Statistical Analysis of Yield Trials by AMMI and GGE

Hugh G. Gauch, Jr.*

ABSTRACT

The Additive Main effects and Multiplicative Interaction (AMMI) model, Genotype main effects and Genotype × Environment interaction (GGE) model, and Principal Components Analysis (PCA) are singular value decomposition (SVD) based statistical analyses often applied to yield-trial data. This paper presents a systematic comparison, using both statistical theory and empirical investigations, while considering both current practices and best practices. Agricultural researchers using these analyses face two inevitable choices. First is the choice of a model for visualizing data. AMMI is decidedly superior, not for statistical reasons, but rather for agricultural reasons. AMMI partitions the overall variation into genotype main effects, environment main effects, and genotype × environment interactions. These three sources of variation present agricultural researchers with different challenges and opportunities, so it is best to handle them separately, while still considering all three in an integrated manner. Second is the choice of a member of a given model family for gaining predictive accuracy. AMMI, GGE, and other SVD-based model families are essentially equivalent, but best practices require model diagnosis for each individual dataset to determine which member is most predictively accurate. Making these two choices well allows researchers to extract more usable information from their data, thereby increasing efficiency and accelerating progress.

THE YIELD TRIAL is one of the most common experiments in agricultural research, typically testing a number of genotypes in a number of environments. Accordingly, effective statistical analysis of yield trials can help agronomists, breeders, and other agricultural researchers to make faster progress. Furthermore, these statistical analyses have broader relevance for agricultural researchers because they pertain to any two-way data matrices, and such data emerge from many kinds of experiments. Two frequently used statistical analyses have been the Additive Main effects and Multiplicative Interaction (AMMI) model and the Genotype main effects and Genotype × Environment interaction effects (GGE) model.

As documented momentarily, the GGE literature claims that (i) GGE graphs are superior to AMMI for visualizing patterns in yield-trial data, especially for showing which genotype won where and thereby delineating megaenvironments, and (ii) GGE is equivalent to AMMI for gaining predictive accuracy. On the contrary, this paper argues that (i) after documenting several errors in the GGE literature, the verdict is reversed, and AMMI is decidedly superior for visualizing data, and (ii) GGE and AMMI can be equivalent for gaining accuracy, given the selfsame principles in both cases that

involve parsimony, but best practices require model diagnosis for each individual dataset.

Literally hundreds of papers have used AMMI or GGE (or occasionally both) to analyze agricultural experiments. There are also books on AMMI by Gauch (1992) and on GGE by Yan and Kang (2002).

What has not yet been produced, however, is a systematic evaluation and comparison of AMMI and GGE for various research objectives and for various dataset properties. This paper combines theoretical considerations and empirical investigations to provide this systematic comparison. Knowledge of the strengths or weaknesses of AMMI and GGE for a given objective and a given dataset can help agricultural researchers to get more from their data.

Research purposes are organized here in two clusters: visualizing data and gaining accuracy. Visualizing yield-trial data includes understanding complex genotype × environment (GE) interactions, determining which genotype won in which environments, and grouping environments with the same winner (or similar winners) into a megaenvironment. Gaining accuracy improves selection success as well as increases repeatability, simplifies conclusions, improves recommendations, and accelerates progress.

When evaluating any statistical methodology, a distinction must be made between the method's current performance given typical practices and its optimal performance given best practices. The former is largely determined by available software and influential publications, whereas the latter is determined by statistical theory and logical principles. This paper critiques current practices and software. But the emphasis is on how AMMI and GGE would perform were they implemented with best practices. Because current practices are both labile and remediable, a comparison of best-practices AMMI with best-practices GGE is especially relevant.

Statistical Theory

Some statistical theory is a prerequisite to understanding and comparing AMMI and GGE. The data structure that these statistical analyses address is a rows × columns, two-way data matrix, that is, a two-way factorial design such as a number of genotypes tested in a

Crop and Soil Sciences, 519 Bradfield Hall, Cornell University, Ithaca, NY 14853. Received 12 July 2005. *Corresponding author (hgg1@cornell.edu).

Abbreviations: AEC, average environment coordinate; AIC, Akaike information criterion; AMMI, Additive Main effects and Multiplicative Interaction; AOV, analysis of variance; ATC, average tester (genotype) coordinate; BIC, Bayesian information criterion; df, degrees of freedom; GE, genotype × environment; GE′, prime, signal-rich part of the interaction; GE*, spurious, noise-rich part of the interaction; GGE, Genotype main effects and Genotype × Environment interaction; PC, principal component; PCA, principal components analysis; SS, sum of squares; SVD, singular value decomposition.

number of environments. The experiment may or may not be replicated. These analyses combine two statistical procedures: analysis of variance (AOV) and singular value decomposition (SVD), which is also sometimes called principal components analysis (PCA).

SVD was invented by Pearson (1901). A data matrix can be envisioned geometrically as points for each row in a space with as many dimensions or axes as there are columns (or the reverse). SVD derives new axes, the principal components (PC). The first PC captures as much of the variation as possible, so it is the least-squares line through the original cloud of points; and likewise, the first two components define the least-squares plane. The variation captured can be expressed as a percentage of the original, total variation. In the frequent case that the first few components capture much of the total variation, SVD provides a useful low-dimensional representation of the data.

AOV was invented by Fisher (1918). For a two-way factorial design, AOV partitions the variation into the row main effects, column main effects, and row $\times$ column interaction effects. For yield trials, the most common outcome is that the environment main effects are largest, followed by the interaction effects and then the genotype main effects.

Fisher and Mackenzie (1923) were first to apply both SVD and AOV, separately, to the same dataset, a potato (*Solanum tuberosum* L.) yield trial. Williams (1952), closely followed by Pike and Silverberg (1952), were first to combine AOV and SVD into an integrated analysis, using AOV to partition the interaction and then applying SVD to this interaction (rather than to the original data). Although AOV involves easy calculations, SVD requires laborious iterative calculations, so necessarily SVD remained obscure until scientists began to have reasonable access to digital computers in the 1960s. Subsequent developments are noted by Gauch (1992:11). But Kempton (1984) was the first publication in the agricultural literature that substantially accelerated interest, and it used both AMMI and GGE. Zobel et al. (1988) built on that work, further popularizing AMMI. About a decade later, several papers popularized GGE, beginning with Yan et al. (2000). At present, AMMI and GGE are among the foremost statistical methods for analyzing yield-trial data.

The research purpose of this literature was visualizing and comprehending complex data, especially complex GE interactions. But Gauch (1988) and Gauch and Zobel (1988) introduced a new research purpose for AMMI models, namely gaining accuracy, and explained this gain by statistical theory that had been well known for several decades. Subsequently, accuracy gain has been demonstrated for many additional SVD-class models, including GGE (Cornelius and Crossa, 1999; Dias and Krzanowski, 2003).

PCA has numerous variants, differing in the transformation applied to the data before performing SVD. GM-centered PCA applies SVD to the original data after subtracting the grand mean:

$$Y_{ger} - \mu = \Sigma_n\lambda_n\gamma_{gn}\delta_{en} + \rho_{ge} + \varepsilon_{ger} \qquad [1]$$

where Y_{ger} is the yield for genotype g in environment e for replicate r, μ is the grand mean, summation is done over components $n = 1$ to N, λ_n is the singular value for principal component n, γ_{gn} is the eigenvector score for genotype g and component n, δ_{en} is the eigenvector score for environment e and component n, ρ_{ge} is the residual for genotype g and environment e, and ε_{ger} is the error for genotype g and environment e and replicate r. Normally just a few components are used, with N usually between 1 and 5, to achieve dimensionality reduction, which leaves a residual. Because paper and hence ordinary graphs are two dimensional, $N = 2$ is common.

GGE, also called environment-centered PCA, applies SVD to the data minus the environment means, which automatically also removes the grand mean:

$$Y_{ger} - \beta_e = \Sigma_n\lambda_n\gamma_{gn}\delta_{en} + \rho_{ge} + \varepsilon_{ger} \qquad [2]$$

where β_e is the mean or average for environment e.

AMMI, also called doubly-centered PCA, applies SVD to the data minus the genotype and environment means, which necessitates adding back the grand mean to avoid removing it twice:

$$Y_{ger} - \alpha_g - \beta_e + \mu = \Sigma_n\lambda_n\gamma_{gn}\delta_{en} + \rho_{ge} + \varepsilon_{ger} \qquad [3]$$

where α_g is the mean for genotype g. Hence AMMI applies SVD to the matrix of residuals from the additive model, that is, the interaction.

