CLUSTERING GENOTYPES BASED ON G $\times$ E INTERACTION IN THE ABSENCE OF HOMOGENEITY OF ERROR VARIANCES

By

CHANDRIKA. C.

THESIS

Submitted in partial fulfilment of the
requirement for the degree

Master of Science in Agricultural Statistics

Faculty of Agriculture
Kerala Agricultural University

Department of Agricultural Statistics
COLLEGE OF HORTICULTURE
Vellanikkara, Thrissur

1992

DECLARATION

I hereby delcare that this thesis entitled "CLUSTERING GENOTYPES BASED ON G X E INTERACTION IN THE ABSENCE OF HOMOGENEITY OF ERROR VARIANCES" is a bonafied record of research work done by me during the course of research and that the thesis has not been previously formed the basis for the award to me of any degree, diploma, associateship, fellowship or other similar title of any other University or Society.

Vellanikkara,
1-4-1992.

Chandrika.C
CHANDRIKA, C.

CERTIFICATE

Certified that this thesis entitled "CLUSTERING GENOTYPES BASED ON G X E INTERACTION IN THE ABSENCE OF HOMOGENEITY OF ERROR VARIANCES" is a record of research work done independently by Miss. CHANDRIKA, C., under my guidance and supervision and that it has not previously formed the basis for the award of any degree, fellowship or associateship to her.

(V.K. GOPINATHAN UNNITHAN)
CHAIRMAN, ADVISORY COMMITTEE
ASSOCIATE PROFESSOR OF
AGRICULTURAL STATISTICS
COLLEGE OF HORTICULTURE
VELLANIKKARA

VELLANIKKARA,
1-4-1992.

ADVISORY COMMITTEE

We, the undersigned members of the Advisory Committe of Miss.Chandrika, C., a candidate for the degree of Master o Science in Agricultural Statistics, agree that this thesis entitle "CLUSTERING GENOTYPES BASED ON G X E INTERACTION IN THE ABSENC OF HOMOGENEITY OF ERROR VARIANCES" may be submitted b Miss.Chandrika, C., in partial fulfilment of the requirement for th degree.

Chairman

Sri.V.K. Gopinathan Unnithan
Associate Professor of
Agricultural Statistics
College of Horticulture
Vellanikkara

Members

Sri.P.V. Prabhakaran
Professor and Head
Dept. of Agricultural Statistics
College of Horticulture
Vellanikkara

Dr.M. Abdul Vahab
Associate Professor (Hort.)
College of Horticulture
Vellanikkara

Smt.T.K. Indirabai
Assistant Professor of
Agricultural Statistics
College of Veterinary & Animal
Science, Mannuthy

External Examiner

(Dr. N Unnikrishnan Nair)

ACKNOWLEDGEMENT

I feel very happy to express my sincere gratitude and indebtedness to Sri.V.K. Gopinathan Unnithan, Associate Professor of Agricultural Statistics and Chairman of the Advisory Committee for his inspiring guidance, generous help and whole-hearted co-operation in the preparation of the thesis.

I would like to profusely thank Sri.P.V. Prabhakaran, Professor and Head, Department of Agricultural Statistics; Dr.M.Abdul Vahab, Assistant Professor (Hort.) and Smt.T.K. Indirabai, Assistant Professor (Ag.Stat.) for their timely help and encouragement as members of the Advisory Committee.

I express my sincere thanks to the Associate Dean, College of Horticulture, Vellanikkara, for providing necessary facilities.

I would be failing in my duty if I do not express my sincere thanks to all the staff members of the Department of Staatistics and my friends for their timely help and co-operation.

It is with pleasure I remember the help extended by Dr.A.V.R. Kesava Rao.

I would like to thank individually Mrs.M.N. Sreekala, Mr. John Thomas and Mr.K. Sukumaran for their help in the collection of references.

I express my sincere thanks to Smt.K.A. Mercy, Smt. Tessy Thomas and Smt. Joicy T. John for their help in the analysis work.

I also take this opportunity to express my whole-hearted appreciation to Mr.Joy who typed out the entire work sincerely.

Lastly, but not the least I express my indebtedness to my family members who had always been a source of inspiration and encouragement to me in all my endeavours.

CHANDRIKA, C.

To my Mother and
in memory of my Father

CONTENTS

APPENDICES

ABSTRACT

Introduction

1. INTRODUCTION

Partition of individuals into a number of groups in such a way that individuals in the same group are alike but distinct from individuals in other groups is very advantageously used in breeding programmes.

Genotype x environment interaction plays a very important role in crop improvement programmes. The phenotype of an individual is completely determined by the genotype and environment. The difference in environments has greater effect on certain genotypes than on others. The interplay of genetic and non-genetic effects on phenotypic expression of individuals is known as genotype-environment interaction.

The widely used regression approach to study genotype-environment (GE) interaction performs satisfactorily only when the regression of the GE interaction on the environmental index has very high predictability, which occurs very rarely. Moreover this approach assumes homogeneity of error variances in different environments. Hence this approach also fails in situations when the error variances in the different environments are heterogeneous.

Stratification of environments or genotypes can be used effectively to achieve low or no genotype-environment interaction within any group. The region for which a breeder is developing improved varieties can be so sub divided that all environments in any sub region are some what similar. Alternatively the genotypes under investigation are formed into different groups such that the genotypes within any group have similar response to differing environments. Another method to reduce the genotype environment interaction is to select stable genotypes that interact less with the environments in which they are to be grown.

Recent are the few attempts on grouping genotypes or environments such that GE interaction within any group is absent but present between any two groups. These clustering techniques have been developed for situations where the error variances in different environments are homogeneous. But often the error variances do become heterogeneous. Therefore it is necessary to have clustering procedure that can be used when error variances in different environments are heterogeneous.

The present investigation is taken up to evolve a procedure to form clusters of genotypes based on their interaction with environment when the error variances in different environments are heterogeneous.

Review of Literature

2. REVIEW OF LITERATURE

Genotype environment interaction has long been known to occur and various methods have been proposed for analysing GE interaction statistically. Clustering approach to study GE interaction is very recent and consequently the literature in this area is limited.

Yates and Cochran (1938) used the regression approach to study the GE interaction in a varietal trial on barley. They regressed the yield of each variety on the environmental means and observed that the regression sum of squares accounted for a large part of the interaction sum of squares.

Many workers (Wood, 1976; Tai, 1971; Eberhart and Russell, 1966; Perkins and Jinks, 1968; Fripp and Caten, 1971; Hardwick and Wood, 1972 etc.) suggested variations of regression approach to study GE interaction. The simplicity of these procedures might have made them so popular among plant breeders for the study of interaction.

Bala Krishnan _et_ _al_. (1978) applied the regression techniques of Finlay and Wilkinson (1963) and Eberhart and Russell (1966) in Pineapple. Suresh Babu (1981) used the statistical techniques proposed by Finlay and Wilkinson (1963), Eberhart and Russell (1966) and Perkins and Jinks (1968) to estimate stability parameters and GE interaction in Bhindi. Devadas (1982) used the stability analysis of Eberhart and Russell (1966) in amaranth varieties. Sulochana (1984) used the stability analysis of Eberhart and Russell (1966) in Cowpea. Ibrahim et al. (1985) applied the regression techniques of Finlay and Wilkinson (1963) and Eberhart and Russell (1966) in black pepper. Ushamani (1987) applied the stability analysis of Eberhart and Russell (1966) and Perkins and Jinks (1968a) in brinjal. Mini (1989) studied different genotypic stability analysis in detail in sesame.

Wricke (1966) proposed ecovalence ratio (W_i) of genotypes grown under several environments as a measure of stability of performance. Ecovalence (W_i) was defined as the percentage contribution of the i^{th} genotype to the genotype-environment interaction sum of squares. The varieties with small W_i values were considered as stable. By this method the genotype-environment interaction sum of squares can be partitioned into components attributable to different [illegible]es, though the performance of genotypes over environments [illegible] be predicted. Ushamani (1987) used this ecovalence ratio [illegible] ess the stability of performance of genotypes in an experiment on brinjal.

Shukla (1972) proposed stability variance (σ_i^2) as a measure of stability of i^{th} genotype and developed an F test taking into account the environmental component of variance (σ_e^2). If the stability variance of genotype and environmental component of variances are equal, that genotype was as judged stable.

John (1984) studied the genotypic stability analysis of Eberhart and Russell (1966), Perkins and Jinks (1968), Freeman and Perkins (1971), Wricke (1966) and Shukla (1972) in detail and pointed out the drawbacks in the analysis of variance of Eberhart and Russell and Perkins and Jinks. She suggested that the regression methods could satisfactorily be used with larger number of genotypes provided the regression explains a substantial part of GE interaction. When regression cannot explain large part of the GE interaction Wricke's ecovalence ratio or Shukla's stability variance could satisfactorily be used.

Lin and Thompson (1975) extended the regression approach to clustering genotypes based on GE interaction. They defined a dissimilarity measure for any subset of 't' genotypes as the variance ratio for testing the null hypothesis of a common regression line against the alternative hypothesis of 't' independent regressions. They proved that this dissimilarity measure equalled the mean of measures for all possible pairs of genotypes in the subset. Thus ᑋe index conformed to the conditions set by Sokal and Michener ı958) for use of their unweighted group average link strategy for clustering.

Lin (1982) proposed a cluster method to group genotypes according to their response to the environments. He defined a dissimilarity index between a pair of genotypes and used Sokal and Michener's (1958) unweighted pair group method in the clustering algorithm. The index was given by

$$d(i, i') = \frac{1}{2(n-1)} \sum_{j=1}^{n} [(Y_{ij} - \bar{Y}_{i.}) - (Y_{i'j} - \bar{Y}_{i'.})]^2$$

where Y_{ij} is the observed mean value of i^{th} genotype in the j^{th} environment ($j = 1, 2 \ldots\ldots\ldots\ldots n$) and $\bar{Y}_{i.}$ is the mean of the i^{th} genotype over 'n' environments. He showed that the dissimilarity index in a cluster of genotypes is nothing but the within group GE interaction mean square, under a two way analysis of variance.

Ramey and Rosielle (1983) used the dissimilarity index proposed by Lin (1982) and proposed a hierarchical agglomerative sums of squares method for clustering genotypes or environments in presence of genotype x environment interaction. The procedure consisted of minimising overall genotype x environment interaction mean squares within cluster at each fusion cycle.

Suresh (1986) proposed a computer oriented interative algorithm for clustering genotype using Mahalanobis D^2 statistic as the dissimilarity index.

Sreekala (1989) proposed three procedures for clustering genotypes based on their interaction with environment using Lin's (1982) dissimilarity index and Mahalanobis D^2 statistic.

Methodology

3. METHODOLOGY

Procedures for clustering genotypes based on their interaction with environments have been suggested by various workers in situation when error variances in different environments are homogeneous. Herein procedures for clustering genotypes when the error variances are heterogeneous in different environments are proposed.

Let us assume that 't' genotypes are tried in each of the 's' environments in a randomised block design with 'r' replications and that the error variances in the 's' environments are found heterogeneous using Bartlett's χ^2 test.

Next step is to have a weighted analysis of variance of the data pooled over the environments in order to test the presence or absence of GE interaction. The weighted analysis can be arised as follows (Panse and Sukhatme, 1978).

