Available online at www.isas.org.in/jisas

JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 65(2) 2011 181-193

GGE Biplot vs. AMMI Graphs for Genotype-by-Environment Data Analysis

Weikai Yan

Eastern Cereal & Oilseed Research Centre, Agriculture and Agri-Food Canada, 960 Carling Ave., Neatby Building, Ottawa, ON, Canada, K1A 0C6.

Received 27 October 2010; Revised 07 November 2010; Accepted 29 November 2010

SUMMARY

Due to the ever-presence of genotype-by-environment interaction (GE), multi-environmental trials (MET) are essential for effective breeding line selection and cultivar recommendation. AMMI (Additive Main effect and Multiplicative Interaction) analysis and GGE (Genotypic main effect plus genotype-by-environment interaction) biplot analysis are two popular graphical analysis systems for MET data analysis. This paper introduces and compares the AMMI graphs and the GGE biplots for three major aspects of MET data analysis: mega-environment delineation, genotype evaluation, and test environment evaluation. The conclusions are: 1) when used properly, both systems are capable of mega-environments delineation and genotype evaluation; 2) the GGE biplot is also effective in test environment evaluation; 3) the GGE biplots are simpler to construct than the AMMI graphs; while different views of the same GGE biplot can be used to address all three aspects of MET data analysis, a different graph has to be constructed in AMMI analysis to address each aspect; 4) the GGE biplots are more informative than the AMMI graphs because of its inner-product property, whereby information on the performance of each genotype in each environment is preserved. Therefore, the GGE biplot graphs are highly preferable over AMMI graphs in MET data analysis.

Keywords : Additive Main effect and Multiplicative Interaction, Genotype main effect, Genotype-by-environment interaction, Interaction principal component, Multi-environment trials, Principal component, Singular-value decomposition, Singular-value partitioning.

1. THREE OBJECTIVES OF MET DATA ANALYSIS

Multi-location trials, or more generally, multi-environment trials (MET), are conducted routinely to generate essential information for breeding line selection, new cultivar release, and cultivar recommendation. MET are essential because of the existence of genotype-by-environment interactions (GE), which complicates genotype evaluation/selection, and for this reason, analysis of genotype-by-environment data from MET trials has been an important component of plant breeding and cultivar recommendation. Data from MET include genotype-by-environment data for multiple traits, although yield is almost universally the most important trait. In the paper, I use 'MET data' and 'genotype-by-environment data' for a trait interchangeably. Within a single year, the term 'multi-environments' is used interchangeably with the term multi-locations.

Early analysis of MET data treated GE as a confounding factor, and the purpose was to identify "stable" cultivars less affected by GE. Many stability indices have been developed to help selecting for stable genotypes (Lin and Binns 1994) thus avoiding GE. Later, it was recognized that some GE can be repeatable across years and therefore are exploitable (Cooper and DeLacy 1994), and a graphical method was developed to delineate mega-environment based on genotype main

E-mail address : yanw@agr.gc.ca

effect (G) and GE to exploit GE (Gauch and Zobel 1997). Further, it was realized that MET data are valuable not only in selecting superior genotypes but also in identifying test environments that are more effective for genotype evaluation (Yan 2001, Yan and Kang 2003). Therefore, a full analysis of MET data should achieve three major objectives: 1) genotype evaluation, 2) test-environment evaluation, and 3) mega-environment delineation (Yan and Hunt 2003, Yan and Tinker 2006, Yan *et al.* 2007). Logically, mega-environment analysis should precede genotype evaluation (including all types of "stability analyses") and test-environment evaluation because the latter two are meaningful only when conducted within mega-environments (Yan *et al.* 2007).

Graphical and visual analysis is always desirable, if not essential, particularly when dealing with complex data structures and patterns. A number of graphical methods have been developed to visually analyze MET data. Among those, the Additive Main effect and Multiplicative Interaction effect (AMMI) graphs (Gauch and Zobel 1997, Gauch *et al.* 2008) and the GGE biplots (Yan 2001, Yan and Kang 2003, Yan and Tinker 2006) are most popular, and there has been a debate on whether the AMMI graphs or the GGE biplots are more appropriate and effective in MET data analysis (Gauch 2006, Yan *et al.* 2007, Gauch *et al.* 2008). The objective of this paper is to compare GGE biplot analysis and AMMI analysis based on on their ability to achieve three aspects of MET data analysis, namely, mega-environment analysis, genotype evaluation, and test-environment evaluation.

2. BIPLOT PRESENTATION OF A TWO-WAY TABLE

The biplot method was originally developed by Gabriel (1971) to graphically display the results from singular value decomposition (SVD) of a genotype by environment two-way table. Any two-way data matrix **Z**, with elements of $[z_{ij}]$, with $i = 1, \ldots g$ rows (genotypes) and $j = 1,\ldots e$ columns (environments), can be decomposed via SVD into t principal components (PC).

$$z_{ij} = \sum_{k=1}^{t} \lambda_k \alpha_{ik} \gamma_{jk} + \varepsilon_{ij}, \qquad (1)$$

with $t \le \min(e, g-1)$. Each PC is composed of an array of genotypic scores α_{ik}, an array of environmental scores γ_{jk}, and a singular value λ_k, the square of which, λ_k^2, is the sum square explained by the kth PC. ε_{ij} is the residue for genotype i in environment j that is not explained by the model. The model is subject to the constraint $\lambda_1 \ge \lambda_2 \ge \ldots \ge \lambda_t \ge 0$ and to orthonormality on the α_{ik} scores, i.e., $\sum_{i=1}^{g} \alpha_{ik}\alpha_{ik'} = 1$ if $k = k'$ and $\sum_{i=1}^{g} \alpha_{ik}\alpha_{ik'} = 0$ if $k \ne k'$, with similar constraints on the γ_{jk}.

When the rank-t matrix **Z** can be sufficiently approximated by a rank-2 matrix, i.e.,

$$z_{ij} = \lambda_1 \alpha_{i1} \gamma_{j1} + \lambda_2 \alpha_{i2} \gamma_{j2} + \varepsilon_{ij}, \qquad (2)$$

it can be graphically presented in a 2-D biplot after an appropriate singular value partitioning

$$z_{ij} = (\lambda_1^f \alpha_{i1})(\lambda_1^{1-f} \gamma_{j1}) + (\lambda_2^f \alpha_{i2})(\lambda_2^{1-f} \gamma_{j2}) + \varepsilon_{ij} \qquad (3)$$

where $f = [0, 0.5, 1]$ is the singular value partitioning (SVP) factor.

