



Interpretation of $G \times E$ interaction for wheat yield in Haryana using GGE Biplots

Bhushan Kumar, Ekta Hooda & B K Hooda*

Department of Mathematics & Statistics

CCS HAU-Hisar (Haryana)

*Corresponding author E-mail: bkhooda@gmail.com

ABSTRACT

Genotype main effect (G) plus genotype-by-environment (GE) interaction i.e. (G+GE) biplot analysis has received a wide applicability in crop breeding improvement programs. It possesses an extra property in evaluation of test environment by discriminating power versus representativeness view and is found superior than AMMI in mega-environment analysis and genotype evaluation. A genotype is considered superior if it has high yield potential in favourable environment and at the same time has a great deal of phenotypic stability. The present study is intended towards the application GGE Biplots for interpretation of genotype versus environment interaction data for wheat rain yield in Haryana. Mega-environment analysis and genotype evaluation has been conducted for 24 genotypes of wheat evaluated at 5 locations of Haryana (Hisar, Karnal, Kaul, Bawal and Sirsa) during 2010-11 cropping seasons under irrigated conditions. A simple R based algorithm has also been suggested for construction of Biplots using R software.

Key Words: Singular Value Decomposition, $G \times E$ interaction, GGE Biplots, Mega-environment analysis, genotype evaluation.

Received 22.03.2018

Revised 22.05.2018

Accepted 11.06.2018

INTRODUCTION

The indigenous cropping pattern is not sustainable as farmers in developing and undeveloped countries want maximum benefit from agriculture through the increased crop productivity. They are always searching for the high yielding varieties. But a high yielding variety in one region may not be high yielding in one or more other regions. The reason of non-uniform performance of a variety is not only due to change in environment but also because of the genotype and environmental interaction. The $G \times E$ interaction is a complex factor and makes the breeding program comprehensive and expensive. In particular, the creation of different genotypes from parents and testing their adaptability in a region makes it multi-year research program. Therefore, breeders are conducting varietal trials for major crops for the targeted regions all over the world in search of new superior genotypes or seeds to meet the climatic and environmental challenges.

Breeders conduct multi-location trials to get improved genotypes in their varietal developmental programmes. They evaluate genotypes on the basis of yield and recommend cultivar for production in a region. The ANOVA technique is not of too much help in selection of cultivars because it does not consider the positivity and negativity of factors. It may happen that a certain genotype may perform better than average in all environments or may be a bad performer in all environments or performing better in some environments and poor in others. It may also happen that the most of the variation is due to performance of genotypes in some specific environments. If the three factors of yield (G, E, and GE) simultaneously come in analysis for every pair of genotype and environment then selection of cultivars becomes simple for the targeted region. It is $G \times E$ which makes the selection of cultivar complex. If $G \times E$ is nonsignificant then a genotype performing best in one environment will perform best in all environments. And there is no need to conduct trials in all environments. But such a situation seldom occurs. Though variation in yield due to environment is greater than variation due to genotype and $G \times E$ interaction but $G+GE$ are relevant to cultivar evaluation as indicated by Yan and Tinker [6].

The genotype by environment data are represented in matrix form in which row represents genotype and column represents environment. Introduced by Gabriel [3], biplot is a pictorial representation of a matrix in a plane by vectors for each row and each column such that scalar product of a row vector and a column vector is the corresponding row-column element of the matrix. In fact, biplot is fusion of two plots, one plot of row factors or genotypes and other plot of environments or column factors. Bradu & Gabriel [2] applied biplot to agricultural data from a cotton performance trial to illustrate their diagnostic role for model selection. Yan *et al.* [5] used GGE biplot to cultivar evaluation and mega-environment investigation with primary goal to identify superior cultivar for the target region.

