SOFTWARE

GGEbiplot—A Windows Application for Graphical Analysis of Multienvironment Trial Data and Other Types of Two-Way Data

Weikai Yan*

ABSTRACT

Plant breeding trials produce quantities of data and finding the useful information within that data has historically been a major challenge of plant breeding. A recently developed graphical data summary, called GGEbiplot, can aid in data exploration. GGEbiplot is a Windows application that performs biplot analysis of two-way data that assume an entry × tester structure. GGEbiplot analyzes the data and outputs the results as an image, and it also produces an interactive show of the data. It allows interactive visualization of the biplot from various perspectives. A multienvironment trial data set, in which cultivars are entries and environments are testers, was used to demonstrate the functions of GGEbiplot. These include but are not limited to: (i) ranking the cultivars based on their performance in any given environment, (ii) ranking the environments based on the relative performance of any given cultivar, (iii) comparing the performance of any pair of cultivars in different environments, (iv) identifying the best cultivar in each environment, (v) grouping the environments based on the best cultivars, (vi) evaluating the cultivars based on both average yield and stability, (vii) evaluating the environments based on both discriminating ability and representativeness, and (viii) visualizing all of these aspects for a subset of the data by removing some of the cultivars or environments. GGEbiplot has been applied to visual analysis of genotype × environment data, genotype × trait data, genotype × marker data, and diallel cross data.

Plant breeding trials produce quantities of data and finding the useful information within that data has historically been a major challenge of plant breeding. Yan (1999) and Yan et al. (2000) presented a versatile graphical approach for analyzing multienvironment trials (METs), called *GGE biplot*. Since the publication of Yan et al. (2000), I have received many positive comments from readers all over the world. It appears that the appreciation and acceptance of the GGE biplot methodology by the readers are immediate. However, while most readers like it, few know how to apply it to their own data. Indeed, it is a tedious, if not difficult, process even for well-trained biometricians. There are a few commercial software programs that can generate biplots, such PC-ORD and Canoco (http://www.ptinet.net/~mjm/canoco.htm; verified 19 Mar. 2001), as well some SAS micros (SAS Inst., 1996), but the biplots generated by these programs are too primitive to be useful. Software that can fulfill biplot analysis as described in Yan et al. (2000) has not been developed. To facilitate the application of the GGE biplot methodology in MET data analysis and in analyses of other types of two-way data, a Windows application, the GGEbiplot software, was developed. This paper describes the functions built in this software and exemplifies their use in MET data analysis.

Crop Sci. Division, Dep. of Plant Agric., Univ. of Guelph, Guelph, ON, Canada N1G 2W1. Received 4 Jan. 2001. *Corresponding author (wyan@uoguelph.ca).

Published in Agron. J. 93:1111–1118 (2001).

WHAT IS A GGE BIPLOT?

The Concept of Biplot

The concept of biplot was first proposed by Gabriel (1971). The main ideas follow. Any two-way table or matrix X that contains n rows and m columns can be regarded as the product of two matrices: A with n rows and r columns and B with r rows and m columns. Therefore, Matrix X can always be decomposed to its two component matrices, A and B. If r happens to be 2, Matrix X is referred to as a rank-two matrix. Each row in Matrix A has two values, which define a point in a two-dimensional plot. Similarly, each column in Matrix B has two values, which also define a point in a two-dimensional plot. When both the n rows of A and m columns of B are displayed in a single plot, this plot is called a biplot. Therefore, the biplot of a rank-two matrix contains $n + m$ points, compared with $n \times m$ values in the original matrix, and yet contains all of the matrix information.

One interesting property of a biplot is that each of the $n \times m$ values can be precisely recovered by viewing the $n + m$ points on the biplot. Assume that we have yield data of three-genotype × three-environment matrix that is a rank-two matrix. After decomposition of the data into its two component matrices, the three genotypes and three environments can be presented in a biplot like Fig. 1. The yield of genotype i in environment j, Y_{ij}, can be recovered by the following formula:

$$Y_{ij} = \overline{\mathbf{OE_j}} \cos\alpha_{ij} \overline{\mathbf{OG_i}} = \overline{\mathbf{OE_j}}\ \overline{\mathrm{OP}_{ij}}$$

where $\overline{\mathbf{OG_i}}$ (or $\mathbf{OG_i}$) is the absolute distance from the biplot origin O to the marker of the genotype i, $\overline{\mathbf{OE_j}}$ (or $\mathbf{OE_j}$) is the absolute distance from the biplot origin O to the marker of environment j, α_{ij} is the angle between the vectors $\mathbf{OG_i}$ and $\mathbf{OE_j}$ and $\overline{\mathrm{OP}_{ij}}$ (or $\mathrm{OP}_{ij} = \cos\alpha_{ij}\ \mathbf{OG_i}$) is the projection of the marker of genotype i to the vector of environment j. To compare the yield of the three genotypes in Environment E1, from Fig. 1, we have

Abbreviations: ATC, average tester coordinate; MET, multienvironment trial; PC, principal component.

