Euphytica 34 (1985) 113–123

MULTIVARIATE ANALYSIS OF A COLLECTION OF SOYBEAN CULTIVARS FOR SOUTHWESTERN SPAIN[1]

R. BARTUAL, E. A. CARBONELL and D. E. GREEN[2]

Instituto Nacional de Investigaciones Agrarias (INIA), José Abascal 56, Madrid-3, Spain

Received 2 April 1984

INDEX WORDS

Glycine max, soybean, variation, varietal stability, multivariate techniques, clustering.

SUMMARY

Multivariate techniques were used to classify 125 soybean lines into clusters. Late maturing varieties belonging to maturity group III showed the best adaptation to the ecological conditions of the area when soybeans were sown at early planting dates. A group of experimental lines, the majority of which had semideterminate stem termination with small leaflets and intermediate maturity, were highly productive when grown as a second crop. Some lines of other groups were identified as likely parents for use in a breeding program to improve agronomic characteristics. The identified groups were quite stable in their performance through changes in environmental conditions (years and planting dates). The analyses indicated that some intercorrelated traits can be omitted in future line evaluation.

INTRODUCTION

Knowledge of the relationship of yield in soybeans (*Glycine max.* (L.) MERR.) with its main components can be of assistance to plant breeders in making selections. In recent years, plant breeders have postulated that selection for components of yield only may not necessarily be the most efficient means to produce lines with improved performance. Both physiological and morphological characteristics of the soybean plant are known to play major and interdependent roles in determining yield (DENIS & ADAMS, 1978; CARBONELL & BARTUAL, 1983), the differences in the range of expression of such characters being a result of the genetic diversity. The utility of multivariate analysis for measuring the degree of divergence between biological populations and for assessing the relative contribution of different characters to the total divergence has been established by several authors, including GHADERI et al. (1979) in mung bean; BHATT (1976), and JOSHI & SINGH (1979) in wheat; NARSINGHANI et al. (1979) in peas; SACHAN & SHARMA (1971) in tomato; SINGH & GUPTA (1968) in cotton; HUSSAINI et al. (1977) in finger millet; SASMAL (1978) in jute; and BROICH & PALMER (1980) in wild and domesticated soybeans.

[1] Experimental work and data analyses supported by the Instituto Nacional de Investigaciones Agrarias of Spain and the IRI Research Institute, Inc. through a World Bank Grant.

[2] INIA, Moncada (Valencia), Spain; INIA, Madrid, Spain; and Iowa State University, Ames, Iowa 50011. Consultant with IRI Research Institute, Inc. (1975–77).

Numerical taxonomy and its related multivariate methods have been used to assist in the selection of parental combinations that would presumably result in high yielding progenies (GHADERI et al., 1979; SNEATH, 1976; WHITEHOUSE, 1971; and SINGH et al., 1980). These studies have been aimed at classifying genotypes into divergent groups. Classifications that are based on an overall combination of characters should result in groups within which genotypes are similar for some characters and dissimilar for others.

The main purposes of this study are to classify one hundred and twenty five soybean lines into groups on the basis of physiological, morphological, seed quality, and productive traits, and to use multivariate techniques to determine which groups are better adapted to specific ecological conditions. This study should also help to identify the groups that would be best suited as a primary crop as well as those that would be best for double cropping or as a second crop. A final purpose is to identify which lines would be the best parents for future breeding programs designed to produce lines with improved specific characteristics.

MATERIAL AND METHODS

One hundred and twenty-five soybean lines introduced from the United States and belonging to maturity groups I to IV, were sown at the 'Haza del Monte' farm, near Sevilla, Spain, on two dates of planting and in two years. The first date was in the middle of May, the second date was in the middle of June. A complete randomized block design with three replications within dates and years was used.