The biplot is constructed by plotting $\lambda_1^f \alpha_{i1}$ as abscissa and $\lambda_2^f \alpha_{i2}$ as ordinate for each genotype, and at the same time plotting $\lambda_1^{1-f} \gamma_{j1}$ as abscissa and $\lambda_2^{1-f} \gamma_{j2}$ as ordinate for each environment. The exponent f is used to rescale the row and column scores to enhance visual interpretation of the biplot for a particular purpose. In the context of MET data, singular values are allocated entirely to cultivar (row) scores if $f = 1$ [this is "cultivar-focused singular value partitioning" or SVP = 1 (Yan 2002)], or entirely to environment (column) scores if $f = 0$ ("environment-focused singular value partitioning" or SVD = 2); and $f = 0.5$ will allocate the square roots of the λ_k values to both cultivar scores and environment scores ("symmetric singular value portioning" or SVP = 3). In GGE biplot analysis, the genotype-focused and the environment-focused SVP are used for genotype evaluation and test environment, respectively. An important property of the biplot is that the rank-2 approximation of any element in the original matrix **Z** can be visually estimated by the inner product of the corresponding genotype and environment vectors and the cosine of the angle between them. This is known as the inner-product property of the biplot.

2.1 Constructing a GGE Biplot

For a MET dataset, each value in the table is the average yield (or any other trait) value of a genotype

in an environment (y_{ij}), which is the sum of the grand mean (μ), the environment main effect (E) for the particular environment (μ_j), the genotypic main effect (G) for the particular genotype (μ_i), and the specific interaction (GE) between the genotype and the environment (ϕ_{ij}), ignoring any random errors

$$y_{ij} = \mu + \mu_j + \mu_i + \phi_{ij} \quad (4)$$

Since only G and GE interaction are pertinent to genotype evaluation, test environment evaluation, and mega-environment delineation, the environment main effect E and the grand mean should be removed from each element to only keep G and GE in the two-way table

$$y_{ij} - \mu - \mu_j = \mu_i + \phi_{ij}. \quad (5)$$

This newly derived two-way data, i.e., environment-centered MET data, after appropriate data scaling, are then subjected to SVD and biplot analysis as discussed above

$$z_{ij} = (\mu_i + \phi_{ij})/s_j \ . \quad (6)$$

Therefore, the resulting biplot contain G and GE and nothing else, and is therefore referred to as a "GGE biplot" (Yan *et al.* 2000). In equation (6), s_j is a scaling factor specific to environment *j*. It can be the standard deviation of means (= square root of phenotypic variance), the experimental error, or the inverse of square-root heritability in each environment, leading to different types of GGE biplots (Yan and Holland 2010). Alternatively, it can be "1" for all environments, leading to an unscaled GGE biplot. This is the type of GGE biplot that is used in this paper for demonstration purposes.

2.2 The Adequacy of a 2-D GGE Biplot

All biplot analysis has an implicit assumption; it is that a 2-D biplot adequately approximates the two-way table under study. The adequacy is usually judged by the goodness of fit of the 2-D GGE biplot, i.e., the percentage of total variation of the two-way table that is explained by the first two PCs. If the goodness of fit is high, say, greater than 70%, then the biplot is a good approximation of the two-way table. However, there is not an objective criterion on what is a good fit.

Yan and Tinker (2006) proposed an "information ratio" to assess the adequacy of a biplot in displaying the patterns of a two-way table. Assume that the two-way table in question has *g* genotypes and *e* environments. The maximum number of PCs that are required to fully represent the two-way table is k = min(e, g-1). If there are no correlations among the environments, all k PCs should be completely independent and the proportion of the total variation explained by each PC should be exactly $1/k$. When some correlations exist among environments, the proportion of variation explained by the first few PCs should be greater than $1/k$, and that explained by other PCs would be less than or equal to $1/k$. An information ratio (IR) can be calculated for each PC, which is the proportion of total variation explained by each PC multiplied by k. The interpretation is: a PC with an IR >1 contains patterns (associations among environments), a PC with an IR ≈ 1 does not contain patterns, but it may contain some independent information, and a PC with an IR < 1 does not contain any pattern or information. A 2-D biplot adequately represents the patterns in the data if only the first two PCs have an IR > 1. If more PCs have an IR > 1, then the 2-D biplot is not adequate. If only the first PC has an IR >1, then biplot analysis is not needed.

2.3 Two Key Properties of a GGE Biplot

2.3.1 The inner-product property

One unique property of the biplot, compared with all other graphical methods, is its inner-product property. That is, the value of each element in the two-way table can be graphically visualized by the product of the length of its row (genotype) vector (L_i), the length of its column (environment) vector (L_j), and the cosine of the angle (A_{ij}) between the two vectors in the biplot

$$z_{ij} \approx L_i L_j \cos A_{ij} \quad (7)$$

The inner-product property is the basis for the GGE biplot to be used in mega-environment delineation, genotype evaluation, and test environment evaluation.

2.3.2 The cosine-correlation equality

Another key property of a GGE biplot is the correlation-cosine equality

$$r_{j'j} \approx \cos A_{j'j}. \quad (8)$$

That is, when the biplot is based on environment-centered data (i.e., a GGE biplot) and when the environment-focused SVP is used, the cosine of the angle between two environments approximates the

genetic correlation between them. This property allows visualization of similarity and dissimilarity between environments in ranking genotypes.

3. GGE BIPLOTS FOR MET DATA ANALYSIS

In this section, an example will be given on the use of GGE biplot in MET data analysis using the data in Table 1, which are environment-centered yield values of 18 winter wheat genotypes tested at nine locations in Ontario in 1993. The environment-centered data were derived by subtracting the mean yield of each environment from the original yield value of each genotype in that environment (equation (5)).

The GGE biplot based on this dataset is presented in Fig. 1. The abscissa of the biplot represents the PC1 scores and the ordinate the PC2 scores, of the genotypes and the environments. The biplot explained 78% of the total G + GE variation. The biplot is based on environment-focused SVP (SVP = 2). According to the information ratios of the first six PC (Table 2), only the first two PC contain patterns (with an IR > 1). Therefore, the 2-D biplot is considered to have adequately represented the patterns in the data.

The 18 genotypes are labeled as *g*1 to *g*18 and the nine environments as *E*1 to *E*9. The straight line drawn from the biplot origin to the placement of a genotype

Table 1. Environment-centered yield data of 18 winter wheat genotypes (*G*1 through *G*18) tested at nine Ontario locations (*E*1 through *E*9) in 1993. *E*1, *E*5, and *E*7 are locations in eastern Ontario and *E*2, *E*3, *E*4, *E*6, *E*8, and *E*9 are locations in southern Ontario.

