CBCS
HTTP://AGROBIOL.SGGW.WAW.PL/CBCS

COMMUNICATIONS IN BIOMETRY AND CROP SCIENCE
Vol. 11, NO. 2, 2016, PP. 127–139

INTERNATIONAL JOURNAL OF THE FACULTY OF AGRICULTURE AND BIOLOGY,
Warsaw University of Life Sciences – SGGW, POLAND

REGULAR ARTICLE

Analysis of sunflower data from a multi-attribute genotype ×environment trial in Brazil

Marisol García-Peña[1]*, Sergio Arciniegas-Alarcón[2], Kaye Basford[3], Carlos Tadeu dos Santos Dias[1]

[1]Universidade de São Paulo, Departamento de Ciências Exatas, Brasil.
[2]Fundación Universitaria Konrad Lorenz, Escuela de Negocios, Bogotá, Colombia.
[3]The University of Queensland, School of Agriculture and Food Sciences, Australia.

*Corresponding author: Marisol García-Peña; E-mail: luzmara@gmail.com

CITATION: García-Peña M., Arciniegas-Alarcón S., Basford K., Dias C.T.S (2016). Analysis of sunflower data from a multi-attribute genotype × environment trial in Brazil *Communications in Biometry and Crop Science* 11, 127–139.

Received: 14 August 2015, Accepted: 29 June 2016, Published online: 19 July 2016

ABSTRACT

In multi-environment trials it is common to measure several response variables or attributes to determine the genotypes with the best characteristics. Thus it is important to have techniques to analyse multivariate multi-environment trial data. The main objective is to complement the literature on two multivariate techniques, the mixture maximum likelihood method of clustering and three-mode principal component analysis, used to analyse genotypes, environments and attributes simultaneously. In this way, both global and detailed statements about the performance of the genotypes can be made, highlighting the benefit of using three-way data in a direct way and providing an alternative analysis for researchers. We illustrate using sunflower data with twenty genotypes, eight environments and three attributes. The procedures provide an analytical procedure which is relatively easy to apply and interpret in order to describe the patterns of performance and associations in multivariate multi-environment trials.

Key Words: *three-way data; genotype-by-environment interaction; clustering via mixtures; principal components.*

INTRODUCTION

According to Basford et al. (1991), when the genotype-by-environment interaction (G×E) is significant, selection and testing strategies for plant breeders are complicated. A portion of the G×E represents differences in adaptation (broad or specific), and in order to make objective selection decisions it is necessary to understand the nature of these interactions.

Generally the breeders are interested in more than one attribute, so it is important to use multivariate analyses of such data. In G×E trials with only one attribute, various methodologies work well, such as the additive main effects with multiplicative interaction models – AMMI. For multivariate multi-environment trials, there are several alternatives. Basford (1982), Basford and McLachlan (1985), Kroonenberg and Basford (1989), Basford et al. (1990, 1991) and van Eeuwijk and Kroonenberg (1995, 1998) present some cluster and ordination techniques to analyse three-way data. Denis and Moro (1995) and Moro and Denis (1995) proposed three-way models for the multivariate analysis of genotype-by-environment interaction in an attempt to generalise factorial regression, biclustering (genotypes and environments simultaneously) and biadditive models.

Interpreting the underlying complex interactions in a three-way array is difficult. If the evaluation of the genotypes is made using only one attribute such as yield, then even though this may be considered to be the most important attribute, much of the available data is being ignored. On the other hand, if separate analyses for each attribute are made, it is difficult to successfully combine the results, even accepting that the correlations among the attributes are ignored (Basford et al. 1990).

In this paper, two multivariate techniques, the mixture maximum likelihood method of clustering - MIXCLUS (Basford and McLachlan 1985) and three mode principal component analysis - TMPCA, Tuckals3 model (Kroonenberg and De Leeuw 1980), are used to analyse genotypes, environments and attributes simultaneously. Three-way sunflower data with twenty genotypes, eight environments and three attributes are used for illustration.

The main objective is to complement previously published papers in this area by demonstrating that it is possible to deal with several attributes in a unique analysis, make both global and detail conclusions about genotype performance, highlight the benefit of using three-way data directly, and provide an alternative analysis for researchers. While these techniques are used to analyse multi-attribute G×E data, they can be used to analyse any similar three-way three-mode data set.

MATERIALS AND METHODS

EXPERIMENTAL DETAILS

We used the results of one year of experimentation from the Embrapa evaluation report on Sunflower 2012/2013. We assessed 20 genotypes, consisting of 18 lines obtained from crossbreeding (hybrids) and 2 cultivars, in eight environments. A randomized complete block design was used with four repetitions. The trials in each of the eight environments were laid out in four-row plots of 6m in length within each block. We considered three attributes, grain yield per hectare (t/ha), oil content (%) and plant height (cm).

