Theor Appl Genet (1991) 82:379–388

© Springer-Verlag 1991

Assessment of a method for cultivar selection based on regional trial data

C. S. Lin*, M. R. Binns

Research Program Service, Research Branch, Agriculture Canada, Ottawa, Ontario K1A 0C6 Canada

Received September 28, 1990; Accepted February 4, 1991
Communicated by A. R. Hallauer

Summary. In this study a method for analyzing regional trial data is investigated for its effectiveness in cultivar selection. The method is a synthesis of three procedures: (1) regression analysis for genotype × environment (GE) interaction, and subsequent cluster analysis for grouping cultivars for similarity of response; (2) superiority measure analysis of cultivar performance based on the distance mean square between the test cultivar and the maximum response; (3) type 4 stability analysis for three-way classification data (cultivar × location × year), to measure a cultivar's stability. Each of these three procedures is aimed at different aspects of the selection problem: the first obtains an overview of the types of cultivar response; the second measures a cultivar's general adaptability within the region; the third assesses a cultivar's ability to withstand unpredictable variation, namely that caused by year effects. Four sets of published data, each originally analyzed by a univariate or a multivariate approach, were used as examples. The results suggest that the present method provides direct and useful information for selection purposes. The superiority measure, which is the core of the method, can be used even if the data do not fit the linear model for GE interaction. Plotting the data with a fixed standard represented by the maximum response provides a simple visual tool for identifying environments where a cultivar performs well.

Key words: Genotype × environment interaction – Cluster analysis – Superiority measure – Stability analysis – Adaptability

Contribution No. C-035 from Research Program Service, Research Branch, Agriculture Canada, Ottawa, Ontario, K1A 0C6
* To whom correspondence should be addressed

Introduction

Several new statistical approaches have been proposed recently for analyzing regional trial data. Since the purpose of such a trial is to compare the responses of cultivars to different environmental conditions, the analysis requires more than an ordinary ANOVA and a comparison of genotype means. Historically, a location was regarded as a qualitative factor, and the cultivar's response property was assessed by non-analytic methods. This practice changed following Finlay and Wilkinson (1963), who applied a regression analysis to the genotype × environment (GE) interaction. Their major contribution was to quantify locations by the environmental index, thus allowing the response characteristic to be assessed quantitatively by a regression coefficient (b). The method is popular among some plant breeders because of its ease for interpretation and cultivar recommendation. However, the method is also known for its limitations. Firstly, the data may not fit the linear model, or, if they do, the residual mean squares form regression may be heterogeneous, so comparisons among b-values are not appropriate. Secondly, the method is good only in identifying a general response pattern, but is unable to detect more subtle differences. To cope with these shortcomings, several authors have proposed a multiplicative model for the GE interaction component of a two-way model (e.g., Zobel et al. 1988) or for the residual component of the regression model (Johnson 1977). The advantage of the principal component approach is that the interrelationships among cultivars can be displayed graphically. Several biplot techniques (e.g. Kempton 1984; Zobel et al. 1988) and other multivariate techniques (e.g., Westcott 1987) have been suggested as descriptive techniques for GE interaction.

We have studied different approaches to cultivar selection, not through model development, but through other devices especially designed to facilitate the selection process. These procedures are: (1) cluster analysis for grouping genotypes based on similarity of the response characteristics; (2) superiority measure for assessing a cultivar's performance relative to a location maximum; (3) stability analysis to assess variation over years. Each of these three procedures is aimed at different aspects of the selection problem. The cluster analysis identifies groups of cultivars that have similar response types. Its primary function is to provide a plant breeder with a general overview of the response characteristics of the data set so that selection is simplified and can proceed on a group basis. The second analysis, the core of the method, measures a cultivar's general adapatability in the region and also helps to identify its specific adaptability. The third analysis (applicable only when the data set is of the form: cultivar × location × time factor) measures a cultivar's homeostatic property (stability) in withstanding unpredictable environmental variation.

The purpose of this paper is to study the effectiveness of the combined use of these three procedures for cultivar selection. For comparison, we use four sets of published data originally analyzed by either a univariate or a multivariate approach. The univariate methods include Shukla's (1972) stability analysis, and the multivariate methods include Kempton's (1984) biplot, Westcott's (1987) spatial analysis (a term coined by Crossa 1988), and Zobel's et al. (1988) AMMI (Additive Main effects and Multiplicative Interaction) analysis. The conclusions derived from the present method and those from the original authors' are compared, and the practicality and limitation of the present method are discussed.

Data and methods

Data sets

Four sets of published data were used as examples:

Set 1. Blackman et al. (1978) yield data (g m^{-2}) for 12 winter wheat cultivars in 14 locations, (seven test sites each with two nitrogen levels: their Table 6). These data were also analyzed by Kempton (1984) and by Westcott (1987) using multivariate approaches.

Set 2. Crossa (1988) yield data (kg ha^{-1}) for 27 maize cultivars in 37 locations (personal communication). These data were originally analyzed using Westcott's (1987) method.

Set 3. Brandle and McVetty (1988) yield data (kg ha^{-1}) for five oilseed cultivars in nine locations for 3 years (their Table 2). These data were originally analyzed using Shukla's (1972) method, based on a 5 × 27two-way classification.

Set 4. Zobel et al. (1988) yield data (kg ha^{-1}) for seven soybean cultivars in 35 environments (their Table 1). These data were originally analyzed by AMMI. Note that the 35 environments they analyzed are only part of the total possible 72 combinations of nine locations by 8 years, and that the seven cultivars were chosen by the original authors from a larger set of cultivars.