Data were collected for 25 different traits. Some are related to the life cycle of the plant, including days to emergence (E), date of blooming (B), beginning seed formation (Se), physiological maturity (PM), and harvest maturity (M). Other characters were lodging at beginning seed formation (LSe), % plants emerged (P), lodging at harvest maturity (LM), plant height at harvest maturity (HM), height of the lowest pod (HLP), number of reproductive nodes at harvesting (NNM), growth habit (GH), leaflet width at beginning seed formation (LWSe), leaflet area at beginning seed formation (LASe), and disease occurrence (DO). Finally, data for characters associated with seed quality, chemical composition of the seed and yield, including mottling (Mot), green cotyledons (GC), shriveling (Sh), wrinkling (Wr), imperfect seedcoats (ISC), weathering (W), protein content (Po), oil content (O), seed size (SS), and yield (Y) were also recorded.

Plots were three meters long and two meters wide, each of them including four rows, spaced 50 cm apart and the planned plant density was 35 plants per linear meter.

Data related to the life cycle of the plant were collected when 50% of the plants in the plot showed the trait in question. Lodging at beginning seed, lodging at the time of harvesting, and all characters associated with seed quality were evaluated on a scale from 1 to 5 (1 = abscence; 5 = maximum expression). Growth habit was evaluated on a 1 to 5 scale (1 = determinate and 5 = indeterminate stem termination).

Data concerning characters such as % plants emerged, leaf area and leaflet width were collected on plants in two central square meters of each plot, with leaf area being evaluated visually by scores ranging from 1 (small size) to 5 (large) and leaflet width from 10 to 50, respectively. Yield was recorded as the weight of the seed harvested

Table 1. Communalities obtained from three factors.

Trait	Communality	Trait	Communality
B (Days to blooming)	0.6394	LASe (Leaflet area at seed formation)	0.5379
Se (Days to seed formation)	0.9022	DO (Disease occurrence)	0.3431
PM (Days to physiological maturity)	0.9974	Mot (Mottling)	0.1059
M (Days to harvest maturity)	0.9513	GC (Green cotyledons)	0.3438
E (Days to emergence)	0.0349	Sh (Shriveling)	0.1697
P (% plants emerged)	0.0165	Wr (Wrinkling)	0.1359
LSe (Lodging at seed formation)	0.9733	ISC (Imperfect seed coat)	0.1592
LM (Lodging at maturity)	0.9452	W (Weathering)	0.3908
HM (Height at maturity)	0.7194	Po (% Protein)	0.9054
HLP (Height of lowest pod)	0.6266	O (% Oil)	0.9511
NNM (Number nodes at maturity)	0.6508	SS (Seed size)	0.1791
GH (Growth habit)	0.1806	Y (Yield)	0.5545
LWSe (Leaflet width at seed formation)	0.5326		

from the two central meters of the two middle rows of each plot, and was converted into kilograms per hectare on a 13% moisture basis. Seed size was the mean weight of two samples of 100 seeds randomly gathered from the harvested seed of each plot, expressed also in grams at 13% seed moisture. Protein and oil content were measured using an infraanalyzer and was expressed as percentage of dry matter. The values for protein and oil were the means of the readings obtained from three samples of 100 randomly gathered seeds from each plot.

Disease occurrence due to spider mites (*Tetranychus urticae* KOCH.) or soybean mosaic virus (SMV) was visually recorded on a subjective scale from 1 (minimum) to 5.

Statistical methods. Maximum likelihood factor analysis (LAWLEY & MAXWELL, 1971) was used on data averaged over years, dates, and replicates. Once the eigenvalues and eigenvectors were obtained, the matrix of factor loadings (correlations between original variables and factors) was submitted to a varimax rotation. Principal components analysis, based on the correlation matrix, was also used on the same set of data and in some subsets of data; namely, averaged over replicates and years or dates as well as over years and replicates for each date separatedly. Cluster analysis using euclidean distance was carried out on the overall means. To aggregate clusters the distance was based on Ward's (1963) method, which optimizes the residual sums of squares of the objective function. Groups were arbitrarily determined from cluster analysis, and thereafter their content was verified by discriminant analysis.

RESULTS

The communalities obtained using three factors are included in Table 1. High values of the communalities correspond to variables related to life cycle, especially those of late stages of development (physiological and harvest maturities). Other variables explained by the factor model are chemical composition of the seed (% protein and %

Table 2. Rotated factor loadings of 25 traits on three factors (those values arbitrarily considered as important are in italics).

