PLANT BREEDING AND SEED SCIENCE
Volume 66 2012

DE GRUYTER OPEN DOI: 10.2478/v10129-011-0063-5

Naser Sabaghnia[1*], Mohtasham Mohammadi[2], Rahmatollah Karimizadeh[2]

[1]Department of Agronomy and Plant Breeding, Faculty of Agriculture, University of Maragheh, Maragheh, Iran.; [2]Dryland Agricultural Research Institute (DARI), Gachsaran, Iran.
[*]Corresponding author: e-mail sabaghnia@maragheh.ac.ir, sabaghnia@yahoo.com

CLUSTERING DURUM WHEAT GENOTYPES IN MULTI-ENVIRONMENTAL TRIALS OF RAIN-FED CONDITIONS

ABSTRACT

For durum wheat genotypes evaluation in multi-environmental trials (MET), measured seed yield is the combined result of effects of genotype (G), environment (E) and genotype by environment GE interaction. The GE interaction structure can be identified if the data are stratified into homogeneous subsets through cluster analysis. A combined analysis to assess GE interactions of 20 durum wheat genotypes across 14 environments was undertaken. The combined analysis of variance for E, G and GE interaction was significant, suggesting differential responses of the genotypes in various environments. Four cluster methods, which differ in the dissimilarity indices depending on the regression model or ANOVA model, were used. According to dendograms of regression methods there were 10 different genotypic groups based on G (intercept) and GE (line slope) sources and 3 different genotypic groups based on GE (line slope) sources. Also, the dendograms of ANOVA methods indicated 11 different genotypic groups based on G and GE sources and 13 different genotypic groups based on GE sources. The above mentioned genotypic groups were determined via F-test as an empirical stopping criterion for clustering. Due to the high values of regression's determination coefficient which ranged from 92.6 to 99.4, using of the linear regression-based clustering was more practical. The genotypes clustering based on similarity of linear regression parameters or ANOVA model indicated that there were considerable variations among durum wheat genotypes and there are different with each other in response to environmental changes. Such an outcome could be regularly applied in the future to clattering durum wheat genotypes and other crops based on regression or ANOVA models in the Middle East and other areas of the world .

Key words: GE interaction, genotypes grouping , *Triticum durum*, seed yield

Communicated by Henryk J. Czembor

INTRODUCTION

Plant breeders perform multi-environmental trials (MET) to study yield stability and identify the most favorable genotypes for target sites. Yield stability in unpredictable environmental conditions can be due to individual, population buffering or both of them (Allard and Bradshaw, 1964). In most of MET, genotype by environment (GE) interaction is commonly observed as the differential ranking of genotypes yields among sites, years and their combinations (environments). Differential genotypic responses to variable environmental conditions limit the identification of superior genotypes especially when involved with changes in genotypic ranking. Identification of causal factors of the GE interaction and its quantification is of prime importance for selecting the most stable genotypes for specific recommendation (Signor *et al.* 2001). There are three important methods for plant breeders to GE interaction analysis in MET data which are based on analysis of variance (Shukla, 1972), principal components analysis (Perkins, 1972) and linear regression model (Finlay and Wilkinson, 1963).

Identifying genotypes that are similar to each other in response to environmental changes but different from genotypes in other groups can be intellectually satisfying, profitable, or sometimes both. Cluster analysis classifies genotypes into categories and its final goal is to identify the actual groups. Although, the cluster analysis does not identify a particular statistical method but it often doesn't need to make any assumptions about the underlying distribution of the data. Although analysis of variance as well as linear regression analysis are used for analyzing two-way data but, they do not give breeders which factor level is responsible or how their responses differ. To achieve these goals, several cluster analyses have been developed which some of them classify individuals for the similarity according to the one-way method (Edwards and Cavalli-Sforza, 1965; Callinski and Corsten, 1985); and some others classify individuals for similarity of interactions based on the two-way method (Lin and Thompson, 1975; Lin and Butler, 1990).

There are two major strategies for grouping genotypes according to their response to environmental factors changes. The first strategy was used by Abou-El-Fittouh *et al.* (1969) which they regarded genotype as a vector of *n* attributes indicated by *n* environments and to use the distance coefficient. Also in this strategy Mungomery *et al.* (1974) has used the squared distance as a similarity index for clustering. In the second strategy, Lin and Thompson (1975) used the deviation mean square from regression analysis of GE interaction (Finlay and Wilkinson, 1963) as dissimilarity index for clustering. As an alternative procedure in the first strategy Lin (1982) used the GE interaction mean square as dissimilarity index for genotypes classification through a slight adjustment of distance coefficient in Abou-El-Fittouh *et al.*

(1969) method. The dissimilarity index of Lin and Thompson (1975) benefits both genotype main effect and GE interaction effects while Lin and Butler (1990) introduced a new dissimilarity index according to regression analysis which benefits only genotype main effect. Also, Lin and Butler (1990) proposed a new dissimilarity index based on mean square of only GE interaction in contrast of dissimilarity index of Lin and Thompson (1975) which uses both effects of genotype and GE interaction in ANOVA procedure.

In every cluster analysis having a well-defined stopping criterion which known as cutoff point is important. The cutoff point must be determined in such a way that it can significantly differentiate two clusters with regard to measured traits relevant to both clusters. The cutoff point is critical because, it is very important to derive the right cut-off point to reduce the risk of Error type II. The cutoff point determination can be most conveniently performed if the dissimilarity index has some relationship with the deviation mean square from regression analysis or GE interaction mean square in ANOVA (Lin, 1982). For four mentioned dissimilarity indices, Lin and Thompson (1975), Lin (1982) and Lin and Butler (1990) defined related F-tests for stopping clustering procedure. The purpose of this investigation is to show the practical importance of cluster analysis at study of GE interaction in MET of durum wheat in rain-fed conditions using four methods which are associated with ANOVA or linear regression procedures in two-way classification dataset. Also, studied durum wheat genotypes were evaluated from both genotype and GE interaction components through these cluster analysis methods .