Method of analyses

The data structures of sets 1 an 2 are two-way classifications, while sets 3 and 4 are three-way classification data. As suggested for type 4 analysis by Lin and Binns (1988 b), sets 3 and 4 were converted to cultivar × location means averaged over years. The data in set 4 are unbalanced, but, since there is a degree of balance between cultivars and locations, namely that comparisons between cultivars are independent of year differences, we used the unadjusted means over years. We did the analyses based on both unadjusted and adjusted cultivar × location means and found only minor differences in the results. Thus, all four sets of data were treated as two-way classifications for cluster and superiority analysis: i.e., sets 3 and 4 were analyzed as 5 × 9 and 7 × 9, respectively. The stability analysis was done for sets 3 and 4 only.

First, Finlay and Wilkinson's (1963) regression analysis was applied to each data set and the heterogeneity of b was investigated. The cultivars were then grouped for similarity of intercepts and slopes using Lin and Thompson's method (1975), or Lin and Butler's (1990) Method 3 where the regression model either did not fit or it did fit but the residuals were not homogeneous. The cutoff points for the dendrograms were determined by the F-ratio between the smallest dissimilarity index at each cluster cycle and the error estimate (see Appendix). Secondly, each cultivar's performance was assessed by a superiority measure (its P-value), defined as the distance MS between the maximum response and the test cultivar, averaged over all locations (Lin and Binns 1988 a). Since a P-value is calculated over all locations, it represents superiority in the sense of general adaptability. If selection is based solely on the P-value, a narrowly adapted cultivar, i.e., poor in general adaptability but good in specific adaptability, might be discarded. Therefore, to prevent discarding potentially useful cultivars with narrow adaptability, a pairwise GE interaction MS between each test cultivar and the maximum response was also calculated. Regarding the maximum response as a dummy cultivar, empirical cutoff points for the P-value and pairwise GE interaction MS were determined by the F-ratio as described by Lin and Binns (1988 a) (see Appendix). Where the pairwise GE interaction MS exceeded the cutoff point, the P-value of this cultivar was not used directly, and the data of these cultivars (and also any others of special interest) were plotted individually across all location means along with the maximum responses. Thirdly, for the three-way classification data (sets 3 and 4), a year within location MS averaged over all locations was calculated for each cultivar as a type 4 stability measure (Lin and Binns 1988 b).

Results and discussion

The regression analyses indicated that all sets of data fitted a linear model. The coefficient of determination (r^2) was generally greater than 80% with the exception of 'WILK' in set 4 (71%). Tests of heterogeneity of regression suggested that the b-values were heterogeneous for all sets except set 4 (Table 1). Note that in set 4 the residual MS from regression was not homogeneous by Bartlett's test (chi-square is 17.6 with 6 df). Since the validity of Lin and Thompson's (1975) cluster analysis depends on homogeneity of residuals, we used Lin and Butler's (1990) Method 3 instead for set 4. This method assumes no model for the GE interaction; however, it comes with a penalty – there is no error estimate if replicated data are not available (Lin and Butler 1990).

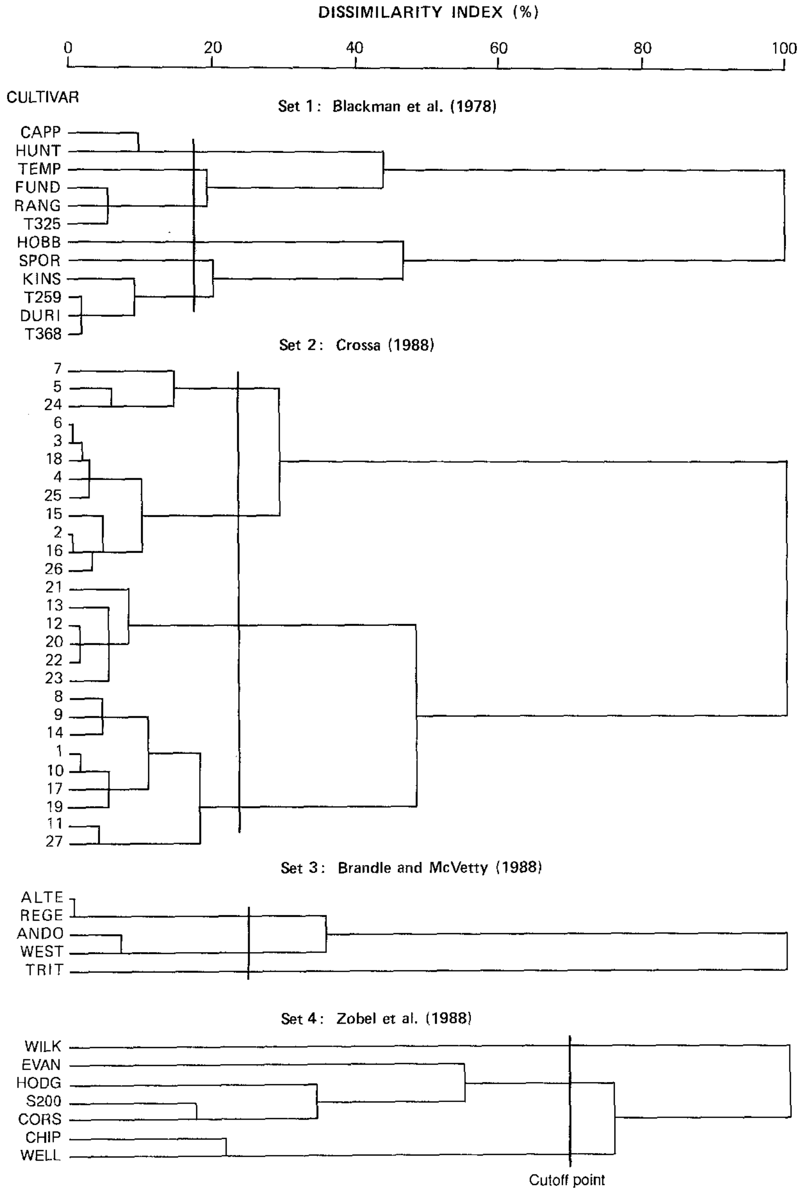

Fig. 1. Dendrograms for all four sets: sets 1, 2, 3 analyzed by Lin and Thompson (1975), set 4 by Lin and Butler (1990) method 3. Groupings are based on cutoff points derived as in the Appendix