Weighted Analysis of Variance of the Pooled Data

Source	Sum of Squares
Total	$\sum_{j=1}^{s} W_j S_j - C$
Environments	$\frac{1}{t} \sum_{j=1}^{s} W_j P_j^2 - C$
Genotypes	$\dfrac{\sum_{i=1}^{t} \left[\sum_{j=1}^{s} W_j Y_{ij} \right]^2}{\sum_{j=1}^{s} W_j} - C$
GE interaction	Total SS - Environments SS - Genotypes SS

where,

Y_{ij} is the mean observation of i^{th} genotype in j^{th} environment

$W_j = r/s_j^2$,

s_j^2 is the error mean square in the j^{th} environment,

$$S_j = \sum_{i=1}^{t} Y_{ij}^2,$$

$$P_j = \sum_{i=1}^{t} Y_{ij},$$

$$G = \sum_{i=1}^{t} \sum_{j=1}^{s} W_j Y_{ij},$$

$$C = \frac{G^2}{t \sum_{j=1}^{s} W_j} .$$

Significance of GE interaction is tested using the χ^2 test

$$\chi^2 = \frac{(n-4)(n-2)\ I_t}{n(n+t-3)} \qquad \text{.............. (3.i)}$$

with degrees of freedom $\frac{(s-1)(t-1)(n-4)}{(n+t-3)}$, where

n is the error degrees of freedom in each environment,

I_t is the interaction sum of squares, which is given by

$$I_t = \sum_{j=1}^{s} W_j S_j - \frac{\sum_{i=1}^{t} \left[\sum_{j=1}^{s} W_j Y_{ij} \right]^2}{\sum_{j=1}^{s} W_j} - \frac{1}{t} \sum_{j=1}^{s} W_j P_j^2 + \frac{\left[\sum_{j=1}^{s} W_j P_j \right]^2}{t \sum_{j=1}^{s} W_j} \qquad \text{...(3.2)}$$

The first step in any clustering procedure is to define a distance function between every pair of members available. For this purpose, the GE interaction sum of squares, when only two genotypes, say 'i' and 'i'' are involved, can be expressed as

$$I_2 = \sum_{j=1}^{s} W_j(Y_{ij}^2 + Y_{i'j}^2) - \tfrac{1}{2}\sum_{j=1}^{s} W_j(Y_{ij} + Y_{i'j})^2 -$$

$$\left[\frac{\left(\sum_{j=1}^{s} W_jY_{ij}\right)^2 + \left(\sum_{j=1}^{s} W_jY_{i'j}\right)^2}{\sum_{j=1}^{s} W_j} - \frac{\left(\sum_{j=1}^{s} W_j(Y_{ij} + Y_{i'j})\right)^2}{2\sum_{j=1}^{s} W_j}\right]$$

$$= \sum_{j=1}^{s} W_j\left(Y_{ij}^2 + Y_{i'j}^2 - \frac{(Y_{ij}+Y_{i'j})^2}{2}\right) - \frac{1}{2\sum_{j=1}^{s} W_j}\left[2\sum_{j=1}^{s} W_j^2Y_{ij}^2 +\right.$$

$$2\sum_{j=1}^{s} W_j^2Y_{i'j}^2 - \sum_{j=1}^{s} W_j^2(Y_{ij} + Y_{i'j})^2 + 4\sum_{j<k} W_jW_kY_{ij}Y_{ik} +$$

$$\left.4\sum_{j<k} W_jW_kY_{i'j}Y_{i'k} - 2\sum_{j<k} W_jW_k(Y_{ij} + Y_{i'j})(Y_{ik} + Y_{i'k})\right]$$

$$= \sum_{j=1}^{s} W_j\frac{(Y_{ij} - Y_{i'j})^2}{2} - \frac{1}{2\sum_{j=1}^{s} W_j}\left[\sum_{j=1}^{s} W_j^2(Y_{ij} - Y_{i'j})^2 +\right.$$

$$\left.2\sum_{j<k} W_jW_k(Y_{ij} - Y_{i'j})(Y_{ik} - Y_{i'k})\right]$$

$$= \sum_{j=1}^{s} W_j (Y_{ij} - Y_{i'j})^2 - \frac{1}{2 \sum_{j=1}^{s} W_j} \left[\sum_{j=1}^{s} W_j^2 (Y_{ij} - Y_{i'j})^2 + 2 \sum_{j<k} W_j W_k (Y_{ij} - Y_{i'j}) (Y_{ik} - Y_{i'k}) \right]$$

$$= \sum_{j=1}^{s} W_j \frac{(Y_{ij} - Y_{i'j})^2}{2} - \frac{1}{2 \sum_{j=1}^{s} W_j} \left(\sum_{j=1}^{s} W_j (Y_{ij} - Y_{i'j}) \right)^2$$

$$= \frac{1}{2} \sum_{j=1}^{s} W_j ((Y_{ij} - \bar{Y}_{i\cdot}) - (Y_{i'j} - \bar{Y}_{i'\cdot}))^2$$

where $\bar{Y}_{i\cdot} = \dfrac{\sum_{j=1}^{s} W_j Y_{ij}}{\sum_{j=1}^{s} W_j}$ and $\bar{Y}_{i.} = \dfrac{\sum_{j=1}^{s} W_j Y_{i'j}}{\sum_{j=1}^{s} W_j}$

This interaction SS can be taken as the dissimilarity index D(i,i') between genotypes i and i'.

$$D(i,i') = \frac{1}{2} \sum_{j=1}^{s} W_j ((Y_{ij} - \bar{Y}_{i.}) - (Y_{i'j} - \bar{Y}_{i'.}))^2 \quad \text{...........} \quad (3.3)$$

The interaction sum squares for a group of 'r' genotypes is given by

$$I_r = \sum_{j=1}^{s} W_j \left(\sum_{i=1}^{r} Y_{ij}^2 \right) - \frac{1}{r} \sum_{j=1}^{s} W_j \left(\sum_{i=1}^{r} Y_{ij} \right)^2 - \frac{\sum_{i=1}^{r} \left(\sum_{j=1}^{s} W_j Y_{ij} \right)^2}{\sum_{j=1}^{s} W_j} +$$

$$\frac{1}{r}\;\frac{\left(\sum_{j=1}^{s} W_j\left(\sum_{i=1}^{r} Y_{ij}\right)\right)^2}{\sum_{j=1}^{s} W_j}$$

$= L_1 - L_2$, where,

$$L_1 = \sum_{j=1}^{s} W_j\left(\sum_{i=1}^{r} Y_{ij}^2\right) - \frac{1}{r} \sum_{j=1}^{s} W_j\left(\sum_{i=1}^{r} Y_{ij}\right)^2$$

$$= \sum_{j=1}^{s} W_j\left(\sum_{i=1}^{r} Y_{ij}^2 - \frac{1}{r}\left(\sum_{i=1}^{r} Y_{ij}\right)^2\right)$$

$$= \frac{1}{r} \sum_{j=1}^{s} W_j\left(r\sum_{i=1}^{r} Y_{ij}^2 - \left(\sum_{i=1}^{r} Y_{ij}\right)^2\right)$$

$$= \frac{1}{r} \sum_{j=1}^{s} W_j\left((r-1) \sum_{i=1}^{r} Y_{ij}^2 - 2 \sum_{i<i'} Y_{ij}Y_{i'j}\right)$$

$$= \frac{1}{r} \sum_{j=1}^{s} W_j\left(\sum_{i<i'=1}^{r} (Y_{ij} - Y_{i'j})^2\right)$$

$$= \frac{1}{r} \sum_{i<i'=1}^{r} \sum_{j=1}^{s} W_j(Y_{ij} - Y_{i'j})^2 \qquad \text{...........(3.4) and}$$

$$L_2 = \frac{\sum_{i=1}^{r} \left(\sum_{j=1}^{s} W_jY_{ij}\right)^2}{\sum_{j=1}^{s} W_j} - \frac{1}{r} \frac{\left(\sum_{j=1}^{s} W_j \sum_{i=1}^{r} Y_{ij}\right)^2}{\sum_{j=1}^{s} W_j}$$

$$= \sum_{i=1}^{r} \frac{\left(\sum_{j=1}^{s} W_j^2Y_{ij}^2 + 2 \sum_{j<k} W_jW_kY_{ij}Y_{ik}\right)}{\sum_{j=1}^{s} W_j} -$$

$$\frac{1}{r}\left[\frac{\sum_{j=1}^{s} W_j^{2}(\sum_{i=1}^{r} Y_{ij})^{2} + 2\sum_{j<k} W_jW_k(\sum_{i=1}^{r} Y_{ij})(\sum_{i=1}^{r} Y_{ik})}{\sum_{j=1}^{s} W_j}\right]$$

$$= \frac{1}{\sum_{j=1}^{s} W_j}\left[\sum_{j=1}^{s} W_j^{2}(\sum_{i=1}^{r} Y_{ij}^{2} - \frac{1}{r}(\sum_{i=1}^{r} Y_{ij})^{2}) + \right.$$

$$\left. 2\sum_{j<k} W_jW_k(\sum_{i=1}^{r} Y_{ij}Y_{ik} - \frac{1}{r}(\sum_{i=1}^{r} Y_{ij})(\sum_{i=1}^{r} Y_{ik}))\right]$$

$$= \frac{1}{r\sum_{j=1}^{s} W_j}\left[\sum_{j=1}^{s} W_j^{2}((r-1)\sum_{i=1}^{r} Y_{ij}^{2} - 2\sum_{i<i'=1}^{r} Y_{ij}Y_{i'j}) + \right.$$

$$\left. 2\sum_{j<k} W_jW_k((r-1)\sum_{i=1}^{r} Y_{ij}Y_{ik} - 2\sum_{i<i'=1}^{r} Y_{ij}Y_{i'k})\right]$$

$$= \frac{1}{r\sum_{j=1}^{s} W_j}\left[\sum_{j=1}^{s} W_j^{2}\sum_{i<i'=1}^{r}(Y_{ij} - Y_{i'j})^{2} + 2\sum_{j<k} W_jW_k\sum_{i<i'=1}^{r}\right.$$

$$\left.(Y_{ij} - Y_{i'j})(Y_{ik} - Y_{i'k})\right]$$

$$= \frac{1}{r\sum_{j=1}^{s} W_j}\left[\sum_{i<i'=1}^{r}(\sum_{j=1}^{s} W_j^{2}(Y_{ij} - Y_{i'j})^{2} + 2\sum_{j<k} W_jW_k\right.$$

$$\left.(Y_{ij} - Y_{i'j})(Y_{ik} - Y_{i'k}))\right]$$

$$= \frac{1}{r\sum_{j=1}^{s} W_j}\sum_{i<i'=1}^{r}(\sum_{j=1}^{s} W_j(Y_{ij} - Y_{i'j}))^{2} \qquad \ldots\ldots\ldots\ldots\ldots(3.5)$$

Using (3.3), (3.4) and (3.5)

$$I_r = \frac{1}{r} \sum_{i < i'=1}^{r} \sum_{j=1}^{s} W_j \left((Y_{ij} - \bar{Y}_{i.}) - (Y_{i'j} - \bar{Y}_{i'.}) \right)^2$$

$$= \frac{2}{r} \sum_{i < i'=1}^{r} D(i,i')$$

This interaction sum of squares can be taken as the dissimilarity index for any subset of 'r' genotypes.

i.e. $$D(1,2,3,\ldots\ldots\ldots,r) = \frac{2}{r} \sum_{i < i'=1}^{r} D(i,i') \qquad \ldots\ldots\ldots\ldots(3.6)$$

3.A. STATISTICAL CLUSTERING

By the method of statistical clustering, it is envisaged to identify maximum subsets of genotypes such that GE interaction is not significant with in any subset while any addition to the set makes the GE interaction significant.

The procedure proposed for the purpose is explained in few steps below.

(a) The dissimilarity indices for every pair of genotypes are calculated using (3.3).

(b) The pair of genotypes having smallest index value is identified. This index gives the interaction sum of squares between the two genotypes.

