GRAPHICAL ANALYSIS OF MULTI-ENVIRONMENT TRIAL (MET) DATA IN SUNFLOWER (*HELIANTHUS ANNUUS* L.) THROUGH CLUSTERING AND GGE BIPLOT TECHNIQUE

IHSAN ULLAH, M. AYUB, *M. RASHID KHAN, M. ASHRAF, M.Y. MIRZA AND M. YOUSAF

Oilseed Programme, National Agricultural Research Centre (NARC) Islamabad
*Department of Biology, Quaid-i-Azam University, Islamabad, Pakistan.
.

Abstract

Multi-environment trial (MET) data of 10 sunflower hybrids were analyzed to: (1) group similar genotypes and similar environments into homogenous groups through cluster analysis, (2) show their interaction with each other through genotype main effects plus genotype x interaction (GGE) biplot graphic display which facilitates the visual evaluation of both the genotypes and environments, and (3) to show the performance of different genotype groups across environments and different environment groups over all genotypes through performance plots. Cluster analysis revealed 5 genotypic and 4 environment groups for seed yield (kg ha^{-1}) and five genotypic and 3 environment group for oil content (%). Pooled analysis showed highly significant results (P< 0.01) with Environments, Genotypes and G X E contributing 81.89, 2.59 and 11.12 % for seed yield kg ha^{-1} and 51.66, 18.11 and 19.2 % for oil content respectively.

The GGE biplot was constructed by plotting the primary effect scores of each genotype (as x-axis) and each environment against their respective secondary effect scores (as y-axis). In these biplots, genotypes are represented by numbers and environments by vectors generating from the origin. Biplots can be used to evaluate cultivars for their yield potential and stability and to evaluate trial sites for their discriminating ability and representiveness.

Genotypes that are close together viz., G1, G4, G5 and G6 are similar in their yield performance across environments, while adjacent environments (E3, E6 and E7) having acute angle between them, are similar in the way they discriminate among genotypes. High yielding and stable genotypes (G1, G4, G5 and G6) usually tend to be on the bottom right quadrant of the joint plot. Regarding oil content (%), E1 and E5 proved to be closely related due to acute angle between them, in the way they discriminated among genotypes. Genotypes G1, G2, G4, G5, G7 and G10 are similar in their behavior across environments in terms of % oil content.

Introduction

For general cultivation of any crop plant, testing at MET is very important to ensure that the selected hybrids/cultivars have acceptable performance in variable environments within the target region (Arshad et al., 2003; Khan et al., 1988; Khan et al., 1987). Effective interpretation and utilization of the MET data in making selection decisions, however, remain a major challenge to researchers. There are two major tasks for researcher to be done in MET data analysis. The first is to determine whether the target region is homogeneous or should be divided into different mega-environments; the second is to select superior cultivars for a given mega-environment on the basis of multiple traits. It is often difficult to determine the pattern of genetic response of different genotypes across the environments without the help of graphical display of the data (*Yan et al.*, 2001).

It is possible to portray the relationship between the genotypes and environments for each attribute graphically. Gabriel (1971, Kroonenberg (1997), Yan *et al.*, (2000) and Yan & Hunt (2001) proposed a GGE biplot that allows visual examination of the GE

interaction pattern of MET data. GGE biplot refers to the genotype main effect (G) and the genotype x environment interaction (GE), which are the two sources of variation that are relevant to cultivar evaluation. In these GGE biplots, genotypes are depicted by numbers and environments by vectors from the origin. It can be used to identify superior cultivars and test environments that facilitate identification of such cultivars (Gwanama *et al.*, 2000).

This GGE biplot can effectively identify the GE interaction pattern of the data. It clearly shows which cultivar won in which environments, and thus facilitates mega-environment identification. Therefore, multi-site trials conducted over years are essential for addressing the mega-environment issue. Ideal cultivars should have a large *PC1* score (high yielding ability) and a small (absolute) *PC2* score (high stability). Similarly, ideal test environments should have a large PC1 score (more discriminating of the genotypes in terms of the genotypic main effect) and small (absolute) PC2 score (more representative of the overall environment) (Yan *et al.*, 2000; Yan & Hunt. 2001).

