

REVIEW & INTERPRETATION

Statistical Analysis of Yield Trials by AMMI and GGE: Further Considerations

Hugh G. Gauch, Jr.,[★] Hans-Peter Piepho, and Paolo Annicchiarico

ABSTRACT

Recent review articles in this journal have compared the relative merits of two prominent statistical models for analyzing yield-trial data: Additive main effects and multiplicative interaction (AMMI) and genotype main effects and genotype \times environment interaction (GGE). This review addresses more than 20 issues that require clarification after controversial statements and contrasting conclusions have appeared in those recent reviews. The AMMI2 mega-environment display incorporates more of the genotype main effect and captures more of the genotype \times environment (GE) interaction than does GGE2, thereby displaying the which-won-where pattern more accurately for complex datasets. When the GE interaction is captured well by one principal component, the AMMI1 display of genotype nominal yields describes winning genotypes and adaptive responses more simply and clearly than the GGE2 biplot. For genotype evaluation within a single mega-environment, a simple scatterplot of mean and stability is more straightforward than the mean vs. stability view of a GGE2 biplot. Diagnosing the most predictively accurate member of a model family is vital for either AMMI or GGE, both for gaining accuracy and delineating mega-environments.

H.G. Gauch, Jr., Crop and Soil Sciences, Cornell Univ., Ithaca, NY 14853; H.-P. Piepho, Bioinformatics Unit, Institut für Pflanzenbau und Grünland, Universität Hohenheim, 70599 Stuttgart, Germany; P. Annicchiarico, CRA-Istituto Sperimentale per le Colture Foraggere, Viale Piacenza 29, 26900 Lodi, Italy. Received 18 Sept. 2007. [★]Corresponding author (hgg1@cornell.edu).

Abbreviations: AEC, average environment coordinate; AMMI, additive main effects and multiplicative interaction; ANOVA, analysis of variance; E, environment; EGE, environment main effects and genotype \times environment interaction; G, genotype; GE, genotype \times environment; GGE, genotype main effects and genotype \times environment interaction; GL, genotype \times location; PC, principal component; PCA, principal components analysis; QTL, quantitative trait locus; RHS, right-hand side; SS, sum of squares; SSG, sum of squares for genotype; SSGE, sum of squares for genotype \times environment; SV, singular value; SVD, singular value decomposition; SVP, singular value partitioning.

THE YIELD TRIAL is one of the most common and important experiments in agricultural research, typically testing a number of genotypes in a number of environments, often for additional traits as well as yield. Accordingly, statistical methods for effective analysis of yield trials have received considerable development and discussion. Two frequently used models for statistical analyses have been the additive main effects and multiplicative interaction (AMMI) model and the genotype main effects and genotype \times environment interaction effects (GGE) model.

Recent review articles in *Crop Science* have compared these analyses. Gauch (2006a) reviewed the AMMI and GGE literatures, favoring AMMI. More recently, Yan et al. (2007) responded to that article, favoring GGE. However, Yan et al. (2007) contains

Published in *Crop Sci.* 48:866–889 (2008).

doi: 10.2135/cropsci2007.09.0513

© 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.

several errors and fails to respond to some critiques of GGE in Gauch (2006a). Consequently, the issues require further consideration.

Gauch (2006a) argues that AMMI's and GGE's relative merits depend on the research purpose. For the research purpose of producing graphical visualizations of patterns in yield-trial data, AMMI is superior because an AMMI1 biplot has a simpler and more informative geometry than a GGE2 biplot, and the AMMI2 biplot, which has no GGE counterpart, can capture substantially more variation. Furthermore, the AMMI1 biplot serves crop and soil scientists simultaneously because it displays genotype (G) main effects, environment (E) main effects, and genotype \times environment (GE) interaction effects, whereas the GGE2 biplot shows only the G and GE effects of interest to crop scientists and soil scientists would need another biplot, which could be called the environment main effects and genotype \times environment interaction effects (EGE) biplot, showing the E and GE effects of interest to them. For the research purpose of delineating mega-environments, both AMMI and GGE are suitable, and comparisons so far indicate similar results, as expected. For the research purpose of gaining accuracy, AMMI and GGE (and the shifted multiplicative model and so on) are all equally capable.

The review by Yan et al. (2007) addresses five main topics. It begins by identifying three major aspects of yield-trial analysis: (i) mega-environment delineation, (ii) genotype evaluation, and (iii) test-environment evaluation. Accordingly, that review's three figures showcase the GGE biplots used to address these objectives, namely, the "which-won-where," "mean vs. stability," and "discriminating power vs. representativeness" views. It then discusses the fundamental differences in theory between the AMMI and GGE perspectives, claiming that (iv) G and GE are best analyzed jointly and they are interchangeable. Finally, that review claims that (v) accuracy gain from model diagnosis should not be overstated.

To facilitate comparison of the Yan et al. (2007) review and the present review, we follow this same order of topics. Because the literature has already been reviewed recently twice, this paper does not again review the literature but rather focuses on Yan et al. (2007).

This paper has four objectives: (i) to correct several errors and to emphasize neglected opportunities in the GGE literature so that researchers who prefer GGE analysis can obtain considerably more reliable and effective results; (ii) to demonstrate that some research objectives accomplished by graphical displays in the GGE literature can be accomplished more simply, directly, and satisfactorily by AMMI-based graphs, scatterplots, or tabular presentations; (iii) to respond to over 20 criticisms of AMMI and counter-claims favoring GGE in Yan et al. (2007); and (iv) to maintain that AMMI is singularly appropriate for crop and soil scientists because AMMI can separate

the main and interaction effects that present researchers with such different problems and opportunities, whereas GGE combines G and GE and thereafter can never separate them as crop scientists require and GGE ignores E of importance to soil scientists.

MODELS AND TERMINOLOGY

Both AMMI and GGE combine analysis of variance (ANOVA) for additive parameters and singular value decomposition (SVD) for multiplicative parameters, the principal components (PCs). In the present context of agricultural research, consider a GE two-way factorial design, with Y_{ge} the yield for genotype g in environment e .

For the additive parameters, let μ be the grand mean, μ_g the mean for genotype g (over environments), and μ_e the mean for environment e (over genotypes). Also let $\alpha_g = \mu_g - \mu$ be the genotype deviation and $\beta_e = \mu_e - \mu$ be the environment deviation.

For the multiplicative parameters, for component n , let λ_n be the singular value for that component (and correspondingly λ_n^2 its eigenvalue), let γ_{gn} be the eigenvector value for genotype g and let δ_{en} be the eigenvector value for environment e , with both eigenvectors scaled as unit vectors, $\sum_g \gamma_{gn}^2 = \sum_e \delta_{en}^2 = 1$. Usually, not all components are retained in the model, so a reduced model also has a residual term, ρ_{ge} .

Let NG be the number of genotypes and NE the number of environments. Then the full model, leaving no residual, has minimum($NG - 1$, $NE - 1$) components for AMMI and min(NG , $NE - 1$) components for GGE. The number of components for a particular member of an AMMI or GGE model family can be indicated by adding that number as a suffix, such as AMMI1 with one component or GGE2 with two. The full models with all components are designated by AMMIF and GGEF. At the opposite extreme, the AMMI model with no components, which has just the ANOVA portion of the model, is designated by AMMI0 (it has no GGE counterpart).

Two additional terms, θ_{ge} and ν_{ge} , will be introduced below. They will serve to focus attention on the values submitted to SVD.

The AMMI model equation may be written as

$$Y_{ge} = \mu + \alpha_g + \beta_e + \sum_n \lambda_n \gamma_{gn} \delta_{en} + \rho_{ge} \quad [1]$$

Another useful form of this model equation results from switching from deviations to means for the additive part, rearranging the terms to put the additive part on the left and the multiplicative part on the right, and defining θ_{ge} to be the interaction

$$Y_{ge} - \mu_g - \mu_e + \mu = \theta_{ge} = \sum_n \lambda_n \gamma_{gn} \delta_{en} + \rho_{ge} \quad [2]$$

Yet another useful form results from switching from actual yields Y to model estimates Y' for the selected member of the model family, such as AMMI2:

$$Y'_{ge} = \mu_g + \mu_e - \mu + \lambda_1 \gamma_{g1} \delta_{e1} + \lambda_2 \gamma_{g2} \delta_{e2} \quad [3]$$

The GGE model may be written as

$$Y_{ge} - \beta_e - \mu = \nu_{ge} = \sum_n \lambda_n \gamma_{gn} \delta_{en} + \rho_{ge} \quad [4]$$

The term defined in this expression, ν_{ge} , does not yet have a name in the GGE literature, so *environment-centered yields* is adopted here. A related quantity, $Y_{ge} - \beta_e$ or the actual yields minus environment deviations, has been termed *nominal yields* in the AMMI literature (Gauch and Zobel, 1997); we propose this name. Obviously, using actual or environment-centered or nominal yields does not affect genotype differences or rankings within each environment. But it does affect SVD.

The GGE2 member of the model family may be written as

$$Y'_{ge} = \mu_e + \lambda_1 \gamma_{g1} \delta_{e1} + \lambda_2 \gamma_{g2} \delta_{e2} \quad [5]$$

The only difference between the AMMI and GGE model equations is that AMMI does, and GGE does not, subtract the genotype effects before performing SVD. Consequently, the multiplicative parameters have different values in Eq. [2] and [4] because SVD is applied to different matrices: interactions for AMMI and environment-centered yields for GGE. Before the name *AMMI* was introduced by Gauch (1988) and Zobel et al. (1988) and the name *GGE* by Yan et al. (2000), these models were called doubly centered principal components analysis (PCA) and environment-centered PCA.

The multiplicative term, $\lambda \gamma_g \delta_e$, requires two multiplications to be evaluated, so for convenience, the singular value λ is usually incorporated into the genotype and/or environment results to obtain “scores” whose products give this term with a single multiplication. Any choice of the arbitrary constant z for $\lambda^z \gamma_g$ and $\lambda^{1-z} \delta_e$ gives the intended product, but certain choices of z are preferable for various purposes.

Three simple choices of z prevail in the literature. The usual partition in the AMMI literature is genotype and environment scores of $\lambda^{0.5} \gamma_g$ and $\lambda^{0.5} \delta_e$. The two common partitions in the GGE literature are genotype-focused singular value partitioning, $\lambda \gamma_g$ and δ_e , and environment-focused partitioning, γ_g and $\lambda \delta_e$. Yan et al. (2007) denote these by singular value partitioning (SVP) = 1 and SVP = 2, respectively. They also mention a symmetric scaling using $\lambda^{0.5} \gamma_g$ and $\lambda^{0.5} \delta_e$.

For any component, for SVP = 1, the sum of squares (SS) for genotype scores equals λ^2 and for environment scores equals 1. For SVP = 2, the SS for genotype scores equals 1 and for environment scores it equals λ^2 . For symmetric scaling, the SS for genotype scores and for environment scores is λ .

Because λ can easily be far larger (or smaller) than 1, a biplot with SVP = 1 or SVP = 2 can have one set of markers reduced to a tiny region, or even a microscopic speck.

Accordingly, biplots are often rescaled to give both sets of markers similar ranges or areas, thereby improving legibility (Digby and Kempton, 1987, pp. 63–67; Yan and Kang, 2002, pp. 42–49). Multiplying genotype scores for both axes by any value s leaves the interpretation of the biplot unaltered, while simultaneously multiplying the environment scores for both axes by $1/s$ also leaves the cross products of genotype and environment scores unaltered (so they approximate θ_{ge} in Eq. [2] for AMMI or ν_{ge} in Eq. [4] for GGE). Several procedures for choosing s have been proposed, but since the distributions of markers are variable, no single default procedure has prevailed. Although the choice of s leaves the interpretation unaltered, to make a biplot reproducible, any axis rescaling must be reported. After both singular value (SV) partitioning and axis rescaling, four sets of scores result: $s \lambda_1^z \gamma_{g1}$, $s \lambda_2^z \gamma_{g2}$, $(\lambda_1^{1-z}/s) \delta_{e1}$, and $(\lambda_2^{1-z}/s) \delta_{e2}$.

Both axes must have the same scale to preserve biplot properties. Otherwise, an unequal scaling “is useless for appreciating distances and angles, including orthogonal projections, which are vital for interpreting biplots” (Gower and Hand, 1996, p. 20).

Because the units of Y are yield (in kg ha⁻¹ or whatever), all terms in the above equations must also be in the units of yield, including $\lambda \gamma_g \delta_e$. But because this term is partitioned into a genotype score and an environment score that need to be discussed and used individually, units must also be assigned to these scores. One choice is to deem the eigenvectors to be without units and attach the units of yield to λ . But for SVP = 1 (or SVP = 2), this choice has the quite awkward implication of giving units of yield to genotype scores but no units to environment scores (or the reverse). Consequently, the preferable choice is to attach units of the square root of yield to both genotype scores and environment scores, regardless of how λ is distributed between them and regardless of any axis rescaling. Because this attribution, giving all scores the same units, best accords with customary practice in both the AMMI and GGE literatures, it is adopted here.

Having recommended attribution of the units of yield^{0.5} to all PC scores—for both AMMI and GGE and for any partitioning of the singular value—products of genotype and environment scores are in the correct units of yield. But for both AMMI and GGE, these products, for either a given component or for a sum of components (including the full model), are in units of yield but are not approximating or equaling actual yields. The multiplicative portion of AMMI concerns θ , that is, interactions. And the multiplicative portion of GGE concerns ν , that is, nominal yields. In the literature, the products $\lambda \gamma_g \delta_e$ are sometimes taken to be approximating “yields,” but such loose terminology is better avoided because these products approximate *interactions* in AMMI and *environment-centered yields* in GGE. Only if SVD were applied directly to the

actual data matrix, without subtraction of anything first, would these products approximate the actual yields.

Because the term *biplot* has been applied by different authors in variously strict or broad senses, we must specify the sense intended here. Biplots provide visualizations for two-way data matrices, namely, GE matrices in the present context of yield-trial research, or row-by-column matrices to use more generic terminology. The foremost feature of a biplot, from which it takes its name, is a graph with markers (points or lines or arrows) for each row and markers for each column of a data matrix. Another important feature of a biplot, emphasized by Gabriel (1971) and Kempton (1984), is the inner-product property. The inner product of the genotype and environment vectors approximates the entries in the matrix submitted to SVD, where the two dimensions ordinarily used to display PC1 and PC2 provide the rank 2 approximation.

Three kinds of biplots are discussed in this paper. The AMMI1 biplot shows genotype and environment means and the grand mean on the abscissa and its PC1 scores for genotypes and environments on the ordinate. The AMMI2 biplot shows its PC1 on the abscissa and PC2 on the ordinate. And the GGE2 biplot shows its PC1 on the abscissa and PC2 on the ordinate.

If markers are shown for rows (or columns) of special interest, rather than all rows to reduce clutter, we would still call such a graph a biplot. Likewise, were row markers shown in one panel of a graph and column markers in another panel, that graph would still be deemed a biplot. And if additional information is added beyond the row and column markers—such as the polygon and rays used to delineate GGE2 mega-environments, or interpretive information on causal factors, or arrows representing environmental covariates (ter Braak, 1986)—our preference is to call such a graph an *augmented biplot*. In a calibrated predictive biplot, the environments are represented by scaled axes rather than vectors so that projections of genotype markers onto environmental axes allow predicted values to be read directly (Gower and Hand, 1996, p. 23–30). If markers for rows and markers for columns are present, but the inner-product property is absent (as in an AMMI1 biplot with means on the abscissa and PC1 on the ordinate), our inclination is to consider this a biplot in the broad sense while acknowledging that some others may prefer to deem this a graph or display. However, if one set of markers is replaced by a set of regions (such as AMMI2 mega-environments with markers for all genotypes replaced by regions for winning genotypes), then *graph* or *display* seems to be a more appropriate name.