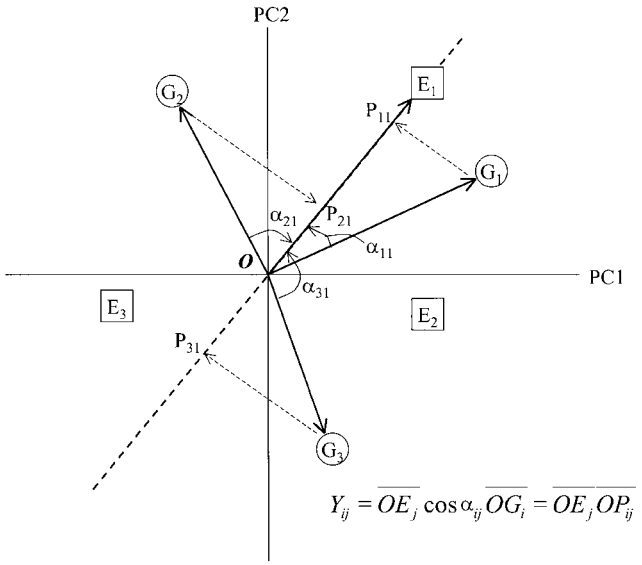

Fig. 1. The geometry of biplot. E1, E2, and E3 are three hypothetical environments; G1, G2, and G3 are three hypothetical genotypes; PC1 and PC2 are first and second principal components, respectively.

$$Y_{11} = (\mathbf{OE_1})\ (\cos\alpha_{21})(\mathbf{OG_1}) = (\mathbf{OE_1})(\mathrm{OP}_{11})$$
$$Y_{21} = (\mathbf{OE_1})\ (\cos\alpha_{21})(\mathbf{OG_2}) = (\mathbf{OE_1})(\mathrm{OP}_{21})$$
$$Y_{31} = (\mathbf{OE_1})\ (\cos\alpha_{31})(\mathbf{OG_3}) = (\mathbf{OE_1})(\mathrm{OP}_{31})$$

where Y_{11}, Y_{21}, and Y_{31} are the yields of Genotypes 1, 2, and 3 in Environment E1; OP_{11}, OP_{21}, and OP_{31} are the projections of the genotype markers onto the vector or its extension of Environment E1. Because $\mathbf{OE_1}$ is common to all genotypes, comparison among Y_{11}, Y_{21}, and Y_{31} can be performed by simply visualizing OP_{11}, OP_{21}, and OP_{31}. In our example of Fig. 1, it is obvious that $OP_{11} > OP_{21} > OP_{31}$; therefore, $Y_{11} > Y_{21} > Y_{31}$. Note that OP_{11} and OP_{21} are above average, whereas OP_{31} is below average because $\cos\alpha_{11}$ and $\cos\alpha_{21}$ are positive and $\cos\alpha_{31}$ is negative.

Approximation of Any Two-Way Table Using a Rank-Two Matrix

Biplot is obviously an elegant display of a rank-two matrix. In reality, however, a two-way data set rarely is exactly a rank-two matrix. Nevertheless, if a two-way data set, e.g., the yield data of a number of cultivars tested in a number of environments, can be approximated by a rank-two matrix, the latter can then be displayed in a biplot (Gabriel, 1971). The process of decomposing Matrix X into its component matrices, A and B, is called singular value decomposition, the result of which is *r* principal components (PCs; *r* equals the smaller of *n* and *m*). Approximation of Matrix X with a rank-two matrix means that only the first two principal components (PC1 and PC2) are used to represent the original Matrix X. If PC1 and PC2 explain a large proportion of the total variation of X, then X is said to be sufficiently approximated by a rank-two matrix and can be approximately displayed in a biplot.

The Concept of GGE

The concept of GGE originates from analysis of METs of crop cultivars. The yield of a cultivar (or any other measure of cultivar performance) in an environment is a mixed effect of genotype main effect (G), environment main effect (E), and genotype × environment interaction (GE). In normal METs, E accounts for 80% of the total yield variation, and G and GE each account for about 10% (Gauch and Zobel, 1996; Yan et al., 2000). For the purpose of cultivar evaluation, however, only G and GE are relevant (Gauch and Zobel, 1996). Furthermore, both G and GE must be considered in cultivar evaluation, thus the term GGE (Yan et al., 2000).

The Model for Constructing a GGE Biplot

The GGE biplot is a biplot that displays the GGE part of MET data. The basic model for a GGE biplot is

$$Y_{ij} - \overline{Y}_j = \lambda_1\xi_{i1}\eta_{j1} + \lambda_2\xi_{i2}\eta_{j2} + \varepsilon_{ij} \quad [1]$$

where

- Y_{ij} is the average yield of genotype *i* in environment *j*
- $\overline{Y}_j$ is the average yield over all genotypes in environment *j*
- λ_1 and λ_2 are the singular values for PC1 and PC2, respectively
- ξ_{i1} and ξ_{i2} are the PC1 and PC2 scores, respectively, for genotype *i*
- η_{j1} and η_{j2} are the PC1 and PC2 scores, respectively, for environment *j*
- ε_{ij} is the residual of the model associated with the genotype *i* in environment *j*

To display PC1 and PC2 in a biplot, the equation is rewritten as

$$Y_{ij} - \overline{Y}_j = \xi^*_{i1}\eta^*_{j1} + \xi^*_{i2}\eta^*_{j2} + \varepsilon_{ij}$$

where $\xi^*_{in} = \lambda_n^{1/2}\xi_{in}$ and $\eta^*_{jn} = \lambda_n^{1/2}\eta_{jn}$, with $n = 1, 2$. This scaling method has the advantage that PC1 and PC2 have the same unit (square root of the original unit, e.g., t ha^{-1} in terms of yield) although other scaling methods are also valid.

