GENOTYPE BY ENVIRONMENT INTERACTION AND INDIRECT SELECTION IN SUNFLOWER FOR ARGENTINA. II. THREE-MODE PRINCIPAL COMPONENT ANALYSIS

Abelardo J. de la Vega, Advanta Semillas S.A.I.C., Ruta Nac. 33 Km 636, CC 294, (2600) Venado Tuerto, Argentina
Fax: +54 3462 435231 ; e-mail: avega@waycom.com.ar

Scott C. Chapman, CSIRO Tropical Agriculture, 120 Meiers Rd., Indooroopilly, Qld 4068, Australia
Fax: +61 7 32142288 ; e-mail: scott.chapman@tag.csiro.au

Summary

The genotype by environment (G×E) interactions observed for oil yield in sunflower in different regions of Argentina can be analyzed in terms of differences among genotypes in individual environments in grain number, grain weight and oil content. Three-mode principal component analysis was applied to a $10 \times 21 \times 4$ (genotypes × environments × attributes) matrix, to collectively interpret the changes of these attributes in a sunflower genotype-environment system, and to assess the relative importance of each trait as underlying determinant of the observed G×E interaction for oil yield.

The $3 \times 2 \times 3$ (genotypes × environments × attributes) principal component model explained about 65% of the variation computed for the multi-environment trial analyzed. The first environment component, that accounted for 54 % of the variation, explained the common pattern over environments and showed that oil content was highly positively correlated to oil yield, while grain number and grain weight showed lack of association with oil yield and were negatively correlated with each other. The second environment component (explaining 11% of the variation) contrasted northern and central environments and showed that grain number is the main underlying determinant of the observed G×E interactions for oil yield.

Introduction

In Argentina, effective identification of superior sunflower genotypes is complicated by the presence of significant genotype × environment (G×E) interactions, such that relative cultivar yields vary across testing environments. In a companion paper (de la Vega *et al.*, 2000), we have used two-way multiplicative models, i.e. pattern analysis (Williams, 1976), to study the response patterns of oil yield and their components for a reference set of sunflower hybrids across a set of growing environments of Argentina. These analysis showed that: (1) the effects of northern (subtropical) and central (temperate) environments on genotype discrimination were orthogonal (poorly correlated), indicating the existence of two mega-environments, and the possibility of selecting for specific adaptation; and (2) some hybrids of contrasting response pattern when considering yield components individually, showed similar patterns of relative performance across environments for oil yield, suggesting the existence of different specific genotypic responses to specific environmental challenges within the same genotype groups. This would complicate the identification of breeding avenues based on traits related to adaptation.

Where more than one attribute is measured on genotypes in multi-environment trials, the collected data can be summarized in the form of genotype by environment by attribute (G×E×A) array of means (Basford *et al.*, 1996). In this case, the generalization of two-way multiplicative models to three-way multiplicative models is a useful strategy for the simultaneous modelling of G×E interaction in more than one trait (van Eeuwijk and Kroonenberg, 1995). In this paper, three-mode (G×E×A) principal component analysis (PCA) (Kroonenberg, 1983) was applied to a $10 \times 21 \times 4$ matrix, to collectively interpret the changes of oil yield and its components in a sunflower GE system, and to assess the relative importance of each trait as an underlying determinant of the G×E interaction for oil yield.

Materials and Methods

The details of the experimental material, test environments, experimental design and measurements are given in de la Vega *et al.* (2000) and in Tables 1, 2 and 3. For three-mode multivariate analysis, the collected data was summarized in the form of a $10 \times 21 \times 4$ (G×E×A) array of means. Three-mode PCA was used for studying the G×E×A matrix. This procedure derives components for each of the three modes (say P, Q, R, of them for G, E, and A, respectively), as well as a three-way matrix of order P by Q by R (the core matrix). In this model each mode is allowed to have a different number of components. The core matrix contains the weights assigned to each of the combinations of the components for the three modes (Kroonenberg and Basford, 1989). 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 and Basford, 1989).

It is possible to portray the relationships between the genotypes and attributes for each component of the environment in a joint plot (Basford *et al.*, 1996). The joint plot diagram was used to depict the component scores of genotypes and attributes associated with the environment modes. In these plots, genotypes are represented by points and attributes by vectors from the origin. 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. Acute angles between any two attribute vectors indicate positive associations, i.e. they are positively correlated; 90º angles indicate no association; and angles greater than 90º indicate negative associations (Kroonemberg, 1997). The three- mode PCA was applied to the

$10 \times 21 \times 4$ (G×E×A) matrix using the program TUCKALS3 (Kroonenberg, 1994). Prior to analysis, the data were standardized within environments.

