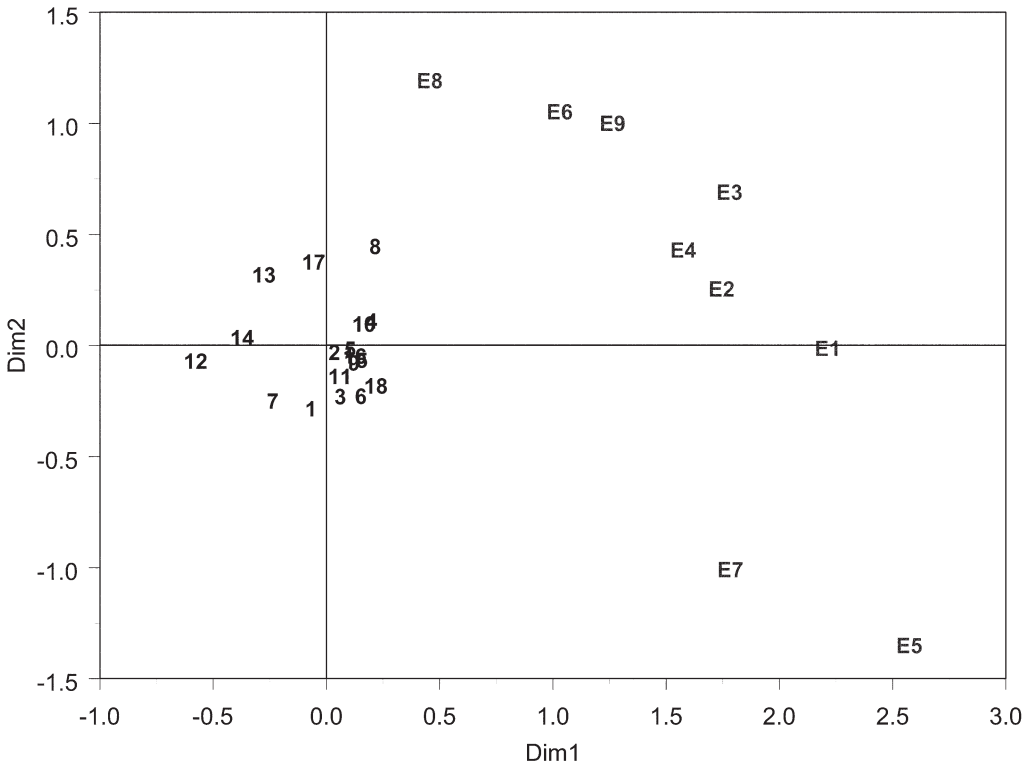

Figure 1. Biplot of the first two latent factors of factor analytic model [FA(2)] based on the two-way GE data in Table 1 with 18 genotypes being identified by numbers 1 to 18 and nine environments being identified by E1 to E9.

variation in yield performance for regional cultivar trials inevitably causes yearly fluctuations in genotypic and environmental scores, thereby trivializing and obscuring any which-won-where pattern that is identified in a single year. Unfortunately, many so-called which-won-where patterns have been based on 1-yr data where *g* genotypes were tested over *e* locations. Strictly speaking, the resulting pattern is one realization among many possible outcomes, and its repeatability in the realization of future years is quite unknown. A repeatable which-won-where pattern over years is the necessary and sufficient condition for mega-environment delineation (Yan et al., 2007).

In carrying out SREG (GGE) biplot analyses of the Alberta hard-red spring wheat regional trials from 1981 to 2002 as described earlier, Navabi et al. (2006) did not find repeatable which-won-where patterns over years and concluded that lack of repeatability of such patterns over years did not allow for further subdivision of spring wheat growing areas in Alberta into distinct mega-environments. Similar observations of unrepeatable which-won-where patterns across years were also made for long-term regional trials of three other major crops [barley (*Hordeum vulgare* L.), canola (*Brassica napus* L.), and field pea (*Pisum sativum* L.)] in the Canadian Prairies (R.-C. Yang, unpublished). Since unpredictable year-to-year weather fluctuation typical in the Canadian Prairies contributes significantly to yield variation and site instability, Yang et al. (2005, 2006) suggested a normalized procedure to filter out much of the random variation among years so that clustering locations would be based on more accurate estimates of location averages. However, it remains largely unknown as to the effectiveness of such a normalization as a general procedure because it depends on the level of imbalance in year × location × genotype combinations and on yearly differences in location × genotype cell means. Thus, it is clear from these discussions that any hasty recommendation regarding the which-won-where pattern on the basis of biplots of one realized set of genotypic and environmental scores without statistical testing may be highly unreliable and misleading.