MIXTURE MAXIMUM LIKELIHOOD METHOD OF CLUSTERING - MIXCLUS

If the genotypes can be clustered or grouped such that those within a group have similar response patterns for each of the attributes across environments, then the breeder can examine a smaller data set and hence more easily understand and integrate the information inherent in the trials. The mixture maximum likelihood method of clustering can be applied to produce a grouping of genotypes based on the simultaneous use of attributes and environments (Basford et al. 1990).

In using the mixture method of clustering, it is assumed in the first instance that there is a specific number, NG, of underlying groups. An initial grouping of genotypes can be obtained using the results of other clustering techniques, like k-means or hierarchical clustering applied to the genotype by environment data for a single attribute, a priori information about the data or simply initial random values. A likelihood is formed under the assumption that the elements are a sample from a mixture in various proportions of these groups (π_{ng}, $ng=1,...,NG$). The details of the mixture method of clustering three-way three-

mode data are given in Basford and McLachlan (1985). The most common assumption and the one used here is that the underlying distribution of the attributes in each group is multivariate normal, that is, the distribution of the vector of attributes ($\mathbf{x}_{ij}$) for genotype i (i=1,...,I) in environment j (j=1,...,J) is given by $f(\mathbf{x}_{ij}) = \sum_{ng=1}^{NG} \pi_{ng} f_{ng}(\underline{\mathbf{x}}_{ij})$, where $f_{ng}(\underline{\mathbf{x}}_{ij}) \sim N(\underline{\boldsymbol{\mu}}_{ngj}, \mathbf{V}_{ng})$. This model assumes that each population has its own mean vector and the response can be different from one environment to another; that is, a group can have good performance in one environment but poor performance in another. The correlation structure between attributes can be different for the different groups. This is to allow for the general situation in which interaction between genotypes and environments results in different correlations among the attributes within the groups.

Following Basford et al. (1990), one of the objectives of the analysis is to estimate these unknown parameters in the model (mean vectors, correlation matrices and mixture proportions). This is achieved by maximizing the log likelihood. The probability that each element belongs to each group is calculated by replacing the unknown parameters in the appropriate probability expression with their maximum likelihood estimates, and this procedure is called the mixture maximum likelihood method. Each element is allocated to the group to which it has the largest estimated probability of belonging.

The method handles data in its original form, i.e. without centring or scaling, as that is effectively taken care of by the multivariate normal distribution (Basford and McLachlan 1985).

THREE-MODE PRINCIPAL COMPONENT ANALYSIS - 3MPCA (TUCKALS3 MODEL)

A model to analyse three-way data was proposed by Tucker (1966). Then, for data classified in three modes (genotypes, attributes and environments)

$$Z_{ijk} = \sum_{p=1}^{P}\sum_{q=1}^{Q}\sum_{r=1}^{R} g_{pqr} a_{ip} b_{jq} c_{kr} + e_{ijk} \text{ with } i=1,2,...,I,\ j=1,2,...,J,\ k=1,2,...,K$$

where I, J and K represent the number of levels of the first, second and third modes, respectively. $\boldsymbol{A}_{I\times P}, \boldsymbol{B}_{J\times Q}$ and $\boldsymbol{C}_{K\times R}$ are marker matrices associated with each mode, with elements a_{ip}, b_{jq} and c_{kr}, and P, Q and R represent the number of components retained in each mode. $\boldsymbol{G}$ is a core array of the three-way data with elements indicating the relationship among the components of each mode (Varela et al. 2009). The element g_{pqr} shows the strength of the relationship among the p^{th} component of the first mode, the q^{th} component of the second mode and r^{th} component of the third mode and its squared value indicates the explained variation for that combination of components. The array $\boldsymbol{G}$ can be considered to be a generalisation of the diagonal matrix of the eigenvalues associated with the singular value decomposition of a two-way matrix (Varela et al. 2008, Araújo et al. 2011).

Tucker (1966) gave a solution for matrices $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ of the model, but the solutions found were non-least squares estimators. For full rank (taking P, Q and R components retained in the first, second and third mode, respectively) it is possible to reproduce the Z_{ijk} value, but with less than the full model the fit may be far away from the true value of Z_{ijk}, for example, when just the first component of each mode is retained (Varela and Torres, 2005).

