ELSEVIER

Field Crops Research 78 (2002) 165–183

Field Crops Research

www.elsevier.com/locate/fcr

Investigating the physiological bases of predictable and unpredictable genotype by environment interactions using three-mode pattern analysis

Abelardo J. de la Vega[a,*], Antonio J. Hall[b], Pieter M. Kroonenberg[c]

[a]*Advanta Semillas S.A.I.C., Ruta Nac. 33 Km 636, CC 294, 2600 Venado Tuerto, Argentina*
[b]*IFEVA, Facultad de Agronomía, Universidad de Buenos Aires, CONICET, Av. San Martín 4453, 1417 Buenos Aires, Argentina*
[c]*Department of Education, Leiden University, Wassenaarseweg 52, 2333 AK Leiden, The Netherlands*

Received 22 February 2002; received in revised form 20 June 2002; accepted 20 June 2002

Abstract

Understanding of the underlying physiology of the genotype-specific responses to predictable and unpredictable environmental variation would improve the efficiency of selection within a complex target population of environments. Three-mode principal component analysis (PCA) can be used for interpreting the complex three-way (genotypes, environments, attributes) trial datasets from which this understanding should emerge. The efficiency of this method largely depends on the right combination between the biological and statistical models used, especially on the attributes selected to describe the genotypic responses and the centring of the three-way input data. In this study, we assessed the scope of yield determination models and double-centring of input data for generating some physiological understanding of the genotype × environment (G × E) interactions observed in a sunflower genotype–environment system and for developing ideotype-based breeding strategies. Double-centring of the three-way arrays permitted the separation of predictable and unpredictable G × E interactions. This, in combination with the use of models that explain the physiological bases of yield variation among genotypes, has served to identify three relevant sources of genotypic variation for use in a breeding program, namely: (i) attributes that can be selected to achieve specific adaptation to the target environment by emphasising predictable interactions (e.g. duration of grain filling, a trait associated with canopy stay green); (ii) attributes that allow the unpredictable G × E interactions to be accommodated, improving the linkage between managed-environments and target production environments (e.g. grain set); and (iii) genotypes of similar response pattern for yield but contrasting relative behaviour for the primary and secondary yield determinants. Breeding projects involving crosses between these genotypes could generate better opportunities for yield improvements for individual mega-environments.

Keywords: Three-mode principal component analysis; Genotype × environment interaction; Sunflower; Centring and normalisation; Yield components

1. Introduction

Effective identification of superior genotypes in multi-environment trials is generally complicated by the presence of significant genotype × environment

---

[*] Corresponding author. Tel.: +54-3462-435-235; fax: +54-3462-435-231.
E-mail address: avega@waycom.com.ar (A.J. de la Vega).

PII: S0378-4290(02)00123-5

(G × E) interactions, such that relative cultivar yields vary across testing environments (Cooper and DeLacy, 1994). Testing regimes may consider such interactions as a source of error or bias in assessing the overall superiority of a genotype. Alternatively, G × E interactions can be viewed as a reflection of differences in genotype adaptation, which may be exploited by selection and/or by adjustments in the testing strategy (Basford et al., 1996).

It has been argued that some understanding of the nature of the G × E interactions is needed to use them effectively through appropriate breeding methodologies (Basford and Cooper, 1998 and references therein). The concept of repeatability or predictability is critical to understand and use adaptations associated with G × E interactions. Allard and Bradshaw (1964) distinguished between predictable and unpredictable interactions, largely associated with locations or management regimes and years, respectively. Adequate models of target genotype–environment systems would enable plant breeders to exploit positive components of G × E interactions, i.e. breeding for specific adaptation associated with predictable interactions, and avoid the pitfalls of the disruptive effects of the sampling variation that are often associated with multi-environment trials, i.e. unpredictable interactions (Basford and Cooper, 1998).

While the record shows that breeding advance can be achieved empirically, more rapid progress could be achieved when some understanding of the physiological bases of crop performance is established and selection criteria are defined in terms of specific traits, i.e. ideotype-based breeding (Lawn and Imrie, 1991). Likewise, some understanding of the underlying physiology of the genotype-specific responses to predictable and unpredictable environmental variation should improve the efficiency of selection within a complex target population of environments.

The study of the physiological bases of G × E interactions implies the analysis of the genotype relative responses across environments for multiple attributes and thus requires a three-way data set of genotypes × environments × attributes (G × E × A). Three-mode principal component analysis (PCA) (Tucker, 1966; Kroonenberg, 1983), an extension of standard PCA to handle such three-way datasets, has been used for handling genotypes, environments and attributes simultaneously (Kroonenberg and Basford, 1989; Basford et al., 1990; Crossa et al., 1995; Chapman et al., 1997), allowing an examination of the relationships between genotypes and attributes associated with specific patterns of environmental variability. Basford and Cooper (1998) highlighted the importance of thinking about the right combination between the biological and statistical models used for generating a more comprehensive understanding of G × E interactions. The decision of which biological–statistical model combination is appropriate to any particular genotype–environment system depends on the researcher's objectives and assessment of the origin of the variability of the data. In the case of three-mode PCA, probably the most relevant biological challenge is the right choice of attributes to describe the genotype-specific responses to contrasting environments. From the statistical point of view, the treatment of the raw data, i.e. centring and normalisation of the G × E × A input data arrays, is a central decision to obtain the 'best' analysis for a particular dataset (Kroonenberg, 1983, Chapter 6). In three-way data, different types of centring and normalisation lead to different solutions. There are also many more ways of centring and normalising to choose from, compared to two-way data. Kroonenberg (1983, Chapter 6) and Harshman and Lundy (1984, pp. 225–253) give detailed discussions on this issue.

Crop yield in a given environment can be explained in terms of its determinants or components. Simple analytical crop models of yield determination probably offer the most scope for improving the efficiency of selection in breeding programs, because an improved physiological understanding of genotype variation in yield performance under differing environmental conditions can be achieved (Wright et al., 1996). However, this criterion is not commonly used to select attributes in the three-mode PCA literature, with the exception of de la Vega and Chapman (2001), who used the primary components and determinants of oil yield (i.e. grain number, grain weight and grain oil proportion; oil-corrected biomass and harvest index) for explaining the bases of G × E interaction for sunflower in two regions of Argentina.

The main purpose of centring is to eliminate from the analysis those means that should not be modelled multiplicatively. In order to understand the implications of different types of centring, one may write three-way data as if they were generated by a three-way analysis of variance:

$$x_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \alpha\beta_{ij} + \alpha\gamma_{ik} + \beta\gamma_{jk} + \alpha\beta\gamma_{ijk}$$

where x_{ijk} is the value of genotype i in the environment j for attribute k; μ the expected mean over all genotypes, environments and attributes, and the remaining terms are the relevant main effects and associated interactions (see also Van Eeuwijk and Kroonenberg, 1998). In all previous studies where three-mode PCA was applied to crops (Kroonenberg and Basford, 1989; Basford et al., 1990; Crossa et al., 1995; Basford et al., 1996; Rincon et al., 1997; Chapman et al., 1997; de la Vega and Chapman, 2001; Cooper et al., 2001), the G × E × A arrays were *jk*-centred (i.e. the G × E × A data were centred by subtracting the across-genotypes environment mean for each attribute; Kroonenberg, 1983, Chapter 6). This led to

$$x'_{ijk} = \alpha_i + \alpha\beta_{ij} + \alpha\gamma_{ik} + \alpha\beta\gamma_{ijk}$$

where the environmental effects, i.e. $\bar{x}_{jk}$, were removed and the genotypic means were maintained within attributes, as well as the G × E and G × A two-way interactions and the G × E × A three-way interaction. This type of centring is useful in pattern analysis to display both genotypic and G × E interaction effects on two-mode biplots. However, in three-mode PCA, uncentred modes tend to have large 1st components, due to the presence of the means, and these complicate interpretation of G × E interactions by confounding their effects with the main effect of the uncentred mode, i.e. genotypes. Consequently, in the cited studies, the large 1st environment components, which accounted for more than 50% of variation in all cases, explained the common pattern over environments, i.e. the genotypic effect, and the 2nd components contrasted the environment types associated with the main pattern of G × E interaction. When locations (Basford et al., 1990), water regimes (Chapman et al., 1997), or regions (de la Vega and Chapman, 2001) were replicated, the 2nd environment component accounted for the predictable portion of the G × E interactions. Unpredictable G × E interactions, which can severely complicate selection, were not retained and described by those models. Our hypothesis is that the *ik*, *jk*-centring (i.e. the G× E × A data are centred within attributes by subtracting both the across-genotypes environment means and the across-environments genotype means, and adding the overall mean; Kroonenberg, 1983, Chapter 6) of the G × E × A arrays, which leads to

$$x''_{ijk} = \alpha\beta_{ij} + \alpha\beta\gamma_{ijk}$$

where both genotypic and environmental effects are removed from the data, and the G × E and G × E × A interaction remain, allows the derivation of environmental components that explain different sources of G × E interaction.