Table 1. ANOVA for GE interaction MS for the four sets of yield data

Data set	Set 1: 12 × 14[a] Blackman et al. (1978)		Set 2: 27 × 37 Crossa (1988)		Set 3: 5 × 9 Brandle and McVetty (1988)		Set 4: 7 × 9 Zobel et al. (1988)	
Source	*df*	MS $g\ m^{-2}$	*df*	MS $Kg\ ha^{-1}$	*df*	MS $Kg\ ha^{-1}$	*df*	MS $Kg\ ha^{-1}$
GE	143	(1,316)	936	(195,307)	32	(34,350)	48	(63,235)
Heter. of b	11	5,234*	26	392,241*	4	79,101*	6	38,321
Residual	132	990	910	189,680	28	27,957	42	66,795

* Significant at the 5% level
[a] 12 cultivars, 14 locations

Table 2. Summary statistics for analyses of Set 1

Group	Cultivar	Regression analysis			Superiority measure		
		Mean[a] g m^{-2}	Slope	r^2 %	P-value	Rank	MS (GE)[c]
[Max. response		569	1.19]				
1	HOBB	561	1.24	98	125	1	99
2	SPOR	532	0.88	94	1,687*	2	1,091
3	KINS	506	1.21	97	2,341*	3	374
	T259	509	0.98	96	2,549*	4	843
	DURI	505	1.03	95	2,732*	5	727
	T368	505	0.99	94	3,125*	6	1,155
	J. G. Reg.[b]	506	1.05				
4	T325	477	1.11	98	4,529*	7	398
	RANG	472	1.09	94	5,068*	8	421
	FUND	461	1.14	95	6,765*	10	1,096
	J. G. Reg.	470	1.11				
5	TEMP	471	0.83	82	(6,627)	9	2,026*
6	HUNT	442	0.83	92	(9,661)	11	1,719*
	CAPP	442	0.68	90	(10,126)	12	2,327*
	J. G. Reg.	442	0.76				

* Significant at nominal 5% level. The cutoff points for P and GE are 1,673 and 1,703 respectively. If GE greater than cutoff, P should be used with caution (see text)
[a] SE of difference between cultivar means = 11.9
[b] The joint group regression
[c] Pairwise GE interactions MS between the cultivar and the maximum response

Table 3. Summary statistics for analyses of Set 2

Group	Cultivar	Regression analysis			Superiority measure		
		Mean[a] kg ha^{-1}	Slope	r^2 %	P-value[c]	Rank	MS (GE)[d]
[Max. response		5,958	1.10]				
1	7	5,495	1.11	95	251	1	148
	5	5,381	1.11	96	254	2	90
	24	5,284	1.13	85	(541)	12	322*
	J. G. Reg.	5,387	1.12				
2	3	5,190	1.05	98	354*	3	60
	6	5,219	1.02	96	359*	4	89
	18	5,208	1.06	95	376*	5	98
	4	5,130	1.06	95	464*	6	125
	25	5,151	1.07	93	489*	7	169
	15	5,182	0.98	93	502 *	8	207
	2	5,109	0.96	94	504*	9	150
	26	5,115	1.02	94	504*	10	153
	16	5,099	0.93	94	506*	11	141
	J. G. Reg.	5,156	1.01				
3	27	4,977	1.07	95	574*	13	96
	19	4,970	0.94	92	615*	14	131
	17	5,027	0.97	93	624*	15	196
	1	4,901	0.89	94	695*	16	143
	10	4,903	0.95	95	703*	17	155
	11	4,897	1.08	95	735*	18	177
	8	4,909	0.89	91	797*	19	254
	9	4,815	0.92	92	831*	20	183
	14	4,812	0.91	91	888*	21	238
	J. G. Reg.	4,913	0.97				
4	21	4,643	1.03	93	1,019*	22	159
	23	4,627	0.96	89	1,096*	23	216
	22	4,571	0.96	91	1,111*	24	152
	12	4,598	0.93	93	1,137*	25	218
	20	4,549	0.93	91	1,236*	26	249
	13	4,486	0.98	95	1,253*	27	175
	J. G. Reg.	4,579	0.97				

* Significant at nominal 5% level. The cutoff points for P and GE are 267 and 269, respectively. If GE is greater than cutoff, P should be used with caution (see text)
[a] SE of difference between cultivar means = 101.3
[b] The joint group regression
[c] Divided by 1,000
[d] Pairwise GE interaction MS between the cultivar and the maximum response. Divided by 1,000

The error estimates used for dendrograms were the residual MS of Table 1 for sets 1, 2, and 3. An approximate error estimate for set 4 was obtained by dividing the pooled error in the original paper (111,000) by 3 (the harmonic mean = 2.7). The resulting clusters gave 6, 4, 3 and 3 groups for sets 1, 2, 3, and 4, respectively (Fig. 1). The summary statistics for sets 1, 2, 3, and 4 are shown in Tables 2, 3, 4, and 5. The intercepts and slopes in each group are approximately homogeneous, indicating that the grouping of cultivars into subsets was satisfactory (the validity of grouping for set 4 is discussed later). The analyses for the superiority measure indicated that only 'TEMP', 'HUNT', and 'CAPP' of set 1 (Table 2), cultivar 24 of set 2 (Table 3), and 'WILK' of set 4 (Table 5) had large pairwise GE interactions with the maximum. The response characteristics of these five cultivars need to be investigated individually either by inspection of b-values or by plotting.