MATERIALS AND METHODS

Experimental data

Twenty durum wheat genotypes were evaluated at five locations from 2007 to 2009, agricultural research stations of Gachsaran, Gonbad, Khoramabad, Ilam and Moghan except Ilam location which evaluated during only two years. At each location, the 20 genotypes were planted using a randomized complete block design with four replications. The soil types were Regosols in Gachsaran, Gonbad, Khoramabad and Ilam, and Cambisols in Moghan. Details of soil properties and geographical characteristics for the five locations are given in Table 1. Each plot consisted of six rows spaced 17.5 cm apart. Row length was 7 m in all locations during all years. Seeding rate was adjusted to obtain ≈20 plants × m^{-1} × row^{-1}. Fertilizer application was 30 kg N_2 × ha^{-1} and 70 kg P_2O_5 × ha^{-1} at planting and 40 kg N_2 × ha^{-1} at stem elongation stage. An area of 4.2 m^2 (4 rows with 6 m long) was harvested and yield (kg × ha^{-1})

was obtained by converting the seed yields obtained from plots to hectares. The experimental plant materials were from the ICARDA durum wheat breeding program. A preliminary analysis showed that environmental variances were homogeneous. Analyses variance was accomplished by the General Linear Model (GLM) procedure of Statistical Analysis System version 6.12 (SAS, 1996).

Table 1

Geographical properties of test location

Location	Longitude	Latitude	Altitude [m]	Soil Texture	Soil Type	Rainfall [mm]
Gachsaran	50°50'E	30°20'N	710	Silty Clay Loam	Regosols	460.8
Gonbad	55°12'E	37°16'N	45	Silty Clay Loam	Regosols	367.5
Khoramabad	48°17'E	23°26'N	1148	Silt-Loam	Regosols	433.1
Ilam	46°36'E	33°47'N	975	Clay-Loam	Regosols	502.6
Moghan	48°03'E	39°01'N	1100	Sandy-Loam	Cambisols	271.2

Stability analysis

Cluster analysis is used to classify data into subgroups which share similar properties. Lin and Thompson (1975) special type of cluster analyses was used to group genotypes for similarity of GE interaction. The data structure is based on cell means of a two way classification (genotype by environment structure). Also, the cluster analysis of Lin (1982) and Lin and Butler (1990), extensions of Lin and Thompson's (1975) cluster analysis, consists of three procedures depending on the linear regression or the ANOVA model were used. These four methods classified the studied genotypes according to joint effect of genotype (G) and GE interaction or GE interaction alone.

Detailed illustration of clustering and computation of dissimilarity index are given in Lin and Butler (1990). The formulas of dissimilarity index in each method and their degrees of freedom are given in Table 2. For formulas 1 and 4; SSR_i, SSD_j indicate the sums of squares (SS) due to the regression and the SS of the deviation from the regression for genotype *i*. Also, $SSR_{(1,2,\ldots,r)}$, $SSD_{(1,2,\ldots,r)}$ show the corresponding SS from the linear regression for genotypes 1, 2, . . . , r and $r \leq m$. For formulas 2 and 3; SSG_i, $SSGE_j$ indicate the sums of squares (SS) due to the genotype and the SS of the GE interaction for genotype *i*. Also, *m* is the number of genotypes, *n* is the number of environments, *r* is the number of genotypes in a newly formed cluster and *rep* is the number of experiment replication. The dis-

similarity indices of methods 1 and 4 are the numerators of the test statistics for a common regression line and for parallelism, respectively. The dissimilarity indices of Methods 2 and 3 are mean squares of genotypes (G) and mean squares of genotypes plus GE interaction (G + GE), respectively. Detailed explanation and simple numerical examples of the procedure can be seen in Lin and Butler (1990). A FORTRAN-77 program which known as Sl16 (Lin *et al.* 1992) is used for all four methods of cluster analysis .

Table 2

Four possible methods of cluster analysis based on regression and ANOVA models

Method	Strategy	Source*	Distance measure	v_1**	v_2***	References
1	Regression	G and GE	$d_{(1,2,\ldots,r)}=\frac{SS_{(1,2,\ldots,r)}-\sum_{i=1}^{r}SS_i}{SS_E(r-1)}$	*2×(r-1)*	*(m-1)×(n-2)*	Lin and Thompson, 1975
2	ANOVA	GE	$d_{(1,2,\ldots,r)}=SS(GE)_i/[(n-1)\times(r-1)]$	*(n-1)×(r-1)*	*m×(rep-1)×(n-2)*	Lin, 1982
3	ANOVA	G and GE	$d_{(1,2,\ldots,r)}=[SS(GE)_i+SS(GE)_j]/[n\times(r-1)]$	*n×(r-1)*	*m×(rep-1)×(n-2)*	Lin and Butler, 1990
4	Regression	GE	$d_{(1,2,\ldots,r)}=\frac{\sum_{i=1}^{r}SS_i-SS_{(1,2,\ldots,r)}}{r-1}$	*(r-1)*	*(m-1)×(n-2)*	Lin and Butler, 1990

*Grouping according to similarity of which sources; ** Degrees of freedom for fraction of F-test;
*** Degrees of freedom for denominator of F-test

RESULTS

Analysis of GE interaction

Table 3

Combined analysis of variance of durum wheat performance trial yield data

Skurce	DF	Mean Squares
Environment (E)	13	177747550.3**
Replication within E	42	826660.4
Genotype (G)	19	544937.2*
G × E	247	304181.0**
(R × G) within E	798	133065.7

** and * significant at the 0.01 and 0.05 probability level, respectively

Conventional combined analysis of variance was conducted to determine the effects of environment (location × year combination), genotype, and their interactions on seed yield of durum wheat genotypes (Table 3). The main effect of envi-

ronment (E) was highly significant ($P < 0.01$), while the main effect of genotypes (G) was only significant at 5 percent probability level ($P < 0.05$). Also, the GE interaction was highly significant at 1 percent probability level ($P < 0.01$). The highly significance of GE interactions of present investigation is indicating the studied genotypes exhibited complicated GE interaction. Seed yield is complex property and a quantitative trait, which expression is the result of genotype, environmental factors and GE interaction (Huehn and Leon 1985). Cooper *et al.* (1999) declared that the large GE interaction cause to the more dissimilar the genetic systems controlling the physiological processes conferring yield stability to different environments. The relative contributions of GE interaction effects for seed yield found in this study are similar to those found in other studies in rain-fed environments (Bertero *et al.* 2004; Sabaghnia *et al.* 2008). Therefore, GE interaction which makes it difficult to select the most favorable genotypes is an important issue in any plant breeding program because it reduces the progress from selection (Kang 1998; Yau 1995).

