Field Crops Research, 27 (1991) 131–157
Elsevier Science Publishers B.V., Amsterdam

Three-way methods for multiattribute genotype×environment data: an illustrated partial survey

K.E. Basford[a], P.M. Kroonenberg[b] and I.H. DeLacy[a]
[a]*Department of Agriculture, University of Queensland, Qld 4072, Australia*
[b]*Department of Education, University of Leiden, Leiden, Netherlands*

(Accepted 9 July 1990)

ABSTRACT

Basford, K.E., Kroonenberg, P.M. and DeLacy, I.H., 1991. Three-way methods for multiattribute genotype×environment data: an illustrated partial survey. *Field Crops Res.*, 27: 131–157.

Several ordination and clustering techniques are discussed with respect to their usefulness in analysing multiattribute genotype×environment data. The methods are briefly described and illustrated by application to data from the Australian Cotton Cultivar Trials (ACCT), a series of regional variety trials designed to investigate various cotton (*Gossypium hirsutum* (L.)) lines in several locations each year. Multivariate techniques applicable to three-way data are necessary to assess these lines using yield and lint-quality data.

By the choice of complementary methods, it is possible to make both global and detailed statements about the relative performance of the cotton lines. These techniques can enhance the researcher's ability to make informed decisions about the genotype×environment data collected from these trials using simultaneous analysis of the attributes of interest.

INTRODUCTION

The existence of significant genotype by environment (G×E) interactions has complicated selection and testing strategies for plant breeders for many years. They reflect differences in adaptation which may be exploited by breeding for specific adaptation (emphasizing favourable interactions) or broad adaptation (minimizing interactions) by selection and by adjustments to the test strategy. However, any objective decision requires a full understanding of the nature of such interactions, and various methodologies have been proposed for their analysis. These include regression on the environment mean (Finlay and Wilkinson, 1963), restriction to similar environments (Horner and Frey, 1957), pattern analysis methods (Byth et al., 1976), principal coordinate analysis (Eisemann, 1981), canonical variate analysis (Seif et al., 1979) and principal component analysis (Goodchild and Boyd,

0378-4290/91/$03.50

1975; Kempton, 1984; Gauch, 1988; Zobel et al., 1988). Each has proved successful in the analysis of univariate G×E data in certain situations.

Because plant breeders are concerned with more than one attribute, it is of interest ot investigate how such analyses are performed. The methods of analysis of data collected on many attributes in one environment (G×A data) have been well developed, and are covered in Plant Breeding and Quantitative Genetics texts. Such standard techniques as correlation, regression, correlated genetic advance and selection indices are used. There has been little in the literature on the simultaneous analysis of multiattribute G×E data. Recent exceptions are Basford (1982), Basford and McLachlan (1985), Kroonenberg and Basford (1989) and Basford et al. (1990). The techniques discussed there, and some other ordination and clustering methods for the analysis of three-way data, are presented here. Our concern is with multivariate or multiattribute G×E interactions which produce a three-way table of performance means, i.e., G×E×A data.

Although this paper is directly concerned with multiattribute G×E data, it must be stressed that the methods being described are generally applicable to three-way data. These techniques are more familiar in the social-science literature, but have not been extensively used in agricultural research. We are bringing them to the attention of agricultural scientists and, by putting them in a common theoretical framework, demonstrate the relationships between them. Their application is demonstrated using the particular case of G×E×A data.

Using the terminology of Carroll and Arabie (1980), these techniques can be characterised as three-way methods because they apply to data which can be classified in three ways: here, genotypes, environments and attributes (G×E×A). Some are called three-way three-mode methods, because they treat the data as they come, i.e. a G×E×A array not condensed or manipulated over any of the ways. Others are called three-way two-mode methods, because one of the entities has been removed or is not measured directly. For example, the G×E×A data could be transformed to G×G×E data by computing per-environment Euclidean distances between each pair of genotypes using their standardized attribute scores.

Two broad classes of analytical methods can be distinguished in the context of three-way data: ordination, and clustering techniques. As stated in Kruskal (1977) and Arabie et al. (1987), the two types are largely complementary, and make use of the same information in different ways. Accordingly, we will always present the results of an ordination and a clustering in conjunction. Multivariate analysis of variance (MANOVA) can also be applied to three-way data, but with a reasonable number of genotypes, environments and attributes, most interaction terms are nearly always significant. DeLacy (1981), Gauch (1988) and Gauch and Zobel (1988) all argued that, even for G×E data on a single attribute, the standard MANOVA was highly uninformative.

The main focus should be on the structure of the interactions and the similarity of the genotypes, which can primarily be evaluated via modelling techniques. The data which are used as an illustration stem from the Australian Cotton Cultivar Trials (ACCT), in particular, the 1981/82 growing season. They consist of the mean performance $x(i,j,k)$ $(i=1,...,25; j=1,...,8; k=1,..,4)$ of 25 cotton lines or entries (referred to as genotypes) in eight locations (referred to as environments) on four attributes (lint yield, lint strength, micronaire – a measure of the fineness of the lint – and lint length).

Before all but one of the ordination and cluster analyses to be presented here, the raw data $x(i,j,k)$ have to be centered and scaled. The chosen form is:

$$\tilde{x}_{ijk} = (x_{ijk} - \mu_k - \beta_j)/s_k \qquad (1)$$

The data are centered by subtracting the sum of the environment mean for that attribute, i.e. the overall attribute effect, $\mu(k)$, and the environment effect, $\beta(j)$. The genotype means are still present in the data. Byth and DeLacy (1989) and Basford et al. (1990) discuss the rationale for this. The data are scaled by dividing by the standard deviation for each attribute, calculated over all environments, $s(k)$.

METHODS OF ANALYSIS

Generalisations of principal component analysis, multidimensional scaling, the mixture method of clustering, and additive clustering will be discussed. Firstly, we consider three-way two-mode methods and then three-way three-mode methods. The results of the cluster analyses will be displayed superimposed on the results from the ordinations to show how the two techniques are complementary and can be used to enhance the understanding of the interactions. Because our major aim is to convey the flavour of what can be done, most details are left unexplained. No mathematical expositions will be given, nor will algorithms for fitting the models be discussed, but the reader will be referred to other publications where these can be found. More detailed interpretations for the analyses of the cotton data are possible, but not presented here.

Three-way two-mode data

One way of trying to analyse three-way data, especially if one is interested in comparing genotypes, is to search for generalisations of standard cluster-analysis techniques. Most cluster techniques take (dis)similarities between elements as their starting point, so in the three-way case one might also start with converting the basic scores into (dis)similarities. Dissimilarities can be defined in terms of distances between genotypes within each environment

over (scaled or standardized) attributes. The most common distance measure used for continuous variables is Euclidean distance, so this was chosen here. Anderberg (1973) and Clifford and Williams (1976) present detailed accounts of the choice of (dis)similarity measures. For each environment j, the dissimilarity between genotypes i and i', $s(i,i';j)$ is defined as

$$s_{ii'}^{(j)} = \sqrt{\sum_{k=1}^{K} (\tilde{x}_{ijk} - \tilde{x}_{i'jk})^2} \qquad (2)$$

Note that the data set is still three-way, but in the form G×G×E. It should be realized that this is not the only way that dissimilarities could be determined. One could calculate the Euclidean distance between genotypes for each attribute over environments, to produce a G×G×A array. However, the G×G×E array is considered more appropriate for these analyses where the emphasis is on the investigation of the genotype response over environments. Very few cluster techniques have been developed to deal with such data. We only know the details of one of them, i.e. the generalisation of the additive cluster technique (Shepard and Arabie, 1979) to individual differences clustering (INDCLUS) by Carroll and Arabie (1983). It is a method for determining overlapping clusters where the elements (genotypes here) can belong to more than one cluster.

Far more ordination techniques are available for similarity data, the most prominent of which is individual differences scaling (INDSCAL) developed by Carroll and Chang (1970). To analyse dissimilarities, a conversion is made to similarities, generally by subtraction or addition of constants. For an overview of other techniques for sets of (dis)similarity matrices, see Carroll and Wish (1974), Carroll and Arabie (1980) or Kroonenberg (1983a, ch. 3). A possible alternative is to compute the scalar or innerproducts between the genotypes across the scaled or standardized attributes, rather than Euclidean distances, and treat these 'covariance' matrices between genotypes with methods such as STATIS, developed in France (e.g., Lavit, 1988).