General assessments of the test cultivars in contrast to those by the original authors are as follows.

Set 1. Among the cultivars in the six groups 'HOBB' appears to be the best (smallest P-value). The excellence of 'HOBB' compared with 'SPOR' and 'KINS' is clearly demonstrated in Fig. 2. All cultivars of groups 5 and 6, which show significant pairwise GE interactions, do so because of poor response, as indicated by their low b-values. Although all methods agree that 'HOBB' is the best cultivar, the present method gives more information. First, it selected out 'HOBB' as a single group, distinguishing it from all others. Second, the near zero P-value of 'HOBB' (125 relative to the residual 990 in Table 1) suggests that this cultivar is incontestably the best. The most clear contrasts among the methods are in the assessment of 'SPOR': based on a projection of the relative yield in the lowest locations (1L and 2L in his Fig. 1), Kempton (1984) indicated that 'SPOR' had the highest

Table 4. Summary statistics for analyses of Set 3

Group	Cultivar	Regression analysis			Superiority measure			Type 4 stability
		Mean[a] kg ha^{-1}	Slope	r^2 %	P-value	Rank	MS (GE)[c]	
[Maximum response		2,579	1.25]					
1	WEST	2,504	1.29	97	6,973	1	4,615	801,176
	ANDO	2,431	1.10	90	24,277	2	14,916	829,199
	J. G. Reg.[b]	2,468	1.20					
2	ALTE	2,200	0.93	98	89,800*	3	19,990	437,839
	REGE	2,187	0.91	87	111,290*	4	38,668	302,588
	J. G. Reg.	2,194	0.92					
3	TRIT	1,768	0.76	90	362,199*	5	36,796	390,384

* Significant at nominal 5% level. The cutoff points for P and GE are 62,624 and 64,021, respectively
[a] SE of difference between cultivar means = 78.8
[b] The joint group regression
[c] Pairwise GE interaction MS between the cultivar and the maximum response

Table 5. Summary statistics for analyses of Set 4

Group	Cultivar	Regression analysis			Superiority measure			Type 4 stability
		Mean[a] kg ha^{-1}	Slope	r^2 %	P-value	Rank	MS (GE)[c]	
[Maximum response		2,900	0.97]					
1	HODG (1)[b]	2,815	1.03	92	12,472	1	9,997	208,517
	CORS (2)	2,700	1.07	94	50,649	2	34,547	207,804
	EVAN (0)	2,605	1.12	87	74,169*	3	34,629	220,903
	S200 (2)	2,601	1.10	87	106,407*	4	69,550	247,522
	J. G. Reg.[c]	2,680	1.08					
2	WELL (2)	2,381	0.90	86	184,654*	5	56,094	224,141
	CHIP (1)	2,297	0.77	98	204,231*	6	25,681	178,372
	J. G. Reg.	2,339	0.84					
3	WILK (0)	2,250	1.01	71	(274,786)	7	71,940*	333,730

* Significant at nominal 5% level. The cutoff points for P and GE are 69,560 and 71,780, respectively. If GE is greater than cutoff, P should be used with caution (see text)
[a] SE of difference between cultivar means = 91.53
[b] Maturity type
[c] The joint group regression
[d] Pairwise GE interaction MS between the cultivar and the maximum response

yield there, although he pointed out that his Fig. 1 might not capture enough of the variability for this estimate to be reliable. On the basis of a distance diagram for the four lowest locations and a table of means (Westcott 1987, Table 3), Westcott (1987) identified 'SPOR' as the best followed closely by 'HOBB'. In contrast, our assessment based on Fig. 2 is that even at the low-yielding locations 'SPOR' is only comparable to 'HOBB', and there is no evidence that it is in fact better: the mean differences ('SPOR' – 'HOBB') at the four lowest locations were 37, 7, 32, and -61 with pooled experimental SED, based on table 2 of Kempton (1984), equal to 36.

Set 2. Group 1 contains the most attractive cultivars (Table 3). Cultivars 7 and 5 have the smallest P-values, and because their slopes and means are about the same, these two cultivars can be considered as almost equivalent (Fig. 3). Cultivar 24 has a large pairwise GE interaction. Although its b-value is about the same as that of cultivars 7 and 5, a plot of the data shows that its perfor-