The objective of this study was to i) graphically summarize the effects of G and GE interaction and to answer the question of "which won where" in the sunflower MET data ii) to examine the possible existence of different mega-environments in the sunflower growing regions, which includes southern parts of Sindh, southern and central parts of Punjab and some regions of NWFP and Baluchistan, iii) to cluster similar genotypes and environments and illustrate the genotype groups' performance in a series of environment groups though performance plots.

Materials and Methods

A set of 10 sunflower single-cross hybrids (Table 1) was evaluated in 7 environments of southern, central and northern Pakistan during 2003-04, to identify patterns of GxE interaction. All the locations were on the public research stations except Monsanto research farm, Manga (Lahore). The set included commercial hybrids widely planted in Pakistan.

Randomized Complete Block Design was used in each environment with four replications. A plot size of 4 rows x 5m and inter-row spacing of 0.75 m was used in all the trials at each location. Tillage and other cultural practices were similar across the sites.

Data for seed yield and oil contents were recorded for all the trials except D.I. Khan where oil data were not recorded. Yield data were determined by hand harvesting two central rows of each plot in each trial discarding the border ones. Oil content was determined on a 10-gram oven-dried sample by nuclear magnetic resonance (NMR) machine (Granlund & Zimmerman, 1975).

Table 1. Hybrids along with their source and origin included in the trials.

S. No.	Hybrid name	Source/origin
1.	G-101	Monsanto Agri. Tech./S. Africa
2.	Hysun-38	ICI Pak seeds/Australia
3.	Mehran-2	ARI, Tandojam/Pakistan
4.	SF-187	Monsanto Agri. Tech./U.S.A
5.	DKS-4040	Monsanto Agri. Tech./ Argentina
6.	Hysun-33	ICI Pak seeds/Australia
7.	FH-106	ORI, Faisalabad/Pakistan
8.	FH-315	ORI, Faisalabad/Pakistan
9.	Bemisal-205	Engro Chemicals, Pakistan/U.S.A
10.	PARSUN-2	Oilseed, NARC/Pakistan

Analysis of variance of mean data from 10 genotypes x 7 environments used for each trait was conducted to examine the partitioning of sums of squares to G, E and GxE interaction. For classification, a hierarchical agglomerative clustering method (Williams, 1976) with incremental sum of squares (Ward, 1963) as the fusion criterion was applied to the matrices of all the studied attributes. Dendograms for each attribute were constructed on the basis of fusion level to investigate similarities in pattern of performance among genotypes (in response to environments) and environments (in discriminating among genotypes). The GGE biplot technique developed by Gabriel (1971) was used to make possible the display in a single graph of the performance of each genotype at each environment for each attribute. In this graphic display, each genotype is represented by a point, called a marker, defined by the genotype's scores on all PCs, and each environment is represented by a vector defined by environment's scores on all PCs. The primary and secondary effects of genotypes and environments were calculated according to the shifted multiplicative model to search for separability in crops cultivars (Cornelius *et al.,* 1992). GGE Biplot can be multidimensional, but two-dimensional biplots, using only the first and the second PCs are most common, both for biological reasons as well as for easy comprehension.

Results and Discussion

A strong across-genotype and across environments variance was found for seed yield (kg ha^{-1}) as well as oil content (%). The partitioning of the sums of squares indicated that environments accounted for 81.89% of the total sums of squares for seed yield and 51.66% of the total sums of squares for (%) oil contents (Table 2). GxE interaction variance was 4.30 and 1.06 times that of the contribution of G for yields and oil content respectively. The sums of squares of oil content components (Table 2) indicated that oil content (%) is likely to have the largest heritability.