Trait	Factor I	Factor II	FactorIII
B	*0.701*	0.351	–0.157
Se	*0.878*	0.363	–0.003
PM	*0.959*	0.274	0.042
M	*0.940*	0.249	0.068
E	0.004	–0.187	–0.003
P	0.086	–0.095	0.015
LSe	0.226	*0.950*	–0.142
LM	0.342	*0.910*	–0.033
HM	*0.685*	*0.500*	0.000
HLP	*0.729*	0.308	–0.018
NNM	*0.667*	*0.453*	–0.009
GH	0.114	0.394	–0.112
LWSe	*0.664*	0.291	–0.085
LASe	*0.672*	0.275	–0.100
DO	0.059	*0.582*	–0.026
Mot	0.083	0.044	*–0.312*
GC	–0.576	0.073	0.084
Sh	–0.360	0.152	–0.132
Wr	–0.334	0.118	0.103
ISC	0.382	–0.098	–0.058
W	0.525	0.243	*–0.236*
Po	0.036	0.056	*–0.949*
O	–0.018	–0.111	*0.969*
SS	0.381	0.035	–0.181
Y	*0.615*	–0.012	*0.419*

oil), lodging at beginning seed formation and lodging at maturity. Those variables connected with plant architecture such as height and number of nodes at maturity, height of the lowest pod, leaflet width and leaflet area at seed formation were explained to a lesser extent by the factor model. Yield showed a moderate communality, and seed quality variables had large residual error terms.

The first factor of the common factor model explained 34% of the total variance; the second and third factors accounted for 10.4% and 7.4% respectively. Hence, a comparatively high variance was absorbed by the first factor in this study. As shown in Table 2, factor I is mainly associated with traits concerning life cycle, with plant architecture, and with yield. Factor II was highly correlated with both lodging at beginning seed formation and at maturity, with traits reflecting height, and with disease occurrence. Factor III was closely related to seed composition, and moderately associated with yield, mottling, and weathering.

Results using principal components, shown in Table 3, were somewhat similar to those from factor analysis. In fact, the first principal axis had the same general composition as the first factor; however, the second and third axes were intermingled as compared to factor analysis. These three principal components were responsible for 57.3% of the variance.

Lines were subjectively classified into six groups, utilizing factors I and II. The mean

Table 3. Composition of the first three principal components for 25 traits (those values arbitrarily judged as important are in italics).

Trait	PC1	PC2	PC3
B	*0.265*	0.072	–0.042
Se	*0.307*	–0.041	0.028
PM	*0.314*	–0.106	–0.014
M	*0.306*	–0.114	–0.004
E	–0.040	–0.037	–0.036
P	0.010	–0.173	–0.069
LSe	0.217	*0.257*	*0.211*
LM	0.240	0.187	*0.262*
HM	*0.287*	–0.009	0.150
HLP	*0.272*	–0.074	0.048
NNM	*0.274*	0.020	0.137
GH	0.126	0.141	0.144
LWSe	*0.256*	–0.004	0.051
LASe	*0.259*	0.043	–0.070
Do	0.103	0.219	0.193
Mot	0.030	*0.263*	–0.111
GC	–0.170	*0.270*	*0.280*
Sh	–0.080	*0.397*	0.194
Wr	–0.071	0.231	*0.306*
ISC	0.115	–0.075	*–0.294*
W	0.194	0.178	0.168
Po	0.048	*0.328*	*–0.437*
O	–0.046	*–0.359*	*0.418*
SS	0.145	0.104	*–0.213*
Y	0.169	*–0.349*	0.128

performance of each group for the 25 traits is presented in Table 4.

Group 1 included mainly late maturing lines (standard maturity groups III and IV), with indeterminate growth habit and high seed yield. The higher yielding lines belonged to maturity group III and also exhibited an acceptable level of seed protein and oil. The most undesirable characteristic of this group was the presence of lodging. However, the most severely lodged lines belonged to group IV. Lines of group 1 had large seeds and hence had a higher degree of imperfect seedcoats. Because the lines of this group were the latest in maturity, weathering of the seed was a problem in some cases.