Finally, it may be emphasized that the data structure addressed by AMMI, GGE, and related analyses is a two-way factorial design having one kind of data (such as yield), either replicated or not. Some yield trials produce additional data structures, including a three-way factorial

of genotypes \times locations \times years, a data matrix with a different variable in each row (or column), and a matrix of environmental measurements as well as the matrix of yield data (Gauch, 1992, p. 102–106). Those complex data structures require other statistical models or in some cases may conveniently be reduced to a two-way data matrix of prominent interest, such as a genotype \times location (GL) matrix for trials repeated over years.

MEGA-ENVIRONMENT DELINEATION

The centerpiece of the GGE literature is GGE2 biplots augmented with polygons to show mega-environments, as illustrated by Fig. 1 in Yan et al. (2007). That review claimed that GGE biplots are superior to AMMI biplots for showing mega-environments, even in cases for which both methods happen to delineate the same mega-environments. This section lists 10 criticisms of AMMI in the order encountered in Yan et al. (2007), followed by our responses in a somewhat different order.

Criticism 1. The GGE2 biplot is more accurate for showing mega-environments because it always captures more G+GE than the AMMI1 biplot.

Criticism 2. The which-won-where patterns are not always easy to visualize in the AMMI1 biplot.

Criticism 3. Although Fig. 6.5 in Gauch (1992) proposed an AMMI2 biplot for showing mega-environments, its usefulness has not yet been demonstrated.

Criticism 4. The AMMI2 representation of mega-environments is better understood as a conclusion-presentation tool, rather than a pattern-discovery tool, because the winning genotypes are identified from a separate table of AMMI2 predicted values and subsequently are superimposed on the graph.

Criticism 5. The AMMI2 biplot is not capable of displaying which-won-where patterns because G and PC scores are not in the same units.

Criticism 6. The AMMI1 biplot lacks the most important property of useful biplots, the inner-product property, because its two axes are not in the same units.

Criticism 7. The AMMI2 biplot also lacks the inner-product property.

Criticism 8. The hypothesis in Gauch (1992) that universal winners would be located near the origin of the AMMI2 biplot is not dependable, especially since universal losers would also be located there.

Criticism 9. The shapes of mega-environment regions in an AMMI1 biplot are completely subjective because its axes are in different units.

Criticism 10. Unlike the GGE2 biplot augmented to show mega-environments, the AMMI1 counterpart also displays the environments E, which are irrelevant to cultivar and test-environment evaluations.

Response to Criticism 1

The first part of this criticism, that more G+GE implies more accuracy, is sometimes true and sometimes false. Capturing more G+GE than AMMI1 does not automatically translate into more accuracy. Most pointedly, sometimes AMMI0 is the most predictively accurate member of the AMMI model family because the GE interactions are small and are buried in the noise. Also, sometimes GE is only marginally significant, in which case AMMI1 can be more accurate than AMMI0, whereas capturing even more GE than AMMI1 by means of either the GGE2 or AMMI2 model only decreases accuracy. Recall Ockham's hill, that overly simple models underfit real signal whereas overly complex models overfit spurious noise (Gauch, 2006b).

The second part of this criticism is true, that a GGE2 biplot always and automatically captures more G+GE than an AMMI1 biplot. This fact was clear in both review articles, Gauch (2006a) and Yan et al. (2007). However, it is only one of a set of three truths that belong together.

The second truth is that model diagnosis is essential for mega-environment delineations and graphs. For instance, consider the Ontario winter wheat (*Triticum aestivum* L.) data in Table 1 of Yan et al. (2007), which were also given in Table 4.4 in Yan and Kang (2002) with one more digit; we have used the latter version. The AMMI1 model recognizes two winners (identical to those of GGE2), AMMI2 recognizes three, AMMI3 recognizes four, and the higher models from AMMI4 to the full model AMMI8 (which equals the actual data in Table 1) recognize five or six, which implies anywhere from two to six mega-environments, depending on model diagnosis and choice. A similar story applies to the GGE model family for those data. Another example is presented in Fig. 5 of Gauch and Zobel (1997).

Consequently, mega-environment delineation and model accuracy are inexorably intertwined. Most pointedly, if a GGE2 biplot is used uncritically as an automatic default for delineating mega-environments, even though in a given case, model diagnosis would reveal that GE is buried in noise, then a growing region may be subdivided into several mega-environments when no subdivision is needed. That mistake increases effort and cost while it decreases accuracy and benefit. Conversely, if the interaction is too complex to be captured by only two axes, a GGE2 or AMMI2 biplot may identify too few mega-environments.

The third of three interrelated truths, already emphasized in Gauch (2006a), is that an AMMI2 mega-environment display always captures more G+GE than a GGE2 mega-environment display—indeed, it automatically captures both more G (100%) and more GE. So, in the fairly common case that AMMI2 is more accurate than AMMI0 and AMMI1 (or even in the still more common case that AMMI3 or a higher model is most accurate, but practical constraints favor sacrificing a little accuracy to gain the fewer mega-environ-

ments of the simpler AMMI2 model), the AMMI2 graph will be more accurate than the GGE2 graph.

Combining the first and third truths, the amount of G+GE captured always follows the rule that $AMMI1 < GGE2 < AMMI2$. For the Ontario wheat example in Yan et al. (2007), AMMI1 captures 75.8%, GGE2 captures 78.0%, and AMMI2 captures 86.7%. For another example, using values from Table 1 in Gauch (2006a), for the G+GE variability of a soybean [*Glycine max* (L.) Merr.] yield trial, AMMI1 captures 85.1%, GGE2 captures 88.6%, and AMMI2 captures 95.1%. For a third example, for the international bread wheat (*Triticum aestivum* L. em Thell.) trial in Crossa et al. (1991) with enormous GE interactions, these three figures are 35.3%, 44.4%, and 47.5%.

Given that GGE2 is flanked on both sides by AMMI models, the only case for which GGE2 is likely to be most predictively accurate is when Ockham's hill has a nearly flat peak around AMMI1 and AMMI2. Such cases are rather rare, but can occur, as in Fig. 4.6 in Gauch (1992, p. 143). And yet, in precisely such cases, the marginal increase in accuracy for GGE2 over the better of AMMI1 and AMMI2 would be expected to be quite small—often too small to have statistical significance, except for unusually large datasets.

Hence, if capturing as much as possible of the signal in G+GE (and as little as possible of the noise) is to be the criterion for choosing a mega-environment display, then AMMI1 or AMMI2 will almost always surpass GGE2 in accuracy. The outflanked GGE2 can win only rarely and even then only marginally. Accordingly, we reject criticism 1, with its opposite conclusion favoring GGE2, because its faulty argument recognizes only one of three interrelated truths.

Response to Criticisms 2, 6, 9, and 10

Four criticisms question the meaningfulness of AMMI1 displays for showing mega-environments. A particular concern is that the two axes of AMMI1 biplots are in different units, such as yield (in kg ha⁻¹ or whatever) for the main effects on the abscissa and the square root of yield for the interaction effects on the ordinate.

Criticism 2 claims that the which-won-where patterns are not always easy to visualize in the AMMI1 biplot. We could not see the point of this criticism. Figures 1–3 present AMMI1 results.

Figure 1 shows the AMMI1 biplot for the Ontario wheat example. The abscissa shows differences in main effects and the ordinate differences in interaction patterns. This biplot captures 93.4% of the yield variability. AMMI calculations were performed by the open source software MATMODEL Version 3.0 (Gauch, 2007) and graphs were rendered by SAS/GRAPH (SAS Institute, 1999). Provided that the genotype effects are large enough to dominate GGE's PC1, the abscissa of an AMMI1 and GGE2

biplot both capture main effects (exactly for AMMI1 and approximately for GGE2), and likewise, both ordinates capture GE interaction effects (exclusively for AMMI1 and mostly for GGE2). Consequently, the broad arrangements of markers in AMMI1 and GGE2 biplots are often similar, as in the present case. Usually PC1 from AMMI1 correlates highly with PC2 from GGE2.

Figure 2 shows the AMMI1 mega-environment display for the same example, obtained by replacing genotype markers with genotype winning regions. There are two mega-environments, identical to those delineated by GGE2 in this instance. The AMMI1 mega-environment geometry is quite simple, involving a single horizontal line at a PC1 score of 0.18158. For comparison, the corresponding GGE2 display in Fig. 1 of Yan et al. (2007) requires a five-sided polygon and five rays from the origin, for a total of 10 lines. Consequently, the opposite of criticism 2 obtains because the AMMI1 geometry is considerably simpler and hence easier to visualize.

Figure 3 shows nominal yields for the 18 wheat genotypes as a function of environment PC1 scores from AMMI1 analysis. The mega-environment boundaries in Fig. 2 can be understood by the responses in Fig. 3, particularly that the two winners are G8 and G18 and that their response lines cross at a PC1 score of 0.18158. Genotype G18 wins in environments E5 and E7, whereas G8, which has the highest overall mean, wins in the other seven environments. Genotypes with a flat response possess greatest stability and are widely adapted if they also possess high mean yield. The genotype with greatest stability is G15 (as readily identified in Fig. 1 as the entry with a PC1 score nearest to 0), shown in Fig. 3 by the thin line starting at 4.237 Mg ha⁻¹ on the left and rising to 4.367. But G15 is so far from winning anywhere that no sacrifice of high performance for the sake of greater stability is warranted in this case. Figures based on AMMI1 and GGE2 are comparable, as both capture GE interaction essentially on one PC axis. But the AMMI1 graph shown in Fig. 3 is simpler and clearer than the GGE2 biplot for visualizing mega-environments, genotype adaptive responses, and similarities among test environments for GE effects. It is analogous to the venerable linear regressions of Fig. 2 in Finlay and Wilkinson (1963), except that environment PC1 scores substitute for environment means along the abscissa. However, AMMI1 always captures more GE than linear regression, in this case, 48.2% instead of only 11.4%. A variant of Fig. 3 is to show interactions rather than nominal yields, as in Fig. 3.10 of Gauch (1992, p. 95). Yet another variant would be to use GGE instead of AMMI, preferably using the perpendicular to the average environment coordinate (AEC) with SVP = 1 rather than any of GGE's original axes to capture as much GE interaction as is possible in GGE analysis. However, even this best of the GGE options captures less GE than does AMMI's PC1, so the AMMI display of nominal yields as in Fig. 3 is ordinarily preferable.

Simple models such as AMMI1 or GGE2 (in which the GE interaction can be taken into account essentially by one principal component) are frequently adequate when analyzing GL interaction in trials repeated in time (Annicchiarico, 1997). In addition, a preference for simple models and fewer mega-environments may arise from the tradeoff between the advantage of greater gains from specific adaptation and the disadvantage of less data and hence less accuracy within each mega-environment that derive from increasing the number of mega-environments (Atlin et al., 2000; Piepho and Möhring, 2005).

Criticism 6 objects that the AMMI1 biplot lacks the inner-product property. This is true. But this is also irrelevant as regards delineating mega-environments. At any rate, sometimes a purpose can be accomplished by various means. A major application of the inner-product property is to approximate the yield and ranking of each genotype for any chosen environment. But this same information is available from a tabular representation. For each environment, the genotypes can be ranked by their AMMI or GGE estimated yields. One can scan down the list of model expected yields much quicker than one can mentally construct the corresponding information from a biplot. Furthermore, the tabular presentation is equally suited to any member of the AMMI or GGE model family, whereas the GGE biplot presentation is restricted to GGE2, which usually is not the most predictively accurate member of the GGE family. Similarly, the commonly used AMMI biplots are restricted to the AMMI1 and AMMI2 models. Finally, the tabular presentation lends itself to useful summary statistics, such as the average yield advantage in each mega-environment for selecting parsimonious model winners rather than actual data winners, on the assumption that the estimated model is more accurate than its data (as demonstrated by cross-validation or other model diagnosis methods).

Yet another alternative is shown in Fig. 4 of Gauch and Zobel (1997). That figure shows environment means on the abscissa and interactions (also in units of yield, not the square root of yield) for each environment's genotype winner on the ordinate. Yield isolines form a simple herringbone pattern. This display shows at a glance how different main and interaction effects can combine to achieve a given yield level. Whereas the GGE2 biplot with inner-product interpretation provides a lot of detail, this AMMI-based figure provides what one might regard as an executive summary.

Criticism 9 objects that the shapes of mega-environment regions in AMMI1 biplots are completely subjective because this biplot's axes are in different units. This is false. Precisely because AMMI1 mega-environments are delineated by horizontal lines, the resulting rectangular shape for mega-environments is invariant even under axis transformations. The relative sizes of mega-environment regions

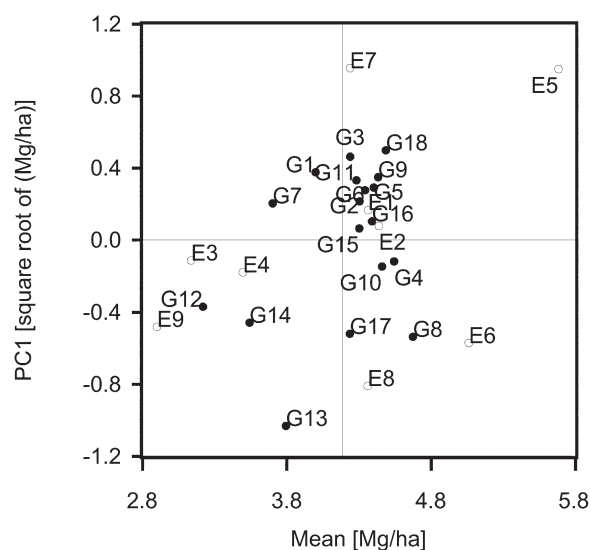


Figure 1. AMMI1 biplot for an Ontario wheat trial. The genotype and environment means are shown on the abscissa, with a vertical line at the grand mean of 4.186 Mg ha⁻¹. The principal component (PC) 1 scores are shown on the ordinate, with a horizontal line at 0. This graph captures 93.4% of the yield variability.

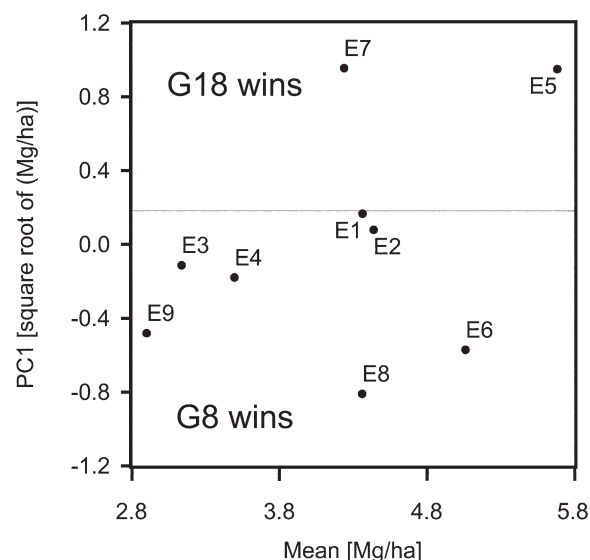


Figure 2. AMMI1 mega-environment display for an Ontario wheat trial. Environment means are shown on the abscissa and principal component (PC) 1 scores on the ordinate. There are two mega-environments, with genotype G18 winning in environments E5 and E7 and genotype G8 winning elsewhere.

could change under axis transformations (which potentially also change axis units), but that would not change the mega-environment assignment of any environment. Given the data, mega-environment boundaries in the AMMI1 display follow inexorably by mathematical equations. There is nothing subjective whatsoever about either the shape or location of mega-environment boundaries.

