Annals of the Sri Lanka Department of Agriculture. 2002. 4: 1-13

COMPARATIVE EVALUATION OF DIFFERENT STATISTICAL TECHNIQUES ON TESTING ADAPTABILITY OF RICE VARIETIES IN VARIED ENVIRONMENTS

G.W.J. CHANDRASIRI
Horticultural Crop Research and Development Institute, Gannoruwa

D.S. De Z. ABEYSIRIWARDENA
Rice Research and Development Institute, Batalagoda, Ibbagamuwa

B.L. PIERIS and H.M.C.K. JAYATHILAKE
Faculty of Agriculture, University of Peradeniya, Peradeniya

ABSTRACT

Determination of adaptability of crop varieties in multi-location variety adaptability trials (MVAT) is a problem in the presence of genotype by environment (G x E) interaction. Various statistical methods are available to analyse the data in MVAT. However, the information on these methods and their relative performance on evaluation of adaptability of rice varieties are limited. Therefore, a study was conducted to compare the statistical methods available for analysis of MVAT through application to MVAT data of rice. Past data of rice MVATS were collected from Rice Research and Development Institute at Batalagoda. It consisted of yield data on performance of 4 varieties at 5 locations over 6 seasons. The statistical techniques such as ANOVA, stability parameters, ranking and multivariate techniques were tested with the data. The results revealed that ANOVA models are not effective in describing pattern of G x E interaction but effective in describing main effects. Different Stability methods consider different aspects of variability across environments and hence produce different results for the same data. Therefore, interpretations of stability of varieties vary according to the parameters considered. Multivariate methods describe G x E interaction effectively with plots that are easy to understand. The first two principle components accounted for 58% and 28% of the variability respectively and described the broad adaptability and specific adaptability of varieties respectively. The PC bi-plots, and plots of factor analysis also described the pattern of G x E effectively. Therefore, consideration of results from ANOVA, stability, and multivariate methods altogether is recommended for evaluation of adaptability of rice varieties.

KEY WORDS: Rice, Variety Adaptability, G x E Interaction, ANOVA, Stability, Rank Methods, Multivariate Methods

INTRODUCTION

The promising rice varieties are tested in different agro-ecological regions to study the adaptability to varying climatic and soil conditions. These trials are commonly referred to as multi-location variety adaptability trials - "MVAT" (Abeysiriwardena *et al.*, 1991) or Multi-environment trials-"MET" (Crossa, 1992). Abeysiriwardena *et al.* (1991) defined the adaptability of a variety as a function of mean production and production stability across environments.

The yield variation due to changing environment is commonly referred to as G x E interaction (Kempton, 1984, Zobel *et al.*, 1988). The incidence of G x E interaction complicates the selection of a rice variety with superior adaptability to diverse environments. Consequently, statistical analysis of

MVAT has received more attention. A wide array of statistical techniques has been proposed to analyse the adaptability of varieties.

Most frequently applied statistical methods for analysis of multi-location trials are classified into 4 classes, namely, (i) Analysis of Variance methods (ANOVA) (ii) Stability methods (iii) Ranking methods and (iii) Multivariate methods (DeLacy *et al.*, 1996, McLaren, 1996). Nevertheless, the information on these methods, information on their application and limitations are scarce among plant breeders and agronomists. Hence, choice of a statistical method(s) relevant to objectives of the MVAT trials is a common problem for plant breeders. Therefore, a study was conducted to review and compare the characteristics of statistical methods available for analysis of MVATs through application to past data on rice.

MATERIALS AND METHODS

Source of Data

The data were collected from the records of MVAT trials at the Rice Research and Development Institute (RRDI), Batalagoda. The data on four varieties, namely, Bg 1639, Bg 2039, Bg 94–1 and At 353 tested continuously for six seasons in five locations such as Mahailluppallama, Ambalantota, Aralaganwila, Girandurukotte and Batalagoda were selected for the investigation. The design of all the MVAT trials were randomized complete block design (RCBD) with four replicates for each individual experiment for a given location and season.

Statistical Analysis of data

The data were analysed using , Statistical Analysis System (SAS) and MINITAB software. The statistical methods given below were used to analyse the data.