Ward's fusion strategy of hierarchical clustering was used on sunflower GxE data of seed yield (kg ha^{-1}). Cluster analysis showed that genotypes (10) and environments (7) can be separated into 5 and 4 groups respectively of different response pattern across environments (genotypes) and across genotypes (environments) (Table 3 & 4). It clearly indicated that hybrids SF-187, Mehran-2 and Bemisal-205 are different from the remaining genotypes in their response over the environments included in the study. Also genotype clusters labeled G1 (G-101, DKS-4040, Hysun-38 and Hysun-33) and G4 (FH-106, FH-315 and Parsun-2), have different response patterns in terms of seed yield over all environments (Table 3 & Fig. 1). Similarly, environment clusters are labeled as G1 (Islamabad & D.I.Khan) and G3 (Peshawar, Faisalabad & Kot Diji) have similar response pattern w.r.t., seed yield (kg ha^{-1}) over all genotypes. Monsanto research farm, Lahore (ind-2) and Tandojam Sindh (ind-4) are different from the rest of the environments in their ability to discriminate among the genotypes (Table 4 & Fig.1).

The results of the ordination analysis are presented in biplot of the 1st and 2nd principal components for seed yield (Fig. 2). The GGE biplot is constructed by plotting the primary effect scores of each genotype (as x axis) and each environment against their respective secondary effect scores (as y axis). In these biplots, genotypes are represented by numbers and environments by vectors generating from the origin. Biplots can be used to evaluate cultivars for their yield potential and stability and to evaluate trial sites for their discriminating ability and representiveness. Entries that are close together are similar in performance across environments, while adjacent environments are similar in the way they discriminate among genotypes. High yielding and stable genotypes usually tend to be on the bottom right quadrant of the joint plot.

Table 2. Analyses of variance for seed yield (kg ha⁻¹) and oil content (%) of 10 sunflower hybrids grown in 7 different environments of Pakistan.

Grain yield (kg ha⁻¹)

Source	Sums of squares	Degree of freedom	% of total sums of squares	F-value	Prob.
Environment (E)	373650182.20	6	81.89	652.6**	0.000
Error	2003921.90	21	0.11		
Genotypes (G)	11821326.60	9	2.59	13.7**	0.000
GxE	50713754.62	54	11.12	7.8**	0.000
Residuals	18070228.60	189	3.96		
Oil contents (%)					
Environment (E)	3099.291	5	51.66	125.0**	0.000
Error	89.266	18	1.48		
Genotypes (G)	1085.810	9	18.10	34.1**	0.000
GxE	1151.992	45	19.20	7.2**	0.000
Residuals	572.768	162	9.55		

Table 3. The groups of genotypes having similar response pattern over all the environments for seed yield (kg ha⁻¹).

Groups	No.	Group members with their code numbers
Group-1	4	G-101 (G1), DKS-4040 (G5), Hyaun-38 (G2) and Hysun-33 (G6)
Indiv-2	1	SF-187 (G4)
Indiv-3	1	Mehran-2 (G3)
Group-4	3	FH-106 (G7), FH-315 (G8) and PARSUN-2 (G10)
Indiv-5	1	Bemisal-2005 (G9)

Table 4. The Groups of similar environments over all genotypes for (%) oil content.

Groups	No.	Group members
Group-1	2	Islamabad (E1) and D.I. Khan (E2)
Indiv-2	1	Monsanto research farm, Manga (E4)
Group-3	2	Faisal Abad (E6), Peshawar (E3) and Kot Diji (E7)
Indiv-4	1	Tando Jam (E5)