Clustering via regression and based intercept and slope

Table 4

Linear regression parameters and regression analysis of variance statistics

Genotype	Intercept	Slope	SS Total	SS Reg.*	SS Res.**	R^2
G1	2520.9	1.029	31707496.9	30552088.1	96283.8	96.4
G2	2697.3	1.023	30824210.9	30232374.4	49319.5	98.1
G3	2453.0	0.946	26739896.0	25817250.0	76887.2	96.5
G4	2635.2	0.949	27420198.4	25982885.4	119775.9	94.8
G5	2509.3	0.938	26648122.9	25391243.7	104740.2	95.3
G6	2528.4	0.961	28045543.4	26663166.4	115198.2	95.1
G7	2644.9	0.998	28920450.9	28748420.8	14335.3	99.4
G8	2580.4	1.017	30067191.2	29843498.9	18641.4	99.3
G9	2564.6	1.005	30419047.4	29143689.6	106279.7	95.8
G10	2637.6	0.986	28803617.4	28099601.8	58668.3	97.6
G11	2513.8	1.051	32904902.4	31901264.4	83636.3	96.9
G12	2493.6	1.002	29454317.4	28984086.6	39186.0	98.4
G13	2397.5	0.991	28846975.5	28378708.2	39021.9	98.4
G14	2562.9	1.016	30326652.9	29786008.8	45053.8	98.2
G15	2680.6	1.139	38946495.4	37444293.0	125183.0	96.1
G16	2376.3	1.019	30738720.9	29991114.5	62300.8	97.6
G17	2564.3	1.016	30476250.9	29833379.5	53573.0	97.9
G18	2641.3	1.021	30741884.9	30122245.6	51636.3	98.0
G19	2745.3	0.943	27712248.9	25668013.4	170352.7	92.6
G20	2470.8	0.954	26688178.4	26263546.6	35385.9	98.4

*Linear regression model sum of squares; ** Residual sum of squares

In this method, simple linear regression for yield stability estimation (Finlay and Wilkinson, 1963) was calculated and for each genotype, the mean yields were regressed against environmental index values; the resulting regression analyses are shown in Table 4. The pooled error estimate is 1465459.2 which is the sum of deviation from linear regression of all genotypes and was used to performing F-test and stopping the clustering process. The clustering cycle, grouped genotypes, dissimilarity index of each clustering cycle, degrees of freedom in each step and related F-test statistic are given in Table 5. According to this table results, F-test statistic was significant in cycle 13 where the dissimilarity index was 11540.23. In this step genotype 12 was grouped with a cluster which containing genotypes 1, 11, 8, 9, 14 and 17 and so there was significant difference between genotype 12 and this cluster based on G and GE sources of linear regression model.

Table 5

The smallest dissimilarity index at each cluster step and the determination of the cutoff point in genotypes clustering through regression model and based on line slope and intercept

Step	Grouped genotypes	Diss.	v_1*	v_2***	F-test
1	14,17	5.67	2	228	0.00
2	9 (14,17)	634.76	4	228	0.10
3	8 (9,14,17)	965.54	6	228	0.16
4	7,10	1115.53	2	228	0.18
5	3,20	1581.21	2	228	0.26
6	1,11	3822.34	2	228	0.63
7	18(7,10)	4669.07	4	228	0.76
8	5,6	5163.75	2	228	0.85
9	13,16	7144.47	2	228	1.17
10	(1,11) (8,9,14,17)	8850.30	10	228	1.45
11	(5,6) (3,20)	9825.96	6	228	1.61
12	2 (7,10,18)	10250.92	6	228	1.68
13	12(1,11,8,9,14,17)	11540.23	12	228	1.89**
14	4(2,7,10,18)	18183.18	8	228	2.98
15	19(4,2,7,10,18)	31533.92	10	228	5.16
16	(12,1,11,8,9,14,17) (5,6,3,20)	32689.11	20	228	5.35
17	(13,16)(12,1,11,8,9,14,17,5,6,3,20)	46595.95	24	228	7.63
18	15 (19,4,2,7,10,18)	74146.71	12	228	12.14
19	(15,19,4,2,7,10,18)(13,16,12,1,11,8,9,14,17,5,6,3,20)	99688.42	38	228	16.33

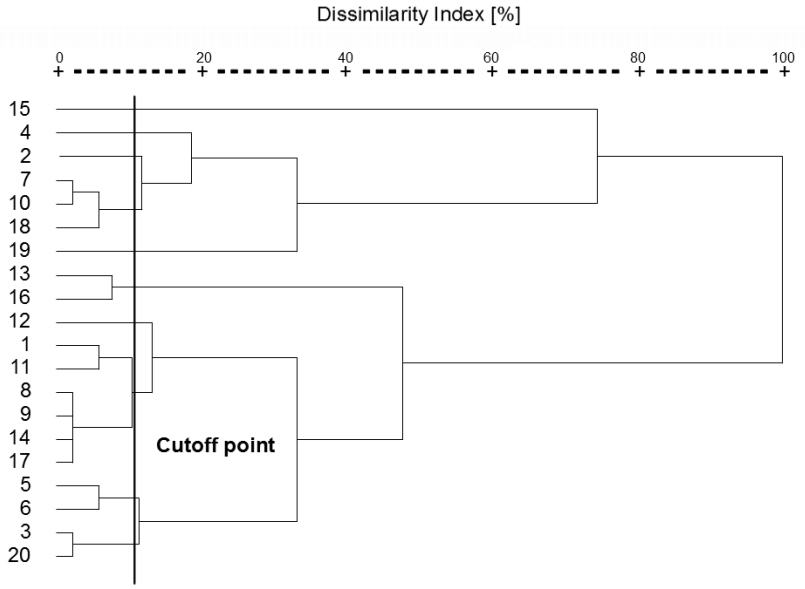

Fig. 1. Dendrogram of dissimilarity indices based on line slope and intercept of regression model for 20 genotypes of durum wheat which were evaluated across 14 environments

The positions of the all studied genotypes and significant cutoff point in this method are seen in Fig. 1. According to this dendogram, there were ten different genotypic groups consist on: genotypes 2, 4, 12, 15 as individual groups; 3 and 20; 5 and 6; 13 and 16; 7, 10 and 18; and 1, 8, 9, 11, 14 and 17 as the composite groups which had more than one genotype. Lin and Thompson (1975) to improve the effectiveness of this cluster method grouping indicate that most of the variation among genotypes is included in the between group component. The determination coefficient of linear regression model for 20 durum wheat genotypes ranged from 92.6 to 99.4 (Table 4). According to Pinthus (1973) genotypes with high coefficient of determination values can be evaluated adequately via linear regression model and the genotype response to environments is predictable to considerable degree. Regarding high values of coefficient of determination (Table 4) for all durum wheat genotypes, it can be conclude that using regression clustering method is useful for this dataset.