(c) Obtain the value of Chi-square statistic for testing the significance of GE interaction using (3.1).

(d) If the Chi-square value is insignificant these genotypes are grouped together.

(e) Each of the remaining genotypes is allocated to the group and the corresponding dissimilarity index for the group is calculated using (3.6). Identify the genotype which gives the smallest dissimilarity index value and repeat steps (c) and (d).

(f) Continue the process in (e) until the Chi-square value becomes significant. Thus formation of the first cluster is completed excluding the genotype last entered which leads to significant Chi-square.

(g) The genotype from among those which are excluded from the cluster/clusters already formed having the least index value with any of the remaining genotypes is identified.

(h) Repeat (c) to (f) irrespective of whether a genotype is included or not in the earlier cluster/clusters.

(i) Repeat steps (g) and (h) until all the genotypes are exhausted.

3.B. CLUSTERING GENOTYPES BY MINIMISING AVERAGE WITHIN CLUSTER GE INTERACTION SUM OF SQUARES USING ITERATIVE RELOCATION ALGORITHM

This method aims at obtaining an optimum cluster configuration which achieves the minimum within cluster GE interaction sum of squares. The proposed procedure is explained in the following steps.

(a) Dissimilarity indices for every pair of genotypes are calculated and the pair of genotypes having maximum index value between them are identified. These two genotypes are considered as the nuclei of two clusters.

(b) Allocate each of the genotype to these clusters in such a way that the index value with the nucleus genotype is minimum.

(c) To increase the number of clusters, by one, identify the two genotypes having maximum index value within the clusters and these genotypes are taken as the nuclei of two clusters in addition to the nuclei of the existing clusters except the clusters containing the newly identified genotypes. Repeat (b).

(d) Repeat (c) until the desired number of clusters is arrived at.

The clustering so obtained for any specific number of clusters can be further optimised by the following iterative relocation algorithm.

(1) Number the genotypes from 1 to t.

(2) Take out genotype 1 from the cluster it belongs to. Allocate it to each of the clusters and calculate the weighted arithmetic mean of the dissimilarity indices of the clusters, weights being the multiplier used to obtain the Chi-square value (3.1). This genotype is fused with the cluster for which the weighted arithmetic mean of the dissimilarity indices is minimum.

(3) Repeat the process with all genotypes.

(4) Repeat steps (1) to (3) until two successive iterations give identical clustering.

OPTIMUM NUMBER OF CLUSTERS

A major problem in this type of clustering is that of finding the number of clusters into which the genotypes or environments are to be grouped. Determination of optimum number of clusters is proposed to be done by the method of maximum curvature as follows.

A graph of weighted arithmetic mean of the dissimilarity indices of the clusters against the number of clusters is drawn with number of clusters on X axis. The point on the X axis which is just beyond the point of maximum curvature of the graph can be taken as the optimum number of clusters.

Illustration

4. ILLUSTRATION

Two sets of secondary data have been made use for illustration of the methodology developed here in. The first set of data was taken from Vahab (1989). They consist of observations on mean yield per plant from 2 replications of an experiment on 55 genotypes of bittergourd, conducted in randomised block design over 3 seasons, at the Department of Olericulture, College of Horticulture, Vellanikkara and are given in Appendix I.

Observations on mean yield per plant from an experiment of 9 genotypes (6 varieties and 3 F_1 hybrids) of brinjal conducted in randomised block design with three replications in 4 seasons, at Department of Olericulture, College of Horticulture, Vellanikkara form the second set of data and are given in Appendix II. These were taken from Varghese (1992).

4.A. FIFTY FIVE GENOTYPES OF BITTERGOURD IN 3 ENVIRONMENTS

The error mean squares (EMS) in the analysis of variance carried out in the three different environments are given below.

Season	1	2	3
EMS	0.0334	0.0873	0.0136

These error mean squares were tested for homogeneity using Bartlett's Chi-square test and were found hetrogeneous. Hence to test GE interaction, weighted analysis of pooled data was carried out and is given below.

Table 4.1 WEIGHTED ANALYSIS OF VARIANCE OF THE POOLED DATA

Source	S.S.
Total	83573.56071
Genotypes	66500.86879
Environments	10280.56263
GE interaction	6792.12475

Calculated value of Chi-square was 3085.18 with 51 degrees of freedem and hence GE interaction was significant. The genotypes were grouped based on their interaction. With environments by the two methods described in 3.A and 3.B.

The dissimilarity indices for every pair of genotypes determined using (3.3) are given in Appendix III.

4.A.1. STATISTICAL CLUSTERING

Cluster configurations obtained using the procedure given in 3.A is given in Table 4.2.

Table 4.2 CLUSTER CONFIGURATION BY STATISTICAL CLUSTERING

Cluster No.	No. of genotypes	Genotypes in the cluster
1	2	3
1	3	42, 27, 32
2	3	36, 29, 30
3	4	40, 7, 50, 16

contd.

Table 4.2 Continued

1	2	3
4	3	47, 34, 55
5	4	48, 46, 22, 28
6	4	37, 3, 9, 53
7	4	10, 6, 1, 2
8	3	21, 18, 52
9	4	31, 25, 26, 53
10	3	11, 16, 2
11	4	13, 22, 46, 48
12	4	39, 2, 1, 10
13	3	33, 12, 20
14	3	5, 6, 10
15	2	54, 51
16	3	14, 47, 34
17	3	8, 2, 1
18	2	41, 4
19	3	19, 36, 29
20	2	43, 14
21	3	35, 5, 6
22	3	45, 19, 36
23	2	49, 13
24	1	15
25	1	17
26	1	23
27	1	24
28	1	38
29	1	44

4.A.2 ITERATIVE RELOCATION ALGORITHM FOR OPTIMUM CLUSTERING

Eleven clusters were formed using IRA procedure. The cluster configuration are given in Table 4.3.

Table 4.3 CLUSTER CONFIRGURATIONS USING IRA

	Serial No. of clusters	Genotypes in the cluster	Weighted AM of average intera 'D'	No. of iteration
	1	2	3	4
		Two clusters		
Initial	1	1 2 3 4 5 6 7 8 9 10 11 12 14 16 17 18 20 21 24 25 26 31 33 34 35 37 39 40 41 43 47 49 50 51 52 53 54 55		
	2	13 15 19 22 23 27 28 29 30 32 36 38 42 44 45 46 48	1255.769	
Final	1	1 2 3 4 5 6 7 8 9 10 11 14 16 17 18 20 21 24 25 26 31 33 34 35 37 39 40 41 43 47 49 50 51 52 53 54 55		
	2	12 13 15 19 22 23 27 28 29 30 32 36 38 42 45 46 48	1254.926	2

Contd.

Table 4.3 Continued

	1	2	3	4
		Three clusters		
Initial	1	3 12 13 14 15 17 18 19 20 21 22 25 26 27 28 29 31 32 33 34 35 36 37 42 43 45 46 47 48 49 52 53 55		
	2	23 30 38 44		
	3	1 2 4 5 6 7 8 9 10 11 16 24 39 40 41 50 51 54	725.703	
Final	1	12 13 14 15 18 20 21 22 27 28 31 32 33 34 42 46 47 48 49 52 55		
	2	19 23 29 30 36 38 44 45		
	3	1 2 3 4 5 6 7 8 9 10 11 16 17 24 25 26 35 37 39 40 41 43 50 51 53 54	560.816	3
		Four clusters		
Initial	1	1 2 3 4 5 6 7 8 9 10 11 12 14 17 18 20 21 25 26 31 33 34 35 37 39 41 43 47 49 51 52 53 55		
	2	15 23 38		

Contd.

Table 4.3 Continued

	1	2	3	4
	3	7 16 24 40 50 54		
	4	13 19 22 27 28 29 30 32 36 42 44 45 46 48	392.014	
Final	1	12 14 17 18 20 21 25 26 31 33 34 37 43 47 49 52 53 55		
	2	15 23 38		
	3	1 2 3 4 5 6 7 8 9 10 11 16 24 35 39 40 41 50 51 54		
	4	13 19 22 27 28 29 30 32 36 42 44 45 46 48	314.384	5
		<u>Five clusters</u>		
Initial	1	1 2 3 4 5 6 8 9 10 14 17 35 39 41 43		
	2	29 38		
	3	7 11 16 24 40 50 51 54		
	4	19 27 29 30 32 36 42 44 45		
	5	12 13 15 18 20 21 22 25 26 28 31 33 34 37 46 47 48 49 52 53 55	246.504	

Contd.

Table 4.3 Continued

	1	2	3	4
Final	1	1 2 3 4 5 6 8 9 10 17 35 39 41		
	2	15 23 38		
	3	7 11 16 24 40 50 51 54		
	4	13 19 22 27 28 29 30 32 36 42 44 45 46 48		
	5	12 14 18 20 21 25 26 31 33 34 37 43 47 49 52 53 55	217.856	6
		<u>Six clusters</u>		
Initial	1	4 5 6 8 10 17 41		
	2	23		
	3	7 16 24 40 50		
	4	19 27 29 30 32 36 42 44 45		
	5	13 15 22 28 38 46 48 49		
	6	1 2 3 9 11 12 14 18 20 21 25 26 31 33 34 35 37 39 43 47 51 52 53 54 55	204.231	
Final	1	1 2 3 4 5 6 8 9 10 17 35 39 41		

Contd.

Table 4.3 Continued

	1	2	3	4
	2	23 38		
	3	7 11 16 24 40 50 51 54		
	4	19 27 29 30 32 36 42 44 45		
	5	13 15 22 28 46 48 49		
	6	12 14 18 20 21 25 26 31 33 34 37 43 47 52 53 55	131.025	4
		Seven clusters		
Initial	1	4 5 6 10 17 41		
	2	23 38		
	3	7 24 40 50		
	4	19 27 29 30 36 42 44 45		
	5	13 15 21 22 28 46 48 49		
	6	1 2 3 8 9 11 16 26 35 37 39 51 53 54		
	7	12 14 18 20 25 31 32 33 34 43 47 52 55	101.174	
Final	1	1 2 4 5 6 10 17 35 41		
	2	23 38		
	3	2 7 8 11 16 24 40 50		
	4	19 27 29 30 32 36 42 44 45		

Contd.

Table 4.3 Continued

	1	2	3	4
	5	13 15 22 28 46 48 49		
	6	3 9 25 26 37 39 51 53 54		
	7	12 14 18 20 21 31 33 34 43 47 52 55	89.885	4
		<u>Eight clusters</u>		
Initial	1	4 5 6 10 41		
	2	23 38		
	3	7 24 40 50		
	4	19 29 30 36 44 45		
	5	13 15 22 28 46 48 49		
	6	1 2 8 11 16 39 51 54		
	7	3 9 14 17 18 20 21 25 26 31 33 34 35 37 43 47 52 53		
	8	12 27 32 42 55	73.161	
Final	1	1 4 5 6 10 17 35 41		
	2	23 38		
	3	2 7 8 11 16 24 40 50		
	4	19 29 30 36 44 45		
	5	13 15 22 28 46 48 49		
	6	3 9 25 26 37 39 51 53 54		
	7	12 14 18 20 21 31 33 34 43 47 52 55		
	8	27 32 42	62.363	5

Contd.