Group 2 was composed mainly of lines of maturity group II (medium-late), with indeterminate growth habit and with less vegetative development than those of group 1. Lines of this group were slightly lower in yield than were those of group 1, were also intermediate in many characters and did not show great promise for use in the area.

Group 3 contained a set of plant introductions that gave a low seed yield, with a high protein content. Hence, they could possibly be used in a breeding program to improve this characteristic.

Group 4 included a set of experimental lines that were slightly earlier than those of group 2 and most of them were semideterminate in growth habit. Their seed had

Table 4. Mean performance for the 25 traits of all lines within groups averaged over years, dates and replications.

Trait	Group					
	1	2	3	4	5	6
B	44.9	38.0	36.4	35.2	34.3	33.3
Se	72.5	62.9	57.9	59.9	54.6	51.6
PM	110.4	98.2	89.4	93.9	83.2	72.6
M	146.2	129.1	115.4	122.3	105.1	93.8
E	10.1	10.1	10.2	10.2	10.1	10.3
P	71.0	64.1	68.7	69.8	68.7	43.6
LSe	2.0	1.8	1.6	1.2	1.4	1.2
LM	2.6	2.2	1.9	1.5	1.5	1.2
HM	119.7	94.5	80.4	80.8	72.1	54.9
HLP	16.1	12.4	9.9	10.9	7.6	3.9
NNM	119.7	16.2	14.8	14.9	13.5	13.7
GH	4.2	3.9	3.7	2.6	3.5	3.0
LWSe	48.3	47.3	45.1	43.2	41.8	34.9
LASe	3.1	2.9	2.6	2.4	2.3	1.8
DO	1.4	1.6	1.5	1.3	1.3	1.3
Mot	1.9	1.6	2.7	1.9	1.4	1.9
GC	1.4	1.5	1.6	1.4	1.8	3.2
Sh	1.5	1.5	1.7	1.4	1.5	2.4
Wr	2.0	2.4	2.5	2.1	2.4	2.9
ISC	2.7	2.3	1.9	1.9	2.0	1.4
W	1.6	1.2	1.2	1.1	1.0	1.1
Po	40.4	39.7	42.2	39.2	39.4	39.1
O	23.5	24.1	22.2	24.4	24.2	24.2
SS	21.9	20.5	20.4	18.0	17.6	15.3
Y	4058	3954	3375	4118	3378	2433

a chemical composition similar to group 2, but yield was higher and seed quality was better. Most of the experimental lines had small leaflets and a low degree of lodging.

Group 5 contained lines belonging mostly to maturity group I that were indeterminate and less productive, possibly because they were early maturing for the area. The % oil of some lines of the group was quite acceptable, but seed quality characters were generally inferior.

Group 6 contained only three very early maturing lines that offered no apparent advantages for the area.

Once the groups of lines had been delimited by cluster analysis, it was interesting to check the validity of the groups and also to obtain the minimum set of variables responsible for the classification. Stepwise discriminant analysis was made up in a progressive way by introducing variables one at a time. The classification functions, based on the minimum set of variables and ranked in descending order of their relative importance, are presented in Table 5.

The most important variable for discrimination was physiological maturity. Once in the model, the rest of the life cycle variables were redundant. Other variables of

Table 5. Classification functions obtained by using stepwise discriminant analysis for six groups of lines.

Trait	Group					
	1	2	3	4	5	6
PM	6.33	5.70	5.41	5.63	5.15	5.24
Po	42.74	42.52	44.41	42.56	41.78	42.34
GC	52.96	48.58	50.45	49.84	51.47	69.14
LWSe	5.36	5.05	4.77	4.13	4.04	1.67
Mot	6.45	4.49	9.45	6.51	4.29	5.16
DO	85.82	94.09	87.63	82.43	81.86	67.09
GH	6.88	4.33	4.29	1.37	3.91	1.74
Y	0.070	0.069	0.066	0.071	0.064	0.057
P	2.06	1.83	1.83	1.81	1.76	1.12
Wr	−11.82	−2.23	−0.55	−2.85	−0.30	7.98
W	−2.84	−13.43	−18.29	−16.69	−15.37	−28.13
Constant	−1661.02	−1550.83	−1579.25	−1485.46	−1392.69	−1302.59

influence were those related to productive traits and seed quality. Using the classification function with the lines used in the present study, only seven of them were misclassified.