ANOVA Method

The ANOVA model given below was fitted for the data.

Model : $Y_{ipqr} = \mu + B_r + G_i + L_p + S_q + (GS)_{iq} + (GL)_{ip} + (LS)_{pq} + (GLS)_{ipq} + e_{ipqr}$

Where μ is the grand mean; G_i, L_p, S_q, represent the effect of the genotype, location, season, and $(GS)_{iq}$ + $(GL)_{ip}$ + $(LS)_{pq}$ + $(GLS)_{ipq}$ + e_{ipqr} represent the genotype x season, genotype x location, location x season, genotype x location x season interactions, random error respectively.

Method of Deviation of Yields From the Maximum

Abeysiriwardena (2001) proposed a method in which deviation of plot yields from the maximum yield of any plot in a given environment are computed and analysis of variance is run on these deviations. Mean deviation (D) and variance in deviation (V^2) across location for each variety (the stability parameters for general adaptability) are computed. Variety with the lowest D and minimum V^2 was declared as the most adaptable variety.

Stability Analysis to Evaluate the Production Stability of Varieties

The five stability statistics mentioned below are cited most frequently in literature and hence they were employed to study the production stability of varieties.

(i) The Variance of a variety across environments (S_i^2)

$$S_i^2 = \sum_{j=1}^{q} \frac{\left(x_{ij} - \bar{x}_i\right)^2}{q-1}$$

(ii) Coefficient of Variance (CV) of varieties across environments

$$CV_i = \left(\frac{S_i}{\bar{X}_i}\right) \times 100$$

Where CV_i is the coefficient of variation of i^{th} variety and S_i and $\bar{x}_i$ are the standard deviation and grand mean of variety yields over environments.

(iii) Wricke's (1962) Ecovalance (W_i^2) parameter

Where Wi is the ecovalance parameter of the i^{th} variety tested in different q environments.

$$W_i^2 = \sum_{j=1}^{Q}\left(x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..}\right)^2$$

(iv) Finlay & Wilkinson's (1963) regression coefficient (b_i)
Where b_i is the regression coefficient for variety i.

$$b_i = \frac{\sum_{j=1}^{Q} (x_{ij} - \overline{x_{i.}})(x_{.j} - \overline{x_{..}})}{\sum_{j=1}^{Q} (\overline{x_{.j}} - \overline{x_{..}})^2}$$

(v) Eberhart and Russell's (1966) deviation parameter (δ_i^2)

$$\delta_i^2 = \frac{1}{q-2}\left[\sum_{j=1}^{q}\left(x_{ij}-\overline{x}_{i.}\right)^2 - b_i^2\sum_{j=1}^{q}\left(x_{.j}-\overline{x}_{..}\right)^2\right]$$

The δ_i^2 is the deviation parameter for the i^th^ variety. The other labels, Xij represent the observed mean value of variety i (i = 1,2...p) in environment j (j = 1,....,q) and $X_{i.}$, $X_{.j}$ and $X_{..}$ denote the marginal means of genotype i, environment j, and overall mean, respectively.

Method of Ranking Yields for Selecting Varieties

The ANOVA method and mean comparison techniques were employed to analyse the data. The mean yield performances of the selected 4 rice varieties in 6 seasons (97 *yala*- 95 *maha*) at 5 locations were separated and tabulated for further analysis. The non-significant groups of variety means according to the mean separation were identified for each location respectively. The group containing the best variety or varieties received the highest score of 4, because there were only 4 varieties in this experiment. A score of 3 was given to the second best group and so on till all the groups in an experiment were scored. Individual varieties in each experiment were next scored on the basis of their availability in different groups. Ex.- a variety belonging to the first group only was assigned the highest score of 4. But variety belonging to both first and second groups received an average score of two groups ((4+3)/2=3.5). Finally, the mean and variance of scores for each variety across locations were computed. The variety showed high mean of scores and minimum variances of scores across locations is selected as the variety for broad adaptability (DAS, 1982).

