THE EVALUATION OF VARIETAL ADAPTABILITY IN SUNFLOWER: SOME GRAPHIC METHODS

THOMAS P. BOGYO [1]
and ALESSANDRO BOZZINI [2]

INTRODUCTION

It is a common practice in plant-breeding and agronomic studies to grow a series of different genotypes (cultivars) in a range of different environments. If the genotypes so tested respond similarly in all the environments, then their relative performance in other environments may be predicted with a certain confidence. Most often, however, their relative performances differ and in this case one would like to establish a pattern of the variation to help predicting the performances of the genotypes in the differing environments.

The traditional analysis of this variation is performed by computing an "Analysis of Variance" of the two-way table where genotypes and environments are rows and columns, taking out the main effects for these and assessing the residual variation by comparing it to the replicates by treatments interaction. In most cases, particularly since the advent of computers through the use of which massive amounts of data can be analysed, this interaction, commonly referred to as *genotype-environment interaction,* is significant, meaning that genotypes do not respond similarly to the environments in which they are tested. A long list of bibliography on the analysis of genotype-environment interaction is available. In this paper we would like to show some possibilities using various graphic methods which will hopefully help to understand the *pattern* of variation.

A number of authors (e.g. Yates and Cochran, 1938; Finaly and Wilkinson, 1963; Eberhart and Russell, 1966) have shown that in many cases the performance of an individual genotype can be expressed as a linear function of an „environmental index", itself a function of the varieties tested in each of the environments. The regression of the yield (or other varietal character) on this environmental index produces a slope, a dimensionless quantity, which can be used as a measure of the sensitivity of the genotype to this compound measure of diverse environmental factors.

Although the method was originally described as early as in 1938 by Yates and Cochran, the graphic method is generally known as the "joint regression analysis" of Finaly and Wilkinson (1963) who, it should be noted emphatically, did not use the method for testing any hypotheses, a pitfall many of their followers fell into.

Rather than that, they used a two-dimensional plot (scatter diagram) with mean yield and regression coefficient as coordinates for each variety (Fig. 1). These diagrams were then broadly interpreted for "adaptability" and "stability" (Fig. 2). Combinations were generalized as follows : those with low yields and high regression (above 1) were regarded as "specifically adapted to unfavourable environments" (having above average stability), those with high yield and average (1.0) regression were deemed "well adapted to all environments". In general they found in an analysis of barley varieties that the English and, to a lesser degree, the Scandinavian varieties were specifically adapted to high-yielding environments. This was to be expected because they were bred for the relatively consistent, high moisture, and high fertility conditions prevailing in England, Denmark, and Sweden. In Southern Australia, however, most of these varieties rarely have the opportunity to show their full potential, and because of their extreme sensitivity to environmental change, in particular to dry seasons or those which cut off early, they often produce very low yields. This is reflected in their low mean yields over all sites and seasons.

Varieties from North Africa, Bolivia, Peru and the South-West of the United States also had average stability, here associated with above-average mean yields, which indicated good general adaptability. Most varieties from these geographic regions showed outstanding performance, producing above-average yields in all sites and seasons ; they far outyielded the Australian varieties.

[1] Professor of Genetics and Agriculture Statistician, Washington State University, Ag. II. 123, Pullman, WA 99164, U.S.A., F.A.O. Consultant.

[2] Chief, Crop and Grassland Production Service, F.A.O., Rome, Italy.

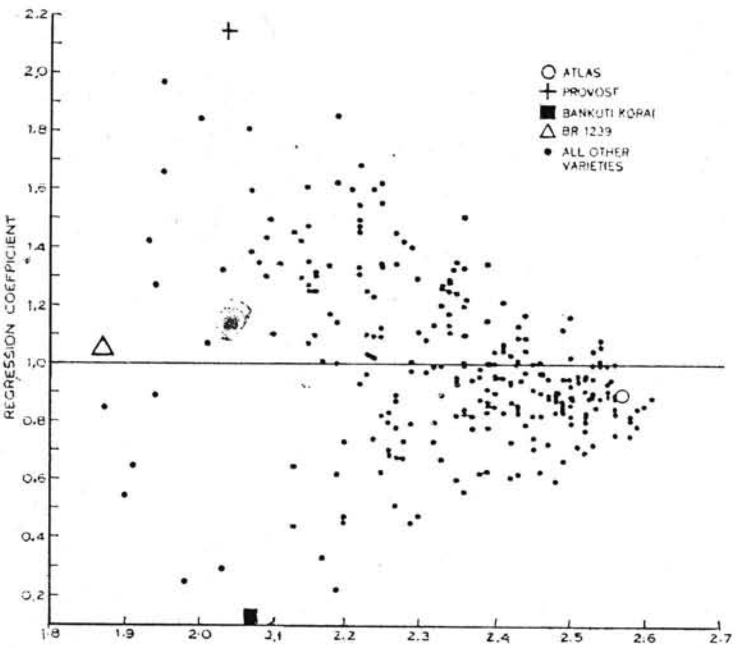

Fig. 1 — The relationship between variety adaptation (regression coefficient) and variety mean yield for 277 barley varieties (From Finaly and Wilkinson, 1963).

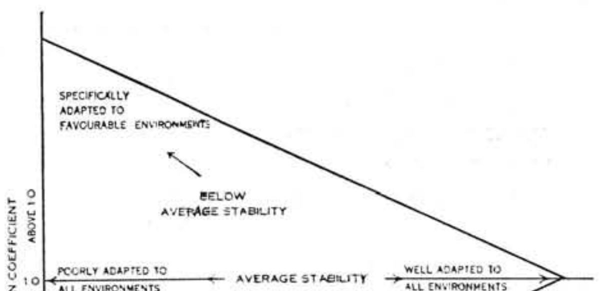

Fig. 2 — A generalized interpretation of the variety population pattern obtained when variety regression coefficients are plotted against variety mean yield (From Finaly and Wilkinson, 1963).

The group of varieties from the South-West of the United States were all introduced or derived from varieties such as California Mariout which originated in Egypt, where it is grown by Arab farmers along the Mediterranean without irrigation and with a rainfall of about 8 in. These varieties are all specifically adapted to low-yielding environments (Fig. 3).

ANALYSIS OF SUNFLOWER YIELD TRIALS : AN EXAMPLE

1. JOINT REGRESSION ANALYSIS

Using the joint regression analysis for the 12 open pollinated sunflower varieties tested in 1979 at 13 locations (see Helia No. 2, page 8) (Fig. 4), the variety Novisad 61 seems well adapted to low yielding environments. With an average yield of 28.4 q/ha and a regression slope of 0.80 it demonstrated that it can maintain a high average yield even under relatively unfavourable conditions. (It ranked 4 out of 12). The same applies to the hybrid H-241, an American single cross, which was tested in the 1978 adaptation trials at 19 locations and ranked 2 out of 20 with a yield of 27.2 q/ha and a regression of yield on environmental index of 0.88 (Fig. 5). It ranked 4, 5 and 1 at Vienna (Austria), Cordoba (Spain) and Ankara (Turkey) respectively, which were the lowest yielding environments in this particular study.

