DIVERGENCE ANALYSIS OF MORPHOLOGICAL AND QUALITY TRAITS IN SUGARCANE

BY

SANTHI, T. E.

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 STATISTICS
COLLEGE OF VETERINARY AND ANIMAL SCIENCES
MANNUTHY - TRICHUR

1989

DECLARATION

I hereby declare that this thesis entitled "DIVERGENCE ANALYSIS OF MORPHOLOGICAL AND QUALITY TRAITS IN SUGARCANE" is a bonafide record of research work done by me during the course of research and that the thesis has not 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.

SANTHI, T.E.

Mannuthy.
29-12-89

CERTIFICATE

Certified that this thesis, entitled "DIVERGENCE ANALYSIS OF MORPHOLOGICAL AND QUALITY TRAITS IN SUGARCANE" is a record of research work done independently by Smt. Santhi, T.S. 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.

Dr. (Mrs.) P. SARASWATHY
(Chairman, Advisory Board)
Associate Professor of
Agricultural Statistics
(NARP - SR)
College of Agriculture
Vellayani - Trivandrum

Vellayani,
29-12-89

APPROVED BY:

Chairman:

Dr. (Mrs.) P. Saraswathi

Members:

Dr. K.C. George

substituted for
Sri. K.L. Sunny

Dr. S.G. Sreekumar

External Examiner

ACKNOWLEDGEMENTS

I am deeply indebted to Dr. (Mrs.) P. Saraswathy, Associate Professor of Agricultural Statistics, College of Agriculture, Vellayani, for suggesting the problem, expert guidance, continued support and encouragement rendered during the entire course of research work.

I owe my deep sense of gratitude to the members of my advisory committee Dr. K.C. George, Professor and Head of Department of Statistics, College of Veterinary and Animal Sciences; Sri. K.L. Sunny, Assistant Professor, K.A.U., Vellanikkara for correcting the manuscript and suggesting modifications and Dr. S.G. Sreekumar, Associate Professor of Plant Breeding, College of Agriculture, Vellayani for his consent to utilize his research data for my M.Sc. programme.

The assistance rendered by Sri. C.E. Ajithkumar, Junior Programmer, College of Agriculture, Vellayani in Computer programming and analysis is deeply acknowledged.

My heartful thanks are also due to my parents, husband and friends for their help at all stages of this work.

Finally, I am thankful to Kerala Agricultural University for the fellowship made available to me for my studies.

Santhi, T.E.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF APPENDICES

INTRODUCTION

INTRODUCTION

Sugarcane is an important cash crop of India. It occupies about an area of 9800 hectares in Kerala with a cane production of 461000 tonnes per year. The complex factors governing sugar production are mainly the area under the crop, varieties, yield, recovery percentage, climate, pest-disease attack etc. A knowledge of the nature and magnitude of genetic diversity in morphological and quality traits is important for careful selection of parents for better production. Correlation studies will help to get a knowledge of association among various traits of the crop.

Multivariate statistical methods are useful in plant breeding programmes to explain the influence of various characters on the phenomenon under study. When multi variables are measured from each unit the analysis is collectively made through this method. The multivariate analysis of dispersion is helpful to find the variation among a number of variables taken together.

Mahalanobi's D^2 statistic is a measure of group distance based on multiple characters. The diverse genotypes for hybridization purposes are identified by this method. The genotypes within a cluster are less divergent than those in other clusters. Clusters separated by the largest statistical distance show the maximum divergence.

Factor analysis is a multivariate analysis used to explain the dependence structure of a set of variables interms of certain common factors. The common factors generate covariances among the observable responses. In factor analysis a hypothesis about the covariance-correlation structure helps to identify fundamental and meaningful dimensions of a multivariate domain. The common factors are necessary to account for the inter-correlation among the variables, a unique factor representing that factor of a variable not ascribable to its correlations with other variables in the set. Maximum likelihood method is found to be the most efficient method of extracting factors, though principal factor analysis is commonly used. Maximum likelihood method also provides tests of significance for the determination of the adequate number of common factors. The present study is conducted with the following objectives:

1. To identify the number of factors responsible for genetic divergence in sugarcane by applying factor analysis method.

2. Comparison of the number of factors of divergence obtained by principal factor analysis and factor analysis by maximum likelihood method.

3. To group the different clones of sugarcane by D^2-analysis.

REVIEW OF LITERATURE

REVIEW OF LITERATURE

Multivariate statistical analysis is very useful in biological research to explain the influence of various factors on the phenomenon under study. Genetic variability is of considerable importance in any plant breeding programme for crop improvement. In plant breeding trials, as a large number of variables are involved, effective breeding calls for the knowledge of genetic variability among parents with regard to those characters which are sought to be improved. Genetic divergence among parents is important because a cross involving genetically diverse parents is likely to produce high heterotic effect and also more variability could be expected in the segregating generations. In such situation, factor analytic methods will give an insight into the fewer causal influences responsible for differentiation among genotypes or populations.

2.1 Theoretical studies

2.1.1 Analysis of dispersion

The multivariate analysis of variance or MANOVA began with the derivation of the simultaneous sampling distribution of the variances and covariances in samples from a multivariate normal population (Wishart, 1928).

Hotelling (1931) found the distribution of a random variable T^2 which is the multivariate extension of student's t distribution in a multivariate normal population.

Wilks (1932) extended the test based on T^2 statistic known as Wilk's lambda criterion.

Bartlett (1934) applied it for testing significance of treatments with regard to two variables in a varietal trial and indicated its general use in multivariate tests of significance. Wilks (1935) and Hotelling (1936) found it useful for testing the independence of several groups of variates.

Bartlett (1947) approximated the distribution of lambda statistic to a chi-square.

2.1.2 D^2-Statistics

Divergence analysis is performed to identify the diverse genotypes for hybridization purposes. Clustering by D^2 statistics is useful in this context.

A measure for group distance based on multiple characters was given by Mahalanobis (1928).

Mahalanobis (1936) published a paper on 'generalized distance', which has become the standard measure of distance between two populations, when all the observed characters are quantitative.

Rao (1948) in his classic work, attempted to generalize the D^2 statistic.

Rao (1952) described Tocher's method of forming the clusters.

Everitt (1979) discussed in detail the unresolved problems of cluster analysis.

Many methods for clustering objects into groups were summarised by Everett in 1980.

Arunachalam (1981) made an exposition of the theoretical concepts behind the genetic distance.

Krzanowski (1983) derived a unique measure of distance between populations on the basis of a mixed data - a mixture of quantitative and categorised data.

2.1.3 Factor analysis

The method of factor analysis is widely used as an exploratory tool to reduce the dimensionality of multivariate data. Factor analysis can explain the causative forces responsible for inter and intra-specific differentiation. The method is potent enough to distinguish the forces of natural and human selection causing the divergence in a particular species.

The theory of factor analysis begins from Spearman's two factor theory, which assumes that the inter-relationships of all the variables involved could be accounted for by a single underlying general factor and group factors which are common to some of the variables but not to all of them. In addition to this, a third type factor which are peculiar to single variables alone called specific factors was also differentiated (Spearman, 1904).

Thurstone (1931) generalized Spearman's approach to more than one causal factor.

Roff's suggestion of insertion of SMC (squared multiple correlation) in the principal diagonal of the correlation matrix has been largely advocated by Guttman (1936) because of the property of SMC that it is the lower bound for the communality.

Holzinger and Harman (1941) presented the principal factor solution of factor loadings.

Thurstone (1947) traced the objective of the factor pattern as follows: "The object of a factor problem is to account for their inter correlations, in terms of a small number of derived variables, the smallest possible number that is consistent with acceptable residual errors".

Kendall (1950) made a useful distinction between dependence and interdependence analysis in multivariate analysis. Analysis of dependence is concerned with how a certain specified group depend on other and analysis of interdependence is concerned with how a group of variables are related among themselves. Factor analysis is latter type of multivariate analysis.

Burt (1952) has given a full amount of tests of significance in factor analysis developed upto that time.

The computation schemes of various factor analysis methods were provided by Fruchter (1954).

Rao (1955) introduced the concept of 'basis' of a vector space for the characterization of factor analysis. In the first charactorization due to him a factor variable explains as much of variation as possible of the data which leading to principal factor analysis. In second characterization, he considered the factor variable as the one which is predictable from the original measurements with the maximum possible precision, leading to canonical factor analysis. For this solution the squared canonical correlation between the linear function of hypothetical factor variable and the linear function of measurable variables is maximised.

Factor analysis as a branch of multivariate analysis is very useful in determining the number and nature of causative influences responsible for the inter-correlation of variables in any population. Essentially, it aims at explaining a p x p correlation matrix (p variates) by means of a fewer number k (k p) of meaningful factors (Maxwell, 1961; Lawley and Maxwell, 1963).

In the two subsequent papers, Cattell (1965 a and b) attempted an excellent nonmathematical introduction to factor analysis. He preferred to call the analysis with closed model which accounts for all variances of variables in terms of what is in the particular sample as component analysis and with the open model, which admits, besides the common factors, unexplained specific factors as factor analysis. The uses of factor analysis in modern research as hypothesis creating and testing method were also discussed.

Hemmevle (1965) in his paper considered the problem of computing estimates of factor loadings, specific variances, and communalities for a factor analytic model. Iterative formulae were developed to solve the maximum likelihood equations and a simple efficient method of its implementation on a digital computer was described.

A general description of the concepts, theories and techniques of factor analysis has been given by Harman (1967).

Joreskog (1969) gave the relevant results for confirmatory factor analysis, where the matrix of factor loadings is uniquely identified by priori restrictions (usually by setting particular loadings to zero).

Mc Donald (1970) made a purely theoretical comparison among the three factor score construction methods namely principal factor analysis, canonical factor analysis and alpha factor analysis. According to him, in choosing a factor model, there are in fact, at least three separate choices to be made which are relatively independent. The first is the choice of basis in common factor space and it is the clearest defining characteristic of the three systems discussed. The second is the choice of an iterative algorithm for the determination of communalities/ uniquenesses. The third is the decision rule for the number of common factors.

Joreskog (1971) has given estimation procedures for factor models involving several populations.

Joreskog and Goldberger (1972) have developed a generalized least-squares procedure. The estimates are scale free and asymptotically equivalent to the maximum likelihood estimates when the distribution is multivariate normal.

A non-metric approach to factor analysis has been considered by Kruskal and Shepard (1974). Although this technique has some attractive theoretical properties, it appears to be very sensitive to random variation in the data.

Swain (1975) considered a class of asymptotically efficient estimators including both generalised least square and maximum likelihood as special cases and derived their large-sample properties.

Joreskog (1977) presents a general, all-encompasing series of methods for orthogonal factor analysis by the least squares and maximum likelihood methods. Many variables in the social sciences involve latent and structural variables and Joreskog (1977) developed estimation procedures for several such methods, working directly from the covariance matrix.

A few of the many methods developed for factor extraction are centroid method (Thurstone, 1947), principal factor method (Karl Pearson, 1901), maximum likelihood method (Lawley, 1940) etc. Here we are considering principal factor and maximum-likelihood methods.

2.1.3.1 Principal Factor Method (PF method)

The literature on factor analysis contain a number of alternative methods and procedures for computation. Among these, principal factor method has several attractive features. Each factor extracts maximum amount of variance and gives the smallest possible residuals. However, this method is preferred in the present study mainly owing to computational facilities.

Hotelling (1933 a) developed the principal axes method which provides an optimal solution at the suggestion of Kelley (1935).

Hotelling (1933 b) suggested the use of this method with either unities in the principal diagonal. The resulting factors are called "principal components" and are used to reproduce the score matrix rather than the correlation matrix. The number of principal components extracted is equal to the number of variables in the study.

Hotelling (1935) developed an iterative method of obtaining the loadings which can be carried to any degree of accuracy.

Principal component analysis is sometimes modified by the insertion of communalities in the diagonal of the correlation matrix and Rao (1955) called this method as principal factor analysis.

Harman (1967) exhibited an outline form of the numerical calculations of the method with an illustrative example. The first requirement in applying the principal factor method is to determine some suitable estimates of communality. According to him PF method can be considered as an excellent reduction of the correlation matrix which provides a basis for rotation to some other form of solution. The method also has the advantage of giving a mathematically unique solution for a given correlation matrix.

Schilderinck (1978) has given a complete picture of the geometric and algebraic approaches of principal factor analysis.

2.1.3.2 Maximum-Likelihood Method (ML method)

The distinction between the solutions obtained by using the principal factor method and maximum likelihood method is that former corresponds to a priori choice of communalities and the latter, the number of common factors. The ML solution is based on fundamental statistical considerations. It considers explicitly the differences between the correlations among the observed variables and the hypothetical values in the universe from which they were sampled.

The efforts to provide a sound statistical basis for factor analysis were made first by Lawley (1940, 1942) who suggested the use of "maximum likelihood method", due to Fisher (1922, 1925), in order to estimate the universe values of the factor loadings from the given empirical data. Lawley's ML method is possible only when the variates are normally distributed. It requires a hypothesis regarding the number of common factors.

Lawley (1940) and Rao (1952) had shown that "ML solution" goes to and fro between communalities and number of factors until it hits on the combination which yields the smallest residual.

Kaiser (1960) recommended (after considering statistical significance, algebraically necessary conditions) the number of common factors as the number of eigen values greater than or equal to one in the correlation matrix. He found this number to be about one-sixth or one-third of the total number of variables. The expression of ML method in factor analysis becomes more meaningful and clear with this foundation.

A more condensed derivation of ML method were appeared in a book by Lawley and Maxwell (1963).

Hemmerle (1965) found that Rao's procedure converges more rapidly than Lawley's procedure. Hemmerle (1965) in his paper considered the problem of computing estimates of factor loadings, specific variances and communalities for a factor analysis model. Iterative formulae were developed to solve the ML equations and a simple and efficient method of implementation of this method on a digital computer was developed by him.

The ML procedure remained impractical for all but for the smallest problems until the work of Joreskog (1967, 1969), as the process converge very slowly.

In Joreskog's (1967) ML method he proceeds systematically, fitting one, two, ... factors and testing at each stage by a chi-square test to see whether further factors are required. It also carries a varimax rotation at each stage. He also presents an example to compare the ML factor estimates with those given by principal components.

Kendall et al. (1983) reported that the ML solution remain scale-free if restrictions are imposed upon the parameters.

2.1.3.3 Factor rotation

Kaiser (1956) proposed the 'varimax' method as a modification of the quartiman method which nearly approximates simple structure. He found that a variable with communalities twice that of another will influence the rotations by four times as much.

As a last step in factor analysis, Cattell (1965 a) explained the rotational technique like 1. Simple structure and 2. Confactor rotation. In simple structure each factor affects only a few variables. But in confactor rotation real factor does happen to operate on all or most of the variables in the sample.

Cattell and Khanna (1977) described different approaches to factor rotation in which he introduced one kind of rotation criterion ie, confactor rotation, which arises when a second factorisation on the same variables with another group is involved.

2.2 Applied Studies

Lawley (1943) applied the ML method to factor analysis of data collected for research in education. This is a satisfactory method and deciding the number of factors required to account for the scores obtained when the number of individuals tested is reasonably large. In this case two general factors are needed to explain eight tests.

Murty and Qudri (1966) studied genetic divergence in a collection of forty self-compatible types of Brassica campestrics varieties using Mahalanobis D^2 statistic. The forty varieties were grouped into nine clusters.

Arunachalam and Jawahar (1967) studied the diversity in a population consisting of eighty genetic stocks of sorghum from 16 countries utilizing ten characters by multivariate analysis using D^2-statistic. The population was divided into three physiological groups.

Murty and Arunachalam (1967) have conducted a multivariate analysis of genetic divergence in the genus Sorghum (wild and cultivated types) using quantitative characters related to fitness under natural and human selection for the diversity found in this genus. The factors were obtained by the centroid method. Factor analysis revealed the adequacy of the three factors for differentiation.

Singh and Gupta (1968) assessed genetic divergence, using Mahalanobis D^2 statistic for yield and its components in thirty three strains of upland cotton evolved from seven diverse crosses. The thirty three strains were grouped into nine clusters.

Singh and Bains (1968) estimated genetic divergence in twenty varieties of upland cotton using Mahalanobis D^2 statistic. The varieties were grouped into five clusters.

Shetty (1969) determined the factors affecting the use of fertilizers among the farmers by using principal component method of factor analysis. The study revealed that the first four factors are sufficient for the explanation of the observed inter-farm variations in the use of fertilizers.

Ram and Panwar (1970) used Mahalanobis D^2 and canonical analysis to assess the nature of divergence and its relationship with the components of genetic variation in rice for four characters. The first two canonical roots accounts for 45 per cent of the total variability.

Gupta and Singh (1970) studied genetic divergence for yield and its components in green gram using Mahalanobis D^2 technique. The varieties differed significantly for the nine characters considered. The 36 strains were grouped in 10 clusters depending on similarities of their D^2 values.

Upadhyay and Murthy (1970) estimated genetic divergence in seventy varieties of pearl millet using Mahalanobis D^2 statistic.

Mehndiratta et al. (1971) studied genetic divergence in thirty varieties of sorghum using Mahalanobis D^2-statistics. The varieties were grouped into seven clusters.

Walton (1972) used factor analysis in identifying the morphological characters related to yield in spring wheats.

Singh (1973) used centroid method of factor analysis in upland cotton to study the evolutionary pattern of this often cross pollinated crop. Thirteen characters were included in the study. The first three factors accounted for 75 per cent or more of the total communality.

Abraham and Noobakht (1974) applied the technique of factor analysis to extract basis factors underlying the observed soil variables. Scores based on four underlying factors could be used for comparison of inter soil variables.

Chaudhary and Singh (1975) estimated genetic divergence in sixty four barley varieties using Mahalanobis D^2 statistic. The varieties were grouped into ten clusters.

Peter and Rai (1976) studied genetic divergence in twenty five varieties of tomato. The study revealed that there is no apparent parallelism between genetic and geographic divergence. The component characters locules per fruit and plant height were found to be important for the expression of genetic divergence.

Martin and Eaves (1977) adapted the analysis of covariance structures to the simultaneous maximum likelihood estimation of genetical and environmental factor loadings and specific variances. The goodness of fit is tested by chi-square and standard errors of parameter estimates can be obtained.

Nair and Gupta (1977) assessed the nature and magnitude of genetic diversity of 32 varieties of oats by multivariate analysis using D^2 statistics. The 32 varieties could be clustered into fourteen groups. Out of these, five clusters were found to be more divergent than the others.

Denis and Adams (1978) performed a principal factor analysis on 22 morphological and yield-determining traits of 16 cultivars and strains of dry beans. There were at least two or three principal factors to be examined for biological meaning and from which to seek insight into the basic structural design of bean plants.

Gaur et al. (1978) studied genetic divergence in potato. Sixty seven potato varieties were grouped in 15 clusters on the basis of D^2 values. The characters least influenced by the selection were mainly responsible for adding divergence to the population.

Tikka and Asawa (1978) used correlation in 28 genotypes of lentil for factor analysis through the principal component method as suggested by Harman. More than 90 per cent of the variability was extracted by two factors. Within each factor, traits were ranked according to the relative magnitude of factor loadings.

Kutiger and Singh (1979) measured the nature and magnitude of genetic diversity using Mahalanobis D^2-statistic for a set of eight characters related to yield and fitness in forty indigenous and exotic strains of chickpea. The population was grouped into ten different clusters.

Dixit (1980) conducted a study on genetic divergence for yield and its components in lentil using Mahalanobis D^2 technique. The 21 varieties were grouped into eight clusters depending on D^2 estimates.

Sundaram et al. (1980) used centroid method of factor analysis in cowpea to study its evolutionary pattern. The analysis divided the nine characters into three groups of factors which accounted for 98 per cent of total variation.

Singh et al. (1980) studied genetic divergence in 30 varieties of tomato using Mahalanobis D^2 technique for yield and its components. The varieties were grouped in eight clusters.

Sawant et al. (1982) utilized phenotypic correlations among seven traits in 90 diversified strains of triticale for factor analysis using principal component method. The factor analysis grouped the seven variables into two main factors which together accounted for about 46 per cent of total diversity

Singh et al. (1982) estimated genetic divergence among 48 exotic and 27 indigenous strains of chickpea using Mahalanobis D^2 statistic and 14 homogeneous genetic groups were formed.

Jatasra and Paroda (1983) studied genetic divergence in 28 hybrids of wheat using Mahalanobis D^2 statistic. All the hybrids got grouped into nine clusters.

Kendall et al. (1983) compared the ML factor estimates with those given by principal components by applying it to fifteen characteristics of 48 applicants for a post.

Anand and Rawat (1984) studied genetic divergence in fifty varieties of brown mustard using Mahalanobis D^2-statistic. The varieties were grouped into nine clusters.