Genotype	Environments									
	*E*1	*E*2	*E*3	*E*4	*E*5	*E*6	*E*7	*E*8	*E*9	Mean
*G*1	0.10	−0.29	−0.29	−0.41	0.26	−0.61	0.11	−0.32	−0.23	−0.19
*G*2	0.05	0.33	−0.23	0.01	0.02	0.09	0.72	0.03	0.04	0.12
*G*3	0.31	0.14	−0.04	−0.04	0.39	−0.04	0.49	−0.46	−0.28	0.05
*G*4	0.37	0.31	0.24	0.41	0.54	0.28	−0.01	0.53	0.55	0.36
*G*5	0.03	0.17	0.37	0.35	0.09	0.36	0.91	−0.26	−0.07	0.22
*G*6	0.82	0.04	−0.15	0.28	0.90	−0.02	−0.25	−0.09	−0.12	0.16
*G*7	−0.99	−0.26	−0.40	−0.34	−0.34	−0.79	−0.08	−0.30	−0.87	−0.48
*G*8	0.49	0.23	1.29	0.46	−0.15	0.77	−0.07	0.70	0.67	0.49
*G*9	0.68	0.30	0.37	−0.06	0.28	−0.20	0.74	0.15	−0.04	0.25
*G*10	0.83	0.22	0.46	0.26	0.25	0.29	−0.34	0.09	0.40	0.27
*G*11	−0.07	0.09	−0.38	−0.07	0.46	0.19	0.62	−0.22	0.25	0.10
*G*12	−1.21	−1.40	−0.75	−1.15	−1.45	−0.80	−0.85	−0.29	−0.80	−0.97
*G*13	−0.26	−0.56	−0.84	0.22	−1.13	0.09	−1.64	0.60	−0.01	−0.39
*G*14	−1.02	−0.58	−0.72	−0.71	−1.05	0.03	−0.96	−0.44	−0.34	−0.64
*G*15	0.01	0.26	0.52	0.10	0.51	0.08	−0.31	−0.15	0.03	0.12
*G*16	0.58	0.26	−0.19	0.40	0.38	0.27	0.06	−0.06	0.13	0.20
*G*17	−0.58	0.53	0.24	−0.14	−0.91	0.24	0.08	0.50	0.48	0.05
*G*18	−0.13	0.22	0.47	0.42	0.96	−0.23	0.78	0.00	0.21	0.30
SD†	0.60	0.45	0.53	0.42	0.69	0.40	0.66	0.35	0.40	

SD†: standard deviation of means within environments.

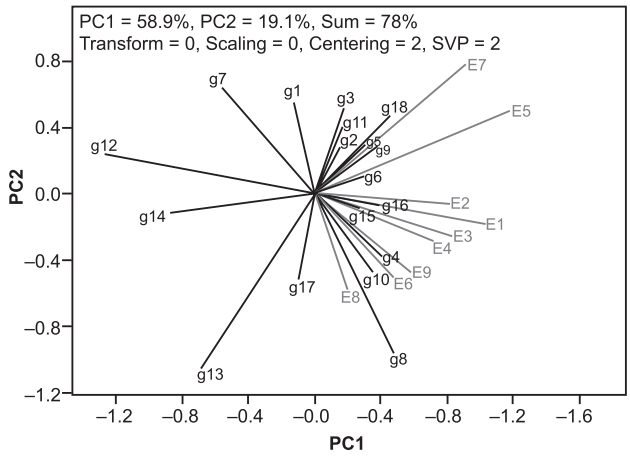

Fig. 1. The inner-product property view of the GGE biplot. This view allows visualization of the relative performance of each genotype in each environment. The biplot explained 78% of the total G + GE variation. It is environment-centered (centering = 2) and un-scaled (scaling =0). It is based on environment-focused singular value partitioning (SVP =2). The genotypes are labeled as *g*1 to *g*18, and the environments as *E*1 to *E*9.

Table 2. Singular value, proportion explained, and information ratio (IR) of the first six principal components (PC)

PC	Singular value	Variation explained (%)	IR
1	5.0	58.9	5.3
2	2.9	19.1	1.7
3	2.1	10.0	0.9
4	1.1	2.9	0.3
5	0.9	1.8	0.2
6	0.3	0.3	0.0

or an environment is called a "vector". From these vectors, the environment-centered yield of each genotype in each environment can be approximately visualized. For example, the genotype *g*12 has obtuse angles with all environments. This means that *g*12 had lower than average yields in all environments (because the cosine of an obtuse angle is smaller than 0). Examining the data in Table 1 confirms this observation. The same statement is true for genotype *g*14. Since *g*12 has a longer vector than *g*14, *g*12 should have lower yields (i.e. more negative values) than *g*14 in all environments; this observation can also be confirmed from Table 1. Deviations from this prediction may occur because the two genotypes do not have exactly the same angles with the environments and because the goodness of fit of the biplot is 78% rather than 100%. As another example, *g*8 has acute angles with all environments except *E*5 and *E*7, suggesting that it yielded higher than average in all environments except *E*5 and *E*7, where it should yield lower than or equal to the environmental means. This again can be confirmed from Table 1.

Owing to the inner-product property, when a MET dataset is sufficiently approximated by a rank-2 matrix, it can be graphically studied in a 2-D GGE biplot for three aspects: 1) differences/similarities among genotypes, 2) relationships among test environments, and 3) specific genotype-by-environment interactions. These correspond to the three objectives of MET data analysis mentioned earlier, namely, genotype evaluation, test environment evaluation, and mega-environment analysis. The GGE biplot analysis system consists of a set of views that were designed to specifically address each of these issues, as discussed below.

3.1 Mega-environment Analysis

A mega-environment is a group of environments or sub-regions in which a single genotype or a group of similar genotypes are specifically adapted and champion in performance (Gauch and Zobel 1997). The purpose of mega-environment analysis is to try to divide a target crop region into meaningful sub-regions so that repeatable GE can be exploited.

When a 2-D GGE biplot is judged as a sufficient approximation of the data, as is in this example, the "which-won-where" view of the GGE biplot (Fig. 2) is an effective tool for visual mega-environment analysis. This view consists of an irregular polygon and a set of straight lines that radiate from the biplot origin to intersect each of the polygon sides at right angles. The vertices of the polygon are the genotype markers located farthest away from the biplot origin in all directions, such that all genotypes are contained within the polygon. A radiate line that perpendicularly intersects a polygon side represents hypothetical environments in which the two cultivars defining that polygon side would perform equally well; the relative ranking of the two cultivars would be reversed in environments on opposite sides of the line (the so-called "crossover GE"). Thereby, the radiate lines divide the biplot into sectors. For each sector, a vertex genotype exists, which is the nominal winner for environments falling in that sector.