A certain caution is necessary when using this method of evaluation. It is possible to fit a regression line to any set of points. The regression coefficient alone, therefore, is not necessarily a meaningful measure of adaptability. It should also be of interest to note the closeness of the points to the fitted line, the measure of which is the *coefficient of determination* (r^2). In the two sunflower trials the r^2 values vary from 0.78 to 0.95 in the hybrids (Table 1) and 0.67 to 0.96 in the open pollinated varieties (Table 2).

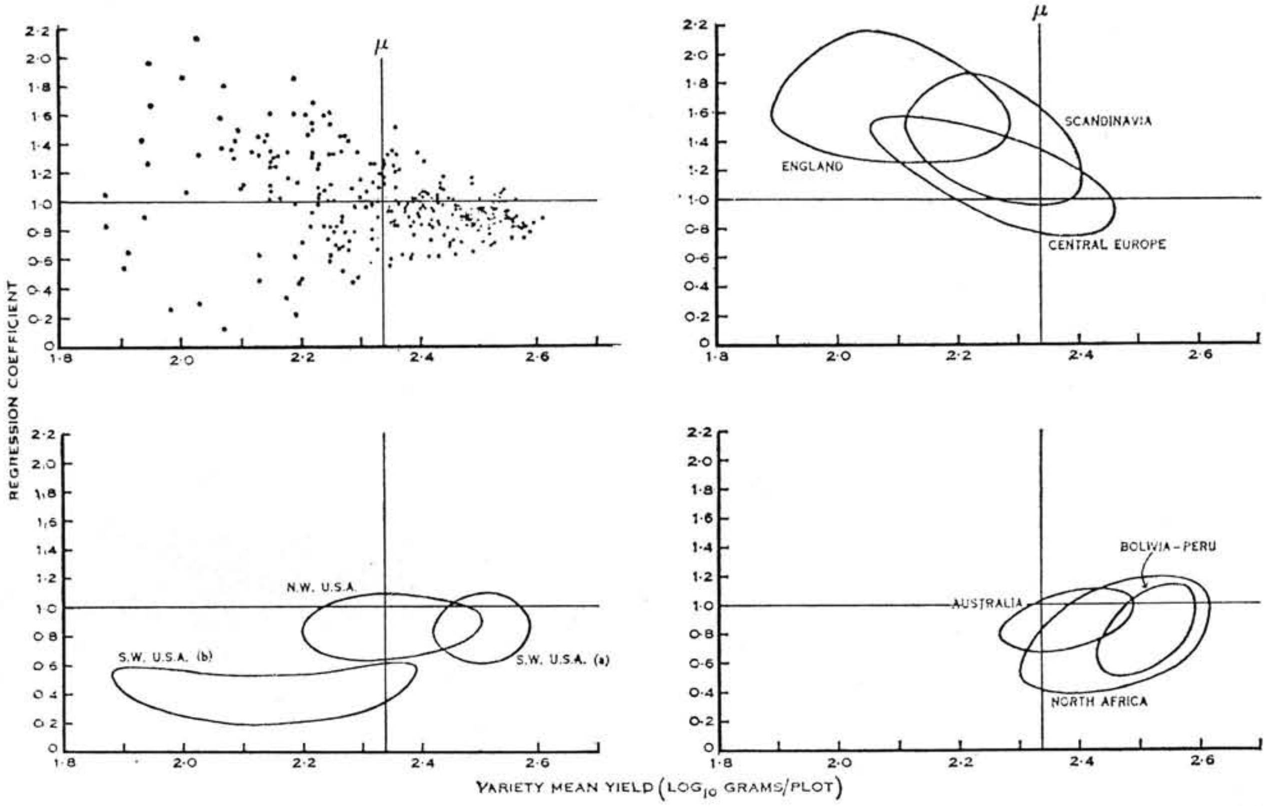

Fig. 3 — The relationship of the geographic area of origin and the type of variety adaptation indicated by position relative to the marker lines regresion coefficient 1.0. (general adaptability) and population mean yield (From Finaly and Wilkinson, 1963).

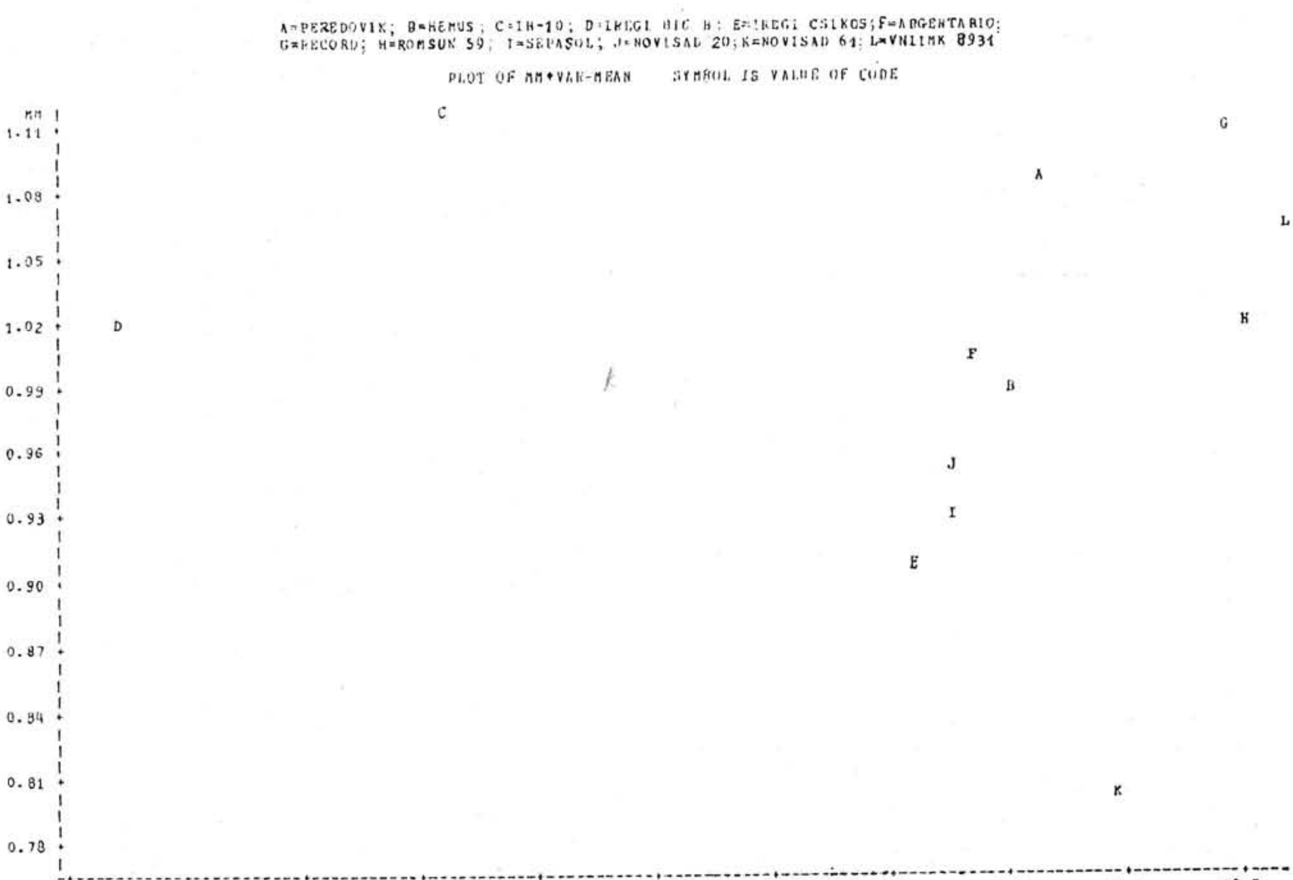

Fig. 4 — The relationship between variety adaptation and variety mean yield for 12 sunflower varieties of the 1979 European co-operative trials with open pollinated varietites (data from Vrânceanu et al., Helia No. 2, page 8).