A GGE biplot is generated by plotting ξ^*_{i1} and η^*_{j1} against ξ^*_{i2} and η^*_{j2}, respectively. The GGE biplot has been used previously in MET data analysis (e.g., Cooper et al., 1997), but methods of interpreting the GGE biplot, as described in this paper, became available only since Yan (1999) and Yan et al. (2000).

In METs, a number of cultivars are tested in a number of environments. To extend the application of GGE biplot to other types of two-way data with similar data structure, the cultivars can be generalized as entries and the environments as testers.

THE GGEbiplot SOFTWARE

The GGEbiplot software was developed to facilitate application of the GGE biplot analysis to MET data and other types of two-way data. GGEbiplot is graphical and interactive; all operations are performed using a pointer (e.g., a mouse). It not only analyzes the data and

displays the GGE biplot, but it also allows the researcher to examine the biplot in various perspectives. The functions built in GGEbiplot are organized under menu entries File, View, Visualization, Format, Model Selection, Data Management, Plot Selection, Other Functions, and Help. These are briefly described below.

File

The File menu contains the following functions:

- The New Job function allows the user to open and visualize an unlimited number of data sets in a single run.
- The Print function allows the biplot image to be printed to a printer; Adobe Writer, which creates a pdf file of the image; or other output devices available to the computer.
- The Exit function quits the program.

View

The View menu provides the following options:

- Display the entries or testers by their full names or by a single letter. If presented in single letters, entries (e.g., cultivars) are presented by the letter C and testers (e.g., environments) by the letter E.
- Show testers only, entries only, or both in the biplot.
- Show or hide the GGEbiplot logo.

Visualization

The Visualization menu bar is the center of the program. It provides the following functions:

- Draw vector lines for the testers, which connect the biplot origin and the markers of each of the testers. The cosine of the angle between two testers (i.e., environments in terms of MET data) approximates the correlation of the two testers. The vectors thus help visualize the similarities among the testers in their differentiation of the entries.
- Show a linear map of the testers. This is a linear display of the angles among the tester vectors. This function is particularly useful when a genotype $\times$ genetic marker data is visualized. The linear map of the markers mimics a genetic map so that groups of genes and quantitative trait loci can be visualized.
- Rank the testers based on the relative performance (adaptation) of any given entry.
- Rank the entries based on their performance with regard to any given tester.
- Discard entries based on a tester. When this function is evoked, entries that had a below-average value with regard to the tester of interest (an environment or a trait) will be removed from the biplot. This mimics independent culling based on performance in an environment or on a trait.
- Compare the performance of two entries with regard to the testers.
- Compare the entries with a check or an *ideal* entry.
- Compare the testers with an *ideal* tester.
- Identify the best entry with regard to each tester.
- Group the testers based on the best entries.
- Visualize the average performance and stability of each entry.
- Visualize the discriminating ability and representativeness of each tester.

Format

The biplot image is designed to be publishable in scientific journals. The resolution of the image is 96 pixels per inch. The following functions were built in under the Format menu bar:

- Give the biplot a title.
- Change the color scheme of the biplot.
- Change the font characteristics of the title and labels in the biplot.
- Change the biplot size. The biplot can fill the screen if the shape of the biplot, which is determined by the data, allows this.

Model Selection

The GGEbiplot software provides options for looking at the same data set using three different models. The first model is presented in Eq. [1]. The second model is

$$(Y_{ij} - \overline{Y}_j)/S_j = \lambda_1\xi_{i1}\eta_{j1} + \lambda_2\xi_{i2}\eta_{j2} + \varepsilon_{ij} \quad [2]$$

where S_j is the standard deviation among the entry means for tester j. This model is particularly useful for analyzing data in which different testers (such as different traits) use different units so that the units are removed.

The third model is based on

$$(Y_{ij} - \overline{Y}_j)/Z_j = \lambda_1\xi_{i1}\eta_{j1} + \lambda_2\xi_{i2}\eta_{j2} + \varepsilon_{ij} \quad [3]$$

where Z_j is the standard error for tester j. Obviously, Z_j can be estimated only with replicated data. This model is preferred for all types of two-way data when replicated observations are available because it adjusts any heterogeneity among the testers.

A function is also built under this menu bar that allows the entries and the testers to switch roles.

Data Management

With the Data Management option, the user can:

- visualize a balanced subset by removing entries or testers that have missing cells and
- visualize a subset of the original data by removing one entry or tester at a time.

Plot Selection

A GGE biplot normally refers to a biplot of PC1 vs. PC2. For large data sets with complex patterns, it may be necessary to also examine the biplot of PC3 vs. PC4. GGEbiplot provides options for examining biplots of PC1 vs. PC2, PC3 vs. PC4, PC5 vs. PC6, PC1 vs. PC3, and PC2 vs. PC3. Users may find that some of these options make sense with their data.

