SPATIAL AND SEASONAL EFFECTS CONFOUNDING INTERPRETATION OF SUNFLOWER YIELDS IN ARGENTINA

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

In Argentina, sunflower is grown on almost four million hectares, including subtropical (northern) and temperate (central and southern) environments. The environments in which hybrids are evaluated can substantially complicate selection due to large genotype × environment (G×E) interactions. For this analysis, we have separated, *a priori*, N and C environments. The spatial and seasonal differences between both regions are investigated through two complementary approaches: (i) a trial dataset, comprising 68 trials, covering 8 years (Y) and 18 locations (L), was analyzed to quantify genotypic, environmental, and G×E variance components for grain yield, oil content and oil yield, and (ii) ordination analysis was applied to a long-term environmental data record, in order to group environments based on the mean value of each variable for the different crop stages.

Analysis of locations over years within regions detected that 83% (C region) and 86% (N region) of the total G×E interaction for oil yield was attributable to G×L×Y, respectively. To achieve repeatabilities of more than 80%, trials would need to be conducted over at least 5 years and 15 locations per year.

Ordination analysis of environmental data separates N and C environments over both years and locations. On average, N environments show shorter photoperiods, less rainfall and lower temperatures during vegetative stages, shorter photoperiods during flowering and higher maximum and minimum temperatures during grain filling. Environmental differences between regions showed a strong degree of repeatability, suggesting that crop mean yields differences and G×E interactions associated to the evaluated attributes would be highly repeatable.

Introduction

The formulation of operational decisions in plant breeding requires knowledge of those environmental factors limiting phenotypic expression and of the nature and magnitude of such genotype × environment (G×E) interactions as do occur (Boyd *et al*., 1976). In Argentina, sunflower is grown on almost four million hectares, including subtropical (northern) and temperate (central and southern) environments. Within this target population of environments (TPE), genotype × environment (G×E) interactions complicate effective identification of superior genotypes and make it difficult to ensure that a multi-environment trial to test new and current hybrids across a small number of locations and years will adequately sample the existing production TPE. However, a portion of these G×E interactions may be repeatable and able to be either controlled or sampled in a stratified manner. Knowledge of the production factors that are responsible for repeatable G×E interactions can lead to the definition of target environments for breeding and selection (Chapman *et al*., 2000).

For the analyses described in this paper, we have separated, *a priori*, subtropical (N) and temperate (C) sunflower growing environments. The spatial and seasonal differences between both regions were investigated through two complementary approaches: (i) a trial dataset was analyzed to quantify genotypic (G), environmental (E), and G×E effects for grain yield, oil content and oil yield, and (ii) principal component analysis (PCA) was applied to a long-term environmental data record, in order to group environments based on the value of each environmental variable for the different crop stages. Our objective was to determine how the relative sizes of genotypic and G×E variance components differ across seasons, locations and regions, to detect differences between regions over years for some environmental variables, and to determine, based on both approaches, to what extent G×E interactions between C and N regions are expected to be repeatable.

Materials and Methods

Trial Dataset, data processing and variance component analysis

A trial dataset from the Advanta Argentina sunflower breeding program was analyzed to quantify genotypic (G), environmental (E), and G×E effects for grain yield (kg ha^{-1}), oil content (%) and oil yield (kg ha^{-1}). The dataset comprises 68 advanced trials, covering 8 years (Y) and 18 locations (L) of northern (N), central (C), and southern (S) regions of Argentina, with the entries being current industry hybrids or experimental hybrids that were being considered for release. The 8 years were seasons 1991/92 to 1998/99. All trials were RCBD or lattice designs, with 3-4 reps and 25-49 entries. Experimental units consisted of plots of 3-4 rows × 6 m long and inter-row spacing of 0.70 m. All yield data is presented at 11% grain moisture. Within years, the same hybrids were grown in all locations within the N or within the C and the S region, and some hybrids were common across all trials. The proportion of hybrids common to all regions in each year varied between 0.12 and 0.43.

