Jackson, P.A. and Hogarth, D.M. (1992) Genotype × environment interactions in sugarcane. I. Patterns of response across sites and crop-years in North Queensland. *Australian Journal of Agricultural Research* 43, 1447–1460.

Lawrence, P.K. and DeLacy, I.H. (1988) Genotype–environment interactions for yield in cotton – contribution of environments to differential genotype response. In: McWhirter, K.S., Downes, R.W. and Read, B.J. (eds) *Proceedings of the Ninth Australian Plant Breeding Conference*. Organising Committee, Wagga Wagga, Australia (27 June–1 July), pp. 181–182.

Lawrence, P.K. and DeLacy, I.H. (1993) Classification of locations in regional cotton variety trials where trial entries change over years. *Field Crops Research* 34, 195–207.

Lin, C.S. and Butler, G. (1988) A data-based approach for selecting locations for regional trials. *Canadian Journal of Plant Science* 68, 651–659.

Mirzawan, P.D.N., Cooper, M. and Hogarth, D.M. (1993a) The impact of genotype × environment interactions for sugar yield on the use of indirect selection in southern Queensland. *Australian Journal of Experimental Agriculture* 33, 629–638.

Mirzawan, P.D.N., Cooper, M. and Hogarth, D.M. (1993b) The magnitude of genotype by environment interactions for cane yield, sugar yield and CCS in southern Queensland and their impact on selection. In: Imrie, B.C. and Hacker, J.B. (eds) *Focused Plant Improvement: Towards Responsible and Sustainable Agriculture. Proceedings Tenth Australian Plant Breeding Conference*. Vol. 1. Organising Committee, Australian Convention and Travel Service, Canberra, pp. 57–61.

Mirzawan, P.D.N., Cooper, M., DeLacy, I.H. and Hogarth, D.M. (1994) Retrospective analysis of the relationships among the test environments of the Southern Queensland sugarcane breeding program. *Theoretical and Applied Genetics* 88, 707–716.

Peterson, C.J. (1992) Similarities among test sites based on cultivar performance in the hard red winter wheat region. *Crop Science* 32, 907–912.

Peterson, C.J. and Pfeiffer, W.H. (1989) International winter wheat evaluation: Relationships among test sites based on cultivar performance. *Crop Science* 29, 276–282.

Rajaram, S., van Ginkel, M. and Fischer, R.A. (1995) CIMMYT's wheat breeding mega-environments (ME). In: Li, Z.S. and Xin, Z.Y. (eds) *Proceedings of the Eighth International Wheat Genetics Symposium*. Vol. 2. China Agriculture Scientech Press, Beijing, China, pp. 1101–1106.

14 Three-mode Analytical Methods for Crop Improvement Programs

K.E. Basford[1], P.M. Kroonenberg[2] and M. Cooper[1]
[1]Department of Agriculture, The University of Queensland, Brisbane, Qld 4072, Australia; [2]Department of Education, Leiden University, The Netherlands

Abstract

Data collected from multi-environment trials conducted for the purpose of comparisons among genotypes are often in the form of a large three-mode array; designated as genotypes by environments by attributes. We consider two complementary ordination and clustering procedures, three-way principal component analysis and three-way mixture approach to clustering, to analyse such data. The application of these techniques enhance the researcher's ability to make decisions in crop improvement programs where several attributes are important and must be considered simultaneously when evaluating the impact of selection strategies. They are illustrated using data from an experiment which examined the grain yield adaptation of a sample of advanced wheat lines from the International Maize and Wheat Improvement Center (CIMMYT) and three Queensland cultivars in a series of water stress environments in Queensland. Although grain yield adaptation was of major concern, examination of other attributes which may influence the adaptation is important and maturity (days to anthesis) is included here. The interpretation of such analysis of multi-environment data to make both general and detailed statements about the relative performance of the lines and differences among the environments is illustrated.

Introduction

The existence of significant genotype by environment (G×E) interactions has been recognized by plant breeders as a complicating factor in selection and testing strategies for many years. The interactions reflect differences in adaptation which may be exploited by breeding for specific adaptation (emphasizing favourable interactions)

 Plant Adaptation and Crop Improvement (eds M. Cooper and G.L. Hammer)

or broad adaptation (minimizing interactions) by selection, and by adjustments to the test strategy. In order to make objective decisions, a full understanding of the nature of such interactions is needed. Various methodologies have been proposed for the analysis of univariate GxE data and they have each proved successful in certain situations.

Our concern is with multivariate or multiattribute GxE interactions where plant breeders measure more than one attribute on genotypes in multi-environment trials (METs). Then the collected data can be summarized in the form of a genotype by environment by attribute (GxExA) array of means which is formally defined as a three-mode three-way data set (Carroll and Arabie, 1983). We shall only discuss techniques which act directly on three-mode data, rather than those that act on a converted two-mode three-way data array, e.g. by computing a difference measure between each pair of genotypes within an environment to form a GxGxE matrix. We want a simultaneous analysis of all three modes in that data set, rather than separate univariate analyses, the results of which would then have to be combined.