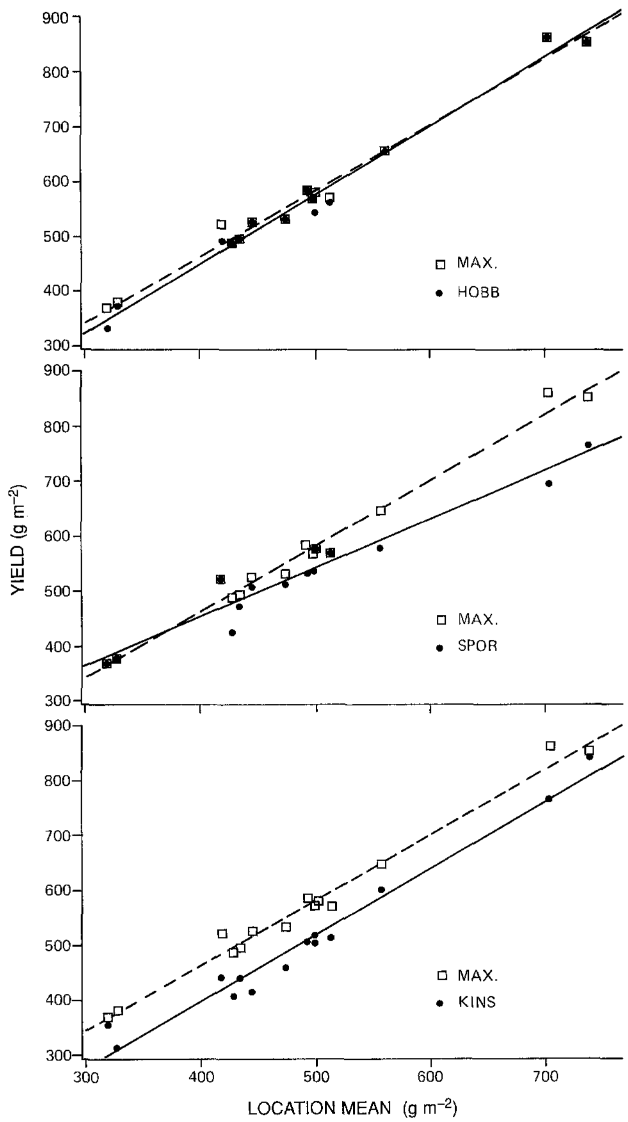

Fig. 2. Maximum and individual cultivar responses plotted against location means along with regression lines: 'HOBB', 'SPOR', and 'KINS' (set 1)

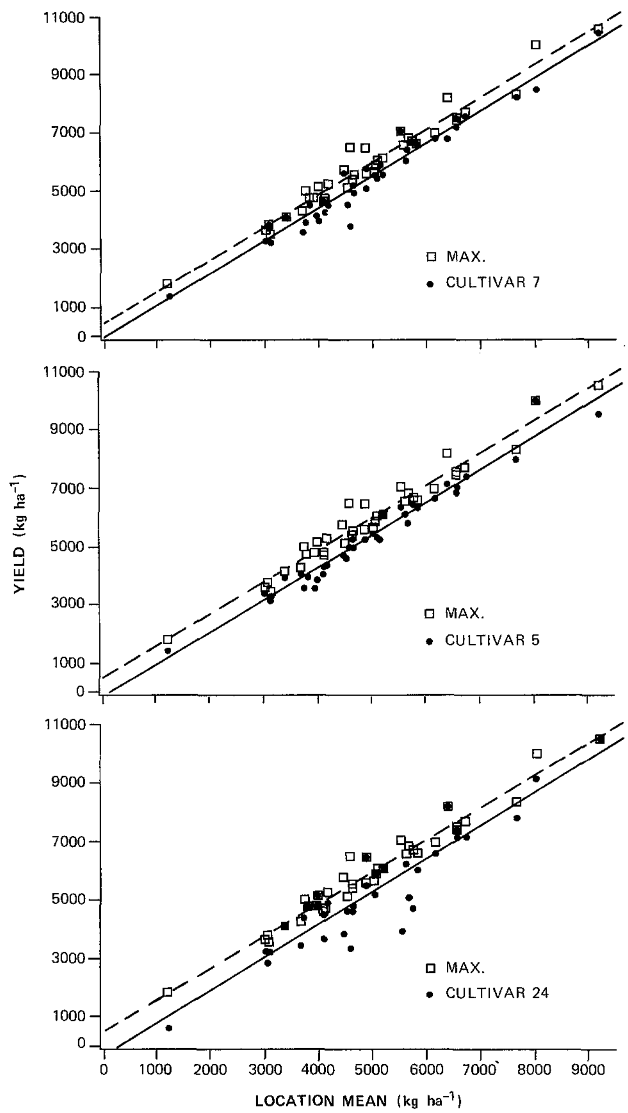

Fig. 3. Maximum and individual cultivar responses plotted against location means along with regression lines: Cultivars 7, 5, and 24 (set 2)

mance is highly variable except in very low- and very high-yielding locations, indicating that cultivar 24 has specific adaptability in part of the middle range of locations (Fig. 3).

Crossa (1988) also identified these three cultivars as the best, but his assessment of the response characteristics is different from ours. Based on a separation into 20 low- and 17 high-yielding locations, he concluded "From the analysis of both sets of sites, it seems that varieties 5 and 7 ..... have relatively good yields, *maintained their yield stability at low-yielding sites,* and responded well to more favorable sites. However, variety 24 ..... performed very well and *maintained its yield stability only at high-yielding sites*" (our italics). The differences are due to the methodology. Westcott's method orders locations by their means from lowest to highest and then performs principal coordinate analyses based on the sequential accumulation of locations starting from both ends according to rank order. In each cycle, the principal coordinate distance diagram is prepared in such a way that the cultivars whose performance is less than average are represented by points that tend to be clustered near the center of the diagram and those whose performances are above average are represented by remote points. A cultivar is considered to be most stable if it shows consistently above average performance throughout the cycles. A potential problem with this method is that at each cycle the data from one further location are added to the accumulated data from all previously included locations. As a result, plotted data at any cycle are correlated with those

Table 6. Separation of GE MS into two subsets of locations as in Crossa (1988); 17 low-yielding and 20 high-yielding locations

Source	*df*	MS	*F*-ratio
GE	936	(195,307)	
GE (low vs high)	26	248,937	1.28[a] ($P>0.15$)
GE (low)	416	179,728	
GE (high)	494	205,603	
(Pooled within subsets	910	193,774)	

[a] Tested against the pooled within subset MS

of the previous cycle, and these correlations increase as the process continues. Thus, Westcott's "stability" pattern, identified through pictorial presentation, is largely influenced by locations towards the two ends, while locations in the middle range are less influential.