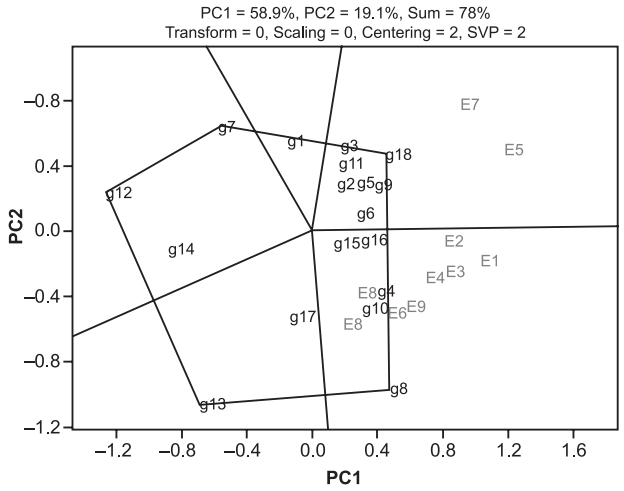

Fig. 2. The which-won-where view of the GGE biplot. This view allows visual grouping of the test environments based on crossover genotype-by-environment interactions among the best genotypes. This is the same GGE biplot presented in Fig. 1 except for the supplementary lines.

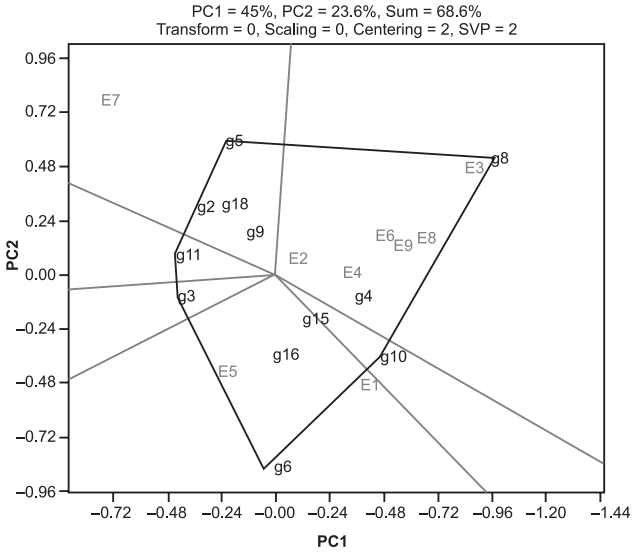

Fig. 3. The which-won-where view of the GGE biplot based on a subset of the genotypes. This view allows visual grouping of the test environments based on crossover genotype-by-environment interactions among the best genotypes. The biplot explained 69% of the total G + GE variation. It is environment-centered (centering = 2) and un-scaled (scaling =0). It is based on environment-focused singular value partitioning (SVP =2).

In Fig. 2, the vertex genotypes that form the polygon are *g*8, *g*18, *g*7, *g*12, and *g*13. The nine environments are cut into two groups by the radiate lines: *E*5 and *E*7 as one group and all other environments as the other. *g*18 is the vertex genotype in sector where *E*5 and *E*7 are placed and is therefore the nominally highest yielding cultivar in these two environments; *g*8 is the vertex genotype in the sector where the other seven environments are placed and therefore the nominal winner in these environments. This crossover GE pattern suggests that the target environments may be divided into two different mega-environments. This pattern appeared to be repeatable across years and the mega-environment delineation is therefore meaningful (Yan *et al.* 2000). *E*1, *E*5, and *E*7 are locations in eastern Ontario and the others are in southern Ontario. Eastern Ontario is characterized by having longer and severe winters than southern Ontario, and therefore requires different winter wheat cultivars for maximum yield. Although belonging to eastern Ontario, *E*1 is different from *E*5 and *E*7 in that it is located by the St. Laurence River and has relatively milder winters.

No environments fell into the sectors of *g*7, *g*12, or *g*13. This means that these genotypes were not the winner in any of the environment; rather, they are likely to be the poorest genotypes in some or all of the environments.

If a 2-D GGE biplot is considered as not adequate in displaying the GE patterns, one option is to construct a GGE biplot based on a subset of the data by removing genotypes that yielded poorly in all or most test environments. This is justifiable from the viewpoint of variety selection as these genotypes will not be selected as superior genotypes (Fig. 3). Deleting low-yielding genotypes lead to reduced proportion of G relative to GE in the new biplot, and therefore greater separation of the test environments. Thus, the nine test environments are separated into three groups in Fig. 3: *E*7 standing alone, with *g*5 as the winner; *E*1 and *E*5 form a group, with *g*6 as the winner; and all other six test environments form a third group, with *g*8 as the winner.

It has to be emphasized that mega-environment delineation must be based on data from multiple years becaue repeatability of a GE pattern is the key for making decisions that have long-term impacts (Yan *et al*. 2011). Appropriate mega-environment analysis should classify the target environment into one of three possible types (Table 3) (Yan *et al.* 2007). Type 1 consists of a single mega-environment with little GE. Theoretically a single test location would suffice to identify the best genotypes for such mega-environments.

Table 3. Three types of target environments based on mega-environment analysis (adapted from Yan *et al.* (2007))

	With Crossover genotype-by-environment interaction	Without crossover genotype-by-environment interaction
Crossover genotype-by-environment interaction repeatable across years	**Type 2**: the target environment consists of multiple mega-environments. Strategy: select specifically adapted genotypes for each mega-environment. A single year multi-location test may be sufficient for making selection decisions.	**Type 1**: the target environment is a single, simple mega-environment. Strategy: test at a single or a few test locations suffices to select for a single best cultivar. A single year test may be sufficient for making selection decisions.
Crossover genotype-by-environment interaction not repeatable across years	**Type 3**: the target environment consists of a single but complex mega-environment. Strategy Type 1: select a set of cultivars (not a single cultivar) for the whole region based on both mean performance and stability; data from multiple years and multiple locations are essential	

Type 2 consists of different mega-environments, which can and should be dealt with individually, whereby the repeatable GE can be converted into productivity by selecting and employing specifically adapted genotypes in each mega-environment. Identifying and exploring such opportunities is a key point in all GE-related analyses. Type 3 consists of a single mega-environment with large but unpredictable GE, which cannot be exploited and must be avoided by selecting widely adapted, high yielding and stable genotypes across years and locations.