Methods of Analysis

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 and Carroll (1980), the two types are largely complementary, and make use of the same information in different ways. Multivariate analysis of variance 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 GxE data on a single attribute, the standard multivariate analysis of variance was largely uninformative. Basford *et al.* (1991) believe that 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.

Hence, we shall discuss a clustering technique and an ordination technique suitable for analysing three-mode three-way data. As well as presenting the individual analyses, the results of the cluster analysis will be displayed superimposed on the results from the ordination to show how the two techniques are complementary and can be used to enhance the understanding of the interactions.

Clustering

If the genotypes can be clustered or grouped such that the genotypes within a group have similar response patterns for each of the attributes across environments, then the plant breeder can examine a much smaller data set and hence more easily integrate the information inherent in the trials. The mixture maximum likelihood method of clustering (Basford and McLachlan, 1985) is a model-based technique which can be applied in such cases to produce a grouping of genotypes (one of the modes) based on the simultaneous use of attributes and environments (the other two modes).

This clustering method 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. It is a non-hierarchical procedure which requires the number of clusters, c, to be specified. Although each cluster is allowed to have a different mean attribute vector in each environment, the covariance matrix (which specifies the correlation structure among the attributes) for each cluster is the same across environments, although it can differ from cluster to cluster. By allowing the mean attribute vector for a cluster to differ across environments, the significant genotype by environment interaction (which is almost always present) can be considered in the identification of groups of genotypes for which a general behavioural description is required. Thus a group could perform well in one environment and poorly in another environment. A covariance matrix particular to each cluster is beneficial as it might be expected that in the underlying group structure, the correlations between attributes might differ across groups of genotypes. For example, there could be a reasonable correlation between two attributes in one group, but virtually no correlation between these attributes in another group. In the current model, the correlation structure for an underlying group does not depend on environment. However, it is possible that significant G×E interactions could result in changes in correlations across environments.

Formally, if there are c groups (clusters) from which the genotypes have been sampled in unknown proportions π_m (m=1,...,c), then the distribution of the vector of attribute values for genotype i (i=1,...,g) in environment j (j=1,...,e) is given by:

$$f(x_{ij}) = \sum_{m=1}^{c} \pi_m f_m(x_{ij}) \tag{14.1}$$

where

$$f_m(x_{ij}) = N(\mu_{mj}, \Sigma_m) \tag{14.2}$$

is the usual assumption of the underlying distribution of the attribute vector in each group being multivariate normal with mean vector μ_{mj} (depending on the group and the environment) and covariance matrix Σ_m (depending on the group). The unknown parameters, i.e. mean vectors, covariance matrices and mixing proportions, are estimated using maximum-likelihood methods. In this process, the genotypes do not have to belong outright to only one of the groups as each genotype has a probability of belonging to each group, i.e. the posterior probability that genotype i belongs to group m, given the parameter estimates, is:

$$\hat{\tau}_{im} = \frac{\hat{\pi}_m \hat{f}_m(x_i)}{\sum_{m=1}^{c} \hat{\pi}_m \hat{f}_m(x_i)} \tag{14.3}$$

where

$$x_i = (x_{i1}, \ldots, x_{ie})' \tag{14.4}$$

is the vector containing the attribute vectors for all e environments. This non-allocation of the individuals to a group during the iterative process is particularly advantageous. Hierarchical procedures have been criticized because some of the initial fusions of individuals into groups (the process starts with n groups of one individual and finishes with one group of n individuals) may prove to be unfortunate at the later stages (when there are few groups). Non-overlapping groups

(clusters) are obtained by allocating each genotype to the group to which it has the highest estimated probability of belonging. The resulting clustering enables an overview of the information inherent in the data.

This mixture method of clustering requires the number of underlying groups or clusters to be specified. From a given starting allocation of the genotypes into groups, the EM algorithm (Dempster *et al.*, 1977) converges to a local maximum of the log likelihood. However, there is no guarantee that a global maximum will be reached. An approximate test on the log likelihood can be used to give an indication of the appropriate number of groups (McLachlan and Basford, 1988), but this is not exact and more research is being undertaken. A subjective assessment of the estimated probabilities of group membership and the rate of increase in the log likelihood values can also be used to determine an appropriate group number to adequately summarize the data. For the purpose of evaluating differences in adaptation among genotypes in METs, it is not necessary that the allocation of the genotypes into groups represents the 'true' grouping of the data, but rather that a satisfactory summarization is obtained. A decision on whether a satisfactory summary is obtained must be judged by the plant breeder in context with the objectives for conducting the METs.

The mixture method of clustering was applied using the program MIXCLUS3, an updated version of that appearing in the Appendix of McLachlan and Basford (1988). A copy of the program can be obtained from the first author of the chapter.