Note that AOV estimates parameters that are added, whereas SVD or PCA estimates parameters that are multiplied. To avoid another multiplication by the singular value, both sets of eigenvectors (for a given component or axis) may be multiplied by the square root of the singular value to produce genotype and environment PC scores whose products then approximate the interactions directly. The square of a singular value is an eigenvalue, which is the sum of squares (SS) captured by a given component.

To distinguish the members of a model family, the number of components is appended, such as GGE2 having two PCs (before relegating higher PCs to the residual) and AMMI1 having one PC. The full model, leaving no residual, is denoted by GGEF and AMMIF.

To understand the implications of these three variants of PCA for practical data analysis, the first and foremost requisite is to understand which effects are captured by which components. Simply multiplying a given component's genotype and environment scores produces a matrix that can then be subjected to AOV to determine its genotype, environment, and interaction sums of squares. A couple of examples may introduce this topic.

Table 1 presents three parallel analyses for the soybean [*Glycine max* (L.) Merr.] yield data given in Gauch (1992:56). This dataset from New York State has seven genotypes, 10 environments, and four replications within each environment. As shown at the bottom of this table, the SS for genotypes is 7117669, for environments is 176360099, and for interactions is 39728708. For GM-centered PCA, every component contains a mixture of all three effects. But in accordance with their relative magnitudes, environment effects dominate PC1, inter-

Table 1. Recovery of genotype, environment, and interaction effects by the parameters of three variants of principal components analysis for a soybean yield trial.

Component	Genotype	Environment	Interaction
GM-centered PCA			
PC1	35 748	175 805 412	5 728 465
PC2	6 605	488 077	29 543 887
PC3	7 046 267	45 437	395 230
PC4	20 643	19 623	2 669 450
Residual	8 406	1 550	1 391 676
Environment-centered PCA (GGE)			
PC1	12 419	0	32 753 455
PC2	6 526 027	0	2 202 430
PC3	568 531	0	3 181 183
PC4	1 217	0	988 822
Residual	9 475	0	602 818
Doubly-centered PCA (AMMI)			
Main	7 117 669	176 360 099	0
PC1	0	0	32 756 258
PC2	0	0	4 681 150
PC3	0	0	1 019 857
Residual	0	0	1 271 443
Sum	7 117 669	176 360 099	39 728 708

action effects PC2, and genotype effects PC3. For GGE, which excludes environment effects, every component contains a mixture of genotype and interaction effects. The interaction effects dominate PC1 and the genotype effects dominate PC2. Because it is less distracted by genotype effects, PC3 captures more interaction than does PC2, so the recovery of interaction is not monotonic. Finally, for AMMI, the main effects simply are the entire genotype and environment effects, leaving PC1 and higher components to capture interaction exclusively in a monotonic sequence that decreases from the first and largest component to the last and smallest component.

It is important to realize, however, that other datasets can display other patterns. For instance, Table 2 presents these same three analyses for the Kentucky bluegrass (*Poa pratensis* L.) quality data discussed in Ebdon and Gauch (2002a, 2002b). This trial for the USA and Canada has 125 genotypes, 69 environments, and three replications within each environment. For GM-centered PCA, PC1 is dominated by environment effects as before, but now PC2 is dominated by genotype effects,

Table 2. Recovery of genotype, environment, and interaction effects by the parameters of three variants of principal components analysis for a Kentucky bluegrass quality trial.

Component	Genotype	Environment	Interaction
GM-centered PCA			
PC1	23.12	22 747.85	411.60
PC2	3303.70	29.49	1315.04
PC3	37.58	6.43	1537.28
PC4	57.68	0.98	1001.15
Residual	57.80	1.31	5608.14
Environment-centered PCA (GGE)			
PC1	3331.91	0	1442.81
PC2	12.20	0	1667.82
PC3	76.79	0	1003.39
PC4	0.50	0	655.40
Residual	58.48	0	5103.79
Doubly-centered PCA (AMMI)			
Main	3479.88	22 786.06	0
PC1	0	0	1944.19
PC2	0	0	1600.67
PC3	0	0	861.01
Residual	0	0	5467.34
Sum	3479.88	22 786.06	9873.21

although it also captures considerable interaction effects. For GGE, the genotype effects appear primarily in PC1, rather than PC2 as before. All of the first four PC contain substantial interaction in amounts that are not monotonic, with PC2 capturing more than PC1. Incidentally, the recovery of genotype effects is also not monotonic, with PC3 capturing over six times as much as PC2. For AMMI, the pattern is the same as before, with genotype, environment, and interaction effects all captured separately.

Generalizing beyond these two examples, datasets can differ in the relative magnitudes of genotype (G), environment (E), and interaction (GE) effects, in whether or not interaction effects can substantially share one or several PCs with one or both main effects when SVD is applied to interaction and main effects combined, and in whether the interaction contains one or several patterns. The number of GE patterns is best determined by the number of statistically significant PCs of AMMI because these PCs address GE exclusively. For GGE analyses and graphs, these dataset differences generate nine cases regarding which effects are captured by PC1 and PC2, as shown in Table 3.

The first three cases concern especially simple GE with a single pattern. In Cases 1 and 2, GE and G cannot substantially share a component. If G has a larger SS than GE, then G appears in PC1 and GE appears in PC2, which is analogous to an AMMI1 graph with G and E in its abscissa and GE in its ordinate. But if GE is larger, then GE appears in PC1 and G in PC2, which is analogous to an AMMI1 graph with its axes reversed. On the other hand, in Case 3, GE and G can substantially share PC1, which leaves PC2 with mere noise. The extent to which GE and G can share a component in GGE analysis depends on the extent of correlation between the AMMI parameters α_g and γ_g, which is circumstantial, varying from one dataset to another. For example, the simple artificial dataset in Gauch (1992:55) exemplifies Case 3 because there happens to be a high correlation (0.845), which allows considerable sharing.

The next three cases concern complex GE, with at least two interaction patterns, GE(1) and GE(2), requiring separate PCs that cannot substantially share a component with G. In Case 4, G is considerably larger than GE(1), so the results are analogous to AMMI1. In Case 5, GE(1) is considerably larger than G, which is analogous to AMMI1 with reversed axes. And in Case 6, both GE(1) and GE(2) are considerably larger than G,

Table 3. The nine cases for the contents of the first two principal components in a GGE2 biplot. These are poles in a continuous gradation of possibilities. Some cases are analogous to AMMI results.

Case	PC1	PC2	Analogy
1	G	GE	AMMI1
2	GE	G	AMMI1 Reversed
3	GE+G	Noise	
4	G	GE(1)	AMMI1
5	GE(1)	G	AMMI1 Reversed
6	GE(1)	GE(2)	AMMI2
7	GE(1)+G	GE(2)	
8	GE(1)	GE(2)+G	
9	GE(1)+G(1)	GE(2)+G(2)	

leaving G mostly deferred to a third and perhaps higher components, which is analogous to an AMMI2 graph with its PC1 and PC2 axes capturing portions of GE. When Case 6 obtains, "GGE analysis" becomes a misnomer for a GGE2 analysis or biplot because GGE2 fails to combine G and GE information, given that G is nearly absent from the first two PCs, so this is actually "GE analysis" instead. The soybean example in Table 1 exemplifies Case 5.

The final three cases concern complex GE that can substantially share a component with G. In Case 7, GE(1) shares its component with G. In Case 8, GE(2) shares its component with G. And in Case 9, distraction caused by capturing interaction splits G into at least two sizable portions, G(1) and G(2), recovered by the two components. The Kentucky bluegrass example in Table 2 exemplifies Case 7. Incidentally, the perennial ryegrass (*Lolium perenne* L.) trial in Ebdon and Gauch (2002a, 2002b) and the winter wheat (*Triticum aestivum* L.) trial in Yan and Kang (2002:59) also exemplify Case 7.

Of course, these nine cases of data representation by GGE are poles in a continuous gradation of possibilities because the relative magnitudes of several sources of variation, the extent of main and interaction effects "substantially" sharing a component, and the extent of an interaction's complexity are all matters that admit of degrees. For instance, given a simple interaction, if the correlation between α_g and γ_g is intermediate, rather than weak as in Cases 1 and 2 or strong as in Case 3, then one of the many new possibilities is splitting G into two sizable portions, resulting in PC1 capturing GE+G(1) and PC2 capturing G(2). And if GE and G are of similar magnitudes, this intermediate correlation will split both GE and G into two portions, making what is actually a simple interaction appear like Case 9, the most complicated case. When SVD is applied to G and GE together in GGE analysis, distraction from G can split a simple GE into two sizable portions and distraction from a simple or complex GE can split G into two or more sizable portions, thus making things seem more complicated than they really are.