According to Yan & Tinker [7] a GGE biplot not only provide effective evaluation of genotypes but also allow for a comprehensive understanding of the target and test environments. They further indicated that a GGE biplot is helpful in understanding the target environment as a whole whether it consists of single or multiple mega environments. Yan *et al.* [8] concluded that genotype main effect (G) plus genotype-by-environment (GE) interaction i.e. (G+GE) biplot analysis has wider adaptability in breeding programs and is superior to AMMI in mega-environment analysis and genotype evaluation. It possesses extra property in evaluation of test environment by discriminating power versus representativeness view which is not possible in AMMI biplot. Grange *et al.* [4] showed that the 'R' package BiplotGUI provides graphical user interface for construction, interaction, and manipulation of biplots. Bishnoi and Hooda [1] studied yield Stability and Association among Parametric and non-Parametric Stability Measures for Wheat in Haryana. Therefore, the present study has been intended towards the construction and application GGE biplots for interpretation of genotype versus environment interactions data for wheat yield in Haryana. Mega-environment analysis and genotype evaluation has been conducted for 24 genotypes of wheat evaluated at 5 locations of Haryana (Hisar, Karnal, Kaul, Bawal and Sirsa) during 2010-11 cropping seasons under irrigated conditions. A simple R based algorithm has also been given for construction of Biplots using R software.

MATERIAL AND METHODS

Data and Methodology

Data

The experimental data on 24 genotypes of wheat evaluated at 5 locations of Haryana (Hisar, Karnal, Kaul, Bawal and Sirsa) in RBD with four replications during 2010-11 cropping seasons under irrigated conditions were used for the present investigation. The multi-environmental trials were conducted by the Wheat section, Department of Genetics and Plant Breeding, Chaudhary Charan Singh Haryana Agricultural University, Hisar. The Mean grain yield of wheat genotypes evaluated at five locations of Haryana is given in Table 1.

Table 1: Mean grain yield (q/ha) of wheat genotype at five locations of Haryana during 2010-11

Genotypes	Hisar	Kaul	Bawal	Karnal	Sirsa
WH 1080	47.08	49.31	54.61	50.43	40.81
WH 1081	47.08	49.31	60.92	61.54	45.95
P 11638	56.25	43.06	57.68	57.69	40.21
P 11850	67.29	42.71	58.48	70.94	51.09
P 11852	46.88	52.08	60.92	66.67	48.67
P 11898	30.42	44.79	53.18	62.82	42.93
P 11933	53.33	41.67	54.91	63.46	47.91
P 11937	55.42	39.58	61.52	56.41	49.58
P 11962	43.96	50.00	56.25	51.92	52.30
P 12025	56.25	41.32	59.97	70.51	49.27
P 12026	46.25	40.63	55.95	50.00	41.72
P 12032	56.67	49.31	60.64	66.67	52.60
P 12033	47.92	43.40	60.49	62.61	50.79
P 12039	54.17	47.22	59.52	64.53	53.20
P 12093	57.50	47.22	56.03	64.32	52.60
P 13014	38.75	44.79	56.25	55.56	59.55
P 13016	50.42	48.61	58.04	63.25	63.18
P 13017	42.50	47.22	54.32	60.90	55.32
DBW 17	56.25	38.19	64.14	63.89	54.11
WH 711	52.92	32.29	59.38	50.00	48.97
PBW 343	49.38	43.40	60.33	57.48	44.14
PBW 550	50.63	47.92	58.32	58.97	50.18
PBW 621	48.33	50.69	57.44	67.09	52.90
HD 2967	60.42	45.49	59.38	62.82	57.13

Model for GGE Biplot Analysis

Let $\mathbf{Y} = (y_{ij})_{n \times p}$ be the GE data matrix representing the mean grain yield of $n = 24$ wheat genotypes evaluated at $p = 5$ locations or environments (Hisar, Karnal, Kaul, Bawal and Sirsa) of Haryana. In terms of effects, the basic model for constructing a GGE biplot from GE data is given by

$$y_{ij} = \mu + g_i + e_j + \phi_{ij} + \varepsilon_{ij} \quad (2.2.1)$$

where, y_{ij} is the average yield of genotype i ($i = 1, 2, \dots, n$) in environment j ($i = 1, 2, \dots, p$), μ is overall or grand mean, g_i genotypic main effect, e_j environmental main effect, ϕ_{ij} is interaction between g_i and e_j and ε_{ij} is the residual of the model associated with the genotype i in environment j . A GGE biplot is constructed by subjecting the environment-centered GGE Data to singular value decomposition (SVD). The GGE data matrix is decomposed into three component matrices as

$$\mathbf{Y} = \mathbf{ULV}^T \quad (2.2.2)$$

Where \mathbf{U} ($n \times p$) and \mathbf{V} ($p \times p$) are column orthonormal matrices, i.e. $\mathbf{U}^T\mathbf{U} = \mathbf{I} = \mathbf{V}^T\mathbf{V}$ and \mathbf{L} is the diagonal matrix of non-zero eigen values of $\mathbf{Y}\mathbf{Y}^T$ or $\mathbf{Y}^T\mathbf{Y}$. The columns of \mathbf{U} (**genotype eigenvector matrix**) are eigenvectors of $\mathbf{Y}\mathbf{Y}^T$ and columns of \mathbf{V} (**environment eigenvector matrix**) are eigenvectors of $\mathbf{Y}^T\mathbf{Y}$.