In order to resolve this situation, Kroonenberg and De Leeuw (1980) proposed an alternating least squares algorithm (Tuckals3), taking as initial values the Tucker solution that is based on the estimators for $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ that minimises the sum of squares of the residual, as in Equation (1):

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(Z_{ijk}-\hat{Z}_{ijk}\right)^2=\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(Z_{ijk}-\sum_{p=1}^{P}\sum_{q=1}^{Q}\sum_{r=1}^{R}a_{ip}b_{jq}c_{kr}g_{pqr}\right)^2 \quad (1)$$

The basic objective of the model underlying the method is to represent each of the ways or modes as well as possible in a low-dimensional space by forming linear combinations (components) of the levels of the modes. Theoretical details and applications can be found in Kroonenberg (1983, 2008).

SELECTING THE NUMBER OF COMPONENTS TO BE RETAINED IN EACH MODE

In the Tuckals3 algorithm, it is necessary to specify a priori the number of components to be retained, so a selection criterion in needed. Timmerman and Kiers (2000) proposed a procedure that indicates the *P*, *Q* and *R* values for the Tuckals3 algorithm to be a global optimal solution, rather than a local one.

The method consists of calculating the adjusted values for all possible solutions obtained from Tuckals3 algorithm. In each solution a model is fitted with the objective of approximating the *Z* values. Then, for each solution the error and explained amount of variation are calculated. The fit coincides with the part of *Z* explained by each solution (Varela and Torres 2005).

All possible solutions must satisfy the following conditions, $P \le QR$, $Q \le PR$ and $R \le PQ$, because a model where, for example, $R > PQ$ gives the same fit as the model with $R = PQ$. In the same way, models with $P > \max(I, JK)$, $Q > \max(J, IK)$ or $R > \max(K, IJ)$ can be omitted because these models do not fit better than those with P, Q and R equal to $\max(I, JK)$, $\max(J, IK)$ and $\max(K, IJ)$, respectively. For instance, if the 3×1×2 solution coincides with the fit of the 2×1×2 solution, the former can be eliminated because it has more axes (Timmerman and Kiers 2000, Kiers and der Kinderen 2003).

After finding the fit values for the possible solutions and for each value of $s = P + Q + R$, the one with best fit is selected. After solutions are selected (one for each value of s), the differences in fit (DifFit) between one solution and the immediately previous one ($dif_s = SQ_s - SQ_{s-1}$) are calculated. This indicates how much is gained by fitting an increasing total number of components (s); dif_s is calculated for *s*=4,...,*S*, where S is the sum of the highest numbers of components used in our analyses, while for s=3, dif_s is equal to the model fit with *P*=1, *Q*=1, *R*=1, which implies that dif_3 is compared with the zero components model (Timmerman and Kiers 2000).

The next step is to eliminate all those solutions for which there exists a solution with more components and a greater associated difference in fit. In this step, the aim is to obtain similarity with the classic PCA in which increasing the number of components produces smaller gains in fit.

For the selected set of solutions, the quotient $b_s = \dfrac{dif_s}{dif_{s+1^*}}$ is calculated, where dif_{s+1^*} is the next largest value after dif_s. The optimal solution is the one with maximum quotient and associated DifFit greater than the critical value $\dfrac{\|Z\|^2}{(S_{\min}-3)}$, where $S_{\min} = \min(I, JK) + \min(J, IK) + \min(K, IJ)$. Models with values below this limit will not be taken into account.

This method permits the researcher to find an optimum balance between the number of components retained in each mode and the variability explained by the model. Timmerman and Kiers (2000) declare that if the number of axes or components has been adequately selected, the Tuckals3 algorithm will rarely lead to a local optimum.

BIPLOT REPRESENTATION

In order to represent the solution graphically, three matrices of principal components or markers are used in a generalisation of the classic biplot proposed by Gabriel (1971), see Gabriel (2002). A biplot is a simultaneous representation of the rows and columns of a matrix, in which the columns (attributes here) are represented by vectors and the rows (genotypes here) are represented by points. The value of an attribute for a particular genotype can be estimated from the projection of the point onto the vector that represents the attribute in question (Varela et al. 2006).

A biplot is designed to work with one matrix (2 modes), so when a third mode is introduced, it is necessary to project on the principal components of one of the modes and draw the same number of biplots as components of that mode. This is called a joint plot (Kroonenberg 1983). For instance, if projecting onto the components of the third mode, it is necessary to consider R biplots, i.e. for R=1 a biplot of the matrix $\boldsymbol{AG}_1\boldsymbol{B}$, for R=2 a biplot of the matrix $\boldsymbol{AG}_2\boldsymbol{B}$ and so on for all the components retained in the third mode. G_r is the part of the three-way array $\boldsymbol{G}$ associated with R=r, i.e. with the r^{th} component of the third mode (Varela and Torres 2005).