Table 4.3 Continued

	1	2	3	4
		Nine clusters		
Initial	1	4 5 6 10 41		
	2	23		
	3	24 40		
	4	19 29 30 36 44 45		
	5	15 38		
	6	1 2 7 8 9 11 16 39 50 51 54		
	7	12 27 28 32 33 42		
	8	3 14 17 25 26 31 34 35 37 43 47 52 53 55		
	9	13 18 20 21 22 46 48 49	78.930	
Final	1	1 4 5 6 10 17 35 41		
	2	23		
	3	2 7 8 11 16 24 40 50		
	4	19 29 30 36 44 45		
	5	15 38		
	6	3 9 25 26 37 39 51 53 54		
	7	27 32 42		
	8	12 14 18 20 21 31 33 34 43 47 52 55		
	9	13 22 28 46 48 49	56.733	5

Contd.

Table 4.3 Continued

	1	2	3	4
		Ten clusters		
Initial	1	4 41		
	2	23		
	3	7 24 40		
	4	19 27 29 30 36 42 44 45		
	5	15 38		
	6	1 2 5 6 8 10 11 16 17 35 39 50		
	7	12 20 32 33 52		
	8	3 9 18 21 25 26 31 37 43 51 53 54		
	9	13 22 28 46 48 49		
	10	14 34 47 55	56.113	
Final	1	4 5 41		
	2	23		
	3	7 16 24 40 50		
	4	19 29 30 36 44 45		
	5	15 38		
	6	1 2 6 8 10 11 35 39		
	7	12 20 27 32 33 42		
	8	3 9 18 21 25 26 31 37 51 52 53 54		

Contd.

Table 4.3 Continued

	1	2	3	4
	9	13 22 28 46 48 49		
	10	14 17 34 43 47 55	46.916	2
		<u>Eleven clusters</u>		
Initial	1	4 5 6 10 41		
	2	23		
	3	7 24 40		
	4	19 27 29 30 36 42 44 45		
	5	15 38		
	6	17 35		
	7	12 20 32 33 52		
	8	1 3 9 18 21 25 26 31 37 39 51 53 54		
	9	13 22 28 46 48 49		
	10	14 34 43 47 55		
	11	2 8 11 16 50	44.728	
Final	1	4 5 6 10 41		
	2	23		
	3	7 24 40 50		
	4•	19 29 30 36 44 45		
	5	15 38		
	6	14 17 35 43		
	7	27 32 42		

Contd.

Table 4.3 Continued

1	2	3	4
8	3 9 18 21 25 26 31 37 51 53 54		
9	13 22 28 46 48 49		
10	12 20 33 34 47 52 55		
11	1 2 8 11 16 39	33.336	4

To determine the optimum number of clusters, the graph of weighted arithmetic mean of average intracluster D values was drawn against the number of clusters and is provided in Fig. 4.1. The optimum number of clusters obtained by the method of maximum curvature is 5.

4.B. NINE GENOTYPES OF BRINJAL IN 4 ENVIRONMENTS

The error mean squares (EMS) obtained from the analysis of variance in different environments are as follows.

Season	1	2	3	4
EMS	315838.24	35418.00	26078	8761.63

They were tested for homogeneity using Bartlett's Chi-square test and were found heterogeneous. The weighted analysis of variance of pooled data is given in Table 4.4.

Table 4.4 WEIGHTED ANALYSIS OF VARIANCE OF THE POOLED DATA

Source	S.S.
Total	711.3815
Genotypes	300.0143
Environments	42.3459
GE interaction	369.0213

The calculated value of Chi-square was 176.1238 with 13 degrees of freedom and hence the GE interaction was significant. The genotypes were then clustered based on their interaction with environments using methods described in 3.A and 3.B.

The dissimilarity indices between every pair of genotypes determined are given in Appendix IV.

4.B.1 STATISTICAL CLUSTERING

Cluster configuration obtained using the clustering procedure given in 3.A is given in Table 4.5.

Table 4.5 CLUSTERS OBTAINED BY STATISTICAL CLUSTERING

Cluster No.	No. of genotypes	Genotypes in the cluster
1	3	5, 3, 4
2	2	6, 2
3	3	9, 5, 3
4	3	1, 3, 5
5	3	8, 1, 3
6	1	7

4.B.2 ITERATIVE RELOCATION ALGORITHM FOR OPTIMUM CLUSTERING

Clusters obtained by IRA procedure is given in Table 4.6.

Table 4.6 CLUSTERS CONFIGURATIONS BY IRA

	Serial No. of cluster	Genotypes in the cluster	Weighted A.M. of average intra-cluster	No. of iteration
	1	2	3	4
		Two clusters		
Initial	1	2 6 7		
	2	1 3 4 5 8 9	52.9474	
Final	1	2 6 7		
	2	1 3 4 5 8 9	52.9474	1
		Three clusters		
Initial	1	7		
	2	1 3 4 5 8		
	3	2 6 9	30.3798	
Final	1	2 7		
	2	1 3 4 5 8		
	3	6 9	27.2883	2
		Four clusters		
Initial	1	7		
	2	1 4 8		
	3	2 6		
	4	3 5 9	9.3162	

Contd.

Table 4.6 Continued

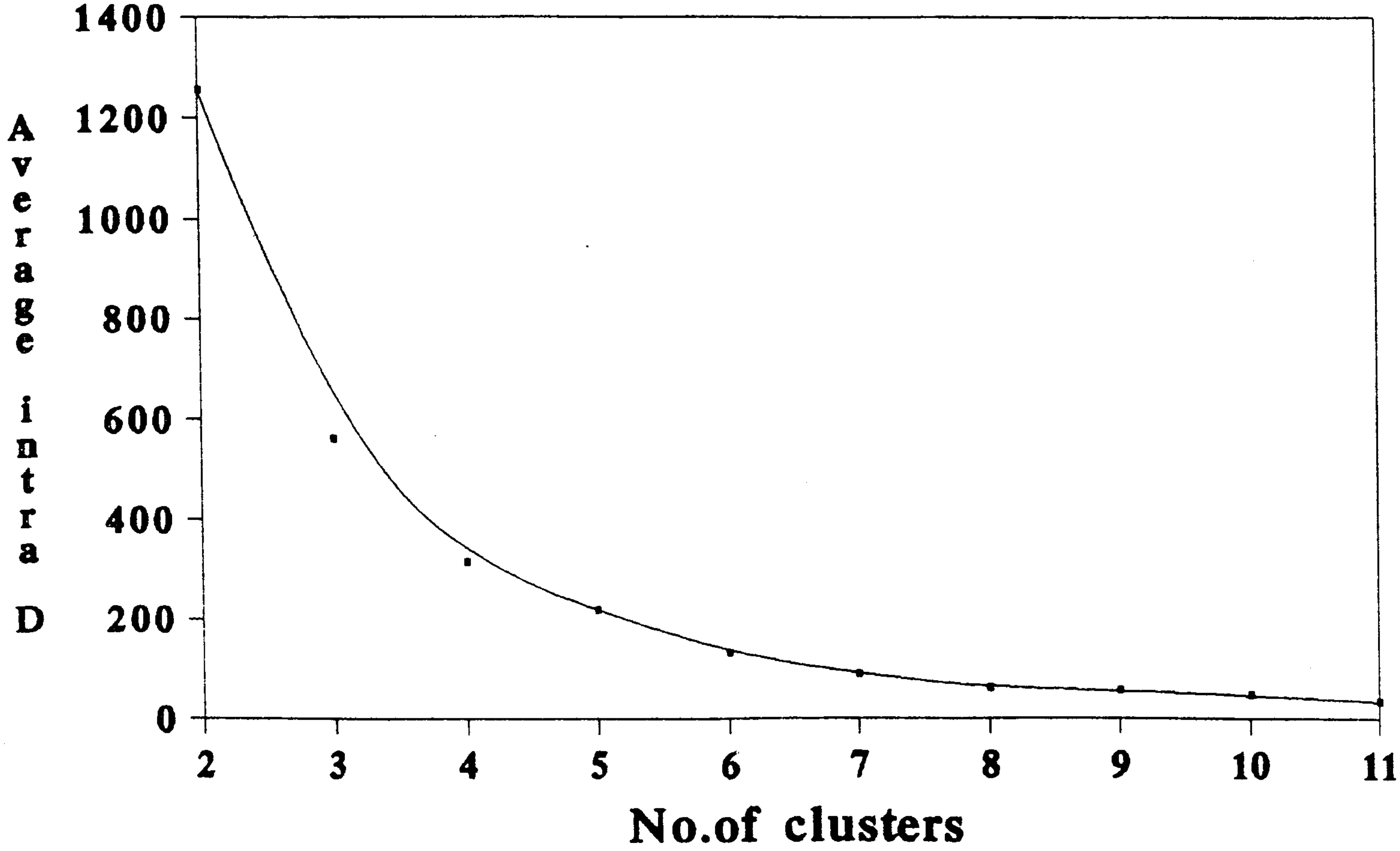
	1	2	3	4
Final	1	7		
	2	1 8		
	3	2 6		
	4	3 4 5 9	7.9885	2
		Five clusters		
Initial	1	7		
	2	1 8		
	3	2 6		
	4	9		
	5	3 4 5	4.5632	
Final	1	7		
	2	1 8		
	3	2 6		
	4	3 9		
	5	4 5	4.3525	2
		Six clusters		
Initial	1	7		
	2	1		
	3	2 6		
	4	9		
	5	3 4 5		
	6	8	3.7776	

Contd.

Table 4.6 Continued

	1	2	3	4
Final	1	7		
	2	1 3		
	3	2 6		
	4	9		
	5	4 5		
	6	8	3.3716	2
		Seven clusters		
Initial	1	7		
	2	1		
	3	2 6		
	4	9		
	5	4		
	6	8		
	7	3 5	1.9305	
Final	1	7		
	2	1		
	3	2 6		
	4	9		
	5	4		
	6	8		
	7	3 5	1.9305	1

The graph of weighted arithmetic mean of average intracluster D values against the number of clusters was drawn and is given in Fig. 4.2. The optimum number of clusters determined by the method of maximum curvature is 5.