Figure 2 shows that genotypes G1 G4, G5 and G6, being in the bottom right quadrant, gave the highest average yields (highest primary scores) and were defined as widely adapted over the sites (small secondary scores) as described by de la Vega *et al.*, (2001). Genotypes G5 and G6 are closely related in their yield potential as well as their response to varying environments as indicated by acute angle formed between G5, origin and G6 as reported by Kroonenberg (1997). Similarly, G5 is closely associated with G1 in its behavior regarding yielding ability and across sites performance. Genotype G2 can also be termed as high yielding and stable because of its large primary effect scores and almost near zero secondary scores. Entries G3, G7 and G8 had the yields which were below the average (primary scores < 0) and were highly unstable (large absolute secondary value). Genotype G3 having a large distance from the origin has a large genotype and plus interaction effect. Genotype G2 gave higher yield than the average but was not very stable, as evidenced by its relatively large secondary effects as is evident from Fig. 2. Cultivars G9 and G10 lying in the left hand bottom quadrant gave yield above average but were highly unstable.
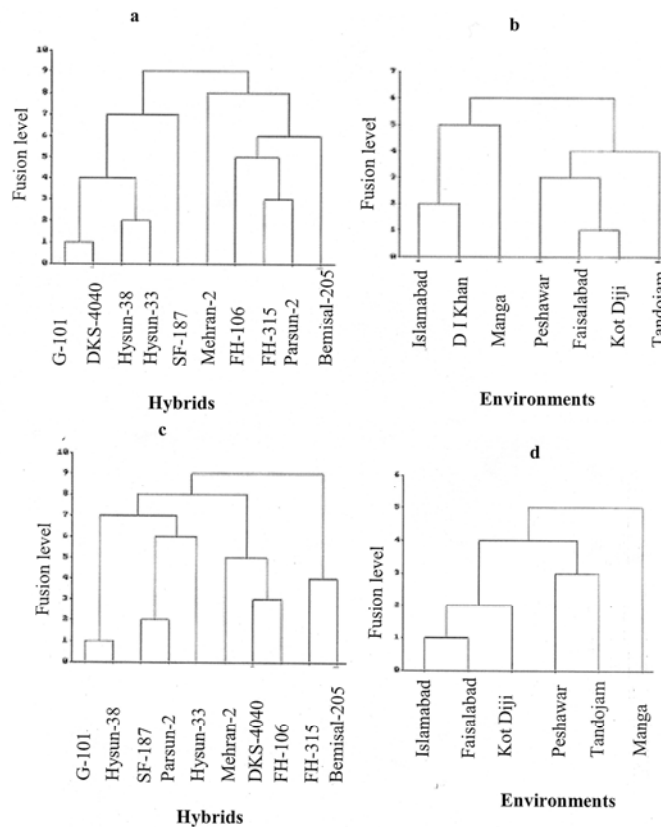
Fig. 1. Dendograms for the classification of 10 sunflower hybrids tested in seven environments for yield kg ha^{-1} (a,b) and six environment for oil content (%) (c,d).
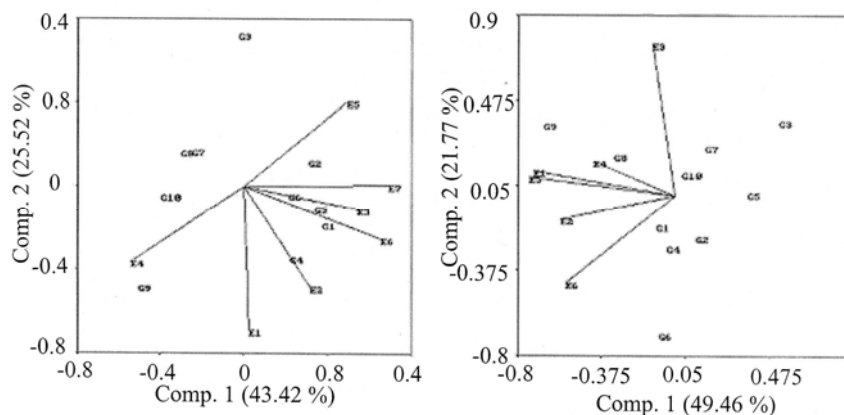