Clustering

The model underlying individual-differences clustering assumes that, for all environments, there is one set of overlapping clusters, but that each environment may weight the clusters differently. In an extreme case, it could happen that each environment has its own cluster, i.e., each cluster has a nonzero weight on only one environment. In general, this will not happen, and most environments weight all clusters, but to a different degree. The model can be formalized as follows:

$$\hat{s}_{ii'}^{(j)} = \sum_{g=1}^{G} w_g^{(j)} \delta_{ig} \delta_{i'g} + w_0^{(j)} \tag{3}$$

where: $\hat{s}(i,i';j)$ is the estimated similarity between genotypes i and i'; $w(g;j)$, the nonnegative numerical weight of the jth environment on the gth cluster; 1-0 $\delta(i,g)$ indicates whether genotype i is in cluster g or not; and w$(0;j)$ is the additive constant for the jth environment, which might sometimes (but not here) be taken to represent the weight of that environment on the cluster containing all genotypes. According to the model, the similarity between two genotypes i and i' for the jth environment is the sum of the weights $w(g;j)$ of those clusters to which they both belong. For instance, if they never belong to the same cluster, then their similarity for the jth environment is estimated as $w(0;j)$, but if they both belong to all clusters, then their similarity is estimated as the sum of all weights $w(g;j)$ plus $w(0;j)$. The estimation of the parameters of the model is very involved, and has both mathematical programming and alternating (or conditional) least-squares features, the details of which can be found in Arabie and Carroll (1980) and Carroll and Arabie (1983). This model has not been widely applied; see, however, Carroll and Arabie (1983), Miller and Gelman (1983) and Soli et al. (1986) for some illustrative applications. Our INDCLUS analyses were carried out with Version 1 of the stand-alone program INDCLUS (Carroll and Arabie, 1982).

Ordination

An ordination counterpart for handling G×G×E data is a three-way generalisation of multidimensional scaling called Individual Differences Scaling (INDSCAL). The model is conceptually similar to the clustering method above. It assumes that there is one set of common (not necessarily orthogonal) genotype dimensions for all environments, but that each environment may weight these dimensions differently. Again in an extreme case, each environment might weight only one dimension, giving as many dimensions as there are environments. In general, each environment will weight each dimension differently. A formal description of the model is as follows:

$$\hat{s}_{ii'}^{(j)} = \sum_{d=1}^{D} w_d^{(j)} a_{id} a_{i'd} \tag{4}$$

where: $\hat{s}(i,i';j)$ is the estimated similarity between genotypes i and i'; $w(d;j)$, the nonnegative numerical weight of the jth environment on the dth dimension; and the $a(i,d)$ indicates the weight of genotype i on dimension d. Note that there exists only one genotype space for all environments (the $a(i,d)$ are not indexed with j), but that each dimension is weighted differently in each environment (the $w(d;j)$ do depend on j).

As explained in Carroll and Chang (1970), a different but equivalent formalisation (and more common one) can be given in terms of weighted Euclidean distances. Parameter estimation for this model can be performed in different ways. A first algorithm was devised by Carroll and Chang (1970) and implemented in their program INDSCAL, a second was constructed by Takane et al. (1977) and implemented in their program ALSCAL and, most recently, yet another has been developed by Kiers (1989), but this is not yet publicly available in a program. Many applications of the INDSCAL model have appeared, especially in the psychological and market-research literature. The one agronomic application known to us is that of Basford (1982), who analysed soybean data. Our analyses were performed with ALSCAL as contained in the general statistical package SPSS (Anonymous, 1987).

Three-way three-mode data

A distinct disadvantage of the previous approaches is that one of the modes disappears from the analysis; in the above formulations it was the attributes. This makes it difficult to relate the information obtained from the analysis back to the particular attributes. For instance, the size of one (dis)similarity might be dominated by differences in one attribute, while that of another (dis)similarity by differences in another attribute. In principle, it seems more appropriate to refrain from eliminating one of the modes and to analyse the untransformed data.

We are aware of only one appropriate cluster technique – the mixture-maximum-likelihood method of clustering for three-mode data (MIXCLUS3) developed by Basford and McLachlan (1985; see also McLachlan and Basford, 1988). Quite a few ordination techniques deal with three-way three-mode data directly; most prominent among these are Three-mode principal component analysis (Three-mode PCA; Tucker, 1966; Kroonenberg, 1983a) and Parallel Factor analysis (PARAFAC; Harshman, 1970; Harshman and Lundy, 1984).

Clustering

The mixture method of clustering uses the measurements on a set of elements (genotypes here) to identify clusters in which the genotypes are relatively homogeneous, while they are heterogeneous between the clusters, assuming the number of clusters, G, is known. Each cluster is assumed to have a different mean attribute vector within and across environments, but the covariance matrix particular to each cluster is the same across environments. This procedure handles the data in the original form $x(i,j,k)$, not centered or scaled. Formally, if there are G groups (clusters) from which the genotypes have been sampled in unknown proportions $\pi(g)$ $(g=1,...,G)$, then the dis-

tribution of the vector of attribute values for genotype $i(i=1,...,25)$ in environment $j(j=1,...,8)$ is given by:

$$f(x_{ij}) = \sum_{g=1}^{G} \pi_g f_g(\underset{\sim}{x}_{ij}) \tag{5}$$

where:

$$f_g(\underset{\sim}{x}_{ij}) \sim N(\underset{\sim}{\mu}_{gj}, V_g), \ (g=1,...,G)$$

is the usual assumption of the underlying distribution of the attribute vector in each group being multivariate normal. The unknown parameters, i.e. mean vectors, covariance matrices and mixing proportions, are estimated using maximum-likelihood methods.

In a sense, the technique is similar to INDCLUS, as it assumes that there exists one cluster structure common to all environments, but that the characteristics of the clusters may vary between clusters and/or environments. In INDCLUS these characteristics are the weights, and in MIXCLUS3 they are the mean vectors and covariance matrices. In both techniques, the genotypes do not have to belong outright to just one cluster: INDCLUS allows overlapping clusters of genotypes, while MIXCLUS3 estimates the model parameters using a probability of cluster membership for each genotype. However in the latter, non-overlapping clusters do result when each genotype is assigned to the cluster to which it has the highest estimated probability of belonging. Obviously, the results and interpretations from these two techniques will be rather different.

The mixture method of clustering has been programmed by, and is available from, the senior author (K.E.B.). The initial version was listed in McLachlan and Basford (1988) as K3MM. The method has been applied to the aforementioned soybean data by Basford and McLachlan (1985) and to the 1980/81 cotton data from the ACCT by Basford et al. (1990).

Ordination

Parallel-factor analysis

One of the simplest models for three-mode data is the so-called PARAFAC model. It is a generalisation of component analysis (PCA), but with the interpretational flavour of factor analysis. Scores are seen as combinations of components or factors, rather than vice versa. For the present data, the model assumes that genotypes have scores on factors and that the factor structure of the genotypes is the same across all environments and attributes. Both the attributes and environments weight these scores independently of each other to estimate the original scores. The formal description of the PARAFAC model for the estimated scores is:

$$\tilde{x}_{ijk} = \sum_{f=1}^{F} (c_{kf} b_{jf}) a_{if} \tag{6}$$

where: $a(i,f)$ are the genotype scores; $b(j,f)$, the environment weights and $c(k,f)$ the attribute weights; and F is the number of factors. As is evident from (6), weights and scores only vary with one mode at a time. Each attribute and each environment weights the genotype scores irrespective of the value for the other mode. Thus, for each attribute and in each environment, the score vectors are parallel; hence the name of the technique. It is not immediately obvious from (6), but the model implies that the genotype factors have the same correlations in each environment. The parameters in the model are determined, and there is no transformational freedom, as in ordinary factor analysis.

The PARAFAC model has been implemented in a computer program called PARAFAC, and is available from the author, Dr. Harshman. Various applications have been published; see Harshman and Lundy (1984) for references.

Three-mode principal-component analysis

In contrast with PARAFAC, where factors were derived for the genotypes and weights for the other two modes, in Three-mode PCA components are derived for each of the modes. Each has its own number of components (P, Q, and R), and these components can be interpreted separately. Generally, the emphasis is not so much on the dimensional interpretation, but rather on the data-reduction aspect of the technique. This is more so because the derived axes may be nonsingularly transformed without loss of model fit. Thus, this approach generalizes a two-mode analysis in which sets of vectors span the vectors of the first few principal components, but need not coincide with them. In addition to the components for each mode, the model also contains parameters $g(p,q,r)$ which weight combinations of components of the three modes. Formally, the model becomes:

$$\tilde{x}_{ijk} = \sum_{p=1}^{P} \sum_{q=1}^{Q} \sum_{r=1}^{R} a_{ip} b_{jq} c_{kr} g_{pqr} \tag{7}$$

The $a(i,p)$, $b(j,q)$, and $c(k,r)$ are the component coefficients for the genotypes, environments and attributes, respectively. When a $g(p,q,r)$ is large compared with other weights, that combination of the pth, qth, and rth component is far more important in estimating the data values than when it is small. Therefore, these weights can be used to select the component combinations for interpretation. In this application, we will only use these weights implicitly to construct more easily interpretable indices, and not discuss them explicitly.