The data for grain yield, oil content and oil yield were analyzed separately. The total number of plots used was 9150. The hybrid means were processed by a residual maximum likelihood method (Patterson and Thompson, 1975) using the ASREML software (Gilmour *et al*., 1995) to estimate the genotypic components of variance, assuming genotypes (*g*) in any season to be a random sample of the current hybrids. As different 'sets' of hybrids were tested in the two regions, we had to analyze these separately. Locations (*l*) and years (*y*) were assumed to be fixed. Within each region, an estimate of phenotypic variance (σ^2_p) was calculated using the variance components from the ASREML analyses:

$$\sigma^2_p = \sigma^2_g + \sigma^2_{gl}/l + \sigma^2_{gy}/y + \sigma^2_{gly}/(l.y) + \sigma^2_{\varepsilon}/(l.y.r)$$

where l, y and r were any given number of locations, years and replicates, respectively. Several sets of values were used for l, y and r to compare different testing strategies. For each of these combinations, heritability (h^2) was estimated as the ratio:

$$h^2 = \sigma^2_g/\sigma^2_p$$

Environmental Dataset, data processing and principal component analysis

The long-term environmental data used in this study was obtained from I.N.T.A. (Instituto Nacional de Tecnología Agropecuaria, Argentina) and from Advanta Argentina. This is comprised of 8 years (90-97) of monthly average rainfall data in 10 locations, 6 years (90-91, 93-95, 97) of monthly average daily maximum and minimum temperature data in 9 locations and monthly average daily photoperiod data in 9 locations. Environmental data was computed for each year and location based on crop phases: pre-planting (P), sowing (S), vegetative period (V), flowering (F) and grain filling (G). Crop phases were monthly based and defined as follows: for the N environments: P (May, Jun, Jul), S (Aug), V (Sep, Oct), F (Nov), G (Dec); for the C environments: P (Jul, Aug, Sep), S (Oct), V (Nov, Dec), F (Jan), G (Feb). Rainfall was the only variable analyzed in crop phase P.

Principal component analysis (PCA) was applied to the trial environment × crop stage matrices of monthly mean data of daily photoperiod, rainfall and daily maximum and minimum temperatures, in order to investigate the interrelations between environments and environmental variables for each crop stage. The principal components (PCs) of the squared Euclidean distance matrix of each environmental variable were estimated using a singular value decomposition procedure (Gabriel, 1971). A biplot of the first two PCs for each weather variable was constructed from this analysis (Gabriel, 1971).

Results

Trial dataset

Analysis of locations over years within regions showed that 83% (C region) and 86% (N region) of the total G×E interaction for oil yield was attributable to G×L×Y (Table 1), respectively (Table 1). This variance component (σ^2_{gly}) was 5.7 and 5.4 times larger than σ^2_g

Table 1. Variance components and standard errors derived from the genotype (g), locations (l) and years (y) model applied to the sunflower hybrid trial dataset from Advanta Semillas

Source	Grain yield			Oil content			Oil yield		
	Variance component $(kg\ ha^{-1})^2$	% g	Standard error	Variance component $(\%)^2$	% g	Standard error	Variance component $(kg\ ha^{-1})^2$	% g	Standard error
Northern region									
g	11177	100	4544	5.318	100	0.749	2286	100	1039
g.l	4025	36	5919	0.587	11	0.226	495	22	1546
g.y	7726	69	4890	0.762	14	0.225	1555	68	1275
g.l.y	39574	354	7439	1.096	21	0.222	12411	543	2065
Rep.	4689	42	1421	0.172	3	0.044	1304	57	380
Residual	97141	869	3143	1.897	36	0.061	24282	1062	786
Central region									
g	24986	100	5432	4.341	100	0.547	3590	100	1129
g.l	1766	7	3840	0.028	1	0.046	660	18	1048
g.y	8757	35	3517	0.326	8	0.069	3503	98	1049
g.l.y	72901	292	6147	0.942	22	0.071	20386	568	1649
Rep.	17802	71	3208	0.223	5	0.038	4583	128	821
Residual	176682	707	4058	1.640	38	0.038	44331	1235	1018