Because of the different units of measurement for the different attributes measured in this genotype × environment trial, it is necessary to adjust the scales so that they can be analysed simultaneously. Hence the data are centred by subtracting the attribute effect and the environmental effect, $\tilde{x}_{ijk} = x_{ijk} - \mu_k - \beta_j$. Then the data are scaled by dividing by the standard deviation for each attribute, calculated over all environments, $s_k = \sqrt{\sum_{i=1}^{I}\sum_{j=1}^{J}\tilde{x}_{ijk}^2}$. Least-squares estimators are used to estimate μ_k and β_j. The centred and scaled data become the Z_{ijk} for the application of three-mode principal component analysis.

COMPUTER PROGRAMS

The mixture method of clustering has been programmed by, and is available from, the third author, Professor K.E. Basford. The three-mode principal component analysis is in the 3WayPack suite which has been written by, and may be obtained from, Professor P.M. Kroonenberg (Leiden University, The Netherlands), see the website of the Three-Mode Company (http://three-mode.leidenuniv.nl).

RESULTS

MIXTURE MAXIMUM LIKELIHOOD METHOD OF CLUSTERING - MIXCLUS

In order to select the number of groups, different initial values for each *NG* value were used. The initial grouping of genotypes was obtained after cluster analysis for each attribute with the k-means method and model-based clustering. *NG*=2 and 3 were evaluated, but for *NG*=3 it was found that the eigenvalues of the individual group covariance matrices were very small or negative which causes problems with the inversion of these matrices. This indicates that it may be necessary to reduce the number of groups, so it was decided to work with *NG*=2.

Table 1 presents the estimated mean vector for each group with their corresponding correlation matrix and group composition. The groups contain 7 and 13 genotypes respectively. In general group 2 had higher means for two of the three attributes than group 1. The correlation between sunflower grain yield and oil content was positive in both groups (0.31 and 0.26). On the other hand, the correlation between sunflower grain yield and plant height was positive in the first group and negative in the second (0.27 and -0.10), the same occurred for the correlation between oil content and plant height (0.24 and -0.06).

Table 1. Estimated means with their corresponding correlation matrix and group composition from the mixture clustering method.

Attribute	Group 1	Group 2
Grain yield (t/ha)	1.746	1.987
Oil content (%)	42.824	39.971
Plant height (cm)	150.698	163.225
Correlation matrix	$\begin{pmatrix} 1 & & \\ 0.31 & 1 & \\ 0.27 & 0.24 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & & \\ 0.26 & 1 & \\ -0.10 & -0.06 & 1 \end{pmatrix}$
Composition	G3-BRSG35[#] G6-BRSG38 G9-BRSG42 G10-Embrapa122[#] G12-EXP26 G13-HELIO358 G15-MG341	G1-BRSG30 G2-BRSG34 G4-BRSG36 G5-BRSG37 G7-BRSG39 G8-BRSG40 G11-EXP24 G14-M734 G16-SRM767 G17-SRM779CL G18-SRMCiro G19-SYN3950HO G20-V100964

[#] Cultivars

Figure 1 displays the estimated environment means for each group for each attribute, using the notation of E1 to represent the first environment, etc. Group 2 had better performance for sunflower grain yield and plant height than group 1, while the opposite was true for oil content. For group 1, the environment with highest sunflower grain yield was environment 3 (E3) and that with the lowest was environment 2 (E2). For group 2, the highest was E3 but the lowest was E1, so the genotypes in group 2 could be a good option to increase sunflower grain yield in environments similar to environment 3. In contrast, E3 had moderate value for oil content for both groups. On the other hand, E1 determined the expression of high and moderate oil concentrations in hybrids of groups 1 and 2, respectively and low sunflower grain yield for both groups. In the case of plant height, again the two groups presented similar behaviour, with E3 having the highest yield and E1 the lowest. Not surprisingly, the pattern for plant height more closely followed that for sunflower grain yield, but with spikes and dips caused by oil content.

In E4 and E5, genotypes in both groups were characterised by moderate values for grain yield and plant height and the highest values for oil content. From Figure 1 it can be concluded that there is not much G×E interaction. Thus, except for grain yield and plant height in some environments, the response patterns across environments were largely similar.