The model for a GGE biplot based on SVD for first two principal components is given by

$$y_{ij} - \mu - \bar{y}_j = \xi_{i1}\lambda_1\eta_{1j} + \xi_{i2}\lambda_2\eta_{2j} + \varepsilon_{ij} \quad (2.2.3)$$

where, y_{ij} is the average yield of genotype i in environment j , \bar{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_{1j} and η_{2j} are the PC₁ and PC₂ scores, respectively for environment j .

To display PC₁ and PC₂ in a biplot, the equation (2.2.3) may be rewritten as

$$y_{ij} - \mu - \bar{y}_j = \xi_{i1}^*\eta_{1j}^* + \xi_{i2}^*\eta_{2j}^* + \varepsilon_{ij} \quad (2.2.4)$$

Where, $\xi_{ir}^* = \lambda_r^k \xi_{ir}$ and $\eta_{rj}^* = \lambda_r^{1-k} \eta_{rj}$, with $r = 1, 2$ and $0 \leq k \leq 1$. GE biplot is constructed by using scores derived from the first two PCs and plotting ξ_{i1}^* and η_{1j}^* against ξ_{i2}^* and η_{2j}^* in the same scatter plot.

Algorithm for Construction of GGE Biplots Using R Package

The prefix 'bi' in biplots refers to the fusion of two plots, one for genotypes and other for locations. Biplot dimensions may be of higher order but their interpretation is complex. Therefore, in practice, only two dimensional biplots are used for analysis and interpretation of $G \times E$ data. Many packages are available for construction of Biplots these days. The UBbiplot package is used for construction of biplots in general while GGEbiplotGUI package for GGE biplot in R system. GGEbiplotGUI may be installed from R cran Package while UBbiplot can be downloaded from www.wiley.com. We suggest the following algorithm for the purpose of construction of biplots using R package.

Step-1: Load the R software of version 3.2.5 or later versions

Step-2: Install GGEbiplotGUI

Step-3: Load packages after starting R in sequence MASS, rgl, tcltk, tkrplot, deldir, numDeriv and GGEbiplotGUI

Step-4: Save data in csv format in excel in working directory and read in R

Step-5: USE function GGEbiplot(data) and run

Step-6: A short screen will open for GGEbiplot(data)

Step-7: Construct biplot by selecting parameters on short screen

RESULTS AND DISCUSSION

The analysis of variance is an important preliminary analysis for confirming the presence of genotype \times environment interaction. The sum of squares and the percent sum of squares along with respective degrees of freedom are presented in Table 2. The analysis of variance of the data showing the sum of squares and the percent sum of squares along with respective degrees of freedom are presented in Table 2.

Table 2: Analysis of variance of $G \times E$ data for Haryana

Source	d.f.	S.S.	% S.S.
Genotypes	23	1205.87	15.85
Environments	4	3998.16	52.57
$G \times E$	92	2402.05	31.58
Total	119	7606.07	100

The effects of genotypes, environments and genotype x environment interaction were observed through combined analysis of variance. Analysis indicates that the maximum variability in grain yield of wheat due to environments (52.57%) followed by G × E interaction is (31.58%) and genotypes (15.85%).

The first two principal components explained 73.51% of the total GGE variation obtained through the SVD of the data converted into GGE data by extracting environments' main effect according to equation (2.2.4). The first principal component explained 46.73% of variation while 26.78% variation was explained by the second principal component. A biplot of G × E data quantifies the factors i.e genotypes and environments and presents these factors as vectors on a plot. The dot product of a pair of genotypes and environment factors provides yield of that factor. In this analysis we use GGE biplot to discuss the following six aspects in reference to the wheat yield data from five locations of Haryana:

- i) Similarity and dissimilarity among genotypes
- ii) Relationships among environments
- iii) Mega-environments analysis (' Which one Where' GGE Biplot)
- iv) Evaluation of test environments
- v) Evaluation of genotypes