within C and N regions, respectively. Several sets of values were used for l, y and r to compare different testing strategies, utilizing the variance components of ASREML analysis to estimate h^2. For a testing program of 5 locations over 3 years, for example, we estimated h^2 of 0.57 and 0.61 for oil yield within the C and N regions, respectively. To achieve repeatabilities of more than 80%, trials would need to be conducted over at least 5 years and 15 locations per year. Oil content was the attribute that showed the largest h^2. To achieve repeatabilities of more than 80% for this trait, a testing regime of three locations over one year would be sufficient.

Environmental attributes

The results of the ordination analysis are presented in biplots of the 1st and 2nd PCs (Figure 1). Environments that are close together show similar values for each environmental variable in each crop stage. For any particular crop stage, environments can be compared by projecting a perpendicular from the environment markers to the crop stage vector, i.e. environments that are further along in the positive direction of the vector showed higher values for the analyzed environmental variable and vice versa (Kroonenberg, 1997). Acute angles between any crop stage vectors indicate positive associations, i.e. they discriminate among environments for a particular environmental variable in a similar manner; 90º angles indicate no association; and angles greater than 90º indicate negative associations (Kroonenberg, 1997).

The 1st and the 2nd PCs for photoperiod accounted for the 100% of the total variation of the crop stage-environment system analyzed (Figure 1A). Crop stages S and V showed higher values for the 1st PC and lower values for the 2nd PC than F and G. The 1st PC (97% of the variation) largely reflects the photoperiod values for the months of S and V, with the environments that showed the highest values for this attribute at the right half of the diagram (C environments) and the environments that showed the lowest values for this attribute located at the left side of the biplot (N environments). The variation explained by the 2nd PC is too small to warrant discussion.

The first two PCs for rainfall accounted for 73% of the total variation (Figure 1B). The vector of crop stage G showed a high score for the 1st PC (47% of the variation), which separates the environments that showed the highest rainfall during G at the right half of the diagram and vice versa. The vector of crop stage S showed a strong negative association with the vector of G (i.e. they form an almost 180º angle between them). The 1st PC clearly separates N environments, with more rainfall for G and less rainfall for stage S, from C environments. Crop stages V and F showed a strong positive association between them and lack of association with S and G (orthogonal vectors). The vectors of crop stages V and F showed high scores for the 2nd PC (26% of the variation), which separates the environments that showed the highest rainfall during V and F, at the bottom half of the diagram, from the environments that showed the lowest values of rainfall during these crop stages, located at the top of this biplot. This PC does not appear to discriminate between C and N environments.

The 1st and the 2nd PCs for maximum temperatures accounted for 77% of the variation (Figure 1C). The vectors of crop stages V and G strongly discriminated between C and N environments. The vector of V showed high scores for the 1st PC (49% of the variation) and the 2nd PC (28% of the variation). Environments that showed the highest maximum temperatures during V associated positively with this vector and tend to be at the bottom left quadrant of this diagram. The vector of crop stage G showed a high score for the 2nd PC, which separates the environments that showed the highest values for maximum temperature during grain filling at the top half of this biplot and vice versa. The angle between the vectors of V and G is larger than 90º, indicating the existence of a negative association between these crop stages in the way they discriminate among environments for maximum temperatures.