Kukadia et al. (1984) conducted a study to determine the importance of various traits for yield improvement in forage sorghum. Genotypic correlations were subjected to factor analysis through the principal component method. Factors accounting for at least 10 per cent variability were retained and arranged in order of variance.

Singh and Gill (1984) assessed genetic divergence among sixty two varieties of upland cotton using Mahalanobis D^2 statistic. The varieties were grouped into twelve clusters.

Bartual et al. (1985) used factor analysis, principal component analysis and cluster analysis to identify sets of varieties better adaptable to the specific environmental conditions. Results obtained from ML factor analysis and principal component analysis were found to be similar.

Dobhal and Harihar Ram (1985) estimated genetic divergence in thirty two varieties of pea using Mahalanobis D^2-statistic. The varieties were grouped into eleven clusters.

Jindal and Gupta (1985) studied genetic divergence in thirty nine strains of fodder cowpea using Mahalanobis D^2-statistic. The strains were grouped into five clusters.

On the basis of multivariate analysis, Valsalakumari et al. (1985) grouped 62 cultivars of banana into 8 clusters considering 22 characters simultaneously. The characters pulp/peel ratio on volume basis followed by weight of fruit contributed the maximum towards divergence.

Morcy and George (1987) studied genetic divergence in 30 culinary varieties of banana by using D^2 analysis and canonical analysis. The varieties were grouped into twelve clusters using D^2 analysis.

Singh et al. (1981) conducted a study on the selection parameters in sugarcane. In 48 varieties of sugarcane there was a wide range of phenotypic variation for six of the eight traits studied, the exceptions being stalk weight and top weight. The phenotypic coefficient of variation was higher than the genotypic coefficient of variation.

Sukumaran et al. (1982) conducted a study to estimate the loss in weight and recovery of sugar in the lodged crop of sugarcane. The length of canes, number of millable canes, weight and recovery of sugar were found to be reduced as the canes lodge.

Nair et al. (1982) conducted a study on the performance of sugarcane varieties in Kerala.

Punia et al. (1983) studied genetic divergence in sugarcane using Mahalanobis D^2 technique and showed that genetic divergence to be high for all the twelve characters studied in 41 genotypes of sugarcane. The 41 genotypes were grouped into 10 clusters depending upon D^2 estimates.

Singh et al. (1983) conducted a study on variability for yield and quality in sugarcane and indicated a wide range of variability for number of tillers, number of millable canes, sucrose percentage in juice and cane yield. The number of tillers and number of millable canes were positively and significantly correlated with cane yield.

Gill et al. (1983) conducted a character association analysis in 28 foreign varieties and two Indian varieties of sugarcane. The study revealed that percentage commercial cane sugar had a positive correlation with cane yield, juice purity, sucrose percentage and number of millable canes.

Nageswara Rao et al. (1983) studied genetic variability and character associations in 19 crosses of sugarcane progenies. Variance was high for stalk length while coefficient of variation was higher for clump weight and millable stalks/clump.

Punia et al. (1983) conducted correlation and path analysis on 41 genotypes of sugarcane. Cane yield/clump was significantly associated with the number of tillers/clump, the number of millable canes/clump, cane thickness and cane weight.

MATERIALS AND METHODS

MATERIALS AND METHODS

3.1 Materials

The morphological and quality traits in the material used for the study consists of 48 clones of sugarcane (*Sacharum officinarum* L.) collected from the germplasm maintained at the Sugarcane Research Station, Thiruvalla. The clones were planted in a randomised block design with three replications, during January 1981. This experiment was the project work of Sreekumar (1986) for his Ph.D. programme. Data on the following characters were collected from the plant crop.

x_1. Germination count: The percentage of sprouts in each plot on the 45th day.

x_2. Shoot count: The number of shoots per plot on the 180th day.

x_3. Brix: One litre of juice was taken and the brix reading recorded using a standard brix spindle. This was estimated at the 12th month.

x_4. Pol percentage: Estimated by Horner's dry lead method.

x_5. Purity percentage: Purity of the juice was expressed as the percentage of pol to brix at the 12th month.

x_6. Number of millable canes: Number of fully mature, healthy canes per plot at the time of harvest.

x_7. Juiciness: Estimated at the 12th month. A sample of two healthy canes was cut from each plot, crushed in a power crusher and the juice extracted. Juiciness was estimated as the volume of juice (ml) obtained from one kilogram of cane.

x_8. Length of internode: Mean length of the middle most internode from the random sample of 5 canes.

x_9. Girth of cane: Mean girth of the middle most internode from the random sample of 5 canes.

x_{10}. Number of internodes: Mean number of internodes per cane from the random sample of 5 canes.

x_{11}. Weight of cane: Mean weight of cane from a sample of 5 canes selected at random from each plot.

x_{12}. Yield of cane: Weight of millable canes per plot at the 12th month.

x_{13}. Length of cane: Mean length of cane from the random sample of 5 canes.

x_{14}. Commercial cane sugar percentage: C.C.S. was determined as per the following formula suggested by Mathur, at the 12th month.

$$C.C.S. = S - [0.4 (B-S)] F$$

where B = Brix

S = Pol percentage

F = 0.73 - Factor relative to fibre percentage of cane

x_{15}. Yield of sugar: Sugar yield per hectare was calculated by multiplying C.C.S. percentage by cane yield per hectare and dividing by 100.

The varieties taken for the study are listed in Table 3.1.1.

Table 3.1.1 Sugarcane varieties taken for the study

Code Number	Name of variety	Code Number	Name of variety	Code Number	Name of variety
1	Coc.774	17	Coc.777	33	Co.7704
2	F.1-2	18	S-105	34	CoA.7601
3	T.67172	19	S-33	35	Co.62198
4	Co.658	20	M.S.6847	36	Co.62101
5	Co.62174	21	Co.740	37	Co.6806
6	Co.997	22	Ic.225	38	Coc.778
7	Co.6807	23	Co.6907	39	B.37172
8	Co.1340	24	Co.6304	40	Co.1305
9	Co.1307	25	CoA.7602	41	Co.785
10	Co.7717	26	S-99	42	Co.453
11	Co.62175	27	Coc.775	43	CoM.7114
12	S-87	28	KHS 3296	44	S-77
13	Co.419	29	Coc.671	45	Co.995
14	Coc.779	30	Coc.771	46	Co.449
15	Co.7219	31	Coc.773	47	CoM.7125
16	Co.527	32	Coc.772	48	Co.527-M-10

3.2 Methodology

3.2.1 Structure of multivariate observations

Multivariate analysis is concerned with analysing multiple measurements that have been made on one or several samples of individuals, and as such it deals with the jointness of p measures on n subjects.

The mathematical model on which most of the multivariate procedures are based is on the assumption of multivariate normal distribution (m.n.d.). This assumption of m.n.d. for multiple measures can be justified by the same central limit theorem argument that leads to the assumption of normality for a univariate measurement. "The multivariate normal distribution often occurs because the multiple measurements are sums of small independent effects" (Anderson, 1958).

Measurements on biometrical characters for n varieties replicated q times were denoted by x_{ijk} where ($i = 1, 2, \ldots, p$; $j = 1, 2, \ldots, q$; $k = 1, 2, \ldots, n$). Suppose the random variables x_i of interest have a multivariate normal distribution with mean $\mu_{px1} = (\mu_1\ \mu_2 \ldots \mu_p)$ and covariance matrix $\Sigma_{pxp} = (\sigma_{ij})$. If the measurements of interest are in widely different units, a more accurate picture of dependence pattern be obtained by standardising variable as $z_i = \frac{x_i - \mu_i}{\sigma_i}$, i = 1,2, ... p. Then analysis of the dependence structure of z_1 .... z_p which is given by the correlation matrix of x_1 .... x_p is done. Thus the observed correlation among variables constitute the original data.

3.2.2 Preliminary statistical analysis

The data were subjected to multivariate analysis of a randomised block design with the ANOVA model as

$$x_{ijk} = \mu_i + t_{ij} + b_{ik} + e_{ijk}, \quad i = 1,2, \ldots p$$

where μ_i is the general mean, t_{ij} is the effect of j^{th} treatment, b_{ik} is the k^{th} block effect and e_{ijk} is the error component, with respect to the i^{th} character and e_{ijk} are normally distributed with mean zero and constant variance.

The least square estimates of the constants of the model are

$$\hat{\mu}_i = \bar{\bar{x}}_{i..}$$

$$\hat{t}_{ij} = \bar{\bar{x}}_{ij.} - \bar{\bar{x}}_{i..}$$

$$\hat{b}_{ik} = \bar{\bar{x}}_{ik} - \bar{\bar{x}}_{i..}$$

Table 3.2.1 ANOVA for RBD

Source	d.f	M.S.
Blocks	q-1	
Treatments	n-1	s_t^2
Error	(n-1) (q-1)	s_e^2
Total	nq-1	

3.2.3 Analysis of dispersion

Multivariate analysis of variance was first developed by Wilks (1932 a). Analysis of dispersion is the process which involves the technique of analysing the variances and covariances of variables in multivariate case (Rao, 1952). The total dispersion is split up into various components as follows.

Table 3.2.2 MANOVA of p variables

Source	d.f.	Dispersion matrix
Deviation from hypothesis	n-1	B
Error	n (p-1)	W
Total	np-1	

The criterion arrived at by Wilks (1932 a) through the generalised likelihood ratio principle is given by $\Lambda = \frac{|W|}{|W+B|}$

where W is the within dispersion matrix

B is the between dispersion matrix

The statistic used for testing the homogeneity of treatment means for all the characters taken together is given by

$$V = -m\log_e \Lambda$$

where V is distributed as χ^2 with (n-1) p degrees of freedom and $m = nq-1 + \frac{(p+n)}{2}$ (Bartlett, 1947).

3.2.4 Estimation of correlation matrix

The phenotypic, environment and genotypic correlations were estimated from the following analysis of variance-covariance of the data.

Table 3.2.3 Analysis of covariance of RBD

Source	d.f	$MS(x_i)$	$MS(x_j)$	$MSP(x_ix_j)$
Replication	q-1			
Treatment	n-1	MSV_i	MSV_j	MSV_{ij}
Error	(q-1) (n-1)	MSE_i	MSE_j	MSE_{ij}
Total	qn-1	MSP_i	MSP_j	MSP_{ij}

Phenotypic correlation coefficient

$$r_{pipj} = \frac{MSP_{ij}}{(MSP_i \; MSP_j)^{\frac{1}{2}}} \quad i \neq j$$

The environment correlation coefficient

$$r_{eiej} = \frac{MSE_{ij}}{(MSE_i \; MSE_j)^{1/2}}, \quad i \neq j$$

Genotypic correlation coefficient

$$r_{gigj} = \frac{(MSV_{ij} - MSE_{ij})/q}{\left[\frac{(MSV_i - MSE_i)}{q} \; \frac{(MSV_j - MSE_j)}{q}\right]^{1/2}}, \quad i \neq j$$

The environment correlation matrix was found to be appropriate for factor analytical studies as it leads to stable factor pattern (Muralidharan, 1986; Tes P.Mathew, 1987). Phenotypic and genotypic correlation matrices failed to give stable factor pattern. So the environment correlation matrix was taken here for the present study.

3.2.5 D^2 - analysis

A measure for group distance based on multiple characters was given by Mahalanobis (1928). With $x_1, x_2 \ldots x_p$ as the multiple measurements available as each individual and d1, d2, ... dp as $\bar{\bar{x}}_1^1 - \bar{\bar{x}}_1^2$, $\bar{\bar{x}}_2^1 - \bar{\bar{x}}_2^2$, ... $\bar{\bar{x}}_p^1 - \bar{\bar{x}}_p^2$ respectively, being the difference in the means of two populations, Mahalanobis' D^2-statistics is defined as follows.

$$D^2 = b_1 d_1 + b_2 d_2 \ldots + b_p d_p \qquad \text{—— (1)}$$

where $\bar{\bar{x}}_i^1$ is the mean value of i^{th} character in the first population and $\bar{x}_i^2$ is the mean value of the i^{th} character in the second population. Here, the bi values are to be estimated such that the ratio of variance between the populations to the variance within the population is maximised. In terms of variances and covariances, the D^2 value is obtained as follows.

$$D^2 = \sum_i \sum_j W^{ij} (\bar{x}_i^1 - \bar{x}_i^2) (\bar{x}_j^1 - \bar{x}_j^2) \quad \text{——} \quad (2)$$

where, W^{ij} is the inverse of estimated variance covariance matrix.

Estimation of D^2 values by the formula given in equation (2) is very complicated when the number of characters being studied becomes large. The computation is very much simplified when the characters under study are independent and are expressed in terms of their respective standard errors. In this case, computation of D^2 value reduces to simple summation of the differences in mean values of various characters of the two populations ie, d_i^2. Therefore, first transformed the correlated variables to uncorrelated ones and then worked out the D^2 values. Transformation was done by using pivotal condensation method. Let Y_1, Y_2 .... Y_p be the transformed variates. For each combination the mean deviation, ie, $Y_i^1 - Y_i^2$ with $i = 1, 2, \ldots p$ was computed and the D^2 was calculated as sum of the squares of these deviations. ie, $\sum (Y_i^1 - Y_i^2)^2$.

3.2.5.1 Test of significance of D^2 values

The D^2 value obtained for a pair of population was taken as the calculated value of χ^2 and was tested against the tabulated value of χ^2 for p degrees of freedom, where p is the number of characters considered.

3.2.5.2 Grouping of varieties into various clusters

Tocher method

The first step in grouping the varieties into distinct clusters was to arrange the populations in order of their relative distances from each other. The two populations having smallest distance from each other were considered first to which a third population having smallest average D^2 value from the first two populations was added. Then the nearest fourth population and so it goes on. At certain stage it was felt that after adding a particular population, there was abrupt increase in the average D^2, this population was not added to that cluster. Similarly, a second cluster was formed. The process was continued till all the populations were included into one or the other cluster.

3.2.6 Factor analysis

Factor analysis is the common term for a number of statistical techniques for the resolution of a set of variables in terms of a small number of hypothetical variables, called factors. It reduces the multiplicity of tests and measures to greater simplicity. The fundamental step in the analysis of a body of observed data is the formulation of a theoretical statistical model. A linear model is used in order to explain observed phenomena in terms of simple theories.

The basic factor analysis model can be written in matrix notation as

$$\underline{Z} = \underline{AF} + \underline{e} \qquad \text{—— (1)}$$

where $\underline{Z}$ is the p x 1 vector of standardised variables

$\underset{\sim}{A}$ is the p x k matrix of factor coefficients

$\underline{F}$ is the k x 1 vector of $(k < p)$ common factors

$\underline{e}$ is the p x 1 vector of specific (unique) factors.

This equation states that the observed variables are weighted combinations of the common factors and the unique factors. The common factors account for the correlations among the variables and the unique factor account for the remaining variance including error of that variable. The total unit variance of a standardised variable Zi is made up of the communality attributable to the common factor and the uniqueness, which is the contribution of the unique factor (Harman, 1967).

In factor analysis it is usual to discard the sample mean vector and to make use of the covariance matrix or correlation matrix alone. The dispersion matrix of the variates in Z is defined as $E(zz')$ and is symmetric and positive definite of order p. The assumptions are

$$E(Fe') = 0 \qquad \text{---} \quad (2)$$

$$E(FF') = I_k \qquad \text{---} \quad (3)$$

& $$E(ee') = \psi \qquad \text{---} \quad (4)$$

where ψ is a diagonal matrix with diagonal elements as ψ_i

Since $E(ZZ') = E\left[(AF + e)(AF + e)'\right]$

We have $R = \Lambda\Lambda' + \psi \qquad \text{---} \quad (5)$

where R is the correlation matrix

In practice A and ψ are unknown parameters which are to be estimated from experimental data.

Principal factor analysis method, centroid method, maximum likelihood method, minimum residual method etc. are some of the methods for estimating the parameters A and ψ. Among these methods some require estimates of communalities while others require estimates of the number of common factors.

3.2.6.1. Exploratory versus confirmatory factor analysis

A particular application of factor analysis is exploratory or confirmatory according as the number of parameters prespecified in the model equation of factor analysis (Joreskog, 1969). In this study exploratory factor analysis is done by the principal factor analysis and maximum likelihood methods.

3.2.6.2 Estimation of communality

Communality is the amount of variance of the characters accounted for by the common factors (Fruchter, 1954).

There are various methods of estimating communality. But the squared multiple correlation (SMC) of each variable with all other variables of the set seems to be the 'Best Possible' systematic estimate of communality (Guttman, 1956).

The SMC of variable Zi is given by SMCi

$$R^2_{i.\ 1,2,\ \ldots.\ (i-1),\ (i+1)\ \ldots.\ p} = 1 - \frac{1}{r^{ii}} \qquad \text{---} \quad (6)$$

where r^{ii} is the diagonal element of R^{-1} corresponding to the variable Zi. The SMC has another important property that it is the lower bound of the communality (Harman, 1967).

The maximum correlations in corresponding row or column may also be taken as initial estimates of communality (Cattell, 1965 a).

3.2.6.3 Principal factor analysis (PFA)

The application of the principal components to the reduced correlation matrix with estimates of communalities in the diagonal instead of unity leads to the principal factor analysis. This method yields a mathematically unique solution of component correlations.

From the classical factor analysis model (1) the relevent portion of the determination of the common factor coefficients may be

$$\underline{Z} = \underline{A}\,\underline{F} \qquad \text{---} \quad (7)$$

or

$$\begin{aligned} Z_1 &= a_{11} F_1 + a_{12} F_2 + \cdots\cdot + a_{1k} F_k \\ &\cdots\cdots\cdots\cdots\cdots \qquad \text{---} \quad (8) \\ &\cdots\cdots\cdots\cdots\cdots \\ Z_p &= a_{p1} F_1 + a_{p2} F_2 + \cdots\cdot + a_{pk} F_k \end{aligned}$$

The sum of squares of factor coefficients gives the communality of a particular variable while a_{im}^2 indicates the contribution of the factor F_m to the communality of Z_i. The principal factor method involves the selection of the first factor coefficients a_{i1} so as to make the sum of the contribution of that factor to the total communality a maximum.

$$\text{ie, } V_1 = a_{11}^2 + \cdots\cdot + a_{p1}^2 \qquad \text{---} \quad (9)$$

is maximum. The coefficients a_{i1} must be chosen such that V_1 is maximum under conditions

$$r_{ij} = \sum_{m=1}^{K} a_{im}\, a_{jm} \quad \text{--- (10)} \quad i, j = 1, 2, \ldots, p$$

where $r_{ij} = r_{ji}$ and r_{ii} is the communality h_i^2 of the i^{th} variable.

This condition implies that the observed correlations are to be replaced by the reproduced correlations, implying the assumption of zero residuals. V_1 is maximised by applying the method of Lagrangian multipliers under the conditions (10).

The maximisation of V_1 leads to the system of p equations in p unknown a_{i1}.

ie, $(R_1^* - \lambda_1 \; I)\, q_1 = 0$ --- (11)

where R_1^* is the reduced correlation matrix

ie, $R_1^* = R_1 - \psi$

q_1 is the latent vector corresponding to the latent root λ_1.

$\lambda_1 = \sum_{i=1}^{p} a_{i1}^2$ and $\lambda_2 = \sum_{i=1}^{p} a_{i2}^2$ and so on.

The linear homogeneous equation system (11) has only a non-trivial solution if its determinant is equal to zero.

$$\text{ie, } \left| R_1^* - \lambda_1 \; I \right| = 0 \qquad \text{--- (12)}$$

The criterion regarding the number of common factors to retain in the factor model is equal to the number of principal components whose eigen values are greater than one. The investigator will usually be satisfied with an even smaller number of factors.

The characteristic equation (12) gives latent roots $\lambda_1, \lambda_2, \ldots \lambda_k \geq 0$ and the associated orthogonal characteristic vectors $\underline{g}_1, \underline{g}_2, \ldots \underline{g}_k$

Jacobi method is used to find out the eigen values and vectors of the matrix A. The idea of the Jacobi's method is to pick up the largest off-diagonal element of the matrix and to annihilate it to zero by applying a proper orthogonal transformation. Then the largest remaining off-diagonal element found out and that is annihilated. The procedure is repeated until the off-diagonal elements were sufficiently close to zero or negligible. The diagonal elements of the matrix is a close approximation to the eigen values. If the successive transformation matrices were multiplied together, they would produce an accurate approximation to the matrix of eigen vectors (Mulaik, 1972).

Substituting the largest characteristic root λ_1 in (11) we get corresponding characteristic vector.

$$\underline{g}_1' = (q_{11}, q_{21}, \ldots q_{p1}) \qquad \text{---} \qquad (13)$$

The normalized characteristic vector $\underline{g}_1$ which fulfil the conditions (9) and (10) is

$$\hat{g}_1 = \frac{g_1}{(g_1' \, g_1)^{1/2}} \qquad (14)$$

then the first column vector of factor loading matrix is determined as $a_1 = g_1\sqrt{\lambda_1}$ — (15)

The second column vector of A is $a_2 = g_2\sqrt{\lambda_2}$ and so on. This shows that a_1, a_2 .... are scaled normalized characteristic vectors.