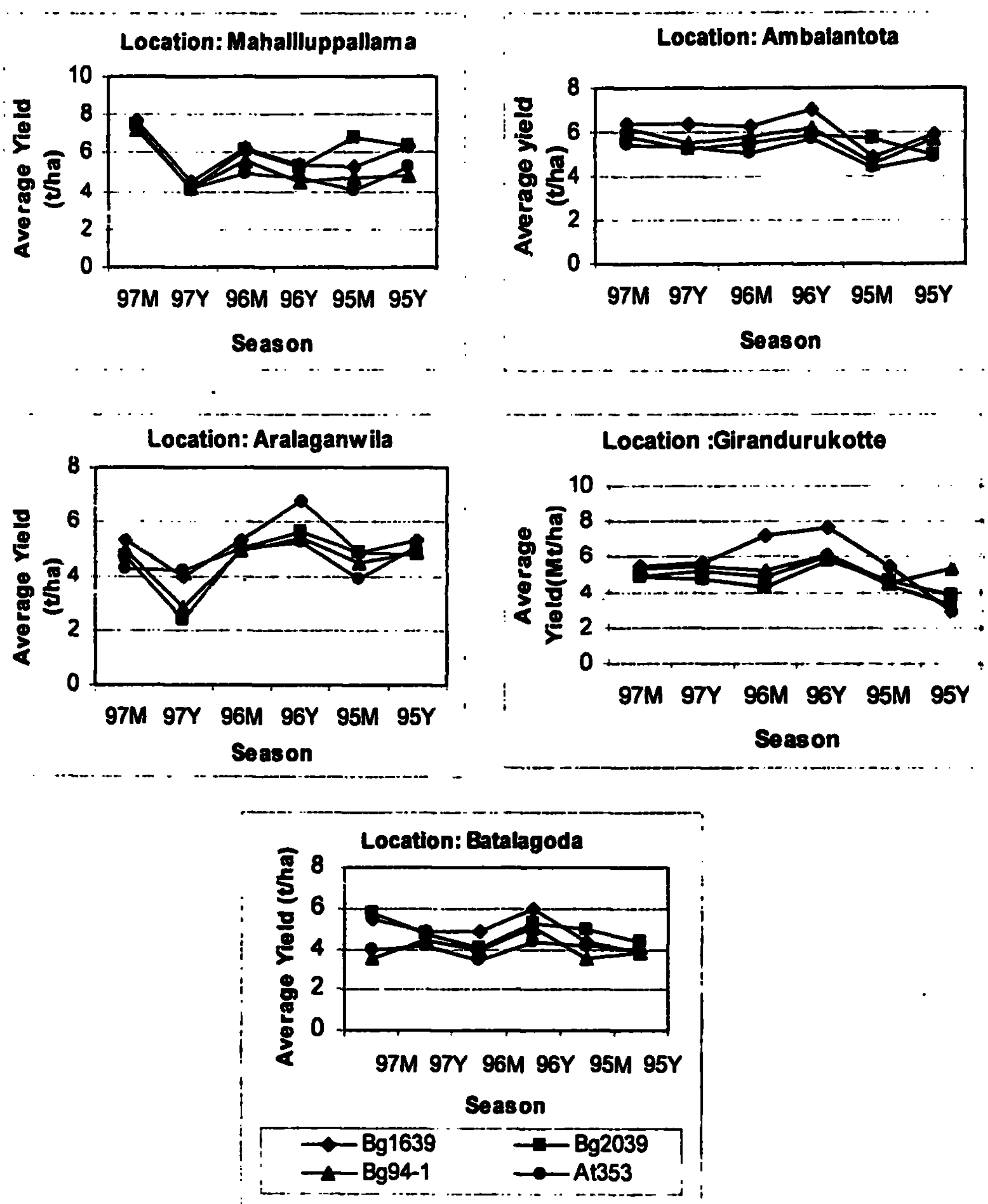
Multivariate Techniques

Principal Component Analysis (PCA), Factor Analysis were employed for data arranged in a two-way table of environment and varieties that contained mean yield of each variety at different environments (Manly, 1985, Broschat, 1979).

RESULTS AND DISCUSSION

ANOVA Method (Analysis of variance for combined seasons and locations)

The SAS output correspondent to the ANOVA model have shown that the P-value for genotype x season x location interaction is 0.0001. It indicate that performance of genotypes at different seasons vary from location to location. This interaction is illustrated in Figure 1 which is not described adequately in conventional ANOVA method.