Genotype 15 with the highest line slope had specific adaptability to favorable environments and clustered as a single group (Fig. 1). Also, genotypes 5 and 6 were stable and had specific adaptability to poor environments due to low line slopes. The genotypes clustering based on similarity of linear regression parameters (both intercept and slope parameters) indicated that there are considerable variation among durum wheat genotypes and there are different with each other in response to environmental changes. This can be due to different origin of these improved genotypes, different pedigree and different breeding procedure. The question of

whether similarity should according to on line slope alone or on both intercept and slope parameters depends on the degree of emphasis the breeder wishes to put on GE interaction. For clustering genotypes, the similarity of both linear regression parameters (intercept and slope) may be more proper but for clustering test locations, the similarity of slope (GE interaction) is often more suitable (Lin and Butler 1990). Of course, Lin and Butler (1990) conclusion is correct in those situations which magnitude of genotype effect (or intercept) is greater than GE interaction (or line slope). Brandle and Brule-Bable (1991) indicated that cluster analysis based on regression analysis may be a suitable tool of selecting stable, high yielding and responsive genotypes in rapeseed.

Clustering via ANOVA and based on GE

Table 6

The smallest dissimilarity index at each cluster step and the determination of the cutoff point in genotypes clustering through ANOVA model and based on GE interaction

Step	Grouped genotypes	Diss.	v_1*	v_2***	F-test
1	14,13	8190.77	13	798	0.31
2	7,8	12472.62	13	798	0.47
3	12 (13,14)	19776.00	26	798	0.74
4	17 (7,8)	20071.38	26	798	0.75
5	2,10	23780.92	13	798	0.89
6	16 (7,8,17)	24569.44	39	798	0.92
7	3 (12,13,14)	27663.59	39	798	1.04
8	18 (7,8,16,17)	32240.62	52	798	1.21
9	20 (3,12,13,14)	32838.15	52	798	1.23
10	11,15	38302.77	13	798	1.44
11	(2,10)(7,8,16,17,18)	38902.15	78	798	1.46**
12	5 (3,12,13,14,20)	39093.66	65	798	1.47
13	6(2,7,8,10,16,17,18)	45722.73	91	798	1.72
14	9 (3,5,12,13,14,20)	46865.23	78	798	1.76
15	4,19	49635.69	13	798	1.87
16	1 (3,5,9,12,13,14,20)	54497.76	91	798	2.05
17	(11,15)(2,6,7,8,10,16,17,18)	58282.67	117	798	2.19
18	(1,3,5,9,12,13,14,20)(2,6,7,8,10,11,15,16,17,18)	66959.64	221	798	2.52
19	(4,19)(1,3,5,9,12,13,14,20,2,6,7,8,10,11,15,16,17,18)	76053.25	247	798	2.86

The dissimilarity index of this procedure is defined in terms of distance adjusted for the average effects of genotypes and it to be equivalent to within group mean square (MS) of GE interaction in ANOVA. The unweighted pair-group method of Sokal and Michener's (1958) was used in the clustering algorithm of this procedure. The clustering process including clustering cycles, genotypes which grouped, dissimilarity index of each clustering cycle, degrees of freedom of F-test in each step and related F-test statistic are given in Table 6. Lin (1982) defined the dissimilarity index as the GE interaction mean square and the new indices constructed in each group cycle are calculated from the data of clustered genotypes. Also, to determine the cutoff point on dendrogram, several methods have been proposed (Baril *et al.*, 1994). In this investigation, we regarded that the optimized number of genotypes must still be very informative for GE interaction interpretation. According to Robert (1997) cutting threshold or cutoff point was fixed 20% of pooled error in combined ANOVA and therefore GE interaction within clusters must thus be less than 20% of total variation. According to results of Table 6, F-test statistic was significant in cycle 11 where the dissimilarity index was 38902.15 and in this step genotypes 2 and 10 were grouped with a cluster which containing genotypes 7, 8, 16, 17 and 18. Thus, there was significant difference between genotypes 2 and 10 with pervious cluster based GE sources of analysis of variance model.

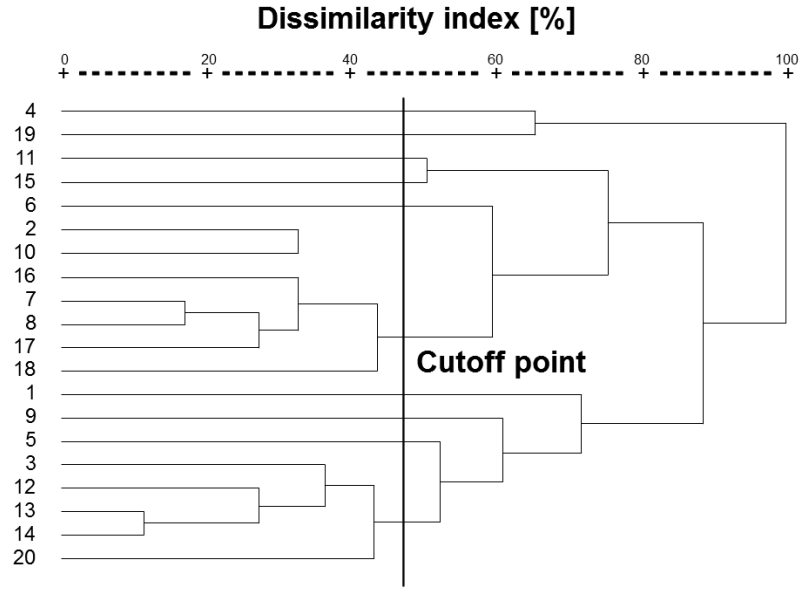

Fig. 2. Dendrogram of dissimilarity indices based on GE interaction of ANOVA model for 20 genotypes of durum wheat which were evaluated across 14 environments

The visualization of this groping method via dendogram and position of the significant cutoff point (Fig. 2) indicated that there were eleven different genotypic groups including; genotypes 1, 4, 5, 6, 9, 11, 15 and 19 as individual groups; 2 and 10; 3, 12, 13, 14 and 20; 7, 8, 9, 16, 17 and 18 as the composite groups which had more than one genotype. Lin (1982) declared that the cluster analysis based on similarity of GE interaction is as an analytical tool for investigating MET data, provides a logical base to compare the individuals within clusters by their average effect, and makes it possible to identify the structure of the GE interaction. The most prominent findings according to Fig. 2 are: genotypes 2 and 10 with the relatively high mean yield and low stability were groped as a same cluster while the other most stable genotypes or high yielding genotypes were cluster individually or mixed to each other. Similar to method 1 the genotypes clustering based on ANOVA and similarity of GE interaction showed huge variation among durum wheat genotypes and so 20 genotypes cluster into 11 groups.