Similarity and dissimilarity of genotypes based on GGE Biplot

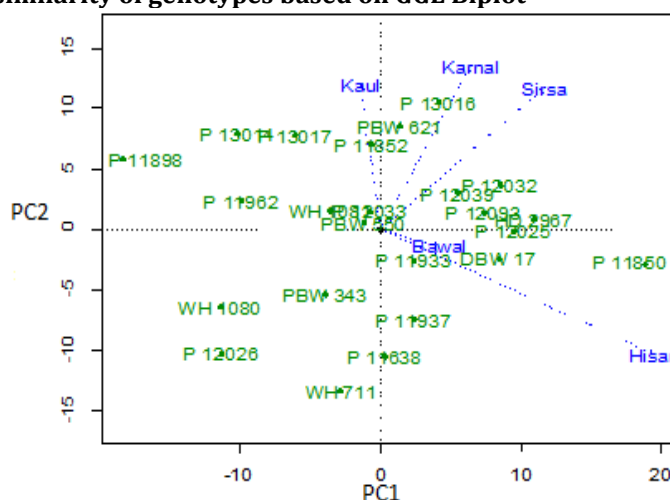


Figure 1 Differentiation of genotypes in GGE biplot for Haryana wheat data

Figure 1 is a GGE biplot of wheat yield data in Table 2 which consist of 24 genotypes and five environments of Haryana. This is row metric preserving biplot and preserves the properties of rows (genotypes) of the data. It indicates that the distance between two genotypes in the biplot approximates the difference between them and can be used for identification of similar genotypes of wheat in Haryana at five locations. For instance, the genotypes P12039, P12032 and P12093 are similar while P11850, P13014 and WH711 are of dissimilar. On the same pattern, many other groups can be identified as similar and dissimilar visualizing only distances among them in Figure 1.

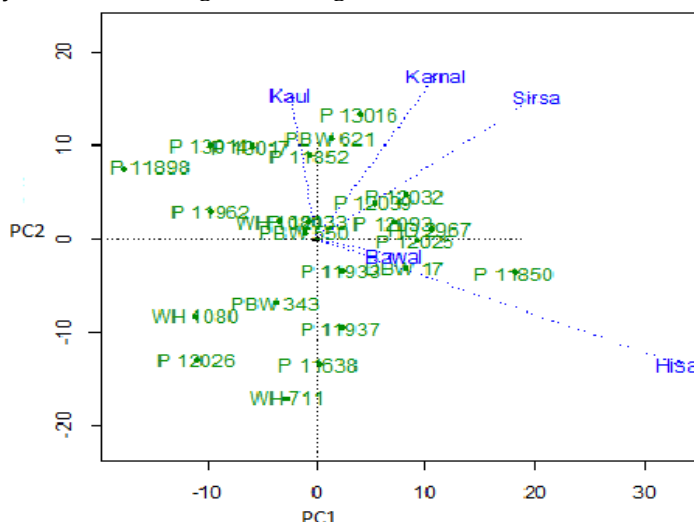


Figure 2 Differentiation of environments in GGE biplot for Haryana wheat data

Figure 2 is column metric preserving GGE biplot of data in Table 2. The correlation between two environments remains preserved since the cosine of the angle between two environments vectors either in row metric preserving or column metric remains unaltered. It is obvious in both biplots that respective angles between environments are equal. It seems that environments' vector in Figure 1 and Figure 2 are same but it is not so because scales of axes are different. Figure 2 indicates that the environments Bawal and Hisar are highly correlated and are similar as for as wheat yield is concerned. Since this biplot is column metric preserving, it explains the properties of environments where lengths from the origin are standard deviations of environments i.e. capacity of differentiation of genotypes. It is observed that the location Hisar and Karnal are superior while Bawal is inferior in genotypes differentiation.