The fact that cultivars are assessed on the basis of particular subsets of locations created by subdivision of the array of ranked locations raises a further potential problem: locations having similar means may have GE interaction, while locations having widely different means may have no GE interaction. In fact, the GE interaction effect between the low- and high-yielding locations in set 2 was not statistically significant (Table 6), indicating that whether the inferences are made based on the subsets or on the whole set should not make much difference.

For these data the present method has further implications. Although cultivar 7 is the best, its overall mean (5,495 Kg ha^{-1}) is much less than the overall mean of the maximum (5,958), indicating that more than one cultivar is needed to achieve maximum production for the entire region. [The ordinary SE is not appropriate for the mean of the maximum. However, a SE can be calculated using standard statistical theory (e.g., Kendall and Stuart 1961, chapter 10, eq 10.29), and is for example, approximately equal to 4 times the average CV percent: if CV = 20%, SE = 80]. We can try to take advantage of the apparent specific adaptability of cultivar 24 by supposing that the correct choice (for higher yield) between cultivars 7 and 24 is made at each location. Then the mean yield for the entire region is 5,716. By contrast, a similar choice between cultivars 7 and 5 gives mean yield of only 5,636.

Set 3. Group 1 contains the best cultivar (Table 4). Since the P-value for 'WEST' is not significantly greater than zero and a plot of the data (Fig. 4) indicates that 'ANDO' is never usefully better, 'WEST' can be recommended for the entire region. A notable difference between us and the original authors is in statements on stability: the original authors identified 'WEST', 'ANDO', 'REGE', and 'TRIT' as less stable than 'ALTE' by Shukla's (1972) stability variance (type 2 stability, see Lin et al. 1986), while we conclude that both 'WEST' and 'ANDO' are equally unstable as compared to the other

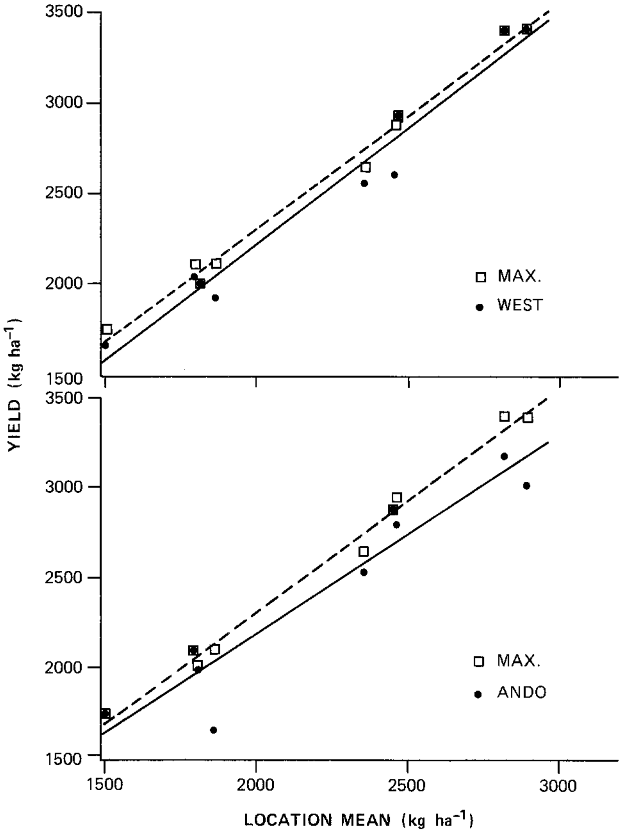

Fig. 4. Maximum and individual cultivar responses plotted against location means along with regression lines: 'WEST' and 'ANDO' (set 3)

three cultivars by type 4 stability. It is important to point out that Shukla's stability variances of 'WEST', 'ANDO', 'REGE', and 'TRIT' are almost identical to each other in magnitude, while the estimates of type 4 stability of 'WEST' and 'ANDO' are about twice as large as those of the other three cultivars. The indication is that these two cultivars are highly suspectable to factors that cause large year to year variations (e.g., drought, lodging, disease, etc.). Note that the weakness of Shukla's stability parameter has already been discussed in the context of type 2 stability (Lin et al. 1986). Our recent research (Lin and Binns 1991) has further confirmed that type 2 stability parameters are non-genetic, and thus not useful for selection purposes.

Set 4. Since the GE interaction for this set of data was not significant (see Table 2 of Zobel et al. 1988), it may be acceptable for this set to compare cultivars by their means, but instead we use P-values (Table 5) for reasons to be discussed later. 'HODG' is best, followed by 'CORS'. A significant pairwise GE interaction shown by 'WILK' was due to the fact that the cultivar was nearly the maximum at one location (the northernmost, Chazy), but was generally ranked lowest at the others (Fig. 5).