Normalisation, which is also named scaling or standardisation, is the process of equalising sums of squares. This process is necessary since attributes are recorded in different scales. Normalisation across slices (Kroonenberg, 1983, Chapter 6), which in this type of study implies the division of the centred data by the standard deviation of each attribute over all genotypes and sites (i.e. $r_{ijk} = x''_{ijk}/S_{..k}$; Kroonenberg and Basford, 1989; Basford et al., 1990), is the most commonly used in three-mode PCA. However, Fox and Rosielle (1982) and Cooper and DeLacy (1994) recommended normalising within environments (i.e. $r_{ijk} = x''_{ijk}/S_{.jk}$), because this type of normalisation causes each environment to have a mean of zero and a standard deviation of 1, this being the most appropriate treatment for reducing the influence of environmental main effects. In further G × E × A three-mode analyses, this criterion was implemented (Basford et al., 1996; Chapman et al., 1997; de la Vega et al., 2001; Cooper et al., 2001).

The objective of this study is to assess the scope of biological and statistical models aimed at improving the efficiency of three-mode PCA for generating a partial physiological understanding of G × E interactions and developing ideotype-based breeding strategies which emphasise predictable interactions and avoid the confounding effect of unpredictable interactions. The models evaluated are: (i) the use of the primary (i.e. grain number, grain weight and oil concentration) and secondary (i.e. floret number, seed set, rate of grain filling, duration of grain filling, kernel proportion and kernel oil proportion) yield components for characterising the genotype responses to contrasting environments; and (ii) the *ik*-, *jk*-centring (Kroonenberg, 1983, Chapter 6) of the G × E × A arrays, i.e. removing both genotypic and environmental effects from the data, allowing the derivation of environmental components that explain different sources of G × E interaction.

The case study consists of a reference set of sunflower hybrids grown in Venado Tuerto (Central Argentina) during 2 years in normal and late planting dates. The hybrids composing the reference set exhibit differential adaptation to Northern and Central regions of Argentina and to normal (October) and late (December) plantings at Venado Tuerto (de la Vega et al., 2001; de la Vega and Hall, 2002a,b). Northern and Central regions are different mega-environments for sunflower, i.e. their differences in terms of genotypic discrimination for oil yield are large and repeatable. Mega-environments were defined as the largest subunits of a crop's growing or target environment within which a particular variety or related practice was useful (CIMMYT, 1989). December planting influences the relative performance of sunflower genotypes in a similar fashion to the Northern mega-environment and could be used as a managed-environment for indirect selection for that region (de la Vega et al., 2001). A potential limitation of this strategy is the likelihood of excluding important genes conditioning adaptation to an undefined challenge/stress not sampled in the managed-environment (Cooper et al., 1995). This would lead selecting genotypes not adapted to the target environment. On the other hand, challenges or stresses specific to the managed-environments could lead discarding valuable hybrids for the target production environment.

In this study, the four levels of the environment (E) mode are a fully crossed design of two levels of planting date (S) and two levels of year (Y). Since planting date is under control of the experimenter, we will consider the portion of the G × E interaction due to the G × S interaction as predictable. Since the influence of the year effect is unpredictable, the portion of the total G × E interaction involving it, i.e. the G × Y and G × S × Y interactions, is considered as unpredictable. Traits conditioning adaptation to December plantings across years, revealed as predictable by three-mode PCA, can be used as phenotypic indicators of specific adaptation to the Northern mega-environment. Unpredictable interactions will result in reduced phenotypic correlation between December plantings and the Northern region in individual years, negatively affecting the efficiency of this managed-environment for indirect selection. The physiological understanding of these interactions would lead to the definition of ideotype-based selection criteria aimed at reducing the risk of discarding valuable genotypes in those years.

2. Materials and methods

2.1. Cultural details

Supplementary irrigated crops of a reference set of nine sunflower hybrids (Table 1) were grown on a deep coarse loam soil (Typic Hapludoll) at the Advanta Semillas Research Center, Venado Tuerto, Argentina (33°41′S, 61°57′W), using October (Oct) and December (Dec) planting dates during 1996/1997 (96) and 1998/1999 (98) seasons. The details of the experimental material, test environments, experimental design, and data collection are given in de la Vega and Hall (2002a,b). The hybrids composing the reference set of genotypes were selected from the Advanta Argentina testing program based on their contrasting relative performance across environments for oil yield. This set includes commercial hybrids widely planted in the Central region, i.e. Contiflor 15, Contiflor 9 and TC 2001, commercial hybrids widely planted in the Northern region, i.e. Morgan 734 and Aguará, and experimental hybrids that showed different patterns of adaptation to those of the commercial hybrids. Taken together, this set of hybrids represent a wide range of genetic diversity according to the origin of their genetic backgrounds and to restricted fragment length polymorphism (RFLP) molecular marker analyses (A. Leon, Advanta Argentina, Balcarce, unpublished data). The years sampled are not a random sample of the environmental conditions at Venado Tuerto. 1996/1997 and 1998/1999 were classified as neutral and La Niña years, respectively, in terms of the El Niño Southern Oscillation effect. Based on this effect, 1996/1997 can be considered representative of the typical summer season at Venado Tuerto. In 1998/1999, rainfall during grain filling for the December planting was higher and incident radiation was lower than that expected in a neutral year (de la Vega and Hall, 2002a). Although the four trials analysed here represent a limited sample of environments, these October and December plantings showed the same pattern of genotypic discrimination for oil yield as found in a much larger analysis (21 trials in 3

Table 1
Mode component scores (with adequacy of fit) for four attributes (oil yield and its primary components) of nine sunflower hybrids grown over four environments in Venado Tuerto, Argentina (3 × 3 × 2 component model for G × E × A)

Code	Name	Group[a] (2-mode)	Group[b] (3-mode)	Component scores			Proportion of SS explained
				1	2	3	
Genotypes (G)							
1	Contiflor 15	2	2	−0.68	−0.24	−0.25	0.68
2	Aguará	1	1	0.37	0.28	0.33	0.25
3	GV23105	1	3	0.49	0.02	−0.59	0.87
4	GV25086	3	2	−0.18	0.06	0.10	0.09
5	TC 2001	2	2	−0.94	0.15	−0.34	0.92
6	GV23146	3	1	0.09	0.96	0.15	0.79
7	GV22510	3	3	0.65	−0.17	0.23	0.56
8	Contiflor 9	2	2	−0.60	−0.53	0.54	0.80
9	Morgan 734	1	3	0.80	−0.52	−0.17	0.70
Proportion of SS explained				0.35	0.18	0.12	0.65
Environments (E)							
Oct 96	October 96			−0.73	−0.39	0.15	0.71
Dec 96	December 96			0.84	0.24	0.12	0.78
Oct 98	October 98			−0.71	0.43	−0.12	0.70
Dec 98	December 98			0.41	−0.45	−0.18	0.40
Proportion of SS explained				0.48	0.15	0.02	0.65
Attributes (A)							
OY	Oil yield (kg ha^{-1})			0.84	0.05		0.71
GN	Grain number per square meter			0.41	−0.62		0.55
GW	1000 grain weight (g)			0.06	0.76		0.58
O%	Grain oil concentration (%)			0.85	0.19		0.76
Proportion of SS explained				0.40	0.25		0.65

[a] Genotype groups identified by hierarchical agglomerative clustering of oil yield across 21 environments of Argentina (de la Vega et al., 2001)—1: Northern-adapted, 2: central-adapted, 3: broadly adapted.

[b] Genotype groups identified by mixture maximum likelihood cluster method of oil yield and its primary components across four environments.

years) of yield trials conducted in Central and Northern regions of Argentina (de la Vega et al., 2001).

For the analysis of the hybrid relative responses to environments, oil yield was considered as the product of its primary and secondary components. Primary components of sunflower oil yield are grain number (achene per square meter), grain weight (mg per achene) and grain oil concentration (%). Secondary components are those affecting primary ones. Thus, grain number is a function of floret number per square metre and seed set (%), i.e. the ratio between filled grains and florets. Grain weight may be regarded as the outcome of the product of rate (mg per achene per day) and duration (days) of grain filling. Sunflower grain comprises pericarp (hull), derived from the ovary wall, and the kernel, which is mostly embryo (Knowles, 1978). As most of the grain oil is deposited in the kernel, grain oil concentration is affected by kernel percentage (%, i.e. kernel mass/grain mass × 100) and kernel oil concentration (%).

For three-mode multivariate analysis, the collected data was summarised in the form of G × E × A arrays of means for the following models of yield analysis:

(1) Oil yield = grain number × grain weight × grain oil concentration.

(2) Oil yield = floret number × seed set % × grain filling rate × grain filling duration × kernel proportion × kernel oil proportion.