Furthermore, as Yan et al. (2001) pointed out, the which-won-where patterns are identifiable only if the target mega-environment is adequately sampled and if the correlation between the genotypic PC1 scores and the genotype main effects is positive and almost perfect (>0.95). Ideally, winning genotypes with high and stable performance should have high PC1 scores but negligible PC2 scores. Similarly, a good test environment should have high PC1 scores and negligible PC2 scores so that it allows for more discrimination among genotypes and is more representative of an average environment. However, a near-perfect correlation is not always possible, particularly when the multi-year data are analyzed. Lack of high correlations limits the ability of GGE2 biplot to identify the which-won-where patterns.

Table 2. Partitioning of the total variability due to the sum of genotype (G) and genotype × location (GL) effects (G + GL) into the first two principal components (PC1 and PC2) and the remainder based on principal component analysis (PCA) for the Alberta hard-red spring wheat regional trials during the years 1981 to 2002.

Year	Number of genotypes	Number of locations	PCA partitioning of G + GL†			SS_{G+GL} ‡ / SS_{L+G+GL}
			PC1	PC2	Remainder	
	n		%			
1981	15	28	43	19	37	13
1982	14	23	63	16	22	12
1983	13	22	47	15	38	17
1984	11	14	52	19	30	33
1985	11	24	46	15	39	2
1986	8	29	49	26	25	5
1987	10	32	50	17	33	6
1988	10	30	49	17	34	15
1989	12	31	38	25	37	6
1990	13	32	38	22	40	5
1991	17	30	30	26	44	8
1992	17	18	47	15	38	8
1993	16	26	28	17	55	11
1994	15	14	44	19	37	11
1995	17	13	29	28	43	10
1996	15	14	34	23	42	6
1997	17	16	39	20	41	11
1998	18	17	45	15	40	6
1999	22	19	49	14	37	8
2000	19	20	42	17	41	11
2001	22	13	45	18	37	14
2002	20	7	61	23	17	22
Average	15	21	44	19	37	11

†PC1 and PC2 are all significant at $P < 0.01$ except that PC2 in 1985 is significant at $P < 0.05$, according to F_{GH} test of Cornelius et al. (1996).

‡SS_{G+GL} is the sum of squares due to G + GL and SS_{L+G+GL} is the sum of squares due to L + G + GL.

Issue 4: What if Genotypes or Environments or Both Are Random Effects?

The AMMI or GGE biplot analysis is based on a strictly fixed-effects model with (additive) main effects for genotypes and environments and multiplicative effects for the interaction all being fixed. The discussion from recent works suggests, however, that either genotypic or environmental effects (or thus interaction effects) should be random (Baker, 1996; Piepho, 1998; Smith et al., 2005; Yang 2007). The determination of whether an effect is fixed or random in GE studies is not always easy and has been debated in the literature. Some statisticians have made a pragmatic suggestion that there should be enough

information in the data to estimate variance and covariance parameters of random effects with sufficient precision. For example, Stroup and Mulitze (1991) suggested that a factor (genotype or environment) should have more than 10 levels before it is considered random. Smith et al. (2005) argued that genotypic effects should be random because selection of best lines or cultivars through rankings rather than comparisons is the main goal either in the early "breeding" phase or in advanced "evaluation" phase. Plant breeders would usually consider that years and their interactions with genotypes are random but debate considerably about how locations should be viewed. Part of the location effect would be "fixed" because it represents known physical properties (e.g., soil type of a location) or long-term average (e.g., precipitation or other agro-climatic patterns) of the same location at some future time. However, the goal of most crop improvement programs is to infer about future performance at many untested locations. Thus, it is our opinion that location effects and their interactions with genotypes should be random as well.