Without going into any details, it can be shown that it is possible to portray

the relationships between the genotypes and attributes for each component of the environment (or the genotypes and environments for each component of the attributes) in one plot. Given an interpretation of an environment component, such a plot indicates which genotypes have comparatively high or low scores on which attribute for that environment component (see also below).

Three-mode PCA has been applied to soybean data (Kroonenberg and Basford, 1989) and to cotton data (Basford et al., 1990). These papers contain more details on the application of this technique to agronomic data and the interpretation of the results. Other applications of Three-mode PCA have been referenced in Kroonenberg (1983b). Computer programs implementing the model have been written by, and may be obtained from, the second author (P.M.K.).

ILLUSTRATION: DATA FROM 1981/82 ACCT

Experimental details

The Australian Cotton Cultivar Trials (ACCT) have been operating since 1974/75 at six to eleven locations per year throughout the major cotton-growing districts in New South Wales and Queensland. In any given year, from 16 to 30 cotton lines are evaluated by measuring lint yield (t/ha) and other lint-quality characteristics, the most important of these being lint strength (g/tex), lint micronaire (combined measure of fibre diameter and maturity), and lint length (inches). Details of the trials, entries and locations are contained in Reid et al. (1989).

In the 1981/82 growing season, the eight locations used in the ACCT were (from north to south) Biloela, Theodore, Darling Downs, St. George, Mooree, Myall Vale, West Namoi, and Warren (Fig. 1). The 25 cotton (*Gossypium hirsutum* (L.)) lines planted are listed in Table 1; the industry standard at the time was dp61. The individual experiments were randomized complete-block designs in Queensland and square-lattice designs in New South Wales, each with three replications per location. Using lint yield and the above lint-quality characters, mean performance can be tabulated in a three-way array, 25 lines (referred to as genotypes) by eight locations (referred to as environments) by four attributes, which plant breeders must interpret.

Organisation of analyses

We have briefly described methods to analyse the above data, either in their G×G×E form or their G×E×A form. Two cluster procedures (INDCLUS, MIXCLUS3) and three ordination procedures (INDSCAL, PARAFAC, Three-mode PCA) have been outlined. As our purpose is to provide an overview of three-way methods, the results of all five analyses cannot be presented in depth. We intend to give the flavour of the results, and comment on some comparisons between the techniques. Such comparisons are necessarily

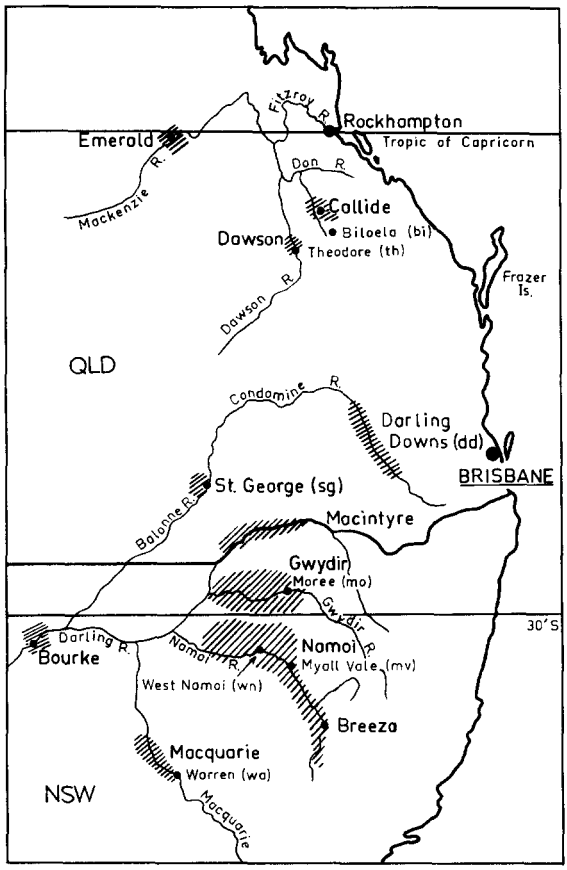

Fig. 1. The eleven locations which represent the major cotton-growing districts in eastern Australia used for the Australian Cotton Cultivar Trials (ACCT).

TABLE 1

Membership and genetic origin of genotypes (from Reid et al., 1989) according to the four cluster MIXCLUS3 solution

	Genotype	*N*	Gentic origin[1]
A)	nam,c310,c315,mo63j	4	UQ,UE,UE,U
B)	m220,10/4,75007	3	UE,A,AS
C)	286f,42/8,37/10,76023	4	AD,A,A,AS
D)	dp61,dp61i,dp16,dp55,dp41,7146n	14	UD,UD,UD,UD,UD,UD
	sicl,siclf,sic2,sic3		AD,AD,AD,AD
	1h,439g,439h		AD,AD,AD
	33/8		A

[1]U, U.S.A.; A, Australian; Q, Quality; E, Eastern; S, Short-season; D, Deltapine.

limited, because only one dataset is considered. To avoid repetition, we have structured the presentation as follows: Firstly, we will discuss the two, rather different, cluster results. Secondly, we will present the Three-mode PCA incorporating the cluster results, primarily because of personal familiarity. These results will then be supplemented with those from the other ordination techniques to illustrate differences and similarities.

1: Clustering

Mixture method of clustering (G×E×A data; non-overlapping clusters)

The mixture method of clustering requires that the underlying number of groups or clusters be specified. Determination of the appropriate number to best represent the data is not straightforward, and much research is being conducted in this area; see, for instance, McLachlan and Basford (1988, section 1.10). Approximate tests on the loglikelihood values indicated that a significant extra amount of variation was being accounted for by going from two to three to four to five to six clusters. However, subjective assessment of the estimated probabilities of group membership, the rate of increase in the loglikelihood values, and because of less-attractive matching of the five and six group solutions with the ordination results, the four-cluster solution was chosen to be presented here. The membership of these groups and genetic origin of the genotypes (from Reid et al., 1989) are given in Table 1.

The four clusters (Table 1 and Fig. 2) had, for each attribute, distinct properties and distinct patterns of response across the environments. The properties and response patterns for the clusters reflected different selectional and genetic backgrounds of the entries within them. All clusters have variable performance in yield across the environments, with the largest cluster (D) having the highest yield in most environments. This cluster consists almost exclusively of genotypes with the Deltapine germplasm, and has relatively weak, reasonably long lint, of average fineness. Cluster A consists of Namcala- and Coker-derived entries of U.S. origin, with strong, long, and reasonably fine lint. Cluster C consists of a mixture of Australian varieties of short, weak, coarse-quality lint. Finally, cluster B is one of mixed genetic origin, and has the finest-quality, reasonably strong, but short lint.

In applying this technique, one can choose a common covariance matrix between the attributes for all groups, or unrestricted covariance matrices for the individual groups. As the latter choice seems more natural, this approach was taken. The estimated covariance matrices for each of the groups is not presented here, and comments are only made about those correlations which had an estimated absolute value greater than 0.4. Groups A, B and C had negative correlation (-0.4, -0.7, and -0.4, respectively) between strength and micronaire. Group A had a negative correlation (-0.6) between micron-