3.2 Genotype Evaluation

A superior genotype should have both high mean performance and high stability across a mega-environment. The "Mean vs. Stability" view of the GGE biplot (Fig. 4) is an effective tool for visual evaluation of genotypes on both aspects. In Fig. 4, the small circle represents the "average environment". It is defined by the averaged coordinates of all test environments in the biplot. The straight line with a single arrow passes through the biplot origin and the average environment, and is referred to as the "average environment axis" or AEA. The arrow points to higher mean performance for the genotypes. The line with two arrows passes through the biplot origin and is perpendicular to the AEA. The arrows point to higher performance variability or less

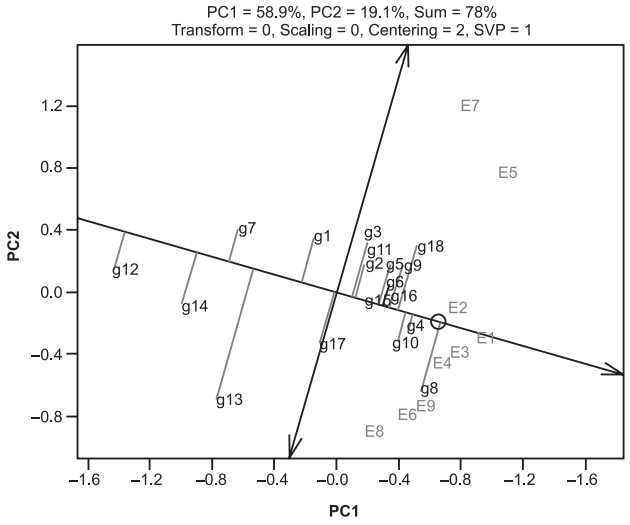

Fig. 4. The mean-vs.-stability view of the GGE biplot. This view allows visual evaluation of genotypes based on their mean performance and stability across environments. It is the same biplot as Fig. 1 and Fig. 2 except that it is based on genotype-focused singular value partitioning (SVP =1).

stability in both directions. This biplot view (or form) is based on genotype-focused SVP, i.e., the singular values are entirely partitioned into the genotypic scores ("SVP = 1") (Yan 2002). Thus, the genotypes are ranked according to their mean yield as follows: $g8 > g4 > g10 > g18 > \ldots > g17 \approx$ grand mean $> g1 > g13 >$

*g*7 > *g*14 > *g*12. The genotype *g*13 was least stable, because it yielded extremely poorly in *E*5 and *E*7 while it performed relatively well in *E*8, *E*6, and *E*9. The genotype *g*8 was not stable either; it yielded only close to the average in *E*5 and *E*7 although it did very well in other environments. Genotype *g*4 yielded the second best and was more stable than *g*8.

The Mean vs. Stability view of the GGE biplot is useful only when the G is sizable. When G is too small relative to GE, this view will not be meaningful. But this does not undermine the usefulness of the GGE biplot. It only reflects the common sense that no generally adapted genotypes can be expected and specifically adapted genotypes must be sought when G is negligibly small.

3.3 Test Environment Evaluation

The purpose of test-environment evaluation is to identify test environments that can be used to effectively select superior genotypes for a mega-environment. An "ideal" test environment should be both discriminating of the genotypes and representative of the target environment. The "Discrimination vs. Representativeness" view of the biplot (Fig. 5) was designed for this purpose.

When the GGE biplot is based on unscaled (un-standardized), environment-centered data ("Scaling = 0"), the length of the vector of an environment is proportional to the standard deviation of cultivar means (SD), which equals the square root phenotypic variance (σ_p) in the test environment, which can be used as a measure of the discriminating power of the environment. Test environments with longer vectors (like *E*5, *E*7, and *E*1) are more discriminating of the genotypes. A test environment with a short vector is less discriminating, meaning that all genotypes tends to perform similarly and little or no information about the genotypic differences can be revealed in such an environment. A short vector could also mean that the environment is not well represented by PC1 and PC2 if the biplot does not adequately display the G+GE of the data.

A second utility of Fig. 5 is to indicate the test-environments' representativeness of the target environment. Since the AEA is the "average-environment axis", test environments that have small angles with AEA, e.g., *E*1, *E*2, *E*3, and *E*4, are more representative of the target environment than those that

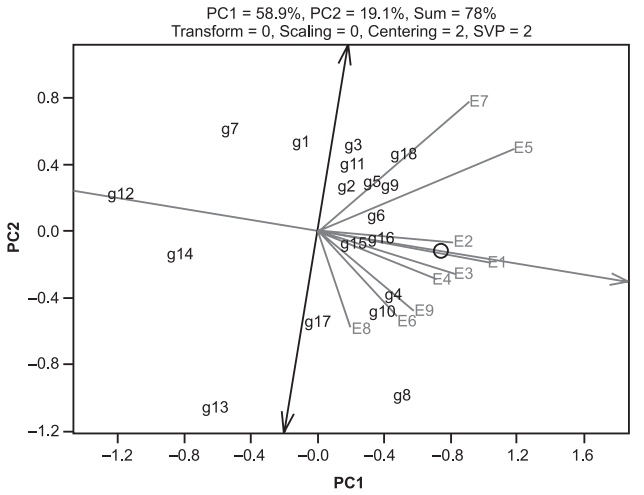

Fig. 5. The discriminating ability vs. representativeness view of the GGE biplot. This view allows visual evaluation of test environments based on their ability to discriminate the genotypes and their representativeness of the target environment. It is the same biplot as Fig. 1 and Fig. 2 except for the complementary lines.

have larger angles with it, e.g., *E*5, *E*7, and E8. This interpretation is based on the cosine-correlation equality property of the GGE biplot (equation (8)). Note that SVP = 2 was used in all biplots presented in this paper except Fig. 4, in which the genotype-focused SVP (SVP=1) was applied for optimal genotype evaluation. The inner-product property (equation (7)) is not affected by the SVP method.

GGE biplot based test environment evaluation can classify a test environment into one of three types (Table 4). Type 1 environments have short vectors and provide little or no information about the genotypic differences and, therefore, should not be used as test environments. Type 2 environments have long vectors and small angles with the AEC abscissa; they are ideal

Table 4. Three types of test environments based on test environment evaluation. (Adopted from Yan *et al.* 2007)

	Discriminating	Non-discriminating
Representative	Type 2: Ideal for selecting superior genotypes.	Type 1: not useful.
Not representative	Type 3: Useful for culling inferior genotypes.	

test environments for selecting superior genotypes. Type 3 environments have long vectors and large angles with the AEC abscissa (e.g., *E*1); they cannot be used in selecting superior genotypes, but are useful in culling unstable genotypes.