The sum of the squares of factor loadings of the variable gives the corresponding communality ie, the squared factor coefficients can be considered as the percentage variance components of the common factor (Harman, 1967). The iteration process is continued with the new estimates of communalities until a specified degree of convergence is occurred. The controlling equation to ensure that no vital information is lost ie

$$R_1^* = AA' \qquad (16)$$

There are many equivalent matrices which all satisfy $R_1^* = AA'$. It implies also the making of a reasonable choice among the many possibilities to perform a final Matrix A, which contains a suitable interpretation of the relation under research. This results in the rotation of the factors of the initial matrix A.

3.2.6.4 Factor rotation

After extraction, the matrix of factor loadings are submitted to varimax orthogonal rotation, the effect of which is to accentuate the larger loadings in each factor and suppress the minor loading coefficients, and in this way improve the opportunity of achieving a meaningful biological interpretation of each factor (Denis and Adams, 1978).

Kaiser's (1958) varimax rotation is one in which factors are rotated in such a way that the new loadings tend to be either relatively large or relatively small in absolute magnitude compared with the original ones. The simplicity of a factor in defined as the variance of its squared loadings.

$$V_k = \frac{p \sum_{i=1}^{p} (a_{im}^2/h_i^2)^2 - (\sum_{i=1}^{p} a_{im}^2/h_i^2)^2}{p^2} \quad \text{—— (17)}$$

where a_{im} is the new factor loading for variable i on factor m, where i = 1,2 .... p and m = 1,2, ... k

For entire factor matrix the normalized varimax criterion is

$$V = \sum_{i=1}^{k} \frac{\left[p \sum_{i=1}^{p} (a_{im}^2/h_i^2)^2 - (\sum_{i=1}^{p} a_{im}^2/h_i^2)^2\right]}{p^2} \quad \text{—— (18)}$$

where h_i^2 is communality of i^{th} variable. The fundamental rationale for attempting to establish the normal varimax criterion is that the normal varimax criterion is that the normal varimax solution is invariant under changes in the composition of the variables.

3.2.6.5 Maximum likelihood factor analysis

Alternate methods that circumvent many of the problems of principal factor analysis have been suggested. One such method is maximum-likelihood factor analysis proposed by Lawley (1940) and later which provides maximum likelihood estimates for the factor loadings. Maximum likelihood solution requires an estimate of the number of common factors. A ML solution has the same general appearance as a PF solution, but it does not have the latter's property of accounting for a maximum amount of variance for a specified number of factors. Also, while a PF solution is unique for a given body of data, a ML solution differs from another by a rotation (Harman, 1967). When estimating a population parameter, if a sufficient statistic exists maximum likelihood estimates are functions of sufficient statistic. Moreover, the ML estimator is a consistent estimator as well as frequently a minimum variance estimator (Mulaik, 1972). A well known property of ML method of factor analysis is that it is independent of the units of measurement in the characters.

The model to be used in this method is (1). Also x follows multivariate normal distribution with mean vector and covariance matrix .

The sample covariance matrix of x is denoted by S where $S = \frac{1}{n}\sum_{k} (x_k - \bar{x})\ (x_k - \bar{x})$

$$\bar{x} = \frac{1}{n} \sum_{k=1}^{n} x_k$$

where x_k is the column vector of random sample of n (p) observations of x. k = 1,2, ..., n, m = n-1. The distribution of S is Wishart with m d.f. ie, $mS \sim W(\Sigma, m)$

Here $E(S) = \Sigma$

The logarithm of the likelihood function for the sample, omitting a function of the observations, is given by

$$\log_e L = \frac{-n}{2} \quad \log_e |\Sigma| + tr\ (S\Sigma^{-1}) \quad \text{---} \quad (19)$$

This is regarded as a function of A and ψ. Considering these as mathematical variables we seek values of A and denoted eventually by $\hat{A}$ and $\hat{\psi}$ that maximise the value of $\log_e L$. It is more convenient to minimise the function.

$$F_k\ (A, \psi) = \log|\Sigma| + tr\ (S\Sigma^{-1}) - \log|\Sigma| - p \quad \text{---} \quad (20)$$

For the purpose of minimising the function F the partial derivatives with respect to the elements of A and the diagonal elements of ψ which is given by

$$\frac{\partial F}{\partial A} = 2\,\Sigma^{-1}(\Sigma - S)\,\Sigma^{-1} A \quad \text{—— (21)}$$

$$\frac{\partial F}{\partial \psi} = \operatorname{diag}\left[\Sigma^{-1}(\Sigma - S)\,\Sigma^{-1}\right] \quad \text{—— (22)}$$

are required.

Equating $\frac{\partial F}{\partial A}$ and $\frac{\partial F}{\partial \psi}$ to zero and solving the resulting equations to get the estimates of A and ψ (Lawley & Maxwell, 1971). The estimation equations are independent of the scale of measurement of the X's and consequently the estimation equations for the a's can be expressed in terms of the correlations rather than the covariances (Lawley, 1970).

$$\text{ie, } R = AA' + \psi \quad \text{—— (23)}$$

$$\text{and } \psi = I - \operatorname{diag} AA' \quad \text{—— (24)}$$

$$A' R^{-1} A \text{ is diagonal} \quad \text{—— (25)}$$

Premultiplying both sides of (23) by $A'\psi^{-1}$ yields

$$(A'\psi^{-1}A + I)\,A' = A'\psi^{-1} R \quad \text{—— (26)}$$

This equation can be simplified to

$$JA' = A\,\psi^{-1}R - A' \quad \text{—— (27)}$$

$$\text{where } J = A'\psi^{-1}A \quad \text{—— (28)}$$

which is amenable to an iterative method of solution (Lawley, 1942)

Starting with an arbitrary factor matrix $A = (\underline{a}_1, \underline{a}_2, \ldots \underline{a}_m)$ (usually loadings obtained from principal factor analysis) and corresponding

$$\psi_1 = I - \text{diag}\ AA' \qquad \text{---} \qquad (29)$$

the factor loadings $B = (\underline{b}_1, \underline{b}_2, \ldots \underline{b}_m)$ are derived from the iterative process, where

$$\underline{b}_1 = \frac{(R\ \psi^{-1} a_1 - a_1)}{a_1^1\ \psi^{-1}\ (R\ \psi^{-1} a_1 - a_1)^{\frac{1}{2}}}$$

$$\underline{b}_2 = \frac{(R\ \psi^{-1} a_2 - a_2 - b_1 b_1^1\ \psi^{-1} a_2)}{a_2^1\ \psi^{-1}\ (R\ \psi^{-1} a_2 - a_2 - b_1 b_1^1\ \psi^{-1} a_2)^{\frac{1}{2}}}$$

....

....

$$\underline{b}_m = \frac{(R\ \psi^{-1} a_m - a_m - \ldots\ b_{m-1} b_{m-1}^1\ \psi^{-1} a_m)}{a_m^1\ \psi^{-1}\ (R\ \psi^{-1} a_m - a_m - \ldots - b_{m-1} b_{m-1}^1\ \psi^{-1} a_m)^{\frac{1}{2}}}$$

$$\psi_2 = I - \text{diag}\ BB'$$

The iterative process is repeated again and again until the convergence is obtained to the desired degree of accuracy. In standardised variates, the convergence criterion has usually be taken as 0.005. The final matrix A contains the ML estimates of factor loadings for the assumed number of common factors. In this iterative method it is tacitly assumed that none of the uniquenesses vanish. In some cases the maximisation of the likelihood function leads to one or more of the variables with uniqueness essentially zero. In the literature of factor analysis this type of improper solutions have usually been known as Heywood case. Joreskog (1967) has made a provision for the Heywood case.

It is assumed that a maximum likelihood factor analysis with a certain value of k has been performed resulting in an improper solution with m ($\leq$ k) of the unique variances zero. Assuming that this has occurred for the first m variables, the dispersion matrix may be partitioned as

$$S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \qquad --- \quad (30)$$

where Matrices S_{11}, S_{12}, S_{21} and S_{22} are of orders m x m, m x (k-m), (p-m) x m and (p-m) x (k-m) respectively. Then the estimates $\hat{A}_{11}$, $\hat{A}_{12}$ and $\hat{A}_{21}$ are defined as

$$\hat{A}_{11} = S_{11} \Gamma \Delta^{-\frac{1}{2}} \qquad --- \quad (31)$$

$$\hat{A}_{21} = S_{21} \Gamma \Delta^{-\frac{1}{2}} \qquad --- \quad (32)$$

and $$\hat{A}_{12} = 0 \qquad --- \quad (33)$$

where Γ is an orthogonal matrix of order m x m that reduces S_{11} to diagonal form and Δ is a diagonal matrix containing latent roots of S_{11}. The matrices $\hat{A}_{22}$ and $\hat{\psi}_2$ are obtained by applying the maximum likelihood method to the conditional dispersion matrix.

$$S_{22.1} = S_{22} - S_{21} S_{11}^{-1} S_{12} \qquad \text{—— (34)}$$

In the analysis of $S_{22.1}$ the number of variables is decreased by m and also the number of factors is decreased by m. Then

$$\hat{A} = \begin{bmatrix} \hat{A}_{11} & \hat{A}_{12} \\ \hat{A}_{21} & \hat{A}_{22} \end{bmatrix} \text{ and } \hat{\psi} = \begin{bmatrix} 0 & 0 \\ 0 & \hat{\psi}_2 \end{bmatrix}$$

are the maximum likelihood estimates of A and ψ.

3.2.6.5.1 Test of significance for the number of factors

One of the main advantages of using the maximum likelihood method of estimation is that it enables us to test the hypothesis H_k that, for specified K, there are K common factors. After obtaining a proper solution the hypothesis is tested by

$$U_k = \left[N-1 - \frac{(2p+5)}{6} - \frac{2k}{3} \right] F_k \ (\hat{\psi})$$

where $$F_k(\hat{\psi}) = \sum_{i<j} \frac{(s_{ij} - \hat{\sigma}_{ij})^2}{\hat{\psi}_i \ \hat{\psi}_j}$$

$s_{ij} - \hat{\sigma}_{ij}$ represents the residual covariance of X_i and X_j

after eliminating K common factors. The criterion U_k is actually a measure of how much the residual covariances differ from zero. Under H_k, for moderately large n, U_k is very nearly distributed as χ^2 with d_k d.F. where d_k = $\frac{1}{2}\left[(p-k)^2 - (p+k)\right]$

This exactly imposes an upper limit on m for given p. ie, The number of common factors cannot exceed the largest integes satisfying $m < \frac{1}{2}(2p+1 - \sqrt{8p+1})$ for a fixed number of p variables.

The non significance of χ^2 means that there would be no point in fitting further factors to the data.

The computations were carried out on the VERSA INS system in the Statistics Department of the KAU.

RESULTS AND DISCUSSION

RESULTS AND DISCUSSION

The results of the present study are given in sections 4.1 to 4.5 under the headings

4.1 Preliminary statistical analysis

4.2 Analysis of dispersion

4.3 Estimated correlation matrices

4.4 D^2 - Analysis

4.5 Factor analysis

4.1 Preliminary statistical analysis

The analysis of variance was done for each character under study. Significant differences were observed among the genotypes with respect to each character. The mean values of the various characters are presented along with their test of significance in Table 4.1.1.

4.2 Analysis of dispersion

Multivariate analysis of variance was performed and the total dispersion matrix was split up into 'between' and 'within' dispersion matrices and the results are given in Appendices I and II respectively. The value obtained for Wilk's lambda statistic was = 4.9×10^{-9}. So that $V = 1566.9$ which is distributed as a chi-square with 705 degrees of freedom and this was significant at one per cent level.

Table 4.1.1 Mean values of various characters and their test of significance with reference to 48 clones of sugarcane

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
1.	45.06	98.33	17.73	14.55	82.26	73.33	444.08	13.01
2.	38.89	74.67	18.99	17.23	90.62	68.00	432.79	11.08
3.	42.90	81.00	14.07	11.47	80.95	70.00	469.91	12.73
4.	53.09	109.33	17.43	15.42	87.86	99.33	443.45	11.50
5.	30.86	49.33	17.20	14.88	86.05	45.67	452.38	10.65
6.	39.20	113.00	20.33	18.63	91.54	99.33	424.89	9.53
7.	37.04	93.33	14.58	11.17	75.75	75.67	450.98	11.72
8.	35.19	96.67	15.65	11.80	75.25	84.00	423.41	12.10
9.	41.98	75.33	15.18	12.86	84.48	69.33	453.22	13.90
10.	61.42	94.67	14.83	12.01	80.76	76.00	458.72	12.98
11.	51.54	89.67	18.53	16.75	90.31	80.67	508.92	11.23
12.	42.59	68.00	19.73	17.64	89.27	59.00	472.71	10.57
13.	49.69	84.00	17.77	15.50	87.14	73.67	481.62	12.19
14.	45.99	90.00	14.69	11.55	78.50	84.00	490.29	12.51
15.	41.36	78.67	18.71	16.03	85.64	66.33	424.42	12.90
16.	28.70	91.67	15.35	12.53	81.65	66.33	453.96	10.85

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
17.	42.90	99.67	17.24	14.73	85.23	81.67	467.97	12.71
18.	50.31	77.00	16.26	13.37	81.68	66.67	457.64	10.69
19.	34.26	94.00	18.20	16.26	89.13	87.33	402.97	12.31
20.	39.81	70.33	12.12	8.53	66.05	54.33	438.89	14.32
21.	52.78	84.00	19.19	15.50	80.56	70.67	433.88	11.90
22.	47.22	96.33	16.62	14.71	87.92	76.33	438.09	11.07
23.	44.14	82.33	19.87	18.43	92.66	79.00	470.17	12.42
24.	47.22	77.33	14.96	11.95	80.04	58.67	429.91	12.49
25.	39.20	87.00	16.97	14.18	83.36	79.33	458.95	15.45
26.	39.81	96.32	19.08	16.79	87.77	82.67	441.14	13.25
27.	44.14	86.00	16.63	14.29	85.01	77.00	377.22	15.79
28.	38.27	64.33	18.55	16.91	91.19	53.67	446.39	10.80
29.	29.01	54.00	19.05	17.85	93.80	46.33	475.40	13.08
30.	62.35	107.33	17.10	14.31	82.55	95.33	378.43	15.81
31.	43.21	66.33	17.27	15.10	87.34	61.00	415.03	12.48
32.	49.38	72.67	16.03	13.49	84.04	66.67	407.99	12.13

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
33.	39.51	66.67	20.19	18.62	92.20	59.33	433.55	11.65
34.	25.31	46.67	15.26	12.99	84.95	45.00	461.81	12.85
35.	43.83	74.67	16.63	14.37	86.31	61.00	409.18	13.39
36.	28.09	62.67	15.46	12.12	75.34	62.67	447.46	12.51
37.	36.11	95.67	18.03	15.15	84.16	78.33	407.68	12.71
38.	47.22	83.67	13.68	10.99	79.49	60.67	462.42	11.98
39.	50.62	107.33	19.47	17.60	90.27	91.00	434.31	12.02
40.	47.53	76.33	17.12	15.04	87.87	69.67	386.84	11.87
41.	54.32	100.00	18.74	15.75	84.16	90.67	428.92	16.66
42.	53.40	91.67	13.50	10.30	75.13	69.33	425.61	14.83
43.	50.93	86.33	16.03	13.25	82.13	68.33	382.64	10.50
44.	47.53	89.00	17.06	14.72	86.12	71.33	376.84	11.21
45.	50.00	101.67	17.89	15.46	86.41	96.67	458.33	11.31
46.	46.91	115.67	14.57	11.44	78.49	102.33	355.04	12.84
47.	35.80	73.00	18.94	16.46	86.86	53.00	454.17	12.10
48.	48.77	96.67	18.09	15.68	86.29	82.00	383.03	11.97
F-values	5.01**	6.94**	3.66**	3.42**	1.91*	4.81**	1.72*	6.89**

* Significant at 5% level ** Significant at 1% level

	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}
1.	8.85	20.53	1.48	103.17	2.43	9.70	8.13
2.	6.91	24.80	1.36	89.87	2.75	12.07	8.87
3.	7.72	22.33	1.60	104.10	2.61	7.61	6.54
4.	7.41	24.33	1.42	144.97	2.65	10.67	12.53
5.	8.61	22.40	1.43	72.22	2.34	10.19	6.30
6.	6.92	24.20	1.07	111.72	2.45	12.81	11.33
7.	7.31	23.67	1.38	108.10	2.59	7.16	5.61
8.	6.94	22.27	1.03	97.61	2.47	7.49	6.01
9.	8.19	23.60	1.64	122.13	3.17	8.71	8.83
10.	8.27	22.53	1.63	111.72	2.66	7.94	7.24
11.	8.82	30.13	1.86	143.57	3.03	11.71	13.58
12.	8.25	29.47	1.83	114.19	3.12	12.25	11.32
13.	8.32	25.13	1.73	133.53	2.83	10.66	11.36
14.	7.99	24.93	1.60	133.12	2.76	7.52	7.98
15.	7.45	22.80	1.36	93.98	2.84	10.92	8.30
16.	6.64	19.73	1.06	56.42	2.26	8.36	3.82

	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}
17.	7.93	24.53	1.47	145.92	2.90	10.02	11.87
18.	8.01	26.73	1.44	101.85	2.76	8.92	7.44
19.	6.49	22.33	1.21	90.86	2.70	11.31	8.28
20.	9.46	22.80	1.99	108.55	2.98	5.18	4.29
21.	7.40	25.00	1.47	107.40	2.78	10.23	8.49
22.	7.26	26.20	1.44	115.71	2.76	10.19	9.77
23.	7.73	25.80	1.56	110.03	2.86	13.03	11.53
24.	7.47	23.40	1.48	83.82	2.64	7.85	5.34
25.	7.96	19.53	1.58	106.87	2.84	9.54	8.23
26.	8.09	23.73	1.61	125.39	2.71	11.59	11.52
27.	7.18	20.47	1.25	94.97	2.87	9.74	7.34
28.	7.98	24.40	1.69	92.13	2.38	11.86	8.91
29.	8.19	21.40	1.66	79.99	2.59	12.69	8.21
30.	8.11	21.07	1.57	156.00	3.12	9.63	12.02
31.	7.64	22.53	1.39	81.97	2.72	10.39	6.83
32.	7.78	24.53	1.65	102.99	2.87	9.11	7.57

	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}
33.	8.95	22.93	1.74	101.19	2.77	13.13	10.62
34.	7.48	22.73	1.71	70.82	2.74	8.87	5.16
35.	6.84	20.13	1.41	72.71	2.42	9.83	5.75
36.	7.49	22.47	1.45	93.24	2.65	7.88	6.09
37.	6.17	20.47	1.09	71.48	2.47	10.22	7.98
38.	7.49	21.67	1.57	103.29	2.58	7.24	5.75
39.	5.98	23.60	1.01	87.32	2.69	12.20	10.75
40.	6.47	18.53	0.95	71.64	2.27	10.37	7.43
41.	7.27	22.40	1.45	120.89	3.14	10.63	12.79
42.	7.37	19.07	1.40	98.14	2.70	6.58	6.55
43.	7.87	27.20	1.63	125.46	3.16	8.86	10.81
44.	7.97	21.07	1.36	108.39	2.24	10.05	10.96
45.	7.18	28.27	1.47	141.84	2.80	10.57	14.99
46.	5.65	17.87	0.90	84.15	2.47	7.43	6.26
47.	7.98	23.67	1.47	73.70	2.59	11.29	8.32
48.	7.27	19.40	1.36	85.46	2.39	10.74	9.16
F-values	5.59**	5.75**	7.52**	7.19**	4.91**	3.19**	4.49**

** Significant at 1% level

4.3 Estimated correlation matrices

The analysis of covariance was done for all the pairs of 15 characters. The phenotypic, genotypic and environment correlation coefficients were estimated and are given respectively in Tables 4.3.1, 4.3.2 and 4.3.3.

4.4 D^2 - Analysis

The genetic distances between the populations were estimated based on 15 variable dimension and the values are presented in Appendix III. Most of these values are significant at 5 per cent level.