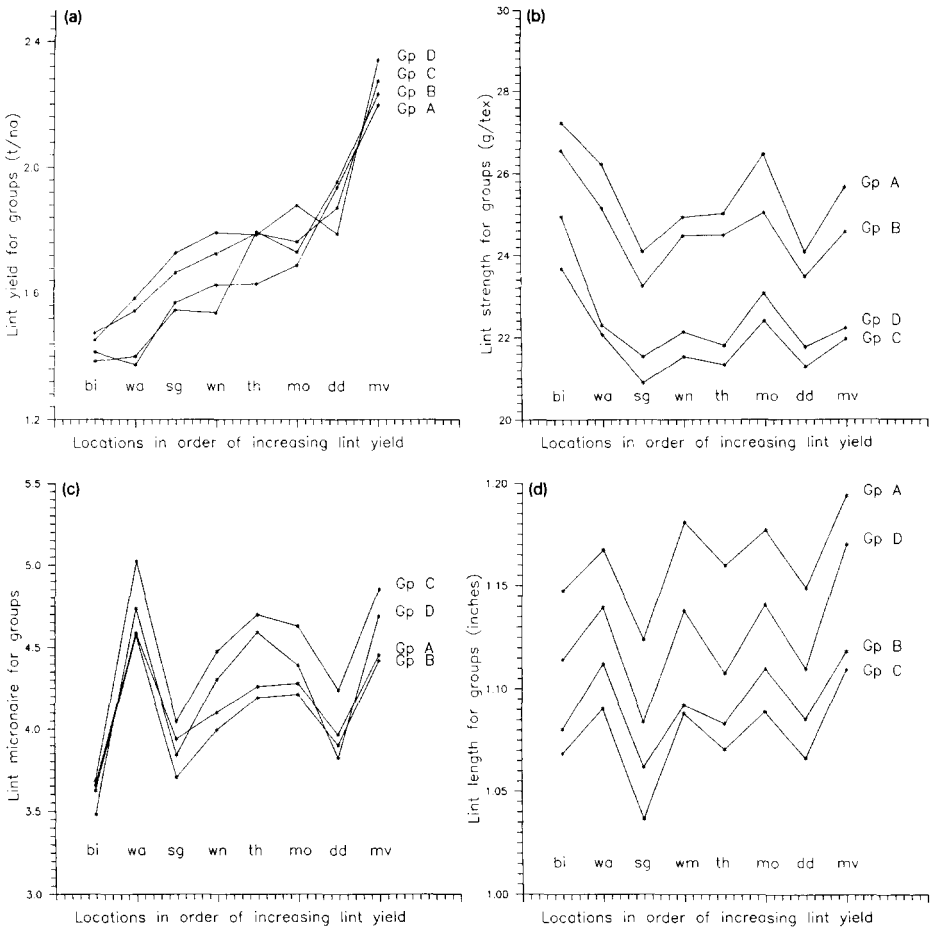

Fig. 2. The expected means for four groups formed by MIXCLUS3 for lint yield and three lint-quality attributes plotted against locations. (For environment, here location, abbreviations see Fig. 1).

aire and length, while Group B had a positive correlation (0.5) between these attributes. Group B also had positive correlations between yield and micronaire (0.7), and yield and length (0.6), and a negative correlation between yield and strength (-0.4).

From Fig. 2 it becomes obvious that there is not much G×E interaction, except for yield, and possibly micronaire. Far more G×A interaction can be observed, i.e. clusters perform differently on different attributes. This is confirmed by the different group-correlation structures outlined above.

Individual-differences clustering (G×G×E data; overlapping clusters)

As with the mixture method, the number of clusters has to be chosen beforehand, and the optimal solution for that number of clusters sought. In all

analyses, the environment weights were restricted to be non-negative, because negative weights have no substantive interpretation. The environments were treated as 'matrix conditional', i.e., they were standardized separately. Due to cost limitations, no comparisons with other options were made. The program was run on a mainframe IBM 3083 in The Netherlands, where computing costs were exorbitant compared with the ordination programs. (The mixture cluster analyses were run in Australia on an IBM mainframe on which time was free, thereby preventing cost comparisons.)

Given our limited experience with this method, choosing the optimal number of clusters was far from easy. The only thing that increases systematically with the increase in the number of clusters from four to five to six to seven is the overall fit (variance accounted for). The fit for the separate environments varied with the number of clusters; that for the seven-cluster solution is shown in Table 2 for comparison with the four-cluster one. The various cluster solutions did not always converge within the specified number of iterations, and sometimes showed one or more instances of negative variances explained. The weights of the clusters for each environment for the four-cluster solution are given in Table 2 with the actual cluster composition in Table 3.

From Table 3, the overlap of the clusters is immediately apparent. Clusters III and IV both contain the 14 genotypes of cluster D from the MIXCLUS3 solution, as well as two additional (42/8 and m220). In addition, clusters III

TABLE 2

Cluster weights of environments (INDCLUS four cluster solution), and fit of INDCLUS solution for four and seven clusters

Environment	Cluster weights[1]					Fit of solutions	
	Clusters					Number of clusters	
	I	II	III	IV	T^2	4	7
Warren	1.16	0.81	1.04	1.40	−1.74	0.59	0.63
Moree	0.91	0.79	1.11	1.25	−1.65	0.55	0.64
Biloela	0.37	0.63	1.38	0.63	−1.31	0.51	0.61
Theodore	0.63	0.88	1.15	1.04	−1.50	0.50	0.59
Myall Vale	0.95	0.64	1.30	0.81	−1.44	0.49	0.62
St. George	0.39	0.67	1.32	0.25	−1.01	0.42	0.48
West Namoi	0.56	0.56	0.89	0.56	−0.99	0.24	0.44
Darling Downs	0.67	0.71	0.90	0.43	−0.92	0.22	0.24
Number of members	8	7	19	21	25		
Overall fit						0.44	0.53
Density of solution						0.55	0.55

[1]A density of 1.00 indicates all genotypes in all clusters; a density of 0.00 indicates each genotype is its own cluster.

[2]T is the additive constant.

TABLE 3

Membership genotypes according to the four cluster INDCLUS solution

Cluster	Genotype	*N*	Genetic origin[1]
I)	nam,c310,c315,m220,mo63j,	8	UQ,UE,UE,UE,A,AS
	10/4,75007		A,AS (=A+B)[2]
	dp55		UD (not A or B)
II)	286f,42/8,37/10,76023	7	AD,A,A,AS (=C)
	75007,dp61,sicotlf		AS,UD,AD (not C)
III)	dp61,dp61i,dp16,dp55,dp41,7146n	19	UD,UD,UD,UD,UD,UD
	sicl,siclf,sic2,sic3		AD,AD,AD,AD
	1h,439g,439h		AD,AD,AD
	33/8		A (=D)
	42/8,m220		A,UE (III&IV)
	c315,75007,286f		A,AS,AD (not IV)
IV)	dp61,dp61i,dp16,dp55,dp41,7146n	21	UD,UD,UD,UD,UD,UD
	sicl,siclf,sic2,sic3		AD,AD,AD,AD
	1h,439g,439h		AD,AD,AD
	33/8		A (=D)
	42/8,m220		A,UE (III&IV)
	10/4,37/10,c310,mo63j,76023		A,A,UE,UE,AS
			(not III)

[1]U, U.S.A; A, Australian; Q, Quality; E, Eastern; S, Short-season; D, Deltapine.
[2]The (A), (B), (C), and (D) refer to the MIXCLUS3 clusters (see Table 1).

and IV each have three and five genotypes, respectively, which are not contained in the other cluster. The five and seven genotypes which III and IV have over and above those of D are all contained in either I or II, or both, while I and II have only one genotype in common.

A reasonable explanation for overlap of clusters in two-way data is that similarity is multidimensional, and that genotypes are similar to each other on different attributes, and therefore can be similar to members of different clusters. In the three-way case, the situation gets even more complex, because genotypes may be similar to different genotypes in different environments. The two large clusters, III and IV, seem to indicate this especially. In some environments, c315, 75005, and 286f are more similar to the deltapine genotypes, while in other environments this is true for 10/4, 37/10, c310, mo63j, and 76023, and in yet other environments it is a bit of both. For instance, the seven-cluster solution shows three big clusters of 20 genotypes each, with an intersection of fourteen, and a medium-sized cluster of 13, with an intersection of nine with the larger clusters.

The cluster weights (Table 2) can be used to evaluate the importance of the clusters for each environment. For most environments, cluster III contributes more to the solution than IV. However, in Warren, cluster IV is noticeably more important. In Biloela and St. George, the prominence of cluster III is particularly clear, and this is probably also reflected in the lower cohesion

of the cluster-I genotypes. Note that the cluster structure is a poor reflection of the situation in Darling Downs and West Namoi, as the clusters found can explain only 22% and 24% of the variability, respectively. For Darling Downs, the situation is not much improved when seven clusters are derived.

MIXCLUS3 and INDCLUS

It is clear that two cluster techniques carry different information about the outcome of the cotton trials. MIXCLUS3 forces a single-cluster solution even if there are differences in cluster composition across environments; but, it gives more information about the behaviour of the clusters in terms of the attributes (Fig. 2). One is able to evaluate the clusters in terms of interest to plant breeders. After the INDCLUS clusters have been derived, attribute means for the clusters can be computed, but in MIXCLUS3, clusters have been derived so that the differences in the cluster means are optimized in a somewhat similar fashion to discriminant analysis. It is therefore to be expected that the INDCLUS version of Fig. 2 would not be as neat.

INDCLUS has the advantage of allowing overlapping clusters, which, in the extreme, allows each environment to have its own arrangement of the genotypes, and does not necessitate forced allocations. It points to differences in cluster composition in the environments, and suggests places to look for the nature of those differences. However, one has to go beyond the cluster procedure to provide the necessary information. On the other hand, INDCLUS can be used in studies where similarities are collected directly, unlike MIXCLUS3 which requires the original G×E×A data.

2: Ordination

In contrast to clustering methods, where the number of clusters must be chosen, the number of dimensions must be decided on when performing ordination techniques. In our view, the number of dimensions used should be determined by the detail with which one wants to examine the data. This is in contrast to the view that a search should be made for the 'correct' number of dimensions. Our approach can be compared to the 'correct' magnification required when using a microscope, where the general rule is to use the lowest magnification compatible with observing the phenomena of interest. Too large a magnification confuses the overall picture with detail; too small a magnification obscures the interaction. For the ordination analyses presented here, we have used a fairly low magnification, since we need a fairly global interpretation with some detail, as our major purpose is to compare and illustrate the various methods.