Other Functions

Under the Other Functions menu entry, the user can:

- provide options for printing out eigenvectors, etc., and
- provide other graphical analyses (still under construction).

Help

A help file will be displayed upon clicking the Help menu bar.

The Output Log File

In addition to graphic outputs, GGEbiplot generates a log file, named GGEbiplot.gge, which is placed in the same folder or directory from which the data was read. The log file contains, among others, the number of entries (cultivars), the number of testers (environments), the number of missing cells, the averaged two-way table that is subjected to singular value decomposition, the variation explained by each of the PCs, and the PC scores for each of the entries and testers that are used to generate the GGE biplot. The correlation coefficient matrix among the testers and eigenvectors can be printed upon selecting appropriate entries under the Other Functions menu entry.

Input Data Format

GGEbiplot is designed for the analysis of balanced two-way data but is tolerant to data sets with missing cells. GGEbiplot can read two types of input data format. The first format is one in which each row contains one observation. Some simple requirements for this format follow:

1) The first line is the header and contains tester name, block name, entry name, and trait name (separated by commas).
2) The data should be in four columns in the order of tester name, block name, entry name, and the measured value (also delimited by commas).

The data do not have to be balanced in terms of entry × tester combinations. The number of blocks can differ with testers. There is no need to indicate missing cells. Missing cells, if any, will be replaced by the respective tester means, and the user will be notified once missing cell(s) are detected.

The second data format that the program can read is one in which the data is presented as a two-way table. Here are some simple requirements:

1) The first row contains the name of the header for the first column and the names of the testers, delimited by commas.
2) Each of the subsequent rows contains the name of the entry and values for each tester with regard to the current entry, also delimited by commas.
3) Missing cells should be indicated by −99. Lack of this information can lead to chaos in the calculation.

The program is designed to accommodate data of 300 entries × 300 testers with three replications although it can be increased or decreased according to the user's requirement.

Computer Requirement and Software Availability

This program works on a Windows 95 platform or later versions of Microsoft Windows. It requires a minimum of 5 megabytes (MB) of random access memory (RAM). For a data set of 20 genotypes × 10 environments with three replications, a minimum 6 MB of RAM is required. The software is available upon request with negotiable charge.

AN EXAMPLE OF MULTIENVIRONMENT TRIAL DATA ANALYSIS USING THE GGEbiplot SOFTWARE

Getting the GGE Biplot

This section exemplifies MET data analysis using GGEbiplot. The sample data are yields from the 1993 Ontario winter wheat (*Triticum aestivum* L.) performance trials, in which 18 cultivars were tested at nine locations. When the data is read correctly, a data-based GGE biplot will appear on the screen (Fig. 2). The GGE biplot contains markers for each of the 18 cultivars in lower case and blue color, as distinguished from markers for each of the nine environments in upper case and red color. The accurate positions of the cultivars and environments are at the beginning of the labels. The model that is used for generating the biplot, along with the percentages of GGE explained by the two axes, are indicated in a rectangular box at the upper-left corner of the biplot. Thus, the GGE biplot for this sample data set, using Model 1 (Eq. [1]), explained 59 + 19% = 78% of the yield variation due to GGE (not to be con-

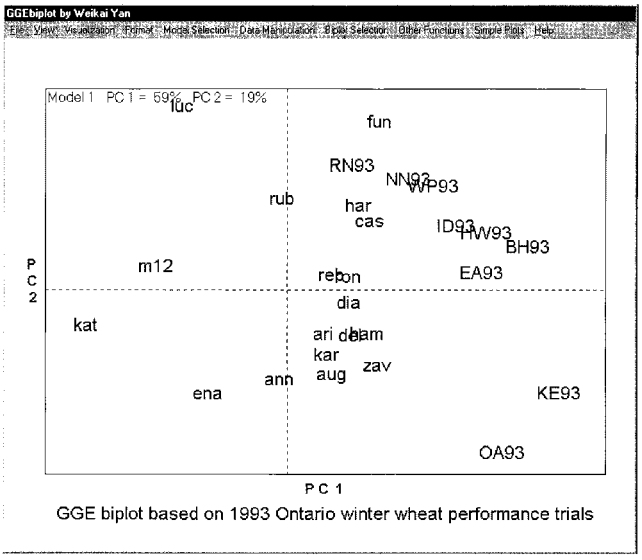

Fig. 2. GGE biplot based on the 1993 Ontario winter wheat performance trials. Cultivars are in lower case, and locations are in upper case. PC1 and PC2 are first and second principal components, respectively.

fused with the total yield variation, which includes E as well as G and GE) (Fig. 2).

The GGEbiplot software provides options to view this biplot in numerous ways to address most questions a breeder or researcher is likely to ask, as will be exemplified below.