Ordination

If we want to know more detail about the relative performance of the genotypes, we need to consider an ordination procedure in which scores on a small number of components or factors are used to summarize the data. Two available techniques are three-mode principal component analysis (Kroonenberg, 1983) and parallel factor analysis (Harshman and Lundy, 1984). We shall only discuss the former, principally because we have more experience with it. In three-mode principal component analysis (which has some of the interpretational flavour of factor analysis), components (or factors) are derived for each of the modes. Each mode has its own number of components, and these components can be interpreted separately. Moreover, a set of parameters is derived which describe the relationships between the components. Generally, the emphasis is not so much on the interpretation of the components themselves, but on the interpretation of the structures of the genotypes, environments and attributes, as well as their interrelationships. The technique is used to reduce the data to such an extent that the main patterns can be inspected.

In order to apply three-mode principal component analysis (or a parallel factor analysis), the mean response of genotype i (i=1,...,g) in environment j (j=1,...,e) for attribute k (k=1,...,a), x_{ijk}, must be centred and scaled (Basford *et al.*, 1991). The chosen form is that recommended by Fox and Rosielle (1982) and Cooper and DeLacy (1994), i.e.:

$$\tilde{x}_{ijk} = (x_{ijk} - \bar{x}_{jk})/s_{jk}. \qquad (14.5)$$

Thus the data are centred by subtracting the environment mean for that attribute, $\bar{x}_{jk}$,

and scaled by dividing by the environment standard deviation for that attribute, s_{jk}.

Formally, given P, Q and R components for genotypes, environments and attributes, respectively, the model becomes:

$$\tilde{x}_{ijk} = \sum_{p=1}^{P} \sum_{q=1}^{Q} \sum_{r=1}^{R} a_{ip} b_{jq} c_{kr} g_{pqr} \qquad (14.6)$$

where a_{ip}, b_{jq} and c_{kr} are the component coefficients for genotypes, environments and attributes, respectively, and the g_{pqr} parameters weight combinations of components of the three modes. When a g_{pqr} value is large compared with other weights, that combination of the pth, qth and rth component is 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 (Kroonenberg, 1983, Section 6.9).

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 a joint plot, a variant of Gabriel's (1971) biplot. The term, joint plot (Kroonenberg, 1983), is used rather than the term biplot, because information from all three modes is used jointly to construct the 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. Thus, a very detailed statement about the relative performance of all the genotypes can be made from this analysis.

Just as the number of underlying groups must be specified for the mixture method of clustering, the three-mode principal component analysis requires the number of components for each mode to be determined. As explained in Basford *et al.* (1991), the number of components 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 components for each mode. The analogy is 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.

The ordination was applied using the program TUCKALS3 (Kroonenberg, 1994). A copy of this program can be obtained from the second author of this chapter.

Application

Experimental details

The data used to illustrate these techniques come from an experiment on 49 advanced wheat lines subjected to a range of water stress environments in a MET conducted in Queensland. The details of the experimental material, test environments, experimental design and measurements were given by Cooper *et al.* (1994a) and are not repeated here in depth. In summary, the 49 wheat lines were 40 advanced lines from the International Maize and Wheat Improvement Center (CIMMYT) in Mexico used in the selection study of Cooper *et al.* (1993), six other CIMMYT lines and three Queensland cultivars (Hartog, Banks and Kite). They were tested in six environments generated by imposing an irrigated and dryland

treatment at three locations. The environments are referred to as Brookstead dryland (BD) and irrigated (BI), Cecil Plains dryland (CPD) and irrigated (CPI), and Gatton dryland (GD) and irrigated (GI). All trials were managed to prevent disease and weeds influencing the relative performance of the lines. The Gatton irrigated environment was considered to provide the yield potential condition for comparison with the other environments.

Although grain yield adaptation was of major concern, grain yield, yield components, phenology and dry matter production and partitioning attributes were measured on all lines in each environment. In the current study, two attributes, grain yield (g m^{-2}) and maturity (days to anthesis), were analysed simultaneously. Significant ($P<0.05$) line variation was reported for both attributes in each environment when the lattice analysis of variance was used (Cooper *et al.*, 1994a). The lattice adjusted data were used in subsequent analyses. From the combined analysis of variance, significant ($P<0.05$) genotype and G×E interaction was identified for both attributes. The relative size of the genotypic (σ_g^2) and G×E interaction (σ_{ge}^2) components of variance estimated using a REML (residual maximum likelihood) procedure were; yield ($\sigma_g^2=287\pm120$; $\sigma_{ge}^2=1082\pm161$) and maturity ($\sigma_g^2=5.58\pm1.33$; $\sigma_{ge}^2=4.68\pm0.52$). Previous analysis of these data by Cooper *et al.* (1994a) was based on correlations between the attributes across environments.

Clustering

Using both the approximate test on the log likelihoods and subjective assessment of the estimated probabilities of group membership for determining underlying group number, the seven-group solution (Table 14.1) was found to be most appropriate for summarizing the variation in the data. Although line 38 was the only one allocated to Group G, other lines had some (small) probability of belonging to this group; otherwise the EM algorithm could not have converged to this solution. (A variance cannot be estimated from a sample of size 1.) Study of the log likelihoods for each starting solution indicated that even though there were many local maxima (depending on starting allocation), that reported in Table 14.1 was by far the best solution. The naming of the groups from A to G is in order of increasing mean yield over all environments.