Criticism 10 objects that the AMMI1 mega-environment display also includes information on E main effects, which is irrelevant to cultivar and test-environment evaluations. Since an axis is used to give the needed main

effects for G, that axis can also present E at no extra cost, so this criticism lacks force. More important, this criticism's logic seems backward. Precisely because the AMMI1 biplot provides G and E main effects, a single biplot serves both crop and soil scientists equally well, whereas comparable information in the GGE paradigm would require a GGE2 biplot for crop scientists and an EGE2 biplot for soil scientists. Given that most sizable agricultural projects involve interdisciplinary teams of crop and soil scientists, a single unified presentation for everyone is far preferable. Although AMMI has been used less frequently in soil sciences than in crop sciences, examples include association of AMMI parameters with soil and other environmental variables (Nachit et al., 1992), analysis of GE responses to soil conditions (Kondo et al., 2003), analysis of soil microbial communities (Thies, 2007), and selecting ideal test sites based on GE responses to the sand and silt contents of soils (Laurentin et al., 2007).

Sometimes the environments are location-year combinations. In such cases, providing information on G and E and GE allows researchers to see whether a test location varies over years in main effects or interaction effects or both or neither (as in Fig. 6.4 of Gauch 1992, p. 218). And by adding mega-environment boundaries (as in Fig. 3 in Gauch and Zobel 1997), researchers can determine whether each test location is predictive for a given mega-environment or else is frequently crossing mega-environment boundaries from year to year (analogous to Fig. 6.5 of Gauch 1992, p. 222).

Response to Criticisms 4, 5, and 8

Three criticisms question the meaningfulness of AMMI2 biplots for showing mega-environments. Beginning with criticism 4, Yan et al. (2007) describe a presumed method for augmenting an AMMI2 biplot with mega-environments. For each winning genotype, its group of environments is determined by consulting a table of predicted yields for the AMMI2 model; those groupings are then superimposed on the biplot. Accordingly, they regard an AMMI2 mega-environment graph as a conclusion-presentation tool rather than a pattern-discovery tool.

However, this presumed method would not work. For instance, again using their Ontario wheat example, the mega-environment with G18 as its winner includes two environments, E5 and E7. But mega-environment regions in AMMI2 graphs are polygons (as in GGE2 as well). Although two points can define a line, they cannot define a region containing any area.

Moving on to the actual methods used to delineate mega-environments in an AMMI2 biplot, a simple method is to cover the biplot with a grid of 70 by 70 or 100 by 100 or whatever evenly spaced points (representing hypothetical environments) and then to note the winning genotype in each pixel. That suffices to delineate mega-environments

within visual accuracy. And if the exact locations of each vertex of a polygon are desired, the genotypes that meet at a vertex provide a system of simultaneous equations that can be solved easily to obtain the exact coordinates.

As previously explained in Gauch (1992, p. 222–230) and Gauch (2006a), these calculations involve not only the first and second PC scores that appear in the AMMI2 graph but also the genotype main effects that remarkably are incorporated without their requiring another axis. But because information beyond that presented directly in the AMMI2 biplot's two axes is used, it is fair to call this a display of mega-environments rather than a discovery. But this distinction seems rather immaterial. In the GGE2 world, either the GGE2 mega-environment graph or the GGE2 table of expected yields can be used independently to “discover” the mega-environments; therefore, neither source of mega-environment information has any inherent priority in terms of discovery.

Rather, the salient distinction is that by incorporating G information without needing an axis, AMMI2 biplots can capture more G+GE than GGE2 biplots. For instance, for the Ontario yield trial of Table 1 in Yan et al. (2007), the which-won-where patterns and mega-environments for AMMI2 are based on 100% of G (automatically) and 71.67% of GE, whereas those for GGE2 are based on 99.45% of G and 53.73% of GE.

Criticism 5 claims that the AMMI2 biplot is incapable of displaying which-won-where patterns because G and PC scores are in different units. This is manifestly false. As just explained, information on G is incorporated without using an axis. Both G and the products of genotype and environment PC scores (which is what enters into the calculations of expected values) are in the same units, namely the units of the data, such as yield.

Figure 4 shows the AMMI2 mega-environment graph for the Ontario wheat trial, which may be compared with the GGE2 graph in Fig. 1 of Yan et al. (2007). The two graphs have a similar geometry. Whereas AMMI2 has a set of irregular polygons covering the graph, GGE2 has a set of irregular polygons arranged like slices of a polygon-shaped pie. However, the AMMI2 display incorporates 86.7% of the G+GE information, whereas GGE2 incorporates only 78.0% (because AMMI2 has two components devoted to GE whereas GGE2 has only one component, PC2, that focuses GE). Genotype G8 wins by having the highest overall mean across environments, as indicated by its mega-environment containing the origin, whereas G6 wins in E1 and E5 and G18 wins in E7 by virtue of sizable positive interactions that more than compensate for these genotypes' smaller means. These two methods provide consistent indications on winning genotypes except for E5, where G6 wins for AMMI2 but G18 wins for GGE2. In the absence of information required for statistical testing of PC axes for this Ontario wheat example, we could not verify whether the additional GE effects taken

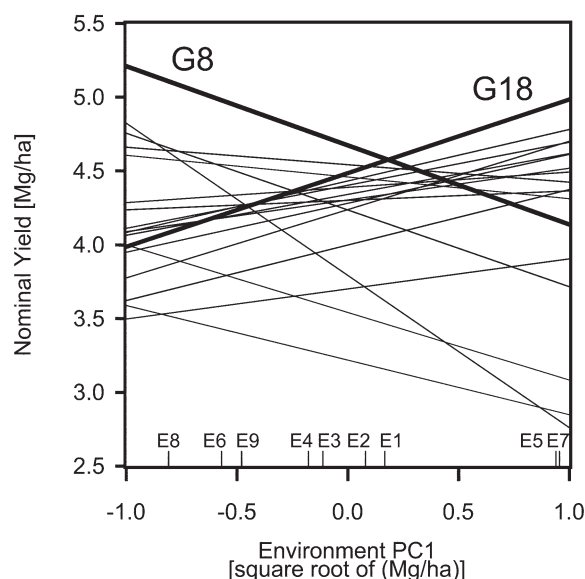


Figure 3. Adaptive responses for Ontario wheat genotypes according to the AMMI1 model. Principal component (PC) 1 scores for the nine environments are shown on the abscissa and nominal yields for the 18 genotypes on the ordinate. The two genotypes, G8 and G18, that win over some range of PC1 scores are identified individually, whereas the responses of the losers are indicated by thin lines. Note that G18 wins in E5 and E7, whereas G8 wins in the other seven environments.

into account by AMMI2 relative to GGE2 (or AMMI1) are dominated by signal or noise.

Criticism 8 claims that Gauch (1992) hypothesizes that universal winners would be located near the origin of the AMMI2 biplot but objects that this is unreliable since universal losers would also be located there. But no such hypothesis is to be found in the pages of Gauch (1992). To the contrary, the markers for genotypes in an AMMI2 biplot (not to be confused with a display of mega-environments, as in Fig. 4) convey no information about universal winners or losers. Precisely because an AMMI2 biplot presents information directly on GE but not G, universal losers (or winners) could appear anywhere and have no requirement to be located near the origin. By contrast, the AMMI1 graph presents information directly on both G and GE, so it can locate potential and universal winners, as shown in Fig. 7 of Gauch and Zobel (1997). (But note the correction that the thick line down the markers for 3165 to 1827 should follow the right side of those points rather than the left.) Also consider a graph of adaptive responses based on AMMI1 as in Fig. 3 (and in Fig. 2 of Gauch and Zobel, 1997).

Response to Criticism 7

Criticism 7 claims that the AMMI2 biplot lacks the inner-product property. That is false. Kempton (1984) first imported the inner-product property of biplots into the agricultural literature from their first description by Gabriel (1971). Kempton used an example of a winter wheat yield trial. He derived

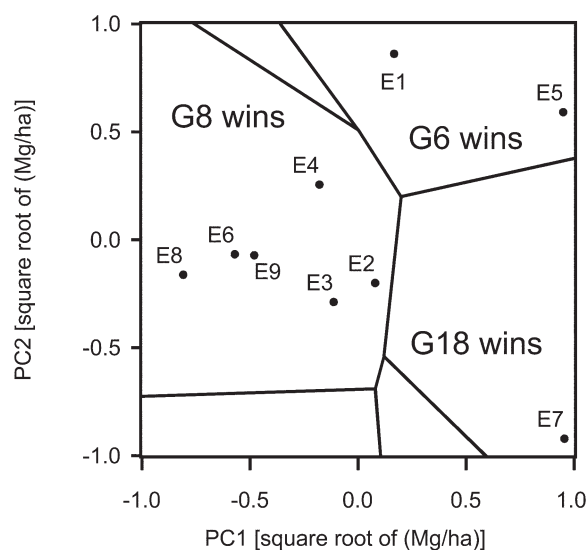


Figure 4. An AMMI2 mega-environment display for an Ontario wheat trial. To avoid clutter, genotype winners are identified only for the three mega-environments that include one or more environments. In addition, G10 would have won for an environment located in the small triangular region at the top, G17 for the bottom left region, G5 for the bottom middle region, and G4 for a region too small to be shown. This graph was constructed with even singular value partitioning (SVP) and no axis rescaling (although other SVP and rescaling choices would leave mega-environment assignments invariant). Products of genotype and environment scores, summed for both principal components, approximate interactions and are in units of yield, namely Mg ha^{-1} . This AMMI2 display incorporates 86.7% of the genotype + genotype \times environment information, whereas its genotype main effects and genotype \times environment interaction 2 (GGE2) counterpart incorporates only 78.0%. (PC, principal component.)

the inner-product formula in his text, showed its geometry in his Fig. 3, and gave a formal proof in his Appendix. Note from his Eq. [9] that the model he used to illustrate the inner-product property is the AMMI2 model. Kempton gave the values obtained by his equation the vague name *residual yields*, but from his definition (Kempton, 1984, p. 127), they are interactions. Accordingly, the “expected response” or “predicted response” to applying the inner-product formula to the AMMI2 biplot for the wheat trial in his Fig. 4 is an interaction, rather than a yield. Kempton was so focused on the GE interactions that his Table 1 presents these interactions rather than the actual yields.

Consequently, it is ironic that Yan et al. (2007) deny that the AMMI2 biplot has the inner-product property because it was precisely the AMMI2 biplot that was used to introduce this property in the agricultural literature. For over two decades, the literature has been clear that the AMMI2 biplot has the inner-product property.

Incidentally, Kempton (1984) included another example of a fungicide trial using the GGE2 biplot, which also has the inner-product property. In this case, these products approximate environment-centered yields, rather than either the actual yields or the interactions.

Continuing the early history of biplots, the very first biplot, Fig. 2 in Gabriel (1971), used neither AMMI2 nor GGE2. Gabriel applied noncentered PCA to a data matrix and then removed the first component because it captures mainly the simple row and column main effects of little interest to him, whereas he was interested in complex “differential” responses captured in components 2 and 3 (also see Jolliffe, 1986, p. 227–228). He explained that his residual matrix “corresponds to interaction residuals after fitting an additive model,” although “corresponds” should not be misinterpreted as “equals.” However, the model that would distinguish main and interaction effects exactly is, of course, AMMI. It is interesting that Gabriel focused on “differential” responses and noted their correspondence with interactions. Gabriel’s approximate interactions were followed by Kempton’s exact interactions.

AMMI2 biplots, which have the inner-product property, have a long history in the statistical literature because researchers frequently need to distinguish average or main effects from “differential” or interaction effects. For agricultural researchers, this corresponds to the fundamentally important distinction between broad and narrow adaptations, which have quite different implications for mega-environments.

Likewise, for ecological data, Digby and Kempton (1987, p. 12) gave several reasons why main effects are usually of less interest than interaction effects. Accordingly, ecologists often construct biplots with AMMI (also called doubly centered PCA). Digby and Kempton (1987, p. 12) and Jolliffe (1986, p. 214) elucidate an interesting relationship between AMMI and correspondence analysis (also called reciprocal averaging): AMMI applies SVD to the residuals from an additive model, whereas correspondence analysis analyzes residuals from a multiplicative model.

An AMMI2 biplot of the GE matrix of interactions provides an inner-product representation of interactions, a GGE2 biplot of the G+GE matrix of environment-centered yields provides an inner-product representation of environment-centered yields, a PCA2 biplot of the data minus grand mean matrix of corrected yields provides an inner-product representation of corrected yields, and a noncentered PCA2 biplot of the actual yields provides an inner-product representation of the actual yields. Which biplots are of greatest interest and utility depends on the research context.

As Yan et al. (2007) observed, the which-won-where view also works for $\text{SVP} = 1$. Indeed, any SV partitioning leaves mega-environment assignments invariant, but because focusing on environments has some advantages in this context, $\text{SVP} = 2$ was used for their Fig. 1.

A future development mentioned in Yan et al. (2007) is a sequential data subdivision method using GGE2 mega-environment displays, guided by formal statistical tests or imposed practical considerations, to produce data subsets that

each comprise an individual mega-environment. Clearly, some such stopping rule or test is needed. For instance, their Fig. 1 subdivides an Ontario wheat trial into two mega-environments, with the larger having seven environments and the smaller having two. However, if a GGE2 biplot with the which-won-where view is constructed for this larger mega-environment, it splits again, with environment E1 separating from the others. Continuing this process leads to no further splits. Hence, depending on what a statistical test or stopping rule would indicate, this Ontario trial may have one, two, or three mega-environments. Which is it? Accordingly, the which-won-where view, as in their Fig. 1, is somewhat uninformative apart from a significance test. Alternatively, mega-environments delineated by a GGE2 biplot (or several sequential biplots) could be justified by considerations of the most predictively accurate member of the GGE model family, perhaps with slight sacrifice of accuracy to gain simplicity as required because of practical constraints requiring few mega-environments.

Response to Criticism 3

Finally, criticism 3 objects that the AMMI2 mega-environment display proposed in 1992 has not yet been shown to be useful. We reply that this neglect is unfortunate, given the above considerations in favor of AMMI2. Hopefully, this communication will encourage researchers to use AMMI2 for mega-environment delineation when the AMMI2 model is deemed predictively accurate. The AMMI2 mega-environment display has been implemented in at least two free software packages used by many breeding programs worldwide, MATMODEL (Gauch, 2007) and CropStat (IRRI, 2008, formerly called IRRISTAT).

GENOTYPE EVALUATION

The principal GGE tool for genotype evaluation is a GGE2 biplot augmented with the AEC axis and its perpendicular (AEC \perp) to separate G from GE, thereby providing the “mean vs. stability” view shown in Fig. 2 of Yan et al. (2007). By definition, the AEC axis passes through the biplot origin and the “average environment,” which has coordinates that are the mean of environment PC1 scores and the mean of environment PC2 scores. Deeming that “genotype evaluation is meaningful only for a specific mega-environment,” this tool is restricted to sets or subsets of environments comprising only one mega-environment. Such biplots use genotype-focused partitioning, SVP = 1.

Yan et al. (2007) made five claims regarding genotype evaluation:

Claim 1. “The AEC abscissa approximates the genotypes’ contributions to G.”

Claim 2. Accordingly, “the AEC ordinate must approximate the genotypes’ contributions to GE, which is a measure of their stability or instability.” That is, AEC and AEC \perp separate G from GE.

Claim 3. Projections of genotype markers onto AEC show “the rank-two approximation of the genotype means.”

Claim 4. “Unless G is too small to be meaningful, the ranking of the genotypes on the AEC abscissa is always perfectly or highly correlated with G.”

Claim 5. Such GGE functionality as allowing genotypes to be evaluated by their mean performance and stability (as in their Fig. 2) has not been shown for any other yield-trial analysis, including AMMI.

Table 1 shows recovery of SS for G (SSG) and SS for GE (SSGE) for various GGE and AMMI biplots. The GGE2 original axes mix G and GE in both PC1 and PC2. The total SSG in the GGE2 biplot is 16.96108 or 99.82% of the dataset’s total of 16.99157. And the total SSGE in the GGE2 biplot is 4.02898 or 42.80% of the dataset’s total of 9.41432. As proven in generality in the Appendix and as shown for the Ontario wheat trial in Table 1, AEC captures all SSG in the GGE2 biplot, regardless of SVP. But for this particular dataset, AEC \perp captures more SSGE with SVP = 1 than with SVP = 2, so the choice of SVP = 1 is preferable.