* The legend for varieties given at Batalagoda is common for other locations

Figure 1. Performance of genotypes over seasons at different locations for four rice varieties

Stability Analysis

The stability describes the variability of yields among environments. The varieties with minimum variability across environments are referred to as stable genotypes. The results of stability analysis are shown in Table 1. The variance (S_i^2) indicates the variability in yields from mean yield of the particular genotype. The genotype with comparatively smaller variance is considered as stable. Hence, the genotype At 353 is the most stable genotype according to the criteria S_i^2. The results of coefficient of variance (CV) are appropriate to compare variability of different varieties. The results indicate that the genotype Bg 1639 is the most stable genotype according the criteria of CV.

Table 1. Results of stability analysis

Variety	S_i^2	*Rank*	*CV*	*Rank*	W_i^2	*Rank*	β_i	*Rank*	δ_i	*Rank*	*Rank total*
Bg 1639	1.11	4	18.88	1	6.48	3	1.22	1	0.19	2	11
Bg 2039	0.96	3	19.06	2	9.11	4	0.92	3	0.32	4	16
Bg 94-1	0.93	2	19.37	4	5.85	2	0.96	2	0.20	3	13
At 353	0.82	1	19.14	3	3.70	1	0.88	4	0.12	1	10

S_i^2 - Variance of a genotype across environment
CV - Coefficient of variance
W_i^2 – Ecovalance (Wricke's)
β_i - Regression coefficient (Finlay & Wilkinson's)
δ_i^2 - Deviation parameter from the regression (Eberhart & Russell's)

The ecovalance (W_i^2) consider the genotype x season interaction mean square as the criteria for stability. The minimum W_i^2 indicates the stable genotype, which is At 353. The Finlay and Wilkinson's regression coefficient (b_i) indicates the response of genotypes to environmental indices developed for different environments. The four different equations for four different genotypes are shown in Table 2. The b_i equal or greater than to 1 is considered as good genotypes. According to this criterion, the genotype Bg 1639 is the best genotype (Table 2). However, the performance of genotypes Bg 2039 and Bg 94-1 also appears to be similar to the performance of genotype Bg 1639. This could be verified statistically by comparing statistical significance of the slopes of the four different equations. This will be carried in the future. The ranking of varieties according to the Eberhart and Russell's deviation parameter (δ_i^2) showed that genotype At353 is the most stable variety followed by Bg 1639.

Table 2. Regression equations for 4 genotypes in Finlay & Wilkinson method

Genotype	*Regression equation*
Bg 1639	Yield = 5.59 + 1.22 I_k
Bg 2039	Yield = 5.14 + 0.928 I_k
Bg 94-1	Yield = 4.98 + 0.967 I_k
At 353	Yield = 4.74 + 0.886 I_k

I_k = environmental index

Comparison of Stability Methods

The comparison of results of stability analysis suggests that description of stability of a genotype vary with the method employed to test the stability. The selection of a best method is a problem as a standard is not available for comparison of different methods. Therefore, further studies on development of indicators to evaluate the efficiency of stability methods are required. Indicators based on power and robustness of different methods would be a sensible indicator statistically and hence the attention should be focused on this line. Use of Monte Carlo method or a bootstrap method would be useful in efficiency studies.

Method of Deviation of Yields From the Maximum

Table 3 shows the deviations of varieties from the maximum yield recorded in different locations averaged over different seasons.. The data indicate that overall mean deviation of variety Bg 1639 is the lowest compared to that of other varieties tested. Moreover, the variance of deviations of this variety is received rank 2 compared to other varieties. The results of deviation method suggest that overall production and production stability of variety Bg1639 is better than other varieties and hence it is the best variety in terms of adaptability. The variety Bg 2039 is in the second place according to this classification.