The Performance of Different Cultivars in a Given Environment

From the Visualization menu bar, select Examine an Entry or Tester. The Entry/Tester Examination Unit will appear on the top of the biplot (Fig. 3). Click the Entry/Tester toggle button until Tester appears in the combo-box. Clicking the drop-down arrow by the combo-box, a list of the environments will appear. Choose any environment of your interest (BH93 in this example) from the list or type the name of the environment in the combo-box. Upon clicking the Look Up button, the following features will appear (Fig. 3):

1) a thick red line that passes through the plot origin and the selected environment (BH93 in this example), which is referred to as the *tester axis*;
2) an oval that surrounds the chosen environment, which indicates the positive direction of the tester axis;
3) a thick blue line that passes through the plot origin and is perpendicular to the tester axis, referred to as the *perpendicular line*; and
4) a group of lines parallel to the perpendicular line.

The cultivars are ranked in the direction of the tester axis, and the parallel lines help visualize the ranking of the cultivars. In this example, 'fun' was the best, followed by 'cas' and 'har', and 'kat' was the poorest in the selected environment BH93. The perpendicular line separates cultivars that performed below average from those performing above average in BH93. Namely, cultivars kat, m12, ena, luc, and ann performed below average, whereas other cultivars, on the same side of the perpendicular line as BH93, performed above average.

The Relative Adaptation of a Given Cultivar in Different Environments

From the Visualization menu bar select Examine an Entry or Tester. The Entry/Tester Examination Unit will appear on the top of the biplot. Click the Entry/Tester toggle button until Entry appears in the combo-box. Clicking the drop-down arrow by the combo-box, a list of the cultivars appears. Choose any cultivar of your interest ('rub' in this example) or type the name of the cultivar in the combo-box. Upon clicking the Look Up button, the following features will appear (Fig. 4):

1) a thick red line that passes through the plot origin and the selected cultivar, which is referred to as the *entry axis*;
2) an oval that surrounds the chosen cultivar (rub, in this example), which indicates the positive direction of the entry axis;
3) a thick blue line that passes through the plot origin and is perpendicular to the entry axis, referred to as the perpendicular line; and
4) a group of lines parallel to the perpendicular line.

The environments are ranked in the direction of the entry axis, and the parallel lines help visualize the ranking of the environments in terms of the relative performance of rub. Thus, rub performed the best in RN93, followed by NN93, WP93, IN93, HW93, BH93, EA93, KE93, and OA93. The perpendicular line separates environments in which rub performed above average from those in which rub performed below average. Thus, rub yielded above average in RN93, NN93, WP93, IN93, and HW93; just average in BH93; and below average in EA93, KE93, and OA93.

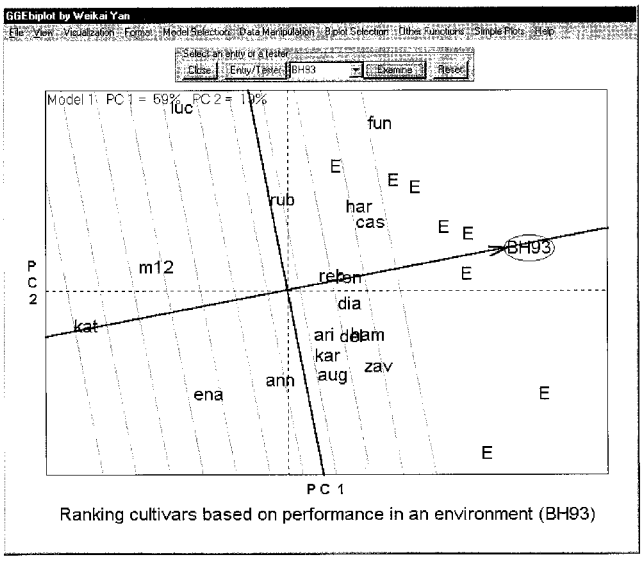

Fig. 3. Performance of different cultivars in environment BH93. Cultivars are in lower case, and locations are in upper case. PC1 and PC2 are first and second principal components, respectively.

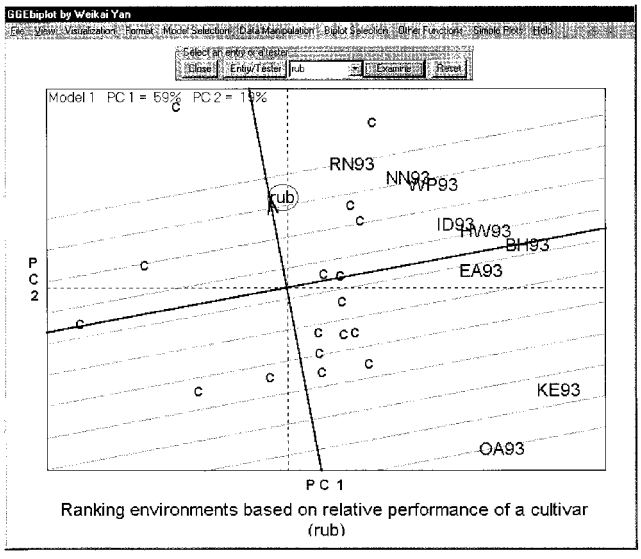

Fig. 4. Relative performance of cultivar rub in different environments. Cultivars are in lower case, and locations are in upper case. PC1 and PC2 are first and second principal components, respectively.