These nine cases, plus their intermediates, need to be borne in mind when interpreting GGE analyses and graphs. It is crucial to know which possibility obtains for a given dataset.

By sharp contrast, AMMI has but one case. The relative magnitudes of G, E, and GE do not matter. Whether α_g and γ_g are correlated or not does not matter. Whether the interaction is simple or complex does not matter. Nor does any other dataset property matter. For each and every dataset, AMMI captures G, E, and GE effects separately. Having but one case dramatically simplifies interpretations of AMMI analyses and graphs.

Research Purposes

When choosing a statistical analysis for agricultural data, the first and foremost considerations are agricultural, not statistical. What are the agricultural questions and objectives? Research purposes are the master, whereas statistical methods are the servant.

Accordingly, the general research purposes mentioned above for SVD-class models—visualizing data and gaining accuracy—must be explored in greater detail. The key is to understand that research purposes may render various portions of the overall variation in the data either relevant or else irrelevant.

There is a venerable division of labor among agricultural researchers. For lack of better names, the two main groups may be called "crop scientists" and "soil scientists." Again, AOV partitions the variation in a yield trial into G, E, and GE. Crop scientists improve genotypes, so their share of the action is G and also GE because these two portions involve genotypes. Analogously, soil scientists improve environments, so their share is E and GE. Of course, many persons and projects span crop and soil sciences. Nevertheless, this recognition of two parallel enterprises is useful.

For crop scientists, including plant breeders, G concerns broad adaptations of benefit throughout a growing region, whereas GE concerns narrow adaptations that can be exploited only by subdivision into two or more megaenvironments if there are crossover interactions with both favorable and unfavorable effects in different environments. Also, GE decreases heritability given a single target region, but that decrease largely disappears given appropriate megaenvironments. Because GE is often larger than G, understanding interactions and implementing megaenvironments can be strategic, accessing several times as much genetic variability for yield and other important traits. Importantly, because G and GE present agricultural researchers with quite different challenges and opportunities, it is desirable for statistical analysis to address both, but separately. Likewise, for soil scientists, it is desirable for statistical analysis to capture, and yet also distinguish, E and GE. Consequently, to serve the needs of all agricultural researchers, the best statistical analyses must distinguish G, E, and GE.

Proper focus in statistical analysis is like using the zoom in a camera to make the object of interest prominent. For instance, E is irrelevant for many objectives in crop research, yet E often comprises 80% or more of the variation, so crop scientists often remove E before applying SVD. Otherwise, the resulting graph will be analogous to a picture of a dog in which the dog is but a tiny object surrounded by uninteresting and distracting things. When high-dimensional data are compressed to two-dimensional graphs, those merely two dimensions are scarce and precious resources that must not be wasted on irrelevant information.

One further partition is important. Because yield-trial data are noisy, as shown by replicates failing to be equal, the data are a mixture of real signal and mere noise. The noise is idiosyncratic and high dimensional. Most of this noise goes into GE because it has most of the data's degrees of freedom (df). By contrast, the signal is ordinarily dominated by relatively few major causal factors, which are usually the imposed and hence known treatments. There are many genotypes and many environments, but the few causal factors write a low-dimensional signal into the high-dimensional data matrix (Gauch, 1992:231–236).

The first few components of a SVD-class model constitute a low-dimensional structure, which perforce can effectively capture low-dimensional signal, but not high-dimensional noise. Therefore, the early components selectively recover signal, whereas the late components selectively recover noise (Gauch, 1988, 1992:111–170, 1993, 2002:269–326, 2006).

Consequently, it is useful to distinguish the prime, signal-rich part of the interaction, denoted here by GE′, from the noise-rich part, denoted by GE*. Because GE′ is real and repeatable, it should be recovered by statistical analysis. But because GE* is spurious and non-repeatable, it should be discarded. The imperfect and yet substantial and beneficial separation of signal and noise, resulting from partitioning GE into GE′ and GE*, requires a careful choice of N in Eq. [1] to [3], the number of components to retain in the model before relegating the higher components to a discarded residual.

Collecting these observations together, the research objectives of agricultural scientists call for statistical analyses to distinguish four portions of the overall variation in a yield trial: G, E, GE′, and GE*. Ideally, crop scientists focus on G and GE′, whereas soil scientists focus on E and GE′. Success in statistical analysis is essentially a matter of selectively focusing on relevant variation and capturing as much of this as is possible, given the inherent limitations of viable display options such as two-dimensional graphs.

An agricultural researcher considering SVD-based statistical analysis of yield-trial data faces two inevitable choices. First, a researcher must choose a model from among GGE and AMMI (and others) for the purpose of visualizing data. Second, a researcher must choose a particular member from the selected model family with a specific number of components for the purpose of gaining accuracy. The next two sections address these two choices.

Choice 1: A Model for Visualizing Data

Graphs are powerful display options for visualizing patterns in large and complex datasets. Biplot graphs have a point or marker for each genotype and each environment (Gabriel, 1971, 1978; Bradu and Gabriel, 1974, 1978). Three kinds of biplot are common in the yield-trial literature. A GGE2 biplot has PC1 for its abscissa and PC2 for its ordinate, as in Fig. 5.2 in Yan and Kang (2002). An AMMI1 biplot has main effects for its abscissa and PC1 for its ordinate, as in Fig. 1 in Zobel et al. (1988) and Fig. 3.8 in Gauch (1992). And an AMMI2 biplot has PC1 for its abscissa and PC2 for its ordinate, as in Fig. 4 in Kempton (1984) and Fig. 3.9 in Gauch (1992). Recall from Eq. [2] that GGE first removes E and then applies SVD to the environment-centered matrix containing G and GE, intentionally mixed together. "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*" (Yan, 2002). By contrast, from Eq. [3], AMMI first partitions G, E, and GE and then applies SVD to the doubly-centered matrix, deliberately containing GE only. Nevertheless, because G and GE are inherently and fundamentally different in their agricultural implications, the GGE literature exhibits energetic and repeated attempts to "repartition" or reseparate G and GE after initially having left them mixed together (Yan et al., 2000; Yan, 2002).

The first proposal for reseparating G and GE takes PC1 in a GGE2 biplot as a surrogate for G, and likewise PC2 as a surrogate for GE, provided that a given dataset has a high correlation between PC1 and G. But this condition raises an acknowledged problem for those datasets not exhibiting that correlation. "However, the requirement for a near-perfect correlation between genotype PC1 scores and genotype main effects is not always met, which restricts the utility of the…GGE biplot…. Consequently, the yielding ability and stability of the genotypes, and the discriminating ability and the representativeness of the test environments cannot be readily visualized," and the recommended criterion for a "near-perfect correlation" is a correlation of "say, $r > 0.95$" (Yan et al., 2001; likewise Yan et al., 2000; Yan and Hunt, 2001; Yan, 2002; Yan and Rajcan, 2002). They further explained that experience with several datasets has shown that "Near-perfect correlation occurs when G is 40% or more of GGE…, and poor correlation occurs when G is 20% or less of GGE" (Yan et al., 2001). So, $r > 0.95$ or G > 40% are the recommended criteria.

But these criteria for deciding when PC1 recovers G and PC2 recovers GE are problematic. Although these criteria may apply reasonably well to the several rather similar datasets used in Yan et al. (2001), which are just different years in the Ontario winter wheat trials, they lack generality and reliability. Because G and GE may or may not substantially share a component, as shown in Table 3, the correlation between G and PC1 is not a reliable indicator of whether PC1 contains G but not GE because it may recover substantial amounts of both. And because GE may be simple or complex, G > 40% is compatible with PC1 being dominated by either GE or G (or else being a substantial mixture of both). Incidentally, these criteria fail to recognize the opposite favorable case for reseparating GE and G, namely Case 2 in Table 3, which has GE dominating PC1 and G dominating PC2 (and G < 20% would suggest this favorable outcome if the interaction has a simple pattern largely captured by a single component).

But again, this first proposal is acknowledged to be dataset dependent, only sometimes reseparating G and GE. This limitation has motivated the search for a better way to reseparate G and GE that would be reliable and robust for any dataset, not just those datasets that happen to meet some special conditions.