Dependence of average intracluster D on No. of clusters

Fig.4.1

Dependence of average intracluster D on No. of clusters

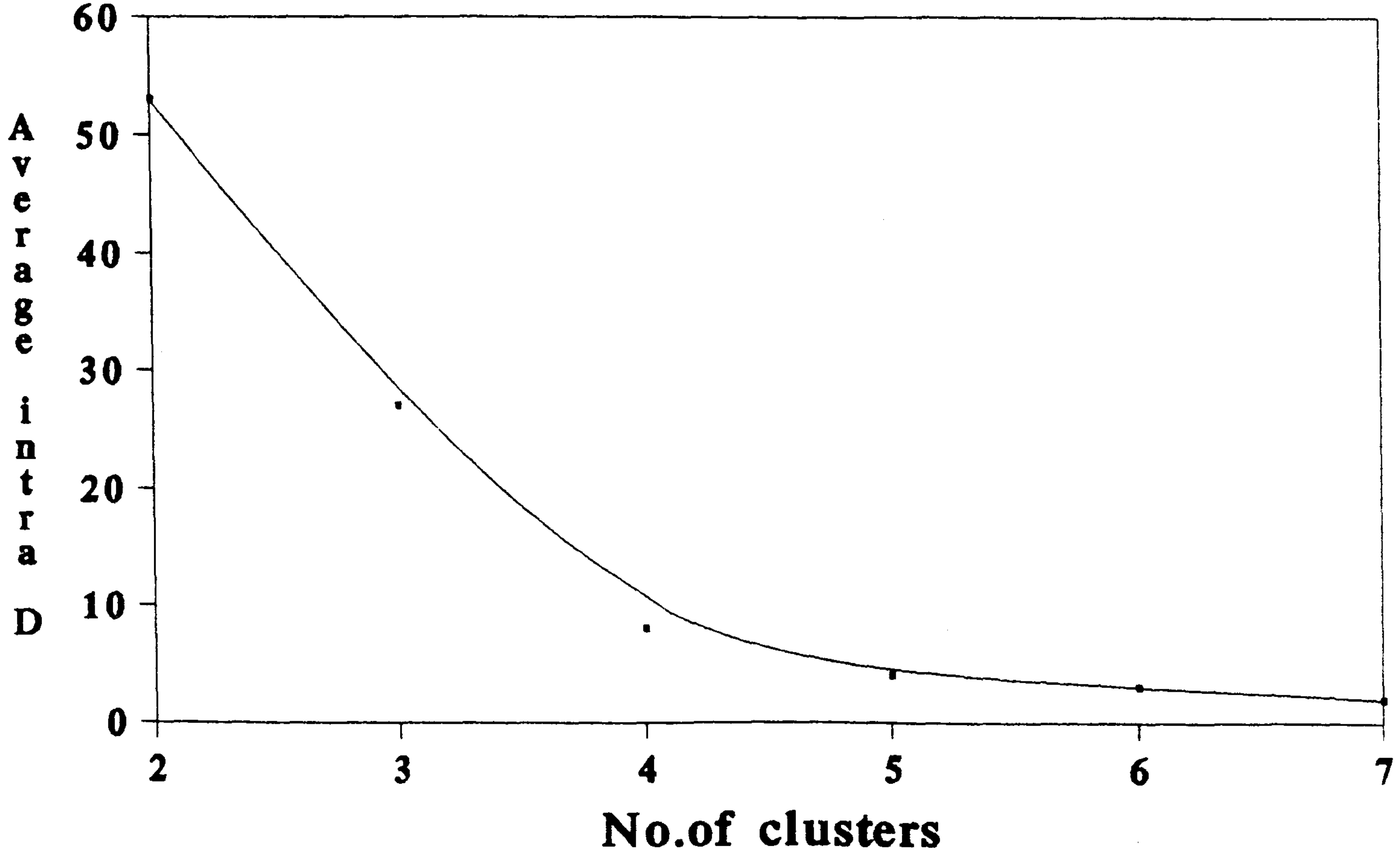


Fig.4.2

Discussion

5. DISCUSSION

Clustering genotypes or environments based on their interaction with environments are being very advantageously used by plant breeders. Clustering techniques based on GE interaction have been developed for situations when the error variances in different environments are homogeneous and these procedures fail in situations when the error variances are heterogeneous. Therefore some new procedures are developed herein to fill this gap.

A distance function between any pair of genotypes which measures their interaction with environments when the error variances in the different environments are heterogeneous was derived herein. The dissimilarity index for any group of genotypes was also derived from these distance measures between every pair of genotypes in the group and it also measures the within group genotype - environment interaction. This characteristic feature of the dissimilarity index is similar to that proposed by Lin (1982) for the case of homogeneous error variances. This property is of great advantage in formation of clusters in that the dissimilarity index of any group of genotypes which also measures within group genotype - environment interaction can be obtained directly from the pairwise distance values without going to the original observed values at every stage. Two methods of clustering using the newly defined dissimilarity index have been proposed herein.

One procedure called statistical clustering technique which aims at identifying maximum sets of genotypes such that the within set genotype - environment interaction is not significant in any set while addition of any genotype to the set makes the within set genotype - environment interaction significant. In other words this procedure does not aim at optimising an objective function unlike any general clustering procedure. What is aimed here is identification of genotypes which have similar response to differing environments. The cluster configuration resulting from this procedure may not be non-overlapping. This should not cause any concern to us, because it is possible that a genotype has a response to environments which is not significantly different from that of two (distinct) genotypes whose responses are significantly different. This is like the comparison of treatments after doing analysis of variance.

Another procedure - the (iterative) relocation algorithm aims at obtaining a cluster configuration by minimising the within cluster genotype - environment interaction sum of squares. This is justified in the sense that the objective of any clustering technique is to form groups of genotypes in such a way that those belonging to any group are similar with respect to some aspect in question while those belonging to different groups are dissimilar. This objective is met with by this procedure with respect to the genotype - environment interaction for any specific number of clusters. By this method the ultimate cluster configuration we get is non-overlapping as in the case of clustering in general. Determination of the optimum number of clusters to which the genotypes are to be grouped is proposed to be done by the graphical method of maximum curvature. It is to be noted that, in general, the objective function, viz., the weighted average of the within cluster genotype - environment interaction sum of squares decreases drastically with the number of clusters initially. With the increase in the number of clusters, the rate of decrease of the objective function also decreases. In other words for initial increase of the number of clusters the decrease in the objective function is substantial and then it decreases gradually. Therefore the point (number of clusters) just beyond the maximum curvature is proposed to be taken as the optimum number of clusters into which the genotypes are to be grouped, because for further increase in the number of clusters the decrease in the value of the objective function will only be marginal.

Though one procedure alone is named 'iterative relocation algorithm', in a way the same type of relocation algorithm is used in the statistical clustering also, though it is not iterative in nature.

So if the objective is to have cluster configuration which minimises the within cluster genotype - environment interaction, one may use the iterative relocation algorithm proposed herein. And if the objective is to identify genotypes having similar response to differing environments, which will be the case most commonly required the statistical clustering can be used.

Both the procedures have already been computerised and hence it will be easy for a user to adopt any of the procedures proposed here in making use of the programmes written in BASIC which are given in Appendix V and VI.

Summary

6. SUMMARY

Cluster analysis that forms clusters of individuals such that there is more homogeneity within clusters than those belonging to different clusters is a widely used technique by plant breeders. Very few attempts were made to group genotypes or environments based on genotype-environment interaction. But these procedures can only be used when the error variances in different environments are homogeneous. New procedures for clustering genotypes based on their interaction with environments which can be used when the error variances are heterogeneous are presented here.

A distance function between two genotypes i and i' which measures their interaction with environments was derived as

$$D(i,i') = \frac{1}{2} \sum_{j=1}^{s} W_j((Y_{ij} - \bar{Y}_{i.}) - (Y_{i'j} - \bar{Y}_{i'.}))^2 \qquad \ldots\ldots\ldots(1)$$

where Y_{ij} is the observed mean value of the i^{th} genotype in j^{th} environment and $Y_{i'j}$ that of i'^{th} genotype in j^{th} environment.

$\bar{Y}_{i.}$ and $\bar{Y}_{i'.}$ are given by

$$\bar{Y}_{i.} = \frac{\sum_{j=1}^{s} W_j Y_{ij}}{\sum_{j=1}^{s} W_j} \quad \text{and} \quad \bar{Y}_{i'.} = \frac{\sum_{j=1}^{s} W_j Y_{i'j}}{\sum_{j=1}^{s} W_j}$$

The dissimilarity index for any set of genotypes was derived from this pairwise distance function as

$$D(1,2,3,\ldots\ldots,r) = \frac{2}{r} \sum_{i<i'=1}^{r} D(i,i') \qquad \ldots\ldots.(2)$$

The dissimilarity indices so defined based on the genotype-environment interaction were utilised for two types of clustering.

6.A. STATISTICAL CLUSTERING

(a) The dissimilarity indices for every pair of genotypes are calculated using (1).

(b) The pair of genotypes having smallest index value is identified. This index gives the interaction sum of squares between the two genotypes.

(c) Calculate the value of the Chi-square statistic for testing the significance of GE interaction.

(d) If the Chi-square value is insignificant these genotypes are grouped.

(e) Each of the remaining genotypes is allocated to the group and the corresponding dissimilarity index for the group is calculated using (2). Identify the genotype which gives the smallest dissimilarity index value and repeat steps (c) and (d).

(f) Continue the procedure in (e) until the Chi-square value becomes significant. Thus formation of the first cluster is completed excluding the genotype last entered which leads to significant Chi-square.

(g) The genotypes having the smallest index value from the remaining indices are identified.

(h) Repeat (c) to (f) irrespective of whether a genotype is included or not in the earlier cluster/clusters.

(i) Repeat steps (g) and (h) until all the genotypes are exhausted.

6.B. MINIMISATION OF AVERAGE WITHIN CLUSTER GENOTYPE-ENVIRONMENT INTERACTION SUM OF SQUARES USING ITERATIVE RELOCATION ALGORITHM

(a) Dissimilarity indices for every pair of genotypes are calculated and the pair of genotypes having maximum index value is identified. These two genotypes are considered as the nuclei of two clusters.

(b) Allocate each of the genotype to these clusters in such a way that the index value with the nucleus genotype is minimum.

(c) To increase the number of clusters, by one, identify the two genotypes having maximum index value within the clusters and these genotypes are taken as the nuclei of two clusters in addition to the nuclei of the existing clusters except the clusters containing the newly identified genotypes. Repeat (b).

(d) Repeat (c) until the desired number of clusters is arrived at.

These clusters are further optimised as follows.

(1) Number the genotypes from 1 to t.

(2) Take out genotype 1 from the cluster it belongs to. Allocate it to each of the clusters and calculate the weighted arithmetic mean of the dissimilarity indices of clusters, weights being the multiplier used to obtain the Chi-square value (3.1). This genotype is fused with the cluster for which the weighted arithmetic mean of the dissimilarity indices is minimum.

(3) Repeat the process with all genotypes.

(4) Repeat steps (1) to (3) until two successive iteractions give identical clusters.

OPTIMUM NUMBER OF CLUSTERS

A method for obtaining the optimum number of clusters into which the genotypes are to be clustered is given herein.

A graph of weighted arithmetic mean of the dissimilarity indices of the clusters, the weights being the multiplier used to obtain the Chi-square value against the number of clusters is drawn with number of clusters on X axis. The point just beyond this one on the X axis at which the graph takes maximum curvature can be taken as the optimum number of clusters.

The methodology developed have been illustrated using two sets of data.

References

REFERENCES

Balakrishnan, S., Prabhakaran, P.V., Nambiar, I.P.S. and Nair, K.K.R. 1978. Seasonal behaviour of plant growth regulators in inducing flowering in pineapple. *Agric. Res. J. Kerala. 16:* 138-141.

Devadas, V.S. 1982. Screening for non-bolting type(s) of amaranthus suited for year round planting. M.Sc. thesis submitted to Kerala Agricultural University, Vellanikkara.

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

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

Fripp, Yvonne, J. 1972. Genotype - environmental interactions in *Schizophyllum commune*. II. Assessing the environment. *Heredity. 28:*223-238.

Fripp, Yvonne, J. and Caten, C.E. 1971. Genotype - environmental interaction in *Schizophyllum commune*. I. Analysis and character. *Heredity. 27:* 393-407.

Hardwick, R.C. and Wood, J.T. 1972. Regression methods for studying genotype - environment interactions. *Heredity. 28:* 209-222.

Ibrahim, K.K., Sukumara Pillai, V. and Sasikumaran, S. 1985. Genotype x season interaction and stability parameters in black pepper (*Piper nigrum* L.). *Agric. Res. J. Kerala. 23:*154-162.

John, Laly, C. 1982. Comparison of different techniques for the estimation of genotype - environment interaction. M.Sc. thesis submitted to Kerala Agricultural University, Vellanikkara.

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. and Thompson, B. 1975. An emprical method of grouping genotypes based on a linear function of the genotype-environment interaction. *Heredity.* *34:* 255-263.

Mahalanobis, P.C. 1936. On the generalised distance in statistics. *Proc. Natl. Inst. Sci. India.* *2*(1):49-55.

Mini, C.J. 1989. A comparative study of genotype-environment interaction in sesame. M.Sc. thesis submitted to Kerala Agricultural University, Vellanikkara.

Perkins, J.M. and Jinks, J.L. 1968a. Environmental and genotype-environmental components of variability. III. Multiple lines and crosses. *Heredity.* *23:* 339-356.