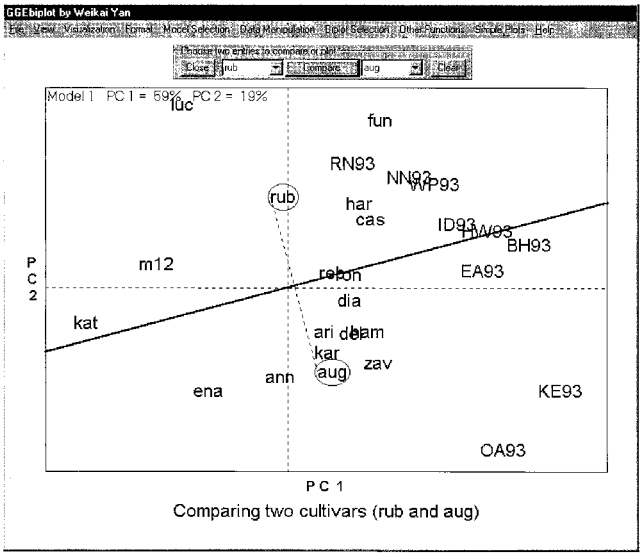

Fig. 5. Comparison of two cultivars in different environments. Cultivars are in lower case, and locations are in upper case. PC1 and PC2 are first and second principal components, respectively.

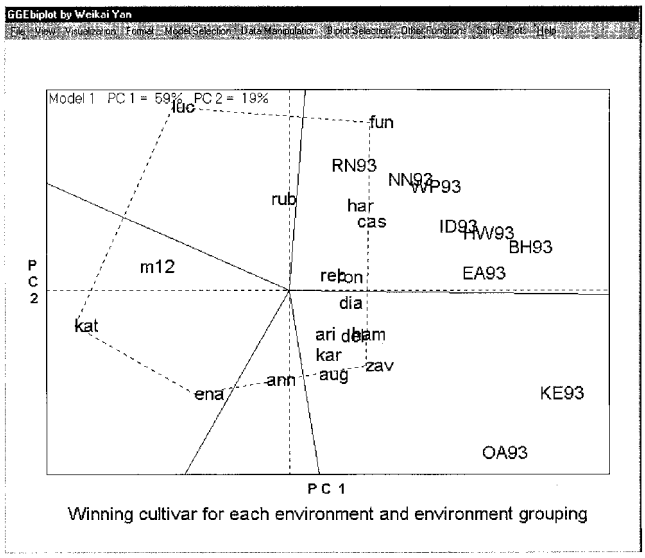

Fig. 6. Convex-hull view of the GGE biplot. Cultivars are in lower case, and locations are in upper case. PC1 and PC2 are first and second principal components, respectively.

Comparison of Two Cultivars

From Visualization menu bar, select Compare Two Entries. The Entry Comparison Unit will appear on the top of the biplot. The unit includes two combo-boxes labeled Entry1 and Entry 2, respectively (Fig. 5). Choose any cultivar from Entry 1, choose a different cultivar from Entry 2, and click the Compare button. The names of the cultivars to be compared can also be typed in the combo-boxes. Upon clicking Compare, the following features will appear (Fig. 5):

1) two ovals that circle the two selected cultivars, respectively;
2) a thick red line, called the *jointer line*, that connects the two ovals; and
3) a blue line that is perpendicular to the jointer line and passes through the plot origin.

In the example of Fig. 5, cultivars aug and rub were compared. We see that tester environments OA93, KE93, EA93, and BH93 were on the aug side of the perpendicular. Thus, aug was better than rub in these environments. Similarly, rub was better than aug in the other five environments, namely HW93, IN93, WP93, NN93, and RN93.

The Best Cultivar(s) in Each Environment

From Visualization menu bar, choose Draw Convex Hull. The GGE biplot will become like Fig. 6. The convex hull in Fig. 6 is drawn on cultivars relatively remote from the biplot origin so that all other cultivars are contained within the convex hull. Figure 6 also contains a set of lines perpendicular to each side of the convex hull. A perpendicular line does not necessarily intersect the convex-hull side; it may only intersect the extension of the convex-hull side, e.g., the convex-hull side that connects cultivars kat and ena. These perpendiculars divide the biplot into several sectors, and the environments inevitably fall into the sectors. There are five sectors in Fig. 6, with cultivars fun, zav, ena, kat, and luc as the corner or vertex cultivars. Environments OA93 and KE93 fell in the sector in which zav was the vertex cultivar. This means that zav was the best cultivar for OA93 and KE93. The other seven environments fell in the sector in which fun was the vertex cultivar, meaning that fun was the best cultivar for these seven environments. No environments fell into sectors with luc, ena, and kat as the vertices, indicating that these cultivars were not the best in any of the environments. Actually, this indicates that they were the poorest cultivars in some or all of the environments.

The Megaenvironments

Another important feature of Fig. 6 is that it indicates environmental groupings, which suggests possible existence of different megaenvironments. Thus, two megaenvironments were suggested in Fig. 6. Based on biplot analysis of 11 yr of data, Yan (1999) proposed that the Ontario winter wheat growing regions consist of two megaenvironments, rather than four as previously believed.