Table 14.1. Membership of the seven group summary of the 49 wheat lines from the mixture method of clustering.

Group	Membership
A	48, 49
B	10, 24, 25
C	1, 2, 7, 12, 13, 17, 21, 33, 41, 42, 43, 44, 47
D	18, 19, 20, 22, 23, 26, 27, 28, 29, 30, 31, 34, 36, 37, 39, 45, 46
E	6, 8, 9, 14, 15, 16, 32, 35, 40
F	3, 4, 5, 11
G	38

The absolute values of the estimated correlation coefficient between yield and maturity for each cluster were generally less than 0.02, although it was 0.33 for Group E and –0.76 for Group A. The latter value should not be interpreted with much confidence as it was effectively calculated from only two lines (48 and 49). This conditional independence of the attributes, i.e. zero correlation among them, is often found in the underlying groups and is sometimes specified in the analysis (Aitkin *et al.*, 1981), although that was not the case here.

For comparison, the composition of the current seven groups is tabulated against that of the six groups obtained from Cooper *et al.* (1994b) who analysed yield alone (Table 14.2) with an hierarchical agglomerative technique (with squared Euclidean distance as the proximity measure and incremental sum of squares as the criterion). Using their composition as an initial allocation for the simultaneous analysis of yield and maturity, a better solution (in terms of log likelihood) at the six group level was obtained using the mixture method of clustering. However, the seven-group solution presented here was chosen as more appropriate. As expected, there were both similarities and differences in the two groupings (Table 14.2) with those of Cooper *et al.* (1994b) being allocated across a number of the groups obtained here.

The response pattern of these seven groups across environments for yield and maturity is shown in Fig. 14.1. The ordering of the environments on the horizontal axis is that of increasing mean attribute value over all lines. Basford *et al.* (1994) investigated the standard errors of the estimated means from the mixture method of clustering. They stated that if the underlying groups are widely spaced and the fitted posterior probabilities of group membership are either close to zero or one, an approximate minimum value could be determined by taking the square root of the estimated variance (of the attribute in question) divided by the sum of the posterior probabilities of belonging to the group. Basford and Tukey (1996) suggest underlap–overlap bars which are ±1.5 times the standard error of plotted means. If

Table 14.2. Comparison of groupings from mixture method of clustering (in the rows) with that obtained from Cooper *et al.* (1994b) (in the columns) using yield alone.

	Grouping from Cooper *et al.* (1994b)					
Table 1 Grouping	91 (7)*	92 (17)	90 (8)	89 (6)	87 (9)	77 (2)
A (2)	48, 49					
B (3)	10, 24, 25					
C (13)		12, 13, 17, 21, 41, 42, 43, 44, 47	2	33	1, 7	
D (17)	19, 37	18, 20, 23, 30, 34	26, 27, 28, 39, 45, 46	29, 36	22, 31	
E (9)		8, 16	32	6, 35, 40	9, 14, 15	
F (4)		11			3, 4	5
G (1)						38

* Number in each group given after group name.