Clustering via ANOVA and based on G and GE

Although, the seed yield of each durum wheat genotypes is a combined result of the of the G, E and GE interaction effects, only G and GE interaction are responsible to genotypes evaluation in MET. Usually, E source describes most of the total seed yield variation, while G and GE interaction are usually small (Yan and Kang, 2003). Lin and Butler (1990) proposed a dissimilarity index using both G and GE interaction in terms of distance adjusted for the average effects of these sources in ANOVA. The numerical results of clustering process: clustering cycles, clustered genotypes, dissimilarity index of each step, degrees of freedom and F-test statistic are given in Table 7. According to the results of this table, F-test statistic was significant in cycle 9 where the dissimilarity index was 37799.31. In this step genotype 20 was grouped with a cluster which containing genotypes 3, 12, 13 and 14 and so there was significant difference between genotype 20 and this cluster based on G and GE sources of ANOVA model.

Table 7

The smallest dissimilarity index at each cluster step and the determination of the cutoff point in genotypes clustering through ANOVA model and based on G plus GE interaction

Step	Grouped genotypes	Diss.	v_1*	v_2***	F-test
1	14,13	13666.29	26	798	0.51
2	7,8	20459.43	26	798	0.77
3	12 (13,14)	21289.14	39	798	0.80
4	17 (7,8)	23865.14	39	798	0.90
5	2,10	25265.14	26	798	0.95
6	16 (7,8,17)	30365.71	52	798	1.14

Table 7

Continued

Step	Grouped genotypes	Diss. index	v_1*	v_2***	F-test
7	3 (12,13,14)	30539.43	52	798	1.15
8	18 (7,8,16,17)	34138.29	65	798	1.28
9	20 (3,12,13,14)	37799.31	65	798	1.42**
10	11,15	39407.54	26	798	1.48
11	(2,10)(7,8,16,17,18)	46035.05	91	798	1.73
12	5 (3,12,13,14,20)	47100.19	78	798	1.77
13	6(2,7,8,10,16,17,18)	49475.43	104	798	1.86
14	9 (3,5,12,13,14,20)	52148.57	91	798	1.96
15	4,19	52361.8	26	798	1.97
16	1 (3,5,9,12,13,14,20)	53771.75	104	798	2.02
17	(11,15)(2,6,7,8,10,16,17,18)	63400.64	130	798	2.38
18	(1,3,5,9,12,13,14,20)(2,6,7,8,10,11,15,16,17,18)	70507.56	234	798	2.65
19	(4,19)(1,3,5,9,12,13,14,20,2,6,7,8,10,11,15,16,17,18)	80351.28	260	798	3.02

Like method 2, for obtaining the dendrogram cutoff point, 20% of pooled error of combined ANOVA (Robert, 1997) was used. According to this dendogram of Fig. 3, there were 13 different genotypic groups consist on: genotypes 1, 4, 5, 6, 9, 11, 15, 16, 19 and 20 as individual groups; 2 and 10; 7, 8, 17 and 18; and 3, 12, 13 and 14 as the composite groups which had more than one genotype. These mentioned results have many similarities to the cluster method involving GE interaction effect based on ANOVA. Karimizadeh *et al.* (2006) reported similar results for clustering different maize hybrids through both ANOVA based procedures and declared that due to small proportion of genotype effect in comparison to GE interaction effect, many similarities are observed in allocation of these two clustering methods. The most prominent differences between Fig. 2 and Fig. 3 were distinguish genotypes 16 and 20 as the single clusters which both of them had low mean yield and were the most constant or stable genotypes. Finally methods 2 and 3 clustering which benefits G and GE interaction effects can be useful for identifying the most stable genotypes according to type I stability (Lin *et al.*, 1986). Although, successful applications of Type I stability have been reported for small area tests (Francis and Kannenberg, 1978) and some international experiments (Mohebodini *et al.*, 2006), but the other stability types (Type 2 and Type 3) are very popular among plant breeders.

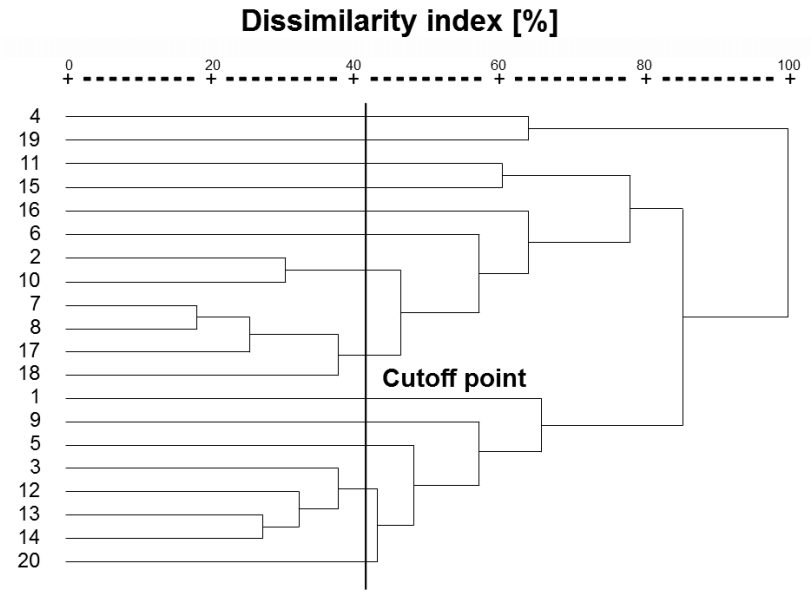

Fig. 3. Dendrogram of dissimilarity indices based on G plus GE interaction of ANOVA model for 20 genotypes of durum wheat which were evaluated across 14 environments

Clustering via regression and based on slope

Similar to method 1, joint linear regression (Finlay and Wilkinson, 1963) was used for clustering. The regression parameters and related sum squares are shown in Table 4. The properties of cutoff point determination in this clustering method and detailed information including clustering cycles, grouped genotypes, dissimilarity indices of cycles and F-test statistic are given in Table 8. According to this table results, F-test statistic was significant in cycle 18 where the dissimilarity index was 34199.52 and in this cycle a cluster which involving genotypes 3, 4, 5, 6, 19 and 20 was grouped with a cluster which containing genotypes 1, 2, 7, 8, 9, 10, 11, 12, 13, 14, 16, 17 and 18. Thus, there was significant difference between these two clusters based on GE sources or lines slopes of linear regression model.