Perkins, J.M. and Jinks, J.L. 1968b. Environmental and genotype-environmental components of variability. IV. Non-linear interactions for multiple inbred lines. *Heredity.* *23:*525-535.

Ramey, T.B. and Rosielle, A.A. 1983. HASS cluster analysis: a new method of grouping genotypes or environments in plant breeding. *Theor. Appl. Genet.* *66:*131-133.

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

*Sokal, R.R. and Michener, C.D. 1958. A statistical model for evaluating systematic relationships. *Univ. Kansas. Sci. Bull.* *38*:1409-1438.

Sreekala, M.N. 1989. Clustering genotypes based on genotype environment interaction. M.Sc. thesis submitted to Kerala Agricultural University, Vellanikkara.

Sulochana, K.A. 1984. Screening for resistance to *Aphis craccivora* in vegetable cowpea (*Vigna anguiculata* var. *sesquipedalis* and var. *cylindrica*) and evaluation of early types. M.Sc. thesis submitted to Kerala Agricultural University, Vellanikkara.

Suresh, K.M. 1986. Standardisation of techniques of clustering genotypes using Mahalanobis D^2 and Wilks 'Λ' criterion. M.Sc. thesis submitted to Kerala Agricultural University, Vellanikkara.

Suresh Babu, K.V. 1981. Phenotypic stability analysis in bhindi (*Abelmoschus esculentus* L.). M.Sc. thesis submitted to Kerala Agricultural University, Vellanikkara.

Tai, C.C. 1971. Genotypic stability analysis and its application to potato regional trials. *Crop Sci.* *11*: 184-190.

Ushamani, P. 1987. Phenotypic stability analysis in bacterial wilt resistant lines of brinjal (*Solanum melongena* L.). M.Sc. thesis submitted to Kerala Agricultural University, Vellanikkara.

Vahab, M.A. 1989. Homeostatic analysis of components of genetic variance and inheritance of fruit colour, fruit shape and bitterness in bittergourd (*Momordica charantia* L.). Ph.D. thesis submitted to Kerala Agricultural University, Vellanikkara.

Varghese, G. 1992. Evaluation of F_1 hybrids resistant to bacterial wilt and inheritance of resistance in brinjal (*Solanum melongena* L.). M.Sc.(Hort.) thesis, Kerala Agricultural University, Vellanikkara.

Wood, J.T. 1976. The use of environmental variables in the interpretation of genotype-environment interaction. *Heredity*. *37*(1): 1-7.

*Wricke, G. 1966. Uber eine Biometrics che methods zur Erfassung der okologischen Anpassung. *Acta. Agri. Scand. Suppl. 16* : 98-101.

*Yates, F. and Cochran, W.G. 1938. The analysis of groups of experiments. *J. Agric. Sci. Camb. 28*:556-580.

* Originals not seen

Appendices

APPENDIX I

Data set I Mean yield per plant (kg) of 55 genotypes of bittergourd over 3 seasons

Genotypes	Seasons		
	S1	S2	S3
1	10.190	9.125	9.625
2	9.550	8.250	9.050
3	10.450	8.700	9.250
4	3.190	3.075	3.375
5	2.500	2.375	2.050
6	3.500	2.975	3.075
7	10.175	8.625	10.250
8	8.650	7.625	8.500
9	8.850	7.250	7.775
10	7.450	6.750	7.025
11	10.400	8.625	9.905
12	10.500	8.750	7.775
13	8.400	4.750	5.265
14	6.150	5.500	4.425
15	8.950	4.125	5.200
16	10.400	8.750	10.115
17	9.800	9.875	8.735
18	10.100	7.575	7.930
19	10.100	7.500	5.945
20	10.600	8.450	8.055
21	7.400	4.625	5.245
22	7.400	3.625	4.050
23	9.800	3.505	4.100
24	10.950	8.775	11.275
25	10.100	8.375	8.485
26	9.750	7.750	8.155
27	9.550	7.500	6.075
28	8.150	4.425	4.455
29	7.100	3.825	2.715
30	8.200	4.700	3.515
31	10.000	8.125	8.190
32	10.500	9.250	7.275
33	9.750	8.000	7.255
34	9.050	8.125	6.925
35	3.300	2.625	2.400
36	7.600	4.400	3.200
37	6.050	4.250	4.735
38	9.700	3.875	5.265
39	6.150	4.675	5.405

Contd.

Appendix-I. Continued

1	2	3	4
40	6.150	4.625	6.305
41	3.050	3.375	3.125
42	6.615	4.550	3.125
43	4.750	3.800	3.300
44	7.050	3.825	1.840
45	6.550	3.875	1.915
46	8.100	4.375	4.645
47	5.600	4.800	3.550
48	8.000	4.400	4.525
49	7.000	3.375	4.350
50	9.750	8.375	9.765
51	9.850	7.375	8.775
52	9.250	7.150	6.965
53	9.300	7.250	7.900
54	10.150	7.200	9.030
55	8.150	7.375	5.845

APPENDIX II

Data set II Mean yield per plant(gm) of 9 genotypes of brinjal over 4 seasons

Genotypes	Seasons			
	S1	S2	S3	S4
1	1566.040	1448.130	957.920	565.000
2	476.680	27.270	33.020	455.000
3	1412.920	1066.040	687.920	527.920
4	938.290	606.880	609.170	312.080
5	1101.920	940.000	655.420	583.750
6	848.420	628.340	357.500	831.670
7	643.540	105.480	71.170	1251.250
8	1763.340	1359.580	1361.250	706.670
9	1671.670	1579.380	917.500	1136.250