Table 3. Mean and variance of deviations of variety yields over locations

Variety	*Mean Yields (t/ha) of Locations*					*Average over locations*	
	MI	*AT*	*AW*	*GK*	*BG*	*Mean*	*Variance*
Bg 1639	0.938	0.914	1.160	1.286	0.944	5.242	0.0275
Bg 2039	1.093	1.126	1.152	1.134	1.236	5.741	0.0028
Bg 94 – 1	1.186	1.341	1.483	1.373	1.006	6.389	0.0343
At 353	1.118	1.230	1.572	1.315	1.292	6.527	0.0280

Average over 6 seasons

Method of Ranking Yields For Selecting Varieties

The results of the method using ranks are shown in Appendix 1. The results indicate that the overall adaptability of Bg1639 is superior compared to that of other varieties according to the ranking method. The problem in this method is that the highest mean score is not associated with the lowest variance of score and vise versa. Therefore, whether to consider mean score or variance of score or both measures simultaneously is a problem in selecting a genotype.

Multivariate methods

The performance of varieties in 30 environments has to be considered to evaluate the adaptability of rice genotypes. It is difficult to understand the information contained in 30 variables. Therefore, construction of indices to summarize the information contained in 30 variables is necessary in analysis of data. Such indices are called principal components. The principal components are the linear combination of original variables (environments). Consideration of values of few principal components instead of the values of the original variables is easy. In general, few principal components (2 to 3) describe the variability in the data.

Principal component analysis

PCA analysis of rice data showed that first three principal components altogether explain 99% of the variability in the data (Figure 2). The first principal component accounts 58% of the variability in the data and hence genotypes are mainly separated on the basis of scores of PCA1. The PCA2 explain 28% of the variability among genotypes. Thus, the first two PCA's altogether explains 86% of the variability in the performance of 4 varieties in 30 environments. Hence, the first two PCs were used to study the GxE interaction.

Principal Component Analysis					
Eigen analysis of the Correlation Matrix					
Eigen value	17.458	8.496	4.046	0.000	0.000
Proportion	0.582	0.283	0.135	0.000	0.000
Cumulative	0.582	0.865	1.000	1.000	1.000

Figure 2. The Eigen analysis of the Correlation Matrix of PCA analysis

The size and sign of variables (environments) determine the scores of a variety for a given principal component. The first principal component (PCA1) has 28 coefficients around 0.2 with negative signs out of 30 environments. It indicates that all the environments (location*Seasons) contribute equally for PCA1. This implies that PCA 1 describes the overall adaptability of genotypes in 30 environments. Thus, PCA1 could be labelled as "Broad adaptability component". The results suggest that a genotype with a higher score according to PCA1 has higher broad adaptability characters than other varieties. The scores in Figure 3 indicate that genotype At 353 has the highest broad adaptability character followed by Bg 94-1. The larger positive coefficients in second principal component (PCA2) are associated mainly with environments in the locations of Mahailluppallama, Batalagoda and Girandurukotte. Therefore, the second principal component could be labelled as "Specific adaptability component". The genotype BG2039 with larger positive score according to PCA2 has specific adaptability to locations of Mahailluppallama, Batalagoda and Girandurukotte.

VAR	PCA1	PCA2	PCA3
Bg 1639	-6.01334	-0.61558	-0.73678
Bg 2039	1.09853	4.27622	0.33924
Bg 94-1	1.24697	- 2.07672	2.58638
At 353	3.66784	- 1.58393	-2.18884

Figure 3. The scores of varieties according to first three principal components

Principal component bi-plot for varieties

The principal component bi-plot of varieties is another method to determine the similar varieties in terms of adaptability. The bi-plot also indibate that varieties At 353 and Bg 94-1 are similar and posses the character of broad adaptability (Figure 4).

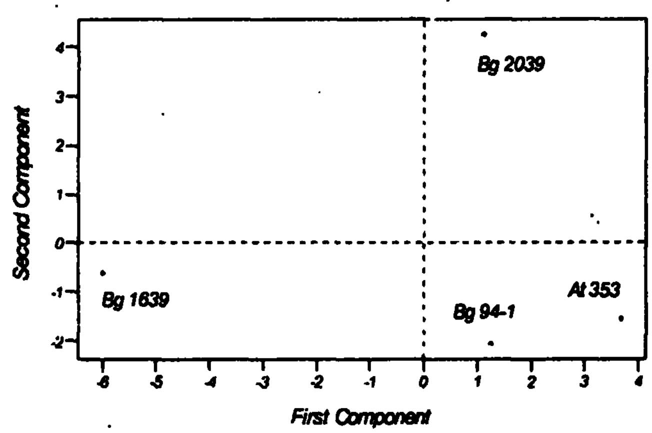


Figure 4. Principal component bi-plot of four varieties

Factor analysis

The principle in factor analysis is similar to that of principal component analysis. The objective is to reduce P variables to n factors where n <<< P. For example, the yield of 4 genotypes could be represented as