A second proposal for reseparating G and GE involves placing special additional axes in a GGE2 biplot. Yan (2001) claimed that the average tester (genotype) coordinate (ATC) has an abscissa that approximates G and an ordinate that approximates GE. "The ATC x-axis passes through the biplot origin and the marker of the average environment, which is defined by the average PC1 and PC2 scores over all environments. … The

average yield of the cultivars is approximated by the projections of their markers to the ATC x-axis" and "The stability of the cultivars is measured by their projection to the ATC y-axis. The greater the absolute length of the projection of a cultivar, the less stable it is"; that is, the more GE it has. For diallel data, Yan and Hunt (2002) interpret the ATC as information on general combining effects and its perpendicular as information on specific combining effects. Similarly, for the average environment coordinate (AEC), Yan (2002) claimed that "As a rule, the genotype projections onto the AEC abscissa are good approximations of the genotype main effects," and again increasing distance in the perpendicular direction indicates less stability, that is, more GE.

This phrase, "As a rule," expresses dataset dependency. To check the adequacy of an AEC approximation of G, Yan (2002) interpreted a correlation of 0.982 between AEC and G as confirmation of a good approximation for a particular dataset, although more might have been said about some general rule or test for adequacy. However, these special axes cannot be a dataset independent and generally satisfactory method for reseparating G and GE in a GGE2 biplot even if for no additional reason than that G can essentially disappear from the first two components, as in Case 6 in Table 3. Regardless whether extensive experience would show only occasional or rather frequent problems with ATC or AEC being adequate, it could be helpful to remind researchers that by switching from GGE2 to AMMI1 graphs, the AMMI1 axes completely separate G and GE, as well as completely recover G, always and automatically for any dataset.

The "ideal environment," which may be actual or hypothetical, is visualized in Fig. 8A in Yan (2001) using an ATC axis, and similarly the "ideal cultivar" in Fig. 2 in Yan (2002) using an AEC axis. The "ideal cultivar" is defined by having "its projection on the ATC… equal to the longest vector of all cultivars" and having its projection in the perpendicular direction be "obviously zero, meaning that it is absolutely stable" (Yan, 2001). That is, it combines a high genotype main effect with a zero interaction. But this definition of an "ideal cultivar" regards interaction only as a problem, not as an opportunity when given two or more megaenvironments. GE interactions create opportunities for different combinations of G and GE effects to provide diverse winners in different megaenvironments, so there are ideal cultivars in the plural, rather than only one ideal cultivar in the singular.

Consequently, the AMMI counterpart, Fig. 7 in Gauch and Zobel (1997), is far superior, showing various potential winners in one or more megaenvironments, plus various universal winners. These more diverse targets give plant breeders more options and hence more likelihood of developing new winners. Furthermore, this AMMI display is robust, having no dataset dependencies. By contrast, the GGE display of a single "ideal cultivar" strangely penalizes rather than rewards further increases in this genotype's main effect, disregards opportunities for increasing yield by exploiting GE interactions, and depends on special dataset properties for the validity of its ATC or AEC axis. Incidentally, thanks to Lei Wang's

close reading of Gauch and Zobel (1997), an error in Fig. 7 has been corrected in his Chinese translation of this paper, which was published in 2001 as an addendum in his Chinese edition of Gauch (1992). The thick line identifying potential winners should follow the lines through the four winners (3165, 3147, 8172, and 1827) on the right side, rather than the left.

Figure 8B in Yan (2001) identifies the "ideal environment," which is "most discriminating." But this figure is quite problematic because, unlike the "ideal cultivar" in Fig. 8A that at least has access to the relevant G and GE information, this "ideal environment" figure lacks access to the relevant E information. For determining winning genotypes, G and GE are relevant, although E is irrelevant, so GGE analysis is appropriate; but for determining desirable locations, E and GE are relevant, although G is irrelevant, so one might instead want "EGE" analysis applying SVD to genotype-centered data. Furthermore, being "absolutely representative of the average environment" means ignoring the problems and opportunities emerging from GE interactions. Hence, what is intended or actually achieved here by a "most discriminating" environment is not obvious.

A third proposal for reseparating G and GE suggests switching from the usual GGE model to a sites (environments) regression (SREG) modification to reseparate G somewhat better, even though this switch involves some other compromises (Yan et al., 2000, 2001). But these regressions usually capture a small amount of GE compared to SVD, so this suggestion cannot be generally or even frequently satisfactory.

A fourth proposal involves redefining GE interaction in new terms that are readily addressed by the parameters that GGE analysis provides (Yan et al., 2000; Yan and Hunt, 2001). Furthermore, this redefinition intentionally and aggressively undermines the statistical merit and agricultural meaningfulness of its predecessor, the standard definition of GE interaction, which features prominently in AMMI concepts. Of course, a redefinition of GE is a wild card that could potentially start the game all over again, rendering wholly irrelevant the preceding critiques of the first three proposals for reseparating G and GE in the GGE2 biplots.

But on two counts, this redefinition fails. First, the initial step in GGE calculations is to remove the E main effects in the perfectly ordinary manner. Hence, the only coherent way to construe the symmetric G main effects is likewise in the perfectly ordinary manner. Indeed, precisely this understanding of G pervades the GGE literature, including frequent references to the correlation between G and PC1 or between G and ATC. But having already construed both E and G in the ordinary manner, what remains is the GE interactions, as ordinarily defined and understood. So, the very first step of removing E has already closed the door on any redefinition of GE! Second, after presenting a redefinition of GE, the subsequent discussion is said to switch to this new definition thereafter. But unmistakably ordinary GE, defined as the residual after subtracting G and E effects from the data matrix, continues to crop up incessantly. The redefinition just cannot be sustained.

Also, the discussion of this redefinition is disturbingly vague and confused.

In review, these four proposals to reseparate G and GE in a GGE analysis or biplot all fail. There just cannot be any generally effective, dataset independent reseparation by SVD after an initial mixing of G and GE. This is clear from Tables 1 to 3. Nevertheless, there is another method that does achieve this desired separation of G and GE, perfectly and automatically for every dataset. AMMI uses AOV to separate G, E, and GE before further partitioning of GE by SVD. In conclusion thus far in this section, AMMI biplots are superior to GGE biplots for visualizing patterns in yield-trial data in a manner that distinguishes between G, E, and GE effects, which is essential because these different effects have different implications for agricultural research objectives.

However, a special and major research objective remains to be examined, namely visualizing which genotypes win in which environments. Grouping those environments with the same (or nearly the same) winner delineates megaenvironments. This research objective is potentially an exception to the preceding conclusion because winners are determined by G and GE jointly. Accordingly, the possibility must be explored that the reseparation of G and GE, which proves so problematic for GGE analysis, is not necessary for this particular research objective. Indeed, the GGE literature has claimed special advantages for visualizing winners and delineating megaenvironments.

The standard verdict on GE interactions has been that they are a problem, reducing heritability and thereby retarding yield gains. Certainly, in a world with no GE interactions, for every crop, one genotype would always rank first worldwide (although there would still be yield differences due to E), and likewise rankings from one test would be applicable always and everywhere. That would simplify plant breeding!

However, this verdict assumes no subdivision into megaenvironments. Three considerations are important.

1. Megaenvironments can restore high heritability because little GE remains within each megaenvironment.
2. This advantage of restored heritability needs to be weighed against the disadvantage of reduced accuracy of genotype means based on less data within smaller regions (Atlin et al., 2000; also see Comstock and Moll, 1963; Basford and Cooper, 1998; Atlin et al., 2001). This tradeoff implies that even when megaenvironments are advantageous, their number will tend to be small. Indeed, merely two or three megaenvironments can suffice for even huge international yield trials with quite diverse genotypes and environments (Crossa et al., 1990, 1991).
3. Predictable GE′ associated with locations increases the number and value of megaenvironments, whereas unpredictable GE′ associated with years decreases this number (Gauch, 1992, p. 217–230; Annicchiarico, 1997a, 1997b; Annicchiarico et al., 2000; Atlin et al., 2000). Given these three considerations, the merit of megaenvironment sub-

division varies from case to case. But sometimes megaenvironments can reverse the conventional verdict on interactions, instead maintaining heritability and accelerating yield gains.

What is at stake in the pursuit of a dependable megaenvironment analysis is the possibility of changing the yield variability due to GE′, which often exceeds G, from a problem into an opportunity. The incremental cost for megaenvironment subdivision differs from case to case, as well as the incremental increase in yield gains. Megaenvironment analysis can help breeders to assess or anticipate these costs and benefits.