Mega-environments analysis ('Which one Where' GGE Biplot)

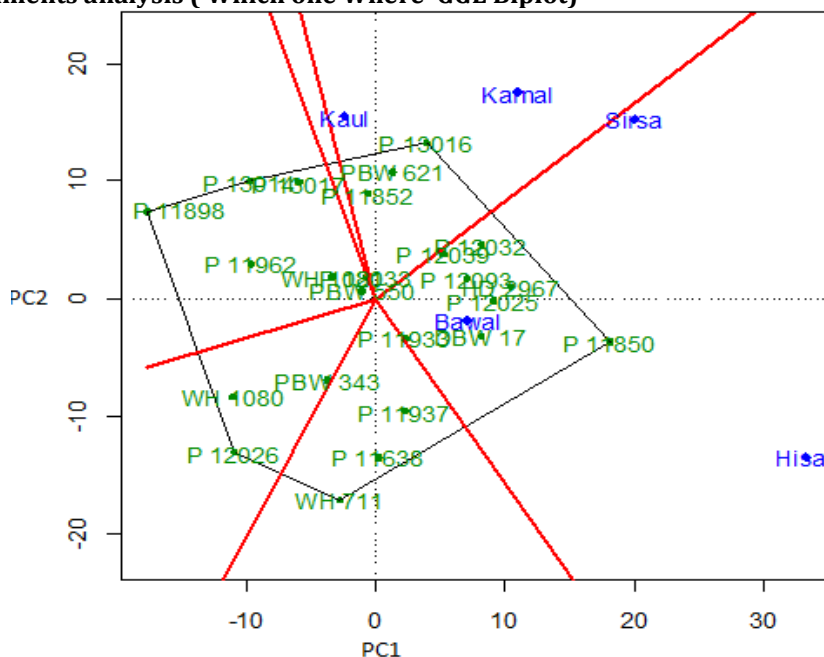


Figure 3 'Which won where' GGE Biplot for Haryana wheat data

For mega-environment analysis, a GGE biplot is constructed by plotting PC1 scores of the genotypes and the environments against their respective scores for PC2 which result from SVD of environment standardized G x E data. The 'which one where' view of the GGE biplot in Fig 3 is an effective tool for mega-environment analysis. It consists of an irregular convex polygon formed such that all genotypes come inside the polygon and a set of perpendicular lines drawn from the biplot origin to all sides of polygon. The vertices of the polygon are the genotypes located farthest away from the biplot origin in different directions. The genotypes and environments in between two perpendicular lines provide positive yield with the environments since angle between any pair of genotype and environment in this region is less than 90 degree. Also, the genotype which is at vertex of the polygon in this region provides maximum yield and each sector formed by the perpendicular lines has a superior genotype at vertex than any other genotype in this sector. The total environment is thus divided into different sectors which have their own superior genotypes. These superior genotypes are called winner genotypes in respective regions. The sectors formed by perpendicular lines enclose similar type environment which has specific performer and those environments which falls in this region are called mega-environment. The biplot in Figure 3 provides superior or winner genotypes in their respective mega environments. The biplot consists two mega-environments identified as Sirsa-Bawal-Hisar consisting one mega environment while Karnal-Kaul form the second mega-environment. It is observed that the genotype P13016 is the winner genotype for the mega-environment Sirsa-Bawal-Hisar, whereas P11850 for the mega-environment Karnal-Kaul.

Evaluation of test environment

Multi-environment trials are generally multi-year programs conducted in large regions to extract more and more information about genotypes and environments. To improve efficiency of trials the experimental locations are evaluated in terms of their representativeness and discriminativeness.