The populations were arranged in increasing order of their relative distances from each other. The forty eight varieties were grouped into thirteen clusters. There were fifteen varieties in the first cluster, five in second, nine in third, seven in fourth and four varieties in the fifth cluster. The varieties Coc.774, Co.997, M.S.6847, Co.740, Coc.771, CoA.7601, Co.1305, Co.449 could not be grouped. The varieties belonging to different clusters and the cluster means are given in Table 4.4.1. Coc.771 had the maximum sugar yield (12.02 kg/plot), germination count (62.35), length of internode (15.81 cm), cane yield (156 kg/plot) and length of cane (3.12 m). Genotypes of Cluster II showed high juiciness (467.70 ml), more number of internodes (27.05) and sugar yield (12.01 kg/plot). The maximum

Table 4.3.1 Phenotypic correlation matrix

	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}
x_1	0.5247	-0.0460	-0.0757	-0.0731	0.4485	-0.1332	0.1825	0.0409	0.1683	0.0815	0.5389	0.3208	-0.0906	0.3107
x_2		0.0817	0.0338	-0.0388	0.8156	-0.2851	0.0842	-0.3557	-0.0113	-0.3420	0.5011	0.0812	-0.0004	0.3592
x_3			0.9665	0.7471	0.1152	-0.0486	-0.1122	-0.0206	0.2309	-0.0217	0.0605	0.0453	0.9429	0.6872
x_4				0.8719	0.0648	-0.0110	-0.1426	-0.0210	0.2349	0.0154	0.0414	0.0305	0.9915	0.7111
x_5					-0.0155	0.0218	-0.1740	-0.0726	0.1714	0.0476	-0.0240	-0.0285	0.8989	0.6200
x_6						-0.1765	0.0968	-0.3709	0.0428	-0.3243	0.5004	0.1327	0.0279	0.3762
x_7							-0.1326	0.2885	0.2931	0.3240	0.0917	0.0720	-0.0006	0.0645
x_8								0.0777	-0.3631	0.0701	0.1059	0.2698	-0.1512	-0.0569
x_9									0.2644	0.6549	0.3092	0.2350	-0.0210	0.1930
x_{10}										0.4609	0.5071	0.5658	0.2362	0.5172
x_{11}											0.3845	0.4765	0.0291	0.2782
x_{12}												0.5538	0.0212	0.7077
x_{13}													0.0256	0.4007
x_{14}														0.7015

Table 4.3.2 Genotypic correlation matrix

	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}
x_1	0.4980	-0.0401	-0.0657	-0.0329	0.4869	-0.3019	0.2179	0.0762	0.1413	0.1090	0.5867	0.3651	-0.1019	0.4193
x_2		0.0052	-0.0534	-0.1692	1.0007	-0.2965	0.1043	-0.4527	-0.1345	-0.5409	0.4019	-0.0690	-0.1225	0.2763
x_3			0.9928	0.9843	0.1858	0.1711	-0.3669	-0.0418	0.3644	-0.1177	0.0447	0.0472	0.9813	0.6125
x_4				1.0008	0.1508	0.1947	-0.3991	-0.0114	0.3711	-0.0780	0.0409	0.0333	0.9979	0.6151
x_5					0.0941	0.2075	-0.4932	-0.0415	0.3822	-0.1160	-0.0076	-0.0135	1.0007	0.5876
x_6						-0.3864	0.1491	-0.4706	-0.0223	-0.5049	0.5380	0.0960	0.1040	0.5229
x_7							-0.2460	0.7831	0.9259	0.9068	0.4960	0.4219	0.2304	0.5467
x_8								0.0505	-0.5655	0.1377	0.1145	0.4422	-0.4134	-0.1605
x_9									0.4150	1.0017	0.4696	0.4626	-0.0009	0.3505
x_{10}										0.5271	0.5591	0.5096	0.3861	0.6826
x_{11}											0.4917	0.6166	-0.0417	0.3337
x_{12}												0.6697	0.0190	0.8114
x_{13}													0.0455	0.5346
x_{14}														0.5993

Table 4.3.3 Environment correlation matrix

	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}
x_1	0.5744	-0.0529	-0.0874	-0.1067	0.3987	-0.0556	0.1274	-0.0095	0.2080	0.0363	0.4676	0.2622	-0.0815	0.1756
x_2		0.1868	0.1460	0.0549	0.5186	-0.3434	0.0446	-0.1888	0.2070	0.0703	0.7020	0.3236	0.1464	0.4929
x_3			0.9449	0.6605	0.0413	-0.1535	0.2187	0.0038	0.0780	0.1102	0.0851	0.0437	0.9140	0.7664
x_4				0.8424	-0.0213	-0.1023	0.1722	-0.0323	0.0880	0.1401	0.0447	0.0281	0.9872	0.8102
x_5					-0.0853	-0.0285	0.0393	-0.1035	0.0491	0.1912	-0.0420	-0.0409	0.8695	0.6921
x_6						-0.0825	0.0155	-0.2328	0.1353	-0.0315	0.4486	0.1800	-0.0449	0.1978
x_7							-0.0851	0.0358	-0.0471	-0.0133	-0.1711	-0.1147	-0.0976	-0.1840
x_8								0.1253	-0.0073	-0.0695	0.0886	-0.0024	0.1526	0.0989
x_9									0.0300	0.0108	0.0263	-0.0859	-0.0429	-0.0164
x_{10}										0.3419	0.4158	0.6482	0.0842	0.2958
x_{11}											0.1576	0.2506	0.1207	0.1982
x_{12}												0.3726	0.0256	0.5643
x_{13}													0.0068	0.2359
x_{14}														0.8050

Tables 4.4.1 Cluster means of various clones of sugarcane for genetic divergence

Cluster number	No. of clusters	Clones	Germination count	Shoot count	Brix %	Pol %	Purity percentage	No. of millable canes	Juiciness (ml)	Length of internode (cm)	Girth of cane (cm)	No. of internode	Weight of cane (kg)	Cane yield per plot (kg)	Length of cane (m)	C.C.S. percentage	Sugar yield per plot (kg)
I	15	T.67172, Co.7717, Co.419, Coc.779, Co.7219, Coc.777, Ic.225, Co.6304, S.99, Coc.773, Coc.772, Co.62198, Co.62101, Coc.778, S.77	45.18	83.13	16.27	13.70	83.51	70.18	441.40	12.44	7.71	23.20	1.52	107.33	2.68	9.25	8.19
II	5	Co.658, Co.62175, S.105, Co.6907, Co.995	49.82	92.0	18.0	15.89	87.78	84.47	467.70	11.43	7.83	27.05	1.55	128.45	2.82	10.98	12.01
III	9	F.1-2, Co.62174, S.87, KHS.3296, Coc.671, Co.7704, Co.785, CoM.7114, CoM.7125	40.02	70.7	18.60	16.51	88.48	60.44	442.11	11.90	8.0	24.30	1.58	96.63	2.76	11.44	9.57
IV	7	Co.6807, Co.1340, Co.527, S.33, Co.6806, B.37172, Co.527-M-10	38.67	96.48	17.05	14.31	83.21	80.67	422.33	11.95	6.69	21.64	1.16	85.32	2.51	9.64	7.37
V	4	Co.1307, CoA.7602, Coc.775, Co.453	44.68	85.0	15.57	12.91	82.00	73.75	428.75	14.99	7.68	20.67	1.47	105.53	2.90	8.64	7.74
VI	1	Coc.774	45.06	98.33	17.73	14.55	82.26	73.33	444.08	13.01	8.85	20.53	1.48	103.17	2.43	9.7	8.13
VII	1	Co.997	39.2	113.0	20.33	18.63	91.54	99.33	424.88	9.53	6.92	24.2	1.07	111.72	2.45	12.81	11.33
VIII	1	M.S.6847	39.81	70.33	12.12	8.53	66.05	54.33	438.89	14.32	9.46	22.8	1.99	108.55	2.98	5.18	4.29
IX	1	Co.740	52.78	84.0	19.19	15.50	80.56	70.67	433.88	11.9	7.4	25.00	1.47	107.4	2.78	10.23	8.49
X	1	Coc.771	62.35	107.33	17.10	14.31	82.55	95.33	378.43	15.81	8.11	21.07	1.57	156.0	3.12	9.63	12.02
XI	1	CoA.7601	25.31	46.67	15.26	12.99	84.95	45.0	461.81	12.85	7.48	22.73	1.71	70.82	2.74	8.87	5.16
XII	1	Co.1305	47.53	76.33	17.12	15.04	87.87	69.67	386.84	11.87	6.47	18.53	0.95	71.64	2.27	10.37	7.43
XIII	1	Co.449	46.91	115.67	14.57	11.44	78.49	102.33	355.04	12.84	5.65	17.87	0.90	84.15	2.47	7.43	6.26

shoot count (115.67) and number of millable canes (102.33) was found for Co.449. M.S.6847 had the maximum girth of cane (9.46 cm) and maximum weight of cane (1.99 kg). Co.997 had the highest values for brix (20.33%), pol (18.6%), purity (91.54%) and c.c.s. (12.81%). The intra and inter cluster D^2 - values are given in Table 4.4.2.

Table 4.4.2 Average intra and inter cluster D^2 - values

Clusters	I	II	III	IV	V
I	21.63	38.52	41.09	46.45	38.17
II		21.75	38.79	55.24	60.60
III			37.52	63.89	57.92
IV				24.89	67.35
V					22.87

The genetic divergence was maximum between clusters IV and V (67.35) followed by III and IV (63.89) and II and V (60.6). Selecting genotypes from such clusters as parents for hybridization will result in the development of superior clones with high productivity. Cluster V was quite divergent from clusters II, III and IV. Though the varieties Co.997 and Coc.771 were not included in any of the clusters, they can be used as parents during crossing programmes. Since Co.997 and Coc.771 had high genetic divergence and yield components. The magnitude of heterosis is expected to be high when crossing the genotypes Co.658, Co.62175, Co.6907, Co.995 of cluster II and S-87. Co.7704, Co.785, CoM.7114 of cluster III and Co.1307, CoA.7602 of cluster V.

4.5 Factor analysis

The clusters I, III and IV were taken for factor analysis and the method was applied for each cluster separately, since the other clusters contained less number of clones.

4.5.1 Cluster I

4.5.1.1 Correlation studies

The environment correlation matrix of cluster I is given in Table 4.5.1.1. The correlations were found to be between -0.2925 and 0.9982. The character x_1 was significantly correlated with x_2, x_6, x_{10}, x_{11}, x_{12} and x_{13}. x_2 was significantly correlated with all the characters except x_3, x_4, x_5, x_7, x_8 and x_{14}. The correlations of x_3 with x_4, x_5, x_8, x_9, x_{11}, x_{14} and x_{15} were significant. x_4 was significantly correlated with x_3, x_5, x_8, x_9, x_{11}, x_{14} and x_{15}. Significant correlations were found to exist for x_5 with x_3, x_4, x_8, x_9, x_{14} and x_{15}. x_6 was correlated with x_4, x_2, x_7 and x_{12} and x_7 with x_6 and x_{11}. The characters x_3, x_4, x_5, x_{13}, x_{14} and x_{15} were found to have significant

Table 4.5.1.1 Environment correlation matrix - Cluster I

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}
x_2	0.5963**													
x_3	-0.1092	-0.0372												
x_4	-0.1147	-0.0378	0.9799**											
x_5	-0.0784	0.0274	0.8386**	0.9216**										
x_6	0.3222**	0.3549**	-0.0619	-0.0695	0.0336									
x_7	-0.0476	-0.1277	0.0976	0.0693	-0.0996	-0.2148								
x_8	0.1530	0.0254	0.3295**	0.3505**	0.3474**	-0.0051	-0.1731							
x_9	-0.1425	-0.2292*	0.3952**	0.4025**	0.3024**	-0.2032	0.1432	0.1377						
x_{10}	0.4319**	0.3218**	0.0836	0.1190	0.1732	0.1483	-0.0895	0.2034	0.0428					
x_{11}	0.3491**	0.3242**	-0.2538*	-0.2324*	-0.1566	0.1944	-0.2170*	0.0572	-0.0929	0.2731**				
x_{12}	0.4264**	0.7357**	0.1864	0.1537	0.0977	0.3689**	-0.0570	0.0937	0.0243	0.4264**	0.3304**			
x_{13}	0.2276*	0.2589*	-0.0372	-0.0192	0.0279	0.1629	0.0710	-0.2925**	-0.1931	0.5507**	0.1161	0.1518		
x_{14}	-0.1148	-0.0374	0.9662**	0.9982**	0.9394**	-0.0712	0.0604	0.3536**	0.4012**	0.1287	-0.2241**	0.1426	-0.0134	
x_{15}	0.1420	0.3727**	0.8191**	0.8278**	0.7530**	0.1153	0.0066	0.2553*	0.3238**	0.3459**	0.0316	0.6474**	0.1371	0.8240

* Significant at 5% level ** Significant at 1% level

correlation with x_8 while x_9 was correlated with x_2, x_3, x_4, x_{15}, x_{14} and x_{15}. Character x_{10} was found to have significant correlation with x_1, x_2, x_{11}, x_{12}, x_{13} and x_{15}. x_{11} was correlated with all the characters except x_5, x_6, x_8, x_9, x_{13} and x_{15}. x_{12} had significant correlated with x_1, x_2, x_6, x_{10}, x_{11} and x_{15}. Correlation of x_{13} with x_1, x_2, x_8 and x_{10} alone were found to be significant. Character x_{14} was significantly correlated with the characters x_3, x_4, x_5, x_8, x_9, x_{11} and x_{14}. The correlations of x_{15} with x_1, x_6, x_7, x_{11} and x_{13} were not significant.

4.5.1.2 Principal factor analysis

Initially the eigen values and corresponding eigen vectors of the environment correlation matrix was found out by Jacobi's method. The latent roots of the matrix are given in Table 4.5.1.2.1. The matrix was found to be positive semidefinite. The first five latent roots of the matrix was greater then one and they altogether contributed about 79.23 per cent to the total variation.

PFA of the environment correlation matrix of order 15 was done with the squared multiple correlation coefficients (SMC) as first estimates of communalities and a five factor solution was extracted. The number of iterations needed for the convergence of communalities was twenty two, with a difference of five units in the third decimal place.

Table 4.5.1.2.1 Latent roots of the environment correlation matrix - Cluster I

Sl. No.	Latent roots	Per cent contribution to variance
1	5.0355	33.5696
2	3.3247	22.1644
3	1.4025	9.3588
4	1.0717	7.1446
5	1.0481	6.9962
6	0.8136	5.4239
7	0.6770	4.5133
8	0.5818	3.8876
9	0.4812	3.2080
10	0.2673	1.7820
11	0.1945	1.2966
12	0.0811	0.5407
13	0.0177	0.1180
14	0.0035	0.0233
15	0.0000	0.0000

The principal factor loadings in the 22nd iteration along with communalities in the 21st and 22nd iterations are given in Table 4.5.1.2.2. The loadings in the 22nd iteration was subjected to varimax rotation to have a more meaningful interpretation of the factors. The rotated loadings are presented in Table 4.5.1.2.3. The important characters associated with each factor were isolated in accordance with the procedure given by Harman (1967).

		Pol at 12th month
Factor	I	C.C.S. percentage
		Brix at 12th month
		Sugar yield per plot
		Purity percentage
Factor	II	Cane yield per plot
		Shoot count
		Germination count
		Number of millable canes per plot
		Weight of cane
Factor	III	Length of cane
		Number of internodes
Factor	IV	Length of internode
		Juiciness at 12th month
Factor	V	Girth of cane

Table 4.5.1.2.2 Principal factor solution in
correlation matrix - Cluster

Variable	Common factor coefficients			
	1	2	3	4
1	0.0023	0.6459	0.1021	0.1180
2	0.1108	0.7959	0.1975	-0.2014
3	0.9443	-0.1866	-0.0237	-0.0854
4	0.9825	-0.1911	-0.0486	-0.0131
5	0.9045	-0.1068	-0.0436	0.1470
6	0.0070	0.4402	0.1154	-0.0201
7	0.0210	-0.1669	-0.2132	-0.2904
8	0.3704	0.0148	0.3318	0.4501
9	0.3862	-0.2409	0.0305	0.0089
10	0.2531	0.6310	-0.3309	0.4345
11	-0.1284	0.4683	0.1471	0.1596
12	0.3306	0.7703	0.2719	-0.2967
13	0.0267	0.5030	-0.8364	-0.0405
14	0.9823	-0.1895	-0.0555	0.0111
15	0.9129	0.3033	0.0503	-0.1999

the 22nd iteration for the environment

I

s	Estimated communality		Origi- nal commu- nality (SMC)
5	21st ite- ration	22nd ite- ration	
-0.0282	0.4423	0.4424	0.5057
0.1272	0.7413	0.7414	0.7608
0.0004	0.9344	0.9344	1.0003
0.0464	1.0064	1.0064	1.0000
0.2714	0.9267	0.9267	0.9813
0.2844	0.2884	0.2884	0.4634
-0.2966	0.2458	0.2460	0.4343
-0.0543	0.4543	0.4531	0.4605
-0.3670	0.3430	0.3429	0.3364
-0.2825	0.8415	0.8402	0.6434
-0.0372	0.2842	0.2643	0.3026
-0.1805	0.8973	0.8971	0.9398
0.1272	0.9663	0.9712	0.6084
0.0604	1.0077	1.0077	1.0000
-0.0830	0.9751	0.9751	0.9757

Table 4.5.1.2.3 Rotated principal factor loadings for the environment correlation matrix - Cluster I

Variable	Common factor coefficients				
	1	2	3	4	5
1	0.0024	0.6044	-0.1994	0.1912	-0.0282
2	0.1108	0.8387	-0.0763	-0.0609	0.1272
3	0.9443	-0.1623	0.0785	-0.1008	0.0004
4	0.9825	-0.1895	0.0363	-0.0429	0.0464
5	0.9045	-0.1424	-0.0417	0.1135	0.2714
6	0.0070	0.4476	-0.0650	0.0535	0.2844
7	0.0210	-0.1677	-0.0302	-0.3586	-0.2966
8	0.3704	0.0362	0.1430	0.5396	-0.0543
9	0.3862	-0.2127	0.1175	0.0028	-0.3670
10	0.2531	0.3796	-0.6671	0.3274	-0.2825
11	-0.1284	0.4482	-0.1041	0.2339	-0.0372
12	0.3306	0.8597	0.0278	-0.1243	-0.1805
13	0.0267	0.1853	-0.9090	-0.3060	0.1272
14	0.9823	-0.1952	0.0222	-0.0228	0.0604
15	0.9129	0.3363	-0.0128	-0.1468	-0.0830
Proportionate variance accounted by each factor	0.3309	0.1797	0.0919	0.0521	0.0360

4.5.1.3 Maximum Likelihood factor analysis

From the principal factor analysis of the data it was hypothesized that a minimum of five factors would suffice to describe the dependence structure. The ML method was applied to extract the factors by Lawley's iterative scheme to get a more meaningful pattern. The sequence terminates either when a proper acceptable solution has been found from the point of view of goodness of fit or when the number of factors agree with the given upper bound.

ML estimation of factor loadings with a five factor model was tried. Forty five iterations were taken for a ± 0.005 convergence criterion. A test of significance of the model gave a χ^2 value of 36.21 which was significant. Since the degrees of freedom for this χ^2 was forty the normal test criterion $\sqrt{2\chi^2} - \sqrt{2n-1}$ was applied to test for the significance, where n is the degrees of freedom. So ML solution of factor loadings was tried with a six factor model. The goodness of fit of this model was tested at 0.01 level and found that six common factors are sufficient to explain the dependence structure. ($\chi^2 30 = 29.39$). Seventy four iterations were required for the convergence with a ± 0.005 convergence criterion. The initial estimates of factor loadings and unique variances obtained from the principal factor method of factor analysis are given in Table 4.5.1.3.1. The ML solutions in the 73rd and 74th iterations are summarised

Table 4.5.1.3.1 Initial estimates of factor loadings and corresponding unique variances for 6 factors of the environment correlation matrix - Cluster I

Variable	Factor loadings						Unique variance
	1	2	3	4	5	6	
1	-0.1148	0.4889	0.0266	-0.1550	-0.0039	0.0031	0.8305
2	-0.0377	0.7765	0.0774	-0.2368	-0.0120	0.2539	0.2690
3	0.9761	-0.0030	-0.0019	0.0015	-0.2177	0.1084	0.0001
4	0.9998	-0.0003	-0.0003	0.0002	-0.0186	-0.0237	0.0002
5	0.9274	0.0156	0.0129	-0.1914	-0.3048	0.3871	0.0001
6	-0.0701	0.3953	0.0007	-0.4944	0.0287	-0.1226	0.4215
7	0.0666	-0.0767	-0.1262	0.5565	-0.1452	-0.3652	0.4904
8	0.3515	0.0330	0.3071	-0.0962	0.0623	0.5669	0.5535
9	0.4022	-0.0862	-0.1680	0.3351	-0.0104	0.2858	0.3915
10	0.1220	-0.5525	0.3376	-0.0120	0.1670	-0.0358	0.4634
11	-0.2310	0.4026	0.0775	-0.0652	0.1405	0.0423	0.2472
12	0.1502	0.9327	0.2514	0.0084	-0.1677	-0.2216	0.0001
13	0.0174	0.4215	0.8887	0.0079	0.0907	-0.0810	0.9826
14	0.9991	0.0004	0.0005	-0.0005	0.0415	0.1445	0.0001
15	0.8269	0.5309	0.0773	0.0395	0.0465	-0.3772	0.0001

in Tables 4.5.1.3.2 and 4.5.1.3.3 respectively. The varimax rotated loadings are presented in Table 4.5.1.3.4. The residual correlation matrix after removal of six factors is given in Table 4.5.1.3.5. The characters more related with each factor are given below.