APPENDIX-III

Pairwise dissimilarity indices for 55 genotypes

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
1	0.00																											
2	0.91	0.00																										
3	9.14	10.45	0.00																									
4	17.76	5.02	[illegible]	0.00																								
5	7.59	[illegible]	14.03	10.59	0.00																							
6	3.40	[illegible]	[illegible]	[illegible]	[illegible]	0.00																						
7	17.80	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.00																					
8	4.74	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.00																				
9	5.84	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.00																			
10	3.14	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.00																		
11	6.06	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	10.96	0.00																	
12	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	53.66	[illegible]	15.26	[illegible]	0.00																
13	146.47	150.00	61.57	246.79	182.76	[illegible]	[illegible]	[illegible]	97.84	[illegible]	[illegible]	[illegible]	0.00															
14	44.49	55.82	28.27	84.62	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.00														
15	238.96	239.96	154.82	367.26	286.95	[illegible]	[illegible]	[illegible]	[illegible]	267.28	[illegible]	39.02	[illegible]	[illegible]	0.00													
16	8.01	3.59	20.94	19.59	18.41	15.92	[illegible]	[illegible]	[illegible]	[illegible]	0.94	[illegible]	[illegible]	[illegible]	260.69	0.00												
17	26.40	[illegible]	[illegible]	[illegible]	[illegible]	20.00	[illegible]	49.44	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	245.32	[illegible]	0.00											
18	55.56	59.35	20.29	[illegible]	77.26	70.75	[illegible]	86.56	[illegible]	[illegible]	61.70	20.67	21.90	39.30	67.42	[illegible]	39.95	0.00										
19	[illegible]	[illegible]	201.92	404.46	292.32	299.44	402.20	357.96	216.40	301.84	317.57	42.79	54.66	126.15	68.68	352.11	205.51	122.25	0.00									
20	84.48	92.74	41.68	158.67	96.35	95.77	162.49	125.48	48.50	95.66	101.50	3.57	19.46	12.99	65.24	121.30	60.81	7.24	[illegible]	0.00								
21	57.26	59.20	20.64	126.77	[illegible]	74.43	107.14	86.62	[illegible]	71.06	58.96	26.00	[illegible]	[illegible]	62.90	74.83	69.75	[illegible]	111.41	11.64	0.00							
22	170.54	174.91	120.90	279.48	225.01	196.20	250.24	[illegible]	117.24	192.28	173.47	[illegible]	[illegible]	[illegible]	9.46	197.91	162.25	[illegible]	45.30	[illegible]	[illegible]	0.00						
23	586.57	592.76	449.59	778.09	647.86	[illegible]	[illegible]	673.52	475.59	[illegible]	587.39	[illegible]	[illegible]	415.29	80.95	278.47	554.75	191.55	120.42	249.17	[illegible]	[illegible]	0.00					
24	48.20	36.53	71.98	45.56	50.51	61.16	7.98	17.64	[illegible]	53.59	14.28	[illegible]	[illegible]	61.20	350.94	17.40	[illegible]	153.75	505.62	219.26	[illegible]	[illegible]	771.62	0.00				
25	[illegible]	27.84	4.73	72.07	35.20	31.47	[illegible]	46.67	6.92	30.46	14.08	32.24	56.57	[illegible]	122.85	45.45	26.94	8.25	146.20	18.98	[illegible]	70.94	387.74	114.60	0.00			
26	33.22	25.57	[illegible]	[illegible]	79.51	33.42	84.94	44.45	3.76	31.39	28.94	38.62	57.65	11.64	115.23	42.00	35.75	[illegible]	154.75	22.05	3.14	66.41	378.13	121.90	0.87	0.00		
27	191.25	207.52	126.20	289.59	195.22	200.19	312.78	157.70	[illegible]	204.81	204.00	12.57	37.47	65.41	70.70	251.52	125.33	56.54	9.85	24.75	65.62	34.20	178.59	769.65	80.67	[illegible]	0.00	
28	212.42	217.44	133.87	227.81	241.33	[illegible]	321.21	[illegible]	146.77	211.58	213.96	37.28	7.75	11.54	11.66	146.66	185.66	49.61	26.26	74.07	52.67	[illegible]	99.20	36.47	94.66	94.79	14.84	0.00
29	[illegible]	[illegible]	221.46	441.44	330.74	[illegible]	[illegible]	157.56	237.64	374.24	338.46	59.56	[illegible]	156.12	49.97	374.80	244.65	111.42	3.80	11.35	[illegible]	35.51	[illegible]	[illegible]	165.06	170.79	[illegible]	[illegible]
30	[illegible]	[illegible]	261.86	524.57	332.98	[illegible]	[illegible]	443.35	281.85	356.65	389.77	80.05	[illegible]	91.17	59.46	428.75	289.19	[illegible]	[illegible]	47.74	[illegible]	[illegible]	[illegible]	555.67	180.24	[illegible]	[illegible]	[illegible]
31	33.04	17.81	5.90	65.39	45.69	[illegible]	[illegible]	55.75	[illegible]	41.19	[illegible]	24.21	44.51	14.16	105.57	57.09	32.64	[illegible]	126.15	[illegible]	6.96	[illegible]	350.26	[illegible]	0.81	[illegible]	69.48	[illegible]
32	[illegible]	158.77	126.06	234.54	171.15	[illegible]	[illegible]	242.55	[illegible]	[illegible]	[illegible]	10.87	62.85	[illegible]	[illegible]	247.88	[illegible]	[illegible]	[illegible]	29.59	[illegible]	[illegible]	241.82	390.56	[illegible]	[illegible]	[illegible]	[illegible]
33	[illegible]	[illegible]	44.40	151.57	69.28	[illegible]	267.84	76.15	50.51	90.00	106.70	1.40	72.16	[illegible]	[illegible]	126.24	50.26	11.69	54.75	1.40	[illegible]	[illegible]	271.63	[illegible]	20.30	25.40	[illegible]	[illegible]
34	67.57	82.34	42.59	120.67	61.27	66.22	[illegible]	106.33	[illegible]	71.47	98.17	9.30	60.30	3.41	133.98	114.68	24.53	15.47	88.24	12.52	74.54	72.39	154.58	320.27	[illegible]	27.73	39.26	[illegible]
35	6.46	11.94	9.55	25.51	4.86	[illegible]	45.60	30.01	7.27	6.36	22.77	[illegible]	[illegible]	19.34	219.42	26.63	8.34	40.59	226.96	58.54	46.55	146.78	540.59	89.94	[illegible]	17.41	[illegible]	[illegible]
36	316.52	331.46	224.88	447.59	332.71	336.41	450.05	391.68	241.33	337.27	340.26	62.29	52.87	156.60	53.95	379.78	245.10	114.35	3.15	73.96	121.38	41.48	84.62	531.19	[illegible]	73.90	[illegible]	21.44
37	11.79	14.20	0.29	51.80	37.57	[illegible]	46.77	[illegible]	1.33	19.11	17.51	[illegible]	[illegible]	25.01	140.46	26.07	31.07	15.70	186.90	35.01	16.49	91.00	423.74	81.72	2.77	[illegible]	118.31	[illegible]
38	360.36	362.30	255.23	514.94	421.82	[illegible]	[illegible]	411.46	275.25	395.58	346.41	146.65	[illegible]	265.25	12.70	381.95	368.47	138.65	96.20	[illegible]	[illegible]	41.97	35.95	462.81	214.55	[illegible]	118.36	[illegible]
39	1.51	1.28	4.40	33.45	15.28	7.57	16.77	7.54	2.44	5.27	3.69	96.64	120.88	44.33	207.36	7.10	34.79	43.22	264.81	72.70	40.22	146.46	539.80	46.38	17.50	15.44	[illegible]	185.40
40	20.79	14.25	44.59	19.97	40.36	27.75	0.15	7.26	38.78	22.96	9.45	218.73	232.77	125.67	327.29	4.48	94.50	119.41	436.23	172.39	115.12	261.98	733.22	6.75	78.21	71.26	[illegible]	318.34
41	17.74	22.31	48.65	3.80	[illegible]	7.83	36.29	15.70	40.81	9.95	36.57	166.86	248.86	69.89	372.34	33.27	32.89	123.48	380.77	149.93	129.51	276.24	774.94	75.35	68.21	71.71	266.36	221.80
42	194.85	209.70	129.36	291.87	197.13	204.05	312.47	255.71	140.25	206.70	225.74	12.99	37.53	66.55	69.90	254.37	126.91	57.25	9.45	24.91	66.32	34.22	176.94	391.83	84.97	93.36	0.00	24.57
43	23.38	70.05	11.20	57.98	[illegible]	23.30	78.75	46.26	11.66	24.72	42.75	34.70	80.72	4.14	164.38	52.98	8.98	22.70	156.45	27.35	26.67	95.26	440.10	130.86	5.05	9.70	88.43	121.60
44	472.50	491.88	361.64	624.69	482.90	492.22	603.95	564.87	381.60	495.07	510.30	132.26	128.86	258.51	117.17	554.33	366.85	218.65	23.84	158.03	229.15	109.78	81.59	707.66	286.46	296.69	64.08	74.50
45	369.42	389.03	274.78	502.04	374.50	384.40	522.98	452.38	291.50	387.78	407.72	79.55	92.68	162.19	96.39	446.68	271.64	155.77	5.63	102.46	166.67	76.97	108.76	617.89	208.62	219.41	29.04	45.27
46	181.50	136.98	189.62	292.72	214.27	206.65	266.27	233.68	122.46	203.18	166.66	33.27	2.42	98.84	10.04	214.13	167.13	35.98	37.60	26.52	35.96	0.39	117.04	317.34	76.96	75.29	29.47	1.53
47	64.92	77.59	40.30	114.70	56.67	64.17	148.81	102.43	43.23	67.67	96.11	11.98	69.67	2.32	143.41	111.99	20.96	26.06	95.23	15.50	37.52	79.39	370.35	216.86	19.75	28.36	43.79	89.41
48	182.42	208.67	110.68	293.08	212.83	206.43	270.45	225.69	123.45	203.47	169.83	29.87	3.74	95.22	12.55	217.55	163.33	36.46	32.55	24.83	37.20	1.25	118.14	324.03	76.77	75.92	24.82	1.04
49	104.50	104.91	52.10	194.09	141.70	129.15	158.19	139.98	61.25	123.98	99.86	38.51	6.06	73.66	27.55	119.90	121.83	10.85	93.48	20.48	7.78	11.67	200.53	192.86	36.79	31.59	62.70	27.58
50	12.54	7.65	33.87	14.01	30.39	18.07	0.57	2.80	28.40	14.19	5.55	185.25	211.13	104.14	307.79	1.95	75.58	103.25	401.17	150.70	100.40	240.92	702.53	12.79	63.28	57.93	292.79	293.70
51	16.86	13.10	7.17	55.92	45.81	31.85	28.26	24.69	7.90	26.79	8.01	95.10	91.15	62.98	154.33	14.00	66.76	30.96	245.61	65.34	27.07	111.16	456.06	46.65	19.47	13.12	169.41	149.92
52	63.31	70.06	26.88	138.35	76.29	74.27	131.62	99.00	32.49	73.67	77.43	8.65	24.61	16.10	74.94	94.85	49.33	2.93	81.50	1.63	6.44	32.84	280.41	183.38	9.66	11.70	38.41	46.31
53	16.42	17.36	1.11	60.20	35.33	27.37	49.39	33.44	2.74	24.68	18.83	153.27	65.44	29.71	130.11	20.04	39.15	12.67	161.95	33.15	12.49	82.09	408.01	82.15	3.34	1.18	114.06	112.08
54	29.49	23.28	16.44	73.80	66.51	46.78	33.52	35.78	17.98	42.11	13.83	111.37	91.92	84.73	147.17	19.99	92.11	37.62	260.86	77.13	31.35	111.77	446.39	45.08	30.86	21.92	186.38	152.57
55	88.47	103.06	57.97	144.63	77.82	87.38	[illegible]	131.57	61.99	91.62	123.67	8.48	68.57	8.01	138.76	141.93	32.72	35.57	75.51	16.20	46.29	75.84	344.92	257.49	31.55	41.61	31.16	80.86

	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55
29	0.00																										
30	1.92	0.00																									
31	140.99	177.67	0.00																								
32	43.51	60.53	70.05	0.00																							
33	76.34	102.43	10.85	20.05	0.00																						
34	114.39	145.34	16.34	27.90	[illegible]	0.00																					
35	258.57	305.00	20.75	126.99	54.16	35.90	0.00																				
36	0.05	1.83	145.47	42.21	77.31	14.59	60.65	0.00																			
37	225.69	246.62	6.00	115.18	77.75	35.53	10.70	209.95	0.00																		
38	65.33	65.99	191.47	173.64	155.10	223.14	[illegible]	69.92	240.85	0.00																	
39	288.30	336.49	25.45	171.67	74.85	64.71	9.26	292.09	6.96	320.52	0.00																
40	461.04	520.56	93.46	317.05	177.30	161.50	50.10	466.63	52.09	459.38	22.14	0.00															
41	425.14	484.10	83.24	237.82	141.17	183.99	11.02	427.31	54.46	522.31	29.60	37.67	0.00														
42	20.44	33.93	70.45	5.55	22.65	39.94	144.30	30.00	117.53	117.47	179.43	326.02	270.62	0.00													
43	183.87	220.32	7.38	76.58	20.27	12.33	6.87	185.40	9.63	267.86	21.44	85.19	49.86	89.70	0.00												
44	18.66	10.20	257.69	87.12	153.97	302.61	397.95	17.09	341.41	115.26	445.08	658.40	594.72	62.98	302.43	0.00											
45	7.57	6.49	184.95	44.15	201.12	134.06	32.33	6.12	257.26	212.74	347.22	540.62	470.25	29.32	218.83	7.24	0.00										
46	31.14	44.65	62.36	56.29	36.36	73.15	154.62	33.79	99.13	40.77	157.34	278.66	289.66	29.33	104.35	97.05	66.45	0.00									
47	122.97	154.80	17.24	20.91	8.36	[illegible]	33.29	120.00	75.42	235.51	62.74	157.69	97.61	44.72	11.09	212.79	142.19	60.32	0.00								
48	27.37	40.66	62.01	50.05	34.17	69.16	151.46	29.75	99.99	44.18	158.93	282.99	288.44	24.67	102.39	90.74	60.22	8.21	76.18	0.00							
49	59.92	113.60	28.74	87.03	32.67	61.76	95.22	93.90	45.85	76.75	83.76	167.52	199.85	63.04	64.92	190.29	141.36	16.09	67.06	18.34	0.00						
50	427.45	485.24	77.42	284.19	154.53	137.53	15.63	432.51	40.43	438.45	14.29	1.06	29.43	294.86	67.85	616.70	501.53	256.33	133.90	259.82	152.42	0.00					
51	257.98	301.73	24.50	176.59	73.95	77.89	29.54	262.92	6.61	249.76	6.37	32.25	68.72	170.76	36.13	413.24	325.69	122.53	78.11	126.83	52.79	26.37	0.00				
52	94.37	123.52	4.93	42.59	1.51	12.17	41.82	96.85	21.58	148.19	52.73	140.54	124.28	39.10	18.03	191.42	129.88	35.63	14.43	34.55	19.37	121.09	46.89	0.00			
53	197.87	237.60	5.82	116.26	27.51	38.96	16.05	201.44	0.52	223.02	9.25	54.92	64.46	115.25	13.36	333.10	251.59	90.33	39.24	91.67	30.11	43.73	6.69	20.12	0.00		
54	267.91	311.02	35.43	199.72	89.16	95.95	[illegible]	273.60	17.95	236.60	18.07	37.37	91.13	187.65	53.86	427.41	342.26	124.08	99.97	125.82	51.56	33.50	2.07	57.94	14.14	0.00	
55	102.93	131.55	26.97	18.34	8.05	[illegible]	[illegible]	102.43	51.70	225.63	85.55	190.19	124.17	31.95	21.80	181.09	116.12	75.10	1.83	70.15	1.58	166.72	100.59	18.52	55.41	123.65	0.00

```
ENDIX IV

y indices for 9 genotypes



    5        6        7        8        9




  0.00
 19.62    0.00
119.66   42.86    0.00
 16.07   68.64  213.44    0.00
  4.57   15.54  106.87   33.27    0.00
```

APPE

Pairwise dissimilarit

```
          1          2          3          4
1      0.00
2     73.99       0.00
3      5.27      40.20       0.00
4     12.09      33.28       3.86       0.00
5     12.07      26.39       1.72       2.71
6     61.17       2.14      31.20      29.55
7    206.14      35.28     146.31     137.05
8      6.09      75.24      10.55       8.49
9     19.37      26.31       6.48      14.22
```