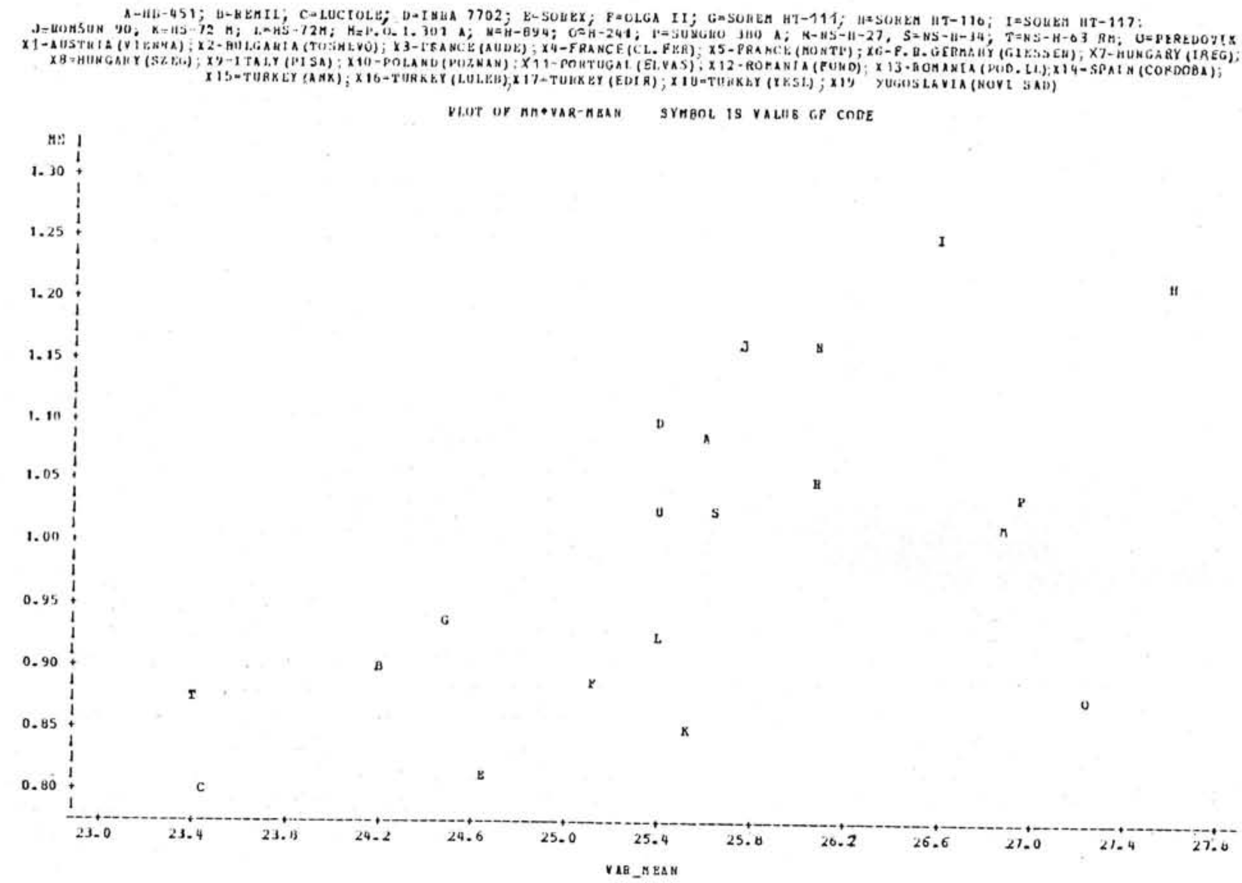

Fig. 5 — The relationship between variety adaptation and variety mean yield for 20 sunflower hybrids of the 1978 European co-operative trials with hybrid varieties (data from Vrânceanu et al., Helia No. 2, page 13).

Table 1

Mean yield, regression slope and r^2 calculated for the 20 sunflower hybrids tested in 1978

Cultivars	Yield	b	r^2
HB-451	25.6	1.09	0.90
REMIL	24.2	0.90	0.91
LUCIOLE	23.4	0.81	0.78
INRA 7702	25.4	1.16	0.95
SOREX	24.6	0.82	0.87
OLGA II	25.1	0.89	0.85
SOREM HT-111	24.5	0.93	0.90
SOREM HT-116	27.6	1.22	0.95
SOREM HT-117	26.6	1.25	0.91
ROMSUN 90	25.8	1.16	0.93
HS-1161	25.5	0.85	0.84
HS-72 M	25.4	0.93	0.86
P.O.I. 301 A	26.9	1.01	0.88
H-894	26.1	1.16	0.90
H-241	27.2	0.88	0.84
SUNGRO 380 A	27.0	1.03	0.91
NS-H-27	26.1	1.04	0.91
NS-H-34	25.6	1.03	0.88
NS-H-63 RM	23.4	0.87	0.88
PEREDOVIK	25.4	1.02	0.90

Table 2

Mean yield, regression slope and r^2 calculated for the 12 open pollinated sunflower cultivars tested in 1979

Cultivars	Yield	b	r^2
PEREDOVIK	27.8	1.09	0.94
HEMUS	27.5	0.99	0.90
IH-10	22.6	1.12	0.78
IREGI 816 B	19.9	1.02	0.67
IREGI CSIKÓS	26.6	0.91	0.67
ARGENTARIO	27.2	1.00	0.97
RECORD	29.3	1.11	0.96
ROMSUN 59	29.5	1.02	0.88
SEPASOL	27.0	0.93	0.94
NOVI SAD 20	27.0	0.95	0.83
NOVI SAD 61	28.4	0.80	0.72
VNIIMK 8931	29.9	1.06	0.90

There is also some argument about the interpretation of the expressions "adaptability" and "phenotypic stability" as introduced by Finaly and Wilkinson. They argue that the ideal variety having general adaptability (highest mean yields over all environments) is the one with maximum yield potential in the most favourable environment and maximum phenotypic stability (low regression coefficient). In their study they found that the varieties with high phenotypic stability all had low mean yields, they were so stable, in fact, that they were unable to exploit higher yielding environments. This is certainly not true for the sunflower studies, although the relatively low number of sunflower cultivars (12 and 20 respectively) compared to the 277 barley varieties tested by Finaly and Wilkinson makes it difficult to compare the two studies.