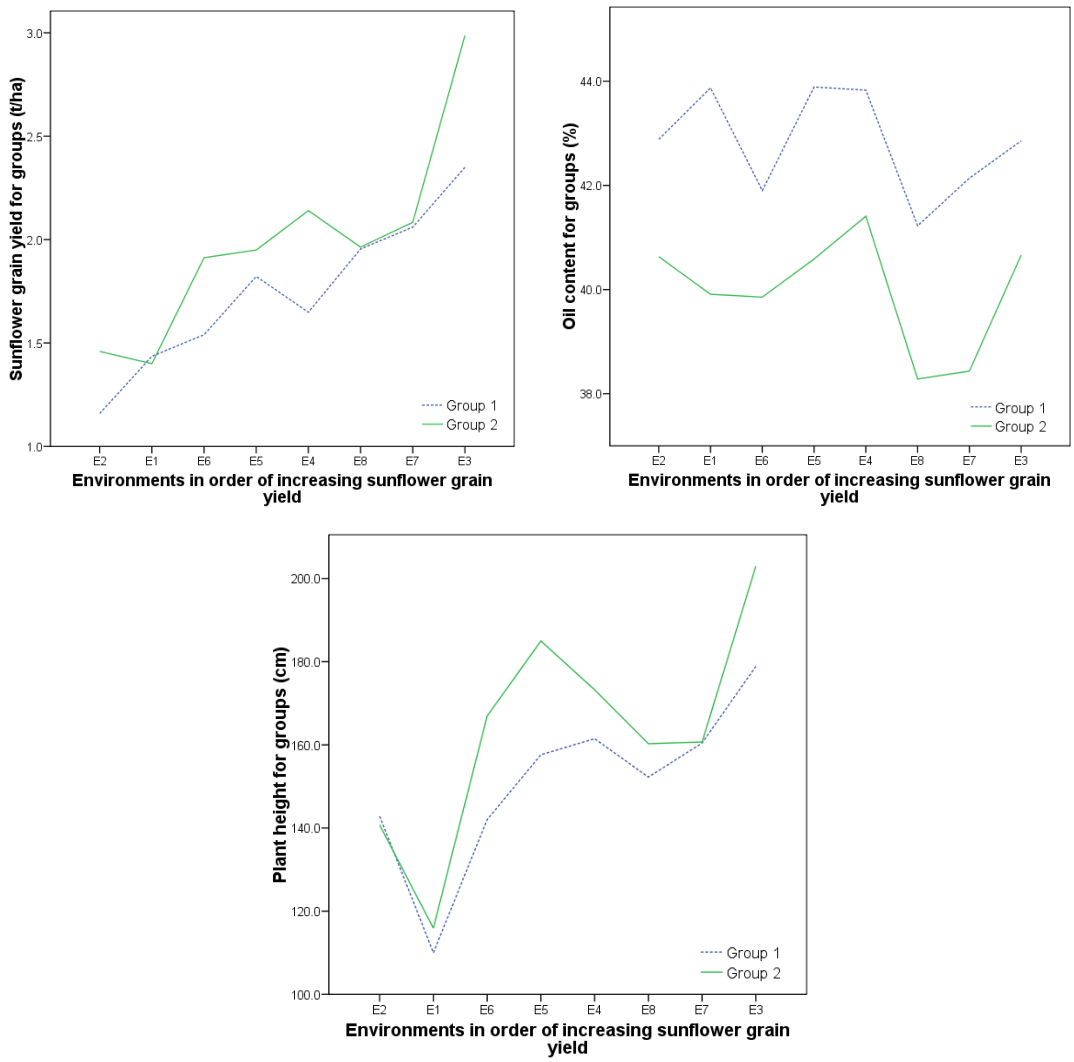


Figure 1. Expected environmental means for the two groups formed by mixture maximum likelihood method for sunflower grain yield, oil content and plant height.

THREE-MODE PRINCIPAL COMPONENT ANALYSIS - 3MPCA (TUCKALS3 MODEL)

Initially, the number of principal components for each mode should be determined by the fit of various Tuckals3 models. As I=20 (number of genotypes), J=8 (number of environments) and K=3 (number of attributes), there exist IJK=480 possible combinations of components. However, we only select those that satisfy the conditions $P \leq QR$, $Q \leq PR$ and $R \leq PQ$, because the others are redundant. For the selected combination of components, the percentage of variation explained by fitting the Tuckals3 model was calculated. Several solutions are obtained for each value of s=P+Q+R, but in Table 2 only the best solution (accounting for most variation) for each value of s is presented. Also, the difference in fit value associated with each solution, the quotient between the differences and the explained percentage of variance, is presented.

The optimal solution was 2×1×2, because it has the highest associated quotient among solutions with difference in fit higher than the critical value, $\left(1/\left[S_{\min} - 3\right] = 1/28 = 0.035\right)$. The retention of just one component for environments reflects the relatively small amount of genotype by environment interaction and this is consistent with the results found in the MIXCLUS analysis.

Tables 3, 4 and 5 show the components for genotypes, environments and attributes for the best fit. The two components retained for genotypes and attributes account for 24% and 20% of the variation, respectively. The first and second components for genotypes and the first component for attributes are contrasts, while the second component for attributes and the one for environments are rough averages.