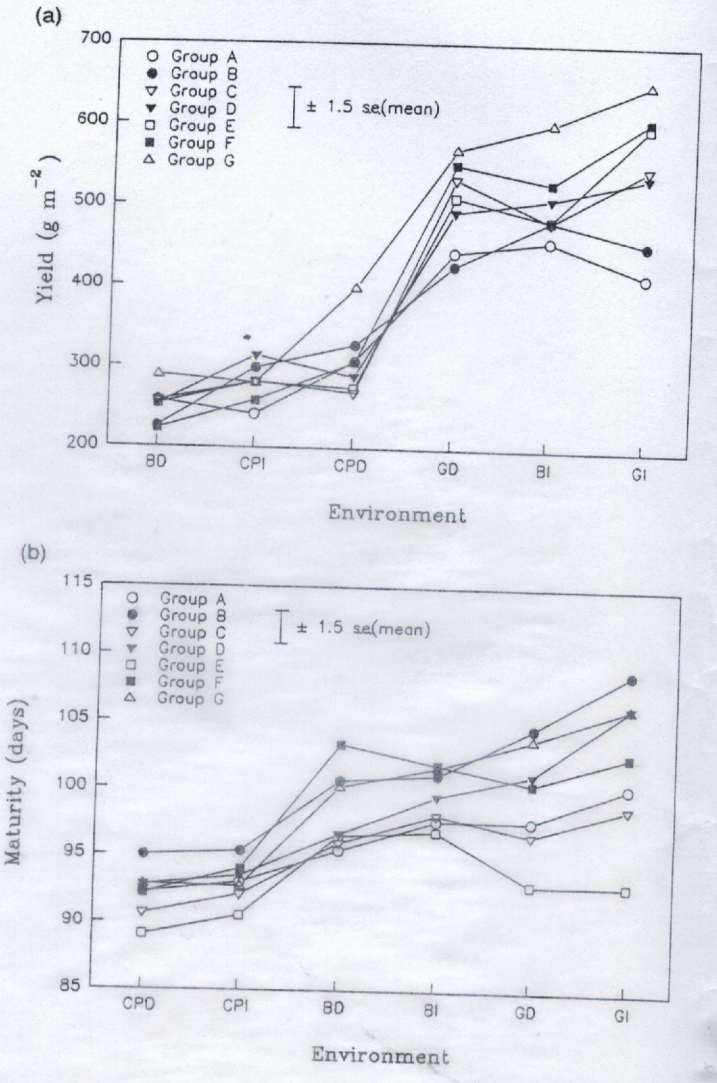

Fig. 14.1. (a) Group mean yields across environments (environment code given in Table 14.3). (b) Group mean maturity across environments (environment code given in Table 14.3).

the bars overlap in comparing any two means, then we are confident that they are not significantly different. The group with the largest estimated standard error of the mean is Group G for both yield (21.0) and maturity (0.79); not surprisingly as this is effectively a single member group. Instead of individual group bars, ±1.5 times this maximum has been used on each plot in Fig. 14.1. This is a very conservative estimate and will stop us interpreting too many differences on these displays. Nevertheless, there is still considerable line group by environment interaction for both attributes. For the two environments BD and CPI, where line mean yield was low (Fig. 14.1a), the conservative standard error suggests that line groups do not differ for yield but do differ for maturity (Fig. 14.1b). Since significant ($P<0.05$)

line variation was identified for yield in each environment, this suggests that the grouping has not adequately described the yield variation among the lines in these two environments.

Ordination

After examining several solutions, it was decided that the 3×3×2 solution (three components for lines, three components for environments and two components for attributes), which accounted for 65% of the variation, was an appropriate summary of the data on the 49 wheat lines.

The two components for attributes were almost equivalent to the original two attributes, and it was decided to consider a varimax rotation for both the environment and attribute components, while leaving the line components unchanged. The transformed (rotated) components for the environments and the attributes are shown in Tables 14.3 and 14.4, respectively. They account for 25%, 25% and 14% of the variation for environments and 28% and 37% of the variation for attributes. The variation accounted for by the various components from the original analysis was in decreasing order, but the rotation can change this (as was the case here). The two (rotated) attribute components are directly representative of the original attributes, yield and maturity, respectively (Table 14.4). By ignoring the small component values in Table 14.3, it can be seen that the first environment component primarily represents BI, CPD and CPI, the second primarily represents GD and GI, while the third primarily represents BD.

When the components are rotated, the core matrix must be counter-rotated in order to see which combinations of (rotated) components account for most of the variability. These are displayed in Table 14.5 where the explained variability is now distributed over a larger number of elements than in the original core matrix, which is not shown. For grain yield (really the yield slice), most weight is on the combination of first line component with the first environment component (0.051) and the second line component with the second environment component (0.088). For maturity (really the maturity slice), most weight is on the combination of first line

Table 14.3. Rotated environment components from the three-mode principal component analysis of the 49 wheat lines.

	Component		
Environment	1	2	3
Brookstead dryland (BD)	–0.02	–0.03	0.91
Brookstead irrigated (BI)	0.42	0.13	0.32
Cecil Plains dryland (CPD)	0.71	0.05	–0.24
Cecil Plains irrigated (CPI)	0.56	–0.13	0.10
Gatton dryland (GD)	–0.04	0.68	0.05
Gatton irrigated (GI)	0.01	0.71	–0.04
R^2	0.25	0.25	0.14

Table 14.4. Rotated attribute components from the three-mode principal component analysis of the 49 wheat lines.

	Component	
Attribute	1	2
Yield	1.00	0.00
Maturity	0.00	1.00
R^2	0.28	0.37

Table 14.5. Counter-rotated core matrix giving the proportion of variation accounted for by the combinations of components.

	Environment components		
	1	2	3
Yield slice			
Line components			
1	0.051	0.020	0.001
2	0.033	0.088	0.028
3	0.010	0.019	0.030
Maturity slice			
Line components			
1	0.156	0.125	0.028
2	0.000	0.000	0.001
3	0.001	0.001	0.055

component with the first and second environment components (0.156 and 0.125, respectively).

When looking at the joint plots with attribute as the reference mode, we interpret the rotated attribute components, and when looking at the joint plots with environments as the reference mode, we interpret the rotated environment components. The joint plots of lines and environments are displayed in Fig. 14.2 for the attribute components, yield and maturity, while the joint plots of lines and attributes are displayed in Fig. 14.3 for the three environment components, (a) mainly BI, CPD and CPI, (b) mainly GD and GI, and (c) mainly BD, respectively. In these joint plots, the wheat lines have been labelled according to the membership of the seven-group solution from the mixture method of clustering.