Fig. 2. Biplot of the 1st and the 2nd principal components for yield kg ha^{-1} and oil content (%) (b) of 10 sunflower hybrids grown in seven environments. Genotypes are represented by numbers and environments by vectors.
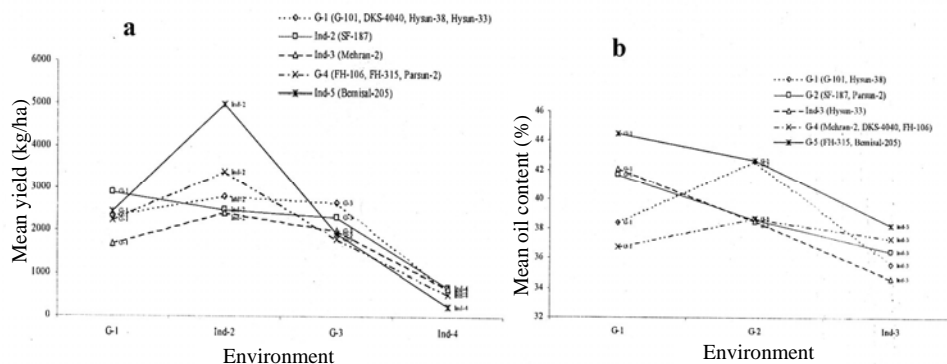
Fig. 3. Performance plots of genotype groups vs. environmental groups based on seed yield kg ha^{-1} (a) and oil content (%) (b).

Table 5. The groups of genotypes having similar response pattern over all the environments for (%) oil content.

Groups	No.	Group members with their code numbers
Group-1	2	G-101 (G1) & Hysun-38 (G2)
Group-2	2	SF-187 (G4) & Parsun-2 (G10)
Indiv-3	1	Hysun-33 (G6)
Group-4	3	Mehran-2 (G3), DKS-4040 (G5) & FH-106 (G7)
Group-5	2	FH-315 (G8) & Bemisal-205 (G9)

Table 6. The groups of similar environments over all genotypes based on (%)oil content.

Groups	No.	Group members
Group-1	3	Islamabad (E1) Faisalabad (E5) and Kot Diji (E6)
Group-2	2	Peshawar (E2) and Tando Jam (E4)
Indiv-3	1	Manga Research Farm (E3)

As regards the cluster analysis of environments, E3, E6 and E7 are included in a separate group (Table 4 & Fig. 1). The maximum angle among the vectors of these environments is far smaller than 90° (Fig. 2). This suggests that these environments are relatively similar in the manner they discriminate among genotypes as reported by Kroonenberg (1995). These environments have large primary effects and would thus facilitate identification of genotypes with better general adaptation. Therefore, from the perspective of selection for high yielding and stable genotypes as is evident from the findings of Yan *et al*., (2001), the best environments should have large primary effect scores, but near zero secondary effects (Fig. 2). So in the above-mentioned environments, E7 and E3 can be termed as best to be used for selecting varieties or hybrids. Environments E4 and E5 have angle around 180° suggesting that they tend to be two different mega-environments and thus indicating larger interactions with the genotypes in them (Fig. 2), as reported by de la Vega (2001).

Performance plots are used to illustrate each genotype group's performance in a series of environment groups based on cluster analysis. Genotypes group G1 consisted of four entries (1, 2, 5, & 6) showed average to good performance at environment groups

G1, ind-2, G3 and poor at environment ind-4 (Fig. 3a). Same pattern was noted in genotypes ind-2 and ind-3 over all the environment groups. The next genotype group G4 having three entries (7, 8 & 10) performed exceptionally well at environment ind-2 and very poorly at environment ind-4 (Fig. 3a). Genotype ind-5 gave highest yield at environment ind-2 but yielded poorly than the other genotype groups at environment ind-4 as is evident from Fig. 3. All the individual genotypes and genotype groups gave relatively high yield at environment (ind-2), confirming their adaptation to this particular environment. But on the other hand, they showed a poor relative performance at environment (ind-4), indicating their poor adaptation to this environment.