Three-mode principal component analysis – Three-mode PCA (G×E×A data)

The choice of number of dimensions of Three-mode PCA is more complicated than in most techniques, because the number of components has to be determined for all three modes. After examining several solutions, it was decided that either a 2×1×2-solution, i.e. two components for the genotypes, one for the environments and two for the attributes, could be used with 53% variation accounted for, or else a 4×2×4-solution with 72% variation accounted for. The former, however, reduces the differences between environments to proportionality of the G×A interaction, i.e. eliminating all G×E interaction. The solution is virtually indistinguishable from an analysis of the 25×4 G×A matrix averaged over environments. It was noted in the MIXCLUS3 analysis (Fig. 2) that there was very little G×E interaction because the curves were largely parallel. This 2×1×2-solution would be equivalent to making the cluster profiles completely parallel. The alternative, i.e., the 4×2×4-analysis, has the advantages that more detail becomes available, and that differences between environments can be investigated. For this particular dataset, the 2×1×2-solution is roughly nested in the 4×2×4, so that both the global and the local picture can be examined at the same time.

In Tables 4, 5, and 6 the (orthogonal) components of the environments, the attributes, and the genotypes are presented for the Three-mode PCA so-

TABLE 4

Environment components[1] from the ordination analyses: Three-mode PCA, PARAFAC, and INDSCAL/ALSCAL

Environment	Three-mode PCA			PARAFAC				INDSCAL/ALSCAL			
	2×1×2	4×2×4		Four factors				Four dimensions			
	1	1	2	1	2	3	4	1	2	3	4
Darling Downs	0.67	0.77	−2.11	0.48	0.94	−0.36	0.99	0.28	0.79	0.82	0.88
Biloela	0.80	0.62	−1.29	0.64	0.98	0.45	0.66	0.59	1.13	1.10	1.38
St. George	0.89	0.95	−0.36	0.83	0.90	1.05	1.15	0.63	1.26	0.73	1.92
Warren	1.13	1.13	0.07	1.25	0.94	0.87	1.38	1.37	0.59	0.65	0.68
Moree	1.06	1.10	0.17	1.05	1.02	1.12	1.26	1.35	1.10	0.94	0.22
Theodore	1.11	1.06	0.38	1.26	1.01	0.45	0.89	1.43	0.69	0.45	0.44
Myall Vale	1.15	1.18	0.86	1.14	1.11	1.48	0.97	1.04	1.08	1.32	0.66
West Namoi	1.07	1.05	0.91	1.05	1.09	1.46	0.07	0.46	1.14	1.53	0.69
R^2	0.54	0.68	0.04	0.27	0.25	0.10	0.09	0.33	0.12	0.11	0.07

[1]All columns are scaled such that the sum of squares is equal to 8. The Three-mode PCA values were multiplied by $\sqrt{8}$; the INDSCAL/ALSCAL values by $1/w$ (dimension weight). The R^2 values of PARAFAC are the squared root-mean-squared contributions which can be used because of the orthogonality of the factors.

TABLE 5

Attribute components[1] from the ordination analyses: Three-mode PCA, and PARAFAC

Attribute	Three-mode PCA						PARAFAC			
	2×1×2		4×2×4				Four factors			
	1	2	1	2	3	4	1	2	3	4
Length	0.49	0.86	0.47	0.86	−0.03	−0.17	−0.00	0.94	0.05	−0.11
Micronaire	−0.41	0.15	−0.44	0.22	0.84	−0.24	−0.46	−0.14	0.00	0.58
Strength	0.71	−0.37	0.69	−0.29	0.55	0.38	0.89	0.32	−0.05	0.08
Yield	−0.30	0.32	−0.33	0.35	−0.01	0.88	−0.31	−0.09	0.64	0.04
R^2	0.37	0.18	0.38	0.18	0.08	0.08	0.27	0.25	0.10	0.09

[1]As the signs of the components are largely arbitrary, they have been oriented in this Table so that the largest value has a positive sign.

lutions, as well as those for the other ordination analyses to be discussed. For the two Three-mode PCA analyses mentioned above, the first component of the environments and the first two of the attributes are very much alike (Tables 4 and 5). This is true for the genotypes as well, but the 2×1×2 components have not been included in Table 6. On their first component, the environments are largely equal, with lower scores for Darling Downs and Biloela. This indicates that the variability of the genotypes over attributes is largely the same in all environments (see Fig. 2). The second component of the environments shows a sharp contrast between Darling Downs and Biloela on the one hand, and Myall Vale and West Namoi on the other.

The need for all four components to describe the variability between the attributes implies that they do not show intense correlation (also evident in the MIXCLUS3 analysis). Because we are dealing with three-mode data, it is not 'improper' to use as many components as there are variables. It simply means that no condensation is necessary or fruitful for that mode. A detailed discussion of the genotypes will be undertaken in conjunction with the attribute components. It is worth noting, from Table 6, that the MIXCLUS3 cluster structure can be observed from the first two genotype components.

It would be useful to express, in a graphical form, the relationships between the genotypes in terms of the attributes, as was done in Fig. 2. This is done in two parts, one for each of the two environment components. The first environment component indicates what the environments have in common, and the second concentrates on the 'true' G×E×A interaction with environment differences reflected in the Biloela/Darling Downs compared with Myall Vale/West Namoi contrast. As explained in Kroonenberg and Basford (1989), an attractive way to present these relationships is by joint plots of the genotype and attributes, one for each environment component (here we include only

TABLE 6

Genotype[1] components from the ordination analyses: Three-mode PCA, PARAFAC, and INDSCAL/ALSCAL

Geno-type	Three-mode PCA				PARAFAC				INDSCAL/ALSCAL			
	Four components				Four factors				Four dimensions			
	1	2	3	4	1	2	3	4	1	2	3	4
nam	−2.85	−0.85	0.05	−1.95	3.23	1.03	0.75	0.94	−3.12	0.94	−0.01	0.31
mo63j	−1.50	−0.20	1.95	0.40	0.64	1.31	−1.68	0.93	−1.10	1.71	−0.73	−0.43
c310	−1.35	1.60	1.10	1.40	−0.58	2.36	−1.55	−0.62	−0.53	2.25	0.24	0.80
c315	−1.25	0.90	0.45	−1.10	0.59	1.56	0.41	0.55	−0.98	1.41	0.68	−0.78
10/4	−1.25	−2.10	−1.70	0.85	2.31	−1.29	−0.48	−1.46	−1.47	−0.66	−1.01	2.11
75007	−0.30	−2.10	−0.35	0.40	1.20	−1.43	−1.04	−0.41	−0.71	−0.72	−1.95	−0.04
m220	−0.30	−0.65	0.25	−0.85	0.62	−0.18	−0.20	0.76	−0.36	0.27	−1.10	−0.79
dp16	−0.10	0.40	−0.70	1.70	−0.36	0.19	−0.44	−1.40	0.21	0.18	0.04	2.03
sic3	−0.05	−0.15	0.50	1.65	−0.43	0.06	−1.45	−0.58	0.18	0.83	−1.47	−0.28
dp6li	−0.05	1.10	−0.05	0.35	−0.53	0.74	0.26	−0.58	−0.02	0.20	1.41	0.74
dp55	−0.05	0.05	−0.85	0.05	0.12	−0.18	0.46	−0.87	−0.56	−0.74	0.45	−1.05
sic2	0.15	0.80	−1.10	−0.60	−0.19	0.28	1.40	−0.24	0.16	−0.06	1.44	−0.82
1h	0.20	1.30	−0.65	−0.30	−0.63	0.69	1.09	−0.42	0.25	0.31	1.47	−0.42
dp41	0.25	0.50	−0.85	1.10	−0.59	−0.05	0.04	−1.48	−0.02	−0.87	0.87	1.08
7146n	0.30	0.35	0.05	0.90	−0.68	0.05	−0.35	−0.74	0.42	−0.18	−0.03	1.61
439g	0.35	0.20	−1.05	0.15	−0.21	−0.31	0.88	−0.73	−0.16	−1.10	0.73	−0.63
33/8	0.35	0.85	0.75	0.35	−0.82	0.52	−0.17	0.22	0.73	0.75	0.21	0.99
siclf	0.50	0.15	1.30	−0.65	−0.57	0.19	−0.37	1.56	0.55	0.45	−0.22	−1.70
439h	0.55	0.75	−2.30	−0.85	−0.25	−0.22	2.40	−1.01	0.29	−1.03	1.84	−0.05
sicl	0.65	0.90	0.15	−0.60	−0.80	0.40	0.67	0.77	1.20	0.53	0.54	0.23
dp61	0.80	0.40	0.25	−1.60	−0.38	−0.16	1.06	1.22	1.19	0.28	0.32	−0.54
42/8	0.65	−1.45	−0.45	−0.00	0.38	−1.60	−0.06	0.02	0.10	−1.48	−0.97	0.38
76023	1.05	−1.25	1.45	1.20	−0.85	−1.28	−1.88	0.03	0.50	−1.29	−1.57	−1.11
286f	1.10	−0.65	1.25	−1.35	−0.33	−0.75	0.02	2.46	1.63	−0.04	−0.75	−0.30
37/10	2.15	−1.00	0.50	−0.60	−0.99	−1.95	0.22	1.09	1.62	−1.92	−0.44	−1.36
R^2	0.38	0.19	0.09	0.06	0.27	0.25	0.10	0.09	0.33	0.12	0.11	0.07

[1]The genotypes have been arranged in order of increasing value of the first component of the Three-mode PCA within each of the groups A, B, D and C from the MIXCLUS3 analysis.

the one for the first environment component which indicates what the environments have in common – Fig. 3), and/or the inner-products of the attributes and genotypes in the space displayed by the plots (Tables 7 and 8).