Consider the joint plots where attributes are the reference mode (Fig. 14.2). We shall initially discuss these, but the same can be said for the joint plots where environments are the reference mode (Fig. 14.3). Arrows (vectors from the origin) are drawn for the environment vectors (Fig. 14.2), while the wheat lines are shown as points in these displays. The length of a vector for a particular environment indicates the importance of that environment to the differences in the component in the reference mode. To describe the performance of any particular wheat line with

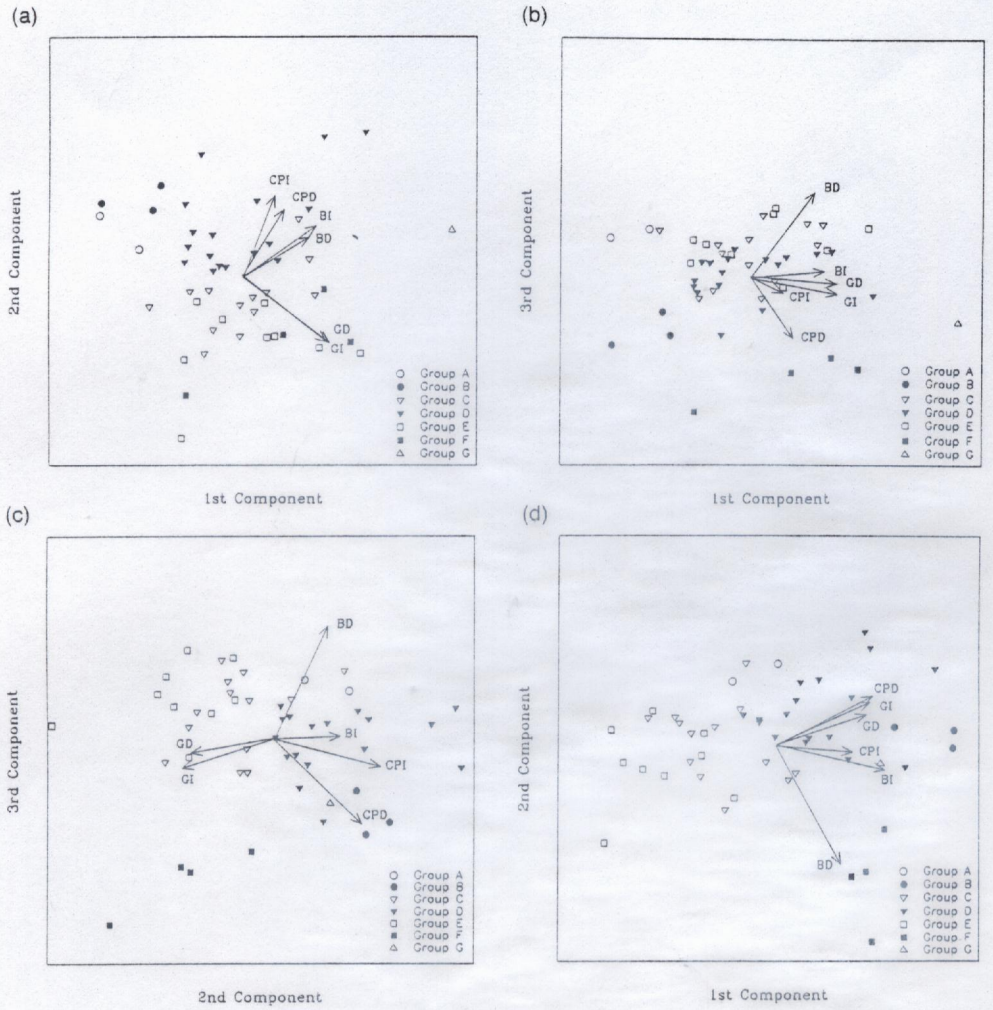

Fig. 14.2. (a) Joint plot of Component 1 vs Component 2 for yield. (b) Joint plot of Component 1 vs Component 3 for yield. (c) Joint plot of Component 2 vs Component 3 for yield. (d) Joint plot of Component 1 vs Component 2 for maturity.

respect to a particular environment, drop a perpendicular to the corresponding vector. This is equivalent to finding the inner product between the vector to a particular wheat line and the environment vector. A positive value indicates a larger than average performance in that environment while a negative value indicates a smaller than average performance in that environment. Parallel environment vectors indicate that the environments influence the performance of lines in a similar way, while vectors at 180° indicate dissimilar performance. Environment vectors at 90° indicate independent performance. For example, in the joint plot of the first versus second components for yield performance (Fig. 14.2a), the dryland and irrigated treatments at each site are quite similar, whereas the performace at Gatton appears to be