Table 2. Best fit solutions for each value of $s=P+Q+R$.

$s=P+Q+R$	Model (Solution $PxQxR$)	Difference in fit	DifFit Quotient	% Fit
3	1×1×1	0.246	1.272	24.64
5	2×1×2	0.194	2.220	44.01
6	2×2×2	0.073		51.29
7	3×1×3	0.027		53.95
8	3×2×3	0.087	1.853	62.68
9	3×3×3	0.047	1.051	67.39
10	4×3×3	0.045	1.157	71.87
11	4×4×3	0.039	1.387	75.74
12	5×4×3	0.023		78.05
13	5×5×3	0.028		80.84

Table 3. Components for genotypes.

Genotype	1	2
G1	-0.352	0.241
G2	-0.226	-0.049
G3	0.230	-0.133
G4	-0.310	-0.010
G5	0.025	0.259
G6	0.159	-0.093
G7	-0.156	-0.232
G8	0.001	0.040
G9	0.112	-0.350
G10	0.215	-0.574
G11	-0.171	-0.139
G12	0.133	-0.041
G13	0.494	0.219
G14	-0.203	-0.175
G15	0.253	0.188
G16	0.129	0.161
G17	-0.348	-0.032
G18	-0.141	0.279
G19	0.080	0.201
G20	0.077	0.241
Variation %	24.0	20.0

Table 4. Components for environments.

Environment	1
E1	0.238
E2	0.229
E3	0.535
E4	0.377
E5	0.409
E6	0.456
E7	0.241
E8	0.169
Variation %	44.0

Table 5. Components for attributes.

Attribute	1	2
Sunflower grain yield	-0.178	0.714
Oil content	0.806	0.500
Plant height	-0.564	0.490
Variation %	24.0	20.0

The relationships among the different modes of data provided by the three-mode principal component analysis are presented in Figure 2, and the strength of these relationships can be measured by the inner products between genotypes and attributes (Table 6). High inner product indicates good performance with the attribute.

The groups found with the mixture maximum likelihood method of clustering are clearly separated in the joint plot in Figure 2. The separation is based mainly on plant height with the genotypes in group 2 having higher values than those in group 1. On the other hand, genotypes G1 and G18 had the highest grain yield and plant height (inner products 0.186, 0.175 and 0.260, 0.174), while G10 and G9 had the lowest (inner products -0.350, -0.210 and -0.321, -0.187). For oil content, G13 and G15 from group 1 had the best performance (inner products 0.423 and 0.246).

Because the selected 2×1×2 model only reflects the average performance across environments, it is useful to look in more detail at differences among the environments. This can be achieved by considering the results corresponding to the 2×2×2 model, even though it was not the optimal solution. The two components for genotypes and attributes were very similar in behaviour to those found with the 2×1×2 model, as was the first environment component. The second environment component accounted for 7.9% of the variation (Table 7) and it was a contrast between environments E1 and E7 and environments E3 and E6. Note that E3 and E7 had high productivity potential while the other two environments had low productivity potential (Figure 1).

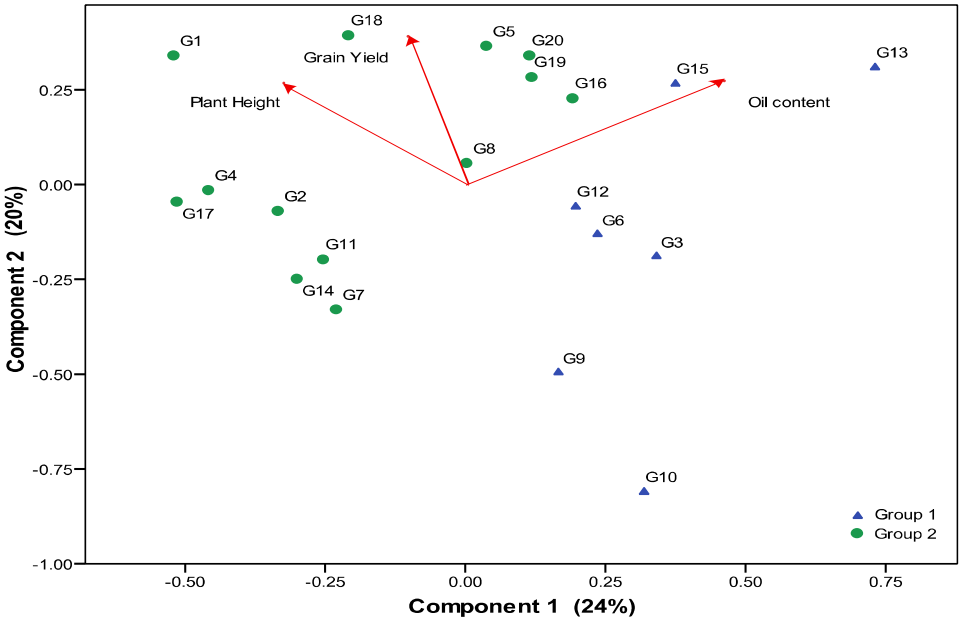


Figure 2. Joint plot for genotypes and attributes for the first environment component for the 2×1×2 solution.