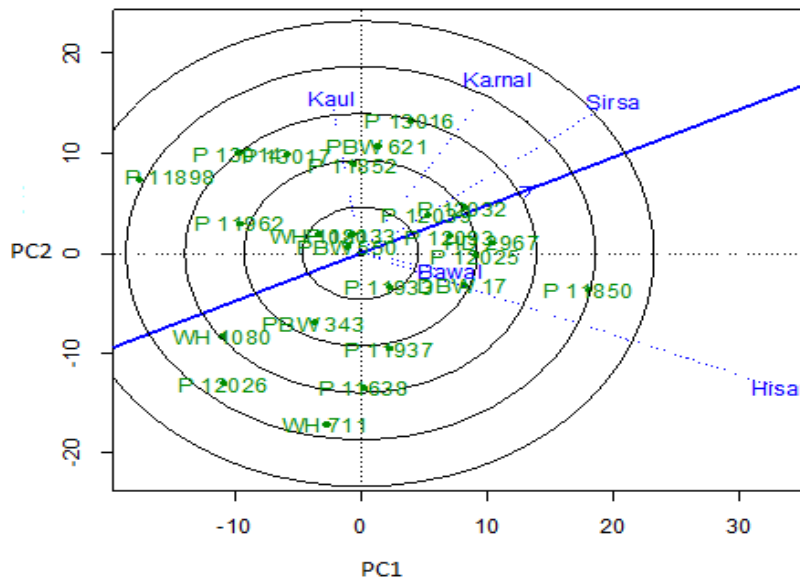


Figure 4 Evaluation of environments for Haryana wheat data

Representativeness determines the closeness of test location to the target environment while discriminativeness is the ability of test environment in differentiating the genotypes in context of yield variation. More the variation comes in genotypes performances more discriminating is the test location and more is the information about genotypes. For evaluation of representativeness, target environment is plotted by taking average of all environments where the angle between target and test environment indicates representativeness of the target environment by the test environment. The discriminativeness is observed by the variance of the environments measured in terms of the lengths of the environment vectors. More the variance of environment more is the discriminating power of environment for genotypes. It is observed that (Fig 4) the environment Sirsa has highest representativeness of the experiment while Hisar has higher discriminating power but low representativeness. The Kaul environment has the least representativeness while Bawal has least discriminating power.

Evaluation of genotypes

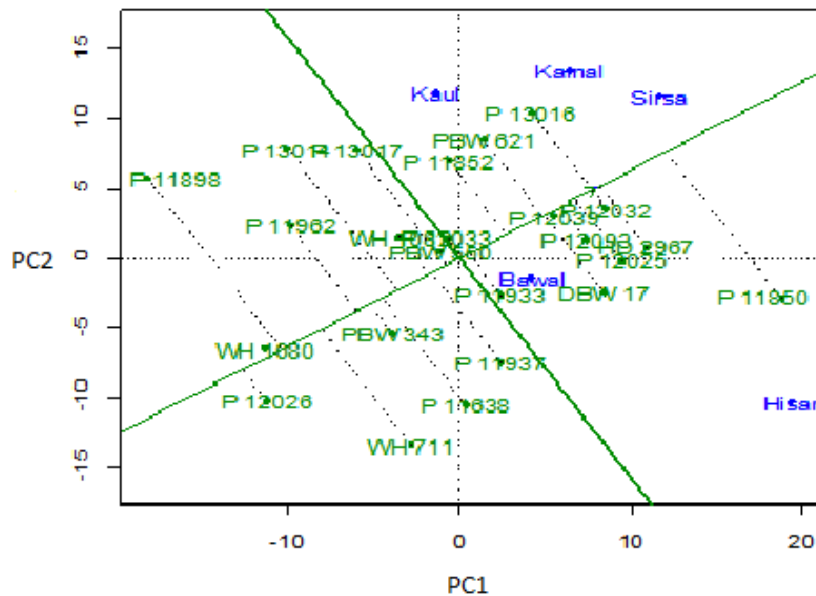


Figure 5 Mean versus stability biplot

Genotypes evaluation is conducted for each mega-environment under the assumption that the mega-environments in Figure 3 are repeatable across years. Genotypes or varieties in breeding programs are

evaluated both for yield and stability within a mega-environment. An ideal genotype [8] should have both high performance in yield and high stability within its mega-environment.

For general release of a breed, evaluation of a genotypes is performed with respect to average performance and stability of all genotypes. The test environment evaluation axis (Figure 4) is useful for this purpose. The axis passing through this virtual environment is called average environment axis (AEA) while a perpendicular axis overlaid on the GGE biplot is called average coordination axis (AEC). According to Yan *et al.* [8] AEA is highly correlated with genotypic performance and indicates genotypes average performance visualized by the projection of a genotype on this axis. More the projection of genotype on AEA more is its mean yield in trials while more projection on AEC indicates more interaction or instability. Since interaction may have either positive or negative effect, stable genotypes with high average yield are selected for further considerations. So genotypes towards target environment from origin and nearby the AEA axis are in race of selection. The present analysis (Figure 5) suggests that the genotypes P12039 and P12032 are favorable for the experimental region considering both average yield and stability of genotypes. Mean yield and stability are simultaneously considered as factors in evaluation of genotypes in Figure 3.5. Simultaneous accounting of factors together is something complex for genotypes in visual applications. This problem is accomplished by defining ideal genotypes. The ideal genotype is a virtual genotype which has highest mean yield and zero instability. The distances from ideal genotype decreases either mean yield or stability or both. So distances are considered as indicator of ranks in evaluation of genotypes. On biplot in Figure 3.6, it is shown by drawing circles from ideal genotype to view distances of different genotypes by inspection. The centre of lowest radius circle is ideal genotype from which relative distances may be observed. P12032 is the most favorable then HD2967 and P12039, P12093 and P13016 are nearly equally preferable since they lie on the boundary of same circle in view of wheat production in five locations of Haryana.