The Average Yield and Stability of the Cultivars

From the Visualization menu bar, choose Show Average Tester Coordinate. An average tester coordinate (ATC) based on the average environment will appear (Fig. 7). The ATC *x*-axis passes through the biplot origin and the marker of the average environment, which is defined by the average PC1 and PC2 scores over all environments. The oval indicates the positive end of the ATC *x*-axis. The ATC *y*-axis passes the plot origin and is perpendicular to the ATC *x*-axis. The average yield of the cultivars is approximated by the projections of their markers to the ATC *x*-axis. Thus, cultivar fun had the highest average yield, and kat had the lowest. The

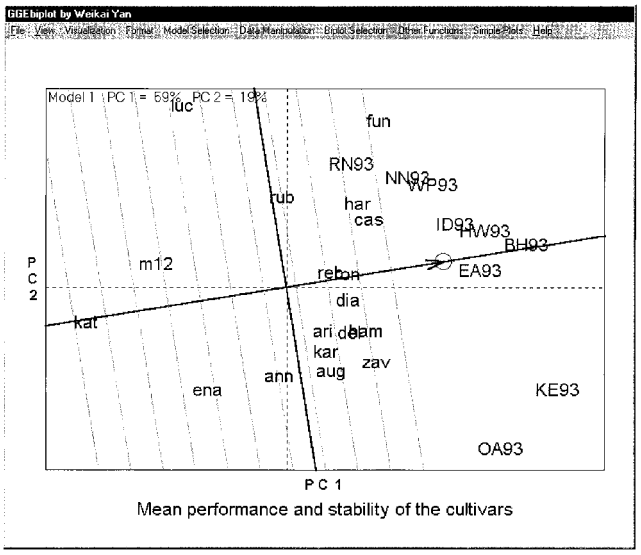

Fig. 7. Average tester coordinate (ATC) view of the GGE biplot. Cultivars are in lower case, and locations are in upper case. PC1 and PC2 are first and second principal components, respectively.

lines parallel to the ATC *y*-axis helps ranking the cultivars in the terms of average yield. The stability of the cultivars is measured by their projection to the ATC *y*-axis. The greater the absolute length of the projection of a cultivar, the less stable it is. Thus luc was the least stable cultivar while 'reb' and 'ron' were the most stable.

The Representativeness and Discriminating Ability of the Environments

Figure 7 also shows the representativeness and discriminating ability of the environments. The vector length, i.e., the absolute distance between the marker of an environment and the plot origin, is a measure of its discriminating ability: the longer the vector, the more discriminating the environment. The absolute length of the projection from the marker of an environment onto the ATC *y*-axis is a measure of its representativeness: the longer the projection, the less representative the environment. Thus, environment BH93 was most representative (as it had a near-zero projection on the ATC *y*-axis) and also highly discriminating (as it had a large projection onto the ATC *x*-axis). Environments KE93 and OA93 were discriminating (far away from the origin) but not representative of the average environment (large projection onto the ATC *y*-axis). Environment RN93 was neither discriminating (small distance from origin) nor representative (large projection onto the ATC *y*-axis).

Cultivar Ranking Based on Both Average Yield and Stability

From the Visualization menu bar, clicking Compare with. . . \The 'Ideal' Entry leads to Fig. 8A. The center of the concentric circles is where an *ideal* cultivar should be; its projection on the ATC *x*-axis was designed to be equal to the longest vector of all cultivars, and its projection on the ATC *y*-axis was obviously zero, meaning that it is absolutely stable. Therefore, the smaller the distance from a cultivar to such a virtual cultivar, the more ideal the cultivar is. Thus, cas was closest to the concentric center, but cultivars zav, dia, ham, ron, reb, cas, har, and fun do not seem to be meaningfully different although other cultivars were apparently inferior.

Environment Ranking Based on Both Discriminating Ability and Representativeness

From the Visualization menu bar, clicking Compare with. . . \The 'Ideal' Tester leads to Fig. 8B. The center of the concentric circles is where an *ideal* environment should be; its projection on the ATC *x*-axis was designed to be equal to the longest vector of all environments; therefore, it is the most discriminating; its projection on