Despite different biological considerations, the only criterion used in modern linear model theory (e.g., Littell et al., 2002) for distinguishing fixed and random effects is as follows. If the effect levels reasonably represent a probability distribution, then the effect is random; if, on the other hand, they do not represent a probability distribution, then the effect is fixed. For random effects, the major focus is always on modeling and estimating variance-covariance GE structure. However, interest may sometimes be in estimation and statistical inference about specific levels of a random factor such as breeding values of randomly chosen bulls for milking ability in dairy cattle. This approach pioneered by Henderson (1984) presents two new features in analyzing the MET data. First, test statistics and CIs for random effects need to be constructed. Second, when specific levels of a random effect (e.g., genetic merits of individual cultivars) are of interest, then best linear unbiased prediction (BLUP) rather than calculation of sample means should be used. The BLUP is also known as a shrinkage estimator because the estimate of a random effect is shrunk to adjust for uncertainty arising from its probability distribution (Yang, 2007). The BLUP is devised to maximize the correlation between estimates of the realized values of the random effects and the "true" realized values of the random effects.

Numerous studies have continued the use of AMMI or GGE biplot analysis even when such an analysis based on the strictly fixed-effects model is obviously inappropriate because either genotypic or environmental effects (and thus interaction effects) should be random. Fortunately, a mixed-model analog of the biplot analysis has recently been developed using the factor analytic (FA) model for approximating the variance-covariance GE structure (Piepho 1998; Smith et al., 2002). Further research work done by Crossa et al. (2006) and Burgueño et al. (2008) have described how to model the variance-covariance GE structure but also the possible covariances between genotypes expressed in the pedigree information. These authors have used mixed linear models for incorporating the additive (relationship **A**) matrix and additive × additive covariance matrix into the FA model, and for modeling GE and GGE interactions. Burgueño et al. (2008) has also described the equivalence between SREG2 and FA(2) for finding subsets of genotypes and environments without COI.

In the FA model, the random effect of the *i*th genotype in the *j*th environment (g_{ij}) is expressed as a linear function of latent variables x_{ik} with coefficients δ_{jk} for $k = 1, 2 \ldots t$, plus a residual, η_{ij}, i.e., $g_{ij} = \mu_j + \sum_{k=1}^{t} x_{ik}\,\delta_{jk} + \eta_{ij}$, so that the *ij*th cell mean can be written as $\gamma_{ij} = g_{ij} + \varepsilon_{ij}$. With only the first two latent factors being retained, g_{ij} is approximated by $g_{ij} \approx \mu_j + x_{i1}\delta_{j1} + x_{i2}\delta_{j2} + \eta_{ij}$. As clearly explained by Burgueño et al. (2008), SREG2 can be perceived as consisting of a set of multiple regression equations, one for each environment, each regression equation consisting of an environmental mean or environmental effect as intercept plus two terms for regression on two genotypic regressor variables, α_{i1} and α_{i2} (either observed or latent) with γ_{j1} and γ_{j2} as the regression coefficients. Thus, there is a clear connection between the SREG2 and the FA(2) models. A similar connection between AMMI2 and FA(2) models was also established in Smith et al. (2002).

The interpretation of the loadings and scores of the first two components of FA(2) is the same as that obtained by the SREG2 fixed-effect model. Under the principal component rotation, the directions and projections of the vectors of FA(2) and SREG2 in the biplot are the same. Therefore, the SREG-based properties that the first principal component of SREG2 accounts for non-COI and the second principal component of SREG2 is due to COI variability should hold for FA(2) as well. The FA(2) biplot for the winter wheat example (Fig. 1) shows essentially the same groupings of environments and genotypes as the GGE2(SREG2) biplot as shown in Fig. 1 of Yan et al. (2007). To recognize the similarity between the two biplots, it should be noted from the FA(2) biplot that (i) the genotypic scores are not scaled and (ii) the scores under PC2 include both positive and negative values. However, the absolute values of genotypic and environmental scores under FA(2) and SREG2 models may not necessarily be the same because shrinkage is involved in BLUPs of random effects in the FA(2) model but not in least squares estimates of fixed effects in the SREG2 model. Furthermore, the standard errors of the estimable functions of fixed effects under the SREG2 differ from those of predictable functions of a mixture of fixed and random effects

under FA(2). It should be noted that the FA models are more flexible in handling unbalanced data. The relevance of FA(2) and SREG2 biplots to detection of COIs will be discussed when addressing Issue 6 later.