Table 1. Recovery of sums of squares for genotype (SSG) and sum of squares for genotype \times environment (SSGE) effects by axes in various genotype main effects and genotype \times environment (GGE) and additive main effects and multiplicative interaction (AMMI) biplots for a winter wheat yield trial with a subset of seven environments (from Yan et al., 2007, their Table 1 and Fig. 2). Although not indicated in this table, the AMMI2 mega-environment display captures 100% of the genotype effects in a manner that does not use or require an axis.

Model and axes [†]	Genotype (SSG)	Interaction (SSGE)
GGE2 original axes		
PC1	16.84816	0.97191
PC2	0.11292	3.05708
Sum	16.96108	4.02898
GGE2 rotated, SVP = 1		
AEC	16.96108	0.32708
AEC \perp	0.0	3.70190
Sum	16.96108	4.02898
GGE2 Rotated, SVP = 2		
AEC	16.96108	0.76145
AEC \perp	0.0	3.26753
Sum	16.96108	4.02898
AMMI1		
Main	16.99157	0.0
PC1	0.0	3.78834
Sum	16.99157	3.78834
AMMI2		
PC1	0.0	3.78834
PC2	0.0	2.32742
Sum	0.0	6.11576
Actual data	16.99157	9.41432

[†]PC, principal component; AEC, average environment coordinate.

Hence, regarding claims 1 and 2, the AEC and AEC \perp axes are only partially successful in separating G and GE. For SVP = 1 as used in their Fig. 2, AEC captures 100% of the SSG in the GGE2 biplot or 99.82% of the SSG in the dataset, although it is contaminated with 8.12% of the SSGE in the biplot or 3.48% of the SSGE in the dataset. And AEC \perp captures 91.88% of the SSGE in the biplot or 39.32% of the SSGE in the dataset.

Regarding claim 3, projections of genotype markers onto AEC are rank-two approximations, but not of “genotype means” (over replications). The G+GE matrix contains the yields minus environment deviations and the grand mean, so these projections approximate the environment-centered yield for each genotype. Incidentally, an exception is a dataset with only three environments, which is rank two after environment centering, meaning that the projection yields an exact prediction.

Regarding claim 4, the AEC axis can be highly correlated with G and can have the same genotype rankings. But the AEC axis cannot possibly be “perfectly” correlated with G because the GGE2 biplot does not capture G completely. For their Fig. 2 with SVP = 1 and consequently AEC at 10.99°, Yan et al. (2007) stated that this correlation is 1.0, but in fact, it is 0.99910.

Regarding claim 5, the intended function of the AEC and AEC \perp axes is to separate G from GE, thereby providing the “mean vs. stability” view of their Fig. 2. They

deny that any analysis other than GGE provides such functionality. But consider the results for AMMI in Table 1 and Fig. 1. AMMI is the optimal method for separating G from GE. The abscissa of an AMMI1 biplot captures 100% of G. And among all SVD-based analyses including GGE, by the least-squares property of SVD applied to the GE matrix, the AMMI1 ordinate with its PC1 captures as much as possible of GE. The ANOVA part of AMMI can do what the PCA part of GGE cannot possibly do: separate G from GE.

Admittedly, the GGE2 biplot does capture more GE than the AMMI1 biplot, namely, 4.02898 compared with 3.78834 for this Ontario wheat trial. This is possible because the GGE2 biplot has two axes capturing some GE, whereas the AMMI1 biplot has only one. But this isolated fact could easily be misinterpreted. Note that what matters for GGE2 in the present context of separating G from GE is not the total amount of GE but rather the amounts in AEC and AEC \perp separately because GE in AEC is undesirable whereas GE in AEC \perp is desirable. For the GGE2 biplot using SVP = 1, as in Fig. 2 of Yan et al. (2007), AEC \perp captures 3.70190, which is somewhat less than PC1 for AMMI1 with 3.78834. This is the relevant comparison. In conclusion, compared with AMMI1 axes, AEC recovers less G, and AEC \perp recovers less GE, meaning that GGE2 is doubly unsuccessful at separating G from GE. Incidentally, were SVP = 2 used instead, then AEC \perp would capture considerably less, only 3.26753.

The AMMI1 biplot in Fig. 1 has additional advantages over the GGE2 biplot for showing mean vs. stability. It is simpler to construct and to interpret because its axes are used directly, rather than needing to be rotated. Also, AMMI separates G from GE perfectly regardless of how simple or complex a dataset may be, whereas for GGE the assumption of a single mega-environment is critical because otherwise a large GE relative to G could drive a large portion of G into the third and higher components that a GGE2 biplot misses. Furthermore, the AMMI1 biplot provides all parameters needed to reconstruct this model’s estimates of yields, whereas the GGE2 biplot can access only environment-centered yields.

Having argued that an AMMI1 biplot is superior to a GGE2 augmented biplot with the AEC axis for displaying each genotype’s mean and stability, it must be emphasized that other methods may also merit consideration when AMMI2 or more complex AMMI models are needed to account for GE effects deemed relevant by model validation or statistical tests. One or more stability statistics could simply be tabulated along with mean values of genotypes, allowing researchers to scan exact values of these estimated parameters. For instance, for the Ontario wheat example in Fig. 2 of Yan et al. (2007), AEC \perp in the GGE2 biplot captures only 39.32% of GE; similarly, PC1 in the AMMI1 biplot captures only 40.24%, whereas a tabulated

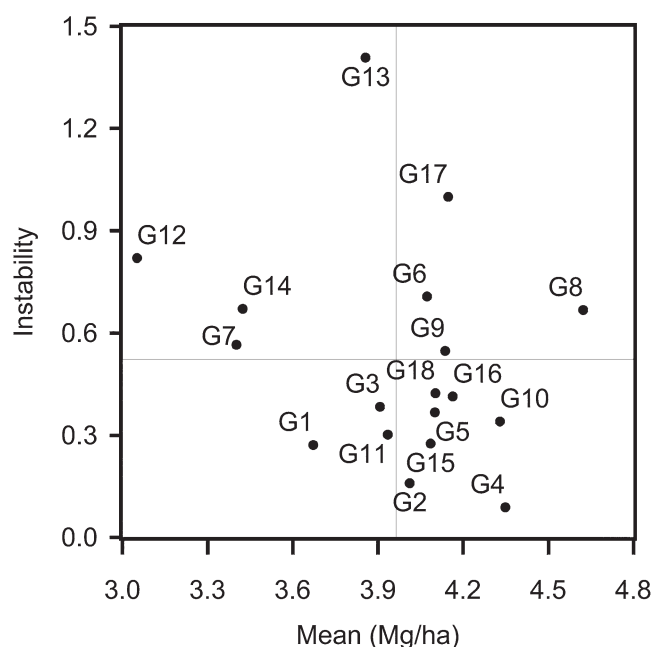


Figure 5. A scatterplot of mean yield and instability for the 18 genotypes in an Ontario wheat trial. Mean yield is a genotype’s average over the environments, including only those seven environments comprising a single mega-environment. Instability is measured as a genotype’s contribution to the sum of squares for genotype-by-environment interaction. The vertical line shows the grand mean of 3.96 Mg ha⁻¹, and the horizontal line shows the average instability of 0.523.

stability measure can access 100% of GE for many such measures, including ecovalence (the SS contribution to GE of each genotype; Wricke, 1962).

Alternatively, a scatterplot could show for each genotype its mean on the abscissa and one desired stability measure on the ordinate. Scatterplots “are not only easy to produce, but they also have the merit of being straightforward to interpret, requiring very little, if any, formal training” (Gower and Hand, 1996, p. 1).

For example, Fig. 5 shows the mean and ecovalence for the 18 wheat genotypes in the seven Ontario locations comprising a single mega-environment in Fig. 2 of Yan et al. (2007). Graphing these two values directly in a scatterplot has advantages over the mean vs. stability view of a GGE2 biplot. Regarding the mean, the scatterplot gives its values for each genotype directly in units of Mg ha^{-1} , while projections onto the tilted AEC axis in the GGE2 biplot involve PC scores with no direct relationship to yield. Regarding instability, the scatterplot gives the contribution to SSGE for each genotype directly, in units of yield squared and with reference to all GE effects. By contrast, projections onto $\text{AEC}\perp$ only approximate those contributions (ecoivalences), perhaps poorly given that $\text{AEC}\perp$ captures only 39.32% of SSGE. In the GGE2 biplot, for instance, the approximated GE contribution for genotype G13 appears to be much smaller than for G17, about a third as large. But from the scatterplot, the fact emerges that genotype G13 makes a larger contribution than G17, actually the largest of all. The reason that the GGE2 biplot seriously underestimates the contribution of G13 is that this genotype has an unusually large score on PC3, which the GGE2 biplot misses.

Although a small dataset is convenient for illustrating statistical techniques and options, the sort of information sought from Fig. 5 here or the comparable Fig. 2 in Yan et al. (2007) requires a larger dataset. The stability measures considered here involve variances, which are second-order statistics. As a rule of thumb, they require at least 20 observations for acceptable accuracy, such as 10 locations for 2 yr or 7 locations for 3 yr. For a simple variance component, such as the environmental variance, which has been used as a stability measure, to be estimated with a coefficient of variation of 20%, 50 environments are required (Piepho, 1998b). Differences in stability should be assessed by significance tests (Piepho, 1996).

More research is needed to compare Fig. 1 and 5 for displaying mean vs. stability. Both cleanly separate G and GE and both capture 100% of G. Although the scatterplot in Fig. 5 has the advantage of capturing 100% of GE, the graph in Fig. 1 has two features that may outweigh that advantage. First, Fig. 5 shows only the relative amounts of GE, whereas Fig. 1 shows the inherent nature of an interaction as a contrast between opposite patterns of differential responses—as well as showing the corresponding interac-

tion patterns for genotypes too. Second, by filtering out noise in the discarded late components and by focusing on the single largest interaction pattern, the AMMI1 biplot in Fig. 1 may be preferable. Further research may even show that accuracy gain from a parsimonious AMMI1 model often allows the parameters in Fig. 1 to be estimated with adequate accuracy and reproducibility for experiments with less data than the 20 environments needed for reliable results in Fig. 5. Agricultural researchers often have an interest in stability (or dependability) despite having experiments with fewer than 20 environments. Consequently, if PC1 scores from AMMI show consistent reproducibility for, say, only 10 or 12 environments, that statistical efficiency could be advantageous.

Yet another possibility to consider for genotype evaluation is combining mean yield and stability into a single parameter of yield reliability (Eskridge, 1990). Also, Annicchiarico (2002) has proposed a simple although approximate method for incorporating temporal stability (assessed from repetition over years) into AMMI-modeled nominal yield responses to locations, thereby modeling yield reliability as the lowest nominal yield of genotypes expected at each site for a given probability of an unfavorable cropping year.

TEST-ENVIRONMENT EVALUATION

The principal GGE tool for test-environment evaluation is a GGE2 biplot augmented with the AEC axis and using environment-focused partitioning $\text{SVP} = 2$ to provide the “discriminating power vs. representativeness” view shown in Fig. 3 of Yan et al. (2007). They state that the “purpose of test-environment evaluation is to identify test environments that effectively identify superior genotypes for a mega-environment” (Yan et al., 2007, p. 647). Hence, presumably, use of this biplot view is limited to datasets or subsets comprising only one mega-environment, thereby ensuring that PC1 and AEC will be dominated by genotype main effects.

Incidentally, the legend to Yan et al.’s (2007) Fig. 3 states that “genotype-focused singular value partitioning” was used, but plainly that is a mistake since numerous other indications are given that environment-focused partitioning was used. Indeed, $\text{SVP} = 2$ is necessary for this figure’s functionality (Digby and Kempton, 1987, p. 63–67). As proven in the Appendix, $\text{SVP} = 2$ uniquely allows SSG and SSGE to be identified for an axis oriented in any direction; $\text{SVP} = 1$ has AEC rotated counterclockwise from the PC1 axis by 10.99° and $\text{SVP} = 2$ has 4.68° . Incidentally, for $\text{SVP} = 2$, the maximum SSG occurs at the AEC of 4.68° , but the maximum SSGE does not occur at the $\text{AEC}\perp$ of 94.68° but rather near 116° . Consequently, as shown in the Appendix, it is impossible to rotate the axes to maximize both SSG in AEC and SSGE in $\text{AEC}\perp$.

AMMI1 and AMMI2 biplots may also be considered for test-environment evaluation. When test locations have

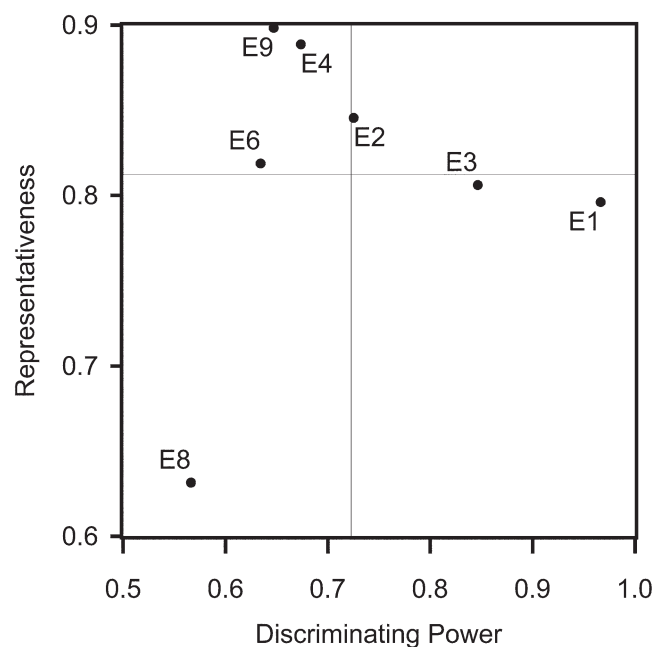


Figure 6. A scatterplot of discriminating power and representativeness for those seven environments comprising a single mega-environment from an Ontario wheat trial. For each environment, discriminating power is measured as the standard deviation of the 18 genotypes in that environment and representativeness is measured as the correlation between genotype values in the environment and genotype means across all environments. The vertical line shows the average discriminating power of 0.723, and the horizontal line shows the average representativeness of 0.812.

been used for several years, AMMI is ideal for discerning different patterns of year-to-year variation and for relating the biplot region occupied by a given location to the mega-environment boundaries at which the winning genotype changes (Gauch, 1992, p. 217–230).

Figure 6 shows a scatterplot for discriminating power and representativeness, which presents more simply and directly the analogous information in Fig. 3 of Yan et al. (2007). The GGE2 biplot feature used for the “discriminating power” of each environment is a vector length that is proportional to “the standard deviation of cultivar means in the environment.” And the feature used for the “representativeness” of each environment is the cosine of an angle between an environment vector and the AEC that approximates “the correlation coefficient between the genotype values in that environment and the genotype means across the environments” (Yan et al., 2007). But if these standard deviations and correlation coefficients are of agricultural interest, these simple quantities can be computed directly from the data and then displayed in a scatterplot. There is no need to perform multivariate analysis, rotate and rescale axes, and visualize lengths and cosines only to get approximate values instead of exact ones.

Although all of the above methods for test-environment evaluation may be useful or complementary, they seem

overly complex and suboptimal. Selection theory proves that just one site characteristic matters for identifying optimal selection sites. The optimal site maximizes the phenotypic correlation between entry yields on the site and entry yields over the target environments (as represented by the test sites for a given mega-environment) because the phenotypic correlation includes the genetic correlation and the broad-sense heritability of the sites (Cooper et al., 1996). An example of optimal site selection in the context of a specific-adaptation strategy is given in Annicchiarico et al. (2005). AMMI analysis was used in combination with cluster analysis to identify two mega-environments for breeding, while phenotypic correlation was used to identify selection locations in each mega-environment.