$$X_i = a_i F + e_i$$

Where X_i is the i^{th} standardized yield with mean zero and a standard deviation of one. The four varieties Bg 1639, Bg 94-1, Bg 2039 and At 353 were considered as the variables. The yields of each variety at 30 environments (*i.e.* yields in 5 locations in 6 seasons) were considered as the data for each variable. The discussion on results of factor analysis is not presented as it is beyond the scope of this paper. The important message is that the loadings plot of factor analysis could be used to identify the varieties that are similar in adaptability based on the performance in all the environments. This is useful to compare the new varieties with the standard or check varieties in the experiments. The loadings plot of factor 1 and factor 2 are shown in Figure 5. The length of co-ordinates from the origin and the angle between coordinates indicate the genotypes of similar production. The results in the plot indicate that over 30 environments yield performance of At 353 and Bg 1639 genotypes are similar.

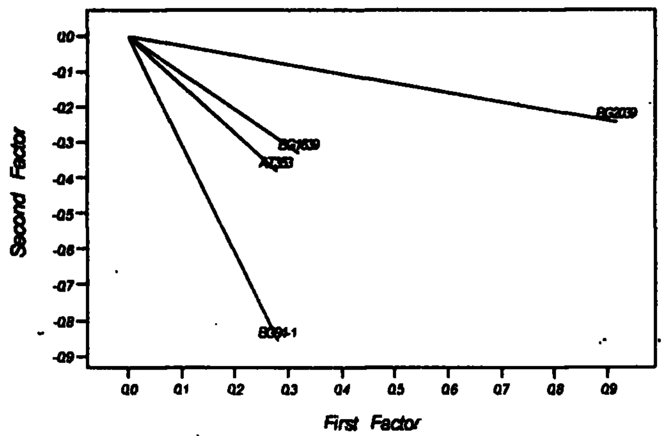


Figure 5. The loading plot for genotypes in Factor analysis

Comparison of results of different Statistical methods

The results of different statistical methods tested in this study are summarized in Table 4. The information in Table 4 indicates that different statistical methods identify different varieties as the most adaptable variety. Therefore, as solution, the variety identified by the majority of statistical methods could be declared as the most adaptable variety. In this study, the variety Bg 1639 has been identified as most adaptable variety in 5 methods and variety At 353 in 4 methods. The methods Factor analysis and Cluster analysis declared both Bg 1639 and At 353 as equally good. Therefore, on the basis of the argument, the variety Bg 1639 is the most adaptable variety to varied environments. However, in this particular situation one can declare that performance of At 353 is also similar to Bg 1639.

Table 4. Comparison of results of different Statistical methods

Method	*Criteria Considered*	*Identified Variety*
ANOVA Model III	Average	Bg 1639
Deviation method	production	Bg 1639
Stability	Average	
S_i^2	production	At 353: (4)*
CV		Bg 1639: (3)**
W_i^2	Variability in	At 353: (3)*
β_i	Production	Bg 1639: (4)**
δ_i^2		At 353: (2)*
Ranking method		Bg 1639
	Ranks of Average	
Multivariate	production	At 353
PCA		Bg 1639, At 353
Factor	Pattern of	
	variability	

CONCLUSIONS

Statistical methods such as conventional analysis of variance (ANOVA), stability parameters, deviation method, ranking methods and multivariate methods are identified for analysis of Multi-location variety adaptability trails

ANOVA method addresses the aspect of average production over seasons and locations as adaptability of varieties. This method describes the main effects of varieties and locations effectively. However, pattern of interactions between variety and environments are not described.

Stability methods address the aspect of variability of the production. Different stability measures consider different aspects of variability. Therefore, rank sum of varieties according to the five stability measures tested is recommended.