Results

The three-mode model with $3 \times 2 \times 3$ components for G, E and A, respectively, was considered adequate for fitting the data ($r^2 = 0.65$), on the basis of informal judgments of the increases in r^2 compared to the increases in dimensions and difficulty of interpretation. In this model, the three components for the genotype mode accounted for 35%, 21% and 9% of the variation, respectively; the two components for the environment mode accounted for 54% and 11%, respectively; and the three components for the attributes accounted for 35%, 24% and 6% of the variation, respectively. Not all genotypes, environments and attributes were fitted equally well by the model (Tables 1, 2, 3).

Treating the components of the three modes separately gives only a partial view of the structure of the variability in the data, it being necessary to look at the components for all modes simultaneously for a full view (Basford and Kroonenberg, 1989). The components of the genotypes (Table 1) and those of the attributes (Table 3) do not have obvious interpretations with respect to mean trait values, and so the lower-dimensional representations serve the purpose of data reduction. While the first environment component (54% of the variation) explained the common pattern over environments, the second component (11% of the variation) contrasted central-type and northern type environments (Table 2). Thus, the joint plot of genotypes and attributes for the first environment component will be used to investigate the interactions between genotypes and attributes for all environments together and the joint plot of genotypes and attributes for the second environment component will be used to investigate the relative merit of each oil yield determinant underlying the observed G×E interactions.

First environment component: Interpreting the attributes contributing to mean oil yield across environments
The analysis of the joint plot of the first two axes (Figure 1) showed that oil content (OC) was positively correlated to oil yield (OY), since their vectors form an acute angle, while grain weight (GW) and grain number (GN) showed lack of association to OY as they are at right angles. The highest yielding genotypes across all environments tend to be on the bottom right quadrant of the joint plot. These hybrids were hence associated with better than average attribute scores for OC. Entries 5 and 7, defined as broadly adapted hybrids by de la Vega *et al*. (2000), were the highest yielding genotypes across environments. There is a strong negative association between GW and GN, since their vectors form an almost 180º angle. The hybrids with the highest values for GN are on the top right quadrant of the diagram, while the hybrids with the highest attribute scores for GW are on the bottom left quadrant. The highest yielding hybrids across environments had intermediate values for GN and GW. The standardized weight of the third axis of this joint plot is too small to warrant discussion.

Second environment component: Contrasting central-type and northern-type environments
Major differences between the central and northern-type environments can be described in a single dimension, corresponding to the first component of the joint plot of genotypes and attributes for the second environment component (Figure 2). This figure displays those aspects of the genotype and attribute relationships that are influenced by the differences between northern and central environments on genotypes and attributes, after the effect of average performance (Figure 1) has been removed.

Table 1. Genotype code and name, genotype group memberships derived from 2-mode analysis, mean values of four attributes (oil yield, grain number, grain weight and oil content) and mode component scores (with adequacy of fit) for 10 sunflower hybrids over 21 environments in Argentina. Genotype groups: northern adapted (1), central adapted (2), broad adapted (3) hybrids

Genotype		Group[a] (2-mode)	Oil yield (kg ha^{-1})	Grain number m^{-2}	1000 grain weight (g)	Oil (%)	Component			Proportion of sums of squares explained (%)
Code	Name						1	2	3	
1	Contiflor 15	2	1117	4929	43.9	48.2	0.61	-0.04	0.18	59
2	Aguará	1	1047	5364	40.2	46.2	0.53	-0.63	-0.38	73
3	GV23105	1	1023	4190	53.9	42.9	-0.75	-0.01	0.05	62
4	GV25015	1	905	4777	47.3	37.5	-0.73	-1.02	0.15	79
5	GV25086	3	1199	4616	49.3	49.1	0.26	0.40	-0.12	40
6	TC 2001	2	1089	4551	44.4	49.0	0.54	0.13	0.46	69
7	GV23146	3	1232	5065	46.8	48.5	0.41	0.16	-0.47	46
8	GV22510	3	1125	4563	49.1	47.4	0.04	0.18	-0.22	20
9	Contiflor 9	2	1080	4320	48.2	47.4	0.18	0.21	0.47	51
10	Morgan 734	1	1057	3729	60.6	44.9	-1.09	0.63	-0.12	81
	Proportion of sums of squares explained (%)						35	21	9	65

[a] Groups in 2-mode identified by hierarchical agglomerative clustering of oil yield (de la Vega *et al.,* 2000)

Table 2. Northern (N), central (C), and managed (M) environments codes, means of four attributes (for 10 hybrids) and mode component scores (with adequacy of fit). Managed environments were: V2- Venado Tuerto, December planting, and VI- Venado Tuerto, December planting with artificially extended photoperiod to 15.5 h during the whole crop cycle

