Indian J. Genet., **67**(2): 161-165 (2007)

Simultaneous selection of cultivars for yield and stability in crop improvement trials

A. R. Rao and V. T. Prabhakaran

Indian Agricultural Statistics Research Institute, New Delhi 110 012

(Received: April 2004; Revised: June 2007; Accepted: July 2007)

Abstract

Assessment of cultivar performance in trials conducted across a range of locations and over years is often difficult because of the presence of significant location × year interaction. Methods for separation of environmental (location × year) variation into predictable and unpredictable components are available in literature. In the present paper an index, based on cultivar yield and the predictable component of environmental variance, MS(Y/L), is proposed for selecting cultivars simultaneously for high yield and stability. Two sets of rice data from All India Coordinated Rice Improvement Program, Directorate of Rice Research, Hyderabad, are used to illustrate the superior performance of the index method *vis-a-vis* the procedure advocated by Lin and Binns [2]. A user-friendly computer programme written in 'C' for judging promising cultivars is developed and available on request.

Introduction

Cultivar × environment interaction continues to be a challenging issue among plant breeders and production agronomists who conduct crop performance trials across diverse locations and over years. Presence of such interaction can reduce progress from selection. The problem gets further aggravated when the effect of location varies considerably from year to year, as evidenced by a significant location × year interaction in the analysis of variance (ANOVA). One approach to solve this problem is to make a single factor out of the location × year combination and to use Finlay and Wilkinson's [1] regression analysis to provide general information on a cultivar performance. However, since the environmental factor is a combination of location and years, it is of little use when recommendation of cultivars to specific locations is desired. Lin and Binns [2] demonstrated a method to separate environmental variance into predictable and unpredictable components, as a part of the environmental effect on a cultivar may be persistent and the remaining part, vary over time. For example, the soil effect is usually persistent from year to year and can be treated as fixed, whereas the weather effect varies from time to time and can be treated as random. In a cultivar × location × year experiment one can assume that the cultivar × location mean averaged over years is a biological equivalent of cultivar × predictable variation, and years within locations an equivalent of cultivar × unpredictable variation. The separation of the environment into predictable and unpredictable was first advocated by Allard and Bradshaw [3], who suggested that while developing cultivars for specific adaptation in predictable environments, plant breeders should also aim to produce cultivars that are adapted to withstand cultivar × unpredictable variation (such as year to year variation).

Apparently, an important criterion needed for cultivar selection is to find cultivars showing negligible years within locations variation, MS(Y/L), apart from being responsive to favourable environments. Lin and Binns [2] termed the MS(Y/L) component, type 4 stability representing the cultivar's homeostatic potential to withstand the unpredictable, adverse weather conditions. These authors, endorsing the view expressed earlier by Lin *et al.* [4] refused to recognize type 3 stability (residual from regression) as a stability parameter on the ground that the regression model in the context of GE interaction is a data-based, descriptive model than a predictive model unless the environmental index is measured independently of the genotypic mean. They argued that type 4 stability is independent of the regression analysis and other genotypes in the test and suggested this parameter to be largely genetically determined because their data revealed that the cultivar rank order under two seeding-rates were approximately the same. In our view, MS(Y/L) could be advantageously used along with the cultivar mean in identifying promising cultivars because these components are truly representative of the physiological adaptability and performance adaptability, respectively. However, integrating these two attributes into a suitable measure (index) - we call it simultaneous selection measure - will go a long way in selecting high yielding and stable cultivars.

Kang [5] proposed the rank sum method for selection of cultivars simultaneously for yield and stability when trials are conducted over locations. But the method has an inherent weakness of weighing heavily in the direction of yield performance, apart from the arbitrariness in the scoring involved. Bajpai and Prabhakaran [6] suggested three new indices, which are superior to Kang's index, for selecting cultivars simultaneously for high yield and stability. However, these indices are not suitable for cultivar × location × year trials, where unpredictable environmental variation is present and need to be computed for each cultivar in the trial.

In India, quite often cultivars of different crops are tested across a wide range of locations for several years under All India Crop Improvement Programmes. However, it is observed that the cultivars are judged largely on the basis of high yield with stability having very little role to play in judging the cultivars' sustainability potential. There is, therefore, an urgent need to devise techniques to select cultivars for both high yield and stability.

The primary objective of this paper is to propose an index based selection utilizing the information on cultivar yield and cultivar contribution to pooled MS(Y/L) and evaluate its performance *vis-a-vis* the procedure advocated by Lin and Binns [2]. A user-friendly computer program developed for executing the index based selection can be made available on request.