The GGE literature claims superiority to AMMI for displaying which-won-where and megaenvironments. Yan et al. (2001) claimed that ability to "graphically show the which-won-where patterns of the data" is a "unique merit of a GGE biplot." Likewise, Ma et al. (2004) reiterated this claim that "The which-won-where pattern can be visualized only by the polygon view of a GGE biplot" and "an AMMI1 biplot…does not explicitly reveal the which-won-where pattern." If true, that would be a strong selling point for GGE.

However, this claim is perplexing because the precedent for which-won-where terminology and graphs is in the AMMI literature (Gauch, 1992, p. 213–230; Gauch and Zobel, 1997), as Yan et al. (2000) had already acknowledged in the GGE literature. The legitimacy of this claim that "an AMMI1 biplot…does not explicitly reveal the which-won-where pattern" depends on the exact meaning of "biplot" and "explicitly." If "biplot" is construed in a strict sense as a graph showing two kinds of points (as originally intended by Gabriel, 1971) and nothing else, then this statement is true. But by this definition, neither can GGE2 show which-won-where explicitly because a polygon with its bisectors must be added. So, the meaning of this claim is quite unclear.

Anyway, for the moment, it may be helpful to distinguish three terms: a "biplot" contains two kinds of points, an "augmented biplot" contains two kinds of points with added lines or items that explicitly convey additional information, and a "graph" is the general term for any two-dimensional visualization of data patterns. That clarified, AMMI1 and GGE2 can both show which-won-where, and in both cases by the selfsame means, an augmented biplot. Figure 1 in Gauch and Zobel (1997) shows an AMMI1 biplot and Fig. 3 shows an AMMI1 megaenvironment graph for the same data. Figure 3 copied the points for the environments from Fig. 1, although it omitted the points for the genotypes to avoid clutter. But of course, the points of both kinds can be carried over from the biplot to produce an augmented biplot, which adds these three horizontal lines to identify which-won-where and to delineate these four megaenvironments.

Indeed, both GGE2 augmented biplots and AMMI1 augmented biplots or graphs can visualize which-won-where patterns in two dimensions. Because G and GE may or may not substantially share a component, as shown in Tables 1 to 3, the variation captured by GGE2 may be anywhere between that of AMMI1 and AMMI2.

For the wheat trial in Table 4.4 and Fig. 5.13 (and the front cover) of Yan and Kang (2002), GGE2 and AMMI1 delineate identical megaenvironments. For the soybean trial in Table 3.2 of Gauch (1992), GGE2 and AMMI1 identify the same winner in five environments and slightly different winners in the other five. For the Kentucky bluegrass data of Ebdon and Gauch (2002a, 2002b), 41 environments have the same winner and 28 have slightly different winners; and for perennial ryegrass, 24 are the same and 36 are slightly different. Hence, limited experience suggests that GGE2 and AMMI1 (or perhaps sometimes AMMI2) will often define similar megaenvironments. Whenever GGE2 and AMMI1 winners differ substantially even though the amounts of variation captured by these models are nearly equal, necessarily the explanation will be that they capture different proportions of main and interaction effects (namely, GGE2 capturing more GE but AMMI1 capturing more G). Further research would be welcome, especially to check whether exceptions are likely for any specific situations.

However, despite an expected functional similarity for which-won-where and megaenvironments, AMMI is decidedly preferable for its conceptual clarity, simpler geometry, and easier interpretation. These further considerations are also important.

Each piece of the GE interaction that a given PC captures is inherently a contrast, as its positive and negative scores directly indicate (except that a GGE component that happens to contain much G but little GE can be unipolar). Genotypes in environments of the same sign have higher yields than the additive model predicts, whereas those of opposite signs have lower yields. Because agricultural researchers conduct sensible and planned experiments, often the interaction contrast has a plausible causal interpretation, such as a wet-to-dry or cold-to-hot environmental gradient. In any case, the intrinsic contrast of a GE interaction component is visualized directly and intuitively by its being the ordinate in an AMMI1 graph, along which successive megaenvironments occur one after another in a simple linear sequence. Also, the overall winner occupies a special and predictable position surrounding zero on PC1, which reinforces the realization that other winners emerge elsewhere precisely because GE interactions have become substantial there. This directness encourages causal hypotheses and interpretations.

By contrast, the megaenvironments in GGE2 take the shape of slices of a pie, giving megaenvironments a circular representation, so the intrinsic polarity of an interaction contrast is lost. The overall winner is undistinguished and may appear in any arbitrary direction from the origin. Also, a GGE2 biplot mixes G and GE information (except for Case 6 in Table 3), which obscures the GE information that is the sole cause of there being any more than just one universal megaenvironment. Horizontal lines that delineate successive megaenvironments in AMMI1 are geometrically simpler than a polygon with its bisectors in GGE.

Another reason why AMMI is superior to GGE for visualizing which-won-where is that there is also a relevant AMMI2 graph, Fig. 6.5 in Gauch (1992). Its axes are the interaction principal components PC1 and PC2 and regions for winning genotypes are polygons. The trial's overall winner occupies a special location, the polygon around and including the origin, where interactions are not yet large enough to declare other winners. (In the peculiar and unlikely case that G is negligible so that all genotype means are virtually tied and hence there is no overall winner, the AMMI2 geometry switches from the usual collection of polygons to the slices of a pie that are like the GGE2 geometry.)

For sizable and realistic datasets with complex interactions, AMMI2 can account for almost twice as much GE as can AMMI1. Remarkably, an AMMI2 graph is only two dimensional, rendering it presentable on ordinary paper, even though it incorporates information from G, PC1, PC2. This is possible because the G information influences winners in a manner that does not require another axis.

When G and GE cannot substantially share a component, a GGE presentation comparable to AMMI2 would need to use the GGE3 model. But each of its three principal components requires its own dimension, so no two-dimensional GGE3 megaenvironment graph could be a competitor to its AMMI2 counterpart.

Having examined GGE in some detail as a statistical method for visualizing data, rather brief remarks may suffice to assess PCA. Whereas GGE is undesirable because its SVD components mix G and GE, GM-centered PCA is even more undesirable because it mixes G, E, and GE. Every component contains some mixture of all three, as shown in Tables 1 and 2. Given various realistic datasets, GM-centered PCA generates many more cases than the nine shown in Table 3 for GGE, which is perplexing. And given the simplest possible interaction, joint regression (Tukey, 1949), the first component of PCA captures everything, G and E and GE, which again is perplexing. Consequently, GM-centered PCA is unsuitable when research objectives necessitate distinguishing G, E, and GE.

Diallel data have been analyzed by GGE (Yan and Hunt, 2002; Narro et al., 2003). The principal distinction in an analysis of a diallel cross is that between general and specific combining ability, that is, additive and interaction effects. AMMI completely and automatically distinguishes general combining ability (for both male and female parents) and specific combining ability (Ortiz et al., 2001; Betrán et al., 2003; Narro et al., 2003; Presterl and Weltzien, 2003). By contrast, GGE discards the male (or else the female) main effect and then combines the other parent's main effect and the interaction effect with the presumption that the latter two effects can be reseparated satisfactorily. However, discarding one of the parent's main effects is peculiar and reseparation may be problematic.

Choice 2: A Member for Gaining Accuracy

Statistical theory and empirical experience combine to provide three findings that guide researchers in selecting an accurate model from a model family. All

SVD-based analyses, including GGE and AMMI, share these findings by way of commonalities, rather than differences.

First, parsimony is one of the general principles of scientific method that pervades science and technology, including yield-trial analysis. It concerns the relationship between predictive accuracy and model complexity for a sequence of more complex (or less parsimonious) members of a model family, such as GGE1, GGE2, GGE3, and so on, until finally the full model equals the actual data, GGEF. As parameters are added, predictive accuracy increases at first, but then declines. MacKay (1992) aptly named this response "Ockham's hill," in honor of the medieval master of parsimony, William of Ockham, whose name also appears in a common synonym for parsimony or simplicity, "Ockham's razor." Gauch (1993) provided a brief introduction to parsimony, and Gauch (1992, p. 111–204; 2002, p. 303–312; 2006) discussed parsimony in the context of yield trials. Gauch (2002, p. 15–19, 269–326, 334–355) summarized the scientific, statistical, and philosophical literatures on parsimony, including a historical survey from Aristotle to Ockham to the present.