2. CLUSTER ANALYSIS

Another graphical way of critically analysing the performance of varieties is through the use of *cluster-analysis*. It is a commonly used term for various techniques which seek to separate the data into consistent groups. Such techniques are generally used for the grouping of objects or individuals under investigation by the similarity or dissimilarity of their performances. In general the objective in cluster analysis is to discover a category structure which fits the observations. There are several different clustering techniques, the simplest and most often used ones are the so-called *hierarchical* techniques in which the classes themselves are classified into groups, the process being repeated at different levels to form a tree. Hierarchical methods may be subdivided into *agglomerative* methods which proceed by a series of successive fusions and *divisive* ones, which partition a set of N entities into finer and finer partitions. The cluster analysis performed on the sunflower data were performed with the former type of method : it begins by forming one cluster for each observations in the analysis. The two closest clusters are combined into one cluster, then the two closest of the new set of clusters are combined into a single cluster and so on.

In adaptation trials cluster analysis can be used for two different purposes, each possibly interesting to the breeder : to develop groups of cultivars which behave similarly in the set of environments where they were tested (clustering genotypes) and for developing groups of environments which, based on the performances of a particular set of genotypes display similarities (clustering environments). As an example based on average yield at 13 locations the varieties Peredovik (USSR) and Hemus (Bulgaria), Argentario (Italy) and Sepasol (Spain), Record (Romania) and Romsun 59 (Romania) show pairwise similarities in their performances (Fig. 6). For the same set of trials when clustering environments Poland and Romania (Fundulea), Bulgaria (Toshevo) and Turkey (Ankara), Spain and Portugal, and the two other Turkish locations Luleburgaz and Edirne show pairwise similarities (Fig. 7). There is very little doubt that the basis of this classification is geographical nearness.

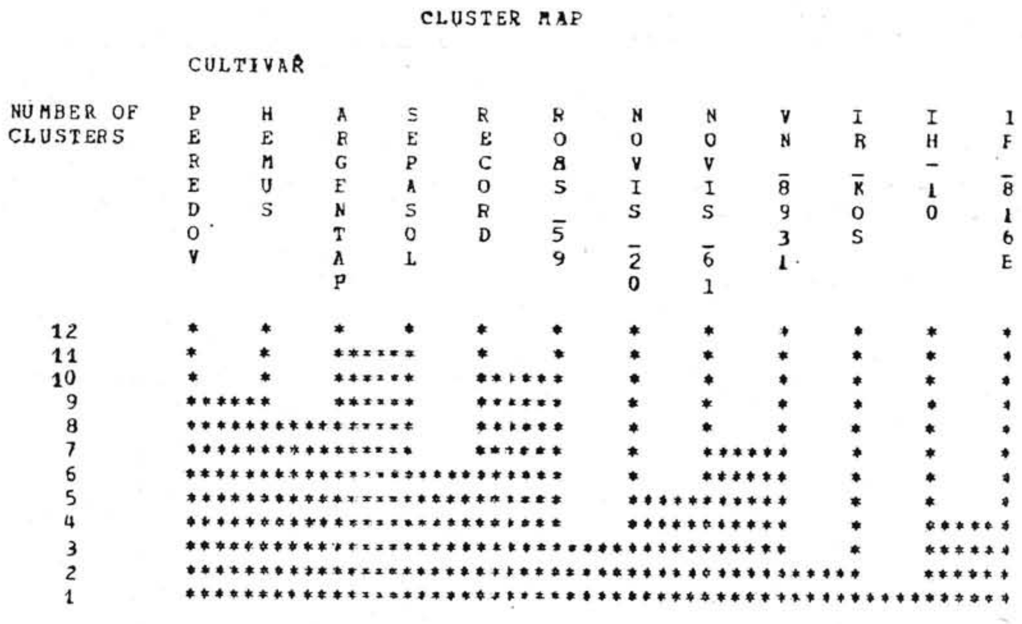

Fig. 6 — Cluster map (dendrogram) of the 12 sunflower varieties of the 1979 European co-operative trials with open pollinated varieties (data from Vrânceanu et al., Helia No. 2, page 8).

The number of groups one would be satisfied with is, of course, left to the investigator. Using our cluster map of environments and deciding on six groups one can find two groups of four countries each : Bulgaria,

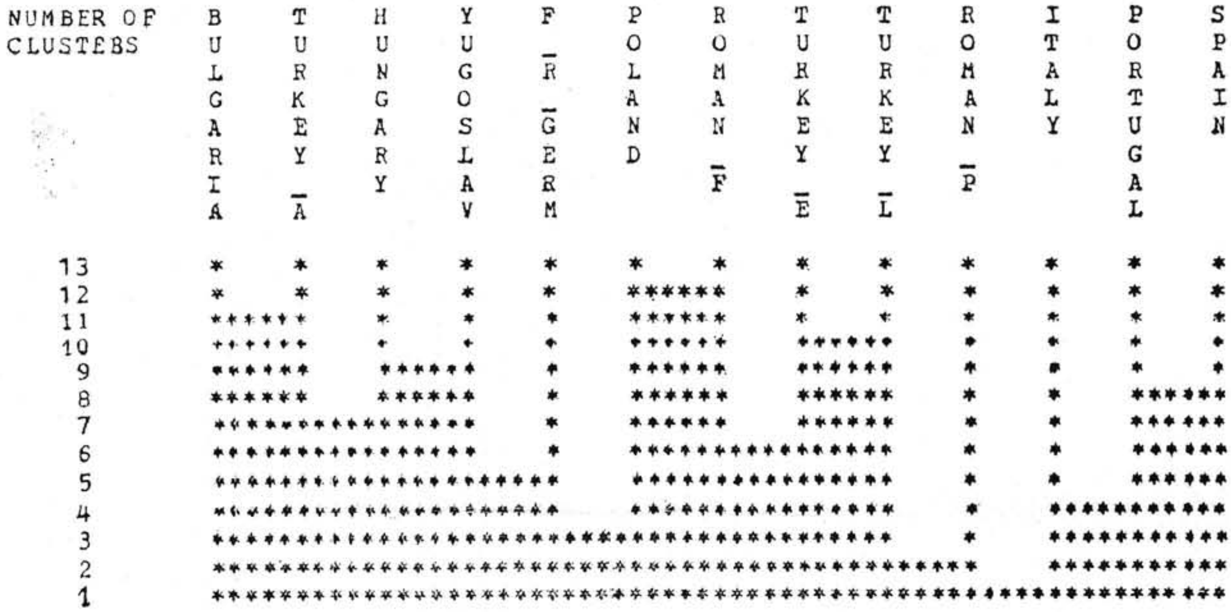

Fig. 7 — Cluster map (dendrogram) of 13 locations at which 12 open pollinated varieties were tested in 1979 (based on mean yield data). Data from Vrânceanu et al., Helia No. 2, page 8.