Material and methods

The statistical model used for cultivar × location × year trials data is given by

$$y_{ijkm} = \mu + g_i + e_j + y_{k(j)} + (ge)_{ij} + (gy)_{ik(j)} + B_{m(kj)} + \varepsilon_{ijkm} \qquad (2.1)$$

where y_{ijkm} is the yield of the *i*th cultivar in *j*th location for *m*th replication in the *k*th year for (i = 1, 2, ..., g; j = 1, 2, ..., l; k = 1, 2, ..., t; m = 1, 2, ..., r), g_i is the *i*th cultivar effect, e_j is the *j*th location effect, $y_{k(j)}$ is the effect of *k*th year within *j*th location, $(ge)_{ij}$ is the interaction effect of *i*th cultivar of *j*th location, $(gy)_{ik(j)}$ is the interaction effect of *i*th cultivar and *k*th year within *j*th location, $B_{m(kj)}$ is the effect of *m*th replication within *k*th year and *j*th location and ε_{ijkm} is the random error deviation distributed as N(0, σ^2). The Skelton ANOVA for analyzing $g \times l \times t$ genotypic means averaged over replications is as presented in Table 1. The average error $\bar{S}_e^2$ shown in the table equals, pooled error/*r* where pooled error is obtained by pooling the per plot variances over *lt* environments.

The significance of M_1 and M_3 are tested using average error while all other components are compared against M_{32}, which will be significant in most of the situations.

Table 1. Skelton ANOVA for cultivar × location × year genotypic mean data

Source	d.f.	MSS
Genotypes (G)	$g - 1$	M_1
Environments (E)	$lt - 1$	M_2
Locations (L)	$l - 1$	M_{21}
Years/locations (Y/L)	$l(t - 1)$	M_{22}
G × E	$(lt - 1)(g - 1)$	M_3
G × L	$(l - 1)(g - 1)$	M_{31}
Heterogeneity of regression	$g - 1$	M_{311}
Residuals	$(g - 1)(l - 2)$	M_{312}
Error: G × (Y/L)	$l(g - 1)(t - 1)$	M_{32}
Average error	$lt(g - 1)(r - 1)$	$\bar{S}_e^2$

Using the years within locations (Y/L) mean squares components of individual cultivars denoted by $MSYL_i$ (i = 1, 2, ..., g), a family of selection indices is obtained, where the cultivar performance and stability are quantified by expressing the individual achievements relative to the mean achievement in the group of cultivars considered. The proposed indices belong to the following family:

$$I = \frac{\bar{Y}_{i...}}{\bar{Y}_{...}} + \alpha \frac{(1/MSYL_i)}{\left(\frac{1}{g} \sum_{i=1}^{g} \frac{1}{MSYL_i}\right)} \qquad \dots (2.2)$$

where $\bar{Y}_{i...}$ is the average performance of the *i*th genotype $\bar{Y}_{...}$ the overall mean, $MSYL_i$ the *i*th genotype's years/locations mean squares and α ($\leq$1) is the weight to be attached to stability component when unit weight is attached to the yield component. The $MSYL_i$ values can be easily computed by forming location × year tables for the *g* genotypes. It must be appreciated that the sum of $l(t - 1)$ times $MSYL_i$, values over genotypes will be exactly equal to the sum of the components, (i) S.S. due to (Y/L) and (ii) the S.S. due to G × (Y/L).

A higher value of $\bar{Y}_{i...}$ is always desirable. If a genotype performs better than the average it will contribute a value > 1 to the Index. In contrast a higher value of $MSYL_i$ (indicating lesser stability) is not desirable and that is why the inverse ratio of this parameter has been used in the index. The value of α is decided by the breeder depending on the importance he would want to attach to the stability component in the light of his professional experience. At the moment we are not in a position to suggest any objective criterion for deciding α. By assigning four hypothetical values for α namely 1, 2/3, 3/7 and 1/4, we get four different indices, which are denoted by I_1, I_2, I_3 and I_4 respectively.

Under the index method the cultivars showing higher index values are selected whereas under the Lin and Binns [2] procedure those with higher $MSYL_i$. values are chosen. The potential of the selected cultivars for general/specific adaptation (under both methods) is decided on the basis of regression slopes determined from regression analysis based on cultivar × location yields averaged over years. A slope significantly > 1 indicates specific adaptation in favourable environments, < 1 indicates specific adaptation in adverse environments and a value close to 1 shows the cultivar's suitability for general adaptation. It must be understood that under these approaches no reliance is placed on the means square deviations or regression slopes in stability/adaptability assessment.