Prior to three-mode analyses, the three-way arrays of means (G × E × A) for each of the analytical models were corrected for hybrid and trial main effects within attributes (residual from additivity, Gabriel, 1978), as in the additive main effects and multiplicative interaction (AMMI) model for two-way tables (Gauch, 1988). Next, the residual arrays were normalised within environments (Fox and Rosielle, 1982; Cooper and DeLacy, 1994). Summarising, the centring and normalisation model applied to each x_{ijk} was

$$r_{ijk} = \frac{x_{ijk} - x_{.jk} - x_{i.k} + x_{..k}}{S_{.jk}}$$

With the aim of simplifying comparisons in three-mode PCA, genotypes having similar response patterns for each of the attributes analysed across testing environments were grouped using the mixture maximum likelihood cluster method (Basford and McLachlan, 1985).

2.2. Mixture maximum likelihood cluster method

This non-hierarchical clustering method is a model-based technique which can be applied to group genotypes that have similar response patterns for each of the attributes analysed across testing environments (Basford and McLachlan, 1985; McLachlan and Basford, 1988; Basford et al., 1996). In this method, genotypes are assumed to be a sample from a mixture of various proportions of a specific number of populations (clusters). 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 in all environments, although it can differ among clusters. The unknown parameters of the groups (i.e. proportions, mean vectors and covariance matrices) are estimated using the maximum likelihood principle and non-overlapping groups (clusters) are obtained by allocating each genotype to the group to which it has the highest estimated probability of belonging. This is calculated by replacing the unknown parameters with their likelihood estimates in the expression for the true posterior probability of belonging to that group. The number of underlying groups or clusters must be specified. From a given starting allocation of the genotypes into groups, the EM algorithm (Dempster et al., 1977) ensures that convergence to a maximum likelihood occurs. In applying this clustering method, either group-specific covariance matrices between attributes can be chosen (used here) or a common covariance matrix for all groups (Basford and McLachlan, 1985; Basford et al., 1996; Chapman et al., 1997). The maximum likelihood method of clustering was applied to the 9 × 4 × 4 (G × E × A) raw data array of model (1) (primary components of oil yield) using the FORTRAN program MIXCLUS3[1], an updated version of that appearing in the appendix of McLachlan and Basford (1988).

2.3. Three-mode PCA

This procedure derives components, i.e. linear combinations of the levels of the modes, for each of the three modes (say, P, Q, and R components for genotypes, environments, and attributes, respectively). It can be assumed that these components together contain the only relevant systematic variation of the three-way array dataset. The components of the three modes can be labelled on the basis of the patterns shown by the levels with high loadings on such components. In this model, each mode is allowed to have a different number of components. The number of components for each mode needs to be simultaneously determined for all modes. Therefore, several solutions have to be inspected to come to an adequate description of a dataset (Kroonenberg, 1983, Chapter 2; Kroonenberg and Basford, 1989).

A three-way array of order P by Q by R (the core array) contains the weights assigned to each of the combinations of the components for the three modes. The complete model is written as (Kroonenberg, 1983, Chapter 2; Kroonenberg and Basford, 1989):

$$x_{ijk} = \sum_{p=1}^{P}\sum_{q=1}^{Q}\sum_{r=1}^{R} a_{ip}b_{jp}c_{kr}g_{pqr} + e_{ijk}$$

where a_{ip} represents the coefficient for genotype i in the genotypic component p $(p = 1, \ldots, P)$, b_{jq} the

---

[1] This program can be obtained from Prof. K.E. Basford, University of Queensland, Australia: http://biometrics.ag.uq.edu.au/software.htm.

coefficient for environment j in the environmental component q $(q = 1, \ldots, Q)$, and c_{kr} the coefficient for trait k in the trait component r $(r = 1, \ldots, R)$. The term g_{pqr} indicates the joint weight for the pth component of the genotypic mode, the qth component of the environmental mode, and the rth component of the attribute mode, and its squared value indicates the variation explained for that combination of components (Kroonenberg, 1983, Chapter 2; Kroonenberg and Basford, 1989). In this way, the core array represents a partitioning of the overall fitted sum of squares into small units through which the complex relationships between the components can be analysed (Kroonenberg, 1983, Chapter 6)[2].

It is very instructive to investigate the component loadings of the genotypes jointly with the component loadings of the attributes, by projecting them together in one space, as it then becomes possibly to display the interaction between genotypes and attributes. The plot of the common space is called a joint biplot, a variant of Gabriel's (1971) biplot (Kroonenberg, 1983; Basford et al., 1996), and is constructed from the core matrix as follows. For each component r of the environment mode (E), the genotype components (G) and the attribute components (A) are scaled by dividing the core slice associated with that component r between them (using singular value decomposition), and weighting the scaled G and A by the relative number of elements in the modes to make the distances comparable. For more detailed discussion see Kroonenberg (1983, Chapter 6).

Three-mode PCA was applied to the $9 \times 4 \times 4$, and $9 \times 4 \times 7$ $(G \times E \times A)$ environment-standardised residual arrays for the yield determination models (1) and (2), respectively, using the program TUCKALS3[3] (Kroonenberg, 1994). Given the complementary character of clustering and ordination techniques (Basford et al., 1996), the genotype groups derived from the mixture maximum likelihood cluster method were superimposed on the bi-dimensional joint biplots derived from the three-mode PCA, in order to enhance the interpretation of the interactions.

The efficiency of the strategies of indirect selection for the Northern region which use attributes selected on the basis of results of three-mode analyses was quantified by the correlation which measured the similarity of genotype discrimination between the indirect selection regime proposed and that for the average performance in the Northern production environment. For that purpose, mean genotypic discrimination of the Northern mega-environment was calculated from 14 trials conducted between 1996/1997 and 2000/2001. The trials from 1996/1997 to 1998/1999 are those described for the Northern region in de la Vega et al. (2001). Additional trials included in this analysis were conducted in Margarita (1999/2000 and 2000/2001), Reconquista (1999/2000 and 2000/2001), and Sáenz Peña (2000/2001).

3. Results and discussion

3.1. Mixture maximum likelihood cluster method

After mixture cluster analysis of oil yield and its three primary components (i.e. grain number, weight and oil concentration), the group composition of the genotypes (Table 1) showed some changes from the initial allocation based on hierarchical agglomerative cluster analysis of oil yield across 21 growing environments of Argentina (de la Vega et al., 2001). Five of the nine genotypes maintained their original group allocation. The major difference in group membership was in the distribution of the Northern- and broadly adapted hybrids which showed relatively good performance in December plantings (Table 1). Two-mode clustering defined two groups of hybrids specifically adapted to Northern and Central mega-environments, and a third group of broadly adapted hybrids (de la Vega et al., 2001). The use of MIXCLUS technique did not generate any interchanges in categorisation among elements of the two specifically adapted groups. On the other hand, shifts between the broadly adapted group and that specifically adapted to the Northern or the Central mega-environments or vice versa did occur. In the three-mode clustering, the broadly adapted hybrid 4 and the central-adapted hybrids were joined in a single group (Group 2, Table 1). The broadly adapted hybrids 6 and 7 and the Northern-adapted hybrids were re-allocated in two

---

[2] The partitioning of the core array into explained variability per element is only true if the three modes are orthonormal (see Ten Berge et al., 1987).

[3] This program can be obtained from the third author: http://www.fsw.leidenuniv.nl/~kroonenb.

groups (Groups 1 and 3, Table 1). This pattern of reallocation is not surprising in view of narrower differences between broadly adapted and either of the specifically adapted categories in comparison to differences between specifically adapted categories.

3.2. Three-mode PCA model fit and description of components

3.2.1. Primary components of oil yield

The three-mode model with 3 × 3 × 2 components for genotypes, environments and attributes, respectively, was considered adequate for fitting the data ($r^2 = 0.65$, Table 1), on the basis of informal judgements of the increases in r^2 as compared to the increases in dimensions and difficulty of interpretation. In this model, the three components for the genotype mode accounted for 35, 18 and 12% of the variation, respectively; the three components for the environment mode accounted for 48, 15 and 2%, respectively; and the two components for the attributes accounted for 40 and 25% of the variation, respectively (Table 1). Not all genotypes, environments and attributes were fitted equally well by the model. For entries 2 and 4, the model accounted for considerably less than 50% of the variability in their response compared to the overall fit of 65% (Table 1). The 1st genotype component (G_1) distinguishes between the hybrids belonging to the MixClus3 Group 2 and those of Groups 1 and 3 (Table 1). The 2nd genotype component (G_2) contrasts hybrids 2 and 6 vs. 8 and 9, and the 3rd genotype component (G_3) reflects the contrast between hybrids 2 and 8 vs. 3 and 5 (Table 1).

Environments were generally well fitted by three components with more than 70% of the variation accounted for in most of them, except for Dec 98 (Table 1). Even though there were differences in fit among the attributes, the model accounted for more than 50% of the variation in all of them (Table 1). The 1st environment component (E_1) contrasted Oct and Dec plantings across years (G × S), i.e. the predictable G × E interaction. The 2nd environment component (E_2) accounted for the G × S × Y interaction, and the 3rd environment component (E_3) explained the G × Y interaction, i.e. $E_2 + E_3$ accounted for the unpredictable part of the G × E interactions. Even though E_3 only explained a small portion of the total variation, it was maintained in the model to show how this procedure can separate different portions of the G × E interactions.