The multivariate method namely, principal component analysis describe the pattern in genotype x environment interaction and highlight the broad adaptability and specific adaptability of genotypes. Factor analysis technique helps to identify the groups of genotypes with similar form of adoption. Hence, multivariate methods are also important for analysis of Multi-location variety adaptability trails data. Multivariate methods consider both production and production stability concurrently. The output is graphical and easy to understand and provide supporting evidence to confirm the results of other methods.

In routine analysis of MVAT data with large number of varieties, the decision of the adaptability should be made considering the results of all 4 classes of methods, namely, Analysis of Variance, Ranking, Stability and Multivariate methods.

REFERENCES

Abeysiriwardena, D.S. de Z. 2001. Statistical analysis of on-farm yield trials for testing adaptability of rice. Euphytica 121:215-222.

Abeysiriwardena, D. S.de Z., R.B. Glenn and E. Paul. 1991. Analysis of multi-environmental yield trials. Tropical Agriculturist 147: 58 - 97.

Broschat, D.K. 1979. Principal Component Analysis in Horticultural Research. Hortscience Vol 14(2).

Crossa, J. 1992. Statistical analysis of multi-location trials. Advances in Agronomy, 44:55-85.

Das, G.R. 1982. A method of scoring yield status for selecting rice varieties. Indian J. Agric. Sci 52 (4): 207-209.

Delacy, I.H., K.E. Basford, M. Cooper, J.K.Bull and C.G. McLaren. 1996. Analysis of Multi-environment trials.In Plant Adaptation and Crop Improvement, Eds. M. Cooper and G.L. Hammer. Pp225-243. CAB International, Wallingford, Oxon OX10 8DE, UK.

Eberhart, S.A. and W.A. Russell. 1966. Stability parameters for comparing varieties. Crop Sci. 6: 36-40.

Finlay, K.W. and G.W. Wilkinson. 1963. The analysis of adaptation in plant breeding programmes. Aust. J. Agric. Res, 14: 742-754.

Kempton, R.A. 1984. The use of bi-plots in interpreting variety by environment interaction. J. Agric. Sci., camb. 103: 123-135.

Lin, C.S., M.R. Binns, and L.P. Lefkovitch. 1986. Stability analysis: Where Do We Stand?. Crop Sci. 26: 894-900.

McLaren, C.G. 1996. Methods of Data Standardization used in pattern analysis and AMMI models for the analysis of international multi-environment variety trials. In Plant Adaptation and Crop Improvement, Eds. M.Cooper and G.L. Hammer. Pp225-243.CAB International, Wallingford, Oxon OX10 8DE, UK.

Zobel, R.W., M.J. Wright and H.G. Gouch. 1988. Statistical analysis of a yield trial. Agronomy J. 80: 388-393.

Appendix 1.

Mean yield, mean score and variance of scores of rice genotypes in DAS methods

Variety	*Mahailluppallama*			*Ambalantota*			*Aralaganwila*			*Girandurukotte*			*Batalagoda*			*over locations*		
	Yield (t/ha)	*mean score*	*variance of* score	*Yield (t/ha)*	*mean score*	*variance of* score	*Yield (t/ha)*	*mean score*	*variance of* score	*Yield (t/ha)*	*mean score*	*variance of* score	*Yield (t/ha)*	*mean score*	*variance of* score	*Yield (t/ha)*	*mean score*	*variance of* score
Bg 1639	5.917	3.83	0.1666	6.139	3.83	0.1666	5.287	4.00	0.0000	5.729	4.00	0.0000	4.888	3.66	0.2667	5.592	3.86	0.1199
Bg 2039	6.037	3.92	0.0417	5.515	3.25	0.1750	4.571	3.58	0.2417	4.701	2.83	0.9664	4.881	3.66	0.2667	5.141	3.45	0.3393
Bg 94-1	5.219	3.25	0.5745	5.671	3.42	0.2416	4.597	3.50	0.2000	5.340	3.25	0.1750	4.063	2.75	0.1750	4.978	3.23	0.2732
At 353	5.114	3.08	0.6414	5.140	2.83	0.1666	4.637	3.50	0.3000	4.778	3.00	0.4000	4.016	2.75	0.3750	4.737	3.03	0.3766