Useful test environments should be further examined for their uniqueness. Some environments may never provide unique information because they are always similar to some other environment(s) in ranking the genotypes. When resources are limited (as they always are), some (not all) of these environments can be dropped without losing much information about the genotypes. Testing cost can be reduced and efficiency improved by using a minimum set of test environments. Like mega-environment analysis, identification and removal of non-informative and redundant test *locations* must be based on multi-year data. An example of identifying essential test locations based on biplot analysis is presented in Yan *et al.* (2010) for testing oat breeding lines in eastern Canada.

To summarize, GGE biplot can effectively address the three major objectives of MET data analysis: mega-environment delineation, genotype evaluation, and test environment evaluation. GGE biplot analysis of MET data can help researchers to better understand their target environment, to establish more cost-effective breeding and testing strategies, and to identify superior genotypes that are widely or specifically adapted.

4. AMMI GRAPHS FOR MET DATA ANALYSIS

AMMI analysis first entered in the literature with Gauch (1988) and Zobel *et al.* (1988). In AMMI analysis, E, G, and GE are all of research interest

$$y_{ij} - \mu = \mu_j + \mu_i + \phi_{ij}, \quad (9)$$

with the GE term further examined by subjecting it to SVD

$$z_{ij} = \phi_{ij}. \quad (10)$$

The PCs derived from AMMI analysis contains nothing but GE, and are referred to as interaction PC or IPC. A biplot composed of IPC1 and IPC2 is a GE biplot (as opposed to a GGE biplot, which contains G + GE), sometimes referred to as an AMMI2 biplot in the AMMI analysis literature (Gauch *et al.* 2008). Since genotype evaluation and mega-environment analysis requires joint consideration of G and GE, a GE biplot is not really useful. In AMMI analysis, it is noted that although G and GE are separated, they are putting together again to form a number of AMMI graphs to address genotype evaluation and mega-environment analysis. So the AMMI graphs are also G+GE graphs. Test environment evaluation has not been a research topic in AMMI analysis.

4.1 Mega-environment Analysis

In addressing mega-environment delineation, AMMI analysis utilizes the information G and the first interaction PC (IPC1) if the AMMI1 model is regarded as the best model for the data. When the AMMI2 model is regarded as the best, information IPC2 is also considered.

4.1.1 Under the Ammi1 model

The AMMI1 regression graph (Fig. 6) was designed to address the "which-won-where" pattern in the data of Table 1. In this graph, the abscissa represents the IPC1 scores for the environments and the ordinate represents the "nominal yield" based on mean yield (G) and the IPC1. Each genotype is represented by a straight line defined by that genotype's mean yield as the intercept and the genotype IPC1 score as the regression coefficient against the environmental IPC1 scores. If

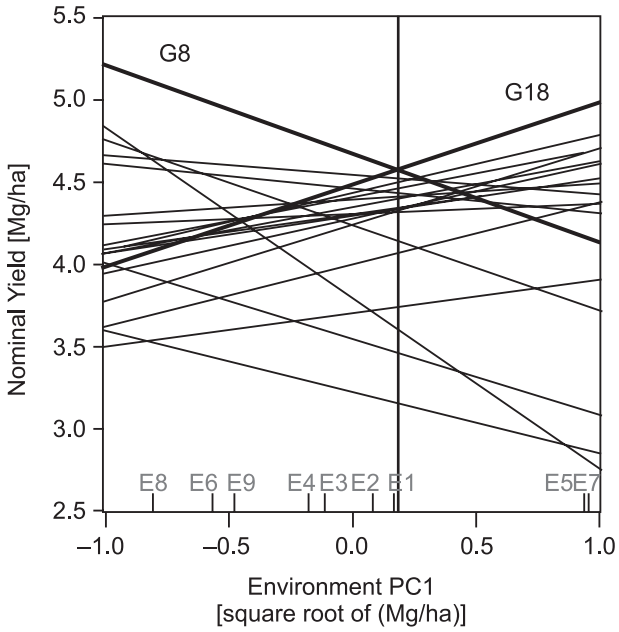

Fig. 6. The AMMI1 regression graph. This graph is used to visualize which genotypes yielded the highest in which environments. The environments are labeled along the abscissa and the genotypes are presented as regression lines to the environmental IPC scores. Adopted from Fig. 3 of Gauch *et al.* (2008)

two straight lines (two genotypes) intersect in the graph, they are said to be involved in a crossover GE. Of the most interesting is the crossover GE between the lines (genotypes) that intersect on the top of the graph. In this example, the genotypes *g*8 and *g*18 intersect on the top, and covers all other genotypes. The vertical line passes through the intersection of the straight lines representing *g*8 and *g*18, it divides the nine environments into two groups: *E*5 + *E*7 vs. the others. The interpretation is clear: *g*18 was the nominal winner in environments *E*5 + *E*7, and *g*8 was the nominal winner in the other seven environments. This result is the same as that in the which-won-where view of the GGE biplot (Fig. 2).

This example supports the statement of Ebdon and Gauch (2002) that mega-environment classification based on this AMMI1 regression graph is virtually the same as that based on the GGE biplot. However, the GGE biplot (Fig. 2) is more preferable to the AMMI1 regression graph (Fig. 6) because it always explains more G+GE. For a rice dataset, the GGE biplot and the AMMI1 graph explained 77.3% and 64.6% of the total G+GE, respectively (Samonte *et al.* 2005). The GGE biplot is also simpler to construct and is a more elegant presentation of the G + GE.

Gauch *et al.* (2008) argued that the geometry of the AMMI1 regression graph is more straightforward to interpret. This may be true for researchers who are not familiar with the biplot theory. However, allowing some experience, the advantages of the GGE biplot will be easily apprehended. First, the which-won-where view is an intrinsic property of the GGE biplot. Once the PC scores are obtained and a GGE biplot is constructed, the only thing the researcher needs to do is to visually draw the polygon and the perpendicular lines. In comparison, constructing the AMMI1 regression graph involves an additional step from the PC scores. Second, in a GGE biplot, each of the environments and each of the genotypes are represented as single points in a 2-D space. In comparison, in the AMMI1 graph the environments are aligned along the abscissa and the genotypes are presented as straight lines. Often the AMMI1 graph is so crowded that only a few genotypes and environments can be clearly labeled. When there are many genotypes and environments, the "which-won-where" patterns can be difficult to visualize, see Fig 2 of Ebdon and Gauch (2002) for an example. Failing to label all genotypes and environments in the graph dramatically reduces the information content and usefulness of the AMMI1 regression graph as a data visualization tool.