The 1st attribute component (A_1) reflects the overall variability of the scores in the levels of this mode, and could be labelled as ‘all attributes together’. This variability is larger for oil yield and grain oil concentration than for grain number and grain weight. The 2nd component of the attribute mode (A_2) is dominated by a strong contrast between grain number and grain weight (Table 1).

3.2.2. Secondary components of oil yield

The three-mode model with 3 × 3 × 3 components for genotypes, environments and attributes, respectively, was considered adequate for fitting the data ($r^2 = 0.65$). The proportion of the total variation explained by the components of each mode, the loadings on the components, and the sum of squares retained by the model for the individual genotypes, environments and attributes are detailed in Table 2.

The 1st genotype component (G_1) again distinguishes between the hybrids belonging to the MixClus3 Group 2 and those of Groups 1 and 3. The 2nd genotype component (G_2) again contrasts hybrids 2 and 6 vs. 8 and 9, but the 3rd genotype component (G_3) contrasts hybrids 2 and 5 vs. 4 and 6 (Table 2). The 1st environment component (E_1) again accounted for the G × S interaction; the 2nd environment component (E_2) again for the G × S × Y interaction; and the 3rd environment component (E_3) explained the G × Y interaction. The 1st attribute component (A_1) reflects the contrast between most attributes and floret number and grain filling rate; the 2nd component of the attribute mode (A_2) contrasts floret number vs. seed set and grain filling rate; and the 3rd attribute component (A_3) contrasts oil yield and grain filling rate vs. grain filling duration (Table 2).

3.3. The core array

The core array indicates the relations among the various components of the three modes. There are several ways in which the elements of the core arrays can be interpreted. In this paper, we will discuss the percentages of explained variation and the three-mode interaction measures, which are, in our judgement, the most useful practical tools for breeders and crop physiologists dealing with G × E × A interactions.

Table 2
Mode component scores (with adequacy of fit) for seven attributes (oil yield and its secondary components) of nine sunflower hybrids grown over four environments in Venado Tuerto, Argentina ($3 \times 3 \times 3$ component model for $G \times E \times A$)

Code	Name (MixClus3 group)	Component scores			Proportion of SS explained
		1	2	3	
Genotypes (G)					
1	Contiflor 15 (2)	−0.61	−0.03	−0.18	0.55
2	Aguará (1)	0.19	0.84	0.30	0.70
3	GV23105 (3)	0.65	−0.20	−0.14	0.59
4	GV25086 (2)	−0.28	−0.15	−0.39	0.37
5	TC 2001 (2)	−0.61	−0.08	0.76	0.84
6	GV23146 (1)	0.06	0.53	−0.34	0.45
7	GV22510 (3)	0.40	0.35	−0.08	0.35
8	Contiflor 9 (2)	−0.70	−0.52	−0.18	0.77
9	Morgan 734 (3)	0.90	−0.74	0.23	0.83
Proportion of SS explained		0.31	0.22	0.12	0.65
Environments (E)					
Oct 96	October 96	−0.78	0.17	0.21	0.69
Dec 96	December 96	0.75	−0.37	0.05	0.70
Oct 98	October 98	−0.72	−0.23	−0.25	0.63
Dec 98	December 98	0.39	0.64	−0.12	0.57
Proportion of SS explained		0.46	0.15	0.03	0.65
Attributes (A)					
OY	Oil yield (kg ha^{-1})	0.79	0.18	0.34	0.76
FN	Floret number per square meter	−0.05	−0.67	0.28	0.53
SS%	Seed set (%)	0.32	0.73	−0.13	0.65
RGF	Grain filling rate (g per achene per day)	−0.35	0.63	0.34	0.63
DGF	Grain filling duration (days)	0.74	−0.02	−0.35	0.66
K%	Kernel proportion (%)	0.79	−0.19	−0.10	0.67
KO%	Kernel oil proportion (%)	0.71	−0.03	0.33	0.62
Proportion of SS explained		0.36	0.21	0.08	0.65

For more detailed description of these and other alternative approaches to interpretation, see Kroonenberg (1983, Chapter 6).

3.3.1. *Primary components of oil yield*

3.3.1.1. Explained variation. The element $G_1E_1A_1$ of the core array (Table 3) indicates the strength of the relationship between the 1st components of the three modes, i.e. 34% of variation explained; $G_2E_2A_1$ indicates the strength of the relationship between the 2nd components of the G and E modes and the 1st component of the A mode, i.e. 4% of variation explained; and so on. From the 18 possible combinations of components of the three modes, which jointly explain 65% of the total variation contained in the system under study, four elements accounted for 58% of such variation, namely $G_1E_1A_1$ (34%), $G_2E_1A_2$ (12%), $G_3E_2A_2$ (8%), and $G_2E_2A_1$ (4%) (Table 3). This analysis of Table 3 is a clear example of how the inspection of the elements of the core array allows the separation of the pattern-rich variability from the noise-rich variability encountered in a $G \times E \times A$ system. Then, the pattern-rich variability can be partitioned as follows:

(i) *Due to $G_1E_1A_1$ (34%).* $G \times S$ interaction for all attributes together, which separates MixClus3 Group 2 hybrids from Group 1 and 3 hybrids (Table 3). This combination of components of the

Table 3
TUCKALS3 core array for the 3 × 3 × 2 (G × E × A) model for four attributes (oil yield and its primary components) of nine sunflower hybrids grown over four environments in Venado Tuerto, Argentina[a,b]

Components of the genotype mode	Components of attribute mode			
	Three-mode interactions		Proportion of variation explained	
	All attributes together (A_1)	Grain number vs. grain weight (A_2)	A_1	A_2
Environment component 1 (E_1): G × S interaction				
G_1	6.56	−0.64	0.34	0.00
G_2	0.06	3.85	0.00	0.12
G_3	−0.77	−1.59	0.01	0.02
$E_1 = 0.48$				
Environment component 2 (E_2): G × S × Y interaction				
G_1	0.02	1.16	0.00	0.01
G_2	2.32	1.48	0.04	0.02
G_3	0.07	3.23	0.00	0.08
$E_2 = 0.15$				
Environment component 3 (E_3): G × Y interaction				
G_1	0.22	−0.11	0.00	0.00
G_2	1.04	−0.44	0.01	0.00
G_3	0.95	−0.53	0.01	0.00
$E_3 = 0.02$				

[a] G: genotype, S: planting date, Y: year.

[b] G_1: Group 2 vs. Groups 1 and 3, G_2: genotypes 2 and 6 vs. 8 and 9, G_3: genotypes 2 and 8 vs. 3 and 5.

three modes accounted for the predictable G × E interactions for oil yield, which can be exploited through selection for specific adaptation.

(ii) *Due to $G_2E_1A_2$ (12%).* G × S interaction which contrasts hybrids 2 and 6 vs. 8 and 9 in terms of their relative responses for grain number and grain weight. This element of the core array accounted for the portion of the predictable G × E interactions associated with the relative responses of hybrids that showed similar patterns of oil yield across environments but achieve these by different physiological pathways. If these differences are a reflection of complementary favourable alleles (Dudley, 1984; Cheres et al., 1999), combining genotypes of similar adaptation profile but of different responses for yield-related traits should bring better opportunities for yield improvements in particular mega-environments. Thus, the relationships between the components associated with this element of the core array may contribute to a more objective selection of parents for breeding projects.

(iii) *Due to $G_3E_2A_2$ (8%) and $G_2E_2A_1$ (4%).* These combinations of components accounted for the unpredictable G × E interactions, i.e. G × S × Y interaction, which may have a disruptive effect on the efficacy of Dec planting as managed-environment for indirect selection for the Northern mega-environment. The analysis of the relationships of the levels the three modes associated to these elements of the core array may help to find phenotypic indicators aimed at avoiding this effect.

3.3.1.2. Three-mode interactions. The percentages of explained variation point to the dominant combinations, but the direction of the relationships must be found in the original, i.e. non-squared, core array (three-mode interactions in Table 3). For

$G_1E_1A_1$, e.g. the three-mode interaction between loadings on components is +6.56 (Table 3). This positive interaction reveals a relative improvement of a particular attribute for a given genotype in an individual environment. Hypothetically this could arise from any of four different combinations of elements of the three modes for (G_1, E_1, A_1), namely (−, −, +), (+, +, +), (−, +, −), and (+, −, −). As A_1 has only positive loadings (Table 1), the number of possible combinations reduces to the first two. This indicates the following. (i) Negative scores for G_1 (Table 1) occur together with negative scores for E_1 and positive scores for A_1. Consequently, the hybrids of the MixClus3 Group 2 showed an improved relative performance in Oct plantings for all attributes, especially for oil yield and grain oil concentration, and vice versa in Dec plantings. (ii) Positive scores for G_1 occur together with positive scores for E_1 and A_1. Consequently, the hybrids of the MixClus3 Groups 1 and 3 showed an improved relative performance in Dec plantings for all attributes. The attribute loadings on component A_1 (Table 1) indicate that oil concentration is strongly underlying the genotype-specific responses to planting date for oil yield. Grain oil concentration is an easy-to-assess trait, which make it suitable for using it as a selection indicator. Since differences in average performance for oil yield in the target population of environments of Argentina are strongly determined by grain oil concentration (de la Vega and Chapman, 2001), this trait is a useful indicator of broad adaptation. However, the strong genotypic effect shown by oil concentration in previous studies, i.e. non-crossover G × S interactions for grain number and size and oil concentration determine cross-over interactions for oil yield (de la Vega and Hall, 2002b), precludes its use as a simple indicator of specific adaptation.