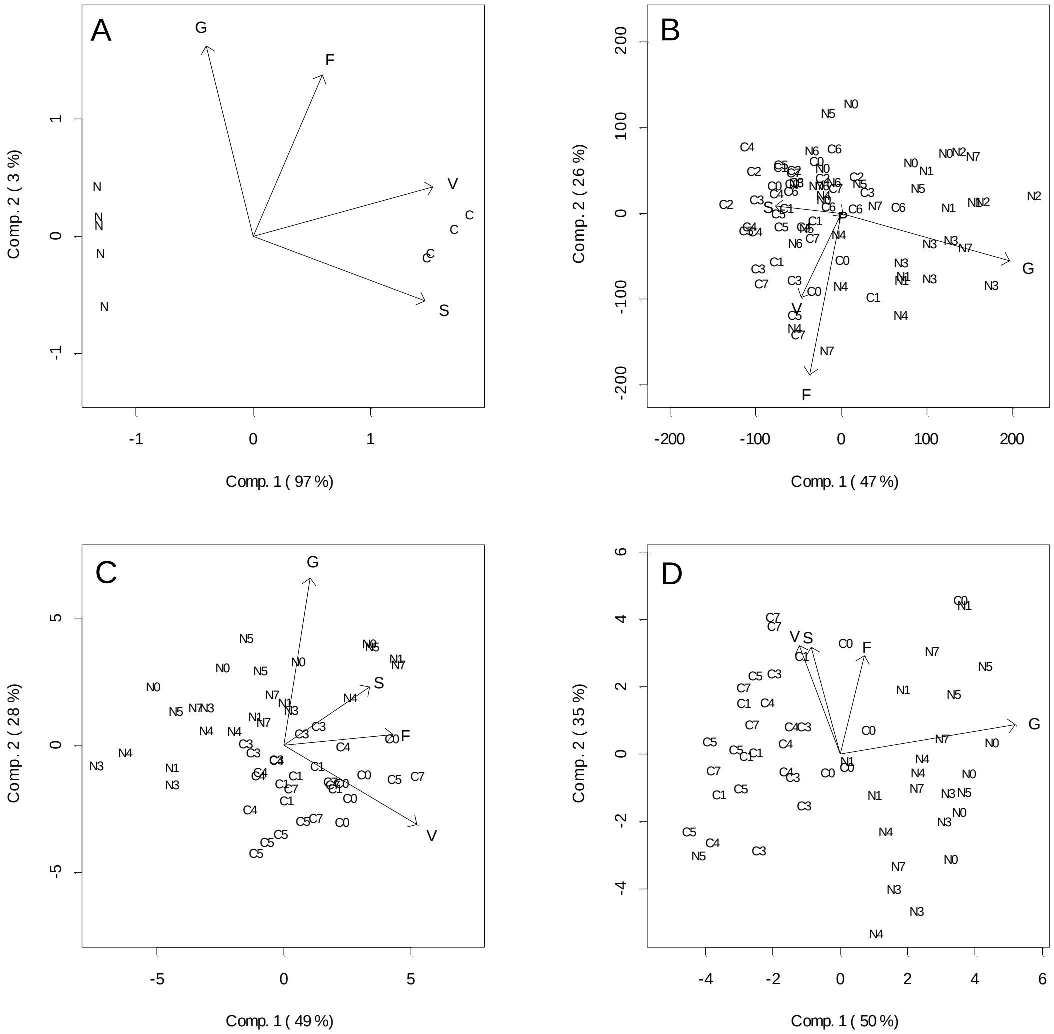

Figure 1. Biplots of the 1st and 2nd principal components for photoperiod (A), rainfall (B), maximum temperature (C), and minimum temperature (D) of several years (numerals) of testing environments from the central (C) and northern (N) sunflower growing regions of Argentina and four crop stages [P: pre-planting (3 months), S: sowing (1 month), V: vegetative stage (2 months), F: flowering (1 month), G: grain filling (1 month)]. Environments are only identified by region and year

For minimum temperature (Figure 1D), only the vector of crop stage G, which showed a very high score for the 1st PC (50% of the variation) and the best fit to the first two PCs (85% of the variation), clearly discriminated between regions. N environments tended to be positively associated with this vector and with the 1st PC and showed the highest values of minimum temperatures during grain filling. The crop stage vectors of S, V and F showed positive association among them, lack of association with G and high scores for the 2nd PC (35% of the variation).