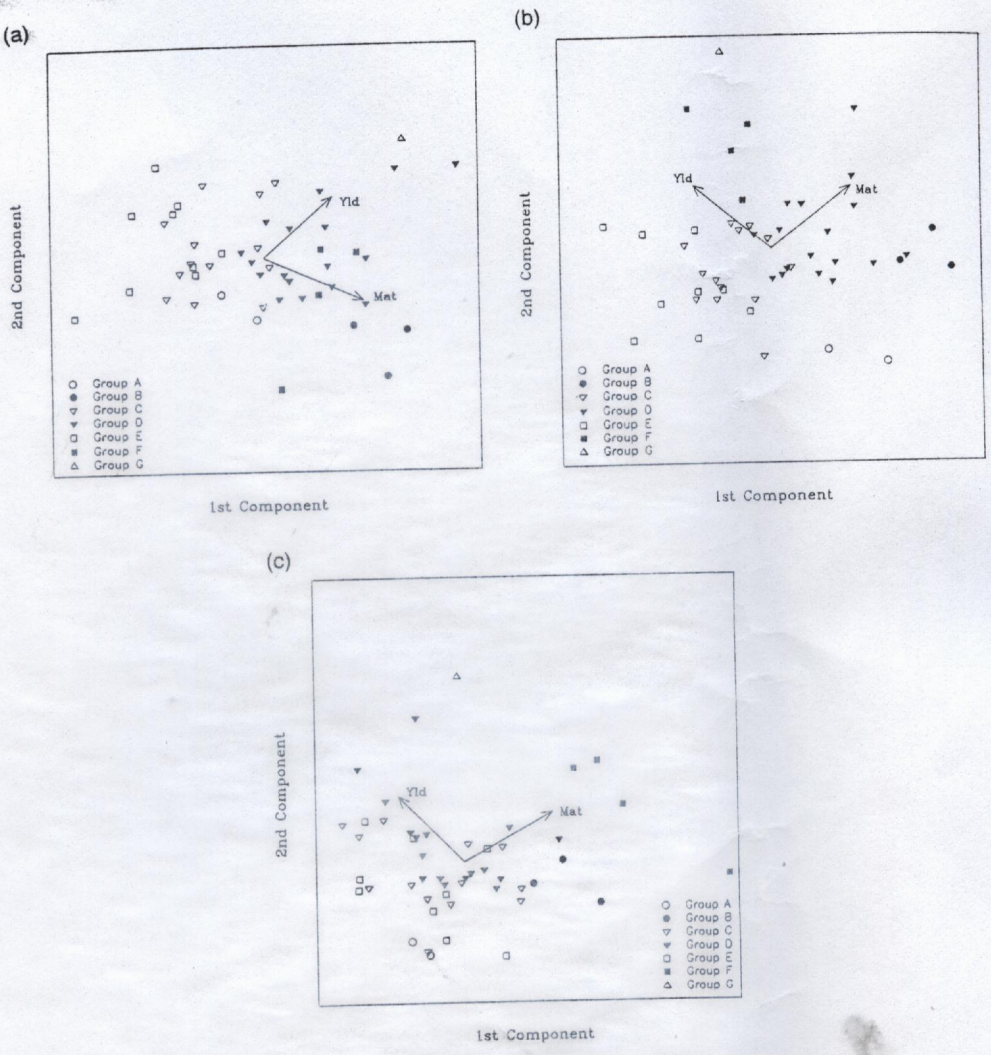

Fig. 14.3. (a) Joint plot of Component 1 vs Component 2 for Environment Component 1 (BI, CPD, CPI). (b) Joint plot of Component 1 vs Component 2 for Environment Component 2 (GD, GI). (c) Joint plot of Component 1 vs Component 2 for Environment Component 3 (BD).

independent of the performance at the other sites. As Brookstead and Cecil Plains are in close proximity on the Darling Downs whereas Gatton is in the Lockyer Valley, this is reflecting genotype by location interactions.

From Fig. 14.2a, the unique response of Group G (containing line 38) is apparent, as is the average to high yield shown by members of Group D in Brookstead and Cecil Plains, but not in Gatton. Similarly, Groups F and E showed average to high yield at Gatton, but low yield at the other locations. This reflects the independence of the response at Gatton compared with that at Brookstead and Cecil Plains. Although yield under the water stress environments (Brookstead dryland and Cecil

Plains, in particular) generally differed from that under yield potential conditions, Group G did well everywhere. In Fig. 14.2b, Group F showed specific adaptation to Cecil Plains dryland, while in Fig. 14.2c, this group does well at Gatton and Cecil Plains dryland, but poorly at Brookstead dryland. Fig. 14.2d emphasizes early versus late flowering groups on the first component, while the second component suggests Group F was particularly later flowering at Brookstead dryland.

It is clear from the joint plots for the three environment components (Fig. 14.3), that yield and maturity are independent of one another as they are at right angles. This is consistent with results from the cluster analysis where these attributes were basically independent for each group, except possibly Group E. This result was somewhat surprising given the importance of phenology for yield in Queensland (Woodruff and Tonks, 1983), but it indicates that variation for grain yield exists which is largely independent of the effects of phenology.