APPENDIX-V

Programme to group genotypes by statistical clustering

```
 10 REM N - error degrees of freedom
 20 REM NE - number of environments
 30 REM NG - number of genotypes
 40 REM C1 - name of the file containing Chi-square values
 50 REM C2 - name of the file containing the dissimilarity indices
 60 DIM D(60,60),CHI(60),IC(60),IN(55)
 70 KL=0
 80 INPUT N,NE,NG
 90 OPEN "i",#1,"C1"
100 FOR ID= 1 TO 30
110 INPUT #1,CHI(ID)
120 NEXT ID
130 NG1= NG-1
140 OPEN "i",#2,"C2"
150 FOR I=1 TO NG
160 IN(I)=0
170 FOR J= 1 TO I
180 INPUT #2,D(I,J)
190 D(J,I)= D(I,J)
200 NEXT J
210 NEXT I
220 FOR I= 1 TO NG
230 IF IN(I)<>0 THEN 250
240 GOTO 280
250 NEXT I
260 LPRINT "clustering over"
270 STOP
280 IA=I
290 KA=I+1
300 IF I=NG THEN KA=I-1
310 SA=D(I,KA)
320 FOR I= 1 TO NG
330 IF IN(I)<>0 THEN 410
340 FOR J= 1 TO NG
350 IF I=J THEN 400
360 IF SA<D(J,I) THEN 400
370 SA =D(J,I)
380 IA=I
390 KA=J
400 NEXT J
410 NEXT I
420 X=((N-4)*(N-2)*SA)/(N*(N-1))
430 DF=((NE-1)*(N-4))/(N-1)
440 ID=DF
450 IF (DF-ID)>=.5 THEN ID=ID+1
460 IF X<=CHI(ID) THEN 490
470 LPRINT "no further cluster can be formed"
480 STOP
490 NC=2
500 IC(1)=IA
510 IC(2)=KA
520 KZ=0
530 FOR I=1 TO NG
```

```
540 FOR J= 1 TO NC
550 IF I= IC(J) THEN 710
560 NEXT J
570 S=0
580 FOR K= 1 TO NC
590 K1= IC(K)
600 FOR K2=K+1 TO NC
610 J1=IC(K2)
620 S= S+D(I,K1)+D(I,J1)+D(K1,J1)
630 NEXT K2
640 NEXT K
650 KZ= KZ+1
660 S= S*2/(NC+1)
670 IF KZ=1 THEN SA=S
680 IF S>SA THEN 710
6[illegible] SA=S
700 IA=I
710 NEXT I
720 X1=((N-4)*(N-2)*SA)/(N*(N+NC-2))
730 DF=(NC*(NE-1)*(N-4))/(N+NC-2)
740 ID=DF
750 IF (DF-ID)>=.5 THEN ID=ID+1
760 IF X1>=CHI(ID) THEN 800
770 NC= NC+1
780 IC(NC)=IA
790 GOTO 520
800 KL=KL+1
810 LPRINT "cluster no.",KL
820 LPRINT "no.of genotypes",NC
830 FOR I= 1 TO NC
840 LPRINT IC(I),
850 NEXT I
860 FOR KK=1 TO NC
870 I=IC(KK)
880 IN(I)=9
890 NEXT KK
900 GOTO 220
910 CLOSE #1,#2
920 END
```

APPENDIX-VI

Programme to group genotypes by minimising GE interaction using IRA

```
10 REM CC - name of the file containing the dissimilarity indices
20 DIM D(60,60),KN(60),KK(20),KS(20,60),Y(60,60),MA(20,60),
   MB(20,60),MO(20),M1(20)
30 INPUT "no. of genotypes",NG
40 INPUT "max.no. of clusters",KZ
50 OPEN "I",#1,"CC"
60 INPUT "error degrees of freedom",NE
70 FOR I= 1 TO NG
80 LPRINT
90 FOR J= 1 TO I
100 INPUT #1, D(I,J)
110 D(J,I) =D(I,J)
120 NEXT J
130 NEXT I
140 KK(1)= NG
150 FOR   I=1 TO NG
160 KS (1,I)=I
170 NEXT I
180 K=1
190 S=0
200 FOR I=1 TO K
210 IF KK(I)=1 THEN 360
220 KL=KK(I) -1
230 KL1= KK(I)
240 FOR J= 1 TO KL
250 J1= J+1
260 FOR JJ= J1 TO KL1
270 K1=KS(I,J)
280 K2=KS(I,JJ)
290 IF S>D(K1,K2) THEN 340
300 S=D(K1,K2)
310 KM=K1
320 KO= K2
330 KI=I
340 NEXT JJ
350 NEXT J
360 NEXT I
370 K=K+1
380 KS(KI,1)=KM
390 KS(K,1)= KO
400 FOR I= 1 TO K
410 KI= KS(I,1)
420 KN(KI) =I
430 KK(I) =1
440 NEXT I
450 FOR I=1 TO NG
460 FOR L=1 TO K
470 IF I= KS(L,1) THEN 620
480 NEXT L
```

```
490 L1 =KS(1,1)
500 S= D(I,L1)
510 LK=1
520 FOR L=2 TO K
530 L1=KS(L,1)
540 IF S<D(I,L1) THEN 570
550 S=D(I,L1)
560 LK=L
570 NEXT L
580 KK(LK)=KK(LK)+1
590 KM=KK(LK)
600 KS(LK,KM) =1
610 KN(I)=LK
620 NEXT I
630 LPRINT "no.of clusters",K
640 GOSUB 680
650 IF K<>KZ THEN 190
660 STOP
670 END
680 REM
690 FOR I=1 TO K
700 MO(I)=0
710 NEXT I
720 FOR I=1 TO NG
730 MI=KN(I)
740 MO(MI) =MO(MI)+1
750 ML= MO(MI)
760 MA(MI,ML) =I
770 NEXT I
780 FOR I=1 TO K
790 M1(I) =MO(I)
800 MI=M1(I)
810 FOR J= 1 TO MI
820 MB(I,J)=MA(I,J)
830 NEXT J
840 NEXT I
850 GOSUB 1730
860 FOR I= 1 TO K
870 LPRINT
880 MI=M1(I)
890 FOR J= 1 TO MI
900 LPRINT MB(I,J);
910 NEXT J
920 NEXT I
930 LPRINT
940 FOR I= 1 TO K
950 LPRINT
960 FOR J= 1 TO I
970 LPRINT Y(I,J);
980 NEXT J
990 NEXT I
1000 LPRINT
```

```
1010 LPRINT "average intra d",XL
1020 KK=1
1030 FOR IL= 1 TO NG
1040 LX= KN(IL)
1050 IF M1(LX)<=1 THEN 1330
1060 MI=KN(IL)
1070 MK=M1(MI)
1080 M1(MI)=M1(MI)-1
1090 MK1= MK-1
1100 FOR J= 1 TO MK1
1110 IF MB(MI,J) = IL THEN 1140
1120 NEXT J
1130 GOTO 1180
1140 FOR KB=J TO MK1
1150 MB(MI,KB)=MB(MI,KB+1)
1160 NEXT KB
1170 DA=XL
1180 FOR L= 1 TO K
1190 M1(L)=M1(L)+1
1200 MI=M1(L)
1210 MB(L,MI)=IL
1220 GOSUB 1730
1230 M1(L)=M1(L)-1
1240 DB=XL
1250 IF DB>=DA THEN 1290
1260 LX=L
1270 KN(IL)=L
1280 DA=DB
1290 NEXT L
1300 M1(LX)=M1(LX)+1
1310 MI=M1(LX)
1320 MB(LX,MI)=IL
1330 NEXT IL
1340 FOR I=1 TO K
1350 IF M1(I)<>MO(I) THEN 1440
1360 NEXT I
1370 FOR I=1 TO K
1380 MI=M1(I)
1390 FOR J= 1 TO MI
1400 IF MB(I,J)<>MA (I,J) THEN 1440
1410 NEXT J
1420 NEXT I
1430 GOTO 1540
1440 KK=KK+1
1450 FOR I=1 TO K
1460 MO(I)=M1(I)
1470 MI=M1(I)
1480 FOR J= 1 TO MI
1490 MA (I,J) =MB(I,J)
1500 NEXT J
1510 NEXT I
1520 GOSUB 1730
```

```
1530 GOTO 1030
1540 LPRINT "no.of iteration",KK
1550 GOSUB 1730
1560 FOR I= 1 TO K
1570 LPRINT
1580 MI= M1(I)
1590 FOR J= 1 TO MI
1600 LPRINT MB(I,J);
1610 NEXT J
1620 NEXT I
1630 LPRINT
1640 FOR I= 1 TO K
1650 LPRINT
1660 FOR J= 1 TO I
1670 LPRINT Y(I,J);
1680 NEXT J
1690 NEXT I
1700 LPRINT
1710 LPRINT "average intra d",XL
1720 RETURN
1730 REM sub bet (x,m,mb,y,nk)
1740 FOR I=1 TO K
1750 FOR J=I TO K
1760 Y(I,J)=0
1770 NEXT J
1780 NEXT I
1790 NK1=K-1
1800 FOR I= 1 TO NK1
1810 MI=M1(I)
1820 IF M1(I)<=1 THEN 1920
1830 MI1=MI-1
1840 FOR KB=1 TO MI1
1850 KO=MB(I,KB)
1860 K1=KB+1
1870 FOR K2=K1 TO MI
1880 K3=MB(I,K2)
1890 Y(I,I)=Y(I,I)+D(KO,K3)
1900 NEXT K2
1910 NEXT KB
1920 I1= I+1
1930 FOR J=I1 TO K
1940 ML=M1(I)
1950 MM=M1(J)
1960 FOR II=1 TO ML
1970 FOR JJ= 1 TO MM
1980 K1=MB(I,II)
1990 K2=MB(J,JJ)
2000 Y(I,J)=Y(I,J)+D(K1,K2)
2010 NEXT JJ
2020 NEXT II
2030 NEXT J
2040 NEXT I
```

```
2050 IF M1(K)<=1 THEN 2160
2060 MI=M1(K)
2070 MI1=MI-1
2080 FOR I=1 TO MI1
2090 KB=MB(K,I)
2100 I1=I+1
2110 FOR J= I1 TO MI
2120 K1=MB(K,J)
2130 Y(K,K)=Y(K,K)+D(KB,K1)
2140 NEXT J
2150 NEXT I
2160 LN=0
2170 XL=0
2180 FOR I=1 TO K
2190 FOR J= I TO K
2200 IF I<>J THEN 2280
2210 KA=(NE-4)*(NE-2)/(NE*(NE+M1(I)-3))
2220 XL=XL+Y(I,J)*KA*2/M1(I)
2230 IF M1(I)=1 THEN KA=0
2240 LN=LN+KA
2250 IF M1(I)=1 THEN 2300
2260 Y(I,J)=Y(I,J)/(M1(I)*.5)
2270 GOTO 2290
2280 Y(I,J)=Y(I,J)/(M1(I)*M1(J))
2290 Y(J,I)=Y(I,J)
2300 NEXT J
2310 NEXT I
2320 XL= XL/LN
2330 RETURN
2340 END
```

CLUSTERING GENOTYPES BASED ON G × E INTERACTION IN THE ABSENCE OF HOMOGENEITY OF ERROR VARIANCES

By

CHANDRIKA. C.

ABSTRACT OF A THESIS

Submitted in partial fulfilment of the
requirement for the degree

Master of Science in Agricultural Statistics

Faculty of Agriculture
Kerala Agricultural University

Department of Agricultural Statistics
COLLEGE OF HORTICULTURE
Vellanikkara, Thrissur

1992

ABSTRACT

A distance function between every pair of genotypes which measures the genotype-environment interaction in the case of heterogeneity of error variances in different environments was derived herein. A dissimilarity index for any set of genotypes which also measures the within group genotype-environment interaction sum of squares in terms of the pairwise distance function was also derived. Two methods of clustering, viz., statistical clustering and clustering by minimisation of average within cluster genotype-environment interaction making use of the proposed dissimilarity index were also proposed.

Statistical clustering helps to group the genotypes such that genotype-environment interaction within any group is insignificant while any addition to the cluster makes it significant. In other words the idea of statistical clustering is to identify genotypes having similar response to varying environments. A point to be noted is that all the groups formed by statistical clustering may not be non-overlapping.

The other procedure helps to form optimum clustering by minimising the average within cluster genotype-environment interaction using an iterative relocation algorithm.

These clustering procedures were illustated making use of two sets of data.