	Pol at 12th month
Factor I	C.C.S. percentage
	Brix at 12th month
	Purity percentage
	Sugar yield per plot
Factor II	Cane yield per plot
	Shoot count
	Germination count
	No. of millable canes per plot
Factor III	Length of cane
	Number of internodes
Factor IV	Juiciness at 12th month
	Girth of cane
Factor V	Length of internode
Factor VI	Weight of cane

In both PFA and ML methods, factor I was found to be highly correlated with pol at 12th month, C.C.S. percentage,

Table 4.5.1.3.2 Maximum likelihood estimates of factor loadings and unique variances in the 73rd iteration - Cluster I

Variable	Factor loadings						Unique variance
	1	2	3	4	5	6	
1	-0.1147	0.4351	0.1776	-0.1397	-0.0791	0.0509	0.7376
2	-0.0378	0.7422	0.1712	-0.2260	-0.1086	-0.0581	0.3523
3	0.9770	-0.0256	-0.0029	0.0031	-0.2087	0.0008	0.0012
4	0.9999	-0.0046	-0.0002	0.0002	-0.0137	0.0001	0.0001
5	0.9261	0.0539	0.0196	-0.2010	0.3009	-0.0348	0.0069
6	-0.0700	0.4752	0.1161	-0.4992	-0.0943	0.0644	0.5786
7	0.0673	-0.1152	0.0884	0.5309	-0.1250	0.2743	0.6017
8	0.3513	0.0851	-0.3105	-0.0951	0.0510	0.2179	0.7138
9	0.4023	-0.0330	-0.1820	0.3290	0.0017	-0.0475	0.6934
10	0.1213	0.4451	0.4921	-0.0010	0.0971	0.1009	0.5254
11	-0.2306	0.4029	0.0597	-0.0400	0.0752	-0.1290	0.7570
12	0.1509	0.9335	0.0533	0.0050	-0.3148	-0.0359	0.0026
13	-0.0179	0.1371	0.9756	0.0051	0.0642	0.0314	0.0239
14	0.9990	0.0039	0.0008	-0.0039	0.0432	-0.0007	0.0001
15	0.8271	0.5012	0.1002	0.0572	-0.1250	-0.1639	0.0089

Table 4.5.1.3.3 Maximum likelihood estimates of factor loadings and unique variances in the 74th iteration - Cluster I

Variable	Factor loadings						Unique variance
	1	2	3	4	5	6	
1	-0.1147	0.4389	0.1768	-0.1377	-0.0745	0.0525	0.7356
2	-0.0378	0.7460	0.1701	-0.2289	-0.1051	-0.0563	0.3465
3	0.9772	-0.0287	-0.0023	0.0033	-0.2107	0.0001	0.0001
4	0.9999	-0.0017	-0.0002	0.0002	-0.0134	0.0001	0.0001
5	0.9258	0.0550	0.0195	-0.2035	0.3026	-0.0312	0.0054
6	-0.0700	0.4762	0.1172	-0.4997	-0.0912	0.0684	0.5772
7	0.0674	-0.1111	0.0875	0.5329	-0.1252	0.2711	0.6023
8	0.3513	0.0870	-0.3104	-0.0936	0.0545	0.2196	0.7127
9	0.4023	-0.0337	-0.1831	0.3271	0.0007	-0.0495	0.6941
10	0.1212	0.4459	0.4907	-0.0006	0.0988	0.1017	0.5256
11	-0.2307	0.4061	0.0582	-0.0434	0.0778	-0.1279	0.7542
12	0.1512	0.9375	0.0302	0.0084	-0.3117	-0.0398	0.0001
13	-0.0179	0.1350	0.9795	0.0022	0.0644	0.0310	0.0170
14	0.9989	0.0063	0.0007	-0.0010	0.0462	-0.0005	0.0002
15	0.8272	0.5043	0.0987	0.0549	-0.1205	-0.1655	0.0068

Table 4.5.1.3.4 Rotated maximum likelihood estimates of factor loadings – Cluster I

Variable	Factor loadings					
	1	2	3	4	5	6
1	-0.1147	0.4389	0.1768	-0.0943	-0.0595	0.1218
2	-0.0378	0.7460	0.1701	-0.2190	-0.1156	0.0727
3	0.9772	-0.0287	-0.0023	0.0252	-0.2044	0.0448
4	0.9999	-0.0017	-0.0002	0.0016	-0.0129	0.0030
5	0.9258	0.0551	0.0195	-0.2284	0.2860	-0.0055
6	-0.0670	0.4762	0.1172	-0.4123	-0.0718	0.2959
7	0.0674	-0.1111	0.0875	0.6074	-0.0556	0.0340
8	0.3513	0.0870	-0.3104	0.0020	0.1062	0.2206
9	0.4023	-0.0337	-0.1831	0.2741	-0.0114	-0.1849
10	0.1212	0.4459	0.4907	0.0318	0.1206	0.0675
11	-0.2307	0.4061	0.0582	-0.1010	0.0445	-0.1102
12	0.1512	0.9375	0.0502	0.0236	-0.3120	0.0297
13	-0.0179	0.1350	0.9795	0.0083	0.0670	0.0120
14	0.9989	0.0063	0.0007	-0.0059	0.0447	-0.0101
15	0.8272	0.5043	0.0987	-0.0073	-0.1571	-0.1421
Contribution of each factor	4.8943	2.4308	1.4277	0.7672	0.3135	0.1924
Proportionate variance accounted by each factor	0.3262	0.1621	0.0952	0.0490	0.0204	0.0155

Table 4.5.1.3.5 Residual matrix after removal of six factors from environment correlation matrix - Cluster I

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}
x_2	0.1981													
x_3	0.0007	0.0001												
x_4	-0.0002	0.0000	0.0000											
x_5	-0.0036	0.0014	0.0000	0.0000										
x_6	0.0492	-0.0684	0.0000	0.0000	0.0029									
x_7	0.0432	0.0669	0.0005	-0.0001	-0.0028	0.0578								
x_8	0.1896	0.0232	-0.0002	0.0000	-0.0053	-0.0338	-0.1628							
x_9	-0.0015	-0.0856	-0.0003	0.0000	0.0002	0.0260	-0.0324	-0.0160						
x_{10}	0.1653	-0.0737	0.0000	0.0000	0.0000	-0.0667	-0.1059	0.2467	0.1041					
x_{11}	0.1406	-0.0063	0.0000	0.0000	-0.0029	0.0129	-0.0939	0.1407	0.0320	0.0967				
x_{12}	0.0034	0.0004	0.0000	0.0000	-0.0001	-0.0006	-0.0001	0.0012	-0.0002	0.0003	0.0011			
x_{13}	-0.0035	0.0000	-0.0001	0.0000	0.0000	0.0009	-0.0002	-0.0040	-0.0012	0.0025	-0.0008	0.0000		
x_{14}	0.0002	0.0000	0.0000	0.0000	0.0000	0.0001	-0.0003	-0.0008	0.0000	0.0000	0.0000	0.0000	0.0000	
x_{15}	0.0053	0.0015	0.0000	0.0000	-0.0001	-0.0003	-0.0013	-0.0004	0.0000	0.0011	0.0024	0.0000	0.0000	0.0000

brix at 12th month, purity percentage and sugar yield per plot. These characters are associated with the quality of the crop and hence can be named as quality factor. Cane yield per plot, shoot count, germination count and number of millable canes per plot were found to be highly correlated with factor II in ML method while weight of cane also contributed for factor II in PF method. Cane yield and its related characters come under this factor. Weight of cane formed an independent factor, factor VI in ML method. Length of cane and number of internodes related to growth of the crop come under factor III as identified by PFA and ML method. So this factor may be named as growth factor. While length of internode and juiciness at 12th month are more contributing to factor IV in PFA girth of cane and juiciness at 12th month contribute to factor IV in ML method. Girth of cane remained independent in factor V in PFA. In ML method length of internode form an independent factor in factor V.

The five common factors in PFA accounted for 69.06 percentage of the variation in the dependence structure while 66.84 percentage variation was explained by the six factor model in ML solution. The proportion of variation accounted by factor I where the characters contributed for this factor being same, accounted about 33.79 per cent in PFA and 32.62 per cent in ML solution. The contribution of the second factor was 17.97 per cent and 16.21 per cent respectively in PFA and ML solutions. While the proportionate variance accounted by factor III in ML was 9.52 per cent it was 9.19 per cent in PFA. The contribution of remaining factors were 8.49 per cent in ML and 8.81 per cent in PF solution.

4.5.2 Cluster III

4.5.2.1 Correlation studies

The environment correlation coefficients were found to lie between -0.4948 and 0.9742 (Table 4.5.2.1). The character x_1 was significantly correlated with x_2, x_6, x_9, x_{12}, x_{13} and x_{15}. x_2 was found to be significantly correlated with all the characters except x_8, x_{10} and x_{11} while x_3 had significant correlation with all the characters except x_1, x_{10}, x_{11} and x_{12}. Correlation of x_4 with the characters except x_1, x_7, x_{10}, x_{11} and x_{12} were found to be significant. Significant correlations were found to exist for x_5 with x_2, x_3, x_4, x_5, x_6, x_7, x_9, x_{15} and x_{14}. x_6 was correlated with all except x_7, x_{10} and x_{11}. The characters x_2, x_3, x_5, x_8, x_{13}, x_{14} and x_{15} were found to have significant correlation with x_7 while x_8 was significantly correlated with x_3, x_4, x_6, x_7 and x_{13}. x_9 was significantly correlated with all except x_7, x_8, x_{10}, x_{11} and x_{12}. Correlation of x_{10} with x_{11}, x_{12} and x_{13} alone were found to be

Table 4.5.2.1 Environment correlation matrix - Cluster III

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}
x_2	0.7376**													
x_3	0.0046	0.3403**												
x_4	0.0508	0.3815**	0.9206**											
x_5	0.2023	0.4707**	0.4378**	0.7021**										
x_6	0.4962**	0.6686**	0.4754**	0.5636**	0.4804**									
x_7	-0.1221	-0.3457**	-0.2348*	-0.1853	-0.2263*	-0.0775								
x_8	-0.1184	-0.0155	0.3630**	0.2579*	-0.1650	0.2381*	-0.4756**							
x_9	-0.2545*	-0.3308**	-0.4225**	-0.4739**	-0.4288**	-0.4948**	-0.1183	-0.0299						
x_{10}	0.1286	0.1660	0.0576	0.0734	0.0270	0.0750	-0.0545	-0.0639	-0.1750					
x_{11}	-0.1655	0.0202	0.1610	0.1091	-0.1176	-0.0070	0.0738	-0.0134	0.1732	0.6450**				
x_{12}	0.5388**	0.6868**	0.0822	0.0686	-0.0141	0.5677**	-0.1482	0.1459	0.0343	0.2759**	0.1500			
x_{13}	0.5258**	0.6729**	0.4023**	0.4850**	0.4612**	0.5248**	-0.3152**	0.3098**	-0.4537**	0.3611**	0.0208	0.5266**		
x_{14}	0.0542	0.4366**	0.9111**	0.9742**	0.7618**	0.5176**	-0.2767**	0.1682	-0.4663**	0.0571	0.0831	0.0507	0.4704**	
x_{15}	0.4190**	0.7644**	0.5984**	0.6375**	0.5072**	0.7994**	-0.2174*	0.0886	-0.2262*	0.1918	0.1280	0.7423**	0.5704**	0.6533**

* Significant at 5% level ** Significant at 1% level

significant. x_{11} is significantly correlated only with x_{10}. Significant correlations were found to exist for x_{12} with x_1, x_2, x_6, x_{10}, x_{13} and x_{15}. x_{13} is significantly correlated with all except x_{11}. Correlation of x_{14} with all except x_1, x_8, x_{10} and x_{11} were significant. The correlations of x_{15} with x_8, x_{10} and x_{11} were not significant.

4.5.2.2 Principal factor analysis

The environment correlation matrix was found to be positive definite. The eigen values and the corresponding eigen vectors of the matrix was found out. The latent roots along with contribution of each to the total variation are given in Table 4.5.2.2.1. First four latent roots of the matrix was greater then one and they altogether contributed 78.03 per cent to the total variation.

Using the principal factor analysis to the environment correlation matrix a four factor model was fitted with squared multiple correlation coefficient as estimates communality. Fifty two iterations were taken for the convergence of communalities with a five unit difference in the third decimal place. The estimates of loadings in the 52nd iteration along with communalities in the 51st and 52nd iterations are given in Table 4.5.2.2.2. Varimax rotation of loadings was applied and the results are given in Table 4.5.2.2.3. The characters which are more correlated with

Table 4.5.2.2.1 Latent roots of the environment correlation matrix - Cluster III

Sl. No.	Latent roots	Per cent contribution to variance
1	6.0411	40.4376
2	2.3828	15.9588
3	1.6671	11.1682
4	1.5642	10.4703
5	0.9865	6.6034
6	0.9473	6.3410
7	0.4015	2.6875
8	0.3541	2.3703
9	0.2917	1.9526
10	0.2031	1.3595
11	0.0478	1.3200
12	0.0394	0.2673
13	0.0115	0.0770
14	0.0011	0.0074
15	0.0001	0.0007

Table 4.5.2.2.2 Principal factor solution in the 52nd iteration for the environment correlation matrix - Cluster III

Variable	Common factor coefficients				Estimated communality		Original communality (SMC)
	1	2	3	4	51st iteration	52nd iteration	
1	0.4532	0.5946	0.2495	-0.0114	0.6215	0.6213	0.8944
2	0.3891	0.4828	0.1252	0.0143	0.8716	0.8716	0.9676
3	0.7427	-0.4855	-0.1384	-0.0075	0.8067	0.8065	0.9997
4	0.8401	-0.5239	-0.0396	0.1192	0.9959	0.9960	0.9942
5	0.6548	-0.2313	0.2605	0.2843	0.6312	0.6309	0.9994
6	0.3809	0.2777	0.0907	-0.0510	0.6522	0.6522	0.9944
7	-0.2969	0.0256	0.0433	0.3318	0.2011	0.2007	0.9292
8	0.2603	-0.2899	-0.3318	-0.4132	1.0000	1.0000	0.9507
9	-0.4819	0.1331	-0.1934	-0.1313	0.3047	0.3046	0.9822
10	0.2058	0.2318	-0.4721	0.1973	0.3583	0.3579	0.9544
11	0.0806	0.0806	-0.0648	0.3676	1.0000	1.0000	0.9599
12	0.5062	0.6920	-0.1574	-0.1873	0.7950	0.7948	0.9961
13	0.7311	0.2174	-0.0181	-0.1264	0.5980	0.5980	0.9729
14	0.8491	-0.5221	0.0095	0.1708	1.0000	1.0001	0.9998
15	0.8613	0.2207	-0.0504	0.0391	0.7945	0.7945	0.9973

Table 4.5.2.2.3 Rotated principal factor loadings for the environment correlation matrix - Cluster III

Variable	Common factor coefficients			
	1	2	3	4
1	0.4472	0.6470	-0.0070	0.0509
2	0.3497	0.5292	-0.1607	-0.0606
3	0.7037	-0.4426	-0.1372	-0.3110
4	0.8468	-0.4810	-0.0892	-0.1991
5	0.7542	-0.1888	0.0991	0.1293
6	0.3410	0.2472	-0.0948	-0.1815
7	-0.2080	-0.0606	-0.0003	0.3922
8	-0.0188	-0.0889	0.0041	-0.4750
9	-0.5376	0.0856	-0.0857	-0.0308
10	0.1091	0.0864	-0.5801	0.0448
11	-0.1045	-0.2300	-0.1028	0.0457
12	0.3614	0.6961	-0.3654	-0.2148
13	0.6508	0.2758	-0.1747	-0.2606
14	0.8769	-0.4801	-0.0604	-0.1406
15	0.7995	0.2421	-0.2732	-0.1489
Proportionate variance accounted by each factor	0.3607	0.1434	0.1249	0.1243

these four factors are given below.

Factor	I	C.C.S. percentage
		pol at 12th month
		Sugar yield per plot
		Purity percentage
		Brix at 12th month
Factor	II	Cane yield per plot
		Germination count
		Shoot count
		Number of millable canes per plot
Factor	III	Number of internodes
Factor	IV	Length of internode
		Juiciness at 12th month

4.5.2.3 Maximum-Likelihood factor analysis

The environment correlation matrix was subjected to ML method of factor extraction under the hypothesis that a four factor-model will suffice to explain the dependence structure. Twenty nine iterations were taken for a ± 0.005 convergence criterion. A test of significance of the factor model showed that four common factors are not sufficient to explain the dependence structure ($\chi^2_{51} = 104.27$). The ML method was then tried for a five factor model which again found to be inadequate to explain the dependence structure

(χ^2_{40} = 72.78). Fifty two iterations were taken for the convergence. A six factor model gave the goodness of fit statistic as χ^2_{30} = 59.14 which was significant. ML solution of factor loadings with six factor model was found to be adequate to explain the dependence structure. Two hundred and seven iterations were taken for the convergence with a ±0.005 convergence criterion. The initial estimates of factor loadings and unique variances obtained from the principal factor method of factor analysis are given in Table 4.5.2.3.1. The ML solutions in the 206th and 207th iterations are summarised in Tables 4.5.2.3.2 and 4.5.2.3.3 respectively. The varimax rotated loadings are presented in Table 4.5.2.3.4. The residual matrix after removal of six factors is given in Table 4.5.2.3.5. The characters dominating the factors are listed below.