Turkey (Ankara), Hungary and Yugoslavia being one and Poland, Romania (Fundulea), Turkey (Luleburghez and Edirne) in the other, the Federal Republic of Germany, Romania (Podu Iloaie) and Italy (Pisa) are all groups themselves, and Portugal and Spain forming the sixth group. This, surely makes sense to everybody. But we would like to go further than this. One would like to know the underlying causes of the similarities or dissimilarities displayed by the cluster map. It should become clear that cluster analysis in genotype-environment interaction studies is not merely a means of reducing the data but also of generating hypotheses that can be tested with different methods. One obvious question to be answered : what kind of clusters would we get if we expressed the data points not in their original form but as deviations from both genotypic, and environmental means ?

Looking at these deviations, for example, it becomes obvious that in Hungary in 1978 (Table 3) the yield of Novi Sad 61 and VNIIMK 8931 was far above what one would have expected based on marginal means (6.61 and 5.03 q/ha more respectively), and the variety Iregi 816 yielded much less than expected (—9.21 q/ha). Looking at this table

Table 3

Deviations from variety and location means together with the first three principal components of variation of the 12 open pollinated varieties of the 1979 European co-operative sunflower trials

Obs	Cultivar	Code	Bulgaria	F. R. Germ.	Hungary	Italy	Poland	Portugal	Roman.-F.
1	PEREDOV	A	—0.7942	0.8391	—3.2609	—1.5609	2.1058	3.0308	1.3391
2	HEMUS	B	0.1673	—0.3994	3.6006	—2.4994	—2.1327	3.9923	1.0006
3	IH-10	C	0.2212	3.7545	0.3545	3.3545	—2.7788	—8.4538	3.4545
4	IR 815 B	D	2.9596	0.4929	—9.2071	8.8929	—0.0404	—5.9154	0.2929
5	IR KOS	E	—1.0788	6.7545	4.4545	2.1545	0.4212	0.6462	1.8545
6	ARGENTAR	F	—0.4404	1.2929	—2.4071	—0.8071	0.0596	—0.0154	—2.1071
7	RECORD	G	—1.4788	1.2545	—1.3455	—3.3455	—0.1788	—0.9538	—0.1455
8	ROMS-59	H	—0.0942	—1.5609	—4.2609	—3.3609	2.8058	4.7308	2.3391
9	SEPASOL	I	—1.2250	1.6083	—1.4917	—1.5917	—0.8250	2.3000	—0.1917
10	NOVIS-20	J	1.2904	—7.3763	1.9237	3.6237	—1.3096	—1.4846	—2.5763
11	NOVIS-61	K	2.0750	—3.3917	6.6081	—1.4917	—0.9250	1.9000	—1.8917
12	VN-8931	L	—1.6019	—3.2686	5.0314	—3.3686	2.7981	0.2231	—3.3686

Obs	Roman-P.I.	Spain	Tur-key-A	Turkey-E	Turkey-L	Yugoslav	Prin 1	Prin 2	Prin 3
1	2.281	—0.3742	—3.2109	2.8391	—0.5692	—2.6942	—0.2335	—0.3628	2.3431
2	—1.358	0.4173	—2.5494	4.3006	—2.5077	—2.0327	0.1521	—0.3344	0.5959
3	8.196	0.0712	—1.5955	—7.7455	—1.0538	2.2212	—3.4880	—0.4409	—1.6255
4	10.135	2.8096	—1.7571	—2.4071	—0.4154	—5.8404	—3.6117	3.9333	0.4082
5	3.796	—1.6288	—3.9955	—9.5455	—8.2538	4.4212	—2.7888	—3.9042	—0.5472
6	0.235	—0.0904	0.8429	0.2929	0.1846	2.9596	0.3984	—0.3478	—0.3011
7	—0.504	0.1712	3.2045	3.2545	1.4462	—1.3788	0.9439	0.2197	0.5278
8	—3.519	—0.1442	—0.0109	2.2391	2.6308	—1.7942	1.1767	0.2232	2.2502
9	—4.150	0.8250	1.0583	1.4083	1.9000	0.3750	0.7884	—0.0006	0.4726
10	—2.235	—1.3596	4.1737	3.8237	2.6154	—1.1096	1.9349	1.7611	—1.9566
11	—10.250	0.4250	3.1583	—0.7917	2.3000	2.2750	2.2272	0.4615	—1.9792
12	—2.627	—1.1519	0.6814	2.3314	1.7231	2.5981	2.5005	—1.2081	—0.1880

another way, e.g. in the same trials the variety Iregi 816 yielded 10.1 and 8.9 q/ha more in Romania (Podu Iloaie) and Italy (Pisa) respectively than what one would have expected based on mean values and the same variety yielded much less than expected both in Hungary (—9.2 q/ha) and Yugoslavia (—5.8 q/ha).

The cluster dendrograms based on deviations give completely different sets of clusters from those based on the original data (Fig. 8). Now the varieties Record, Sepasol, Argentario and VNIIMK 8931 fall into a cluster, Peredovik, Hemus and Romsun 59 into another and the remaining varieties are clusters by themselves showing individual behaviour. The clusters of environments based on deviations rather than original data (Fig. 9) are not based on geographical distribution anymore but show certain similarities between Bulgaria, Spain, Romania (Fundulea) and Poland, Italy and Podu Iloaie (Romania), Hungary and Yugoslavia, Ankara and Luleburgaz in Turkey showing hidden similarities while the remaining three environments (Germany, Portugal) and Edirne (Turkey) are clusters by themselves. For the successful utilization of plant-breeding material we need to understand the nature of these similarities and dissimilarities. These are probably related to the various components of the environment, such as rainfall, temperature, evapotranspiration, photoperiod,