Here it must suffice to explain that the data comprise a mixture of real signal and mere noise, with early model parameters selectively recovering signal and late model parameters selectively recovering noise. The best model, identified by model diagnosis and usually fairly parsimonious, strikes the ideal compromise to retain much signal but discard much noise. Other members of the model family with too few parameters are less accurate because they underfit signal, whereas too many parameters will overfit noise, so the result is Ockham's hill. By contrast, if the successive members of a model family are judged by criteria other than predictive success, such as the percentage of variation accounted for, a strikingly different pattern emerges: performance increases automatically and continually until reaching the full model. Another statistic, an F test applied to each PC, is intermediate, usually choosing a higher model than predictive accuracy indicates, and yet stopping before reaching the full model. However, a suitably modified F test can better match tests for predictive accuracy (Cornelius, 1993). Other statistics, such as the Bayesian information criterion (BIC), respect Ockham's hill by combining a reward for model fit and a penalty for model complexity. BIC and the similar Akaike information criterion (AIC) have been used for AMMI analysis of yield trials (Casanoves et al., 2005). There is also a very simple heuristic test that agrees fairly often with more computationally intensive tests (Gauch, 1992, p. 147–149; Gauch and Zobel, 1996a). A respected assessment of predictive accuracy is cross validation. This requires replication, which is usually unproblematic because most yield trials have replication, especially those published in peer-reviewed journals. There are many different schemes for cross validation and model choice (Gauch, 1988; Gauch and Zobel, 1988; Piepho, 1994, 1995, 1997, 1998; Cornelius and Crossa, 1999; Dias and Krzanowski, 2003; Moreno-González et al., 2003a, 2003b, 2004). Regrettably, there

has not yet been a systematic comparison. In a linear model context, AIC and BIC are asymptotically equivalent to certain variants of cross validation (McQuarrie and Tsai, 1998).

Second, best practices require model diagnosis for each individual dataset because datasets vary in the best member of a model family to optimize predictive accuracy. Theory and experience combine to identify the dataset properties causing this variation. In general, the order of the best model increases with the size of the data matrix and the magnitude and complexity of the interactions, but decreases with the noise level. For instance, for a large international yield trial with diverse genotypes and environments (and hence sizable and complex GE interactions) and with modest noise (such as a coefficient of variation of 10% or less), the best model is ordinarily at least AMMI3 (or GGE4). Because prior experience with other datasets does not indicate a single choice, such as GGE2 being most accurate almost always, there is no avoiding model diagnosis. There is far too much variability in the order of the best model for the past to guide the future.

Third and finally, all SVD-based analyses are essentially equivalent in their ability to gain predictive accuracy, provided of course that best practices include model diagnosis for individual datasets. Cornelius and Crossa (1999) and Dias and Krzanowski (2003) compared AMMI, GGE, and other SVD-based methods for accuracy gain. The result, on the basis of several datasets, is small and inconsistent differences. Diverse models with nearly equal numbers of parameters (or df) achieve nearly equivalent accuracy gain.

What is at stake in this choice of the best member of a model family is predictive accuracy, which also affects efficiency and repeatability for every agricultural research purpose, including determining which-won-where and delineating megaenvironments. Therefore, "predictive accuracy and agricultural interpretation should match up" (Gauch, 1992, p. 151). Understand that GGE and AMMI are model families that offer many different which-won-where patterns, depending on the order of the model. For instance, GGE2 offers one which-won-where pattern, but other members including GGE1 and the actual data, GGEF, offer others, so this choice matters. Likewise, megaenvironments change with the order of the model, as in Fig. 5 in Gauch and Zobel (1997) that exhibits Ockham's hill. A good model choice increases the repeatability of megaenvironment conclusions and recommendations (Crossa et al., 1990, 1991). Model diagnosis, preferably by cross validation, is essential for best practices because predictive accuracy promotes success with each and every agricultural research objective.

Given best practices, as already explained, GGE and AMMI are essentially equal for gaining accuracy. A potentially different matter, however, is what happens in actual practices in the current GGE and AMMI literatures, as considered next.

Both purposes, visualizing data and gaining accuracy, have been prominent in the AMMI literature. Perhaps the availability of AMMI software serving both pur-

poses has helped (Gauch and Furnas, 1991). Nevertheless, current practices reflect best practices only occasionally. Frequently, AMMI is used only to present an AMMI1 biplot for visualizing data patterns, but no model diagnosis checks the suitability of that model for drawing agricultural conclusions, no accuracy gain is achieved, and no megaenvironment analysis follows.

The foremost purpose in the GGE literature has been visualizing patterns in yield-trial data, though some attention has also been given to gaining accuracy. Perhaps the availability of GGE software offering numerous kinds of graphs, but not implementing cross validation, has contributed to this outcome (Yan, 2001, 2002). Satisfactory model diagnosis and choice is much rarer in the GGE literature than the AMMI literature. Although nothing prevents a better scenario in the future, the present situation merits critique, which hopefully would help to move current practices toward best practices.

In the GGE literature, the GGE2 model has served as a virtual default, with limited utilization of other choices and inadequate discussion of this choice's implications for reliable agricultural conclusions and recommendations. Model choice for GGE has largely relied on the earlier literature on AMMI model choice, making the simple adjustment that AMMI with N components corresponds to GGE with $N + 1$ components. Incidentally, the reason for this offset of 1 principal component is that AMMI, but not GGE, has G+E parameters for the genotype and environment main effects. Likewise, each principal component, for either method, has G+E parameters. So, AMMI with N components has the same number of parameters as GGE with $N+1$ components. Reckoning by df gives essentially the same numbers and exactly the same model correspondences. But there are at least three different schemes for assigning df to principal components and the literature is unsettled (Gauch, 1992, p. 112–129), so it is better here simply to reckon by numbers of parameters to match corresponding models.

Researchers using GGE have been encouraged to use the heuristic test from the AMMI literature to check whether biplots using PC3 and higher components require consideration (Yan et al., 2000; Yan and Hunt, 2002). Yet in actual practice, GGE2 overwhelmingly dominates that literature, no doubt because of the happenstance that ordinary graphs printed on paper have two dimensions.

The argument given for this virtual default of GGE2 is that experience in the corresponding AMMI literature, particularly Gauch and Zobel (1996a), has shown that either AMMI1 or AMMI2 is usually most predictively accurate, so GGE2 is indicated because its number of parameters and hence general performance are closest (Yan et al., 2000). For instance, Yan and Hunt (2001) claimed that "Only two PC, PC1 and PC2, are retained in the model because such a model tends to be the best model for extracting patterns and rejecting noise from the data."

Furthermore, what is taught by precept is also taught by example. Particularly for the prominent objective of determining which-won-where, GGE2 has been used countless times, though I am unaware of a single use of any other member of the GGE family (other than occasional comparisons of GGE2 winners with the actual data, the GGEF winners, though these papers' agricultural conclusions uniformly feature the GGE2 winners). That uniform use of GGE2 occurs even despite hints, from occasionally reporting F tests or heuristic tests, that GGE2 is not the ideal choice for optimizing accuracy for a given dataset. However, for several reasons, this argument for default GGE2 analysis fails.

First, even if it were true that AMMI1 or AMMI2 is usually most accurate, that would not obviate the need for model diagnosis in the AMMI literature because these two models can have substantially different accuracies for a given dataset. For one dataset, AMMI1 might be unacceptably inaccurate, but for another, AMMI2.

Second, again supposing that either AMMI1 or AMMI2 is most accurate, that would not support a default choice of GGE2. Rather, this suggests success of any member from GGE1 to GGE3.

Third, wide reading of the currently expansive AMMI literature shows that diagnoses outside AMMI1 and AMMI2 are rather frequent. Hard numbers are elusive, but my rough estimates are 20% for AMMI0, 40% for AMMI1 or AMMI2, and 40% for AMMI3 to AMMIF. Clearly, different members of the AMMI (or GGE) family are most predictively accurate for different datasets, so model diagnosis is needed. Nor is the cause of this outcome a mystery, as explained already.