Environment			Group[b] (2-mode)	Oil yield (kg ha^{-1})	Grain number m^{-2}	1000 grain weight (g)	Oil (%)	Component		Proportion of sums of squares explained (%)
Code	Location	Region[a]						1	2	
9J7	9 de Julio	C	5	1125	5207	44.3	46.9	0.81	0.26	71
9J8	9 de Julio	C	6	764	3493	44.4	49.4	0.81	0.06	67
9J9	9 de Julio	C	6	1628	5501	61.4	50.6	0.80	0.31	74
DX8	Daireaux	C	4	1037	4260	51.8	47.0	0.76	0.39	73
DX9	Daireaux	C	6	1782	6123	56.7	50.6	0.85	0.26	78
SA9	Sampacho	C	5	1449	4600	67.6	48.2	0.77	0.34	70
VT7	V. Tuerto	C	4	1608	6547	50.5	48.1	0.78	0.28	69
VT8	V. Tuerto	C	5	1007	4558	47.4	46.8	0.69	0.24	54
VT9	V. Tuerto	C	5	1992	6259	64.9	48.4	0.77	0.35	71
V27	V. Tuerto	M	1	1058	5696	43.5	41.1	0.57	-0.63	73
V29	V. Tuerto	M	2	950	5321	45.1	40.8	0.81	-0.23	71
VI9	V. Tuerto	M	4	709	4505	36.8	39.4	0.60	0.23	42
LB7	Las Breñas	N	1	441	1868	54.4	41.3	0.34	-0.66	55
MA8	Margarita	N	3	346	2748	30.8	43.8	0.62	-0.14	40
MA9	Margarita	N	6	1602	5478	56.4	51.2	0.83	-0.01	68
OR7	Orán	N	2	1290	6243	45.9	43.0	0.80	-0.51	89
OR9	Orán	N	2	534	3347	39.7	39.2	0.73	-0.35	66
PA9	Paraná	N	1	1206	4989	49.9	47.9	0.70	-0.35	61
RE7	Reconquista	N	2	573	3164	36.6	47.1	0.67	-0.33	56
RE8	Reconquista	N	3	337	2255	33.1	45.9	0.63	-0.19	43
RE9	Reconquista	N	6	1366	5002	53.7	50.4	0.88	0.05	78
	Proportion of sums of squares explained (%)							54	11	65

[a] C: central region, N: northern region, M: managed environments

[b] Environment groups in 2-mode identified by clustering of oil yield (de la Vega *et al.,* 2000)

Central and northern-type environments had positive and negative scores, respectively, for the component 1 of this joint plot. Mega-environment effects were very strong for OY, this attribute being the best indicator of the contrasting effects of the two types of environments. Regarding the OY determinants, mega-environment effects were stronger for GN than for GW and OC, indicating that this attribute accounted for most of the observed G×E interactions for OY.

Table 3. Scores and fit for first three components of four attributes for 10 hybrids and 21 environments

Attribute	Component			Proportion of sums of squares explained (%)
	1	2	3	
Oil yield (kg ha^{-1})	0.52	-0.42	0.23	50
Grain number m^{-2}	0.51	0.37	0.32	51
1000 grain weight (g)	-0.51	-0.68	0.18	76
Oil (%)	0.79	-0.41	-0.24	84
Proportion of sums of squares explained (%)	35	24	6	65

A product term to compare scores may be calculated as a product of any combination of the scores of the three modes given in Figure 2, e.g. entry 6 (score 2.1) will have a positive product for oil yield (score 1.8) in a central environment as 9J9 (score 0.3). For the same environment-type and attribute, entry 10 (score −1.5) would have a negative product. Entry 10 in a northern-type environment as LB7 (score −0.6) will have a positive product for OY and for the other attributes, considering that all of them had positive scores. As OY is the attribute that showed the highest score, the magnitude of its G×E×A products will be larger than the product terms of the OY determinants. The distinction between the environment types seems primarily due to the northern and central specifically adapted genotype groups, rather than the broadly adapted hybrids, being located relatively close to the origin (Figure 2).

Discussion

Yield component traits have been the subject of numerous physiological and genetical analyses, but the results of these studies are often equivocal and of limited value as guides to plant breeders. This may be in part because of compensation, that can be genetically determined and also be a consequence of environmental conditions (Austin, 1993). The three-mode PCA used here has helped to interpret the genotype main effects and the G×E interaction for yield in sunflower for Argentina in terms
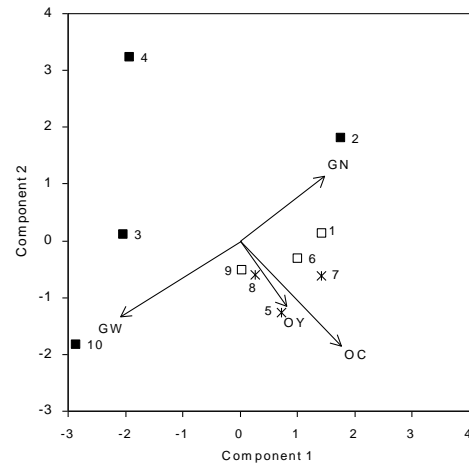