Discussion

The variable and unpredictable nature of much of the G×E interaction within N and C regions implies that broad adaptation within each region is required. Considering the low repeatabilities estimated for oil yield, a large sample of environments is needed to identify superior hybrids and testing across locations and years are equally important. To sample the TPE with as few trials as possible, the objective should be to ensure that the frequencies of occurrence of different 'types' of environments match those being experienced in the TPE (Chapman *et al*., 2000). One alternative approach would be to classify testing environments as members of different 'environment types' and weight their results to account for the mismatch with the TPE (Fox and Rosielle, 1982; Basford and Cooper, 1998). This classification could be direct, utilizing environmental indices (e.g. Woodruff, 1981), o indirect, utilizing the relative responses of a reference set of genotypes (e.g. Fox and Rosielle, 1982).

In general, N environments had shorter photoperiods until grain filling, less rainfall and lower temperatures during the vegetative stages and higher temperatures and more rainfall during grain filling than the C environments. Rainfall is the environmental attribute that showed the highest seasonal and spatial variability. Lack of radiation data precluded an exploration of the effects of this variable, which should be included in future analyses. For all analyzed variables, N environments showed a larger spatial and seasonal variability, whereby the sets of environment points of this region showed a larger dispersion in the biplots than the C environments. This suggest that mean yield variability and G×E interactions associated to the variation for these variables would be relatively larger within the N region. Environmental differences between C and N regions showed a strong degree of repeatability, whereby there is almost no overlapping between the areas of environment distribution of C and N regions in the biplots of the four attributes analyzed. This suggests that the environmental differences between C and N regions associated with the effect of these four variables on crop mean yields and G×E interactions would be highly repeatable. Assessing a reference set of genotypes in C and N locations is a reliable strategy to achieve a better environmental grouping definition and to define the best alternatives to deal with the current G×E interactions. This is the aim of a companion paper (de la Vega *et al*., 2000).

Acknowledgments

We would like to thank Advanta Semillas (Ariel Lorenzo, Ricardo Siciliano, Alan Scott) for making this research possible. We also thank Aldo Martínez, Sergio Solián, Daniel Kennedy, Carlos Ghanem, Ney Flores and César Sánchez for collaborating in the field experiments, Joanne Walker, for her assistance in the multivariate analyses, and Dr. Antonio Hall for his valuable comments.

References

Basford, K.E., and Cooper, M., 1998. Genotype x environment interactions and some considerations of their implications for wheat breeding in Australia. Aust. J. Agric. Res. 49: 53-174.

Boyd, W.J.R., Goodchild, N.A., Waterhouse, W.K., and Singh, B.B., 1976. An analysis of climatic environments for plant-breeding purposes. Aust. J. Agric. Res. 27: 19-33.

Chapman, S.C., Cooper, M., Butler, D., and Henzell, R., 2000. Genotype by environment interactions affecting grain sorghum. I. Characteristics that confound interpretation of hybrid yield. Aust. J. Agric. Res. 50: (in press).

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.

Fox, P.N., and Rosielle, A.A., 1982. Reference sets of genotypes and selection for yield in unpredictable environments. Crop Sci. 22: 1171-1175.

Gabriel, K.R., 1971. The biplot-graphical display of matrices with applications to principal component analysis. Biometrika 58: 453-467.

Gilmour, A.R., Thompson, R., and Cullis, B.R., 1995. Average information REML, an efficient algorithm for variance parameter estimation in linear fixed models. Biometrics 51: 1440-1450.

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.

Patterson, H.D., and Thompson, R., 1975. Maximum likelihood estimation of components of variance. In Proccedings of the 8th International Biometrics Conference. pp. 197-207.

Woodruff, D.R., 1981. The prediction of plant growth and development using various environmental indices. I. Review. In D.E. Byth and D.E. Mungomery (Editors), Interpretation of Plant Response and Adaptation to Agricultural Environments. Queensland Branch, Australian Institute of Agricultural Science, pp. 145-157.