With the coexistence of FA(2) based on a mixed-effects model and SREG2 or AMMI2 based on a fixed-effects model, the question naturally arises as to which model should be used in practical crop improvement programs. In the early phase of the breeding programs, hundreds or thousands of breeding lines need to be evaluated. The large number of breeding lines is considered as a random sample from a breeding population. It is thus reasonable to assume that genotypic and GE effects are random. After several cycles of selection, however, the number of lines is considerably reduced and these lines are now ready for a comparison with standard "check" cultivars. At this stage, the breeding lines or cultivars may be reasonably considered as fixed [but as discussed earlier, Smith et al. (2005) argued against this consideration]. Thus, FA(2) biplot is preferred at the early stage of breeding programs whereas the SREG2 or AMMI2 biplots may be useful at the late stage of breeding programs.

Issue 5: How Relevant Is the Biplot Analysis to Understanding of Nature and Causes of Interaction?

Descriptions of two-way tables of MET data through biplot or other descriptive statistical analyses are just the very first step toward more in-depth understanding of GE variability. In other words, mere descriptions of GE interactions are of limited value if no further analysis is performed to determine which underlying biological factors cause the observed environmental differentiation (Baker, 1996). In an attempt to address this deficiency, Yan and Tinker (2005) recently suggested a covariate-effect biplot which is generated from the first two PCs of SVD of correlations of target trait (say yield) with every other trait in each of the test environments. Yan and Tinker (2005) argued that the covariate-effect biplot would enable one to determine if GE for yield can be exploited by indirect selection for the other traits. These other explanatory traits may be replaced or augmented by other genetic covariables, such as quantifications of pedigree information, presence/absence of QTLs, or gene expression profiles in a microarray. However, this covariate-effect biplot focuses only on inclusion of phenotypic and genetic covariables and ignores the presence of GE interactions in the explanatory traits. More importantly, the covariate-effect biplot, like its predecessor, remains merely a descriptive graphic tool and provides no functional relationship and predictability with explanatory covariables necessary for any understanding of underlying biological causes of GE interactions. To examine effects of both genetic and environmental covariables and to develop functional relationships and predictability with explanatory covariables, factorial regression or partial least squares analysis (e.g., Vargas et al., 1999; Brancourt-Hulmel and Lecomte, 2003; van Eeuwijk et al., 2005) may be more useful. These analyses are particularly useful in obtaining more parsimonious models for predicting the GE variability. Thus, initial biplot or any other descriptive statistical analyses should generally be followed up by the development of prediction models that allow for identifying causative factors of GE variability.

The linearity assumption required for the above analysis of the response to genotypic and environmental covariables may not be warranted as most physiological and developmental processes during plant growth and production proceed in a nonlinear fashion (Baker, 1988a; Wu et al., 2004; van Eeuwijk et al., 2005). Classical nonlinear response functions include yield-density curves, sigmoid curves, and asymptotic curves (Baker 1988a). Yield-density functions are needed to describe the relationship between crop yield and seeding density; sigmoid functions (e.g., logistic curve) are often used to model different patterns of vegetative growth; and asymptotic functions are useful for specifying the relationships between yield and amount of fertilizer application. More elaborated response functions are also described in Wu et al. (2004) and van Eeuwijk et al. (2005). As acknowledged in Yan and Tinker (2006), the biplot analysis is PCA-based, using only linear correlations among genotypes or environments, and thus it is not capable of detecting the nonlinear relationships such as those described above. In this case, nonlinear biplots based on Euclidean or non–Euclidean dissimilarity measures (Gower and Hand, 1996, Chapter 6) may be useful for exploiting nonlinear relationships among genotypes or environments.

Issue 6: Can the Biplot Analysis Contribute to Detection of Crossover Interaction?