To assess the performance of the proposed indices, cultivar trial data is taken from All India Coordinated Rice Improvement Programme. The data pertains to the Initial varietal trial (one year) and Advanced Varietal Trials (two years) of Irrigated-mid-early (IME) cultivars tested across different locations. Here, it may be noted that only complete data, i.e. for those locations and varieties where data were present for all the three years is taken for the illustration. Two data sets of the type 15 × 5 × 3 and 19 × 4 × 3 have been considered. The set 1 consist of variety Nos. 14258, 14264, 14265, 14274, 14275, 14276, 14277, 14278, 14280, 14281, 14282, 14283, 14285, 14290 and 14292 for the locations CBT, SKL, PNT, KUL and RSP over the years 1994, 1995, 1996 in Zone 1. The second set consist of variety nos. 13261, 13379, 13380, 13722, 13729, 13730, 13731, 13732, 14031, 14032, 14034, 14036, 14037, 14038, 14039, 14040, 14041, 14042, 14044 for the locations BBN, CHP, CHN, and JDP over the years 1993, 1994 and 1995 in Zone 2.

Results and discussion

From the combined ANOVA (Table 2) we note that the main effects of genotypes and environments are highly significant. So also is the G × E component. The significance of the (Y/L) component clearly indicates yield instability over years, which we are interested to analyze. The regression slopes are not homogeneous for set 1 but are homogeneous for set 2. This means that in Set 1 the values of regression coefficients are of immense value in identifying promising cultivars for favourable environments, using the index. Also, since residual mean squares is highly significant and the regression slopes homogeneous, in set 2, almost the entire interaction is of non-linear type and so no useful prediction for genotypes is possible under the conventional regression technique. However, heterogeneity of residuals is not important in any analysis based on type 4 stability, because residuals have no bearing on type 4 stability. In fact, the years within locations mean squares, representing unpredictable environmental variation, is used for the estimation of type 4 stability (Lin and Binns [2]).

The summary statistics of the type 4 analysis are shown in Tables 3 and 4 for sets 1 and 2. These tables also indicate the ranks obtained by the varieties based on yield, $MSYL_i$ as well as index values. It is observed that in general the selection indices have shown significant positive association with both yield and the stability parameter. For data set 1, the rank correlations of the indices I_1, I_2, I_3 and I_4 with yield are 0.48, 0.53, 0.61 and 0.71 respectively. Significant rank correlations of order 0.98, 0.96, 0.95 and 0.89 are also observed between index-based ranks and stability ranks of the cultivars. For data set 2, a similar trend is observed with regard to these correlations. The lower order correlations of the indices with cultivar yield do not really mean that the index method is not very effective. A look at the cultivar yields in Tables 3 and 4 reveals that the yields do not show much variation and still different ranks had to be assigned to the cultivars depending on the observations, which in fact forms 3-4 clusters containing values of similar magnitude.

Table 2. Combined ANOVA of cultivar × location × year data (Kg/ha) for two different sets under All India Coordinated Rice Improvement programme

Set 1 (15 × 5 × 3)			Set 2 (19 × 4 × 3)		
Source	d.f.	MSS^a	Source	d.f.	MSS^a
Genotype (G)	14	4.53**	Genotype (G)	18	6.96**
Environment (E)	14	147.57**	Environment (E)	11	134.64**
Location (L)	4	104.39**	Location (L)	3	441.89**
Year Y/L	10	164.84**	Year (Y/L)	8	19.43**
G × E	196	5.41**	G × E	198	2.72**
G × L	56	7.34*	G × L	54	4.03**
Heterogeneity of regression	14	10.34**	Heterogeneity of regression	18	3.51^{NS}
Residuals	42	6.34^{NS}	Residuals	36	4.29*
G × (Y/L) (Error)	140	4.64**	G × (Y/L) (Error)	144	2.23**
Average error	420	0.59	Average error	432	0.39

a: Each entry is divided by 10^5; *(P<0.05); **-(P<0.01); NS - Non-significant

Table 3. Summary statistics for the data set1 (15 × 5 × 3) on mean yield (kg/ha), type 4 stability and index values, along with the ranks