Table 6. Inner products between genotypes and attributes.

Group	Genotype	Grain yield	Oil content	Plant height
1	G3	-0.108	0.106	-0.161
	G6	-0.075	0.073	-0.112
	G9	-0.210	-0.059	-0.187
	G10	-0.350	-0.075	-0.321
	G12	-0.043	0.075	-0.080
	G13	0.046	0.423	-0.154
	G15	0.066	0.246	-0.050
2	G1	0.186	-0.148	0.260
	G2	0.007	-0.174	0.090
	G4	0.041	-0.216	0.145
	G5	0.139	0.117	0.086
	G7	-0.105	-0.197	-0.013
	G8	0.022	0.016	0.015
	G11	-0.051	-0.171	0.029
	G14	-0.066	-0.207	0.031
	G16	0.070	0.151	0.000
	G17	0.035	-0.250	0.155
	G18	0.175	0.011	0.174
	G19	0.099	0.132	0.038
	G20	0.121	0.146	0.055

Table 7. Components for environments in the 2×2×2 model.

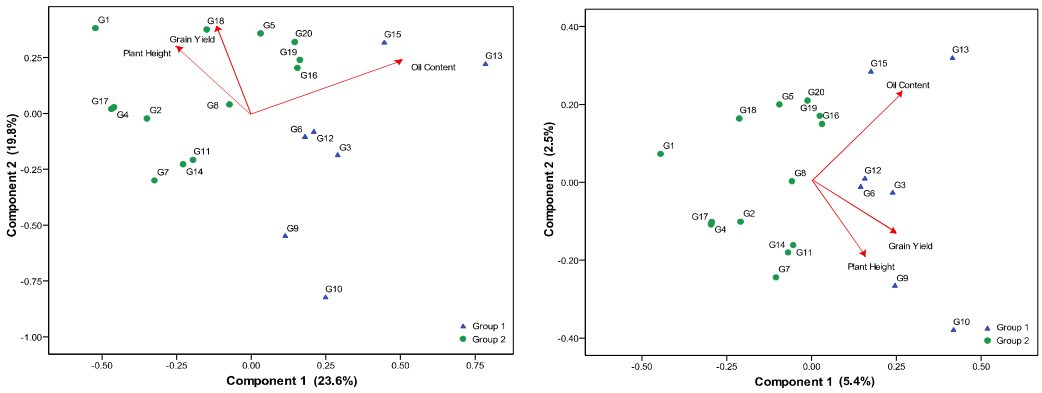
Environment	1	2
E1	0.244	0.54
E2	0.234	0.088
E3	0.534	-0.377
E4	0.385	-0.087
E5	0.397	0.14
E6	0.446	-0.373
E7	0.257	0.556
E8	0.174	0.289
Variation %	43.4	7.9

Figure 3 shows the joint plots for genotypes and attributes for the first and second environment component. Figure 3(a) is very similar to Figure 2 (as would be expected), while Figure 3(b) shows how the relationships between attributes and genotypes are different for environments E1 and E7 in comparison with environments E3 and E6. Compared with their overall performance in all environments, G9 and G10 had lower yields and heights in E3 and E6 than in E1 and E7, while G1 and G18 had relative higher yields and heights in E3 and E6 than in E1 and E7 (Table 8). Furthermore, G13 and G15 had higher oil contents in E1 and E7 than in E3 and E6, with the reverse pattern for G17, G4 and G1 (Table 8).

Table 8. Inner products between genotypes* and attributes (second joint plot).

Group	Genotype	Grain yield	Oil content	Plant height
1	G9	0.095	0.003	0.089
	G10	0.152	0.022	0.138
	G13	0.061	0.184	0.005
	G15	0.006	0.112	-0.026
2	G1	-0.119	-0.101	-0.084
	G2	-0.038	-0.079	-0.014
	G4	-0.059	-0.102	-0.027
	G7	0.006	-0.085	0.030
	G17	-0.059	-0.104	-0.026
	G18	-0.074	-0.018	-0.065

* Only genotypes with at least one value ≥ |0.07|



(a) First environment component (b) Second environment component

Figure 3. Joint plot for genotypes and attributes for the first and second environment component for the 2x2x2 solution.

DISCUSSION

The information obtained from joint analysis of sunflower data can be summarised in the following way.