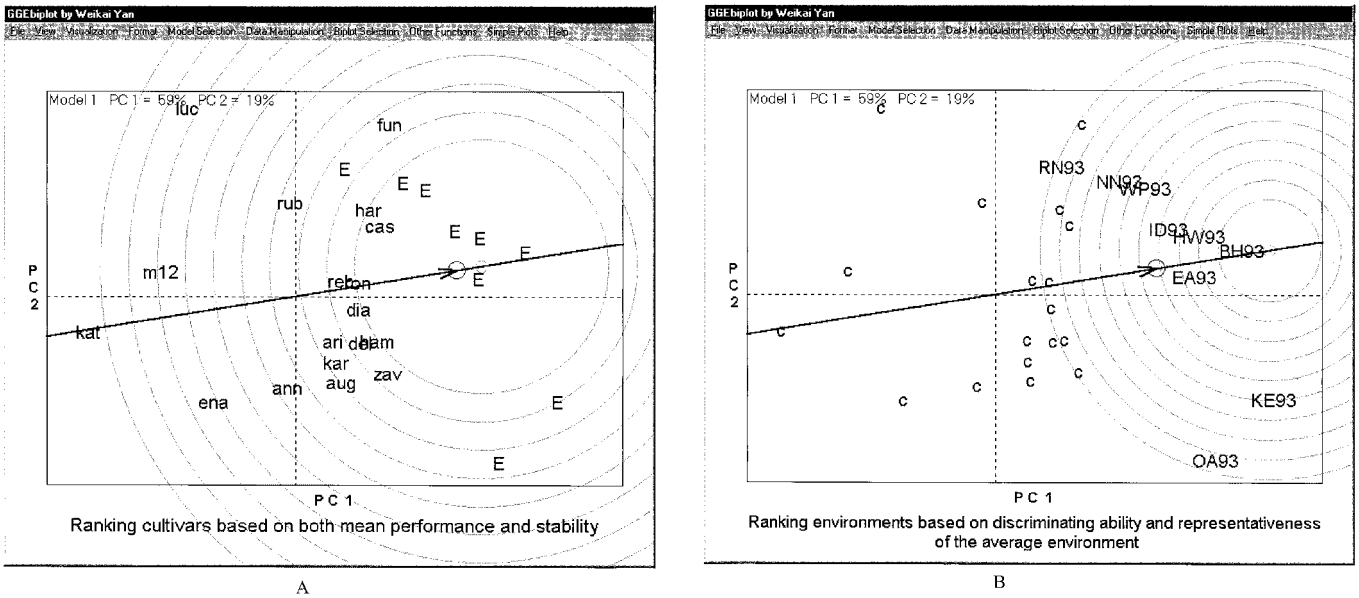

Fig. 8. (A) Comparison of the cultivars with the *ideal* cultivar and (B) comparison of the environments with the *ideal* environments. Cultivars are in lower case, and locations are in upper case. PC1 and PC2 are first and second principal components, respectively.

the ATC *y*-axis was obviously zero, meaning that it is absolutely representative of the average environment. Therefore, the closer an environment is to this virtual environment, the better it is as test environment. Thus, BH93 was the best, followed by EA93, HW93, and IN93 as a group; WP93, NN93, and KE93 as a group; and OA93 and RN93 as the poorest test environments.

Visualizing a Subset of the Data

A subset of the data can be analyzed by removing some of the cultivars or environments. This can be done by clicking Remove an Entry or Remove a Tester under the Data Management menu bar. Figure 6 suggests that OA93 and KE93 belonged to a different megaenvironment from the rest of the environments. When these two environments are removed from the data, the resulting biplot is shown in Fig. 9. This biplot indicated that cultivar fun performed the best for all remaining environments except BH93, in which cultivars har and dia were the best. Note that this conclusion is slightly different from that based on Fig. 6, in which the comparison was primarily between fun and zav.

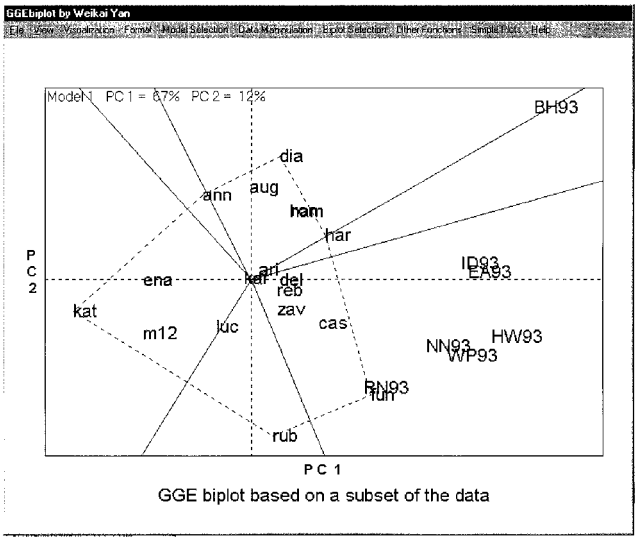

Fig. 9. GGE biplot after environments OA93 and KE93 were removed. Cultivars are in lower case, and locations are in upper case. PC1 and PC2 are first and second principal components, respectively.

SUMMARY

It is clear that the GGEbiplot software is an excellent tool for visual MET data analysis. It graphically addresses the questions that a researcher is likely to ask. The GGE biplot methodology was developed originally for analyzing MET data (Yan, 1999; Yan et al., 2000). However, it can also be used to visualize other types of two-way data. For example, it was satisfactorily used to visualize diallel cross data (Yan and Hunt, unpublished data, 2001), genotype × trait data (Yan and Rajcan, unpublished data, 2001), and genotype × genetic marker data (Yan and Falk, unpublished). Generally, the GGE biplot methodology is equally applicable to all types of two-way data that assume an entry × tester structure. Thus, the GGEbiplot software is not only an excellent tool for visual analysis of MET data, but also a generic tool for visual analysis of other types of two-way data.

REFERENCES

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

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

Gauch, H.G., and R.W. Zobel. 1996. AMMI analysis of yield trials. p. 1–40. *In* M.S. Kang and H.G. Gauch (ed.) Genotype-by-environment interaction. CRC Press, Boca Raton, FL.

SAS Institute. 1996. SAS/STAT user's guide. SAS Inst., Cary, NC.

Yan, W. 1999. Methodology of cultivar evaluation based on yield trial data—with special reference to winter wheat in Ontario. Ph.D. diss. Univ. of Guelph, ON, Canada.

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