The clustering cycles were summarized graphically in dendogram of Fig. 4 and significant cutoff point position was shown. According to Fig. 4, there were three different genotypic groups consist on: genotype 15 as individual group; 3, 4, 5, 6, 19 and 20 as a composite group and other reaming genotypes (10, 13, 7, 9, 12, 1, 2, 8, 11, 14, 16, 17, 18) as an independent group. Lin and Butler (1990) to improve the effectiveness of Lin and Thompson (1975) cluster method proposed this clustering method based on only lines slopes or GE source. Like to method 1, the validity of regression model for this present data is proved by high values of determination coef-

ficient (Table 4). Karimizadeh *et al.* (2006) also declared that there were good agreements between clustering methods which are based on similarity of both slopes and intercepts or only slopes in studying the different maize hybrids in MET.

Table 8

The smallest dissimilarity index at each cluster step and the determination of the cutoff point in genotypes clustering through regression model and based on line slope

Step	Grouped genotypes	Diss.	v_1*	v_2***	F-test
1	14,17	3.33	1	228	0.00
2	8 (14,17)	0.02	2	228	0.00
3	2,18	44.40	1	228	0.01
4	16 (8,14,17)	59.98	3	228	0.01
5	9,12	99.65	1	228	0.02
6	3,19	111.04	1	228	0.02
7	4(3,19)	249.07	2	228	0.04
8	(2,18) (8,14,16,17)	266.91	5	228	0.04
9	10,13	349.12	1	228	0.06
10	7 (9,12)	360.85	2	228	0.06
11	5 (3,4,19)	610.48	3	228	0.10
12	1 (2,8,14,16,17,18)	619.31	7	228	0.10
13	6,20	751.88	1	228	0.12
14	(10,13) (7,9,12)	1604.48	4	228	0.26
15	(6,20) (3,4,5,19)	1955.61	5	228	0.32
16	11 (1,2,8,14,16,17,18)	3986.67	7	228	0.65
17	(10,13,7,9,12) (1,2,8,11,14,16,17,18)	8505.79	12	228	1.39
18	(10,13,7,9,12,1,2,8,11,14,16,17,18) (6,20,3,4,5,19)	34199.52	18	228	5.60**
19	15 (10,13,7,9,12,1,2,8,11,14,16,17,18,6,20,3,4,5,19)	63147.79	19	228	10.34

The GE interaction in linear regression model was partitioned to heterogeneity (randomized variation) and residual components. In other word, GE interaction split to heterogeneity with df=19, SS=1199680.0 and residual with df=228, SS=17585540.0. The heterogeneity component as the randomized variation was not significant and indicated contribution of non-random effects was grater than random effects in GE interaction nature. Mandel (1961) indicated that if the lines slopes were identical for all studied genotypes, this heterogeneity component is distributed as χ^2 and is independent of environmental effects. Therefore, with considering high values of determination coefficient for linear regression, the model is appropriate and the GE interaction partitioning provides a method of testing for systematic GE interaction.

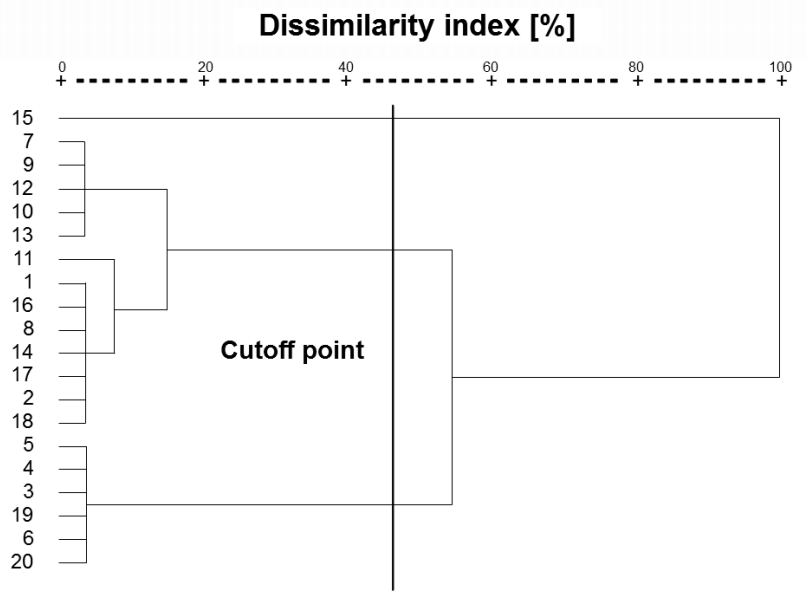

Fig. 4. Dendrogram of dissimilarity indices based on line slope of regression model for 20 genotypes of durum wheat which were evaluated across 14 environments

DISCUSSION

MET data analysis for durum wheat indicated homogeneous error variance in each of the 14 location × year environments and significant GE interaction. The environment accounted for a high percentage (96%) of sums of squares for seed yield (G+E+GE) while genotype accounted for a small percentage of sums of squares (about 1%). The GE interaction effects accounted for a relatively small amount of the sums of squares remaining for seed yield (3%). However, the GE sums of squares component was three times larger than the genotype component for seed yield. The relative contributions of G and GE interaction effects to the total variation for durum wheat seed yield found in research study are similar to those found in other crops MET studies in rain-fed environments (Cooper *et al.*, 1999; Bertero *et al.*, 2004; Sabaghnia *et al.* 2008). The GE interaction makes difficult to select the best performing and most stable genotypes (Hill, 1975). Also, GE interaction reduces the progress from selection in plant breeding programs (Yau, 1995).