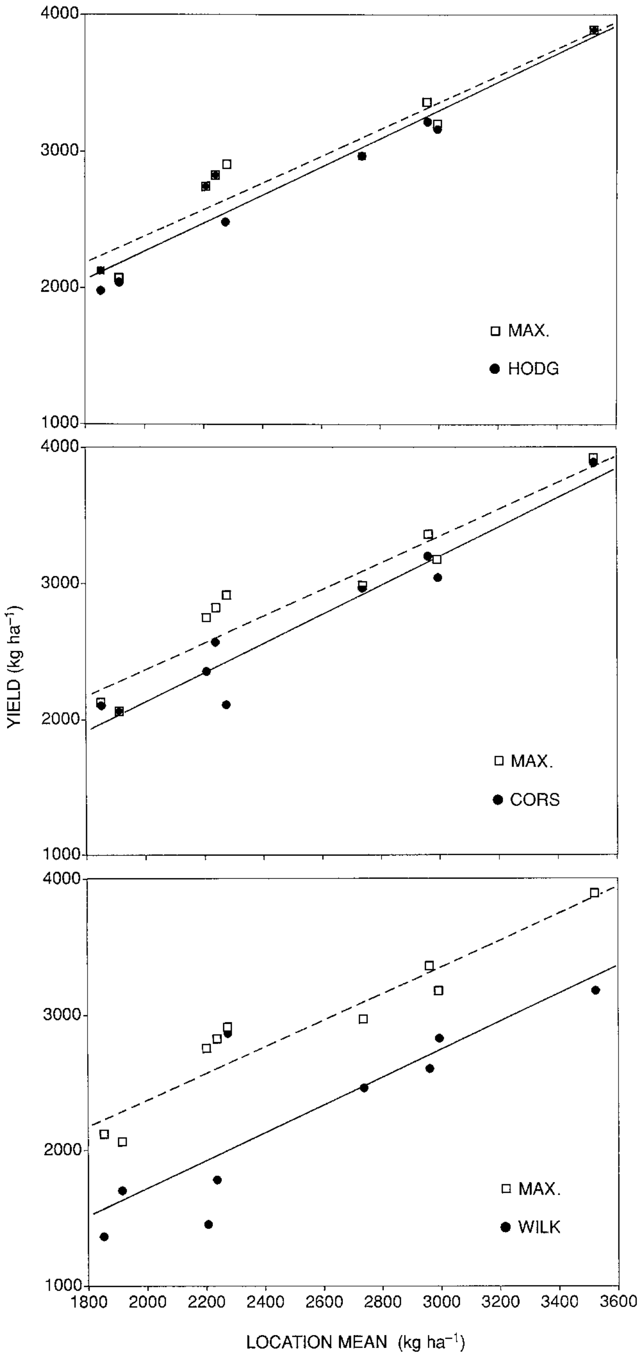

Fig. 5. Maximum and individual cultivar responses plotted against location means along with regression line: 'HODG', 'CORS', and 'WILK' (set 4)

Type 4 stability was about the same for all cultivars except 'WILK', which appeared to be the least stable. The original authors found that three maturity groups were visibly separated by the principal axis in their biplot, while our cluster grouping is not. To check which of these two classifications is more compatible with respect to the entire data structure, two ANOVAs were compared (Table 7). The grouping by maturity does subdivide the interaction very effectively (C × L/G is very small) but not the main effect (C/G is large), whereas the present grouping is consistent for both. However, if a multiplicative analysis of the GE interaction gives results in some disagreement with a cluster grouping using both G and GE, a more detailed examination of the cultivars may be illuminating. We analyzed the 7 × 9 data also by AMMI. The resulting distribution pattern in the biplot was very similar to that of 7 × 35, indicating that the relationships between cultivars and locations were not greatly influenced by the (unpredictable) year effects.

Table 7. Combined ANOVA for Set 4 of cultivar groupings based on method 3 of Lin and Butler (1990) and on three maturity types of Zobel et al. (1988)

Source	Method 3		Maturity	
	df	MS	*df*	MS
Location (L)	8	2,287,505	8	2,287,505
Cultivar (C)	6	(412,548)	6	(412,548)
Group (G)	2	1,085,022	2	110,588
C/G	4	76,311	4	563,528
C × L	48	(63,236)	48	(63,236)
G × L	16	99,904	16	149,754
C × L/G	32	44,902	32	19,977
Error[a]	667	37,000	667	37,000

[a] Approximation (see text)

Conclusions

The characteristics of the present method can be summarized as follows:

(1) The cluster analysis provides a *mechanical* way of sorting the data by similarity of means and slopes, or by a joint effect of G and GE if method 3 of Lin and Butler (1990) is used. It helps a plant breeder identify general response types among the test cultivars. The procedure can be used as a useful screening process when the number of test cultivars is large. However, if the number is small, this procedure may be redundant. Note that for Lin and Thompson's method, since the same environmental indices are used for both individual regression and the joint regression of each group, the intercept and slope characteristics of each group are the arithmetic means over the individual regressions.

(2a) One of the difficulties for comparing cultivars is that one normally wishes to take into consideration both "level" (mean) and "shape" (GE interaction) aspects of a data structure. This problem, common to regression and multivariate plot techniques, is to a great extent resolved by using a single parameter P-value based on the maximum response as the standard. It may be regarded as a transformation of a two-dimensional statistical space into a one-dimensional biological space. Plant breeders often use the cultivar mean averaged over all locations as a criterion for selection. Such an approach is reasonable if there is no GE interaction, but if there is interaction (usually the case in a regional trial), then it can be misleading. In contrast, comparison of P-values is always sensible whether experimental-wise GE interaction is repesent or not, provided the pairwise GE interaction with the maximum is properly checked. One of the most serious fears that a breeder has in any selection program is the possibility of rejecting a potentially useful cultivar whose mean may not be high but which has good adaptability to a relatively narrow niche of environments (e.g., 'WILK' of set 4) or accepting a cultivar whose mean may be high but which has a large variation at certain loctions (e.g., cultivar 24 of set 2). The present method provides a safety net to prevent this occurring, while selection by overall means does not have such a protective mechanism. This is why the P-value can be generally recommended instead of the mean to assess cultivar performance.