Cluster analysis of oil content indicated that the genotypes can be grouped into 5 classes of different performance over the sites included in the study (Table 5). Similarly, environments have been divided into three clusters based on the way they discriminate against the genotypes for oil contents (%) as is shown in Table 6. Entries G1, G2, G4, G5, G7 and G10 which are close together in the biplot (Fig. 2b), have stable and similar performance behavior across environments while environments E1 (Islamabad) and E5 (Faisalabad), having an acute angle between are very closely related in the way they discriminate among genotypes. The angle between environments E3 (Monsanto research farm, Manga) and E6 (Kot Diji, Sindh) tends to be larger than 90°, which suggests that these are two different environments as reported by de la Vega, (2001), regarding their response over the genotypes for % oil content, (Fig. 2b).

The response plots of five genotype groups for % oil content indicated different patterns of performance across environment groups (Fig. 3b). Group G1 including two entries (G1 & G2) performed relatively well at environment group G1, exceptionally well at G2 but poorly at environment ind-3. The genotype group G4 (having entries G3, G5 & G7) showed relatively low but stable values for % oil content across the environment groups (Fig. 3b). Genotype group G5 having two hybrids 8 & 9 proved to be adapted to all environment groups by giving highest yield at every environment group. Environment (ind-3) recorded lowest % oil content regardless of the genotype groups (Fig. 3b).

References

Arshad, M., A. Bakhsh, A.M. Haqqani and M. Bashir. 2003. Genotype –Environment Interaction for grain yield in Chickpea (*Cicer arietinum* L.). *Pak. J. Bot*., 35(2): 181-186.

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

De la Vega, A.J. and S.C. Chapman. 2001. Genotype by environment interaction and indirect selection for yield in sunflower. II. Three mode principal component analysis of oil and biomass yield across environments in Argentina. *Field Crops Res*., 72: 39-50.

Gabriel, K.R. 1971. The biplot graphical display of matrices with applications to principal component analysis. *Biometrica*, 58: 453-467.

Granlund, M. and D. C. Zimmerman. 1975. Effect of drying conditions on oil contents of sunflower (*Helianthus annuus* L.) seed determined by wide-line Nuclear Magnetic Resonance (NMR). *North Dakota Acad. Sci. Proc*., 27(2): 128-132.

Gwanama, C., M.T. Labuschagne and A.M. Botha. 2000. Analysis of genetic variation in *Cucurbita moschata* by random amplified polymorphic DNA markers. *Euphytica*, 113: 19-24.

Khan, I.A., B.A. Malik and M. Bashir. 1988. Investigation of genotype X environment interaction for seed yield in chickpea (*Cicer arietinum* L.). *Pak. J. Bot*., 20(2): 201-204.

Khan, I.A., B.A. Malik and M. Tahir 1987. Phenotypic stability for yield in chichpea. *Pak. J. Sci. Ind. Res*., 30(6): 455-456.

Kroonenberg, P.M. 1997. Introduction to biplots for GXE tables. Research Report No. 51. Centre for statistics, the Uni. of Queensland, Brisbane, Qld.

Ward, J.H. 1963. Hierarchical grouping to optimize an objective function. J. *Am. Stat. Assoc.*, 58, 236-244.

Weikai Yan, Paul L. Cornelius Jose Crossa and L.A. Hunt. 2001. Two types of GGE biplots for analyzing multi-environmental trial data. *Crop Sci.*, 41: 656-663.

Williams, W.T. 1976. *Pattern Analysis in agricultural sciences*. Elsevier, Amsterdam.

Yan, W. and L.A. Hunt. 2001. Genetic and environmental causes of GXE interaction for winter wheat yield in Ontario. *Crop Sci.*, 41: 19-25.

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