Figure 3 (top) and (bottom) show the first against the second dimension, and the third against the fourth dimension, respectively. The genotypes have been labelled according to the membership of four-cluster MIXCLUS3 solution, and in Fig. 3 (top) this group composition has been further highlighted. It appears that the clusters arise primarily because the genotypes have similar

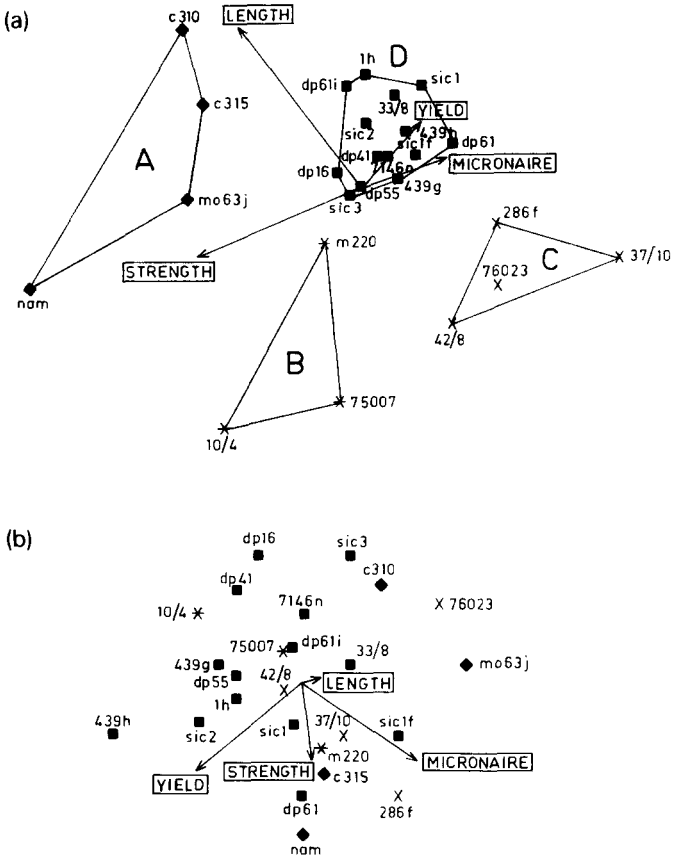

Fig. 3. Common structure for all environments: Top, Joint plot of first and second components of genotypes and attributes for first environment component from Three-mode PCA 4×2×4-solution (54% explained variation). A, B, C and D are the clusters from a four-cluster MIXCLUS3 analysis. (♦), cluster A; (*), cluster B; (×), cluster C; (■), cluster D.
Bottom, Joint plot of third and fourth components of genotypes and attributes for first environment component from Three-mode PCA 4×2×4-solution (13% explained variation). A, B, C and D are the clusters from a four cluster MIXCLUS3 analysis. (♦), cluster A; (*), cluster B; (×), cluster C; (■), cluster D.

profiles on the attributes in all environments, and that lint length and lint strength contribute more to the separation between clusters than micronaire and lint yield. This is also evident in Fig. 2, where the cluster profiles are not as clearly distinguishable on the latter two variables. Yield and micronaire contribute more to the differences in the third and fourth dimensions (Table 5), and this is reflected in Fig. 3 (bottom) by their longer arrows. The higher dimensions bear no relationship to the four-cluster solution. Even though the plots are helpful in interpreting comparisons between genotypes based on the

attributes, they are difficult to use. This is especially so given that a four-dimensional space is required.

The inner products between (the vectors of) each genotype and attribute (Table 7) are a more usual interpretational device for Three-mode PCA, and more experience has been gained with their use. It is reasonably easy to see the consistency between this Table and the grouping obtained from the

TABLE 7

Inner products between genotypes and attributes.[1] First environment component (within cluster ordered with respect to yield)

Cluster[2]	Genotype	Length	Strength	Micronaire	Yield	Selected 1981[3]
A)	c315	4.1	3.3	−0.3	0.1	yes
	nam	2.7	9.3	−2.9	−2.0	yes
	c310	6.0	1.0	−1.2	−2.5	
	mo63j	2.8	3.5	0.1	−3.8	
B)	m220	−0.8	1.9	0.3	−0.3	yes
	75007	−3.9	2.1	−1.5	−2.0	yes
	10/4	−3.0	4.0	−4.9	−2.3	
C)	37/10	−5.3	−4.1	3.9	1.7	
	286f	−2.7	−1.0	3.9	0.8	
	42/8	−4.2	−0.5	0.1	−0.0	
	76023	−3.6	−2.7	2.3	−2.2	
D)	439h	−0.2	−1.2	−1.4	4.0	yes
	dp61	−0.5	−0.9	2.5	2.4	yes
	sic2	0.9	−0.5	−0.6	2.3	yes
	1h	2.1	−1.2	−0.1	2.0	yes
	sicl	1.0	−1.8	1.7	1.9	yes
	439g	−0.6	−1.2	−0.8	1.3	yes
	dp55	−0.2	0.0	−1.2	0.7	yes
	dp6li	2.3	−1.0	−0.1	0.6	
	dp41	0.4	−1.9	−1.3	0.3	yes
	siclf	−0.0	−0.8	2.7	0.1	yes
	33/8	1.5	−1.9	1.4	0.1	
	7146n[4]	0.4	−1.8	0.0	−0.2	
	dp16	0.8	−1.4	−2.0	−0.8	yes
	sic3	0.1	−1.1	−0.6	−2.1	

[1]A value of zero indicates average on an attribute.

[2]The clusters, from the four-cluster MIXCLUS3 analysis, may be characterised as:

A) long, strong lint, rather fine micronaire, low yield;

B) short, strong lint, rather fine micronaire, low yield;

C) weak, short lint, coarse micronaire, variable yield;

D) average length, weak lint, variable micronaire, generally good yield.

[3]Yes in column 1981 means selected from 1981/82 trials.

[4]Probably 7146n is the closest to an 'average' cotton plant.

TABLE 8

Inner products between genotypes[1] and attributes: Second environment component, i.e. West Namoi/Myall Vale versus Biloela/Darling Downs (within cluster ordered with respect to yield)

Cluster	Genotype	Length	Strength	Micronaire	Yield	Selected 1981
A)	nam	−0.7	−0.4	−1.9	−1.0	yes
	mo63j	−0.3	−0.0	−1.1	−1.5	yes
B)	75007	0.3	−1.0	0.0	−1.0	yes
C)	76023	0.6	0.4	−0.5	−1.1	
D)	439h	−0.1	−0.2	1.1	1.6	yes
	sic2	−0.1	−0.1	0.7	1.0	yes
	1h	−0.2	−0.1	0.9	1.0	yes

[1]Only genotypes with at least one value over |1.0| included.

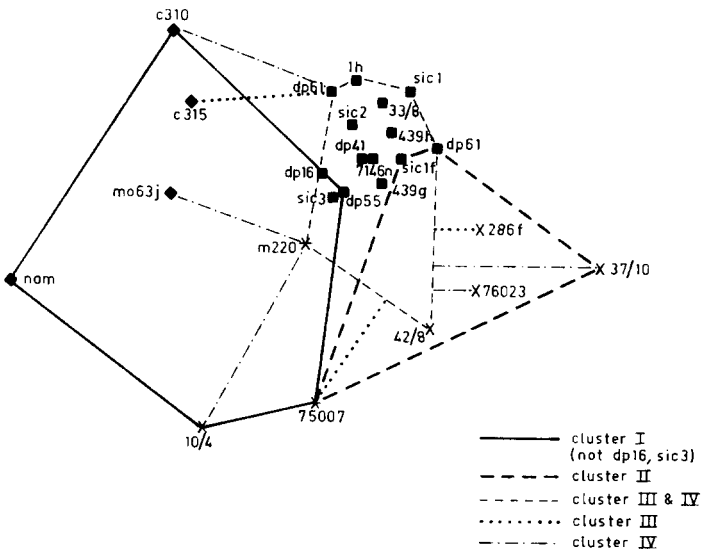

Fig. 4. As for Fig. 3 (top), but with four-cluster INDCLUS solution.