Variety	Yield		Type 4 stability		Selection Index Value							
					I_1 (α = 1.00)		I_2 (α = 0.67)		I_3 (α = 0.43)		I_4 (α = 0.25)	
	Mean	Rank	Value	Rank	Value	Rank	Value	Rank	Value	Rank	Value	Rank
14258	44.27	12	109.59	2	1.14	4	1.10	4	1.07	4	1.03	5
14264	48.11	3	115.11	4	1.15	3	1.13	3	1.11	3	1.09	3
14265	43.08	15	128.97	6	1.03	6	1.01	7	0.99	9	0.97	9
14274	48.82	1	151.84	9	1.00	7	1.02	6	1.03	6	1.04	4
14275	44.89	11	195.17	12	0.86	12	0.88	12	0.91	12	0.93	13
14276	46.05	8	125.27	5	1.07	5	1.06	5	1.05	5	1.03	6
14277	46.14	7	215.06	14	0.84	14	0.87	14	0.90	13	0.94	11
14278	44.89	10	197.68	13	0.85	13	0.88	13	0.90	14	0.93	14
14280	43.62	14	166.85	11	0.90	11	0.91	11	0.92	11	0.93	12
14281	45.04	9	162.64	10	0.93	10	0.94	10	0.95	10	0.96	10
14282	46.68	5	146.72	8	1.00	9	1.00	8	1.01	7	1.01	7
14283	48.28	2	109.72	3	1.18	2	1.15	2	1.13	2	1.10	2
14285	46.30	6	145.15	7	1.00	8	1.00	9	1.00	8	1.01	8
14290	44.19	13	234.75	15	0.82	15	0.82	15	0.86	15	0.89	15
14292	46.86	4	94.20	1	1.27	1	1.22	1	1.17	1	1.12	1

NB: All the figures under columns Mean and Value should be taken as 10^2

Table 4. Summary statistics for the data set2 (19 × 4 × 3) on mean yield(kg/ha), type 4 stability and index values, along with the ranks

Variety	Yield		Type 4 stability		Selection Index Value							
					I_1 (α = 1.00)		I_2 (α = 0.67)		I_3 (α = 0.43)		I_4 (α = 0.25)	
	Mean	Rank	Value	Rank	Value	Rank	Value	Rank	Value	Rank	Value	Rank
13261	33.77	12	27.72	9	0.98	8	0.98	8	0.99	8	0.99	9
13379	33.88	11	39.50	15	0.84	15	0.87	14	0.90	14	0.93	14
13380	34.55	8	16.38	2	1.33	3	1.27	3	1.20	3	1.14	4
13722	29.36	18	22.87	7	1.02	7	0.99	7	0.96	10	0.93	15
13729	36.18	6	36.99	14	0.90	11	0.93	10	0.96	9	1.00	8
13730	35.46	7	51.97	18	0.78	18	0.83	17	0.89	16	0.94	12
13731	36.73	2	53.61	19	0.79	17	0.85	16	0.91	13	0.97	10
13732	32.95	14	48.14	17	0.76	19	0.81	19	0.85	19	0.89	18
14031	36.21	5	16.41	3	1.35	2	1.30	2	1.24	2	1.18	2
14032	36.57	4	18.62	4	1.26	4	1.22	4	1.19	4	1.15	3
14034	36.80	1	14.39	1	1.48	1	1.40	1	1.32	1	1.24	1
14036	32.31	15	41.91	16	0.80	16	0.83	18	0.86	18	0.89	16
14037	33.26	13	34.86	12	0.88	13	0.90	13	0.92	12	0.94	13
14038	36.64	3	33.32	11	0.94	9	0.97	9	1.00	7	1.02	6
14039	34.35	9	21.04	5	1.15	5	1.12	5	1.09	5	1.06	5
14040	30.86	17	33.13	10	0.86	14	0.87	15	0.88	17	0.89	17
14041	34.28	10	35.35	13	0.89	12	0.91	11	0.93	11	0.96	11
14042	29.06	19	27.41	8	0.92	10	0.91	12	0.89	15	0.88	19
14044	32.26	16	22.40	6	1.08	6	1.05	6	1.03	6	1.00	7

NB: All the figures under columns Mean and Value should be taken as 10^2

This has rendered the ranking ineffective and resulted in low correlation.