Plant breeders usually use joint linear regression model to explore GE interaction in MET. The joint linear regression model major contribution was to quantify an environment effect using an environmental index (Yates and Cochran, 1938; Finlay and Wilkinson, 1963). This procedure permits

the response property to be assessed quantitatively by a regression coefficient. Lin and Thompson (1975) and Lin and Butler (1990) developed special types of cluster analyses to group genotypes for similarity of GE interaction plus G effect or only GE interaction via linear regression model. Brandle and Brule-Bable (1991), Lin and Lin (1994) and Karimizadeh *et al.* (2006) showed that this cluster analysis based on regression analysis has good ability for distinguish of similarities and dissimilarities. According to Lin *et al.* (1986) regression models of MET data analysis have Type II stability concept and a genotype is considered to be stable if its response to environment is parallel to the mean response of all genotypes in the trial and this type of stability beside Type III are very popular among plant breeders.

Clustering of 20 durum wheat genotypes via regression and based intercept and slope parameters (both G and GE interaction) determined 10 distinct genotypic groups while clustering genotypes through regression and based slope parameter (only GE interaction) reveled 3 distinct genotypic groups. Although, some authors (Robert, 1997; Karimizadeh *et al.* 2006) report that there are relatively similar results from the above mentioned clustering regression-based procedures, but we did not achieve similar results. This difference could be associated with the nature of the crop, environmental conditions or diverse genetic background of durum wheat genotypes obtained from different sources. However, these clustering methods could be more useful due to high amounts of determination coefficient for all studied genotypes. Furthermore, regression model-based procedures for GE interaction studies benefits type II stability of Lin *et al.* (1986) would be acceptable for the farmers and agronomists and equivalent to the dynamic concept of stability (Becker, 1981).

The choosing of cluster method depends on the choice of regression model or ANOVA for interpreting GE interaction in MET. The advantages of regression based clustering are removing random error in comparison of genotypes, and criterion of comparison is provided via regression's deviation mean squares (Lin and Butler, 1990). In most cases, plant breeders prefer to use of joint regression analysis instead of ANOVA-based procedures because they would prefer an agronomic (dynamic) concept of stability. In this concept of stability, it is not needed that the genotypic response to environmental conditions must be equal for all genotypes (Becker and Leon, 1988). But, how do the breeders understand that the linear fit is good or poor. Usually, the significant test of slope and the size of the coefficient of determination are regarded as the criteria. Lin and Butler (1990) advised that unless the coefficient of determination is more than 70% and the residuals are relatively homogeneous, the uses of regression based methods are not appropriate, and the alternative ANOVA procedures are recommended.

Clustering of studied durum wheat genotypes through ANOVA-based methods determined 11 distinct genotypic groups according to GE interaction, and 13 distinct genotypic groups according to G plus GE interaction. These clustering methods benefits mostly from real amounts and nature of GE interaction and so use Type I stability (Lin *et al.* 1986) or static concept of stability (Becker and Leon, 1988). The results of these both ANOVA clustering methods had most similarities to each other and this finding is in agreement with the other reports of ANOVA-based clustering methods (Robert, 1997; Karimizadeh *et al.* 2006). Many univariate parametric procedures of stability have such property, for example Wricke's (1962) ecovalance and stability variance of Shukla (1972). Although, most plant breeders do not prefer using of Type I stability-based methods but, analysis of the genetic properties of each stability type indicated that Type I useful for selection the most favorable genotype in stability analysis and GE interaction studies due to is good heritability and predictability (Lin and Binns, 1991).

If GE interaction does not fit the linear model well (determination coefficient lower than 50%) we should use the ANOVA-based clustering methods. In such situations, Lin and Butler (1990) proposed that for grouping genotypes, the similarity of both G and GE may be more suitable; but for grouping environments, the similarity of GE alone is more proper. Corsten and Denis (1990) recommended a simultaneous clustering method for classification of genotypes and test environments. Also, if the GE interaction is known and its magnitude is greater than G source, the genotypes they must be clustered by GE alone using ANOVA method Lin (1982) and regression method Lin and Butler (1990). In contrast, if G is greater in magnitude than GE, a joint assessment may lead to suitable clustering. The unique property of these clustering procedures is that the dissimilarity index constructed at any step of clustering is equivalent to the GE interaction mean squares for all genotypes in the group.

There are many ways to define the distance or similarity between two clusters with more than one case in a cluster. Also some clustering methods have been proposed to classification of genotypes or environments (Abou-El-Fittouh *et al.* 1969; Mungomery *et al.* 1974; Lin, 1982; Corsten and Denis, 1990). Whatever method is selected, the question concerning the determination of cutoff point for the dendrograms is raised. The link between the cluster analysis and the ANOVA in the cluster methods which used in this investigation provides a comfortable way of determining the cutoff point based on the F-test of the smallest dissimilarity index and the error estimate. Once this cutoff point is determined, one genotype per group is chosen, which makes it possible to propose a new experimental plant materials, less costly and largely informative in the GE interaction studies, and useful for future investigations. The all mentioned clustering methods en-

able plant breeders to explain the dataset into homogeneous subsets and to find out the GE interaction structure via the group pattern. These procedures have been reported to be useful not only for two-way data classification, but also for multi-way classification data (Lin *et al.* 1984).

The results of this investigation indicate that there are complex GE interactions in durum wheat MET in rain-fed conditions of Iran. Seed yield in the abiotic stress conditions was a poor predictor of yield in the favorable environments and so yield stability analysis of such environments must be done in similar conditions and not in the favorable environments (Cooper *et al.* 1997). Also, the clustering results showed that there are distinct genotypic groups for studied durum wheat genotypes from both G and GE interactions aspects. Finally, the results of this research indicate that cluster analysis may be a suitable means of choosing genotypes that are stable, high yielding, and responsive. Of further interest is the fact that the breeding genotypes in some groups were more stable with low mean yield or high yielding genotypes with low stability, indicating that the cluster analysis was successful in identifying variations among studied durum wheat genotypes. Such an outcome could be regularly applied in the future to delineate predictive, more rigorous recommendation strategies as well as to help define stability concepts for recommendations for durum wheat and other crops in the Middle East and other areas of the world.

ACKNOWLEDGEMENTS

Sincere gratitude goes to Iran's Research and Education Organization (AREO) and its Agricultural Research Stations for providing plant materials, experimental sites, and technical assistance.

REFERENCES

Abou-El-Fittouh, H.A., Rawlings, J.O., Miller, P.A., 1969. Classification of environments to control genotype by environment interactions with an application to cotton. Crop Sci. 9, 135–140.