(2b) Plotting the data with a fixed standard represented by the maximum response is in complete contrast to multivariate techniques. The advantages of using the maximum response as a standard are twofold. First, because the maximum response represents the optimum response of a population (the test cultivars), it has a more stable and broader inferential base as a standard for comparison than check varieties. Second, it removes much of the dependence among cultivars (certainly it is independent of inferior cultivars) and thus provides a more heuristically ideal comparison than, for example, the location mean. Comparison of response patterns between the maximum and a candidate cultivar has a more direct and unambiguous visual representation. Not only can a cultivar's superiority be visually assessed in terms of general adaptability (closeness of cultivar's response across locations to the maximum), but also its specific adaptability can be identified (i.e., on which segment of locations the observed values are close to the maximum).

(3) The merit of type 4 stability can be assessed from two standpoints: statistical and biological. Statistically, this parameter is independent of the superiority measure and the regression analyses that are used to identify a cultivar's general adaptability. Thus, stability defined by type 4 becomes a selection criterion distinct from the criterion of adaptability. Biologically, among the four types of stability parameter, only type 1 and type 4 can measure a cultivar's homeostatic property (Lin and Binns 1988b). Our recent research also shows that these two parameters are heritable (Lin and Binns 1991). From an agronomic point of view type 4 is better than type 1 for two reasons: type 4 measures stability with respect to unpredictable environmental variation over which a breeder has no control, and, unlike type 1, its practicality would not be effected by the geographical size of test area. Unfortunately, type 4 is measurable only when the experiment is of the type cultivar × location × time (Lin and Binns 1988b). This leaves type 1 as the only sensible but not always useful choice of stability parameter for a cultivar × location experiment.

In conclusion, there are two important general considerations relating to the present method. Firstly, although regression analysis and grouping play important roles in the present method, they are by no means essential. The core of the present method is the superiority measure, which requires no assumption of linearity. Even where the data do not fit the linear model for GE interaction, the analyses for superiority measure and subsequent plotting are still appropriate and sensible. Secondly, the merit of the present method becomes more apparent as the geographical area covered by the test sites increases in scope. For example, if the test sites are chosen from a county within a state, selection by cultivar means or by type 1 stability may be adequate, but if the test sites are chosen from a large area such as in national or international trials, selection by these conventional parameters are not meaningful. Under such circumstances, the utility of the present method becomes significant.

Acknowledgements. We should like to thank Dr. J. Crossa, CIMMYT, for sending us the mean table of cultivars × locations, and T. Vanasse for computational assistance. We should also like to thank the referees for many useful comments.

Appendix. Determination of cutoff points. Let m be the number of locations in the set, and r be the number of cultivars in the smallest dissimilarity index in each cluster cycle. The cutoff point (C) for the dissimilarity index in cluster analysis, and for the P-value and GE MS in the superiority measure, can be calculated as follows:

$$C = (\text{MS of error}) \times F\ (v_1, v_2)$$

where $F\ (v_1, v_2)$ is the tabular 5% value of the F-statistic with *df* v_1 and v_2 defined as

$v_1 = 2(r\text{-}1)$ for dissimilarity index of Lin and Thompson (1975),

$v_1 = m(r\text{-}1)$ for dissimilarity index of Lin and Butler (1990),

$v_1 = m$ for P-value, (Lin and Binns 1988a),

$v_1 = m\text{-}1$ for the pairwise GE interaction (Lin and Binns 1988a),

and v_2 is the *df* for the error.

References

Blackman JA, Bingham J, Davidson LJ (1978) Response of semi-dwarf and conventional winter wheat varieties to the application of nitrogen fertilizer. J Agric Sci 90:543–550

Brandle JE, McVetty PBE (1988) Genotype × environment interaction and stability analysis of seed yield of oilseed rape grown in Manitoba. Can J Plant Sci 68:381–388

Crossa J (1988) A comparison of results obtained with two methods for assessing yield stability. Theor Appl Genet 75:460–467

Finlay KW, Wilkinson GN (1963) The analysis of adaptation in plant breeding programs. Aust J Agric Res 14:742–754

Johnson GR (1977) Analysis of genotypic similarity in terms of mean yield and stability of environmental response in a set of maize hybrids. Crop Sci 17:837–842

Kempton RA (1984) The use of biplots in interpreting variety by environment interactions. J Agric Sci 103:123–135

Kendall MG, Stuart A (1961) The advanced theory of statistics, vol. 1, 2nd edn. Griffin, London

Lin CS, Binns MR (1988a) A superiority measure of cultivar performance for cultivar × location data. Can J Plant Sci 68:193–198

Lin CS, Binns MR (1988b) A method of analyzing cultivar × location × year experiments: a new stability parameter. Theor Appl Genet 76:425–430

Lin CS, Binns MR (1991) Genetic properties of four types of stability parameter. Theor Appl Genet (in press)

Lin CS, Butler G (1990) Cluster analyses for analyzing two-way classification data. Agron J 82:344–348

Lin CS, Thompson BK (1975) An empirical method of grouping genotypes based on a linear function of the genotype-environment interaction. Heredity 34:255–263

Lin CS, Binns MR, Lefkovitch LP (1986) Stability analysis: where do we stand? Crop Sci 26:894–900

Shukla GK (1972) Some statistical aspects of partitioning genotype-environmental components of variability. Heredity 29:237–245

Westcott B (1987) A method of assessing the yield stability of crop genotypes. J Agric Sci 108:267–274

Zobel RW, Wright MJ, Gauch HG (1988) Statistical analysis of a yield trial. Agron J 80:388–393