For $G_2E_1A_2$, the positive three-mode interaction (3.85, Table 3) can be interpreted as follows. Hybrids 2–6 showed an improved relative performance for grain number in Oct plantings, i.e. (+, −, −) for (G_2, E_1, A_2) (Table 1), and for grain weight in Dec plantings, i.e. (+, +, +) for (G_2, E_1, A_2). Hybrids 1, 7–9 showed an improved relative performance for grain weight in Oct plantings, i.e. (−, −, +) for (G_2, E_1, A_2), and for grain number in Dec plantings, i.e. (−, +, −) for (G_2, E_1, A_2). Note that the hybrids adapted to Dec plantings, i.e. MixClus3 Groups 1 and 3, which showed similar patterns of adaptation for oil yield and consequently grouped together in G_1, can be separated based on their contrasting responses for grain number and grain weight in G_2. It is to be expected that, if these differences are the reflection of complementary favourable alleles (Dudley, 1984; Cheres et al., 1999), combining genotypes of same interaction direction on G_1 and opposite interaction direction on G_2 in breeding projects should bring higher opportunities of genetic progress than combining genotypes of the same interaction direction on both genotypic components. The first attempts we have made in this direction tend to confirm this hypothesis (data not shown). It is important to note, however, that the F2 populations from this type of crosses must be large enough to have a reasonable chance of identifying the desired genotypes in the segregating population.

The positive three-mode interactions of the core array elements $G_3E_2A_2$ (3.23, Table 3) and $G_2E_2A_1$ (2.32, Table 3) indicate the effect of the unpredictable G × E interactions, which could result in a reduced efficacy of Dec plantings for indirect selection. The Northern-adapted genotypes 2 and 6, e.g. showed a better relative performance for oil yield in Dec 96 (+, +, +) than in Dec 98 (+, −, +) (Table 1), for both the positive elements $G_3E_2A_2$ and $G_2E_2A_1$. This G × S × Y interaction for oil yield was determined by the relative responses of grain weight and grain oil concentration, according to the interaction directions for $G_3E_2A_2$ and $G_2E_2A_1$ (Table 1). However, for the positive $G_3E_2A_2$, both genotypes showed a relative increase in grain number in Dec 98, i.e. (+, −, −) (Table 1). In consequence, the environmental factor driving the observed G × S × Y interaction, which determined the relatively poor performance of hybrids 2 and 6 for oil yield in Dec 98, affected grain weight and oil concentration rather than grain number.

3.3.2. *Secondary components of oil yield*

3.3.2.1. Explained variation and three-mode interactions. Four elements of the core array for model (2), from the 27 possible combinations of components of the three modes, accounted for 52% of the total variation contained in the system (Table 4), namely $G_1E_1A_1$ (27%), $G_2E_1A_2$ (14%), $G_2E_2A_1$ (6%), and $G_3E_2A_2$ (5%). The combination of the 1st components

Table 4
TUCKALS3 core array for the 3 × 3 × 3 (G × E × A) model for seven attributes (oil yield and its secondary components) of nine sunflower hybrids grown over four environments in Venado Tuerto, Argentina[a]

Components of the genotype mode	Components of attribute mode					
	Three-mode interactions			Proportion of variation explained		
	All attributes vs. floret number and grain filling rate (A_1)	Floret number vs. seed set and grain filling rate (A_2)	Oil yield and grain filling rate vs. grain filling duration (A_3)	A_1	A_2	A_3
Environment component 1 (E_1): G × S interaction						
G_1	7.78	0.02	0.30	0.27	0.00	0.00
G_2	−0.39	5.50	1.19	0.00	0.14	0.01
G_3	0.49	1.18	−2.96	0.00	0.01	0.04
$E_1 = 0.46$						
Environment component 2 (E_2): G × S × Y interaction						
G_1	−1.29	0.75	1.84	0.01	0.00	0.02
G_2	−3.69	0.21	−1.24	0.06	0.00	0.01
G_3	1.19	3.26	−1.33	0.01	0.05	0.01
$E_2 = 0.15$						
Environment component 3 (E_3): G × Y interaction						
G_1	0.93	−0.46	−1.05	0.00	0.00	0.01
G_2	−0.59	−1.47	0.17	0.00	0.01	0.00
G_3	−1.15	0.96	−0.05	0.01	0.00	0.00
$E_3 = 0.03$						

[a] G_1: Group 2 vs. Groups 1 and 3, G_2: genotypes 2 and 6 vs. 8 and 9 , G_3: genotypes 2 and 5 vs. 4 and 6.

of the three modes, i.e. $G_1E_1A_1$, explains the G × S interaction for oil yield, which contrast the MixClus3 genotype Group 2 vs. Groups 1 and 3. These reflect the predictable G × E interactions that can be exploited through selection for specific adaptation. The positive three-mode interaction for this combination of components (7.78, Table 4) indicates that the hybrids of the MixClus3 Group 2 showed an improved relative performance for oil yield in Oct plantings and those of Groups 1 and 3 in Dec plantings. The high loadings of grain filling duration, kernel proportion and kernel oil proportion on component A_1 (Table 2) indicate that the genotype-specific responses for these attributes are strongly determining the observed G × S interaction for oil yield. The contrasting response patterns shown by floret number and grain filling rate indicate that these attributes are not underlying the genotype relative responses to planting date for oil yield. Kernel proportion and kernel oil proportion are difficult-to-assess traits in large breeding populations and do not conform to usual requirements for screening procedures (Shorter et al., 1991). However, the duration of grain filling in Dec plantings, which is also a hard-to-assess trait, is positively associated with the duration of green leaf area (canopy stay green) in post-anthesis (de la Vega and Hall, 2002a). This latter is a trait that can easily be evaluated by a breeder and can be used as a surrogate for duration of grain filling.

The positive element $G_2E_1A_2$ (5.50, Table 4) explains the portion of the G × S interaction which contrasts hybrids 2 and 6 vs. 8 and 9 in terms of their relative responses for floret number, grain set and grain filling rate (see loadings on components in Table 2). As in the case of the primary components of oil yield, this element of the core array accounted for the portion of the predictable G × E interactions that contrasts hybrids of similar response patterns for oil yield in terms of their relative responses for the oil yield components. The relative improvement of hybrids 2 and 6 in Dec plantings in terms of grain weight is explained by their relative increase in terms of rate of grain filling. By contrast, the improved performance of hybrids 8 and 9 in Dec plantings in terms of grain number is explained by their specific response to this environment type in terms of floret number.

The combinations of components $G_2E_2A_1$ (−3.69, Table 4), and $G_3E_2A_2$ (3.26, Table 4) explain most of the observed G × S × Y interaction. The negative element $G_2E_2A_1$ reveals that the Northern-adapted hybrids 2 and 6 showed a poorer relative performance for oil yield in Dec 98 than in Dec 96, i.e. (+, +, +) for (G_2, E_2, A_1) in Dec 98 for a negative element of the core array (see loadings on components in Table 2). Grain filling duration, kernel proportion and kernel oil proportion (positive loadings on A_1, Table 2) would have been negatively affected in Dec 98 in these hybrids, determining the relative reduction observed for oil yield and its primary components grain weight and grain oil concentration. The environmental factors negatively affecting grain filling of hybrids 2 and 6 in Dec 98 did not affect grain number. However, as in the case of oil concentration for G × S interactions, the strong genotypic effect showed by grain number (de la Vega and Hall, 2002b) precludes its use as phenotypic indicator of indirect selection aimed at avoiding the disruptive effect of testing environments such as that sampled in Dec 98. When the secondary components of grain number are considered, the positive element of the core array $G_3E_2A_2$ (Table 4 and the loadings on components in Table 2) reveal that hybrid 2 improved its relative performance for seed set and hybrid 6 improved its relative performance for floret number in that environment. Although both secondary components of grain number contributed to its determination, floret number shows a strong genotypic effect (de la Vega and Hall, 2002b), again negating its value as a phenotypic selection indicator. By contrast, seed set showed a strong G × E interaction effect and it is a trait easy to assess, since fertilisation failures and embryo abortion are concentrated in the central portion of the floral disk (de la Vega and Hall, 2002b).