Figure 1. Joint plot of the first two components of 10 sunflower hybrids and four attributes associated with the first environment component. Genotypes are represented by numbers (see Table 1 for genotype names). Genotype symbols represent groups identified by hierarchical agglomerative clustering of oil yield (de la Vega et al., 2000): northern adapted hybrids (2, 3, 4, 10), central adapted hybrids (1, 6, 9), broadly adapted hybrids (5, 7, 8). Vectors for attributes: oil yield (OY), grain number (GN), grain weight (GW), and oil content (OC)

of its components. Different yield determinants accounted for most of the variation in genotype average performance (i.e. yield potential and broad adaptation) and specific response patterns of genotypes across environments (i.e. specific adaptation). OC variation was strongly associated to variation in mean OY across environments. GN was the main underlying determinant of the contrasting genotypic performance between central and northern environment-types. This could be in part because GN in sunflower is determined between floral initiation and
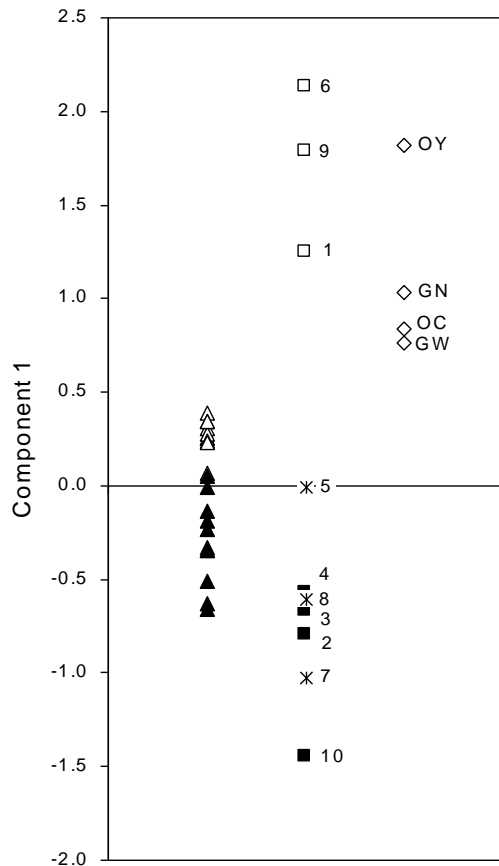
Figure 2. Joint plot of the first two components of 10 sunflower hybrids and four attributes associated with the first environment component. Genotypes are represented by numbers (see Table 1 for genotype names). Environment symbols: N-type (black triangles), and C-type environments (open triangles). Genotype symbols represent groups identified by hierarchical agglomerative clustering of oil yield (de la Vega *et al*., 2000): northern adapted hybrids (2, 3, 4, 10), central adapted hybrids (1, 6, 9), broadly adapted hybrids (5, 7, 8). Vectors for attributes: oil yield (OY), grain number (GN), grain weight (GW), and oil content (OC)

the start of rapid kernel growth (Connor and Hall, 1997), a longer period compared with the windows of determination of GW and OC. Future sunflower breeding avenues for yield potential should target the trade-off between GN and GW. Specific adaptation to both environment-types (perhaps driven by effects of photoperiod) could be addressed by selecting for larger relative GN in particular environments.

References

Austin, R.B., 1993. Augmenting yield-based selection. In: Plant Breeding: Principles and prospects. (M.D. Hayward, N.O. Bosemark and I. Romagosa). Chapman and Hall, London. pp. 391-405.

Basford, K.E., Kroonenberg, P.M., and Cooper, M., 1996. Three-mode analytical methods for crop improvement programs. In: M. Cooper, and G.L. Hammer (Editors), Plant Adaptation and Crop Improvement. CAB INTERNATIONAL 1996, pp. 291-305.

Connor, D.J., and Hall, A.J., 1997. Sunflower physiology. In: A.A. Schneiter (Editor), Sunflower Technology and Production. American Soc. of Agronomy, Crop Science Soc. of America, Soil Science Soc. of America. Madison, WI, USA. pp. 113-182.

de la Vega, A.J., Chapman, S.C., and Hall, A.J., 2000. Genotype by environment interaction and indirect selection in sunflower. I. Multi-attribute two-mode pattern analysis. This volume.

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. Computational Statistics and Data Analysis 18: 73-96.

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

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

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

Williams, W.T., 1976. Pattern analysis in agricultural science. Elsevier Scientific Publishing Company, Amsterdam.