There is less localization of the groups for the joint plots for the environment components (Fig. 14.3) than for the attribute components (Fig. 14.2). For Brookstead irrigated and Cecil Plains (Fig. 14.3a), the higher yielding lines tended to be later flowering ones in Groups F and G and some individual lines from Group D. They could be taking advantage of irrigation at these locations and the rainfall which occurred at flowering at Cecil Plains. For Gatton (Fig. 14.3b), the low pre-anthesis stress could have ensured that both early and later flowering lines had high yield. For Brookstead dryland (Fig. 14.3c), the severe water stress could have resulted in the high yield being generally associated with quicker flowering lines. The possible exception to this would be Group G.

These results were consistent with the analyses of Cooper *et al.* (1994a) in that there was general independence of yield and days to anthesis with only weak relationships. Looking at the distribution of the wheat lines on the joint plots provides a much clearer interpretation than that obtained by examining correlations.

Discussion

Both the clustering and ordination procedures gave a sensible and useful summarization of the data from the trial on the 49 wheat lines subjected to water stress environments. Considerably more detail and interpretation were available through the complementary use of these techniques, especially in examining the relationships and variation among and 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. When selection has to be made for multiple traits, tandem selection, independent culling levels or selection indices are often used. Where independent culling levels are attempted, it is extremely difficult to assess jointly information on multiple attributes integrated across environments. Similarly, it is hard to visualize what is happening with selection indices. Joint plots provide a powerful graphic to assist in this process. Alternatively, they could be used to study the patterns once selections have been made.

As argued by Basford *et al.* (1991), the major advantage of these methods is that they allow the data set to be treated in the form of a three-mode three-way array. An overall picture of response is obtained by studying the groups from a

clustering procedure in which the important G×E interactions present in such trials have been incorporated directly into the underlying models. Similarly, the representation of the wheat lines in a reduced space allows a quicker appreciation of the major differences inherent in the data. In addition, the ordination technique does allow more detailed information and possible structure in the environments and attributes to be extracted. For the example considered here, an enhanced interpretation of the influence of flowering time on yield was obtained over that obtained by Cooper *et al.* (1994b).

These techniques provide complementary information which can be readily displayed in common figures. They can be interpreted with relatively limited training and effectively improve and refine the information obtained by plant breeders from their trials. Hence they are very useful techniques which could be frequently employed in the statistical analysis of such three-mode three-way data.

Acknowledgements

The experimental work was supported by the Australian Wheat Research Council and the farmers of Queensland. The work of P.M. Kroonenberg was partially supported by the Netherlands Organization of Scientific Research (NWO).

References

Aitkin, M., Anderson, D. and Hinde, J. (1981) Statistical modelling of data on teaching styles (with discussion). *Journal of the Royal Statistical Society, Series A* 144, 419–461.

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

Basford, K.E. and McLachlan, G.J. (1985) The mixture method of clustering applied to three-way data. *Journal of Classification* 2, 109–125.

Basford, K.E. and Tukey, J.W. (1996) Graphical profiles as an aid to understanding plant breeding experiments. *Journal of Statistical Planning and Inference*, Special Issue on Robust Statistics and Data Analysis (in press).

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

Basford, K.E., Greenway, D.R. and McLachlan, G.J. (1994) *Standard Errors of Fitted Component Means of Normal Mixture Models*. Centre for Statistics Research Report No 27, The University of Queensland, Brisbane.

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

Cooper, M. and DeLacy, I.H. (1994) Relationships among analytical methods used to study genotypic variation and genotype-by-environment interaction in plant breeding multi-environment experiments. *Theoretical and Applied Genetics* 88, 561–572.

Cooper, M., Byth, D.E., DeLacy, I.H. and Woodruff, D.R. (1993) Predicting grain yield in Australian environments using data from CIMMYT international wheat performance trials. 2. The application of classification to identify environmental relationships which exploit correlated response to selection. *Field Crop Research* 32, 323–342.

Cooper, M., Byth, D.E. and Woodruff, D.R. (1994a) An investigation of the grain yield adap-

tation of advanced CIMMYT wheat lines to water stress environments in Queensland. I. Crop physiological analysis. *Australian Journal of Agricultural Research* 45, 965–984.

Cooper, M., Byth, D.E. and Woodruff, D.R. (1994b) An investigation of the grain yield adaptation of advanced CIMMYT wheat lines to water stress environments in Queensland. II. Classification analysis. *Australian Journal of Agricultural Research* 45, 985–1002.

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

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977) Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B* 39, 1–38.

Fox, P.N. and Roseille, A.A. (1982) Reducing the influence of environmental main-effects on pattern analysis of plant breeding environments. *Euphytica* 31, 645–656.

Gabriel, K.R. (1971) The biplot graphic display of matrices with application to principal com-ponent analysis. *Biometrika* 58, 453–467.

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. *Theoretical and Applied Genetics* 76, 1–10.

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

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

Kroonenberg, P.M. (1994) The TUCKALS line: A suite of programs for three-way data analysis. *Computational Statistics and Data Analysis* 18, 73–96.

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

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

Woodruff, D.R. and Tonks, J. (1983) Relationship between time of anthesis and grain yield of wheat genotypes with differing developmental patterns. *Australian Journal of Agricultural Research* 34, 1–11.