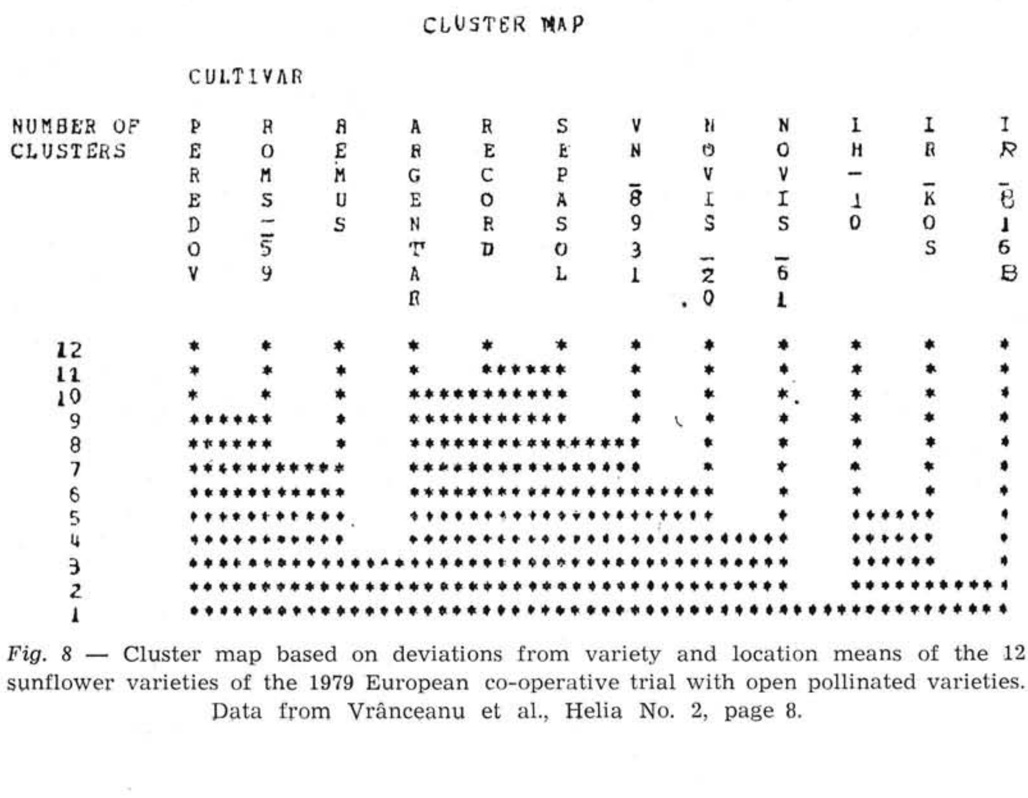

Fig. 8 — Cluster map based on deviations from variety and location means of the 12 sunflower varieties of the 1979 European co-operative trial with open pollinated varieties. Data from Vrânceanu et al., Helia No. 2, page 8.

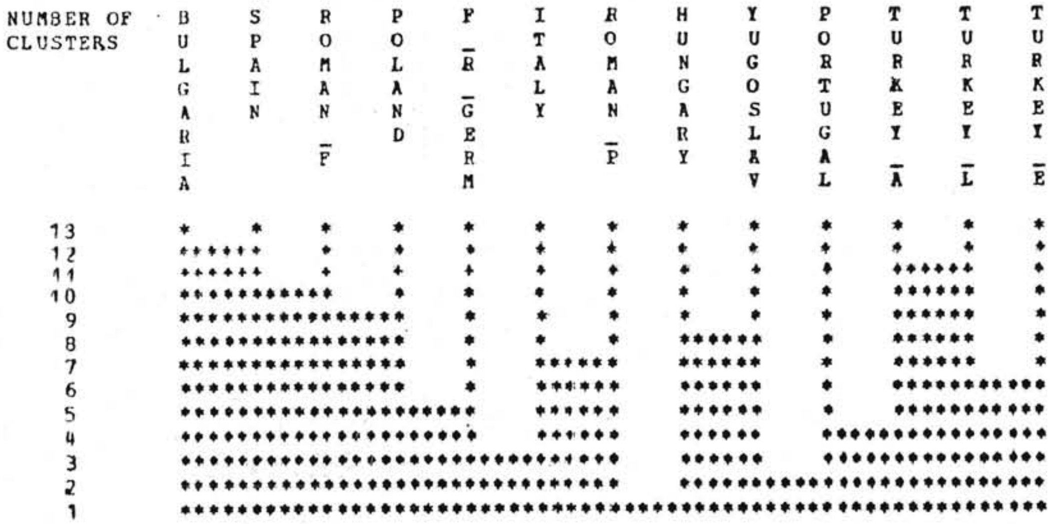

Fig. 9 — Cluster map based on deviations from variety and location means of the 13 locations at which 12 open pollinated varieties were tested in 1979 as part of the European co-operative trials with open pollinated varieties. Data from Helia No. 2.

soil type and depth or any other environmental factor one could think of. These data are absolutely necessary if we wish to understand the relationships of genotypes with the environments they are grown in. The aim of the standardized adaptation trials organized by the Sunflower Network in collaboration with F.A.O.'s Crop and Grassland Production Service is to obtain this information together with the other agronomic data. In order to understand the relationships of genotypes grown in a set of environments as well as the relationships of the environments which were involved in the testing of a set of genotypes it is useful to look at a third system of classification, those obtained by the use of principal component analysis.

3. PRINCIPAL COMPONENTS ANALYSIS

Given a set of n variables, some or all of which may be intercorrelated, it is possible to construct a set of n or fewer orthogonal (independent) variables which are linear combinations of the original variables and which account for the variance of the data. To put it another way, the axes representing the original variables may be rotated individually to be orthogonal to each other and in the process it may be found that the dimensionality of the data can be reduced. A particular technique of finding such a set of orthogonal variables is the method of principal components. They are of great interest not only because their coefficients define a set of orthogonal vectors, but because of their maximum variance properties. The first principal component has the largest variance of any linear combination of the variables represented in the data matrix ; the second principal component has the largest variance of any linear combination orthogonal to the first one ; the third one has the largest variance of any linear combination orthogonal to the first two, and so on. The degree of variation explained by the first k principal components can be calculated. Should the first two components, for example, explain about 90% of the existing variation then the grouping obtained by the use of the principal components will be very powerful. This was not the case with our sunflower data : only 61% of the 1979 (open pollinated varieties) data can be explained by the first two principal components ; for the 1978 hybrid experiment this figure is only 45%. Thus the separation of varieties is not as good as one would like it. Nevertheless the relationship between the cluster analysis dendrogram and the two dimensional plots of genotypic scores evaluated at the first, second and third principal components is quite obvious (Fig. 10, 11, 12).

The advantage of the principal components over the cluster analysis is not only that it is easier to separate groups in two dimensions than it is in one, but that one can do tests of hypotheses. Freeman and Dowker (1973) have considered the components of variation within genotypes and then doing