Geneticists and plant breeders (e.g., Haldane, 1947; Gregorius and Namkoong, 1986; Baker, 1988b, 1996; Cornelius et al., 1993) have long recognized that GE is of little consequence in selection programs unless COIs are predominant. In the absence of COIs, GE is simply caused by differences in scales, and the best genotype in one environment remains the best in all other environments. Statistical tests for COIs have been described under the fixed-effect model (Baker, 1988b; Cornelius et al., 1992), random-effect model (Yang, 2002), and mixed-effect model (Yang, 2007). However, the usual AMMI2 or GGE2 biplot analysis does not distinguish if interactions are COIs or non-COIs. Thus, it is important to determine whether the GE variability captured in the biplots is relevant to predicting the presence of COIs. Zobel et al. (1988) first popularized the AMMI1 biplot which allows for simultaneous views of the main performance of genotypes and their stability (interaction) as characterized by AMMI PC1 scores. In

other words, such a biplot can identify some genotypes with different main performance but little interaction, or others with little main difference but large interaction. However, once again it is unknown whether or not such interaction involves COIs. The use of a SREG1 (or GGE1) biplot based on a constrained SVD non-COI PC1 solution (e.g., Crossa and Cornelius, 1997) has been made to predict the absence of COIs based on the earlier work of rank-one shifted multiplicative model (SHMM1) by Cornelius et al. (1992). If the SHMM1 model is an adequate approximation to the two-way GE data, and primary effects of the environments (i.e., PC1 scores) are either all non-positive or non-negative, then the SHMM1 model has the two proportionality properties. First, differences between genotypes in any single environment are proportional to genotypic differences in any other environment. Second, differences between environments in terms of the performance of any single genotype are proportional to those in terms of the performance of any other genotype. The second proportionality restriction is not required for assessing genotypic non-COI status and is thus removed in the SREG1 model. The SHMM1 or SREG1 biplot portrays the graph consisting of a set of regression lines, one for each genotype, all concurring at a point either at the boundary or outside of the region containing the plotted points. If, on the other hand, the PC1 scores are of different signs, the SHMM1 and SREG1 biplots show the presence of COIs (i.e., the regression lines intersect at a point within the region containing the plotted points).

More recently, Crossa et al. (2002) pointed out that if the PC1 scores of environments from SREG2 or SHMM2 models are of the same sign, then the interaction variation of the PC1 is due largely to the presence of few or no COIs and the variation accounted for by the PC2 is due to COI; if, on the other hand, the PC1 scores of environments are of different signs, then the interaction variation is due largely to the presence of numerous COIs. In other words, the SHMM2 and SREG2 biplots of the first two PCs would represent the graph of non-COI variation (PC1) vs. COI variation (PC2). This is possible only if the environment scores of PC1 are all of the same sign or if the environment scores of PC1 are of different signs but a constrained SHMM2 or SREG2 solution ensures that the first multiplicative term should show a non-COI pattern. Thus, COIs should be negligible in the winter wheat example (Table 1) because the environmental PC1 scores are of the same sign. This is confirmed from statistical tests for COIs (Baker, 1988b; Cornelius et al., 1992; Yang, 2007). In the winter wheat example, the total number of 2×2 interaction contrasts (quadruples) for COI evaluation is 5508 ($18 \times 17 \times 9 \times 8/4$). With four replications and the error mean square of 0.067 with 483 degrees of freedom as obtained from Table 2 of Yan et al. (2000), the percentages of the quadruples with significant COIs ranged from 3.9% for the original Azzalini-Cox test (the most conservative test) to 12.2% for the most sensitive interaction-wise test (Cornelius et al., 1992). All of these tests are performed under the fixed-effects model. The COI frequency would be further reduced under the mixed- and random-effects models (Yang, 2007). These discussions serve to emphasize that a display of a biplot alone is not sufficient to determine the presence of COIs. The use of the proportionality properties along with a constrained SHMM or SREG solution for non-COIs must be made. Previous research (e.g., Cornelius et al., 1993; Crossa et al., 1995; Crossa and Cornelius, 1997; Crossa et al., 2004) has used the proportionality properties and constrained SHMM and SREG solutions for subsetting groups of environments or genotypes within which COIs are negligible. The use of FA(2) model for subsetting genotypes and environments in the winter wheat data on the basis of the approach of Burgueño et al. (2008) shows that there are only 30 significant COIs, which represents 0.54% of the total quadruples. In other words, the majority of the interactions are due to changes in scale.