Factor I
- C.C.S. percentage
- pol at 12th month
- Brix at 12th month
- Purity percentage
- Sugar yield per plot

Factor II
- Cane yield per plot
- Shoot count
- Germination count
- Number of millable canes per plot

Table 4.5.2.3.1 · Initial estimates of factor loadings and corresponding unique variances for 6 factors – Cluster III

Variable	Factor loadings						Unique variance
	1	2	3	4	5	6	
1	−0.0703	0.5677	0.0170	−0.1867	−0.4642	0.3238	0.3173
2	−0.4456	0.6884	0.0133	−0.1788	−0.2107	0.1201	0.2366
3	0.9203	−0.0143	−0.0119	0.3881	−0.0411	0.0694	0.0001
4	0.9771	−0.0037	−0.0052	0.0513	−0.0395	−9.0405	0.0394
5	0.7515	−0.0141	−0.0116	0.6579	0.0379	−0.1059	0.0002
6	−0.5362	0.5617	−0.3660	−0.0771	−0.3719	−0.4484	0.0001
7	0.2743	−0.1303	0.2800	−0.0323	−0.0168	0.3100	0.7320
8	0.1846	0.0789	0.2400	−0.4766	0.2891	−0.3095	0.4956
9	0.4767	−0.0475	−0.0657	0.0858	0.4516	−0.1341	0.5369
10	−0.0580	0.2704	0.3063	−0.0271	0.0325	−0.0789	0.8217
11	0.0805	0.1131	0.1105	−0.2571	0.2386	0.1066	0.8341
12	−0.0598	0.9902	0.0210	0.1093	−0.0201	−0.0424	0.0013
13	−0.4820	0.5143	0.4877	0.1214	−0.4480	0.3903	0.0001
14	0.9992	0.0080	0.0097	−0.0187	0.0307	−0.4869	0.0001
15	0.6603	0.7129	0.1963	0.0007	−0.0432	−0.2080	0.0002

Table 4.5.2.3.2 Maximum likelihood estimates of factor loadings and unique variances in the 206th iteration - Cluster III

Variable	Factor loadings						Unique variance
	1	2	3	4	5	6	
1	-0.2060	0.4489	0.2100	-0.1848	-0.3850	0.1687	0.5011
2	-0.5222	0.6148	0.0184	-0.0301	-0.2586	-0.0792	0.2749
3	0.6209	-0.3243	-0.0108	0.0239	-0.2110	0.0171	0.4638
4	0.8353	-0.2121	-0.0039	0.0241	-0.4712	0.0206	0.0342
5	0.9734	-0.1342	-0.0126	0.0112	0.1684	-0.0097	0.0057
6	-0.5674	0.3898	-0.3916	-0.2867	-0.0891	0.3005	0.1923
7	-0.2550	-0.1425	0.0791	0.1214	-0.0285	0.2845	0.8119
8	0.0499	0.3381	0.2112	-0.2807	0.3918	-0.2314	0.5527
9	0.4801	-0.0447	-0.3199	0.4108	0.1631	-0.1315	0.4525
10	-0.0529	0.2743	0.5955	-0.3741	0.0648	0.0254	0.4225
11	0.0654	0.2316	0.5833	-0.2093	0.2035	0.4239	0.3369
12	-0.0529	0.9322	0.0347	0.0786	-0.3531	-0.0014	0.0002
13	-0.5241	0.4185	0.0949	0.3748	-0.5250	0.5322	0.0001
14	0.8815	0.1614	0.0409	-0.0695	0.4331	-0.0048	0.0028
15	0.6177	0.3452	0.0314	0.1413	-0.0384	-0.1571	0.4522

Table 4.5.2.3.3 Maximum likelihood estimates of factor loadings and unique variances in the 207th iteration - Cluster III

Variable	Factor loadings						Unique variance
	1	2	3	4	5	6	
1	-0.2070	0.4450	0.2120	-0.1847	-0.3830	0.1659	0.5059
2	-0.5225	0.6129	0.0203	-0.0264	-0.2616	-0.0762	0.2760
3	0.6195	-0.3264	-0.0107	0.0192	-0.2131	0.0140	0.0005
4	0.8323	-0.2104	-0.0038	0.0201	-0.4732	0.0176	0.0384
5	0.9756	-0.1360	-0.0127	0.0091	0.1714	-0.0057	0.0005
6	-0.5653	0.3861	-0.3936	-0.2890	-0.0850	0.2986	0.1965
7	0.2569	-0.1443	0.0763	0.1244	-0.0315	0.2813	0.8118
8	0.0452	0.3413	0.2083	-0.2776	0.3949	-0.2275	0.5533
9	0.4829	-0.0438	-0.3239	0.4138	0.1591	-0.1306	0.4464
10	-0.0496	0.2752	0.5986	-0.3778	0.0688	0.0233	0.4155
11	0.0623	0.2356	0.5863	-0.2074	0.2005	0.4278	0.3306
12	-0.0513	0.9291	0.0366	0.0756	-0.3560	-0.0003	0.0004
13	-0.5274	0.4155	0.0905	0.3770	-0.5290	0.5251	0.0001
14	0.8794	0.1655	0.0445	-0.0734	0.4365	-0.0027	0.0014
15	0.6147	0.3404	0.0281	0.1454	-0.0362	-0.1543	0.4592

Table 4.5.2.3.4 Rotated maximum

Cluster III

Variable	1	2
1	0.0755	0.5894
2	-0.2768	0.7562
3	0.8853	-0.1137
4	0.9201	-0.1600
5	0.6601	-0.1587
6	-0.5125	0.4849
7	0.1467	-0.1539
8	-0.1183	0.1118
9	0.4077	-0.0662
10	0.0249	0.2827
11	-0.1516	0.1297
12	0.0941	0.9788
13	-0.1791	0.4930
14	0.9333	0.1450
15	0.5603	0.3031
Contribution of each factor	4.4031	3.0614
Proportionate variance accounted by each factor	0.2575	0.2377

likelihood estimates of factor loadings -

Factor loadings			
3	4	5	6
+0.1671	-0.3362	-0.0105	0.0040
0.0581	-0.2080	0.0854	-0.1472
-0.0153	0.0121	-0.3618	0.2677
-0.0060	0.1753	-0.0810	0.2281
-0.0661	0.4564	-0.5518	-0.1475
-0.1517	-0.3114	-0.1179	0.3638
0.0781	0.0652	-0.0296	0.3631
0.0861	-0.1395	0.6210	-0.0872
-0.1515	0.5636	0.2008	-0.0461
0.6819	-0.1289	0.0638	0.1356
0.5422	-0.2226	0.1165	0.4687
0.0003	0.1375	-0.0979	-0.0638
0.0678	0.1029	-0.6645	0.4672
0.0240	-0.1440	0.0226	-0.2910
0.0675	0.0524	-0.0325	-0.0760
1.5467	1.3436	0.7647	0.7946
0.0972	0.0871	0.0631	0.0516

Table 4.5.2.3.5 Residual matrix after removal of six factors from the environment correlation matrix - Cluster III

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}
X_2	0.2346													
X_3	0.0032	0.0024												
X_4	-0.0339	-0.0576	-0.0005											
X_5	0.0006	0.0004	0.0000	0.0000										
X_6	-0.0015	-0.0011	0.0000	0.0004	0.0000									
X_7	-0.0092	-0.1147	-0.0006	0.0666	-0.0002	-0.0004								
X_8	-0.2429	-0.1266	-0.0020	0.0361	-0.0004	0.0029	-0.4120							
X_9	-0.0292	-0.0658	0.0018	0.0173	0.0005	-0.0016	-0.2142	0.2381						
X_{10}	0.0305	-0.0426	0.0001	-0.0035	0.0000	-0.0007	-0.0121	-0.1033	-0.1906					
X_{11}	0.0003	0.0012	0.0000	-0.0001	0.0000	0.0000	-0.0004	0.0000	0.0026	0.0002				
X_{12}	0.0001	0.0002	0.0000	0.0002	0.0000	0.0000	0.0000	0.0004	0.0000	-0.0002	0.0000			
X_{13}	-0.0014	-0.0013	0.0000	0.0001	0.0000	0.0000	0.0010	0.0011	-0.0004	0.0001	0.0000	0.0000		
X_{14}	-0.0117	-0.0075	0.0000	0.0023	0.0000	0.0000	0.0001	0.0081	-0.0094	-0.0009	0.0000	0.0000	0.0000	
X_{15}	-0.0051	-0.0005	0.0008	-0.0124	0.0002	-0.0003	0.0188	-0.0523	0.0260	0.0233	-0.0001	0.0000	-0.0002	-0.0030

Factor III	Number of internodes
Factor IV	Girth of cane
Factor V	Length of cane
	Length of internode
Factor VI	Weight of cane
	Juiciness at 12th month

The characters C.C.S. percentage, pol at 12th month, brix at 12th month, purity percentage and sugar yield per plot were found to be highly correlated with Factor I in both PFA and ML methods. The characters are related to quality aspects of the crop and hence named as the quality factor. The second factor is associated with cane yield per plot, shoot count, germination count and number of millable canes per plot in both methods. The third factor is characterised by number of internodes in PFA and ML methods. Length of internode and juiciness at 12th month are more contributing to factor IV in PFA while girth of cane alone contribute factor IV in ML method. In ML method length of cane and length of internode formed fifth factor and 6th factor is highly correlated with weight of cane and juiciness at 12th month.

The four common factors is PFA accounted for about 75.33 percentage of variation in the dependence structure while 79.44 percentage variation was explained by the six factor model in ML solution. Factor I accounted 36.07 percentage of variation in PFA and 25.75 in ML method. Proportionate variance accounted by the second factor was 14.34 per cent in PFA and 23.77 per cent in ML method. The contribution of the third factor was 12.49 per cent and 9.72 per cent respectively in PFA and ML method. Contribution of remaining factors were 12.43 per cent in PFA and 20.20 per cent in ML method.

4.5.3 Cluster IV

4.5.3.1 Correlation studies

The environment correlation matrix is given in Table 4.5.3.1 and the correlation coefficients were found to lie between -0.8199 and 0.9948. Character x_1 was significantly correlated with all the characters except x_3. The correlation of x_2 with all the characters were significant except for x_4, x_5, x_7 and x_{14}. Correlation of x_3 with the characters except x_1, x_8, x_9, x_{10}, x_{12} and x_{13} were found to be significant. Significant correlations were found to exist for x_4 with x_1, x_3, x_5, x_7, x_9, x_{11}, x_{14} and x_{15}. x_5 had correlation with all except x_2, x_8 and x_{10}. Correlation of x_6 with x_4, x_7, x_8, x_{11} and x_{14} were found to be non-significant. Characters x_2, x_9, x_{10}, x_{12} and x_{13} were found to have non-significant correlation with x_7 while characters x_1, x_2, x_4, x_5, x_6, x_{11}, x_{12}, x_{13} and x_{15} have significant correlation with x_8. x_9 was correlated with all except x_3,

Table 4.5.3.1 Environment correlation matrix - Cluster IV

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}
x_2	0.6376**													
x_3	-0.1236	0.3404**												
x_4	-0.2258*	0.1678	0.9433**											
x_5	-0.3795**	-0.1999	0.5667**	0.7850**										
x_6	0.6611**	0.8525**	0.2721**	0.0878	-0.3207**									
x_7	0.2282*	-0.1246	-0.6430**	-0.6373**	-0.4044**	0.0647								
x_8	0.4042**	0.3195**	0.0806	0.0196	0.0334	0.1948	0.2545*							
x_9	0.4464**	0.2351*	-0.1989	-0.3652**	-0.5818**	0.2934**	0.1978	-0.1151						
x_{10}	0.2811**	0.5392**	0.0271	-0.0166	-0.1986	0.4494**	-0.0430	0.0471	-0.1035					
x_{11}	-0.2600*	0.1023	0.4992**	0.6044**	0.6387**	0.0567	-0.2554*	0.2452*	-0.8199**	0.4015**				
x_{12}	0.6800**	0.7623**	0.0820	-0.1204	-0.4631**	0.7343**	0.1364	0.4444**	0.2875**	0.6688**	0.0619			
x_{13}	0.4582**	0.6282**	-0.1171	-0.1677	-0.2628*	0.3236**	-0.0605	0.4366**	-0.0997	0.6957**	0.1686	0.3460**		
x_{14}	-0.2224*	0.1465	0.9084**	0.9948**	0.8246**	0.0548	-0.6348**	0.0189	-0.4033**	-0.0013	0.6280**	-0.1472	-0.1389	
x_{15}	0.2509*	0.5604**	0.8808**	0.8129**	0.4606**	0.4786**	-0.4982**	0.3100**	-0.1318	0.2605*	0.5166**	0.4419**	0.1813	0.8039**

** Significant at 5% level * Significant at 1% level

x_7, x_8, x_9 and x_{14}. x_{10} was not correlated with x_3, x_4, x_5, x_7, x_8, x_9 and x_{14} and significantly correlated with other characters. x_{11} had significant correlation with all the characters except x_2, x_6, x_{12} and x_{13} and x_{12} with all except x_3, x_4, x_7, x_{11} and x_{14}. x_{13} was significantly correlated with x_1, x_2, x_5, x_6, x_8, x_{10}, x_{12} and x_{13}. Correlation of x_{14} with x_1, x_3, x_4, x_5, x_7, x_9, x_{11} and x_{15} were found to be significant. x_{15} was significantly correlated with all except x_9 and x_{13}.

4.5.3.2 Principal factor analysis

The environment correlation matrix was found to be positive semi definite. The eigen values and the corresponding eigen vectors of the matrix was determined. The eigen values along with contribution of each latent root to the total variation are given in Table 4.5.3.2.1. First four latent roots of the matrix was greater than one and they altogether contributed 87.37 per cent to the total variation.

A four factor model was extracted using principal factor analysis with squared multiple correlation coefficient as initial estimate of communality. The number of iterations needed for the convergence of communalities was six, with a difference of five units in the third decimal place. The principal factor loadings in the 6th iteration

Table 4.5.3.2.1 Latent roots of the environment correlation matrix - Cluster IV

Sl. No.	Latent roots	Per cent contribution to variance
1	5.4288	36.1920
2	4.6725	31.1500
3	1.8167	12.1113
4	1.1881	7.9207
5	0.7128	4.7520
6	0.4009	2.6727
7	0.3359	2.2393
8	0.2319	1.5460
9	0.0994	0.6627
10	0.0728	0.4853
11	0.0315	0.2100
12	0.0087	0.0580
13	0.0000	0.0000
14	0.0000	0.0000
15	0.0000	0.0000

along with communalities in the 5th and 6th iterations are given in Table 4.5.3.2.2. Factors in the 6th iteration was subjected to varimax rotation. The rotated loadings are presented in Table 4.5.3.2.3. The characters which are highly correlated with the four factors are given below.

	pol at 12th month
	Brix at 12th month
Factor I	C.C.S. percentage
	Sugar yield per plot
	Purity percentage
	Cane yield per plot
Factor II	Shoot count
	Number of millable canes per plot
	Germination count
Factor III	Number of internodes
	Length of cane
Factor IV	Length of internode
	Juiciness at 12th month

4.5.3.3 Maximum Likelihood factor analysis

ML estimation of factor loadings with a four factor model was done. Twenty eight iterations were taken for a ± 0.005 convergence criterion. A test of significance of the model gave a χ^2 value of 108.02 for fifty one degrees

Table 4.5.3.2.2 Principal factor solution in the 6th iteration for the environment correlation matrix - Cluster IV

Variable	Common factor coefficients				Estimated communality		Original communality (SMC)
	1	2	3	4	5th iteration	6th iteration	
1	-0.1930	0.7569	0.1713	0.2206	0.6889	0.6882	0.9999
2	0.2300	0.8846	0.1361	-0.0699	0.8588	0.8586	1.0000
3	0.9085	0.0929	0.3285	-0.0064	0.9418	0.9420	1.0000
4	0.9792	-0.0950	0.2056	0.0076	1.0000	1.0000	1.0000
5	0.7663	-0.4253	-0.1133	0.1680	0.8096	0.8091	1.0000
6	0.1149	0.7964	0.2265	-0.0490	0.7012	0.7011	0.9999
7	-0.5971	0.0911	-0.1842	0.2951	0.4885	0.4859	0.9998
8	0.1109	0.4434	-0.3034	0.7350	0.8390	0.8412	0.9999
9	-0.4983	0.3048	0.6901	-0.0095	0.8202	0.8175	0.9999
10	0.1256	0.6526	-0.4310	-0.4970	0.8715	0.8744	1.0000
11	0.7463	0.0241	-0.5953	-0.0165	0.9099	0.9123	0.9999
12	-0.0202	0.9430	-0.0537	-0.0036	0.8918	0.8925	1.0000
13	-0.0058	0.6782	-0.4120	-0.0926	0.6411	0.6384	1.0000
14	0.9780	-0.1119	0.1504	0.0062	0.9913	0.9917	1.0000
15	0.8438	0.4446	0.2111	0.1089	0.9660	0.9661	1.0000

Table 4.5.3.2.3 Rotated principal factor loadings for the environment correlation matrix - Cluster IV

Variable	Common factor coefficients			
	1	2	3	4
1	-0.1151	0.7569	0.2310	0.2206
2	0.2642	0.8846	0.0405	-0.0699
3	0.9655	0.0929	-0.0342	-0.0064
4	0.9852	-0.0950	-0.1747	0.0076
5	0.6686	-0.4253	-0.3911	0.1680
6	0.1911	0.7964	0.1673	-0.0490
7	-0.6227	0.0911	0.0520	0.2952
8	-0.0103	0.4434	-0.3229	0.7350
9	-0.2048	0.3078	0.8262	-0.0096
10	-0.0443	0.6526	-0.4468	-0.4970
11	0.4702	0.0241	-0.8309	-0.0165
12	-0.0388	0.9430	-0.0423	-0.0037
13	-0.1591	0.6782	-0.3801	-0.0926
14	0.9635	-0.1119	-0.2254	0.0062
15	0.8616	0.4446	-0.1190	0.1089
Proportionate variance accounted by each factor	0.3213	0.2981	0.1439	0.0653

of freedom, which was significant. So ML solution of factor loadings was tried with a five factor model which again found to be significant (χ^2_{40} = 57.97). Fifty two iterations were taken for the convergence. ML solution of factor loadings with six factors was found to be adequate to explain the dependence structure (χ^2_{30} = 54.52). Sixty eight iterations were required for the convergence with a ±0.005 convergence criterion. The initial estimates of factor loadings and unique variances are given in Table 4.5.3.3.1. The ML solutions in the 67th and 68th iterations are given in Tables 4.5.3.3.2 and 4.5.3.3.3 respectively. The varimax rotated loadings are presented in Table 4.5.3.3.4. The residual matrix after removal of six factors is given in Table 4.5.3.3.5. The characters with high loadings in each factor are given below.

	Pol at 12th month
	C.C.S. percentage
Factor I	Brix at 12th month
	Sugar yield per plot
	Purity percentage
	Cane yield per plot
Factor II	Shoot count
	Number of millable canes per plot
	Germination count

Table 4.5.3.3.1 Initial estimates of factor loadings and corresponding unique variances for 6 factors - Cluster III

Variable	Factor loadings						Unique variance
	1	2	3	4	5	6	
1	-0.2202	0.6192	0.3038	-0.1875	-0.4302	0.0497	0.2531
2	-0.1652	0.7787	0.1868	-0.1264	-0.0153	0.3766	0.1734
3	0.9318	-0.1817	-0.2158	0.1236	-0.1783	0.1766	0.0001
4	0.9991	-0.0005	-0.0269	0.0157	-0.0253	-0.0495	0.0002
5	0.8002	-0.2507	-0.2937	-0.0275	0.2840	-0.4709	0.0001
6	-0.0802	0.7342	-0.2260	-0.1000	-0.0560	0.2219	0.3510
7	0.6363	-0.0431	0.0893	-0.2131	-0.1970	0.2443	0.5586
8	0.0251	0.4171	0.0003	0.4443	0.4770	-0.0257	0.3998
9	0.3831	-0.1673	-0.7414	-0.3452	0.0910	-0.0051	0.1481
10	-0.0045	0.2573	0.4994	-0.3345	0.2404	-0.1128	0.5020
11	0.6193	0.2131	0.6923	-0.2936	0.0195	0.1062	0.0003
12	-0.1233	0.9829	0.1139	0.0227	-0.0489	-0.1165	0.0002
13	-0.1505	0.6691	0.2831	0.2577	-0.1603	0.0335	0.3563
14	0.9980	0.0216	0.0339	-0.0327	0.0346	-0.2885	0.0001
15	0.8201	0.5248	0.1733	0.0493	-0.1256	-0.0733	0.0002

Table 4.5.3.3.2 Maximum likelihood estimates of factor loadings and unique variances in the 67th iteration - Cluster IV

Variable	Factor loadings						Unique variance
	1	2	3	4	5	6	
1	-0.2501	0.6139	0.3417	-0.0511	-0.4072	0.2568	0.2094
2	-0.1479	0.7741	0.2459	-0.0951	-0.0213	-0.5171	0.0415
3	0.9217	-0.1810	-0.2398	0.0840	-0.2285	0.0222	0.0004
4	0.9971	-0.0019	-0.0889	0.0037	-0.0229	0.0070	0.0002
5	0.8159	-0.2357	-0.2459	0.0551	0.2989	-0.0066	0.8741
6	-0.0644	0.7311	-0.2326	-0.0243	-0.0976	0.3852	0.2487
7	0.6265	-0.0576	0.1787	0.2538	-0.1651	0.0574	0.4773
8	0.0299	0.4400	0.0959	-0.5661	0.2635	-0.1162	0.3929
9	0.4392	-0.1423	-0.7782	0.1379	0.0398	-0.0451	0.1586
10	-0.0094	0.3489	0.3727	-0.5520	0.0569	-0.0093	0.4313
11	0.6665	0.2385	0.6951	-0.0902	0.0112	0.0059	0.0074
12	-0.1349	0.9818	0.1216	0.0386	-0.0146	-0.0082	0.0013
13	-0.1489	0.6698	0.1870	0.2568	-0.2247	0.2841	0.2971
14	0.9979	0.0196	0.0459	-0.0336	0.0532	-0.0082	0.0001
15	0.8079	0.5264	0.2242	0.1065	-0.0792	-0.0091	0.0022

Table 4.5.3.3.3 Maximum likelihood estimates of factor loadings and unique variances in the 68th iteration - Cluster IV

Variable	Factor loadings						Unique variance
	1	2	3	4	5	6	
1	-0.2517	0.6090	0.3446	-0.0471	-0.4040	0.2598	0.2141
2	-0.1433	0.7711	0.2490	-0.0936	-0.0252	-0.5200	0.0431
3	0.9188	-0.1809	-0.2437	0.0806	-0.2333	0.0197	0.0024
4	0.9953	-0.0052	-0.0928	0.0007	-0.0227	0.0048	0.0002
5	0.8192	-0.2326	-0.2400	0.0587	0.3014	-0.0018	0.1229
6	-0.0609	0.7281	-0.2365	-0.0274	-0.0997	0.3893	0.2480
7	0.6242	-0.0607	0.1835	0.2564	-0.1610	0.0545	0.4784
8	0.0335	0.4433	0.0981	-0.5699	0.2606	-0.1194	0.3856
9	0.4422	-0.1384	-0.7811	0.1360	0.0366	-0.0482	0.1531
10	-0.0130	0.3528	0.3698	-0.5540	0.0540	-0.0055	0.4287
11	0.6706	0.2416	0.6956	-0.0881	0.0081	0.0020	0.0002
12	-0.1375	0.9816	0.1243	0.0417	-0.0117	-0.0036	0.0002
13	-0.1457	0.6699	0.1849	0.2569	-0.2286	0.2892	0.2939
14	0.9965	0.0157	0.0475	-0.0338	0.0564	-0.0044	0.0001
15	0.8035	0.5283	0.2281	0.1097	-0.0773	-0.0073	0.0052