MIXCLUS3 analysis, particularly for lint length and lint strength. Attention is also focussed on any genotypes which may have a somewhat different response pattern to the rest of the group to which it was assigned in the cluster analysis. For instance, sicl, siclf, dp61 and 33/8 stand out in cluster D because of coarse micronaire, while 1h and dp61i have particularly long lint for that group.

A similar Table (8) has been prepared for the genotype/attribute relations corresponding to the environment differences between Darling Downs and Biloela in Queensland and the Namoi locations, Myall Vale and West Namoi,

in New South Wales. Compared with their overall performance in all environments, nam, mo63j, 75007 and 76023 had relatively lower yields in West Namoi/Myall Vale than in Biloela/Darling Downs, while 439h, sic2 and 1h had relatively higher yields in West Namoi/Myall Vale than in Biloela/Darling Downs. Furthermore, nam and mo63j had relatively finer lint in West Namoi/Myall Vale than in Biloela/Darling Downs, and 75007 was weaker in West Namoi/Myall Vale than in Biloela/Darling Downs.

It is instructive to display Fig. 3 (top) again, but with the four-cluster INDCLUS results portrayed (Fig. 4) instead of the MIXCLUS3 results. As remarked earlier, the many similar clusters in the INDCLUS solution probably represent differences in performance across environments, i.e., some genotypes performed more alike in some environments compared with others so that for certain environments they belong to the main (Deltapine) cluster, while in others they do not.

Parallel factor analysis – PARAFAC (G×E×A data)

Unlike Three-mode PCA, PARAFAC has only one set of (genotype) components (or 'factors', as they are called by the originator (Harshman, 1970)), and the elements (i.e. attributes and environments) weight these axes according to the importance of that factor to the element in question. As in ordinary factor or component analyses, the interpretation of the axes is primarily derived from the attributes (variables) and the factors are interpreted by themselves rather than by investigating the space they span (as for Three-mode PCA), even though that remains a distinct possibility; moreover they are generally (but not here) oblique. (The full rationale for the interpretation is clearly explained in Harshman and Lundy, 1984.) The model is conceptually simpler (only one kind of factor, rather than three) than Three-mode PCA, and therefore often more easily interpreted.

To gain an impression of the interpretation, we will look at the PARAFAC results (Tables 4, 5 and 6) but, for simplicity, only consider those genotypes which have values equal to or greater than 1.0 in absolute value. This provides an oversimplified picture, but is unavoidable in a paper of this kind.

Factor 1: Nam (3.2), 10/4 (2.3) and 75007 (1.2) stand out in all environments (but less so in Biloela (0.6), Darling Downs (0.5) and St. George (0.8)) in that they have particularly strong lint (0.9) and fine micronaire (−0.5), yet low yield (−0.3) and average length (−0.0). The reverse is true for 37/10 (−1.0).

Factor 2: c310, c315, nam and mo63j have particularly long lint of above-average strength in all environments, whereas 37/10, 42/8, 75007, 10/4 and 76023 have particularly short lint of below-average strength in all environments.

Factor 3: 439h, sic2, dp61 and 1h have particularly high yields in Myall Vale, West Namoi, Moree and St. George, and low yields in Biloela, Theodore and, especially, Darling Downs. The situation is reversed for 76023, mo63j,

c310, sic3 and 75007 having low yields in Myall Vale, etc., and high yields in Biloela, etc.

Factor 4: dp41, 10/4, dp16 and 439h have particularly fine micronaire (negative values) in Warren, Moree and St. George, while 286f, siclf, dp61 and 37/10 have coarse micronaire in those environments. The reverse is true for West Namoi.

Without going into a full comparison of the Three-mode PCA and PARAFAC results, there are several points that should be noted. The sum of the R^2 values of the PARAFAC and the Three-mode PCA solutions are almost equal (0.72 and 0.71), and Table 9 (to be discussed below) shows that the PARAFAC genotype factors can be predicted quite well from the Three-mode PCA genotype components. Thus, the two models predict the same variability, but organise their information in different ways. Moreover, the first two PARAFAC factors essentially span the same space as the first two Three-mode PCA components, i.e., the former are a rotation of the latter. The same can be said of the third and fourth factors (components) of the two models. Both the PARAFAC factor descriptions and the inner-product descriptions come to essentially the same conclusions.

The environment factors in Table 4 show that the models present the differences between environments in another way. Three-mode PCA has two components, one to show what the environments have in common and one to show what their major differences are. In PARAFAC, such differences are represented in the different weights the environments attach to the factors. Because the values are all positive (except one), the trend is the same for all

TABLE 9

Regression of PARAFAC AND INDSCAL/ALSCAL genotype coordinates[1] on Three-mode PCA genotype components

	Predictors[2]	Criteria							
		PARAFAC				INDSCAL/ALSCAL			
		1	2	3	4	1	2	3	4
b	T3/1	−0.75	−0.60	0.17	0.16	0.92	−0.63	0.02	−0.30
	T3/2	−0.54	0.76	0.34	−0.12	0.22	0.50	0.80	0.04
	T3/3	0.21	0.25	−0.64	0.67	0.17	0.48	−0.45	−0.32
	T3/4	−0.28	−0.01	−0.65	−0.69	0.00	−0.02	−0.25	0.49
R^2		0.99	1.00	0.98	0.96	0.93	0.88	0.90	0.43

[1]Due to centering of the data, all axes have zero means, and thus all regression constant terms are zero.

[2]b is the unstandardized regression weight.

R^2 is the squared multiple correlation between criterion and predictors.

T3/i is the ith genotype component of the Three-mode PCA.

environments; only the extent of the trend is different (the inner products of the factors, which indicate the cosines between them, range from 0.84 to 0.97). In this way, the similarities and differences between the environments are represented in all factors.

Individual Differences Scaling – INDSCAL (G×G×E)

Unlike the previous two techniques, INDSCAL starts from (dis)similarity matrices, so the G×E×A data were transformed to G×G×E data using Euclidean distances. This rather hampers the interpretation, as the present data have a fairly large G×A interaction. The genotype coordinates on the dimensions are given in Table 6 and the weights of the environments in Table 4. As the overall (ALSCAL) fit of the INDSCAL model to the G×G×E data has an R^2 value of 0.62, the four dimensions fit less of the transformed data than the other models do the original data. However, such a comparison is really not justified, as the data being fitted are different. An explicit interpretation of the dimensions will not be given, because the information they carry is largely the same as the genotype components of the previous analyses – as will become clear below. The environment weights provided by the INDSCAL model indicate to what extent the configuration defined by the genotypes is enlarged or reduced by each of the environments, with the direction of extension being along the coordinate axes.

To evaluate the differences between the three ordination techniques, we have regressed the PARAFAC factors and the INDSCAL/ALSCAL dimensions on the Three-mode PCA components (Table 9). As mentioned before, all PARAFAC factors are very well predicted by the components. The agreement is somewhat less for the INDSCAL/ALSCAL dimensions, but even here only the last deviates in a really noticeable manner. The orientation of the dimensions is generally different, and there is not such a direct split into the first two and the last two.

DISCUSSION

The information obtained from the various analyses of the 1981/82 data from the Australian Cotton Cultivar Trials can be summarized as follows:

(1) Both the clustering and ordination procedures gave a sensible and useful integration of the data from this regional variety trial. Considerably more detail and interpretation were available through the complementary use of the ordinations, especially in examining the relationship among, and the variation within clusters. This addresses the practical problem for plant-breeders that, although such clusters are easier to look at than many individual lines, selection has to be made for individual lines.

(2) The methods have successfully integrated the yield and quality data.

The analyses point to a decision in favour of either high yields of moderate to good quality lint or moderate yield but superior lint quality.

Before lines are entered in the ACCT, they have been previously tested in trials at two to three locations for approximately two years. These data, together with the ACCT data, are used to select entries for the next year's trials. From the above analyses, the 'best' members from cluster D would be selected on high yield and adequate quality, and the best from cluster A (and maybe B) on the basis of good quality and reasonable yield. This is consistent with what happened in practice (see Table 7).