DISCUSSION

Because significant line differences were found for all traits considered (CARBONELL & BARTUAL, 1983), they were included in the subsequent multivariate analysis. Previous estimates of correlations among traits indicated that there is a high positive correlation between traits associated with length of life cycle and those related to plant architecture. Seed quality, especially green cotyledons and shriveling, had a moderate negative correlation with yield. Seed mottling and weathering were also correlated with % protein and % oil (CARBONELL & BARTUAL, 1983).

Factor and principal components analyses. Average data set. When working with factor analysis (or principal components) one may question how many factors (or components) are needed to achieve a certain degree of explanation of the biological and agronomic nature of the data. No definite answer exists, the decision is usually quite arbitrary, and a compromise must be established. In the present study, the variance attributable to the first factor was three times higher than that of the second factor, with a steady decline in the contribution to the total variance by each factor as the number of factors increased.

A common method to estimate the factor loadings is the principal factor method as used with dry beans by DENIS & ADAMS (1978). It chooses the first factor so that it accounts for as much of the communal variance as possible, with the second factor accounting for as much as possible of the remaining communal variance, etc. If the estimates of the communalities are chosen to be unity, then the method reduces to principal components analysis (CHATFIELD & COLLINS, 1980). One useful advantage of the maximum likelihood method used in the present study as compared to principal

factor analysis is that the estimates are scale invariant. Factor analysis should be used with caution in analyses of biological data because it requires a number of questionable assumptions about the data and a proper underlying statistical model. LAWLEY & MAXWELL (1971) stressed that the model is useful only as an approximation to reality, but that it could be used as an aid to interpret the relationships among variables and to identify clusters of variables. From the results, it is clear that the first factor is related to traits reflecting life cycle and plant architecture. Both life cycle traits and plant architecture traits were positively correlated as mentioned above. Loadings in the rank of 0.60–0.96 indicate a high positive relationship between the two types of characters and this factor, suggesting that it will delimit groups of varieties differing essentially in maturity, late varieties with large vegetative growth having high positive values for that factor. Yield was associated with life cycle and plant architecture traits and negatively correlated with shriveling, wrinkling and green cotyledons (CARBONELL & BARTUAL, 1983). The association between seed size and the presence of imperfect seedcoats can be explained by the fact that the larger the seed, the higher the possibility of having a ruptured or imperfect seedcoat. Some of the individual variables used in the present study had strong associations including physiological and harvest maturity, leaflet area and leaflet width, indicating that one of the two traits involved in each pair may be disregarded in future studies and thus reduce the number of traits to be evaluated.

A statistical tool with somewhat similar objectives compared to factor analysis is principal components analysis. The two techniques differ substantially. The latter does not rely on any statistical model and a number of assumptions implied by factor analysis (multinormality, etc.) are not required. However, principal components analysis suffers from other drawbacks such as the absence of any 'error' structure and the dependence upon the scales used to measure the variables. Results from principal components analysis were of the same nature as those from factor analysis revealing that, in spite of the shortcomings of each method, conclusions obtained from their application are valid in the context of this study.

Discriminant analysis reassigned 7 of the 125 lines to a group different from that obtained by cluster analysis. Discriminant analysis was based on a subset of the original traits and maximized a different objective function, so some degree of disagreement was expected. Two lines, belonging originally to group 2, were classified as group 3. In fact, in a two components representation the reassigned lines were located on the border between the two adjacent groups as far as physiological maturity (variable highly correlated with the first principal component) is concerned, but their protein content was more similar to that of group 3. One line was moved from group 2 to group 1 because of wrinkling. Another line had a life cycle similar to group 2, but was maintained in group 1 because of the influence of wrinkling. Two lines were moved from group 2 to group 1 due to their lateness. One plant introduction was moved from group 3 to group 2, and a line was reclassified from group 4 to group 2 because of its yield and growth habit (indeterminate rather than semideterminate).