These observations leave one wondering how often GGE2 is not the best member of the GGE family. For example, Yan et al. (2000) comes close to providing the validation information sought here, although for a somewhat modified version of the regular GGE analysis. Its Table 2 provides a particular kind of F tests for five datasets, indicating that one component is best for one dataset (1994), but for the other four datasets, PC2 is significant at the 0.0000 level. Given an extremely significant PC2 and the usual gradual decline in the magnitude of successive principal components, even though tests are not reported for higher components, it is extremely likely that at least PC3 is also significant at the 0.05 level (or even the 0.01 level) in all four of these cases. This suggests that GGE1 would be best once and GGE3+ four times, but the default GGE2 bats zero for five. This paper also used the heuristic test of Gauch and Zobel (1996a), which diagnosed two components for three datasets (including the 1994 dataset) and three components for the other two datasets. This suggests that the similar GGE2 model might be best for three of these five datasets. However, a model diagnosis that quantifies predictive accuracy, such as cross validation, could provide more satisfactory conclusions. Nevertheless, these other tests provide a strong hint that definitive diagnosis would find GGE2 the most predictively accurate member of its model family for only some of these datasets. Even after model diagnosis has called the two-component model into question, it alone serves as the basis for all graphs, tables, and agricultural conclusions. That is current practices not in alignment with best practices.

This is not to say that GGE2 is automatically suboptimal, but rather it is to say that GGE2 is not automatically optimal. Consequently, readers of the GGE literature who are aware of the best-practices requirement for individual model diagnoses for any SVD-based yield-trial analysis may rightly feel unsettled about agricultural conclusions based on GGE2 defaults.

Fourth, in the sad absence of model diagnosis, the default model is not GGE2, nor GGE1, nor GGE3, but rather, the actual data, GGEF. Unless a specific reduced model has been proven to be more accurate for the dataset at hand, the actual data should be used to determine which-won-where and to delineate megaenvironments. This proper default has the great practical advantage of requiring no statistical analysis or expertise or software whatsoever. For each environment, merely look at the list of yield measurements and pick the largest number! Then group environments with the same (or similar) winner into the same megaenvironment.

Fifth and finally, most of the GGE literature exhibits a missed opportunity to gain accuracy. There is frequent awareness, gained from the AMMI literature, that reduced models can gain accuracy, together with the added assumption that GGE2 ordinarily achieves this benefit. But there is inadequate attention to the necessity of model diagnosis for each dataset to actually achieve this potential for gaining accuracy.

Several summary statistics are useful for quantifying the practical differences between the various members of a model family, such as GGE2 and its actual data, GGEF. One is simply the fraction of environments for which GGE2 and GGEF pick different winners. Another is the average difference in yield estimates, expressed as a percentage of the grand mean, which can be calculated both ways by assuming that either the reduced or else the full model is more to be trusted. For realistic and sizable datasets (as contrasted with tiny datasets often used in the literature to illustrate calculations), the typical outcome in the AMMI literature has been that AMMI1 (or AMMI2) pick different winners than AMMIF for a majority of the environments. Furthermore, selections and recommendations based on the less accurate of these two models typically cause an average yield loss of several percent of the grand mean, often about 5% (Gauch, 1992, p. 171–204; Gauch and Zobel, 1996b). Similar outcomes may be expected for the GGE family.

Two considerations can place this value of 5% in perspective. First, if a representative farmer has gross income which goes 90% to expenses and 10% to profit, then a 5% yield loss translates to a 50% net income loss. Even if the profit margin is the more favorable 20%, there is still a 25% loss of income, which few farmers could easily afford. Second, for most of the world's major crops, plant breeders currently achieve annual yield gains averaging around 1%. Hence, losing 5% of a crop's yield from suboptimal cultivar choices is like turning the clock back 5 yr on plant breeding. Using GGE2 to determine which-won-where is unfortunate when model diagnosis would show that GGE1 or GGEF or whatever determines a different, more accurate, and more repeatable which-won-where pattern.

Incidentally, the differences in a model family's successive which-won-where patterns can also be displayed graphically. Figure 6 in Gauch and Zobel (1997) shows changes in yield estimates for a given environment for the AMMI1 and AMMIF models, with quite different which-won-where patterns. But this graph could easily be extended to show a model sequence, say AMMI0 to AMMI5 plus AMMIF, as well as having several panels for several environments. If necessary to avoid clutter, such a graph could focus on those genotypes that win at least once, while ignoring complete losers. The legend for such a graph should indicate which member of this model family is most predictively accurate, as well as specify this best model's statistical efficiency.

An analogy between validation and replication is instructive. No one would discount the importance of replication, at least for research published in peer-reviewed journals. Yet model validation often has as much or even more potential for increasing accuracy and repeatability.

Failure to perform model diagnosis to increase predictive accuracy and research efficiency constitutes a serious opportunity cost, often more serious than failure to replicate. Best practices for SVD-based analyses of yield trials must include diagnosing the most predictively accurate member of the model family, be that GGE or AMMI or whatever. The only exception to this rule may be the use of biplots or graphs merely for dimensionality reduction to see which genotypes (or environments) are similar or dissimilar and to detect clusters or trends, quite apart from any more specific objectives, such as determining which-won-where, that depend critically on model choice.

From extensive experience, AMMI routinely achieves a statistical efficiency of 1.5 to 4 (Gauch, 1992, p. 134–153; Gauch and Zobel, 1996a). Ebdon and Gauch (2002b) reported an exceptional statistical efficiency of 5.6 for the AMMI2 model for a large perennial ryegrass quality trial. Achieving this accuracy gain by a parsimonious AMMI model required less than a minute of computer time on an ordinary personal computer. Ebdon (2002) estimated that the same accuracy gain through brute-force collection of 5.6 times as much data would have cost over $1 000 000. Furthermore, winners and megaenvironments determined by AMMI are more reliable and repeatable over years than those determined by the original data.

Accuracy matters. Therefore, both replication and validation matter.

Eight Summary Assertions

This paper's comparison of the GGE and AMMI literatures is complicated and technical, which could easily leave some readers bewildered about where their own convictions lie. This complexity could also easily beget a fruitless exchange wherein the issues are never joined. For both of these reasons, it is imperative to discern that despite many details, the basic issues are mercifully few. Indeed, by judging the following eight assertions to be either true or else false, any reader can quickly locate his

or her own position and can readily compare anyone else's position.

The first four assertions concern the first choice between GGE and AMMI for visualizing patterns in yield-trial data.

1. For a two-way factorial design of genotypes × environments, analysis of variance partitions the variation into three sources: the genotype main effects, environment main effects, and genotype × environment interactions; that is, G, E, and GE. (T, F).
2. These three sources of variation present different problems and opportunities to agricultural researchers: G pertains to broad adaptations that require no subdivision by plant breeders of a growing region into several megaenvironments, E is irrelevant for genotype rankings and selections although it is relevant for soil scientists, GE pertains to narrow adaptations, and G and GE jointly determine megaenvironments. (T, F).
3. AMMI produces biplots with a single set of interpretive principles rather than numerous different cases and AMMI is uniquely the statistical method among SVD-based analyses that treats G, E, and GE separately. (T, F).
4. AMMI is better suited than GGE for visualizing patterns in yield-trial data. (T, F).

The next four assertions concern the second choice of a member within the GGE or AMMI model family for gaining accuracy.

5. For different datasets, different members of a GGE or AMMI model family can be most predictively accurate, such as GGE2 for one dataset but GGE4 for another. (T, F).
6. Accuracy affects numerous agricultural research purposes, including selecting superior genotypes (which-won-where) and delineating megaenvironments. (T, F).
7. Model diagnosis is required for each individual dataset. (T, F).
8. Parsimonious GGE and AMMI models, selected with model diagnosis, are essentially equivalent for gaining accuracy. (T, F).

Transparently, my answers, which side with the AMMI literature, are that "All eight assertions are true." But regardless of one's views, the key feature of these assertions remains: they provide for a concise, one-sentence characterization of one's basic views, thereby promoting clear communication before delving into countless details. For instance, another view, which I would identify as the general posture of the GGE literature, could be that "Assertions (4) and (7) are false, but the remainder are true." One sentence does the job.

I would respectfully request that all responses to this paper include one sentence that judges each of these eight assertions to be either true or else false, preferably before communicating a mass of details and equations. I believe that honoring this request would best serve the interests of this journal's readers.

CONCLUSIONS

The research purposes of agricultural scientists conducting yield trials require distinguishing G, E, and GE because they present different challenges and opportunities. Accordingly, this raises the statistical question: How can an agricultural researcher separate G, E, and GE and then produce a low-dimensional visualization of complex GE with least-squares errors? This is a well-posed question that has a unique, known, definitive answer: AMMI.