FUTURE OUTLOOK

There is no doubt that a biplot, whether it is based on AMMI, GGE, or any other linear-bilinear model, is a useful visualization technique to quickly explore patterns of similarity or dissimilarity among genotypes or environments, and extract useful information from complex GE data. As demonstrated constantly in recent reviews (e.g., Yan and Tinker, 2006; Gauch, 2006; Yan et al., 2007; Gauch et al., 2008), visual information displayed in a given biplot is indeed easier and faster to convey and grasp than tabular numerical information, thereby facilitating appropriate interpretation of complex GE data sets. However, like any other statistical method used for elucidating GE patterns in MET data, the biplot analysis is not without its limitations. In this paper, we are particularly concerned with the utility and interpretations of such a biplot analysis beyond its functionalities and capabilities. Specifically, we discuss the six key issues concerning possible misuse or overutilization of AMMI or GGE biplot analysis and raise questions about the validity and scope of the functionalities and capabilities if such inadvertent applications of the biplot analysis are allowed. To raise these issues is not necessarily meant to discredit the usefulness of the biplot analysis. In fact, we believe the resolutions to these and other issues will help enhance the value that is already captured by the biplot analysis. For example, inclusion of the FA(2) biplot would expand the functionalities and capabilities of a biplot from a strictly fixed-effects model to a mixed-effects model, thereby allowing for the use of biplot analysis for modeling variance-covariance GE structure or incorporating pedigree information on relationships between relatives. However, proponents and users of the biplot analysis must recognize and acknowledge its fundamental limitations.

We offer two take-home messages for future endeavors with biplot analyses. First, a biplot is simply a descriptive, graphical tool and cannot be used for hypothesis testing because there is no uncertainty measure. Selection of superior genotypes during breeding or recommendation of best cultivars from MET data are critical decisions that plant breeders, agronomists, or crop producers have to make but which require sound scientific bases rather than based merely on subjective judgment calls from visualization. The parametric approach (Denis and Gower, 1994, 1996) as well as our bootstrapping non-parametric approach to obtaining uncertainty measures for genotypic and environmental scores will both definitely help put those critical decisions on sound statistical and scientific bases. The development and application of these parametric and non-parametric approaches should be a focus of future research to add statistical inference capability to the biplot analysis. This is analogous to the usual difference between descriptive vs. inferential statistics. Descriptive statistics focus on collecting, organizing, summarizing, presenting, and analyzing data, but without drawing any conclusion or inference. On the other hand, inferential statistics is the science of decision making in the face of uncertainty. The biplot analysis alone is like descriptive statistics in that it is insufficient to test hypotheses and draw sound, definitive conclusions. Second, biplots alone are not sufficient for fully assessing the complex GE structure. More complete understanding of the GE structure requires important supplementary information such as adequacy of the number of multiplicative terms being retained, features of different biplots, genetic and environmental covariables, constrained bilinear solutions for non-COIs, and complete assessment of the genetic correlation among environments and relationship between genotypes.

Finally, while this paper focuses on a critical evaluation of the biplot analysis as a statistical tool, it should not be forgotten that the biplot analysis is only one of many perspectives in the long and rich GE literature and the recent overemphasis on visual description and characterization of GE interaction by biplot proponents has sidetracked other important perspectives centering around impacts of GE on selection response and breeding strategy (e.g., Atlin et al., 2000; Piepho and Mohring, 2005). Thus, the major future research effort should be directed toward reinstalling those other perspectives, including (i) definition and delineation of the target population of environments for optimal selection response, (ii) incorporation of GE and multiple traits in marker-assisted selection and genomic selection, (iii) use of mixed-model and empirical Bayesian approaches for ranking genotypes in the presence of random GE effects, and (iv) development of linear and nonlinear models for predicting response to environmental and genetic covariables.

Acknowledgments

We thank Gary Atlin and three anonymous reviewers for helpful comments. This research was supported by Alberta Agriculture and Rural Development and the Natural Sciences and Engineering Research Council of Canada grant.

References

Atlin, G.N., R.J. Baker, K.B. McRae, and X. Lu. 2000. Selection response in subdivided target regions. Crop Sci. 40:7–13.

Baker, R.J. 1988a. Differential response to environmental stress. p. 492–504. *In* B.S. Weir et al. (ed.) Proceedings of the Second International Conference on Quantitative Genetics. Sinauer, Sunderland, MA.

Baker, R.J. 1988b. Tests for crossover genotype-environmental interactions. Can. J. Plant Sci. 68:405–410.