3.4. Joint biplots

The joint biplots of genotypes and attributes associated with individual environment components provide an alternative way for visualising patterns of G × E × A interaction. The 1st environment components for yield determination models (1) and (2) explain the G × S interaction and the 2nd environment component for both models explain the G × S × Y interaction (Tables 1 and 2). Thus, the joint biplots of genotypes and attributes for the 1st and the 2nd environment components for both models will be used to investigate the relationships between genotypes and attributes associated with the predictable and unpredictable G × E interactions, respectively. The proportion of variability explained by the 3rd environment components of both yield determination models is too small to justify discussion.

3.4.1. First environment components: Interpreting the attributes contributing to predictable G × E interactions

The component weights for the 1st and 2nd axes of the joint biplots of genotypes and attributes in the 1st environment component were 0.34 and 0.14, respectively, for yield determination model (1) and 0.27 and 0.14, respectively, for model (2). The bi-dimensional joint biplots of Fig. 1A and B display those aspects of the relationships between genotypes and attributes that are influenced by the differences between Oct and Dec plantings across years, after the genotypic and environmental effects have been removed. These effects can be selected in breeding for specific adaptation to late plantings or to the Northern mega-environment.

In these bi-dimensional joint biplots, genotypes are represented by points and attributes by vectors from the origin. Genotypes located near the origin have all their values close to the attribute means, given that the data are attribute-centred. Genotypes that are close together are similar in their specific responses for all attributes analysed. For any particular attribute, genotypes can be compared by projecting a perpendicular from the genotype points to the attribute vector, i.e. entries that are further along in the positive direction of an attribute vector show higher values for this attribute and vice versa. The bi-dimensional joint biplots also display the strength of the associations among attributes. Acute angles between any two attribute vectors indicate positive associations, i.e. they are positively correlated; 90° angles indicate no association; and angles greater than 90° indicate negative associations (Chapman et al., 1997; Kroonenberg, 1997).

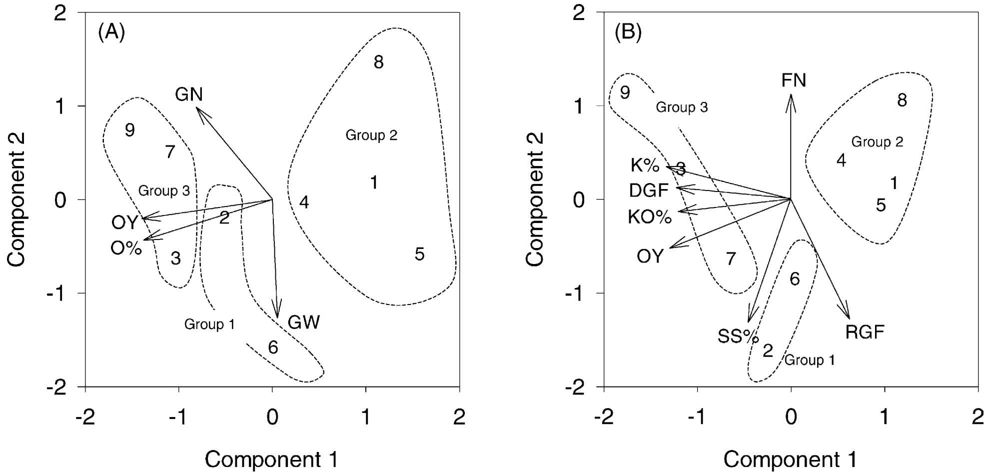

Fig. 1. Joint plots of the 1st two components of sunflower hybrids and attributes associated with the 1st environment component for the primary (A) and secondary (B) oil yield determinant models. Genotypes are represented by numbers (see Table 1 for genotype names) and attributes by vectors (positive direction shown by arrow head) from the origin. Genotype groups identified by mixture maximum likelihood cluster method of oil yield and its primary components are superimposed on plots. Attribute abbreviations are: OY: oil yield, GN: grain number, GW: grain weight, O%: grain oil proportion, FN: floret number, SS%: percentage of seed set, RGF: rate of grain filling, DGF: duration of grain filling, K%: kernel proportion, KO%: kernel oil proportion.

In the joint biplots of Fig. 1A and B, the positive direction (indicated by arrow heads in figures) of the attribute vectors indicates improved relative performance in Dec plantings. This is because the loadings on the 1st environment component are positive for the Dec 96 and Dec 98 (Table 1). Hybrids 2, 3, 6, 7, and 9 (i.e. MixClus3 Groups 1 and 3) showed a relative increase in oil yield in Dec plantings, since their perpendicular projections on this attribute vector intercept it in positive direction (Fig. 1A). The angle formed by the attribute vectors of grain oil concentration and oil yield indicates that these traits are strongly and positively correlated (Fig. 1A). This means that the genotype-specific responses for oil concentration are strongly underlying the predictable G × E interactions for oil yield. Oil yield showed a weak positive association with grain number. The attribute vectors of oil yield and grain weight form almost a right angle (Fig. 2A), suggesting that the variation in grain weight has little effect on the G × S interactions for oil yield. Grain number and grain weight showed a negative association between them in terms of their G × S interaction effects. Hybrids 7 and 9 showed a relative improvement in grain number in Dec plantings, and hybrid 6 showed a relative improvement in grain weight in that environment.

Oil yield showed a positive association with kernel proportion, kernel oil proportion, duration of grain filling, and grain set, a lack of association with rate of grain filling, and a slightly negative association with floret number (Fig. 1B). Both kernel proportion and kernel oil proportion positively determined the genotypic relative responses for oil concentration. A relative increase in floret number in Dec plantings (Fig. 1B) determined the relative increase in grain number (Fig. 1A) shown by hybrids 8 and 9. Contrarily, hybrids 7 and 2 showed a relative increase in grain number in Dec plantings (Fig. 1A) determined by a relative improvement in seed set (Fig. 1B). However, only the relative responses for seed set are positively correlated to those for oil yield (Fig. 1B). Similar observations can be made for grain weight and its secondary determinants rate and duration of grain filling.

Although these joint biplots show that the patterns of relative performance for oil yield become blurred when the focus moves towards its primary and secondary components, they allowed the identification of seed set (determining grain number) and duration of

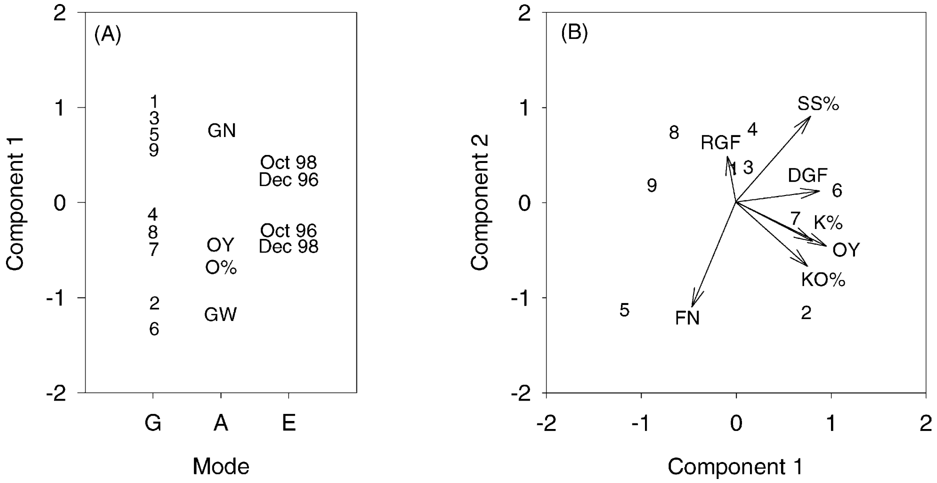

Fig. 2. Joint plots of the 1st (A) and the 1st and the 2nd (B) components of sunflower hybrids and attributes associated with the 2nd environment component for the primary (A) and secondary (B) oil yield determinants models. Genotypes and attributes as in Fig. 2.

grain filling (determining oil concentration) as useful putative indicators of adaptation to Dec plantings and the Northern region. These analyses were also useful for detecting materials of similar response pattern for oil yield but contrasting relative behaviour for some primary and secondary yield determinants (e.g. hybrids 2 and 9, which showed similar responses for oil yield and contrasting responses for rate of grain filling and floret number; Fig. 1B).

3.4.2. Second environment components: interpreting the attributes contributing to unpredictable G × E interactions

The effects associated with the 2nd environment component for models (1) (Fig. 2A) and (2) (Fig. 2B), i.e. G × S × Y interactions, complicate indirect selection in Dec plantings for specific adaptation to the Northern mega-environment. The component weights for the 1st and the 2nd axes of the joint biplot of genotypes and attributes in the 2nd environment component for model (1) were 0.12 and 0.03, respectively. Therefore, major effects of unpredictable interactions on the primary components of oil yield can be described in a single dimension, corresponding to the 1st component of the joint biplot. In a case like this, when one dimension effectively explains most of variability, the joint biplots collapse into a single line, where it is possible to include the component loadings of the environment mode as well. In such a case, a product term to compare scores may be calculated as a product of any combination of the scores of the three modes (Basford et al., 1990; Chapman et al., 1997). Hybrid 6 (score −1.32), e.g. will have a negative product for oil yield (score −0.46) in Dec 98 (score −0.45) (Fig. 2A). For the same environment type and attribute, hybrid 9 (score 0.55) will have a positive product.