Stability over dates or years. Line stability was evaluated by components analysis considering each line in two different environments (dates or years) as two different lines and studying their relative location in the principal axes. Another way to look at the

problem is by investigating the correlation between the factors obtained by the analysis through different environments (DENIS & ADAMS, 1978). If the correlation is high it may be concluded that the causes for variety differences remain the same in different environments.

Principal components analysis over dates showed that the first axis maintained its meaning and the same was true for the second axis, except for % protein that when compared with the overall analysis shows an important negative effect. The negative effect was due to the general increase in protein in the late planting (FEASTER, 1942; OSLER & CARTTER, 1954). A general trend was indicated by a very noticeable displacement of varieties according to the second component. The first planting date was located toward the large values and the second date had lower values. However, this trend was not the same for all lines due to the fact that the life cycles of the late lines were much more affected by delay of planting than those of early lines (CHAPMAN & BLAMEY, 1979; GREEN et al., 1965, among others). Factor analysis showed the same trend over both dates and therefore, the same as overall data resulted, except for disease occurrence which played a less important role in the second factor of the first date as compared with second date and overall. Therefore, the presence of this trait as important characteristic in defining the second factor in the averaged data set is mainly due to the marked influence of this variable in the second dat of the first year, given that year 2 had no disease incidence at all. On first date, the attack of red spider mites and/or soybean mosaic virus was so severe that all varieties showed a high degree of infestation whereas for second date after the chemical treatment only those varieties that were susceptible and had a high degree of lodging showed large values for this trait. Hence, its variability among varieties was large as compared with first date, thus being an important characteristic to differentiate them. The association between lodging and disease occurrence found by factor analysis is also revealed by its correlation coefficients with values of 0.212 and 0.071 for early and late lodging on first date and 0.675 and 0.708 for second date, respectively.

Relative performance of lines by principal components analysis was maintained over years (MUNGOMERY et al., 1974). Only a slight difference was found in the response of the later and earlier varieties. This difference was indicated by a small but significant lines × years interaction (CARBONELL & BARTUAL, 1983). No definite comparison exists by factor analysis given that disease occurrence was not included in year 2; however, the composition of the factors was very similar in both years, indicating that variables clustered together in a similar fashion over years.

In general, groups defined by the average data set analysis were quite stable over years and/or dates, the only exception being those lines located near the border that separates groups.

Analysis of dates. In order to reveal which groups of lines of soybeans would be chosen to grow as a single (primary) crop or to be sown in double cropping (or as a second crop) the effect of planting date was evaluated. Principal components analyses for data averaged over years and replications, corresponding to two different planting dates, showed a similar pattern for each date as mentioned above. The mean performance for the productive traits of the six groups defined previously is included in Table 6. The better adapted higher yielding lines for use as a single crop belonged

Table 6. Mean performance of all groups for the productive traits in two planting dates averaged over years and replications.

Group	Date 1			Date 2		
	yield	% protein	% oil	yield	% protein	% oil
1	4016	39.9	23.9	4094	40.5	23.3
2	3656	40.2	23.8	4076	40.1	23.6
3	3255	40.8	23.4	3503	43.2	21.6
4	3866	39.0	24.7	4275	39.6	24.0
5	3020	39.0	24.6	3741	39.9	23.8
6	2284	37.4	26.4	2573	40.6	22.3

to group 1 and were in maturity group III. The lines in group 4 might also be promising, especially when sown as a second crop. Group 4 was considered to be superior to group 1 when planted in te second date due to a lower degree of lodging and small leaflets, possibly resulting in a more efficient use of solar energy as the daylength shortens in autumn. In the first date, there should be a greater amount of solar illumination and the plants could fulfill their needs with only the upper leaves photosynthesizing. Therefore, light penetration into the canopy may not be so important in producing a higher yield in the first date of planting. It may be a good agronomic practice to plant a high percentage of the second crop land with lines belonging to group 4 and the remaining with lines belonging to group 1 in order to program a gradual harvesting. In locations where late season weathering could be a problem, lines belonging to group 2 may also be chosen for planting as a second crop, instead of those of group 1 because their yield performance was essentially equal.