Table 4.5.3.3.4 Rotated maximum likelihood estimates of factor loadings – Cluster IV

Variable	Factor loadings					
	1	2	3	4	5	6
1	0.1278	0.6023	0.4071	−0.4036	−0.2783	0.0251
2	−0.2156	0.8919	0.1898	−0.1875	−0.0986	−0.1843
3	0.9484	−0.1328	−0.0641	0.0512	−0.2694	0.0353
4	0.9725	−0.0023	−0.2315	0.0056	−0.0222	0.0061
5	0.6989	−0.3038	−0.4902	0.0386	0.3079	−0.1278
6	−0.1336	0.7613	−0.2045	−0.2498	−0.2021	0.0981
7	0.6501	−0.0606	0.0266	−0.3003	−0.0576	0.0334
8	−0.0002	0.2215	0.1036	0.7437	0.0051	−0.0352
9	0.1681	−0.1539	−0.8817	0.0980	0.0011	−0.1182
10	0.1064	0.3277	0.3544	0.2139	0.0239	0.3597
11	−0.4119	0.1752	0.8740	0.1585	0.0693	0.0727
12	0.0903	0.8174	0.1618	−0.3919	−0.1372	−0.3533
13	−0.1973	0.7731	0.1284	−0.1009	−0.2062	0.0160
14	0.9590	0.0004	0.2748	0.0146	0.0660	−0.0047
15	0.8342	0.4067	0.0418	0.3121	0.0365	−0.1818
Contribution of each factor	5.3084	4.1750	1.7402	0.8161	0.5931	0.4793
Proportionate variance accounted by each factor	0.3139	0.2714	0.1560	0.0933	0.0248	0.0247

Table 4.5.3.3.5 Residual matrix after removal of six factors from the environment correlation matrix - Cluster IV

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}
x_2	-0.0022													
x_3	-0.0011	0.0000												
x_4	0.0000	0.0000	0.0000											
x_5	-0.0122	0.0130	-0.0017	0.0004										
x_6	0.0894	0.0211	-0.0064	0.0010	-0.0425									
x_7	0.0343	0.0201	-0.0177	0.0040	0.0195	0.1321								
x_8	0.0133	-0.0048	0.0062	-0.0008	0.0181	-0.1429	0.0490							
x_9	0.0161	0.0068	-0.0028	0.0008	0.0338	0.0542	0.0999	0.0193						
x_{10}	0.0020	-0.0002	0.0000	0.0000	0.0003	-0.0005	0.0002	0.0011	-0.0002					
x_{11}	0.0001	0.0000	0.0000	0.0000	0.0001	0.0005	0.0002	0.0000	-0.0001	0.0000				
x_{12}	-0.0009	0.0001	0.0000	0.0000	-0.0002	0.0000	0.0013	0.0004	0.0004	0.0000	0.0000			
x_{13}	-0.0781	0.0099	0.0090	-0.0022	-0.0183	-0.1943	-0.1813	0.1787	-0.1304	0.0002	-0.0006	0.0001		
x_{14}	-0.0007	0.0000	0.0000	0.0000	-0.0005	-0.0039	0.0033	-0.0012	-0.0006	0.0000	0.0000	0.0000	0.0027	
x_{15}	0.0183	-0.0029	0.0007	-0.0001	0.0031	-0.0014	-0.0278	-0.0105	-0.0082	0.0000	0.0000	0.0000	-0.0007	0.0000

Factor III Girth of cane
Weight of cane

Factor IV Length of internode

Factor V Length of cane

Factor VI Number of internodes

In both PFA and ML methods, factor I was found to be highly correlated with pol at 12th month, C.C.S. percentage, brix at 12th month, sugar yield per plot and purity percentage which was designated as quality factor. Cane yield per plot, shoot count, number of millable canes per plot and germination count were found to be highly correlated with factor II in both the methods. The third factor was dominated by number of internodes and length of cane in PFA and girth of cane and weight of cane in ML method. The fourth factor is characterised by length of internode and juiciness at 12th month in PFA and length of internode only in ML method. In ML method, length of cane formed an independent factor, factor V and number of internodes another factor, factor VI.

In PFA 82.86 percentage of variation in the dependence structure was explained by the four common factors while in ML method 87.41 percentage was explained by the six factor model. The proportion of variation accounted by factor I was about 32.13 per cent in PFA and 31.39 per cent in ML solution. Second factor accounted a proportion of variance of 29.81 per cent in PFA and 27.14 per cent in ML solution. The contribution of factor III was 14.39 per cent in PFA and 15.6 in ML solution. Contribution of remaining factors were 6.53 and 13.28 respectively in PFA and ML methods.

PFA and ML methods were tried for each cluster. The clones within each cluster are less divergent than those between clusters. The two methods were tried for each cluster with the purpose of identifying the factors in general. All the clusters gave more or less the same result when tried with both the methods. However ML method is preferred as it allows the testing of the adequacy of the factor model for generating the observed correlations.

The characters pol at 12th month, C.C.S. percentage, brix at 12th month, sugar yield per plot and purity percentage remained the same in factor I for all the three clusters. This factor is clearly a factor associated with quality aspect and contributes a major share of the variation of the dependent structure of the morphological and quality traits of the crop. The characters which are more amenable to change in this factor are pol at 12th month, C.C.S. percentage and brix at 12th month. Second factor were dominated with the characters cane yield per plot, shoot count, germination count and number of millable canes in all the three clusters while weight of cane is found to be an additional character in this factor for the first cluster. In this factor the characters cane yield per plot and shoot count are found to be the characters which are more amenable to changes. In the case of other factors the characters are not the same in the three clusters. However in all the clusters 6 factors are found to be necessary to explain the covariance structure as revealed by the ML method.

From this study it is clear that the main factor of divergence in sugarcane is the quality factor of which the characters pol at 12th month, C.C.S. percentage and brix at 12th month contributing more towards divergence. Hence this factor has to be given more importance in breeding programmes on divergence. In the second factor the characters cane yield per plot, shoot count and number of millable canes contributing more towards divergence.

SUMMARY

SUMMARY

Multivariate statistical techniques are very much useful in plant breeding programmes on sugarcane as they estimate the degree of divergence in morphological and quality traits which are intercorrelated to varying degrees. Factor analysis is considered as the best analytic method due to its power and elegance in studies of this type. Principal factor analysis and maximum likelihood method are two ways to extracting the factors of divergence, of which maximum likelihood method is considered as the best one as it satisfies certain properties of a best estimator and allows for the determination of an adequate number of stable factors from the point of view of goodness of fit of the factor-model.

The available data on various morphological and quality traits in sugarcane with respect to forty eight varieties were utilized for the study. The analysis of dispersion revealed significant differences among the varieties for aggregate effect of all the characters indicating considerable variability among the experimental material.

Divergence analysis is performed to identify the diverse genotypes for hybridization purposes. The forty eight genotypes were grouped into thirteen clusters by

D^2-analysis. The first cluster consisted of fifteen varieties, second five varieties, third nine, fourth seven and fifth cluster consisted of four varieties. The other genotypes were not able to cluster.

Various factor-models were tried for the environment correlation matrix as factor analysis aims to explain the intercorrelations among the numerous variables in terms of simple relations. Factor analysis was done separately for the first, third and fourth clusters.

Principal factor analysis allows for the determination of a m-factor pattern where m refers to the number of principal components whose eigen values are greater than or equal to one (Harman, 1967). As such a five-factor model was fitted to the environment correlation matrix of cluster I and four factor models for third and fourth clusters. The first factor was the same for the three clusters which was a quality factor. The characters pol at 12th month, C.C.S. percentage, brix at 12th month, sugar yield per plot and purity percentage belonging to this factor. Second factor was dominated with cane yield per plot, shoot count, germination count and number of millable canes in the three clusters while weight of cane also was in this factor for the first cluster. Third factor was the same for first and fourth clusters which consisted of the characters length of cane and number of internodes while number of internodes only in the third factor of the second cluster. The characters length of internode and juiciness at 12th month belonging to fourth factor which was the same for the three clusters. The additional factor for the first cluster consisted of girth of cane only.

The maximum likelihood method resulted in fitting a six factor model to explain the correlation structure in all the three clusters. The first two factors are the same as that obtained by PFA. Third factor consisted of length of cane and number of internodes in the first cluster while number of internodes remained alone in the third cluster. In the fourth cluster third factor was dominated with the characters girth of cane and weight of cane. The characters juiciness at 12th month and girth of cane belonging to fourth factor in the first cluster and girth of cane remained independently in the third cluster. But in the fourth cluster length of internode formed as the fourth factor. Fifth factor was dominated by length of internode in the first cluster while length of cane and length of internode dominated this factor in the third cluster. Length of cane formed the fifth factor in fourth cluster. Sixth factor consisted of weight of cane in the first cluster, juiciness at 12th month in the third cluster and number of internodes in the fourth cluster.

The characters which are more amenable to changes due to selection are pol at 12th month, C.C.S. percentage and brix at 12th month in the quality factor and cane yield per plot and shoot count in the second factor which found to be the same in all the three clusters studied. It is clear that the quality factor is the main factor of divergence in sugarcane.

REFERENCE

REFERENCES

Abraham, T.P. and Hoobakht, A. (1974). An application of factor analysis for interpretation of soil analysis data. J. Indian Soc. agric. Stat., 26(1): 103-112.

Anand, I.J. and Rawat, D.S. (1984). Genetic diversity, combining ability and heterosis in Brown mustard. Indian J. Genet., 44(2): 226-234.

Anderson, T.W. (1958). An Introduction to Multivariate Statistical Analysis. John Wiley and Sons, Inc. New York, pp. 374.

Arunachalam, V. and Ram, J. (1967). Geographical diversity in relation to genetic divergence in cultivated sorghum. Indian J. Genet., 27: 369-380.

Arunachalam, V. (1981). Genetic distance in plant breeding. Indian J. Genet., 40(1): 226-236.

Bart, C. (1952). Tests of significance in factor analysis. Brit. J. Psychol. Stat. Sec., 5: 109-133.

Bartlett, M.S. (1934). The vector representation of a sample. Proc. Camb. Phil. Soc., 30: 327-340.

Bartlett, M.S. (1947). Multivariate analysis. J. Roy. Stat. Soc., (B) 9: 176-197.

Bartual, B., Carbonell, E.A. and Green, D.E. (1985). Multivariate analysis of a collection of soybean cultivars for South Western Spain. Euphytica 34(1): 113-123.

Catell, R.B. (1965 a). Factor analysis. An introduction to essentials 1. The purpose and underlying models. Biometrics 21(1): 190-215.

Cattell, R.B. (1965 b). Factor analysis. An introduction to essentials II. The role of factor analysis in research. Biometrics 21(2): 405-435.

Cattell, R.S. and Khanna, D.K. (1977). Principles and procedures for unique rotation in factor analysis. Mathematical methods for digital computers. 3. Ed. Enslein, K., Ralston, A. and Wilf, H.S., John Wiley and Sons, Inc. New York.

Cheudhary, B.D. and Singh, V.P. (1975). Genetic divergence in some and exotic barley varieties and their hybrids. Indian J. Genet., 35: (409-413).

Denis, J. and Adams, M.W. (1978). A factor analysis of plant variables related to yield in dry beans. Crop Sci., 18(1): 74-78.

Dixit, R.H. (1980). Genetic divergence for yield and its components in lentil. Indian J. agric. Sci., 50(3): 222-224.

Dobhal, V.K. and Ram, H. (1985). Genetic divergence in pea. Indian J. agric. Sci., 55(2): 67-71.

Everitt, B. (1979). Unresolved problems in cluster analysis Biometrics 35(1): 169-181.

Everitt, B. (1980). Cluster analysis. 2nd Edn. Heinemann Educational Books, London.

*Fisher, R.A. (1922). On the mathematical foundations and theoretical statistics. Phil. Trans. Roy. Soc. 222: 309-368.

*Fisher, R.A. (1925). Theory of statistical estimation. Proc. Camb. Phil. Soc., 22: 700-725.

Fruchter, B. (1954). Introduction to Factor Analysis. D. Van Nostrand Company, Inc. New York, pp. 280.

Gaur, P.C., Gupta, P.K. and Kishore, H. (1978). Studies on genetic divergence in potato. Euphytica 27: 361-368.

Gill, S.D., Tripathi, B.K., Misra, G.P. and Lal, S. (1983). Character association analysis in some foreign varieties of sugarcane. Indian J. agric. Sci., 53(12): 985-987.

Gupta, M.P. and Singh, R.B. (1970). Genetic divergence for yield and its components in greengram. Indian J. Genet., 30(1): 212-221.

Guttman, L. (1956). "Best possible" systematic estimates of communalities. Psychometrika, 21(3): 273-285.

Harman, H.H. (1967). Modern Factor Analysis. 2nd edn. University of Chicago Press, Chicago, pp. 473.

Hemmerle, W.J. (1965). Obtaining maximum likelihood estimates of factor loadings and communalities using an easily implemented iterative computer procedure. Psychometrika, 30(3): 291-302.

Holzinger, K.J. and Harman, H.H. (1941). Factor Analysis. University of Chicago Press, Chicago. pp. 417.

Hotelling, H. (1931). The generalization of student's ratio. Ann. Math. Stat., 2: 360-364.

Hotelling, H. (1933 a). Analysis of a complex of statistical variables into principal components. J. Educ. Psychol., 24: 417-441.

Hotelling, H. (1933 b). Analysis of a complex of statistical variables into principal components. J. Educ. Psychol., 24: 498-520.

*Hotelling, H. (1935). The most predictable criterion. J. Educ. Psychol., 26: 139-142.

Jatasra, D.S. and Paroda, R.S. (1983). Genetic divergence in wheat. Indian J. Genet., 43(1): 63-67.

Jindall, S.K. and Gupta, B.S. (1985). Genetic divergence in fodder cowpea. Indian J. agric. Sci., 55(1): 1-3.

Joreskog, K.G. (1967). Some contributions to maximum likelihood factor analysis. Psychometrika, 32: 443-482.

Joreskog, K.G. (1969). A general approach to confirmatory maximum likelihood factor analysis. Psychometrika, 34(2): 183-202.

Joreskog, K.G. (1971). Simultaneous factor analysis in several populations. Psychometrika, 36: 409.

Joreskog, K.G. and Goldberger, A.S. (1972). Factor analysis by generalized least squares. Psychometrika, 37.

Joreskog, K.G. (1977). Factor analysis by least squares and maximum-likelihood methods. Mathematical Methods for Digital Computers. 3 Ed. Enslein, K., Ralston, A. and Wilf, H.S., John Wiley and Sons, Inc. New York.

*Kaiser, H.F. (1956). The varimax method of factor analysis. Unpublished Ph.D. Thesis, University of California.

*Kaiser, H.F. (1960). The application of electronic computers to factor analysis. Edc. Psychl. Measurement, 20: 141-151.

Katiyar, R.P. and Singh, S.P. (1979). Genetic divergence in chickpea. Indian J. Genet., 39(2): 354-358.

*Kelley, T.L. (1935). Essential traits of mental life. Harward Studies in Education, 26 Harward University Press, Cambridge.

Kendall, M.G. (1950). Factor analysis as a statistical technique. J. Roy. Stat. Soc., (B) 12: 60-75.

Kendall, M., Stuart, A. and Ord, J.K. (1983). The Advanced Theory of Statistics. 3 (4th ed.) Charles Griffin and Company, London.

Kruskal, J.B. and Shepard, R.N. (1974). A non metric variety of linear factor analysis. Psychometrika 39: 123.

Krjanowski. (1983). Distance between populations using mixed continuous and categorical variables. Biometrika 70(1): 235-243.

Kukadia, M.V. and Singhania, D.L. (1984). Factor analysis in sorghum for forage yield. Indian J. Agric. Sci., 54(11): 1001-1003.

*Lawley, D.N. (1940). The estimation of factor loadings by the method of maximum likelihood. Proc. Roy. Soc. Edin., 60: 64-82.

*Lawley, D.N. (1942). Further investigations in factor estimation. Proc. Roy. Soc. Edin., 61: 176-185.

Lawley, D.N. (1943). The application of the maximum likelihood method for factor analysis. Brit. J. Psych., 33: 172-175.

Lawley, D.N. and Maxwell, A.E. (1963). Factor Analysis as a Statistical Method. Butterworth, London, pp. 177.

*Mahalanobis, P.C. (1928). A statistical study at Chinese head measurement. J. Asiatic Soc. Bengal, 25: 301-377.

Mahalanobis, P.C. (1936). On the generalised distance in statistics. Proc. natn. Acad. Sci. India, 2: 49-55.

Martin, N.G. and Eaves, L.J. (1977). The genetical analysis of covariance structure. Heredity 38(1): 79-95.

*Maxwell, A.E. (1961). Recent trends in factor analysis. J. Roy. Stat. Soc., (A) 124: 49-59.

Mc. Donald, R.P. (1970). The theoretical foundations of principal factor analysis, canonical factor analysis and alpha factor analysis. Brit. J. Math. Stat. Psychol., 23(1): 1-21.

Mehndiratta, P.D., Phul, P.S. and Arrora, N.D. (1971). Genetic diversity in relation to fodder yield and its components in sorghum. Indian J. Genet., 31(2): 300-304.

Mercy, K.A. and George, K.C. (1987). Genetic divergence in culinary varieties of banana. Agric. Res. J. Kerala, 25(1): 11-16.

Morrison, D.F. (1978). Multivariate Statistical Methods. Mc Graw-Hill. International Book Company.

Mulaik, S.A. (1972). The Foundations of Factor Analysis. Mc. Graw Hill. Inc. New York. pp. 453.

Muralidharan, K. (1986). Assessment of genetic divergence by factor analysis in groundnut. Unpublished M.Sc. (Ag.Stat.) Thesis. K.A.U.

Murty, B.R. and Qudri, M.I. (1966). Analysis of divergence in some self-compatible forms of Brassica campertris var. Brown Sarson. Indian J. Genet., 26(1): 45-58.

Murty, B.R. and Arunachalam, V. (1967 a). Computer program for some problems in biometrical genetics. 1. Use of Mahalanobis D^2 in classificatory problems. Indian J. Genet., 27(1): 60-67.

Murty, B.R. and Arunachalam, V. (1967). Factor analysis of diversity in the genus sorghum. Indian J. Genet., 27: 123-135.

Nageswara Rao, P., Rehman, M.A., Panduranga Rao, C. (1983). Genetic variability and character associations in sugarcane progenies. Indian J. agric. Sci., 53(8): 621-623.

Nair, P.K.C., Sukumaran, S. and Namboodiri, K.M.N. (1982). Performance of sugarcane varieties in Kerala. Agric. Res. J. Kerala, 20(1): 26-31.

Nair, P.S. and Gupta, Y.K. (1977). Genetic diversity based on components of fodder yield in oats. Agric. Res. J. Kerala, 15(2): 160-164.

*Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. Philosophical Magazine (Series 6) 2: 559-572.

Peter, K.V. and Rai, B. (1976). Analysis of genetic divergence in tomato. Indian J. Genet., 36: 379-383.

Punia, M.S., Chaudhary, B.S. and Hooda, R.S. (1983). Genetic divergence in sugarcane. Indian J. agric. Sci., 53(6): 434-436.

Punia, M.S., Chaudhary, B.S. and Hooda, R.S. (1983). Correlation and path analysis of cane yield in sugarcane. Indian J. Genet., 43(1): 109-112.

Ram, J. and Panwar, D.V.S. (1970). Intraspecific divergence in rice. Indian J. Genet., 30(1): 1-10.

Rao, C.R. (1948). The utilisation of multiple measurements in problems of biological classification. J. Roy. Stat. Soc. (B) 10(2): 159-203.

Rao, C.R. (1952). Advanced Statistical Methods in Biometric Research. John Wiley and Sons, Inc. New York, pp. 246-250.

Rao, C.R. (1955). Estimation and tests of significance in factor analysis. Psychometrika 20(2): 93-111.

Roff, M. (1936). Some properties of the communality in multiple factor theory. Psychometrika 1: 1-6.

Sawant, A.R., Asawa, B.M. and Rawat, G.S. (1982). Factor analysis of ninety diversified strains of triticale. Indian J. agric. Sci. 52(4): 209-211.

Schilderinik, J.H.F. (1978). Regression and Factor Analysis in Econometrics. 2nd ed. pp. 236.

Shetty, N.S. (1969). A factor analysis of use of fertilizers by farmers. Indian J. agric. Eco., 24: 50-61.