The most remarkable observation in the one-dimensional joint biplot of Fig. 2A is the negative association observed between the Northern-adapted hybrids 2 and 6 and Dec 98 for oil yield. This effect could have led breeders to discard these hybrids based on a single test effected in Dec 98. According to the product terms obtained between the scores of the levels of the three modes, the relative response of hybrids 2 and 6 to Dec 98 in terms of oil yield was determined by a reduction in their relative performance for grain weight and grain oil concentration in that environment. In contrast, both hybrids showed a positive score for grain number in Dec 98.

The component weights for the 1st and the 2nd axes of the joint biplot of genotypes and attributes in the 2nd environment component for model (2) were 0.08 and 0.06, respectively. Given the similarity of both axes in terms of explained variability, we will use the bi-dimensional joint biplot (Fig. 2B) to interpret the unpredictable interactions for oil yield in terms of its secondary components. In this joint biplot, the positive direction of the attribute vectors indicates improved relative performance in Dec 96 and Oct 98. The primary determinants of yield, grain weight and oil concentration, are partially determined by the duration of grain filling, which was negatively affected in Dec 98 in hybrids 2 and 6, since their perpendicular projections on this attribute vector intercept it in positive direction (Fig. 2B). Kernel proportion and kernel oil proportion, which jointly determine oil concentration, showed the same type of response. This means that the environmental factor underlying the poorer performance of hybrids 2 and 6 in Dec 98 than in Dec 96, negatively affected the duration of grain filling of these hybrids in Dec 98 more than that of the Northern-adapted hybrids 3 and 9.

Rainfall excess at the end of the crop cycle affected the duration of grain filling of late-maturity hybrids in Dec 98 (de la Vega and Hall, 2002a). In order to assess potential effects of rainfall excess on the usefulness of Dec 98 as a part of an indirect selection procedure for adaptation to the Northern mega-environment, we contrasted the relationship between oil yield and time to anthesis in both situations, using data from trials in the Northern mega-environment for hybrids well adapted to those conditions (Fig. 3). Clearly, Dec 98 did not provide a good image of the relationship typical of the Northern target production area and its use could have promoted discarding valuable late-maturity hybrids (i.e. 2 and 6) for the Northern target production environment. By contrast, Dec 96 showed the same type of relationship between oil yield and time to anthesis of that of the Northern mega-environment (data not shown).

Finally, the negative effect of Dec 98 on the duration of grain filling in some Northern-adapted hybrids is reflected in a reduced, in contrast to Dec 96, phenotypic correlation between Dec planting and the Northern mega-environment in terms of oil yield (Fig. 4A and B). The joint biplot of Fig. 2B reveals that hybrids 2 and 6 improved their relative performance for grain set and floret number, respectively, in Dec 98 (i.e. their perpendicular projections intercept these attribute vectors in negative direction). As discussed above, rapid assessment of these attributes is difficult. To examine the robustness of the diameter of the empty centre as a phenotypic indicator of seed set that might allow the disruptive effect of the $G \times S \times Y$ interactions to be avoided, we assessed the similarity between the genotypic discrimination of seed set in the central portion of the capitulum in Dec plantings and oil yield in the Northern region. For both years evaluated, the diameter of the empty head centre in Dec influenced the performance of the genotypes in a more reliable fashion relative to oil yield in the Northern region than oil yield in the Dec plantings (Fig. 4). The use of this indirect selection strategy instead of oil yield would have avoided discarding the Northern-adapted hybrids 2 and 6 in

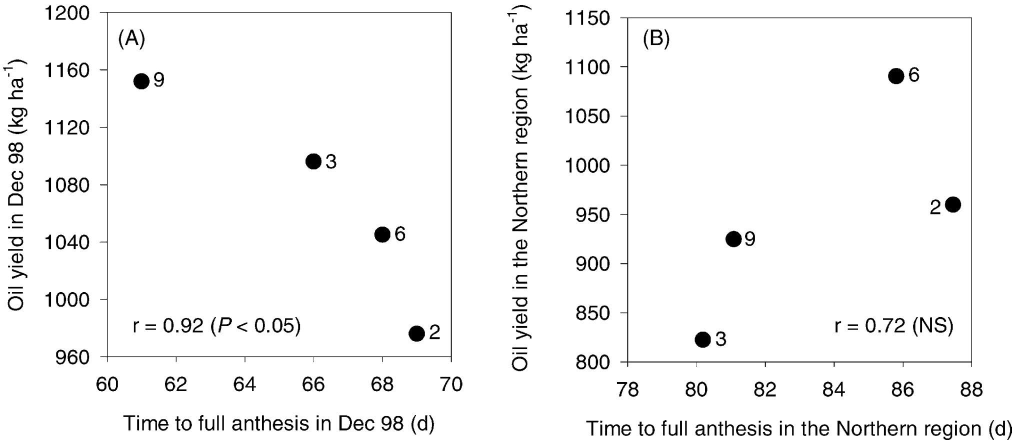

Fig. 3. Association between time to full anthesis and oil yield for the best adapted hybrids to the Northern region in Dec 98 (A) and in the Northern mega-environment (average 14 trials conducted from 1996/1997 to 2000/2001 seasons) (B). See Table 1 for hybrid names.

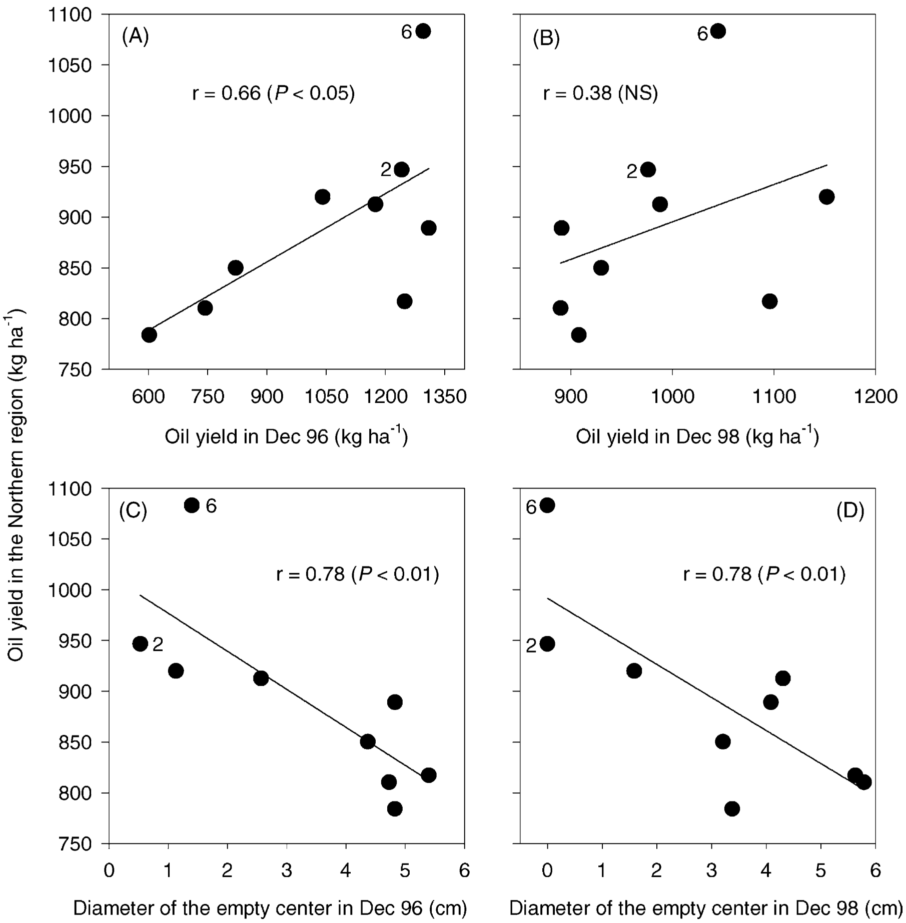

Fig. 4. Association between oil yield in the Northern mega-environment (average of 14 trials conducted from 1996/1997 to 2000/2001 seasons) and oil yield in Dec 96 (A), and Dec 98 (B) plantings in Venado Tuerto, and diameter of the empty centre of the capitulum in Dec 96 (C) and Dec 98 (D). Hybrids 2 and 6 are indicated to compare the effect of different selection strategies on them.

Dec 98 (Fig. 4B and D), increasing the realised response to selection.