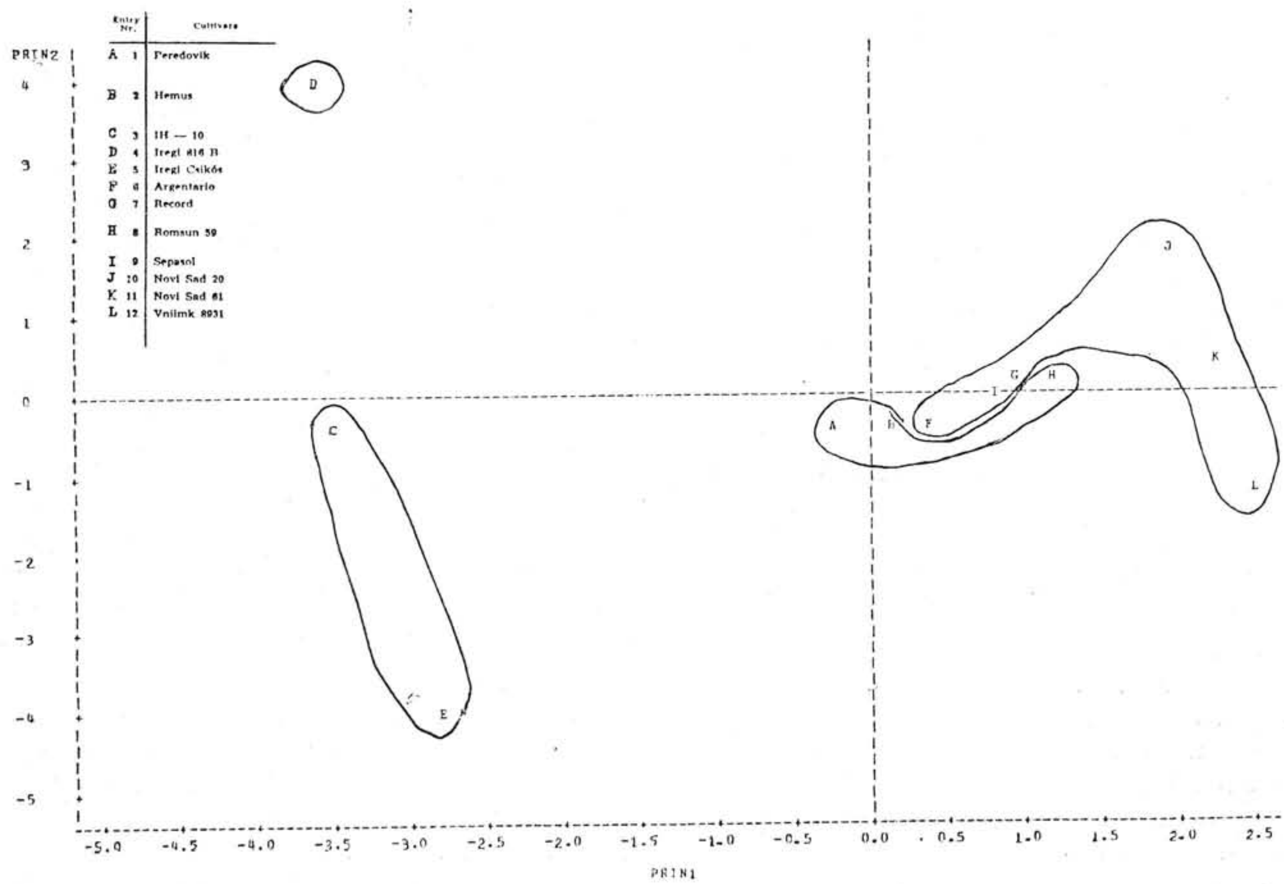

Fig. 10 — Plot of the first two principal components of the variation of deviations from variety and location means of the 12 open pollinated varieties in the 1979 European co-operative sunflower trials. Superimposed on the plot are the clusters obtained by the cluster analysis of the same data (Fig. 8).

analysis of variance on the resulting scores. The variation within environments can be examined similarly. They found that in these studies the first principal component of the variation within genotypes is closely related to the main effect of environments, but that the converse does not necessarily hold : the first component within environments is not always the main effect of genotypes. In our sunflower data we cannot say unambigously if any of the above statements are true. The best way to examine these questions is to regress the principal components on the various components of the environment. For this we need the meteorological and soil data, which we did not have available at the time of preparing this paper. One of the main advantages of the standardized adaptation trials are that we shall be able to clarify these points.

With appropriate analyses we hope to shorten the period of testing possibly by several years and lessen the number of locations where cultivars are grown thus saving time, money and effort, matters particularly important in developing countries.

REFERENCES

Eberhart, S. A., Russel W. A., 1966, *Stability Parameters for Comparing Varieties,* Crop Sci., 6, 36—40.

Finaly K. W., Wilkinson G. N., 1963, *The Analysis of Adaptation in a Plant-Breeding Programme,* Aust. J. Agric. Res., **14**, 742—754.

Freeman G. H., Dowker B. D., 1973, *The Analysis of Variation Between and Within Genotypes and Environments,* Heredity, **30**, 97—109.

Yates F., Cochran W. G., 1938, *The Analysis of Groups of Experiments,* J. Agric. Sci., Camb., **28**, 556—580.

MÉTHODES GRAPHIQUES POUR ÉVALUER L'ADAPTABILITÉ DES VARIÉTÉS ET DES HYBRIDES DE TOURNESOL

Résumé

Les essais comparatifs internationaux des cultivars de tournesol implantés par le réseau de recherches F.A.O., donnent la possibilité d'établir par l'analyse de la variance, les interactions génotype-environnement, la mesure dans laquelle les conditions pédo-climatiques sont favorables à cette culture etc. Le présent ouvrage évalue les méthodes de calcul et d'interprétation des données sur les performances des géno-

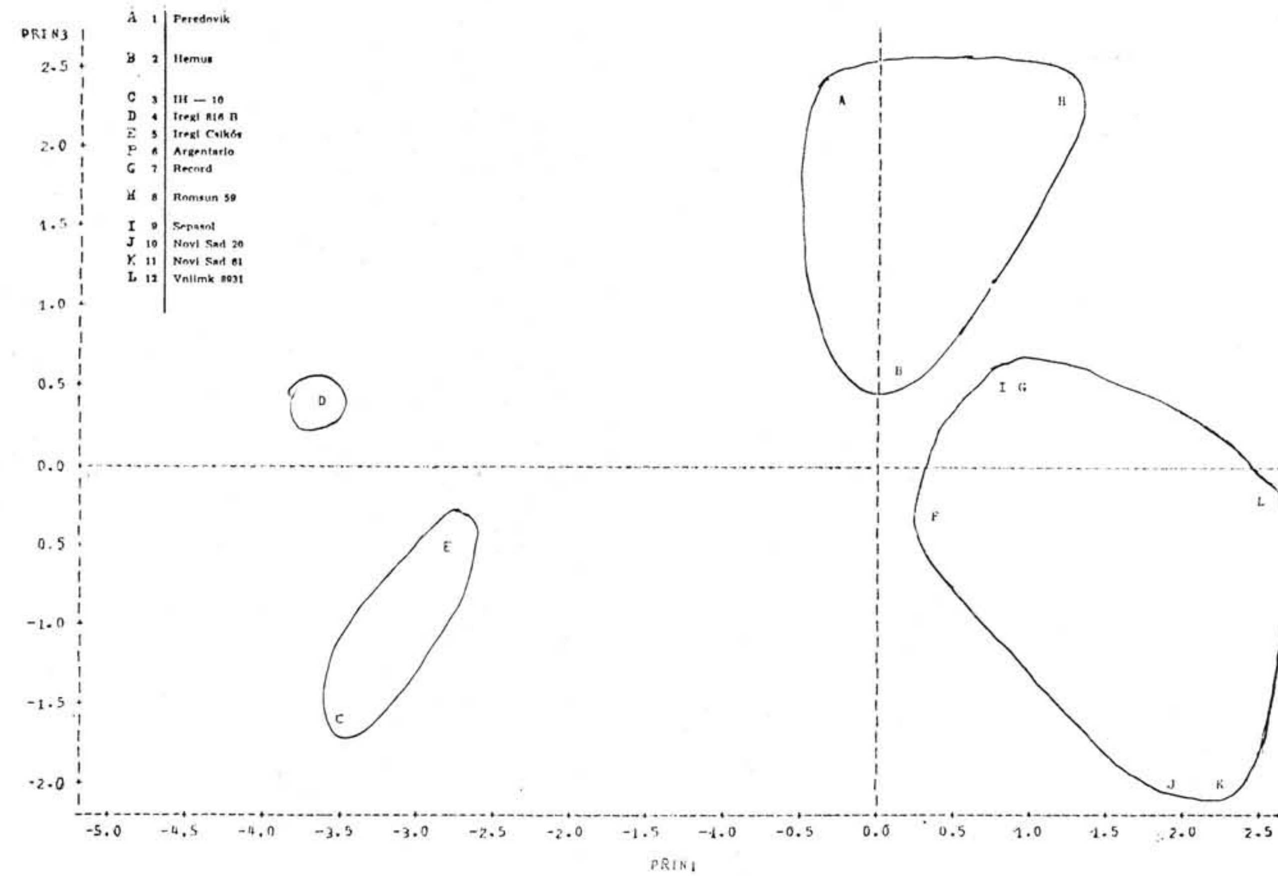

Fig. 11 — Plot of the first and third principal components of the variation of deviations from variety and location means of the 12 open pollinated varieties of the 1979 European co-operative sunflower trials. Superimposed on the plot are the clusters obtained by the cluster analysis of the same data (Fig. 8).