Finally, researchers may want to consider artificial environments. Analyzing G and GE separately using AMMI for alfalfa (*Medicago sativa* L.) identified two major environmental factors associated with GE interactions across test sites (Annicchiarico, 1992). Subsequently, adaptive responses could be reproduced by means of only a few artificial environments established at a single site and reproducing different mega-environments by managing the levels of the two environmental factors (Annicchiarico and Piano, 2005). This approach allowed breeding for a few different subregions at a much lower cost than the usual strategy requiring more test sites (also see Annicchiarico, 2007a,b).

ANALYZING GENOTYPE AND GENOTYPE × ENVIRONMENT JOINTLY OR SEPARATELY

The most fundamental disagreement between the AMMI and GGE paradigms concerns the relative advantages of analyzing both G and GE jointly in GGE analysis, or else analyzing both G and GE separately in AMMI analysis. Accordingly, this choice merits especially careful attention.

According to Yan et al. (2007), G and GE are interchangeable and their presumed separation in AMMI is based on merely a mathematical distinction lacking biological interpretability and agricultural meaningfulness. In fact, the distinctions between G, E, and GE in ANOVA have both mathematical validity and agricultural relevance, whereas the presumed interchangeability of G and GE renders the GGE position self-refuting. To review and critique the argument in Yan et al. (2007) in favor of GGE, it is helpful to organize their material in the format of a formal argument having four premises and two conclusions.

Premise 1. G and GE can be interpreted as wide and specific adaptations, respectively, only within the confines of a particular experiment’s set of cultivars and set of environments.

Premise 2. GE becomes G if the scope of the environments is narrowed, whereas G becomes GE when

the scope of the environments is widened, so there is no clear biological boundary between G and GE and hence, they are interchangeable.

Premise 3. Furthermore, the only possible basis for a clear biological distinction between G and GE would be that G and GE are controlled by distinct genes or genetic interactions. But quantitative trait locus (QTL) studies and analyses of adaptation show that such is not the case, so again G and GE are interchangeable.

Conclusion 1. Recognition of the interchangeability between G and GE is the sole justification for mega-environments as delineated by GGE analysis, which intentionally combines G and GE.

Premise 4. The distinction between G and GE, having no biological or agricultural meaning, is merely a mathematical distinction in the ANOVA.

Conclusion 2. Recognition that separating G and GE is based on an irrelevant mathematical distinction, rather than a meaningful biological or agricultural distinction, invalidates the case for AMMI at its very theoretical foundations.

This argument may appear to be strong and compelling. But consider the above six elements, one by one.

Critique of Premise 1

Granted, a different experiment with different cultivars and different environments can reach different results. If ever there was an automatic, guaranteed truism, this is it.

But the possibility of different experiments reaching different results in general and the actuality of results being unrepeatable in particular instances constitute two different matters. After all, the story of statistics is the story of using sizable and representative samples to make reasonably reliable inferences to a specified and larger population of inference, such as a dozen experimental locations in Iowa to make a crop recommendation for that state. Hence, premise 1 is overstated.

The valid content in this premise, even if trivial, is that yield-trial research is an experimental science with results that depend on the data. But the invalid implication from this premise is that somehow this dependency of results on data counts against AMMI solely. As the numerous citations from the Gauch (2006a) and Yan et al. (2007) reviews already documented, both G and GE effects—that is, both wide and specific adaptations—found in one experiment are routinely confirmed in additional experiments. Despite some failures, agricultural research is replete with repeatable results. In addition, nonrepeatable GL interaction revealed by repetition over years can be taken into account and used as the error term in selecting the AMMI model, thereby modeling only repeatable GL effects (Annicchiarico, 1997).

Critique of Premise 2

The language in premise 2, that GE “becomes” G (or the reverse), is vague and confusing. Instead, what should be stated is simply that in general similar environments induce small GE interactions whereas diverse environments induce large GE interactions (and likewise for similar and diverse genotypes). The commonplace observation that GE interactions are sometimes small and sometimes large does not in the least imply that G and GE are “interchangeable.”

Critique of Premise 3

Yan et al. (2007) are quite correct in noting that major QTLs “often” affect both G and GE. But it is equally true that major QTLs often cause only GE, as shown by the results in their citations (e.g., Romagosa et al., 1996) and other articles (e.g., Romagosa et al., 1999; Ribaut et al., 2007). Cho et al. (2007), for example, detected 29 QTLs for main effects, 13 QTLs for GE effects, and 6 QTLs for both main and GE effects (ignoring several other QTLs detected with only minor or marginal statistical significance) in rice (*Oryza sativa* L.). And conventional breeding, with its longer history and more extensive experimentation, decisively reinforces the findings from molecular breeding. Specific phenological and physiological traits routinely cause an inverse genetic correlation between yield potential and adaptation to drought tolerance or other stresses. Hence, within reasonably large or diverse target regions, such traits (or combinations of traits) induce positive interactions in a particular kind of environment but negative interactions in other environments. Consequently, they contribute largely or even mostly to GE rather than to G. For instance, see Ludlow and Muchow (1990) and Wright and Rachaputi (2004) regarding drought tolerance and Worku et al. (2007) regarding nitrogen response. Ortiz et al. (2001) and Emebiri and Moody (2006) found high heritability for GE interaction effects captured by AMMI’s PC1 scores.

Yan et al. (2007) advocated the possibly strong correlation (in absolute value) between genotypes’ mean yields and AMMI PC1 scores as a further indication of common genetic control over G and GE effects, citing Ebdon and Gauch’s (2002a) study of turfgrass quality as an example of that correlation. Studies on grain yield of major crops have shown this correlation in the presence of wide variation in site mean yield because of the contrast between high- and low-yielding sites for GE effects and the positive relationship of genetic variance with mean yield of sites. This implies that genotypes with large positive interactions with high-yielding sites also possess high mean yields (as in Annicchiarico et al., 2006). This scale effect of heterogeneity of genotypic variance among environments may account for most of the GE interaction variance, may produce noncrossover interactions, and may be

removed by a suitable data transformation (Cooper et al., 1996; Annicchiarico, 2002, p. 51–54). Thus, the relationship of practical relevance to crop scientists, namely, that between G and crossover GE effects, cannot be revealed simply by correlations between mean yield and GE interaction parameters of genotypes. On the whole, there is little evidence from QTL studies, physiological information, and analyses of adaptation for common genetic control of G and relevant GE effects.

Critique of Conclusion 1

Indeed, an “interchangeability” between G and GE would be the best justification, even if not the sole justification, for GGE analysis. But the premise that G and GE are interchangeable is false, so this justification fails.

Critique of Premise 4

Genotype and GE effects are defined and distinguished by a mathematical conceptualization of yield trials, namely, ANOVA, as invented by Sir Ronald Fisher nearly a century ago. Admittedly, G and GE are not “things” in the real world, unlike the cultivars and environments that are physical things. The relevant issue here, however, is not whether G and GE are concepts or things but whether they are useful concepts or useless concepts.

To clarify the issue, consider an equivalent example. Speed or velocity is also a mathematical construct, namely, the first derivative of distance with respect to time (acceleration and shock are the second and third derivatives). There is no such “thing” as motion or velocity—rather, what are real are just objects in different places at different times. So, is velocity (or acceleration or shock) a useless concept, just because it is a mathematical construct involving differential calculus?

Critique of Conclusion 2

The persistent claim in the GGE literature that G and GE are indistinct or interchangeable is offered as a decisive criticism of AMMI. However, on two counts, when followed to its logical conclusion, this claim makes the GGE literature self-refuting.

First, given Eq. [4] for GGE or Eq. [5] for GGE2 more specifically (or Eq. [2] in Gauch 2006a or Eq. [1] in Yan et al., 2007), the very first step in GGE analysis is removing E. This step is necessary because the GE data matrix from a yield trial contains G+E+GE; so, to obtain the G+GE portion for GGE analysis, E must be removed. But removing E and removing G are entirely symmetric operations.

Consequently, the GGE literature’s verdict that G is inseparable from (or interchangeable with) GE must apply equally to E being inseparable from GE, which in turn applies equally to E being inseparable from G+GE. And bear in mind that the GGE literature takes inseparability to mean a lack of biological and agricultural mean-

ingfulness. The inexorable implication follows that GGE analysis begins with a meaningless calculation, which is self-refuting for the GGE position.

Gauch (2006a) already expressed this criticism. Again in other and fewer words, to question the meaningfulness of the distinction between G and GE is to question ANOVA, which the GGE literature needs to distinguish and separate E from G+GE. Readers of both reviews will notice that Yan et al. (2007) did not respond to this criticism in their review.

Second, although the boundaries between mega-environments depend on both G and GE, as everyone agrees, the existence of mega-environments depends on GE alone. If there were no GE interactions, then the rankings for genotypes tested in one environment would automatically apply to any other environment; thus, a new genotype would need to be tested just once against the current leader. The separation of GE from G and E should therefore be regarded as a relevant mathematical distinction because of the enormous agricultural significance and precise biological interpretation of GE as the sole cause of the existence of mega-environments. Accordingly, a GGE literature that persisted in claiming that G and GE are indistinct and interchangeable would be a literature that could not explain the existence of mega-environments. But this is a high price to pay because handling mega-environments is the centerpiece of Yan et al. (2007), as in their Fig. 1. Likewise, their claim that the AEC view of the biplot in their Fig. 2 and 3 “does reparate G from GE whenever G is sizable” necessitates that G and GE be distinct and meaningful portions of the overall variation. Otherwise, no meaning could be given to their “mean vs. stability” or “discriminating power vs. representativeness” view of a GGE biplot, which again is a high price to pay. More generally, “G” and “GE” separately, as contrasted with “G+GE” jointly, appear literally dozens of times in Yan et al. (2007). Consequently, one can hardly imagine how a person who really thought that G and GE are interchangeable could manage to comprehend that paper.

There is only one way for the GGE literature to preserve (i) the validity of its first step of removing E from G+E+GE to obtain the necessary G+GE portion of the total variance and (ii) an explanation for the existence of mega-environments, necessarily in terms of GE alone. The GGE literature must retract the error of claiming that G and GE are inseparable and interchangeable, and it must reinstate the mathematical, biological, and agricultural meaningfulness of the distinctions in ANOVA between G, E, and GE.

MODEL DIAGNOSIS AND ACCURACY GAIN

The review article by Gauch (2006a) emphasized that AMMI and GGE (and other variants of PCA) are essentially equal in their ability to gain accuracy by partitioning

a parsimonious and signal-rich model from a discarded and noise-rich residual. Accordingly, there would seem to be no need for any further discussion of the relative merits of AMMI and GGE for gaining accuracy. However, the review by Yan et al. (2007) maintained a huge discrepancy between the AMMI and GGE literatures in their perspective on, and implementation of, opportunities to gain accuracy. So, despite no difference in theory and potential, there is a marked difference in practice that merits discussion. The position on model diagnosis and accuracy gain in the GGE literature, as reviewed by Yan et al. (2007), has three principal claims:

Claim 1. Accuracy gain should not be overstated.

Researchers usually recommend a set of top cultivars, not just one cultivar, to the growers in a given region, so the exact rankings among these nearly equal cultivars are not important. Indeed, in practice it suffices to classify genotypes into just a few coarse categories, such as excellent, acceptable, and unacceptable. Also, breeders usually select for multiple traits, which reduces the relevance of great accuracy for any individual trait, including yield.

Claim 2. More pointedly, there is a fundamental difference between truly predicting future performance and merely “predicting” past performance. The latter, which has been prominent in the AMMI literature, uses data splitting or cross validation in which part of the data are used to assess the accuracy of AMMI models based on the remaining data from the same experiment. For real prediction from one experiment to other experiments, three specific conditions (discussed below) must be met, and all three are untenable in actual agricultural research.

Claim 3. Model diagnosis is useful to avoid underfitting or overfitting of mega-environments. Whereas claims 1 and 2 reiterate the ongoing position of the GGE literature, claim 3 introduces new research currently in progress.

Critique of Claim 1

In fact, accuracy matters. Gauch (2006a,b) gives many reasons why accuracy matters for both researchers and growers, so those reasons are not repeated here.

Classifying genotypes into just three categories—such as, good, medium, and poor—may suffice for the first of several stages of selection. Nevertheless, this seemingly modest task may be harder than one might suspect. Assume, for example, that numerous genotypes are to be screened with the objective of keeping the top 10 or 20%, but with only a single location and with little or no replication, the coefficient of variation is fairly large, for example, 15% or more. That noise level will cause many truly superior genotypes to be discarded summarily, as well as many truly inferior ones to be advanced, only to

be discarded eventually after conducting more extensive and expensive tests (Gauch, 2006a). Thus, even coarse screening benefits from accuracy. On the one hand, given an early stage of selection without multiple environments, AMMI and GGE analyses are inapplicable, so a parsimonious model will not be a resource for gaining accuracy. On the other hand, were AMMI or GGE modeling applied to later stages with multiple environments for accuracy gain, that efficiency gain could liberate resources for improving the early stages of selection by means of more replication or other refinements (Gauch and Zobel, 1996).

A major focus of the GGE literature is delineating mega-environments by identifying the winning genotype in each environment, which contrasts with the objective advocated here of targeting several genotypes in each site. Indeed, sometimes the objective is to recommend a single genotype. On the other hand, predictively accurate yield estimates can be used to define mega-environments on the basis of more than one high-yielding genotype (Annicchiarico et al., 2006).

Furthermore, Yan et al. (2007) offered the premise that breeders often optimize several traits simultaneously in support of the conclusion that advantages from accuracy gain should not be overstated, but this logic seems backward. The more plausible conclusion regarding this challenging selection task is that using AMMI or GGE to gain accuracy for all of the traits would be enormously helpful.

Critique of Claim 2

Yan et al. (2007) objected that claims of AMMI accuracy gain depend on three conditions that are all implausible in actual agricultural research. But these conditions are overstated. They are examined here, starting with the second condition.

Their condition 2 requires that the most accurate member of the AMMI family must be used for recommendations, rather than substituting a simpler model for the sake of obtaining a smaller roster of winners that suits practical constraints. They observe that this was done in Ebdon and Gauch (2002b) and argue that this substitution “renders the model diagnosis completely irrelevant” (Yan et al., 2007, p. 651).

Recall the relationship between accuracy and parsimony, which has been well understood in theory and repeatedly demonstrated in practice. Although SVD-based analyses had been applied to yield trials for elucidating patterns at least as early as Fisher and Mackenzie (1923), the first application for the purpose of gaining accuracy was apparently Gauch (1988). Figure 1 in Gauch (1988) depicts the selective recovery of signal in early model components and the selective recovery of noise in late model components, such that overly simple models underfit real signal and overly complex models overfit spurious noise, whereas an intermediate range of relatively parsimonious models

is more accurate than the data, that is, the full model. Figure 2 in Gauch (1988) gave the example of a soybean yield trial, with AMMI1 to AMMI13 more accurate than the data (or full model, AMMI14) and AMMI0 less accurate. Among the 13 models that gained accuracy, AMMI1 achieved the highest accuracy gain. Four years later, MacKay (1992) gave this accuracy response, which occurs in countless contexts across the sciences, a deep theoretical understanding and an apt name, Ockham's hill. Subsequently, Gauch (1993, 2002, 2006b) presented numerous examples of Ockham's hill. All of these examples follow the pattern of the Gauch (1988) example, that several members of the AMMI family are more accurate than the data, not just one member. Likewise, Table 1 in Ebdon and Gauch (2002b) showed that AMMI2 to AMMI7 are more accurate than the Kentucky bluegrass (*Poa pratensis* L.) data, with AMMI7 most accurate, whereas AMMI0 to AMMI7 are more accurate than the perennial ryegrass (*Lolium perenne* L.) data, with AMMI2 most accurate.