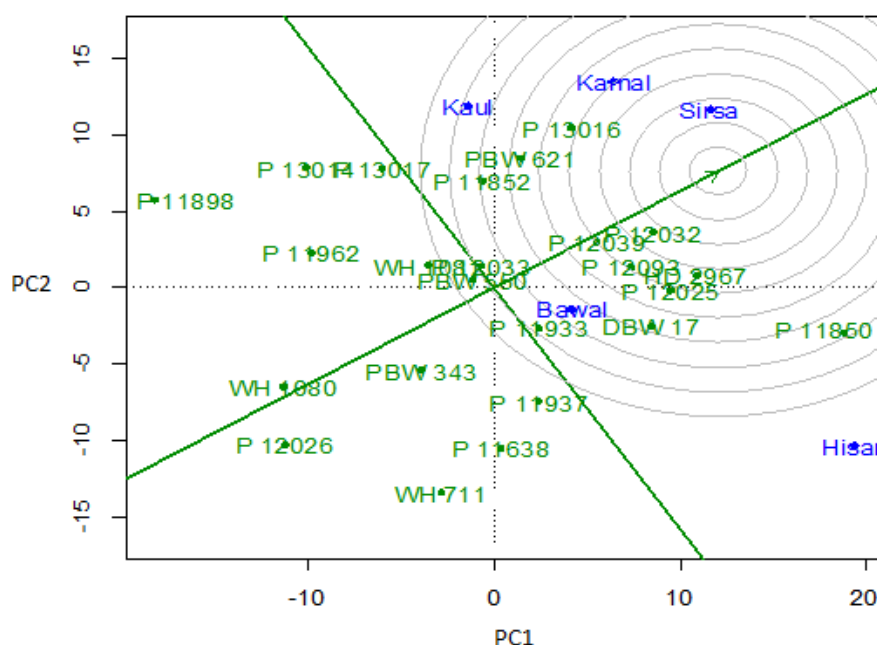


Figure 6 Ranking of genotypes in biplot

ACKNOWLEDGEMENTS

The authors are grateful to Dr. S. S. Dhanda, Ex- Plant Breeder, Department of Genetics and Plant Breeding CCS Haryana Agricultural University, for providing the data used in this paper. Help and cooperation of Dr. Dhanda is dully acknowledged.

REFERENCES

1. Bishnoi, S. & Hooda, B.K. (2018). Yield stability and association among parametric and non-parametric stability measures for wheat (*Triticumaestivum*) genotypes in Haryana. *Int. Jour. of Agril. and Stat. Sci.*, **14** (1): 323-330.
2. Bradu, D. & Gabriel, K.R. (1978). The biplot as a diagnostic tool for models of two-way tables. *Techno.*, **20**: 47-68.
3. Gabriel, K.R. (1971). The biplot graphic display of matrices with application to principal component analysis. *Biomet.*, **58** (3): 453-465.
4. Grange, A.L., Roux, N.L. & Lubbe, S.G. (2009). Biplot GUI: Interactive biplots in R. *J. of stat. soft.* **30** (12): 1-37.

5. Yan, W. & Kang, M.S. (2003). GGE biplot analysis: A graphical tool for breeders, geneticists, and agronomists. CRC Press, Boca Raton, FL.
6. Yan, W. & Tinker, N.A. (2005). An integrated biplot system for displaying, interpreting, and exploring genotype by environment interactions. *Crop Sci.*, **45**:1004-1016.
7. Yan, W. & Tinker, N.A. (2006). Biplot analysis of multi-environment trial data. *Canad. J. of Plant Sci.*, **86**:623-664.
8. Yan, W., Kang, M.S., Ma, B., Woods, S. & Cornelius, P.L. (2007). GGE Biplot vs. AMMI Analysis of Genotype-by-Environment data. *Crop Sci.*, **47**:641-653.

CITATION OF THIS ARTICLE

Bhushan Kumar, Ekta Hooda & B K Hooda. Interpretation of G × E interaction for wheat yield in Haryana using GGE Biplots. *Bull. Env. Pharmacol. Life Sci.*, Vol 7 [9] August 2018: 28-35