AMMI separates G, E, and GE before applying SVD to GE for a least-squares, low-dimensional visualization. This always works for all datasets. By contrast, GGE separates E before SVD and then attempts to reseparate G and GE after applying SVD to both. But numerous and valiant attempts to reseparate are all futile because inevitably they depend on special dataset properties not found in all datasets.

One research purpose that does not require separating G, E, and GE is gaining accuracy, which instead requires separating signal from noise. For this purpose, all SVD-based analyses, including AMMI and GGE, are essentially equivalent. Some validation or test procedure is required to select the most predictively accurate member of the AMMI or GGE model family for each individual dataset.

For no research purpose, however, is GGE superior to AMMI for analyzing yield-trial data. Rather, AMMI is always superior or equal to GGE. Therefore, to avoid needless and cumbersome multiplication of methodologies, there is no call for a mix-and-match strategy using both AMMI and GGE. Among SVD-based statistical analyses, AMMI is uniquely the analysis that completely and always separates G, E, and GE as required for most agricultural research purposes, and furthermore it also separates signal from noise as well as any other method for the purpose of gaining accuracy. Consequently, any further search for a better SVD-based analysis is doomed in advance to failure. The most suitable SVD-based method for yield-trial analysis was already invented in 1952, half a century ago.

REFERENCES

Annicchiarico, P. 1997a. Additive main effects and multiplicative interaction (AMMI) analysis of genotype-location interaction in variety trials repeated over years. Theor. Appl. Genet. 94:1072–1077.

Annicchiarico, P. 1997b. Joint regression vs AMMI analysis of genotype-environment interactions for cereals in Italy. Euphytica 94:53–62.

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

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

Atlin, G.N., M. Cooper, and Å. Bjørnstad. 2001. A comparison of formal and participatory breeding approaches using selection theory. Euphytica 122:463–475.

Basford, K.E., and M. Cooper. 1998. Genotype × environment interactions and some considerations of their implications for wheat breeding in Australia. Aust. J. Agric. Res. 49:153–174.

Betrán, F.J., D. Beck, M. Bänziger, and G.O. Edmeades. 2003. Genetic analysis of inbred and hybrid grain yield under stress and nonstress environments in tropical maize. Crop Sci. 43:807–817.

Bradu, D., and K.R. Gabriel. 1974. Simultaneous statistical inference on interactions in two-way analysis of variance. J. Am. Statist. Assoc. Theory Meth. 69:428–436.

Bradu, D., and K.R. Gabriel. 1978. The biplot as a diagnostic tool for models of two-way tables. Technometrics 20:47–68.

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

Comstock, R.E., and P.H. Moll. 1963. Genotype-environment interaction. p. 164–196. In W.D. Hanson and H.F. Robinson (ed.) Statistical genetics and plant breeding. National Academy of Sciences, Washington, DC.

Cornelius, P.L. 1993. Statistical tests and retention of terms in the additive main effects and multiplicative interaction model for cultivar trials. Crop Sci. 33:1186–1193.

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.

Crossa, J., P.N. Fox, W.H. Pfeiffer, S. Rajaram, and H.G. Gauch. 1991. AMMI adjustment for statistical analysis of an international wheat yield trial. Theor. Appl. Genet. 81:27–37.

Crossa, J., H.G. Gauch, and R.W. Zobel. 1990. Additive main effects and multiplicative interaction analysis of two international maize cultivar trials. Crop Sci. 30:493–500.

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.

Ebdon, J.S. 2002. Scientists look for ways to improve turf evaluations. Turfgrass Trends 11(8):1–5.

Ebdon, J.S., and H.G. Gauch. 2002a. Additive main effect and multiplicative interaction analysis of national turfgrass performance trials: I. Interpretation of genotype × environment interaction. Crop Sci. 42:489–496.

Ebdon, J.S., and H.G. Gauch. 2002b. Additive main effect and multiplicative interaction analysis of national turfgrass performance trials: II. Cultivar recommendations. Crop Sci. 42:497–506.

Fisher, R.A. 1918. The correlation between relatives on the supposition of Mendelian inheritance. Trans. Roy. Soc. Edinburgh 52:399–433.

Fisher, R.A., and W.A. Mackenzie. 1923. Studies in crop variation. II. The manurial response of different potato varieties. J. Agric. Sci. (Cambridge) 23:311–320.

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

Gabriel, K.R. 1978. Least squares approximation of matrices by additive and multiplicative models. J. Roy. Statist. Soc. Ser. B 40:186–196.

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; Chinese edition 2001, China National Rice Research Institute, Hangzhou, China.

Gauch, H.G. 1993. Prediction, parsimony and noise. Am. Sci. 81:468–478.

Gauch, H.G. 2002. Scientific method in practice. Cambridge Univ. Press, Cambridge, England; Chinese edition 2004, Tsinghua Univ. Press, Beijing, China.

Gauch, H.G. 2006. Winning the accuracy game. Am. Sci. 94:133–141.

Gauch, H.G., and R.E. Furnas. 1991. Statistical analysis of yield trials with MATMODEL. Agron. J. 83:916–920.

Gauch, H.G., and R.W. Zobel. 1988. Predictive and postdictive success of statistical analyses of yield trials. Theor. Appl. Genet. 76:1–10.

Gauch, H.G., and R.W. Zobel. 1996a. AMMI analysis of yield trials. p. 85–122. In M.S. Kang and H.G. Gauch (ed.) Genotype-by-environment interaction. CRC Press, Boca Raton, FL.

Gauch, H.G., and R.W. Zobel. 1996b. Optimal replication in selection experiments. Crop Sci. 36:838–843.

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

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

Ma, B.L., W. Yan, L.M. Dwyer, J. Frégeau-Reid, H.D. Voldeng, Y. Dion, and H. Nass. 2004. Graphic analysis of genotype, environment, nitrogen fertilizer, and their interactions on spring wheat yield. Agron. J. 96:169–180.

MacKay, D.J.C. 1992. Bayesian interpolation. Neural Comput. 4:415–447.

McQuarrie, A.D.R., and C.-L. Tsai. 1998. Regression and time series model selection. World Scientific, Singapore.

Moreno-González, J., J. Crossa, and P.L. Cornelius. 2003a. Additive main effects and multiplicative interaction model: I. Theory on variance components for predicting cell means. Crop Sci. 43:1967–1975.

Moreno-González, J., J. Crossa, and P.L. Cornelius. 2003b. Additive main effects and multiplicative interaction model: II. Theory on shrinkage factors for predicting cell means. Crop Sci. 43:1976–1982.

Moreno-González, J., J. Crossa, and P.L. Cornelius. 2004. Genotype × environment interaction in multi-environment trials using shrinkage factors for AMMI models. Euphytica 137:119–127.

Narro, L., S. Pandey, J. Crossa, C. De León, and F. Salazar. 2003. Using line x tester interaction for the formation of yellow maize synthetics tolerant to acid soils. Crop Sci. 43:1718–1728.

Ortiz, R., S. Madsen, W.W. Wagoire, J. Hill, S. Chandra, and O. Stølen. 2001. Additive main effect and multiplicative interaction model for a diallel-cross analysis. Theor. Appl. Genet. 102:1103–1106.

Pearson, K. 1901. On lines and planes of closest fit to systems of points in space. Philos. Mag. Sixth Series 2:559–572.

Piepho, H.-P. 1994. Best linear unbiased prediction (BLUP) for regional yield trials: A comparison to additive main effects and multiplicative interaction (AMMI) analysis. Theor. Appl. Genet. 89:647–654.

Piepho, H.-P. 1995. Robustness of statistical test for multiplicative terms in additive main effects and multiplicative interaction model for cultivar trial. Theor. Appl. Genet. 90:438–443.

Piepho, H.-P. 1997. Analyzing genotype-environment data by mixed models with multiplicative effects. Biometrics 53:761–766.

Piepho, H.-P. 1998. Empirical best linear unbiased prediction in cultivar trials using factor analytic variance-covariance structures. Theor. Appl. Genet. 97:195–201.

Pike, E.W., and T.R. Silverberg. 1952. Designing mechanical computers. Machine Design 24:131–137, 159–163.

Presterl, T., and E. Weltzien. 2003. Exploiting heterosis in pearl millet for population breeding in arid environments. Crop Sci. 43:767–776.

Tukey, J.W. 1949. One degree of freedom for non-additivity. Biometrics 5:232–242.

Williams, E.J. 1952. The interpretation of interactions in factorial experiments. Biometrika 39:65–81.

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

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

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

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

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

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 analysis of test sites and trait relations of soybean in Ontario. Crop Sci. 42:11–20.

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