types : indice des conditions pédo-climatiques (environmental index), analyse de la régression articulée (joint regression analysis) et le diagramme de dispersion (scatter diagram). Comme exemple on cite des méthodes graphiques d'analyse de la réaction des cultivars de tournesol aux conditions des essais : analyse de la régression articulée en utilisant le coefficient de détermination et l'analyse cluster (cluster analysis) et la technique hiérarchique (hierarchical techniques). L'analyse cluster peut servir aux sélectionneurs pour séparer des groupes de cultivars ayant un comportement semblable dans des conditions différentes d'environnement (clustering genotypes) ou bien des groupes d'environnement ayant des conditions semblables (clustering environments). La méthode peut être utilisée pour l'analyse des principaux éléments composants, car comparée à l'analyse cluster, elle donne la possibilité de séparer plus facilement les groupes (ayant deux dimensions) et permet de tester toutes les hypothèses.

EVALUACIÓN DE LA ADAPTABILIDAD DE LAS VARIEDADES E HÍBRIDOS DE GIRASOL : ALGUNOS MÉTODOS GRAFICOS

Resúmen

La experimentación en culturas comparativas internacionales de los cultivares de girasol, efectuada en el marco de la Red de Investigación F.A.O., ofrece amplias posibilidades para establecer con la ayuda del analisis de la variación las diversas interacciones genotipo-medio, la favorabilidad de las condiciones de medio para esta cultura etc. En este trabajo se hacen referencias y apreciaciones sobre diferentes métodos de cálculo e interpretación de las performancias de los genotipos : el índice de las condiciones de medio (environmental index) el análisis de la regresión articulada (joint regression analysis) y el diagrama de la dispersión (scatter diagram).

Se proponen y se dan como ejemplos algunos métodos gráficos de análisis de la respuesta de diferentes cultivares de girasol a diferentes condiciones de experimentación, como el análisis de la regresión articulada, empleando el coeficiente de determinación y el análisis cluster (cluster analysis) empleando la técnica jerárquica (hierarchical techniques). El análisis cluster puede ser empleado por los mejoradores tanto para establecer grupos de cultivares de comportamiento semejante en un set de condiciones diferentes de medio (clustering genotypes) como también grupos de medios semejantes (clustering environments).

Está sugerido también el empleo del método de análisis de los principales componentes que, en comparación al análisis cluster, separa más facilmente los grupos (ya que cuenta con dos dimensiones) y puede testar los hipótesis.

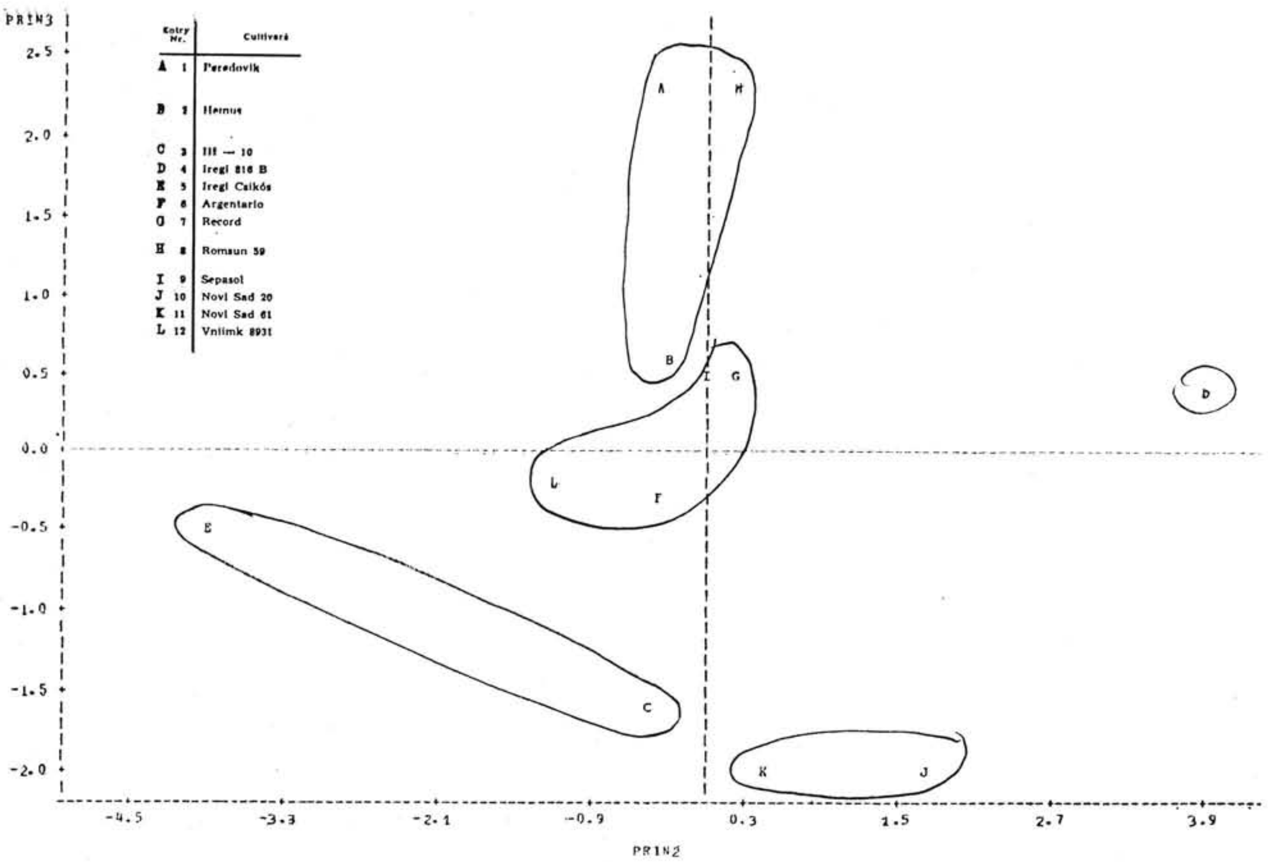

Fig. 12 — Plot of the second and third principal components of variation of deviations from variety and location means of the 12 open pollinated varieties of the 1979 European co-operative sunflower trials. Superimposed on the plot are the clusters obtained by the cluster analysis of the same data (Fig. 8).