Of particular interest are the rather common instances in which Ockham's hill has a broad peak, in which case a slight decrease in accuracy can provide the tradeoff of a substantial increase in parsimony. And in multi-environment trials, greater parsimony means fewer mega-environments that are more manageable given various practical constraints, as Fig. 5 of Gauch and Zobel (1997) shows.

Hence, the logic in Yan et al.'s (2007) condition 2 seems backward. A switch from the most predictively accurate AMMI model to a simpler one accommodating practical constraints does not render model diagnosis "completely irrelevant," but rather, a switch is particularly well justified precisely when model diagnosis has shown that the simpler AMMI model also gains accuracy relative to the actual data, perhaps even with little sacrifice relative to the very best model. Also, note that Ebdon and Gauch (2006b) did use the higher-order AMMI models for estimating and ranking yields in their Fig. 1, whereas for the different research purpose of delineating mega-environments and displaying adaptive responses, they used the simpler AMMI1 model in their Fig. 2 to accommodate practical constraints. When a complex AMMI model provides a sizeable advantage over a simpler one in terms of predictive accuracy, condition 2 can easily be satisfied by targeting cultivars on the basis of tabulated expected yields for this model.

In much of the AMMI literature, model diagnosis has involved the error mean square, either directly (Gauch, 1992, p. 147) or indirectly by means of cross validation (p. 134–146). But when the research purpose is mega-environment delineation for trials repeated over years, the units to be grouped into a mega-environment are the locations rather than the environments (location-year combinations), so the genotype \times location \times year interaction (that is, the nonrepeatable GL interaction) is the relevant

error term (Annicchiarico, 1997). In the context of trials repeated over years, using that error term improves the prediction of future responses and tends to diagnose simpler AMMI models, thereby reducing the amount of further adjustment required because of practical constraints.

Returning to Yan et al.'s (2007) condition 1, they insist that the accuracy gain from the best model is irrelevant in practice unless it leads to different cultivar recommendations than those indicated by the original data. This is a partial truth requiring some clarification. As explained above, accuracy gain may result from the most accurate member of a model family, but alternatively, a slight sacrifice in accuracy may be a worthwhile tradeoff to simultaneously achieve a small and workable number of mega-environments. Given that clarification, a parsimonious model must indeed make some different recommendations from the actual data to matter, which needs to involve some environments though not necessarily all. But far from being problematic, in practice this condition is met routinely.

For example, Table 1 in Yan et al. (2007) involves 18 genotypes, of which five are winners in one or more environments. But five mega-environments for Ontario winter wheat are decidedly excessive and unwieldy. Hence, there is merit in their parsimonious GGE2 model recognizing just two winners and hence two mega-environments—recall that Gauch (2006a) duplicated that result with the AMMI1 model.

Finally, Yan et al.'s (2007) condition 3 insists that future performances must be exactly the same as those indicated by the current data. And the authors argue that this condition is almost always false because of GE interactions and the fundamental difference between really predicting future performance and merely "predicting" past performance.

In reply, past and future performances need not be exactly the same, so condition 3 is excessive. Rather, it suffices for future performances to be substantially correlated with past performances. This adequate condition is met when past experiments guide growers' future choices within the same mega-environment. Indeed, the big payoff for delineating mega-environments successfully is that it allows past experimental results to be applied appropriately to future farming decisions. Only if a statistical method fails to delineate sensible mega-environments would it follow that past performance cannot guide future decisions.

That predicting future performance is harder than estimating past performance is a very good reason for doing more than fitting a parsimonious AMMI or GGE model. It is also a very bad reason for doing less.

Crossa et al. (1991) provided a dramatic example of the benefit of accuracy gain from AMMI analysis, specifically in the context of stratifying genotypes into three classes: top, middle, and bottom. They studied an international bread wheat trial with 18 genotypes tested in 25

locations. Remarkably, although these locations across several continents were extremely diverse, the quite parsimonious AMMI1 model was most predictively accurate, and it identified merely two mega-environments, differentiated by the degree of terminal heat stress. Correspondingly, the genotypes were placed in three groups having early, medium, and late maturity; clearly, the early maturity group is favored in locations with severe terminal heat stress. Their Fig. 2 shows that for three genotypes in the medium maturity group having small GE interactions, the “AMMI1 estimation had a profound effect, producing sharper, stratified ranking patterns” (Crossa et al., 1991). The AMMI1 yield estimates, which discarded the higher components containing mostly noise, gave each of these genotypes much more consistent rankings than did the noisy raw data, so that each genotype could be placed in the top, middle, or bottom group with much greater confidence. Turning to other genotypes with early or late maturity, which generated large GE interactions related to terminal heat stress, their opposite performances in the two mega-environments was very clear using AMMI1 adjusted yield estimates.

For another example, J. Scott Ebdon of the University of Massachusetts, Amherst, and Gauch have obtained preliminary results from a multiyear experiment to compare the predictive accuracy of AMMI estimates and actual data for the National Turfgrass Evaluation Program. At each of six locations, using 3-yr average quality ratings for the 1997 to 1999 Kentucky bluegrass tests, five winning genotypes were selected based on the actual data and another five winners based on AMMI estimates (with the occasional exception that a genotype appearing in both lists was bypassed). These 10 genotypes (in each location) were established in fall 2003. Quality data were subsequently collected in 2005 and 2006 and then averaged. The main result was that AMMI selections from the 1997 to 1999 data were five times as successful as the raw-data selections in recommending those genotypes that performed best (ranks 1 and 2) in the future, 2005 to 2006. Likewise, the raw-data selections were twice as likely as AMMI to recommend genotypes that performed worst (ranks 9 and 10). Hence, AMMI analysis increased predictive accuracy substantially (Ebdon and Gauch, unpublished data).

For a third example, Annicchiarico et al. (2006) found repeatable GL interactions over years for durum wheat (*Triticum durum* Desf.) in Algeria and hence, consistent mega-environments with substantially predictable interactions. Variety recommendations were based on 2 yr of data, using both the raw averages and the AMMI1 estimates. Both sets of recommendations were then compared using a third, independent test year. The yield gain across Algeria relative to the currently grown cultivars obtained by AMMI1 modeling was 10.3% compared with 6.2% for the full model (the observed 2-yr data). Thus, AMMI1 analysis provided

an advantage of $(10.3 - 6.2)/6.2$ or 66% in yield gain at the very modest cost of the data analysis as compared to the high cost of multi-environment testing. Therefore, using a parsimonious model to gain accuracy was a highly cost-effective means for increasing efficiency and accelerating progress. Also, AMMI1 implied simpler recommendations with just 3 different winners and hence just 3 mega-environments, whereas the raw data indicated 15 winners and mega-environments. A more complex analysis, factorial regression (van Eeuwijk et al., 1996), which requires additional environmental data, achieved a slightly greater yield gain of 11.8%. Annicchiarico (2007b) also compared selection gains in breeding for broad and narrow adaptation using artificial environments that reproduced three mega-environments defined by AMMI and cluster analysis for alfalfa in northern Italy. The adoption of distinct genetic bases for the two most-contrasting mega-environments allowed for more than doubling the selection efficiency over the three mega-environments. AMMI analysis allowed identification of specific crop traits, germplasm pools, and testing conditions that together defined a specific-adaptation breeding strategy (Annicchiarico and Piano, 2005; Annicchiarico, 2007a). Repeatable interactions and consistent mega-environments are the basis for one experiment having predictive relevance for other experiments under similar conditions.

Empirical results thus confirm what statistical theory would suggest, that increasing accuracy for a given experiment increases predictive success for reasonably similar environments. That said, clearly much more could and should be done, including the following four opportunities. First, the AMMI literature has so far emphasized the model family comprised of truncated models—AMMI1, AMMI2, and so on—and the resulting Ockham’s hill. Particularly because AMMI1 and AMMI2 can also be used for informative biplots, these models are quite useful. Nevertheless, Piepho (1998a) and Cornelius and Crossa (1999) showed that a wider class of shrinkage models (for AMMI or GGE or other models) can achieve even greater accuracy. Second, Piepho and Möhring (2005) explored best linear unbiased predictors that help researchers exploit both broad and narrow adaptations simultaneously for growing regions subdivided into mega-environments. Having compared several research strategies, they found that simple means within each mega-environment—which are what researchers use most frequently—are consistently the worst option because “yield data from neighboring subregions [mega-environments] may be exploited to improve yield estimates for the subregion of interest.” Third, planned experiments have two designs: the treatment design that a parsimonious model such as AMMI or GGE can address, and the experimental design involving layout and replication. Overwhelmingly, the most common experimental design for yield trials is the randomized complete block design, which reduces the error mean

square (and thereby increases statistical significance) but which does not adjust yield estimates closer to the true values. By contrast, better experimental designs can make adjustments that increase accuracy, including lattice and row-column designs (John and Williams, 1995), designs optimized with a spatial analysis in mind (Cullis et al., 2006; Williams et al., 2006), improved methods of analysis for individual traits such as mixed model analysis for recovery of information (Federer and Wolfinger, 1998), and geostatistical analyses (Gilmour et al., 1997). Furthermore, this opportunity to gain accuracy obtains even for single-environment yield trials, for which AMMI and GGE are inapplicable. Finally, when selection of the best genotypes or treatments is a major research objective, the numbers of genotypes, environments, and replications should be chosen carefully to optimize selection gains (Gauch and Zobel, 1996), while also satisfying any additional research requirements and constraints.

Critique of Claim 3

Regarding claim 3, we only disagree that model diagnosis is unimportant for accuracy gain, as already explained, but we agree that model diagnosis is (also) important for delineating mega-environments. Figure 5 of Gauch and Zobel (1997) is relevant in this context. As Gauch (2006a) already emphasized, a model sequence such as GGE1, GGE2, GGE3, and so on leads to different groups of mega-environments; thus, this unavoidable model choice should be guided by relevant criteria and model diagnosis.

In our opinion, the most disputable claim in Yan et al. (2007) is that “understanding the patterns in a GED [genotype \times environment data] set is more important than getting some ‘accurate’ estimates.” Clearly, these objectives are interrelated, with the latter being the prerequisite for the former. To understand patterns in a dataset, researchers should always require accurate estimates initially. Then they can explore the patterns by some biplot or other methods. The success of this endeavor will depend crucially on having accurate estimates.

When the sole purpose of an analysis is accurate prediction (as opposed to understanding complex GE interactions), the GGE model may have merit. For a given dataset, AMMI or GGE or EGE or another SVD-based analysis may be most accurate, as judged by the Akaike or Bayesian information criterion—although the differences are often rather small (Piepho, 1998a; Cornelius and Crossa, 1999). In the mixed model context for modeling genetic covariance between different environments, GGE is suitable (Piepho and Möhring, 2005).

In conclusion, accuracy does matter. The accuracy gain that AMMI or GGE can deliver routinely equates to having two or three times as many replications. Whereas the cost of data collection is great, the cost of statistical analysis is trivial; therefore, extra accuracy at trivial cost

is a great bargain. Furthermore, given sensible mega-environments, delineated by either AMMI or GGE with appropriate model diagnosis to avoid underfitting or overfitting, past experimental results have relevance for future cultivar choices within a mega-environment. Consequently, the persistent tendency in the GGE literature to discount accuracy gain, even though GGE has just as much ability in this regard as any competitor, can only be a disservice to research practitioners.

DISCUSSION

We have already responded here to Yan et al.’s (2007) review, addressing its five main topics in the same order. But several additional matters merit discussion. Gauch (2006a) critiqued the application of GGE for analyzing heterotic patterns in diallel crosses because “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 re-separated satisfactorily. However, discarding one of the parent’s main effects is peculiar and re-separation may be problematic.” By contrast, “AMMI completely and automatically distinguishes general combining ability (for both male and female parents) and specific combining ability.” But Yan et al. (2007) have not responded to that critique. Instead, they favorably cite four papers using GGE for analysis of heterotic patterns.

The most extensive critique in Gauch (2006a), which involves all three of its tables, is that interpretation of GGE2 biplots is problematic because there are nine cases (plus intermediate gradations) for how G and GE information can appear in the two axes of a GGE2 biplot. Perhaps the intended solution in Yan et al. (2007) is to restrict augmented biplots with an AEC axis to data subsets containing only one mega-environment, although no such restriction is mentioned for other GGE2 biplots. A more forthcoming response would be helpful.

Yan et al. (2007) propose sequential data subdivision using GGE2 biplots guided by formal statistical tests or imposed practical constraints to produce subsets comprising individual mega-environments. But this seems rather unsatisfactory. First, producing separate biplots for each mega-environment seems both cumbersome and unambitious. The exciting potential of an effective biplot is to reveal the overall patterns in a yield trial as a whole. Second, if GGE2 biplots are pressed into service as a classification technique, requiring numerous sequential steps and tests, researchers may question whether some long-established hierarchical classification technique would provide nicer results in a single step. Third and perhaps most important, producing cultivar recommendations for individual mega-environments considered in isolation has been shown to be remarkably inaccurate and inefficient (Atlin et al., 2000; Piepho and Möhring, 2005).

Increasingly, contemporary agricultural research to improve crops uses a combination of traditional breeding and molecular techniques. In their discussion of data centering options, Yan et al. (2007) noted that double-centered data are best “for studying gene expression data where it is the relative change of gene expression levels, as opposed to the absolute levels of the genes or of the treatments, that is the research focus”; that is, “if GE is of sole interest.” We agree that AMMI is appropriate for microarray data because interest usually focuses on the differential expression of genes in the various tissues or times or whatever treatments are under study. It has also been useful with terminal restriction fragment length polymorphism analyses of extracted soil DNA to show differential changes in microbial communities associated with environmental differences (Culman et al., 2006; Thies, 2007). Furthermore, AMMI has been used in QTL scans for GE interaction (differential) effects (Romagosa et al., 1996; Emebiri and Moody, 2006; Cho et al., 2007).

Multivariate analyses such as AMMI and GGE have substantial ability to partition a signal-rich model from a noise-rich discarded residual, thereby gaining accuracy (Gauch, 1988, 2006b; Cornelius and Crossa, 1999). But other analyses, especially agglomerative hierarchical classifications, lack that ability and hence are quite vulnerable to noise (Smith and Gauch, 1992). Accordingly, a noise-susceptible analysis can be improved by first using a noise-reducing analysis like AMMI to preprocess the data, replacing the noisy actual data with model estimates from the most predictively accurate member of the AMMI model family.

Ongoing software development is critical. For instance, Gower and Hand (1996) and Gauch and Zobel (1997) showed several potentially useful kinds of graphs that are not yet implemented in any available software. Trials repeated over years have special requirements for model diagnosis and special opportunities for mega-environment delineation that need further software development. And a particularly urgent imperative is to make user-friendly software readily available to those whose crop environments are most diverse and whose food supplies are most insecure: third-world agricultural researchers. There are at least three free implementations of AMMI (Gauch, 2007; Onofri and Ciricifolo, 2007; IRRI, 2008). Greater consensus on the best statistical methods for yield-trial analysis would be enormously beneficial because it could focus limited resources for software development on the most promising statistical analyses and graphical displays.

CONCLUSIONS

The fundamental reason AMMI is appropriate for agricultural research is that the ANOVA part of AMMI can separate the G and E main effects and the GE interaction effects that present researchers with such vastly dif-

ferent problems and opportunities, whereas separation of G from GE is a mathematical impossibility for GGE analysis. Furthermore, the principal components part of AMMI can separate the prime, signal-rich portion of the interaction in the early components, GE', while discarding the noise-rich portion in the late components, GE*. Crop scientists focus on G and GE', whereas soil scientists focus on E and GE', while both ignore the noise in GE*. Thereby, in an integrated manner, AMMI suits interdisciplinary research teams with both crop and soil scientists. In addition, AMMI offers better opportunities than GGE for graphically displaying mega-environments and adaptive responses. The reward for effective statistical analysis of agricultural trials is faster improvement of yield, disease resistance, nutritional quality, and other important traits.