The methods successfully integrate the attributes measured in the multi-environment trial. The analysis helps the breeder make decisions in favour of moderate to high yields and oil content with low to moderate heights in general or in selected environments. In this particular case, selecting the "best" genotypes from group 2 according to high yield, acceptable height and adequate oil content, and the best from group 1 on the basis of good oil content and reasonable yield and height would be G5, G18 (Group 2) and G13 (Group 1). In terms of specific environmental responses, G10 yields higher in environments like E1 and E7 (in comparison with those like E3 and E6), while the reverse is true for G1 and G18. On the other hand, G13 has higher oil content in environments like E1 and E7 (in comparison with those like E3 and E6). Both methods facilitated a simple interpretation about the relationship between and within the groups and this is of practical assistance to breeders.

While the groups are easier to look at than the individual genotypes, it should be remembered that the selection has to be made on individuals. The major advantage of these three-way methods is that they allow the breeders to simultaneously analyse the three-way data in a direct way and obtain a global vision of the genotype response across the different environments. Additionally, the information provided by both the clustering and ordination methods can be easily presented in figures which facilitate a relatively simple interpretation.

ACKNOWLEDGEMENTS

The first author thanks the National Council of Technological and Scientific Development, CNPq, Brazil, and the Academy of Sciences for the Developing World, TWAS, Italy, (CNPq-TWAS program) for the financial support.

REFERENCES

Araújo L.B., Varela M., Araújo M.F.C., Dias, C.T.S. (2011). Multiattribute response of maize genotypes tested in different coastal regions of Brazil. *International Journal of Agronomy* Vol. 2011, ID 215843, 1–6.

Basford K.E. (1982) The use of multidimensional scaling in analysing multi-attribute genotype response across environments. *Australian Journal of Agricultural Research* 33, 473–480.

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

Basford K.E., Kroonenberg P.M., Delacy I.H., Lawrence P.K. (1990). Multiattribute evaluation of regional cotton variety trials. *Theoretical and Applied Genetics* 79, 225–234.

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

Denis J.B., Moro J. (1995). Multivariate generalizations for modeling two-way interaction. I. Defining and estimating models. *Biuletyn Oceny Odmian* 26–27, 43–56.

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

Gabriel K.R. (2002). Le biplot-outil d'exploration de données multidimensionelles. *Journal de la Societe Francaise de Statistique* 143, 5–55.

Kiers H.A.L., der Kinderen A. (2003). A fast method for choosing the numbers of components in Tucker3 analysis. *British Journal of Mathematical and Statistical Psychology* 56, 119–125.

Kroonenberg P.M. (1983). Three-mode principal component analysis: *Theory and applications*. Leiden, DSWO Press.

Kroonenberg P.M. (2008). *Applied multiway data analysis*. New Jersey, Wiley-Interscience.

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

Kroonenberg P.M., De Leeuw J. (1980). Principal component analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika* 45, 69–97.

Moro. J., Denis J.B. (1995). Multivariate generalizations for modeling two-way interaction. II. Interpreting models and examples. *Biuletyn Oceny Odmian* 26–27, 57–72.

Timmerman M.E., Kiers H.A.L. (2000). Three-mode principal components analysis: Choosing the numbers of components and sensitivity to local optima. *British Journal of Mathematical and Statistical Psychology* 53, n.1, 1–16.

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

van Eeuwijk F.A., Kroonenberg P.M. (1995). The simultaneous analysis of genotype by environment interaction for a number traits using three-way multiplicative modelling. *Biuletyn Oceny Odmian* 26–27, 83–96.

van Eeuwijk F.A., Kroonenberg P.M. (1998). Multiplicative models for interaction in three-way ANOVA, with applications to plant breeding. *Biometrics* 54, 1315–1333.

Varela M., Torres V. (2005). Aplicación del análisis de componentes prinicipales de tres modos en la caracterización multivariada de somaclones de king-grass. *Revista Cubana de Ciencia Agrícola* 39, 543–552.

Varela M., Crossa J., Rane J., Joshi A.K., Trethowan R. (2006). Analysis of three-way interaction including multiattributes. *Australian Journal of Agricultural Research* 57, 1185–1193.

Varela M., Vicente J.L., Galindo P., Blázquez A., Castillo J.G., Estévez A. (2008). Una generalización de los modelos AMMI basada en el algoritmo de Tuckals3 para el análisis de componentes principales de tres modos. *Cultivos Tropicales* 29, 69–72.

Varela M., Crossa J., Joshi A.K, Cornelius P.L., Manes Y. (2009). Generalizing the sites regression model to three-way interaction including multi-attributes. *Crop Science* 49, 2043–2057.