Allard, R.W., Bradshaw, A.D., 1964. Implications of genotype-environment interactions in applied plant breeding. Crop Sci. 4, 503–508.

Baril, C.P., Denis, J.B., Brabant, P., 1994. Selection of environments using simultaneous clustering based on genotype × environment interaction. Can. J. Plant Sci. 74, 311–317.

Becker, H.C., 1981. Correlations among some statistical measures of phenotypic stability. Euphytica 30, 835–840.

Becker, H.C., Leon, J., 1988. Stability analysis in plant breeding. Plant Breeding 101, 1–23.

Bertero, H.D., de la Vega, A.J., Correa, G., Jacobsen, S.E., Mujic, A., 2004. Genotype and genotype-by-environment interaction effects for grain yield and grain size of quinoa (*Chenopodium quinoa* Willd.) as revealed by pattern analysis of international multi-environment trials. Field Crops Res. 89, 299–318.

Brandle, J.E., Brule-Bable, A.L., 1991. An integrated approach to oilseed rape cultivar selection using phenotypic stability. Theor. Appl. Genet. 81, 679-684.

Calinski, T., Corsten, L.C.A., 1985. Clustering means in ANOVA by simultaneously testing, Biometrics 41, 39–4

Cooper, M., Rajatasereekul, S., Immark, S., Fukai, S., Basnayake, J., 1999. Rainfed lowland rice breeding strategies for Northeast Thailand. I. Genotypic variation and genotype × environment interactions for grain yield. Field Crops Res. 64, 131–151.

Corsten, L.C.A., Denis, J.B., 1990. Structuring interaction in two way tables by clustering. Biometrics 46, 207 –215.

Edwards, A.W., Cavalli-Sforza, L.L., 1965. A method for cluster analysis. Biometrics 21, 362–375.

Finlay, K.W., Wilkinson, G.N., 1963. The analysis of adaptation in a plant breeding programme. Aust. J. Agric. Res. 14, 742–754.

Francis, T.R., Kannenberg, L.W., 1978. Yield stability studies in short-season maize: I. A descriptive method for grouping genotypes. Can. J. Plant Sci. 58, 1029–1034.

Hill, J., 1975. Genotype–environment interaction, a challenge for plant breeding. J. Agric. Sci. 85, 477–493.

Huehn, M, Leon, J., 1985. Phenotypic yield stability depending on plant density and on mean yield per plant of winter rapeseed varieties and of their F_1 and F_2-generations. J. Agron. Crop Sci. 162,172–179.

Kang, M.S., 1998. Using genotype-by-environment interaction for crop cultivar development. Adv. Agron. 62, 199–252.

Karimizadeh, R., Dehghani, H., Dehghanpour, Z., 2006. Using Cluster Analysis for Stability of Maize Hybrids 2, 2006; J. Crop Produc. Process.10, 337–348.

Lin, C.S., 1982. Grouping genotypes by a cluster method directly related to genotype–environment interaction mean square. Theor. Appl. Genet. 62, 277–280.

Lin, C.S., Binns, M.R., 1991. Genetic properties of four types of stability parameters. Theor. Appl. Genet. 82, 505–509.

Lin, C.S., Binns, M.R., Lefkovitch, L.P., 1986. Stability analysis: where do we stand?. Crop Sci. 26, 894–900.

Lin, C.S., Butler, G., 1990. Cluster analyses for analyzing two-way classification data. Agron. J. 82, 344–348.

Lin, C.S., Butler, G., Hall, I., Nault, C., 1992. Program for investigating genotype-environment interaction. Agron. J. 84, 121–124.

Lin, C.S., Thompson, B., 1975. An empirical method of grouping genotypes based on a linear function of the genotype environment interaction. Heredity 34, 255–263.

Lin, C.S., Williams, C.J., Binns, M.R., 1984. Investigation of interchromosomal interaction among three major chromosomes of *Drosophiln melangaster* in response to environments and the relationship between multi-line and two-line analyses: Reexamination of Caligari and Mather data. Heredity 52, 403–414.

Lin, C.Y. Lin, C.S., 1994. Investigation of genotype-environment interaction by cluster analysis in animal experiments. Can. J. Animal Sci. 74, 607–612.

Mandel, J., 1961. Non-additivity in two-way analysis of variance. J. Am. Statist. Ass. 56, 878–888.

Mohebodini, M., Dehghani, H., Sabaghpour, S.H., 2006. Stability of performance in lentil (*Lens culinaris* Medik) genotypes in Iran. Euphytica 149, 343–352.

Mungomery, V.E., Shorter, R., Byth, D.E., 1974. Genotype environment interactions and environmental adaptation. 1 Pattern analysis - application to soya bean populations. Aust. J. Agric. Res. 25, 59–72.

Perkins, J.M., 1972. The principal component analysis of genotype-environmental interactions and physical measures of the environment. Heredity 29, 51–70.

Pinthus, J.M., 1973. Estimate of genotype value: a proposed method. Euphytica 22, 121–123.

Robert, N., 1997. Structuring genotype × environment interaction for quality traits in bread wheat, in two multi-location series of trials Euphytica 97, 53–66.

Sabaghnia, N., Dehghani, H., Sabaghpour, S.H., 2008. Graphic analysis of genotype × environment interaction of lentil yield in Iran. Agron. J. 100, 760–764.

SAS Institute., 1996. SAS/STAT user's guide. v. 6, 4th ed. SAS Inst., Cary, NC.

Shukla, G.K., 1972. Some statistical aspects of partitioning genotype-environmental components of variability. Heredity 29, 237–245.

Signor, C.E., Dousse, S., Lorgeous, J., Denis, J.B., Bonhomme, R., Carolo, P., Charcosset, A., 2001. Interpretation of genotype × environment interactions for early maize hybrids over 12 years. Crop Sci. 41, 663–669.

Wricke, G., 1962. Über eine methode zur erfassung der ökologischen streubreite in feldversuchen. Z. Pflanzenzüchtung 47, 92–96.

Yan, W., Kang, M.S., 2003. GGE biplot analysis: A graphical tool for breeders, geneticists, and agronomists. CRC Press. Boca Raton, FL.

Yates, F., Cochran, W.G., 1938. The analysis of groups of experiments. J. Agr. Sci. 28, 556–580.

Yau, S.K., 1995. Regression and AMMI analyses of genotype × environment interactions: An empirical comparison. Agron. J. 87, 121–126.