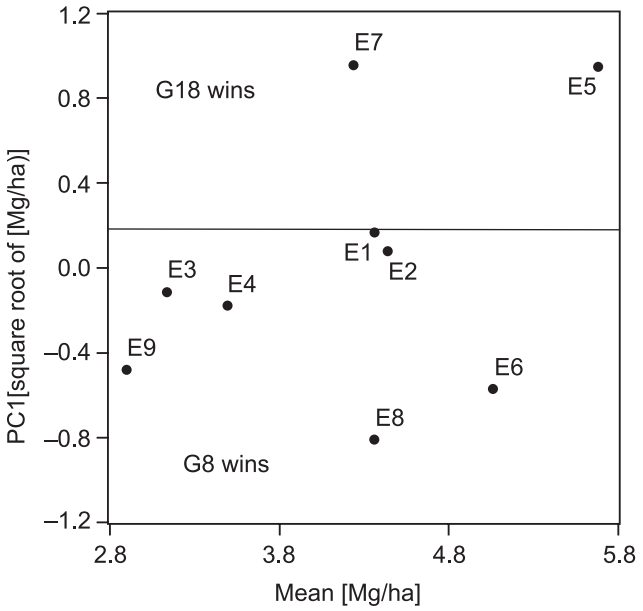

Fig. 7. The AMMI1 mega-environment display. This graph is a display of the which-won-where results identified in Fig. 6. This is not an essential graph because it contains no additional useful information beyond Fig. 6. Adopted from Fig. 2 of Gauch *et al.* (2008).

The AMMI1 mega-environment display (Fig. 7) shows the which-won-where result identified from Fig. 6 except that the information on the mean yield of the environments is also incorporated. However, since this information is not useful for test environment evaluation, this graph adds little to Fig. 6.

4.1.2 Under the AMMI2 model

When two IPCs are required to approximate the data, AMMI2 mega-environment display can be used to present the which-won-where results from AMMI analysis (Fig. 8). In this graph, the environments are defined by their IPC1 and IPC2 scores. This is useful information; it indicates which environments contribute more (or less) to GE. For example, *E*8, *E*5, and *E*7 had long distances from the plot (not biplot) origin, indicating that they contributed more to GE than others. This is consistent to the conclusion from the GGE biplot (Fig. 5) that these environments were less representative of the average environment. When information on the mean yield (G) and the genotypic IPC1 and IPC2 scores are incorporated, the graph is

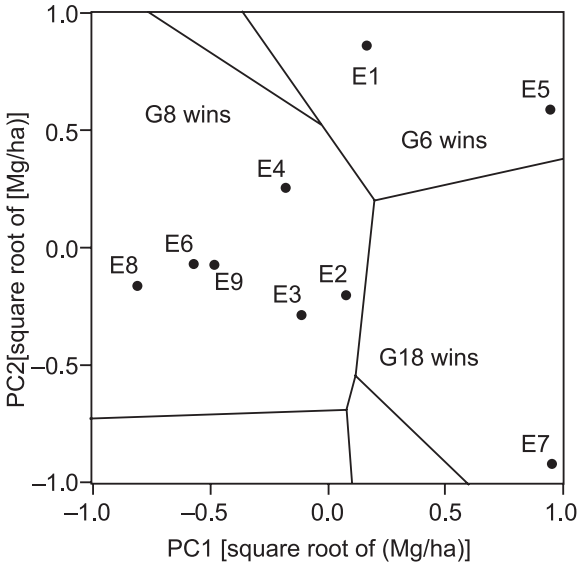

Fig. 8. The AMMI2 mega-environment display. This graph is a display of the which-won-where results identified based on information of G, IPC1, and IPC2. Adopted from Fig. 4 of Gauch *et al.* (2008)

divided into several irregular areas, and for each area, a nominal winner genotype is identified. According to Gauch *et al.* (2008, p.872), the AMMI2 mega-environment display is constructed as follows. "….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 (based on G and the two IPC). 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 can be seen from this description, considerable effort is needed to construct an AMMI2 display, in addition to SVD, whereas SVD is all that is needed to construct a GGE biplot. Once the GGE biplot is constructed, the which-won-where pattern can be drawn by the eye.

Based on G, IPC1, and IPC2, the nine test environments are divided into three sub areas: *E*7 stands alone, with *g*18 as the winner; *E*1 and *E*5 form a group, with *g*6 as the winner; and the other six environments form a third group , with *g*8 as the winner. Note again this which-won-where pattern is very similar to the GGE biplot based on a subset of higher-yielding genotypes (Fig. 3).

So the AMMI1 mega-environment display (Fig. 6 and/or Fig. 7) is equivalent to the which-won-where view of the GGE biplot based on the full dataset (Fig. 2), and the AMMI2 mega-environment display (Fig. 8) is equivalent to the which-won-where view of the GGE biplot based on a subset of higher-yielding genotypes. However, the GGE biplot displays not only the mega-environment delineation *result* but also the genotype-by-environment data that lead to the result. The GGE biplot contains information on all genotypes whereas the AMMI displays contain only the "winning" genotypes. Thus, although AMMI analysis can achieve the same or similar mega-environment delineation, the GGE biplot method is simpler, more informative, and more elegant.

4.2 Genotype Evaluation

The graph for genotype evaluation in AMMI analysis is the AMMI1 "biplot" (Fig. 9). Its abscissa represents the main effects (G and E) and its ordinate represents the IPC1 scores representing GE of the genotypes and environments. This graph provides a means to simultaneously visualize the mean performance (G) and the stability (IPC1) of the genotypes. Although it is often called a biplot because it presents both genotypes and environments, the

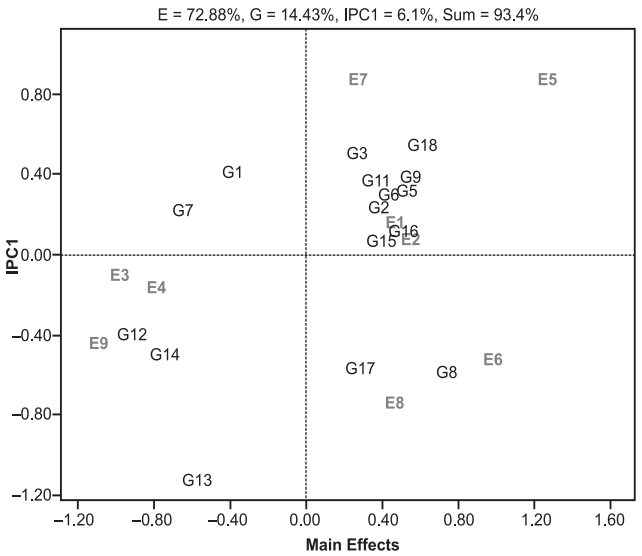

Fig. 9. The AMMI1 biplot. It can be used to visualize the mean performance and stability of the genotypes across environment. It also represent the environmental main effect and the environments' contribution to genotype-by-environment interaction, but this information is not useful for test environment evaluation. This "biplot" is not a true biplot because it does not have the inner-product property.