4. Conclusions

In three-mode PCA, different types of centring lead to different solutions (Kroonenberg, 1983). Double-centring of the G × E × A arrays of means, as used here, permitted the identification of environmental components that allowed separation of predictable and unpredictable G × E interactions. This would have been difficult to achieve if genotype means had been retained. In addition, the use of three-mode PCA in combination with biological models capturing the physiological bases of yield variation among genotypes, rather than attributes of which the product is not yield, has served to identify three relevant sources of genotypic variation for a breeding program, namely: (i) attributes that can be selected to achieve specific adaptation to the target environment by emphasising predictable interactions (e.g. duration of grain filling, a trait associated with canopy stay green; Fig. 1B); (ii) attributes that allow the unpredictable G × E interactions to be accommodated, improving the linking between managed-environments and target production environments (e.g. grain set; Fig. 4); and (iii) genotypes of similar response pattern for yield but contrasting relative behaviour for the primary and secondary yield determinants (e.g. hybrids 2 and 9; Fig. 1A and B), which combined in breeding projects could bring better opportunities for yield improvements in individual mega-environments. Although the elements of the core array and the joint biplots of genotypes and attributes in environmental components are alternative ways of interpreting the results of a three-mode PCA, hitherto only joint biplots have been extensively discussed in the plant breeding literature. In this paper, we have also demonstrated how inspection of the explained variation and three-mode interactions of the elements of the core array allows the separation of the pattern-rich variability from the noise-rich variability, and further permits dividing the first one in useful sources of variation for a breeding program.

Fischer (1981) distinguished between two general approaches for investigating the physiological basis of genetic variation for quantitative traits, namely top-down and bottom-up. This paper describes part of a top-down study, where the investigation commenced with the identification of repeatable patterns of variation for yield (de la Vega et al., 2001) and attempted to work from these differences towards identification of traits lower in the hierarchy of plant organisation that explain yield variation. Although the reference genotypes used in this study were selected on the basis of their contrasting relative performance and the environments sampled reflect the genotypic discrimination of Central and Northern mega-environments (de la Vega et al., 2001), it is to be expected that other traits can be detected if more genotypes and environments are used. There is also no strong evidence to suggest that it is not possible to recombine some of the components of specific adaptation in a way that will contribute to an improvement in broad adaptation, which is the main target of a breeder. Cooper (1999) recommended evaluating the trait-based selection criteria emerged from this type of analysis by testing experimental populations relevant for a breeding program in environments representative of the target population of environments, i.e. bottom-up approach. This approach is under examination at present.

Acknowledgements

We would like to thank Advanta Semillas (Steven Quast, Alan Scott) for supporting this research. We also thank Carlos Ghanem, Aldo Martínez, Sergio Solián, Daniel Kennedy, Ney Flores and César Sánchez for collaborating in the field experiments. The work of the third author was partially supported by a grant from the Netherlands Organisation for Scientific Research (NWO).

References

Allard, R.W., Bradshaw, A.D., 1964. Implications of genotype–environmental interactions in applied plant breeding. Crop Sci. 4, 503–508.

Basford, K.E., Cooper, M., 1998. Genotype × environment interactions and some considerations of their implications for wheat breeding in Australia. Aust. J. Agric. Res. 49, 153–174.

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

Basford, K.E., Kroonenberg, P.M., DeLacy, I.H., Lawrence, P.K., 1990. Multiattribute evaluation of regional cotton variety trials. Theoret. Appl. Genet. 79, 225–324.

Basford, K.E., Kroonenberg, P.M., Cooper, M., 1996. Three-mode analytical methods for crop improvement programs. In: Cooper, M., Hammer, G.L. (Eds.), Plant Adaptation and Crop Improvement. CAB International/ICRISAT & IRRI, Wallingford, UK, pp. 291–305.

Chapman, S.C., Crossa, J., Basford, K.E., Kroonenberg, P.M., 1997. Genotype by environment effects and selection for drought tolerance in tropical maize. II. Three-mode pattern analysis. Euphytica 95, 11–20.

Cheres, M.T., Miller, J.F., Knapp, S.J., 1999. Donor inbred lines for enhancing the performance of single-cross sunflower hybrids. Crop Sci. 39, 1325–1331.

CIMMYT, 1989. An account of how priorities are set among mega-environments from a breeding perspective. Internal Document Number 17, Mexico, DF.

Cooper, M., 1999. Concepts and strategies for plant adaptation research in rainfed lowland rice. Field Crops Res. 64, 13–34.

Cooper, M., DeLacy, I.H., 1994. Relationships among analytical methods used to study genotypic variation and genotype-by-environment interaction in plant breeding multi-environment experiments. Theoret. Appl. Genet. 88, 561–572.

Cooper, M., Woodruff, D.R., Eisemann, R.L., Brennan, P.S., DeLacy, I.H., 1995. A selection strategy to accommodate genotype-by-environment interaction for grain yield of wheat: managed-environments for selection among genotypes. Theoret. Appl. Genet. 90, 492–502.

Cooper, M., Woodruff, D.R., Phillips, I.G., Basford, K.E., Gilmour, A.R., 2001. Genotype-by-management interactions for grain yield and grain protein concentration of wheat. Field Crops Res. 69, 47–67.

Crossa, J., Basford, K.E., Taba, S., DeLacy, I.H., Silva, E., 1995. Three-mode analysis of maize using morphological and agronomic attributes measured in multilocational trials. Crop Sci. 35, 1483–1491.

de la Vega, A.J., Chapman, S.C., 2001. Genotype by environment interaction and indirect selection for yield in sunflower. II. Three-mode principal component analysis of oil and biomass yield across environments in Argentina. Field Crops Res. 72, 39–50.

de la Vega, A.J., Hall, A.J., 2002a. Effects of planting date, genotype, and their interaction on sunflower yield. I. Determinants of oil-corrected grain yield. Crop Sci. 42, 1191–1201.

de la Vega, A.J., Hall, A.J., 2002b. Effects of planting date, genotype, and their interaction on sunflower yield. II. Components of oil yield. Crop Sci. 42, 1202–1210.

de la Vega, A.J., Chapman, S.C., Hall, A.J., 2001. Genotype by environment interaction and indirect selection for yield in sunflower. I. Two-mode pattern analysis of oil and biomass yield across environments in Argentina. Field Crops Res. 72, 17–38.

Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm (with discussion). J. Roy. Stat. Soc. Ser. B 39, 1–38.

Dudley, J.W., 1984. A method of identifying lines for use in improving parents of a single cross. Crop Sci. 24, 355–357.

Fischer, R.A., 1981. Optimising the use of water and nitrogen through breeding of crops. Plant Soil 58, 249–278.

Fox, P.N., Rosielle, 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-graphical display of matrices with applications to principal component analysis. Biometrika 58, 453–467.

Gabriel, K.R., 1978. Least squares approximation of matrices by additive and multiplicative models. J. Roy. Stat. Soc. Ser. B 40, 186–196.

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

Harshman, R.A., Lundy, M.E., 1984. Data preprocessing and the extended PARAFAC model. In: Law, H.G., Snyder Jr., C.W., Hattie, J.A., McDonald, R.P. (Eds.), Research Methods for Multimode Data Analysis. Praeger, New York, pp. 216–284.

Knowles, P.F., 1978. Morphology and anatomy. In: Carter, J.F. (Ed.), Sunflower Science and Technology. Agron. Monogr. 19. ASA, Madison, WI, pp. 55–87.

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

Kroonenberg, P.M., 1994. The TUCKALS line: a suit of programs for three-way data analysis. Comput. Stat. Data Anal. 18, 73–96.

Kroonenberg, P.M., 1997. Introduction to biplots for G × E tables. Research Report #51. Center for Statistics, The University of Queensland, Brisbane, Qld 4072, Australia.

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

Lawn, R.J., Imrie, B.C., 1991. Crop improvement for tropical and subtropical Australia: designing plants for difficult climates. Field Crop Res. 26, 113–139.

McLachlan, G.J., Basford, K.E., 1988. Mixture Models: Inference and Applications to Clustering. Marcel Dekker, New York, 253 pp.

Rincon, F., Johnson, B., Crossa, J., Taba, S., 1997. Identifying subsets of maize accessions by three-mode principal component analysis. Crop Sci. 37, 1936–1943.

Shorter, R., Lawn, R.J., Hammer, G.L., 1991. Improving genotypic adaptation in crops—a role for breeders, physiologists and modellers. Exp. Agric. 27, 155–175.

Ten Berge, J.M.F., De Leeuw, J., Kroonenberg, P.M., 1987. Some additional results on principal components analysis of three-mode data by means of alternating least squares algorithms. Psychometrika 52, 183–191.

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

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.

Wright, G.C., Nageswara Rao, R.C., Basu, M.S., 1996. A physiological approach to the understanding of genotype by environment interactions—a case study on improvement of drought adaptation in groundnut. In: Cooper, M., Hammer, G.L. (Eds.), Plant Adaptation and Crop Improvement. CAB International/ICRISAT & IRRI, Wallingford, UK, pp. 365–381.