Baker, R.J. 1996. Recent research on genotype-environmental interaction. p. 235–239. *In* Proc. V Intl. Oat Conf. and VII Intl. Barley Genet. Sym., Vol. 1. Invited papers. Univ. Ext. Press, Univ. of Saskatchewan, SK, Canada.

Brancourt-Hulmel, M., and C. Lecomte. 2003. Effect of environmental variates on genotype × environment interaction of winter wheat: A comparison of biadditive factorial regression to AMMI. Crop Sci. 43:608–617.

Burgueño, J., J. Crossa, P.L. Cornelius, and R.-C. Yang. 2008. Using factor analytic models for joining environments and genotypes without crossover genotype × environment interaction. Crop Sci. 48:1291–1305.

Casanoves, F., J. Baldessari, and M. Balzarini. 2005. Evaluation of multienvironment trials of peanut cultivars. Crop Sci. 45:18–26.

Cornelius, P.L., and J. Crossa. 1999. Prediction assessment of shrinkage estimators of multiplicative models for multi-environment cultivar trials. Crop Sci. 39:998–1009.

Cornelius, P.L., and M.S. Seyedsadr. 1997. Estimation of general linear-bilinear models for two-way tables. J. Statist. Comput. Simulation 58:287–322.

Cornelius, P.L., J. Crossa, and M.S. Seyedsadr. 1996. Statistical tests and estimators of multiplicative models for genotype-by-environment interaction. p. 199–234. *In* M.S. Kang and H.G. Gauch (ed.) Genotype-by-environment interaction. CRC Press, Boca Raton, FL.

Cornelius, P.L., J. Crossa, M.S. Seyedsadr, G. Liu, and K. Viele. 2001. Contributions to multiplicative model analysis of genotype-environment data. Statistical Consulting Section, American Statistical Association, Joint Statistical Meetings. 2001 Proceedings, American Statistical Association, Aug. 7, 2001 [CDROM]. Atlanta, GA.

Cornelius, P.L., M. Seyedsadr, and J. Crossa. 1992. Using the shifted multiplicative model to search for "separability" in crop cultivar trials. Theor. Appl. Genet. 84:161–172.

Cornelius, P.L., D.A. Van Sanford, and M.S. Seyedsadr. 1993. Clustering cultivars into groups without rank-change interactions. Crop Sci. 33:1193–1200.

Crossa, J., J. Burgueno, P.L. Cornelius, G. McLaren, R. Trethowan, and A. Krishnamachari. 2006. Modeling genotype × environment interaction using additive genetic covariances of relatives for predicting breeding values of wheat genotypes. Crop Sci. 46:1722–1733.

Crossa, J., P.L. Cornelius, K. Sayre, and J.I. Ortiz-Monasterio. 1995. A shifted multiplicative model fusion method for grouping environments without cultivar rank change. Crop Sci. 35:54–62.

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

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

Crossa, J., R.-C. Yang, and P.L. Cornelius. 2004. Studying crossover genotype × environment interaction using linear-bilinear models and mixed models. J. Agric. Biol. Environ. Stat. 9:362–380.

Denis, J.-B., and J.C. Gower. 1994. Asymptotic covariances for parameters of biadditive models. Util. Math. 46:193–205.

Denis, J.-B., and J.C. Gower. 1996. Asymptotic confidence regions for biadditve models: Interpreting genotype-environment interactions. Appl. Stat. 45:479–493.

Denis, J.-B., and A. Pazman. 1999. Bias of LS estimators in nonlinear regression models with constraints. Part II: Biadditive models. Appl. Math. 44:375–403.

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

Eberhart, S.A., and W.A. Russell. 1966. Stability parameters for comparing varieties. Crop Sci. 6:36–40.

Efron, B. 1982. The jackknife, the bootstrap and other resampling plans. S.I.A.M., Philadelphia, PA.

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

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. 1992. Statistical analysis of regional yield trials: AMMI analysis of factorial designs. Elsevier, Amsterdam, The Netherlands.

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

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

Gower, J.C., and D.J. Hand. 1996. Biplots. Chapman and Hall, London.

Gregorius, H.-R., and G. Namkoong. 1986. Joint analysis of genotypic and environmental effects. Theor. Appl. Genet. 72:413–422.

Haldane, J.B.S. 1947. The interaction of nature and nurture. Ann. Eugen. 13:197–205.