AMMI1 biplot is not a true biplot because it does not have the inner-product property of a biplot (Equation (7)). In terms of genotype evaluation, the AMMI1 graph is similar to the mean vs. stability view of the GGE biplot (Fig. 4). For example, it shows that the genotype *g*8 had highest mean yield while *g*12 had the lowest. It also shows *g*13 to be least stable. However, the AMMI1 biplot is less useful than the GGE biplot because it always explains less G+GE than the GGE biplot. In addition, the commonly used AMMI1 biplot (such as Fig. 1 of Gauch *et al*. 2008) has different units in the two axes: while the units of abscissa are in yield per unit area, those of the ordinate are in its square root. This makes the shape of the graph completely subjective, which may mislead the researcher on the relative importance of the mean (G) vs. the stability (GE). The AMMI1 biplot presented in Fig. 9 uses square root of yield as units for both axes, which partially removes this problem.

4.3 Test Environment Evaluation

As mentioned above, test environment evaluation has not been a research topic in AMMI analysis. The AMMI1 biplot (Fig. 9) displays the test environments by their main effects E and IPC1 scores, but it provides no information on the ability of the test environments in identifying superior genotypes.

5. CONCLUSIONS

This paper leads to the following conclusions:

1. Graphical analysis of MET data is a major component in plant breeding programs and regional variety performance trials. Complete MET data analysis involves three major objectives: mega-environment analysis, genotype evaluation, and test environment evaluation.

2. AMMI analysis and GGE biplot analysis are two popular systems for MET data analysis, mainly due to their graphical presentation of the data.

3. Both AMMI analysis and GGE biplot analysis use the information of G+GE in mega-environment analysis and genotype evaluation. In AMMI analysis, G and GE are separated first and then re-assembled together. In GGE biplot analysis, no attempt is made to separate G from GE; they are dealt with jointly.

4. Both AMMI analysis and GGE biplot analysis have graphics that are capable of mega-environment analysis. The GGE biplots is more informative because it explains more G+GE and preserves information on the performance of each genotype in each environment. The GGE biplot is also simpler to construct because no additional calculations are needed beyond constructing the biplot.

5. Both AMMI analysis and GGE biplot analysis have graphics that allow visual evaluation of genotypes in two aspects: mean performance and stability across environments. The GGE biplot is more effective because it contains more G+GE variation than the AMMI1 “biplot”, which is not a true biplot. The GGE biplot is also more informative due to its inner-product property.

6. GGE biplot analysis is effective in graphical evaluation of test environments in two aspects: their ability to discriminate the genotypes and their representativeness of the target environment represented by all test environments. AMMI analysis does not have graphics for test environment evaluation.

7. Overall, GGE biplot analysis is a preferred graphical system for MET data analysis.

REFERENCES

Ebdon, J.S. and Gauch, H.G. Jr. (2002). Additive Main Effect and Multiplicative Interaction Analysis of national turfgrass performance trials: II. Cultivar recommendations. *Crop Sci*., **42**, 497-506.

Cooper, M. and DeLacy, I.H. (1994). Relationships among analytical methods used to study genotypic variation and genotype-by-environment interaction in plant breeding multi-environment experiments. *Theo. Appl. Genetics,* **88**, 561-572.

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. (2006). Statistical analysis of yield trials by AMMI and GGE. *Crop Sci*., **46**, 1488-1500.

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

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

Lin, C.S. and Binns, M.R. (1994). Concepts and methods of analyzing regional trial data for cultivar and location selection. *Plant Breed. Rev*., **12**, 271-297.

Samonte, S.O.P.B., Wilson, L.T., McClung, A.M. and Medley, J.C. (2005). Targeting cultivars onto rice growing environments using AMMI and SREG GGE biplot analyses. *Crop Sci*., **45**, 2414-2424.

Yan, W. (2001). GGEBiplot—A Windows application for graphical analysis of multi-environment trial data and other types of two-way data. *Agron. J*., **93**, 1111-1118.

Yan, W. (2002). Singular-value partition for biplot analysis of multi-environment trial data. *Agron. J*., **94**, 990-996.

Yan, W., Pageau, D., Fregeau-Reid, J. and Durand, J. (2011). Assessing the representativeness and repeatability of test locations for genotype evaluation. *Crop. Sci*. (in press)

Yan, W., Frégeau-Reid, J., Pageau, D., Martin, R., Mitchell-Fetch4, J., Etienne, M., Rowsell, J., Scott, P., Price, M., de Haan, B., Cummiskey, A., Lajeunesse, J., Durand, J. and Sparry, E. (2010). Identifying essential test locations for oat breeding in eastern Canada. *Crop Sci*., **50**, 504-515.

Yan, W. and Hunt, L.A. (2003). Biplot analysis of multi-environment trial data. p. 289–303. In : M. S. Kang (ed.) *Quantitative Genetics, Genomics, and Plant Breeding*. CAB International, Wallingford, Oxon, UK

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

Yan., W. and Holland, J.B. (2010). A Heritability-adjusted GGE biplot for test environment evaluation. *Euphytica*, **171(3)**, 355-369.

Yan, W. and Kang, M.S. (2003). *GGE Biplot Analysis* : *A Graphical Tool for Breeders, Geneticists, and Agronomists*. CRC Press. Boca Raton, FL.

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

Yan, W. and Tinker, N.A. (2005). An integrated system of biplot analysis for displaying, interpreting, and exploring genotype by environment interactions. *Crop Sci*., **45**, 1004-1016.

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

Zobel, R.W., Wright, M.J. and Gauch, H.G. (1988). Statistical analysis of a yield trial. *Agron. J*., **80**, 388-393.