The merit of the index method can be assessed directly from the outcomes of index based selection. From Table 3 it is seen that all the indices have identified cultivar 14292 as the best in Set 1. The same conclusion is reached also from Lin and Binns [2] criterion of type 4 stability ranking. The yield of this candidate is also very close to the highest yield in the set. The second best candidate according to the index ranking is 14283 which is a top yielder as well. Although the type 4 stability ranking of this candidate is 3, its $MSYL_i$ is almost equal to the 2nd best candidate. This means that this candidate deserves to be ranked 2nd based on stability ranking. The index has chosen cultivar 14264 as the 3rd best candidate which has 3rd position also in terms of mean yield. The actual stability rank of this candidate is 4, which can be upgraded to rank 3 because the 2nd and 3rd positions are showing little difference in stability values and hence both can be placed in second position.

From the above discussion it is quite clear that both the index method and the Lin and Binns[2] approach select the same set of three best cultivars numbered 14292, 14283, 14264 from Set 1. It can also be observed that all the selected varieties have higher yield than the local checks. The regression slopes (Table 5) of the first two candidates are 1.67 and 1.99(respectively) showing that they are suitable for cultivation in high yielding environments. As regards the third candidate no clear cut recommendation can be given because the regression slope in this case is as small as 0.12.

If three cultivars are chosen from Set 2, both procedures will choose the same 3 candidates. The superiority of the index method lies on the fact that it is free from the ambiguities regarding stability ranks to be assigned to different cultivars when the stability values are almost equal, as encountered in the application of Lin and Binns [2] procedure. Moreover, the breeder will be in a position to execute index based selection in a few minutes by running the program (SYSGIS) developed by the authors.

Computer program for judging desirable genotypes

For the benefit of breeders, geneticists and production agronomists a user-friendly computer program has been developed. This program (SYSGIS), written on C platform, calculates mean yield, type4 stability and simultaneous selection index value for each cultivar under test. The program requires replication-wise genotypic performance from individual environments and over years. The menu prompts the user to provide the following information: (i) number of genotypes (ii) number of environments (iii) number of years and (iv) number of replications. The input for the programme should be arranged in a nested pattern, as replication within year within location within genotype. The relevant data should be created in an ASCII file (*data.prn*) so that the output will be generated in another ASCII file (in the name of 'result').

The various steps used in calculating the index value for the ith genotype by SISGYS are: (i) the mean performance of each genotype (ii) the stability measure, that is, mean square due to year within location ($MSYL_i$) (iii) Index value of the genotypes (iv) ANOVA indicating partitioning of GEI into variation due to heterogeneity of regression and deviation from regression. The genotype with highest index value will be given rank 1. The SISGYS programme also works for the mean data of cultivars (over replications) collected over locations and over years, in which case the mean squares due to pooled error and replication cannot be calculated. The SYSGIS can be obtained by sending an e-mail to either author (arrao@iasri.res.in or vtp@iasri.res.in).

Table 5. The regression slopes (b_i) computed for different cultivars from cultivar × location yields averaged over years for two sets of data

Set 1 (15 × 5 × 3)															
Variety	14258	14264	14265	14274	14275	14276	14277	14278	14280	14281	14282	14283	14285	14290	14292
b_i	-0.19	0.12	1.32	0.89	0.84	0.49	0.79	0.90	1.49	0.95	1.74	1.99	1.43	0.60	1.67
Set 2 (19 × 4 × 3)															
Variety	13261	13379	13380	13722	13729	13730	13731	13732	14031	14032	14034	14036	14037	14038	14039
b_i	1.27	0.71	0.81	0.89	0.54	1.07	1.00	0.92	1.07	1.04	0.76	0.92	1.36	1.21	0.75

Acknowledgements

The authors are extremely grateful to the referees for their useful and constructive comments which have helped us in improving the quality of the presentation.

References

1. **Finlay K. W. and Wilkinson G. N.** 1963. The analysis of adaptation in a plant breeding programme. Aust. J. Agr. Res., **14**: 742-754.
2. **Lin C. S. and Binns M. R.** 1988. A method of analyzing cultivar × location × year experiments: a new stability parameter. Theor. Appl. Genet., **76**: 425-430.
3. **Allard R. W. and Bradshaw A. D.** 1964. Implications of genotype-environment interactions in applied breeding. Crop Sci., **4**: 503-508.
4. **Lin C. S., Binns M. R. and Lefkovitch L. P.** 1986. Stability analysis - where do we stand ? Crop. Sci., **26**: 894-900.
5. **Kang M. S.** 1993. Simultaneous selection for yield and stability in crop performance trials: Consequences for growers. Agron. J., **85**: 754-757.
6. **Bajpai P. K. and Prabhakaran V. T.** 2000. A new procedure of simultaneous selection for high yielding and stable crop cultivars. Ind. J. Genet., **60**(2): 141-146.