Henderson, C.R. 1984. Applications of linear models in animal breeding. University of Guelph Press. Guelph, Canada.

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

Littell, R.C., W.W. Stroup, and R.J. Freund. 2002. SAS for linear models, 4th ed. SAS Institute Inc., Cary, NC.

Perkins, J.M., and J.L. Jinks. 1968. Environmental and genotype-environmental components of variability III. Multiple lines and crosses. Heredity 23:339–356.

Navabi, A., R.-C. Yang, J. Helm, and D.M. Spaner. 2006. Can spring wheat-growing megaenvironments in the northern Great Plains be dissected for representative locations or niche-adapted genotypes? Crop Sci. 46:1107–1116.

Piepho, H.P. 1998. Methods for comparing the yield stability of cropping systems—a review. J. Agron. Crop Sci. 180:193–213.

Piepho, H.P., and J. Mohring. 2005. Best linear unbiased prediction of cultivar effects for subdivided target regions. Crop Sci. 45:1151–1159.

Smith, A., B.R. Cullis, and R. Thompson. 2002. Exploring variety-environment data using random effects AMMI models with adjustment for spatial field trends: Part 1: Theory. *In* M.S. Kang (ed.) Quantitative genetics, genomics and plant breeding. CABI Publishing, Oxford, UK.

Smith, A.B., B.R. Cullis, and R. Thompson. 2005. The analysis of crop cultivar breeding and evaluation trials: An overview of current mixed model approaches. J. Agric. Sci. 143:1–14.

Stroup, W.W., and D.K. Mulitze. 1991. Nearest neighbor adjusted best linear unbiased prediction. Am. Stat. 45:194–200.

Timmerman, M.E., H.A.L. Henk, and A.K. Smilde. 2007. Estimating confidence intervals for principal components loadings: A comparison between the bootstrap and asymptotic results. Br. J. Math. Stat. Psychol. 60:295–314.

van Eeuwijk, F.A., M. Malosetti, X. Yin, P.C. Struik, and P. Stam. 2005. Statistical models for genotype by environment data: From conventional ANOVA models to eco-physiological QTL models. Aust. J. Agric. Res. 56:883–894.

Vargas, M., J. Crossa, F.A. van Eeuwijk, M.E. Ramírez, and K. Sayre. 1999. Using partial least squares regression, factorial regression, and AMMI models for interpreting genotype × environment interaction. Crop Sci. 39:955–967.

Wu, R.L., C.-X. Ma, M. Lin, and G. Casella. 2004. A general framework for analyzing the genetic architecture of developmental characteristics. Genetics 166:1541–1551.

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 N.A. Tinker. 2005. An integrated system of biplot analysis for displaying, interpreting, and exploring genotype-by-environment interactions. Crop Sci. 45:1004–1016.

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

Yan, W., L.A. Hunt, Q. Sheng, and Z. Szlavnics. 2000. Cultivar evaluation and mega-environment investigation based on GGE biplot. Crop Sci. 40:596–605.

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., M.S. Kang, B. Ma, S. Woods, and P.L. Cornelius. 2007. GGE biplot vs. AMMI analysis of genotype-by-environment data. Crop Sci. 47:643–655.

Yang, R.-C. 2002. Likelihood-based analysis of genotype–environment interactions. Crop Sci. 42:1434–1440.

Yang, R.-C. 2007. Mixed model analysis of crossover genotype-environment interactions. Crop Sci. 47:1051–1062.

Yang, R.-C., F.C. Yeh, and A.D. Yanchuk. 1996. A comparison of isozyme and quantitative genetic variation in *Pinus contorta* spp. *latifolia* by F_{ST}. Genetics 142:1045–1052.

Yang, R.-C., S.F. Blade, J. Crossa, D. Stanton, and M.S. Bandara. 2005. Identifying isoyield environments for field pea production. Crop Sci. 45:106–113.

Yang, R.-C., D. Stanton, S.F. Blade, J. Helm, D. Spaner, S. Wright, and D. Domitruk. 2006. Isoyield analysis of barley cultivar trials in the Canadian Prairies. J. Agron. Crop Sci. 192:284–294.

Yates, F., and W.G. Cochran. 1938. The analysis of groups of experiments. J. Agric. Sci. 28:556–580.

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