Singh, B.B. (1973). Factor analysis in Gossypium hirsutum L. Madras agric. J., 60: 1535-1939.

Singh, D. (1980). Genetic diversity for some quantitative characters in barley. Indian J. Genet., 40: 391-395.

Singh, H.N. (1983). Variability for yield and quality in sugarcane. Indian J. agric. Sci., 59(9): 786-789.

Singh, H.N., Singh, S.B. and Singh, T.K. (1981). Selection parameters in sugarcane. Indian J. agric. Sci., 51(8): 562-566.

Singh, K.P., Kakar, S.N., Singh, V.P. and Chaudhary, D.D. (1982). Genetic divergence in chickpea. Indian J. agric. Sci., 52(8): 492-495.

Singh, R.B. and Bains, S.S. (1968). Genetic divergence for gining out turn and its components in upland cotton varieties obtained from different geographic locations. Indian J. Genet., 28: 262-268.

Singh, R.B. and Gupta, M.P. (1968). Multivariate analysis of divergence in upland cotton. Indian J. Genet., 28: 151-157.

Singh, R.R. and Singh, H.N. (1980). Genetic divergence in tomato. Indian J. agric. Sci., 50(8): 591-594.

Singh, T.H. and Gill, S.S. (1984). Genetic diversity in upland cotton under different environments. Indian J. Genet., 44(3): 506-513.

Spearman, C. (1904). General intelligence objectively determined and measured. Amer. J. Psychol., 15: 201-293.

Sreekumar, S.G. (1986). Genetic variability, path analysis and stability parameters in sugarcane. Unpublished Ph.D. Thesis, K.A.U.

Sukumaran, S.N., Jacob, K.M. and Nair, P.K.C. (1982). Lodging of sugarcane and its influence on yield and quality of certain cultivars. Agric. Res. J. Kerala, 20(2): 8-15.

Sundaram, A., Ramakrishnan, A., Marappan, P.V. and Balakrishnan, G. (1980). Factor analysis in cowpea. Indian J. agric. Sci., 50(3): 216-218.

Swain, A.J. (1975). A class of factor analysis procedures with common asymptotic sampling properties. Psychometrika 40: 315.

Tes P.Mathew (1988). Factor analysis of genetic divergence in Sesame. Unpublished M.Sc.(Ag. Stat.) Thesis, K.A.U.

Thurstone, L.L. (1931). Multiple Factor Analysis. University of Chicago Press, Chicago.

Thurstone, L.L. (1947). Multiple Factor Analysis. Psychol. Rev., 38: 406-427.

Tikka, S.B.S. and Asawa, B.M. (1979). Factor analysis in lentil. Indian J. agric. Sci., 48(11): 643-646.

Upadhyay, M.K. and Murty, B.R. (1970). Genetic divergence in relation to geographical distribution in pearl millet. Indian J. Genet. 30: 704-715.

Valsalakumari, P.K., Nair, P.C.S. and Prabhakaran, P.V. (1985). Genetic divergence in banana. Agric. Res. J. Kerala, 23(2): 146-149.

Walton, P.D. (1972). Factor analysis of yield in spring wheat. Crop Sci., 12: 731-733.

Wilker, S.S. (1932 a). Certain generalizations in the analysis of variance. Biometrika, 24: 471.

Wilks, S.S. (1935). On the independence of k sets of normally distributed statistical variables. Econometrics 3: 309.

Wishart, J. (1928). The generalized product moment distribution in sample from a normal multivariate population. Biometrika, 20(A): 32-36.

* Originals not seen

APPENDICES

Appendix - I

Between dispersion matrix

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}
x_1	201.27	199.56	-2.07	-4.09	-7.54	164.76	-157.16	7.35	1.13	10.11	0.56	257.24	2.04	-4.35	22.05
x_2		762.01	3.86	-1.31	-24.95	630.41	-436.98	6.78	-15.08	-10.28	-5.15	393.66	-0.01	-5.54	36.59
x_3			11.51	13.56	27.31	12.67	6.46	-2.14	-0.14	4.71	-0.10	5.66	0.07	10.42	9.31
x_4				16.67	36.11	10.88	14.92	-2.88	-0.09	5.75	-0.06	5.36	0.05	12.99	11.44
x_5					94.21	7.16	42.81	-7.58	-0.75	11.47	-0.10	-5.05	-0.09	29.05	24.37
x_6						613.55	-355.85	7.97	-14.01	0.88	-4.29	413.84	1.17	5.17	47.42
x_7							3265.70	-25.23	35.89	140.46	12.69	457.71	4.89	14.99	59.47
x_8								6.59	0.21	-5.67	0.11	9.13	0.39	-2.33	-1.23
x_9									1.78	2.14	0.47	17.07	0.20	-0.05	1.56
x_{10}										21.49	0.94	79.79	1.04	4.60	11.86
x_{11}											0.17	5.85	0.09	-0.01	0.53
x_{12}												1030.19	8.29	2.05	103.54
x_{13}													0.18	0.05	0.83
x_{14}														10.23	8.79
x_{15}															17.87

Appendix - II

Within dispersion matrix

X_1	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}
X_1	40.19	38.16	-0.60	-1.22	-4.74	28.54	-15.34	0.79	-0.03	2.55	0.03	35.46	0.31	-0.93	2.23
X_2		109.79	3.47	3.38	4.04	61.37	-156.65	0.46	-1.12	4.19	0.11	87.99	0.64	2.75	10.29
X_3			3.15	3.70	8.22	0.83	-11.85	0.38	0.00	0.27	0.03	1.81	0.01	2.90	2.71
X_4				4.88	13.05	-0.53	-9.84	0.37	-0.04	0.38	0.05	1.18	0.01	3.90	3.57
X_5					49.21	-6.76	-8.72	0.27	-0.41	0.67	0.20	-3.53	-0.05	10.92	9.67
X_6						127.53	-40.56	0.17	-1.48	2.95	-0.05	60.60	0.38	-0.91	4.45
X_7							1895.59	-3.62	0.88	-3.96	-0.09	-89.13	-0.94	-7.61	-15.96
X_8								0.96	0.07	-0.01	-0.01	1.04	0.00	0.27	0.19
X_9									0.32	0.03	0.00	0.18	-0.01	-0.04	-0.02
X_{10}										3.74	0.10	9.61	0.24	0.29	1.14
X_{11}											0.02	0.28	0.01	0.03	0.06
X_{12}												143.10	0.84	0.55	13.45
X_{13}													0.04	0.00	0.09
X_{14}														3.21	2.87
X_{15}															3.97

Appendix-III.

1	2	3	4	5	6	7	8	9
1.	53.23	21.76	31.76	44.55	61.56	27.93	29.22	61.37
2.		34.49	38.49	23.37	69.25	39.79	45.56	53.81
3.			22.21	32.47	87.30	21.01	36.45	23.97
4.				47.48	48.27	26.94	36.40	53.59
5.					79.41	41.93	57.81	60.11
6.						41.56	34.05	151.96
7.							19.93	63.60
8.								87.50
9.								
10.								
11.								
12.								
13.								
14.								
15.								

D^2-values

10	11	12	13	14	15	16
19.99	53.04	70.86	37.15	27.43	29.48	46.59
48.36	48.52	27.79	32.55	34.83	14.52	73.04
11.89	36.39	48.02	16.61	9.29	20.16	64.37
27.33	27.01	50.65	19.92	18.12	32.69	74.22
44.43	45.19	25.82	29.60	38.49	23.72	31.36
98.55	96.10	103.35	97.88	79.07	69.22	41.14
38.41	55.49	61.00	41.75	21.43	29.00	39.19
54.93	82.41	91.09	67.94	40.45	35.16	16.58
40.73	51.46	52.59	23.38	26.78	31.61	136.79
	38.94	54.95	20.77	20.64	28.35	82.13
		19.22	12.55	26.88	47.96	127.51
			18.99	40.44	31.43	132.70
				11.89	23.22	114.18
					24.86	82.75
						64.16

D^2-values (contd.)

	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
1.	29.78	41.21	35.90	70.13	41.56	28.99	36.02	26.31	35.61	15.89	47.61	48.02	52.22	71.35	41.63	45.71
2.	39.10	14.44	27.35	102.13	20.59	21.23	19.31	26.20	66.97	36.04	51.83	25.14	33.90	92.02	15.82	22.51
3.	17.03	28.44	31.81	40.50	38.10	17.62	21.23	12.09	22.98	13.57	28.21	32.05	29.78	47.69	21.82	17.15
4.	14.33	31.68	32.57	93.52	38.42	11.50	19.39	34.68	54.88	15.51	54.50	38.39	59.53	59.16	40.10	35.31
5.	48.62	18.64	48.82	82.86	27.66	30.02	25.62	25.87	75.81	39.29	68.59	8.81	23.39	105.18	15.91	21.62
6.	73.44	72.61	34.62	200.57	71.90	40.44	64.45	80.15	131.38	62.46	115.86	76.71	114.89	176.66	86.58	108.10
7.	28.50	39.27	18.36	74.17	42.89	16.16	29.83	27.71	53.35	21.27	50.47	38.72	48.16	92.26	38.44	43.69
8.	39.02	45.49	16.90	124.38	49.01	24.54	50.57	31.49	64.20	34.13	52.75	66.84	83.75	113.40	52.78	65.48
9.	24.35	57.54	62.93	31.68	64.46	54.49	36.52	46.72	19.30	33.69	28.44	67.01	32.14	23.40	33.62	24.96
10.	33.06	28.01	51.04	53.06	32.92	24.66	29.18	14.46	38.30	26.76	39.70	39.55	47.30	45.33	26.69	19.33
11.	28.64	29.25	74.31	75.67	45.72	28.63	19.23	48.64	74.16	27.76	90.01	33.98	54.39	80.13	50.99	34.80
12.	45.16	22.01	70.08	83.35	25.88	35.71	18.70	44.42	86.20	43.03	87.67	20.13	33.86	95.70	30.08	22.28
13.	14.73	23.71	54.80	50.59	27.18	24.33	12.33	30.24	40.78	16.03	52.56	22.56	25.98	42.08	24.55	13.38
14.	9.51	26.19	37.82	51.93	33.80	19.30	21.66	26.75	33.10	13.15	44.25	35.20	36.72	50.45	30.57	22.86
15.	24.32	21.44	18.73	70.22	13.85	19.85	13.17	14.35	33.04	20.04	21.60	29.18	22.18	54.40	5.99	14.19

D^2-values (contd.)

	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48
1.	52.85	80.96	33.35	41.78	38.13	25.14	56.51	60.59	51.54	35.41	57.97	26.16	43.06	61.65	23.61	19.68
2.	38.03	36.38	25.61	27.22	45.63	32.51	48.72	34.51	67.19	64.86	25.71	33.42	30.24	87.39	21.47	36.02
3.	40.29	34.00	17.63	14.62	48.99	5.02	67.31	50.22	37.37	18.02	34.75	25.11	29.38	67.62	29.16	28.46
4.	51.83	61.13	41.24	37.02	55.34	26.45	51.39	50.63	56.60	48.52	40.21	18.87	8.10	67.91	46.05	30.74
5.	23.07	44.28	31.08	25.04	66.58	30.01	77.16	39.99	93.73	71.17	31.05	20.88	43.78	116.34	18.35	38.74
6.	111.69	160.69	83.57	90.45	38.64	85.69	28.83	67.79	137.55	125.41	94.73	53.14	45.93	67.02	53.85	47.86
7.	59.57	64.28	32.31	26.06	30.13	20.71	41.85	51.26	69.67	46.57	46.31	29.18	28.58	55.74	24.07	27.15
8.	98.39	97.87	38.87	37.81	11.90	41.52	30.42	47.94	69.37	52.32	71.81	38.34	34.18	27.88	33.71	31.95
9.	38.99	30.71	51.54	27.98	101.75	33.37	121.01	89.03	29.89	35.83	37.84	67.94	63.69	122.53	61.99	73.60
10.	44.38	64.45	22.88	42.34	62.00	10.54	66.59	50.78	41.00	19.58	36.07	24.81	36.64	76.33	35.34	24.33
11.	37.29	72.28	70.98	47.92	109.05	44.57	100.93	101.72	78.81	83.06	37.12	48.57	20.47	150.76	52.09	70.48
12.	20.48	52.47	59.78	39.68	102.63	46.61	96.01	80.40	87.89	94.00	19.17	50.87	36.13	163.79	35.73	68.95
13.	18.40	44.36	39.32	25.40	85.52	19.17	86.52	66.12	48.09	47.34	21.19	28.96	23.22	120.95	37.71	45.97
14.	40.91	47.60	37.36	22.42	62.17	13.67	73.54	65.01	46.45	39.83	27.65	30.71	19.82	88.88	34.36	41.43
15.	27.93	40.53	14.04	16.37	32.42	19.35	40.63	27.56	31.55	31.59	22.65	26.70	33.15	66.63	12.80	21.46

D^2-values (contd.)

	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
16.	82.98	76.03	32.11	165.47	84.33	47.51	83.74	50.71	99.44	69.07	84.95	93.14	111.69	175.67	60.76	103.81
17.		37.60	35.44	63.76	42.50	21.26	21.80	36.89	29.05	8.97	40.52	48.50	43.35	42.46	38.64	34.09
18.			45.64	93.87	13.42	14.01	22.76	15.52	77.49	37.92	65.32	21.05	48.42	92.93	19.07	17.00
19.				105.52	43.81	20.86	27.35	31.56	46.80	25.67	30.28	51.90	49.45	87.68	29.94	45.76
20.					102.48	90.67	68.83	68.70	37.71	54.16	64.61	84.08	48.56	54.13	68.50	51.85
21.						24.73	24.03	20.66	72.10	37.99	59.15	24.25	44.83	82.46	17.97	19.98
22.							15.36	14.55	60.74	19.79	49.01	24.90	48.48	81.65	24.26	24.67
23.								25.84	42.96	13:03	40.90	19.56	21.92	60.09	16.43	17.53
24.									47.27	29.77	32.79	28.18	40.74	72.89	13.71	15.05
25.										22.95	17.72	81.73	40.02	27.28	45.93	46.28
26.											36.03	34.81	31.47	48.53	32.71	31.07
27.												77.24	45.50	33.68	27.10	36.29
28.													23.66	102.85	21.27	20.34
29.														60.50	18.33	23.57
30.															59.98	18.06
31.																6.41

D^2-values (contd.)

	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48
16.	134.69	125.91	55.28	73.96	13.66	64.64	33.71	63.18	113.73	77.33	113.05	61.41	75.20	30.29	45.30	42.76
17.	45.32	58.08	46.80	23.06	61.99	24.37	73.07	70.92	37.93	40.90	36.02	36.64	20.45	85.07	43.11	48.48
18.	41.62	57.40	33.17	34.76	58.57	24.64	55.56	41.60	74.96	60.36	17.82	23.81	21.44	96.97	21.47	37.60
19.	63.72	65.26	25.44	31.01	11.53	35.96	20.82	32.81	51.61	47.15	52.84	36.42	35.32	31.30	24.64	19.50
20.	53.65	44.31	76.35	50.20	141.08	49.80	175.62	141.43	68.69	53.30	72.82	93.47	107.93	166.63	84.90	94.99
21.	38.90	71.21	27.50	37.34	49.97	29.41	46.76	34.37	61.04	58.66	21.91	26.44	29.51	94.64	18.69	30.22
22.	48.72	63.76	25.58	29.35	34.97	16.99	32.32	36.96	61.37	45.16	26.05	16.61	8.99	60.01	20.87	23.52
23.	19.87	43.62	28.10	22.13	54.26	26.51	50.37	47.12	43.88	49.66	28.01	28.75	18.12	92.25	21.54	30.29
24.	49.14	47.60	8.17	23.36	33.99	10.07	45.10	30.16	46.60	22.63	30.71	21.40	33.14	58.87	16.54	20.82
25.	56.26	46.97	42.24	31.83	68.47	36.65	99.19	87.41	14.56	21.59	71.46	69.00	72.16	82.58	57.90	52.99
26.	33.68	51.24	33.18	19.49	47.49	23.41	62.18	64.66	35.53	38.26	42.27	28.31	21.34	75.96	27.67	29.68
27.	64.70	51.37	22.87	32.52	46.11	35.96	64.96	46.72	13.22	14.26	60.24	56.17	68.07	52.42	50.67	39.56
28.	21.33	48.65	30.48	32.50	72.21	27.06	74.93	45.99	95.54	75.84	29.39	18.39	33.32	124.73	21.04	36.43
29.	12.09	20.88	31.83	24.34	79.46	30.61	95.86	61.03	62.01	58.19	40.37	46.09	64.49	131.42	28.22	48.51
30.	63.86	81.36	68.32	63.52	123.40	55.49	135.69	100.96	22.51	35.35	66.70	78.41	84.92	122.01	102.67	78.15
31.	22.50	34.49	11.57	21.61	48.17	18.55	53.01	21.45	46.54	36.30	18.61	23.43	41.74	81.39	17.76	23.91

D^2-values (contd.)

	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48
32.	20.31	32.35	19.88	22.80	70.47	15.20	75.58	40.23	47.35	36.97	10.70	25.54	35.97	100.69	29.22	34.03
33.		41.93	47.76	37.67	100.57	39.31	105.24	69.36	73.88	73.48	27.60	39.88	58.79	150.26	36.22	48.70
34.			44.78	24.03	97.84	40.85	127.85	86.61	70.01	61.45	54.22	74.08	80.40	139.72	54.20	77.39
35.				25.63	28.51	14.64	42.00	17.76	42.89	21.13	43.78	22.41	46.87	51.25	20.53	13.84
36.					51.70	22.76	76.98	52.87	45.68	37.46	39.24	36.57	35.83	83.73	30.53	43.64
37.						46.62	13.98	34.36	67.47	52.61	82.16	47.62	55.64	19.35	28.76	22.23
38.							61.86	39.48	49.33	20.63	26.41	19.67	34.73	70.39	25.44	24.05
39.								31.86	81.87	72.98	79.53	52.33	49.14	33.97	37.76	27.96
40.									79.05	52.48	54.54	24.59	58.68	51.04	38.39	20.93
41.										20.87	72.55	77.57	67.62	76.69	66.91	57.85
42.											63.93	48.57	67.11	50.91	54.66	36.16
43.												33.49	36.88	117.64	36.10	47.74
44.													27.87	69.81	27.73	15.02
45.														83.32	39.70	42.88
46.															78.79	38.70
47.																21.14

The numbers 1 to 48 in the table refers to the code number of varieties

DIVERGENCE ANALYSIS OF MORPHOLOGICAL AND QUALITY TRAITS IN SUGARCANE

BY

SANTHI, T. E.

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 STATISTICS
COLLEGE OF VETERINARY AND ANIMAL SCIENCES
MANNUTHY - TRICHUR

1989

ABSTRACT

Multivariate analytical techniques are found to be very useful in plant breeding research to explain the influence of various factors on the phenomenon under study. Factor analysis is found to be an appropriate tool to identify the factors of genetic divergence. D^2-analysis is helpful to group the divergent genotypes into various clusters when measurements on a number of related characters are available on a large number of genotypes such that the genotypes within a cluster are homogeneous with respect to these characters and heterogeneous between the clusters.

The present study is aimed at identifying the factors of divergence in relation to morphological and quality traits in forty eight clones of sugarcane. The fifteen clones T.67172, Co.7717, Co.419, Coc.779, Co.7219, Coc.777, Ic.225, Co.6304, S-99, Coc.773, Coc.772, Co.62198, Co.62101, Coc.778 and S-77 are able to group into one cluster. Four more clusters are able to form respectively with five varieties (Co.658, Co.62175, S-105, Co.6907, Co.995) in the second cluster, nine (F.1-2, Co.62174, S-87, KHS 3296, Coc.671, Co.7704, Co.785, CoM.7114, CoM.7125) in the third, seven (Co.6807, Co.1340, Co.527, S-33, Co.6806, B.37172, Co.527-M-10) in the fourth and four varieties (Co.1307,

CoA.7602, Coc.705, Co.453) in the fifth cluster. The remaining clones are not able to group. Among these clusters are utilized for factor analysis.

A factor related to quality is extracted as the first factor in all the three clusters. The characters pol at 12th month, C.C.S. percentage, brix at 12th month, purity percentage and sugar yield per plot dominated this factor. Among these characters pol at 12th month, C.C.S. percentage and brix at 12th month are found to be more amenable to changes due to selection. The second factor is identified by the characters cane yield per plot, shoot count germination count and number of millable canes. Apart from these characters weight of cane is also included in this factor in cluster I. The characters which are more amenable to change due to selection are cane yield per plot and shoot count. The characters are not common in the remaining four factors. These six factors are able to explain 66.84 percent, 79.44 percent and 87.41 percent of variation respectively in the first, third and fourth cluster.