APPENDIX

A GGE biplot is based on the SVD $\mathbf{Y}^* = (\mathbf{I} - p^{-1}\mathbf{1}\mathbf{1}')\mathbf{Y} = \mathbf{U}\mathbf{S}\mathbf{V}'$, where \mathbf{Y} is a $p \times q$ matrix of yields of p genotypes and q environments, \mathbf{I} is an identity matrix, $\mathbf{1}$ is a vector of ones, \mathbf{S} is a diagonal matrix of singular values ($\lambda_1, \lambda_2, \dots$), and \mathbf{U} and \mathbf{V} containing eigenvectors γ_{gn} and δ_{en} , respectively, obey the conditions $\mathbf{U}'\mathbf{U} = \mathbf{I}$ and $\mathbf{V}'\mathbf{V} = \mathbf{I}$. Let $\mathbf{A} = \mathbf{U}\mathbf{S}^z$ and $\mathbf{B} = \mathbf{V}\mathbf{S}^{1-z}$ be scores for G and E, respectively, where z is a scalar determining the partition of the singular values among E and G scores, so that $\mathbf{Y}^* = \mathbf{A}\mathbf{B}'$. A GGE2 biplot is based on the first two columns of \mathbf{A} , $\mathbf{A}_2 = \mathbf{U}_2\mathbf{S}_2^z$ (markers for G) and \mathbf{B} , $\mathbf{B}_2 = \mathbf{V}_2\mathbf{S}_2^{1-z}$ (markers for E), so that we may approximate $\mathbf{Y}^* \approx \mathbf{A}_2\mathbf{B}_2'$.

It will now be investigated whether a GGE2 biplot can separate G from GE, and how this depends on rotation and scaling (z) of biplot axes. We first show that identification of G and GE is possible in the direction of the unrotated principal component axes. We then show that identifiability also holds for any rotation of the axes, when $z = 0$, but not generally otherwise. Finally, the special case of a rotation to the AEC axis is considered. In this case, G is fully captured by AEC, regardless of the choice of z , while GE can be identified only when $z = 0$ or $z = 1$, generally being split among both AEC and AEC \perp . An important consequence is that GE is never fully captured by AEC \perp alone.

Unrotated Principal Component Axes

The sum of squares for G (SSG) and the sum of squares for GE (SSGE) for the approximation $\mathbf{Y}^* \approx \mathbf{A}_2\mathbf{B}_2'$ are given by

$$\text{SSG} = \text{trace}(\mathbf{A}_2\mathbf{B}_2'\mathbf{K}\mathbf{B}_2\mathbf{A}_2') \quad [\text{A1}]$$

and

$$\text{SSGE} = \text{trace}(\mathbf{A}_2\mathbf{B}_2'\mathbf{P}\mathbf{B}_2\mathbf{A}_2') \quad [\text{A2}]$$

where $\mathbf{K} = q^{-1}\mathbf{1}\mathbf{1}'$ and $\mathbf{P} = \mathbf{I} - \mathbf{K}$ (also see Laffont et al., 2007). The SSG can be written as

$$\begin{aligned} \text{SSG} = & \text{trace}(\mathbf{a}_1 \mathbf{b}'_1 \mathbf{K} \mathbf{K} \mathbf{b}_1 \mathbf{a}'_1) + \text{trace}(\mathbf{a}_2 \mathbf{b}'_2 \mathbf{K} \mathbf{K} \mathbf{b}_2 \mathbf{a}'_2) \\ & + \text{trace}(\mathbf{a}_1 \mathbf{b}'_1 \mathbf{K} \mathbf{K} \mathbf{b}_2 \mathbf{a}'_2) + \text{trace}(\mathbf{a}_2 \mathbf{b}'_2 \mathbf{K} \mathbf{K} \mathbf{b}_1 \mathbf{a}'_1) \end{aligned} \quad [\text{A3}]$$

where $\mathbf{A}_2 = (\mathbf{a}_1 \ \mathbf{a}_2)$ and $\mathbf{B}_2 = (\mathbf{b}_1 \ \mathbf{b}_2)$. The first two terms on the right-hand side (RHS) of Eq. [A3] correspond to the SSG for the first two unrotated axes. The last two terms always vanish, because (Harville, 1997, p. 50)

$$\begin{aligned} \text{trace}(\mathbf{a}_1 \mathbf{b}'_1 \mathbf{K} \mathbf{K} \mathbf{b}_2 \mathbf{a}'_2) &= \text{trace}(\mathbf{b}'_1 \mathbf{K} \mathbf{K} \mathbf{b}_2 \mathbf{a}'_2 \mathbf{a}_1) \\ &= \text{trace}(\mathbf{b}'_1 \mathbf{K} \mathbf{K} \mathbf{b}_2 \mathbf{0}) = 0 \end{aligned} \quad [\text{A4}]$$

The same result is obtained for SSGE by replacing \mathbf{K} with $\mathbf{P} = \mathbf{I} - \mathbf{K}$. These results show that G and GE can be separated in the direction of the principal axes, regardless of the choice of z .

Orthogonally Rotated Axes

Consider an orthonormal rotation matrix $\mathbf{\Omega} = (\mathbf{o}_1 \ \mathbf{o}_2)$, which rotates G and E scores according to $\mathbf{A}_2 \Rightarrow \mathbf{A}_2 \mathbf{\Omega}$ and $\mathbf{B}_2 \Rightarrow \mathbf{B}_2 \mathbf{\Omega}$, respectively. The SSG based on the rotated axis scores becomes

$$\begin{aligned} \text{SSG} = & \text{trace}(\mathbf{A}_2 \mathbf{\Omega} \mathbf{\Omega}' \mathbf{B}'_2 \mathbf{K} \mathbf{K} \mathbf{B}_2 \mathbf{\Omega} \mathbf{\Omega}' \mathbf{A}'_2) \\ = & \text{trace}(\mathbf{A}_2 \mathbf{o}_1 \mathbf{o}'_1 \mathbf{B}'_2 \mathbf{K} \mathbf{K} \mathbf{B}_2 \mathbf{o}_1 \mathbf{o}'_1 \mathbf{A}'_2) \\ & + \text{trace}(\mathbf{A}_2 \mathbf{o}_2 \mathbf{o}'_2 \mathbf{B}'_2 \mathbf{K} \mathbf{K} \mathbf{B}_2 \mathbf{o}_2 \mathbf{o}'_2 \mathbf{A}'_2) \\ & + \text{trace}(\mathbf{A}_2 \mathbf{o}_1 \mathbf{o}'_1 \mathbf{B}'_2 \mathbf{K} \mathbf{K} \mathbf{B}_2 \mathbf{o}_2 \mathbf{o}'_2 \mathbf{A}'_2) \\ & + \text{trace}(\mathbf{A}_2 \mathbf{o}_2 \mathbf{o}'_2 \mathbf{B}'_2 \mathbf{K} \mathbf{K} \mathbf{B}_2 \mathbf{o}_1 \mathbf{o}'_1 \mathbf{A}'_2) \end{aligned} \quad [\text{A5}]$$

The first two terms on the RHS of [A5] correspond to the SSG for the first two rotated axes. The last two terms vanish when $z = 0$, because in this case, $\mathbf{A}'_2 \mathbf{A}_2 = \mathbf{I}$ and hence

$$\begin{aligned} & \text{trace}(\mathbf{A}_2 \mathbf{o}_2 \mathbf{o}'_2 \mathbf{B}'_2 \mathbf{K} \mathbf{K} \mathbf{B}_2 \mathbf{o}_1 \mathbf{o}'_1 \mathbf{A}'_2) \\ = & \text{trace}(\mathbf{o}'_2 \mathbf{B}'_2 \mathbf{K} \mathbf{K} \mathbf{B}_2 \mathbf{o}_1 \mathbf{o}'_1 \mathbf{A}'_2 \mathbf{A}_2 \mathbf{o}_2) \\ = & \text{trace}(\mathbf{o}'_2 \mathbf{B}'_2 \mathbf{K} \mathbf{K} \mathbf{B}_2 \mathbf{o}_1 \mathbf{o}'_1 \mathbf{o}_2) = 0 \end{aligned} \quad [\text{A6}]$$

and

$$\text{SSG} = \text{trace}(\mathbf{o}'_1 \mathbf{B}'_2 \mathbf{K} \mathbf{K} \mathbf{B}_2 \mathbf{o}_1) + \text{trace}(\mathbf{o}'_2 \mathbf{B}'_2 \mathbf{K} \mathbf{K} \mathbf{B}_2 \mathbf{o}_2) \quad [\text{A7}]$$

This shows that the contribution of an axis to SSG can be identified when $z = 0$, which corresponds to environment-focused partitioning SVP = 2 in Yan et al. (2007). When $z \neq 0$, this result will not generally hold (a special case, where identifiability holds, occurs for the rotation to AEC; see below). The same kind of derivation can be made for SSGE by replacing \mathbf{K} with \mathbf{P} .

Rotation to AEC and AEC \perp

We next turn to the special case where the coordinate system is rotated such that the first axis coincides with AEC. First note that $\mathbf{K} \mathbf{B}_2 = \mathbf{1} \mathbf{1}' \mathbf{B}_2 = \mathbf{1} \mathbf{c}'$, where $\mathbf{c}' = \mathbf{1}' \mathbf{B}_2$ is the vector of AEC coordinates. Thus, SSG can be written as

$$q^{-2} \text{trace}(\mathbf{A}_2 \mathbf{c} \mathbf{1}' \mathbf{1} \mathbf{c}' \mathbf{A}'_2) = q^{-1} \text{trace}(\mathbf{A}_2 \mathbf{c} \mathbf{c}' \mathbf{A}'_2) \quad [\text{A8}]$$

This is proportional to the sum of squares of projections of G markers onto the AEC axis, given by the inner product of \mathbf{A}_2 (G markers) and \mathbf{c} (AEC). Note that this result holds regardless of the choice of z , so the projection onto AEC recovers all G in the GGE biplot, regardless of the partition of singular values among environments and genotypes (choice of z).

The same result for SSG is obtained from the general results for rotated axes in the previous section on orthogonally rotated axes. The cosine and sine of the angle between the first principal component and AEC are

$$\begin{aligned} \cos(\varphi) &= \frac{(1 \ 0) \mathbf{B}'_2 \mathbf{1}}{\sqrt{|\mathbf{B}'_2 \mathbf{1}|}} = \frac{\mathbf{b}'_1 \mathbf{1}}{\sqrt{|\mathbf{B}'_2 \mathbf{1}|}} \\ \text{and } \cos(90 - \varphi) = \sin(\varphi) &= \frac{(0 \ 1) \mathbf{B}'_2 \mathbf{1}}{\sqrt{|\mathbf{B}'_2 \mathbf{1}|}} = \frac{\mathbf{b}'_2 \mathbf{1}}{\sqrt{|\mathbf{B}'_2 \mathbf{1}|}} \end{aligned} \quad [\text{A9}]$$

Thus, the rotation matrix that rotates G scores \mathbf{A}_2 and E scores \mathbf{B}_2 from the original coordinates to those of AEC and AEC \perp , is given by

$$\begin{aligned} \mathbf{\Omega} &= \begin{pmatrix} \cos(\varphi) & -\sin(\varphi) \\ \sin(\varphi) & \cos(\varphi) \end{pmatrix} \\ &= \frac{1}{\sqrt{|\mathbf{B}'_2 \mathbf{1}|}} \begin{pmatrix} \mathbf{b}'_1 \mathbf{1} & -\mathbf{b}'_2 \mathbf{1} \\ \mathbf{b}'_2 \mathbf{1} & \mathbf{b}'_1 \mathbf{1} \end{pmatrix} = (\mathbf{o}_1 \ \mathbf{o}_2) \end{aligned} \quad [\text{A10}]$$

With this particular rotation, we have (for any z)

$$\begin{aligned} \mathbf{K} \mathbf{B}_2 \mathbf{o}_2 &= q^{-1} \mathbf{1} \mathbf{1}' \mathbf{B}_2 \mathbf{o}_2 \\ &= \frac{1}{q \sqrt{|\mathbf{B}'_2 \mathbf{1}|}} \mathbf{1} (\mathbf{1}' \mathbf{b}_1 \ \mathbf{1}' \mathbf{b}_2) \begin{pmatrix} -\mathbf{1}' \mathbf{b}_2 \\ \mathbf{1}' \mathbf{b}_1 \end{pmatrix} = \mathbf{0} \end{aligned} \quad [\text{A11}]$$

It follows that SSG in the rotated coordinate system with axes AEC and AEC \perp is

$$\begin{aligned} \text{SSG} &= \text{trace}(\mathbf{A}_2 \mathbf{\Omega} \mathbf{\Omega}' \mathbf{B}'_2 \mathbf{K} \mathbf{K} \mathbf{B}_2 \mathbf{\Omega} \mathbf{\Omega}' \mathbf{A}'_2) \\ &= \text{trace}(\mathbf{A}_2 \mathbf{o}_1 \mathbf{o}'_1 \mathbf{B}'_2 \mathbf{K} \mathbf{K} \mathbf{B}_2 \mathbf{o}_1 \mathbf{o}'_1 \mathbf{A}'_2) \end{aligned} \quad [\text{A12}]$$

because the other three trace terms in Eq. [A5] involve the term $\mathbf{K} \mathbf{B}_2 \mathbf{o}_2$ and thus vanish. This shows that AEC, being associated with \mathbf{o}_1 , captures all SSG in the GGE biplot. The result holds for any z .