As in 1980/81, nam (Namcala) has very strong lint and is among the best lines for long lint and fine micronaire. Although it is included in the trials as a benchmark for high-quality lint, it does not yield enough to be acceptable. The dp61 and sic2 quality is 'good enough' for most 'good' quality cotton. Dp16 is also retained in the trials for genetic reasons.

As mentioned earlier, we cannot give details of the individual problems one might encounter while executing these analyses. The prospective user should look at the original publications for comprehensive information. The relevant programs and documentation are generally in the public domain or available on request.

Although the overlapping clusters gave some additional insight by pointing to differences in cluster composition in the different environments, it is not straightforward to obtain this extra information. The response plots from MIXCLUS3 (Fig. 2) are particularly useful in displaying the differing response patterns of the clusters in the individual environments. When taken in conjunction with the Three-mode PCA, relationships between the lines within the clusters can be explored. The other ordination techniques did not add any further significant information.

The major advantage of these methods is that they allow the data set to be treated in the form of a three-way array. An overall picture of response is obtained and, in the case of the clustering approaches, used to allocate the cotton lines to either overlapping or non-overlapping groups. The important G×E interaction present in such trials is incorporated directly into the underlying models. Similarly, the representation of the cotton lines in a reduced space allows a quicker appreciation of the major differences inherent in the data. The ordination techniques allow possible structure in the environments and attributes to be extracted. The techniques provide complementary information which can be readily displayed in common figures. They are useful techniques which could be commonly employed in the statistical analysis of such three-way data.

ACKNOWLEDGMENT

Research by the second author, P.M.K., was partially supported by a grant from the Netherlands Organization of Scientific Research (NWO).

REFERENCES

Anderberg, M.R.C., 1973. Cluster Analysis for Applications. Academic Press, New York.

Anonymous, 1987. SPSS Statistics Guide. McGraw-Hill, New York.

Arabie, P. and Carroll, J.D., 1980. MAPCLUS: A mathematical programming approach to fitting the ADCLUS model. Psychometrika, 45: 211–235.

Arabie, P., Carroll, J.D. and DeSarbo, W.S., 1987. Three-way scaling and clustering. Sage Publications, Beverly Hills, CA.

Basford, K.E., 1982. The use of multidimensional scaling in analysing multi-attribute genotype response across environments. Aust. J. Agric. Res., 33: 473–480.

Basford, K.E. and McLachlan, G.J., 1985. The mixture method of clustering applied to three-way data. J. Class., 2: 109–125.

Basford, K.E., Kroonenberg, P.M., DeLacy, I.H. and Lawrence, P.K., 1990. Multiattribute evaluation of regional cotton variety trials. Theor. Appl. Genet., 79: 225–234.

Byth, D.E. and DeLacy, I.H., 1989. Genotype by environment interaction and the interpretation of agricultural adaptation experiments. In: I.H. DeLacy (Editor), Analysis of Data from Agricultural Adaptation Experiments. Thai/World Bank National Agricultural Research Project, Bangkok, pp. 186–194.

Byth, D.E., Eisemann, R.L. and DeLacy, I.H., 1976. Two-way pattern analysis of a large data set to evaluate genotypic adaptation. Heredity, 37: 215–230.

Carroll, J.D. and Arabie, P., 1980. Multidimensional scaling. In: M.R. Rosenzweig and L.W. Porter (Editors), Annual Review of Psychology. Annual Reviews, Palto Alto, CA, pp. 607–649.

Carroll, J.D. and Arabie, P., 1982. How to use INDCLUS, a computer program for fitting the individual differences generalization of the ADCLUS model. AT&T Bell Laboratories, Murray Hill, NJ.

Carroll, J.D. and Arabie, P., 1983. INDCLUS: An individual differences generalization of the ADCLUS model and the MAPCLUS algorithm. Psychometrika, 48: 157–169.

Carroll, J.D. and Chang, J.J., 1970. Analysis of individual differences in multidimensional scaling via an N-way generalization of Eckart-Young decomposition. Psychometrika, 35: 283–319.

Carroll, J.D. and Wish, M., 1974. Models and methods for three-way multidimensional scaling. In: D.H. Krantz, R.C. Atkinson, R.D. Luce and P. Suppes (Editors), Contemporary Developments in Mathematical Psychology (Vol. II). Freeman, San Francisco, CA, pp. 57–105.

Clifford, H.T. and Williams, W.T., 1976. Similarity measures. In: W.T. Williams (Editor), Pattern Analysis in Agricultural Science. Elsevier, Amsterdam, pp. 37–46.

DeLacy, I.H., 1981. Analysis and interpretation of pattern of response in regional variety trials. In: D.E. Byth and V.E. Mungomery (Editors), Interpretation of Plant Response and Adaptation to Agricultural Environments. Australian Institute of Agricultural Science, Brisbane, pp. 27–50.

Eisemann, R.L., 1981. Two methods of ordination and their application in analysing genotype-environment interactions. In: D.E. Byth and V.E. Mungomery (Editors), Interpretation of Plant Response and Adaptation to Agricultural Environments. Australian Institute of Agricultural Science, Brisbane, pp. 293–307.

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

Gauch, H.G., 1988. Model selection and validation for yield trials with interaction. Biometrics, 44: 705–715.

Gauch, H.G. and Zobel, R.W., 1988. Predictive and postdictive success of statistical analyses of yield trials. Theor. Appl. Genet., 76: 1–10.

Goodchild, N.A. and Boyd, W.J.R., 1975. Regional and temporal variations in wheat yield in Western Australia and their implications in plant breeding. Aust. J. Agric. Res., 26: 209–217.

Harshman, R.A., 1970. Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-mode factor analysis. UCLA Work. Pap. Phonetics, 16: 1–84. (Reprinted by Zerox University Microfilms, Ann Arbor, MI; order no. 10,085).

Harshman, R.A. and Lundy, M.E., 1984. The PARAFAC model for three-way factor analysis and multidimensional scaling. In: H.G. Law, C.W. Snyder Jr., J.A. Hattie, and R.P. McDonald (Editors), Research Methods for Multimode Data Analysis. Praeger, New York, pp. 122–215.

Horner, T.W. and Frey, K.J., 1957. Methods for determining natural areas for oat varietal recommendations. Agron. J., 49: 313–315.

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

Kiers, H.A.L., 1989. Three-Way Methods for the Analysis of Qualitative and Quantitative Two-Way Data. DSWO Press, Leiden, The Netherlands.

Kroonenberg, P.M., 1983a. Three-Mode Principal Component Analysis: Theory and Applications. DSWO Press, Leiden, The Netherlands.

Kroonenberg, P.M., 1983b. Annotated bibliography of three-mode factor analysis. Br. J. Math. Stat. Psychol., 36: 81–113.

Kroonenberg, P.M. and Basford, K.E., 1989. An investigation of multi-attribute genotype response across environments using three-mode principal component analysis. Euphytica, 44: 109–123.

Kruskal, J.B., 1977. The relationship between multidimensional scaling and clustering. In: J. Van Ryzin (Editor), Classification and Clustering. Academic Press, New York, pp. 17–44.

Lavit, C., 1988. Analyse conjointe de tableaux quantitatifs: Methodes et programmes. Simultaneous Analysis of Quantative Tables: Methods and Programs. Masson, Paris.

McLachlan, G.J. and Basford, K.E., 1988. Mixture Models: Inference and Applications to Clustering. Marcel Dekker, New York.

Miller, K. and Gelman, R., 1983. The child's representation of number: A multidimensional scaling analysis. Child Devel., 54: 1470–1479.

Reid, P.E., Thomson, N.J., Lawrence, P.K., Luckett, D.J., McIntyre, G.T. and Williams, E.R., 1989. Regional evaluation of cotton cultivars in eastern Australia 1974–85. Aust. J. Exp. Agric., 29: 679–689.

Seif, E., Evans, J.C. and Balaam, L.N., 1979. A multivariate procedure for classifying environments according to their interaction with genotypes. Aust. J. Agric. Res., 30: 1021–1026.

Shepard, R.N. and Arabie, P., 1979. Additive clustering: Representation of similarities as combinations of discrete overlapping properties. Psychol. Rev., 86: 87–123.

Soli, S.D., Arabie, P. and Carroll, J.D., 1986. Discrete representation of perceptual structure underlying consonant confusions. J. Acoust. Soc. Am., 79: 826–837.

Takane, Y., Young, F. and De Leeuw, J., 1977. Nonmetric individual differences multidimensional scaling: an alternating least squares method with optimal scaling features. Psychometrika, 42: 7–67.

Tucker, L.R., 1966. Some mathematical notes on three-mode factor analysis. Psychometrika, 31: 279–311.

Zobel, R.W., Wright, M.J. and Gauch, H.G., 1988. Statistical analysis of a yield trial. Agron. J., 80: 388–393.