The SSGE in the rotated coordinate system with axes AEC and AEC \perp is

$$\begin{aligned} \text{SSGE} &= \text{trace}(\mathbf{A}_2 \mathbf{\Omega} \mathbf{\Omega}' \mathbf{B}'_2 \mathbf{P} \mathbf{P} \mathbf{B}_2 \mathbf{\Omega} \mathbf{\Omega}' \mathbf{A}'_2) \\ &= \text{trace}(\mathbf{A}_2 \mathbf{o}_1 \mathbf{o}'_1 \mathbf{B}'_2 \mathbf{P} \mathbf{P} \mathbf{B}_2 \mathbf{o}_1 \mathbf{o}'_1 \mathbf{A}'_2) \\ &+ \text{trace}(\mathbf{A}_2 \mathbf{o}_2 \mathbf{o}'_2 \mathbf{B}'_2 \mathbf{P} \mathbf{P} \mathbf{B}_2 \mathbf{o}_2 \mathbf{o}'_2 \mathbf{A}'_2) \\ &+ \text{trace}(\mathbf{A}_2 \mathbf{o}_1 \mathbf{o}'_1 \mathbf{B}'_2 \mathbf{P} \mathbf{P} \mathbf{B}_2 \mathbf{o}_2 \mathbf{o}'_2 \mathbf{A}'_2) \\ &+ \text{trace}(\mathbf{A}_2 \mathbf{o}_2 \mathbf{o}'_2 \mathbf{B}'_2 \mathbf{P} \mathbf{P} \mathbf{B}_2 \mathbf{o}_1 \mathbf{o}'_1 \mathbf{A}'_2) \end{aligned} \quad [\text{A13}]$$

The last two terms on the RHS of [A13] vanish when $z = 1$. This is now shown for the third term, which can be rewritten as

$$\begin{aligned} & \text{trace}(\mathbf{A}_2 \mathbf{o}_1 \mathbf{o}_1' \mathbf{B}_2' \mathbf{P} \mathbf{P} \mathbf{B}_2 \mathbf{o}_2 \mathbf{o}_2' \mathbf{A}_2') \\ &= \text{trace}(\mathbf{A}_2 \mathbf{o}_1 \mathbf{o}_1' \mathbf{B}_2' \mathbf{P} \mathbf{B}_2 \mathbf{o}_2 \mathbf{o}_2' \mathbf{A}_2') \\ &= \text{trace}(\mathbf{A}_2 \mathbf{o}_1 \mathbf{o}_1' \mathbf{B}_2' \mathbf{B}_2 \mathbf{o}_2 \mathbf{o}_2' \mathbf{A}_2') \\ & - \text{trace}(\mathbf{A}_2 \mathbf{o}_1 \mathbf{o}_1' \mathbf{B}_2' \mathbf{K} \mathbf{B}_2 \mathbf{o}_2 \mathbf{o}_2' \mathbf{A}_2') \end{aligned} \quad [\text{A14}]$$

The last term in [A14] has already been shown to vanish for any z . It is readily verified that

$$\mathbf{M} = \mathbf{o}_1 \mathbf{o}_1' \mathbf{B}_2' \mathbf{B}_2 \mathbf{o}_2 \mathbf{o}_2' = \frac{\lambda_1^{1-z} - \lambda_2^{1-z}}{|\mathbf{B}_2' \mathbf{1}|} \begin{pmatrix} d^2 e^2 & -d^3 e \\ d^3 e & -d^2 e^2 \end{pmatrix} \quad [\text{A15}]$$

where $d = \mathbf{b}_1' \mathbf{1}$ and $e = \mathbf{b}_2' \mathbf{1}$. We have $\mathbf{M} = \mathbf{0}$ when $z = 1$ (SVP = 1 in Yan et al., 2007), whence $\text{trace}(\mathbf{A}_2 \mathbf{o}_1 \mathbf{o}_1' \mathbf{B}_2' \mathbf{B}_2 \mathbf{o}_2 \mathbf{o}_2' \mathbf{A}_2') = \text{trace}(\mathbf{A}_2 \mathbf{M} \mathbf{A}_2') = 0$. Thus, the third term on the RHS of [A14] vanishes. The same can be shown for the fourth term on the RHS of [A13]. Thus, we have

$$\begin{aligned} \text{SSGE} &= \text{trace}(\mathbf{A}_2 \mathbf{o}_1 \mathbf{o}_1' \mathbf{B}_2' \mathbf{P} \mathbf{P} \mathbf{B}_2 \mathbf{o}_1 \mathbf{o}_1' \mathbf{A}_2') \\ & + \text{trace}(\mathbf{A}_2 \mathbf{o}_2 \mathbf{o}_2' \mathbf{B}_2' \mathbf{P} \mathbf{P} \mathbf{B}_2 \mathbf{o}_2 \mathbf{o}_2' \mathbf{A}_2') \end{aligned} \quad [\text{A16}]$$

This proves that AEC and AEC \perp allow identification of SSGE also when $z = 1$. It is also evident that SSGE is always split between both axes; that is, SSGE is never entirely captured by AEC $\theta\perp$ alone.

Does AEC \perp Maximize the Contribution to SSGE?

Contributions to SSGE are identifiable for arbitrary rotations only when $z = 0$, so we consider only this case. The rotation \mathbf{o}_1 that maximizes the contribution of the second rotated axis to SSGE, $\text{trace}(\mathbf{o}_2' \mathbf{B}_2' \mathbf{P} \mathbf{P} \mathbf{B}_2 \mathbf{o}_2)$, must minimize the contribution to the first axis, $\text{trace}(\mathbf{o}_1' \mathbf{B}_2' \mathbf{P} \mathbf{P} \mathbf{B}_2 \mathbf{o}_1)$. Defining $\mathbf{o}_1 = [\pm\theta, \sqrt{(1-\theta^2)}]$ it can be shown that computation of the first derivative of $\text{trace}(\mathbf{o}_1' \mathbf{B}_2' \mathbf{P} \mathbf{P} \mathbf{B}_2 \mathbf{o}_1)$ with respect to θ and equating to zero yields a quadratic polynomial in θ^2 of the form

$$\begin{aligned} & \theta^4 [4h_{12}^2 + (h_{22} - h_{11})^2] \\ & + \theta^2 [-4h_{12}^2 - (h_{22} - h_{11})^2] + h_{12}^2 = 0 \end{aligned} \quad [\text{A17}]$$

where

$$\mathbf{H} = \begin{pmatrix} h_{11} & h_{12} \\ h_{12} & h_{22} \end{pmatrix} = \mathbf{B}_2' \mathbf{P} \mathbf{P} \mathbf{B}_2 \quad [\text{A18}]$$

Thus, the rotation minimizing $\text{trace}(\mathbf{o}_1' \mathbf{B}_2' \mathbf{P} \mathbf{P} \mathbf{B}_2 \mathbf{o}_1)$ is obtained as one of the two solutions for θ^2 of Eq. [A17]. Obviously, neither of these two solutions equals the AEC.

Acknowledgments

We appreciate helpful suggestions on earlier versions of this paper from Scott Ebdon, Cajo ter Braak, and two anonymous reviewers. We also thank Martin Wells for helpful discussions about identifiability in the GGE model.

References

- Annicchiarico, P. 1992. Cultivar adaptation and recommendation from a set of alfalfa trials in northern Italy. *J. Genet. Breed.* 46:269–278.
- Annicchiarico, P. 1997. 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. 2002. Genotype \times environment interactions: Challenges and opportunities for plant breeding and cultivar recommendations. p. 85–86. Available at <http://www.iao.florence.it/training/adaptation/references.php> (verified 27 Feb. 2008). FAO Plant Production and Protection Paper No. 174. FAO, Rome.
- Annicchiarico, P. 2007a. Lucerne shoot and root traits associated with adaptation to favorable or drought-stress environments and to contrasting soil types. *Field Crops Res.* 102:51–59.
- Annicchiarico, P. 2007b. Wide- versus specific-adaptation strategy for lucerne breeding in northern Italy. *Theor. Appl. Genet.* 114:647–657.
- Annicchiarico, P., F. Bellah, and T. Chiari. 2005. Defining subregions and estimating benefits for a specific-adaptation strategy by breeding programs: A case study. *Crop Sci.* 45:1741–1749.
- Annicchiarico, P., F. Bellah, and T. Chiari. 2006. Repeatable genotype \times location interaction and its exploitation by conventional and GIS-based cultivar recommendation for durum wheat in Algeria. *Eur. J. Agron.* 24:70–81.
- Annicchiarico, P., and E. Piano. 2005. Use of artificial environments to reproduce and exploit genotype \times location interaction for lucerne in northern Italy. *Theor. Appl. Genet.* 110:219–227.
- Atlin, G.N., R.J. Baker, K.B. McRae, and X. Lu. 2000. Selection response in subdivided target regions. *Crop Sci.* 40:7–13.
- Cho, Y.-G., H.-J. Kang, J.-S. Lee, Y.-T. Lee, S.-J. Lim, H. Gauch, M.-Y. Eun, and S.R. McCouch. 2007. Identification of quantitative trait loci in rice for yield, yield components, and agronomic traits across years and locations. *Crop Sci.* 47:2403–2417.
- Cooper, M., I.H. DeLacy, and K.E. Basford. 1996. Relationships among analytical methods used to analyse genotypic adaptation in multi-environment trials. p. 193–224. *In* M. Cooper and G.L. Hammer (ed.) *Plant adaptation and crop improvement*. CAB Int., Wallingford, UK.
- 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.
- Cullis, B.R., A.B. Smith, and N.E. Coombes. 2006. On the design of early generation variety trials with correlated data. *J. Agric. Biol. Environ. Stat.* 11:381–393.
- Culman, S.W., J.M. Duxbury, J.G. Lauren, and J.E. Thies. 2006. Microbial community responses to soil solarization in Nepal's rice-wheat cropping system. *Soil Biol. Biochem.* 38:3359–3371.

- Digby, P.G.N., and R.A. Kempton. 1987. Multivariate analysis of ecological communities. Chapman and Hall, New York.
- Ebdon, J.S., and H.G. Gauch. 2002a. Additive main effect and multiplicative interaction analysis of national turfgrass performance trials: I. Interpretation of genotype \times 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 recommendation. *Crop Sci.* 42:497–506.
- Emebiri, L.C., and D.B. Moody. 2006. Heritable basis for some genotype-environment stability statistics: Inferences from QTL analysis of heading date in two-rowed barley. *Field Crops Res.* 96:243–251.
- Eskridge, K.M. 1990. Selection of stable cultivars using a safety-first rule. *Crop Sci.* 30:369–374.
- Federer, W.T., and R.D. Wolfinger. 1998. SAS code for recovering intereffect information in experiments with incomplete block and lattice rectangle designs. *Agron. J.* 90:545–551.
- Finlay, K.W., and G.N. Wilkinson. 1963. The analysis of adaptation in a plant-breeding programme. *Aust. J. Agric. Res.* 14:742–754.
- Fisher, R.A., and W.A. Mackenzie. 1923. Studies in crop variation: II. The manurial response of different potato varieties. *J. Agric. Sci.* 23:311–320.
- 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. (Chinese ed., China National Rice Research Inst., Hangzhou, China, 2001.)
- 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, UK. (Chinese ed., Tsinghua Univ. Press, Beijing, 2004.)
- Gauch, H.G. 2006a. Statistical analysis of yield trials by AMMI and GGE. *Crop Sci.* 46:1488–1500.
- Gauch, H.G. 2006b. Winning the accuracy game. *Am. Sci.* 94:133–141.
- Gauch, H.G. 2007. MATMODEL version 3.0: Open source software for AMMI and related analyses. Available at <http://www.css.cornell.edu/staff/gauch> (verified 27 Feb. 2008). Crop and Soil Sciences, Cornell Univ., Ithaca, NY.
- Gauch, H.G., and R.W. Zobel. 1996. 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.
- Gilmour, A.R., B.R. Cullis, and A.P. Verbyla. 1997. Accounting for natural and extraneous variation in the analysis of field experiments. *J. Agric. Biol. Environ. Stat.* 2:269–293.
- Gower, J.C., and D.J. Hand. 1996. Biplots. Chapman and Hall, London.
- Harville, D.A. 1997. Matrix algebra from a statistician's perspective. Springer, New York.
- IRRI. 2008. CropStat for Windows, version 5. Available at <http://www.irri.org/science/software/irristat.asp> (verified 27 Feb. 2008). Biometrics and Bioinformatics Unit, Int. Rice Research Inst., Los Baños, Philippines.
- John, J.A., and E.R. Williams. 1995. Cyclic and computer generated designs. 2nd ed. Chapman and Hall, London.
- Jolliffe, I.T. 1986. Principal component analysis. Springer, New York.
- Kempton, R.A. 1984. The use of biplots in interpreting variety by environment interactions. *J. Agric. Sci.* 103:123–135.
- Kondo, M., P.P. Pablico, D.V. Aragonese, R. Asbisit, J. Abe, S. Morita, and B. Courtois. 2003. Genotypic and environmental variations in root morphology in rice genotypes under upland field conditions. *Plant Soil* 255:189–200.
- Laffont, J.-L., M. Hanafi, and K. Wright. 2007. Numerical and graphical measures to facilitate the interpretation of GGE biplots. *Crop Sci.* 47:990–996.
- Laurentin, H., D. Montilla, and V. Garcia. 2007. Interpreting genotype \times environment interaction in sesame (*Sesamum indicum* L.). *J. Agric. Sci.* 145:263–271.
- Ludlow, M.M., and R.C. Muchow. 1990. A critical evaluation of traits for improving crop yields in water-limited environments. *Adv. Agron.* 43:107–153.
- MacKay, D.J.C. 1992. Bayesian interpolation. *Neural Comput.* 4:415–447.
- Nachit, M.M., M.E. Sorrells, R.W. Zobel, H.G. Gauch, R.A. Fischer, and W.R. Coffman. 1992. Association of environmental variables with sites' mean grain yield and components of genotype-environment interaction in durum wheat. *J. Genet. Breed.* 46:369–372.
- Onofri, A., and E. Ciricifolo. 2007. Using R to perform the AMMI analysis on agriculture variety trials. Available at http://cran.r-project.org/doc/Rnews/Rnews_2007-1.pdf (verified 27 Feb. 2008). *R News* 7(1):14–19.
- Ortiz, R., W.W. Wagoire, J. Hill, and S. Chandra. 2001. Heritability of and correlations among genotype-environment stability statistics for grain yield in bread wheat. *Theor. Appl. Genet.* 103:469–474.
- Piepho, H.-P. 1996. Analysis of genotype-by-environment interaction and phenotypic stability. p. 151–174. *In* M.S. Kang and H.G. Gauch (ed.) *Genotype-by-environment interaction*. CRC Press, Boca Raton, FL.
- Piepho, H.-P. 1998a. Empirical best linear unbiased prediction in cultivar trials using factor analytic variance-covariance structures. *Theor. Appl. Genet.* 97:195–201.
- Piepho, H.P. 1998b. Methods for comparing the yield stability of cropping systems: A review. *J. Agron. Crop Sci.* 180:193–213.
- Piepho, H.-P., and J. Möhring. 2005. Best linear unbiased prediction of cultivar effects for subdivided target regions. *Crop Sci.* 45:1151–1159.
- Ribaut, J.-M., Y. Fracheboud, P. Monneveux, M. Banzinger, M. Vargas, and C. Jiang. 2007. Quantitative trait loci for yield and correlated traits under high and low nitrogen conditions in tropical maize. *Mol. Breed.* 20:15–29.
- Romagosa, I., S.E. Ullrich, F. Han, and P.M. Hayes. 1996. Use of the additive main effects and multiplicative interaction model in QTL mapping for adaptation in barley. *Theor. Appl. Genet.* 93:30–37.
- Romagosa, I., H. Feng, S.E. Ullrich, P.M. Hayes, and D.M. Wesenberg. 1999. Verification of yield QTL through realized molecular marker-assisted selection responses in a barley cross. *Mol. Breed.* 5:143–152.
- SAS Institute. 1999. SAS/GRAPH version 8. SAS Inst., Cary, NC.
- Smith, M.F., and H.G. Gauch. 1992. Effects of noise on AMMI and hierarchical classification analyses. *S. Afr. Statist. J.* 26:121–142.

- ter Braak, C.J.F. 1986. Canonical correspondence analysis: A new eigenvector technique for multivariate direct gradient analysis. *Ecology* 67:1167–1179.
- Thies, J.E. 2007. Soil microbial community analysis using terminal restriction fragment length polymorphisms. *Soil Sci. Soc. Am. J.* 71:579–591.
- van Eeuwijk, F.A., J.B. Denis, and M.S. Kang. 1996. Incorporating additional information on genotypes and environments in models for two-way genotype by environment tables. p. 15–50. *In* M.S. Kang and H.G. Gauch (ed.) *Genotype-by-environment interaction*. CRC Press, Boca Raton, FL.
- Williams, E.R., J.A. John, and D. Whitaker. 2006. Construction of resolvable spatial row-column designs. *Biometrics* 62:103–108.
- Worku, M., M. Bänziger, G.S. Erley, D. Friesen, A.O. Diallo, and W.J. Horst. 2007. Nitrogen uptake and utilization in contrasting nitrogen efficient tropical maize hybrids. *Crop Sci.* 47:519–528.
- Wricke, G. 1962. Über eine Methode zur Erfassung der ökologischen Streubreite. *Z. Pflanzenzücht.* 47:92–96.
- Wright, G.C., and N.C. Rachaputi. 2004. Drought and drought resistance. p. 386–390. *In* R.M. Goodman (ed.) *Encyclopedia of plant and crop science*. Marcel Dekker, New York.
- Yan, W., L.A. Hunt, Q. Sheng, and Z. Szlavncs. 2000. Cultivar evaluation and mega-environment investigation based on GGE biplot. *Crop Sci.* 40:596–605.
- 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., 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.
- Zobel, R.W., M.J. Wright, and H.G. Gauch. 1988. Statistical analysis of a yield trial. *Agron. J.* 80:388–393.