SUMMARY AND CONCLUSIONS

A collection of one hundred and twenty-five soybean lines, sown in southwestern Spain on two dates of planting in two consecutive years, was evaluated for 25 traits. Factor analysis, principal components analysis, and cluster analysis were used to investigate the diversity among lines and identify sets of varieties better adapted to the specific environmental conditions.

Results, using factor analysis and principal components, were similar. The most important factor contained traits related to life cycle, plant architecture, and seed yield. Important traits in factor II were lodging and other traits reflecting plant height, and the presence of diseases. Factor III was mainly related to chemical composition of the seed and also to seed yield. Several traits were redundant, meaning that in future studies they could be omitted, simplifying data collection. Cluster analysis identified six groups of lines, differing mainly in lateness, plant height, yield, % protein and % oil as well as in seed quality traits. The most promising groups were numbers 1 and 4, the former being favored for growing soybeans as a primary crop and containing lines belonging to maturity group III, while the latter was favored for double cropping. Group 3 contained some lines that could be used in a breeding program to improve protein content. Some varieties of group 5 showed a very high oil content. Groups

were quite stable over differing environmental conditions (years and dates). Discriminant analysis confirmed the composition of the groups, with minor changes, and reduced the number of significant variables to eleven, the most important being physiological maturity, protein content and green cotyledons.

REFERENCES

BHATT, G. M., 1976. An application of multivariate analysis to selection for quality characters in wheat. Austral. J. Agr. 27: 11–18.

BROICH, S. L. & R. G. PALMER, 1980. A cluster analysis of wild and domesticated soybean phenotypes. Euphytica 29: 23–32.

CARBONELL, E. A. & R. BARTUAL, 1983. Valoración agronómica y clasificación de una colección de lineas de soja sembrada en dos fechas en el bajo Guadalquivir. Comunicaciones INIA Ser. Prod. Veg. no. 57.

CHAPMAN, J. & F. P. C. BLAMEY, 1979. Phasic development of eight soybean cutivars in response to planting time in northern Natal. Agroplantae 11: 19–24.

CHATFIELD, C. & A. J. COLLINS, 1980. Introduction to multivariate analysis. Chapman & Hall. London. 246 pp.

DENIS,, J. C. & M. W. ADAMS, 1978. A factor analysis of the plant variables related to yield in dry beans. I. Morphological traits. Crop Sci. 18: 74–78.

FEASTER, C. V., 1942. Influence of planting date on yield and other characteristics of soybean grown in southeast Missouri. Agron. J. 41: 57–62.

GHADERI, A., M. SHISHEGAR & B. EHDAIE, 1979. Multivariate analysis of genetic diversity for yield and its components in mung bean. J. Amer. Soc. Hort. Sci. 104: 728–731.

GREEN, D. E., E. L. PINNELL, L. E. CAVANAH & L. F. WILLIAMS, 1965. Effect of planting date and maturity date on soybean seed quality. Agron. J. 57: 165–168.

HUSSAINI, S. J., M. M. GOODMAN & D. H. TIMOTHY, 1977. Multivariate analysis and the geographical distribution of the world collection of finger millet. Crop Sci. 17: 257–263.

JOSHI, M. G. & B. SINGH,1979. Genetic divergence among tetraploid *Triticum* species. Indian J. Genet. Plant Breed. 39: 188–193.

LAWLEY, D. N. & A. E. MAXWELL, 1971. Factor analysis as a statistical method. American Elsevier. New York.

MUNGOMERY, V. E., R. SHORTER & D. E. BYTH, 1974. Genotype × environment interactions and environmental adaptation. I. Pattern analysis. Application to soya bean populations. Aust. J. Agric. Res. 25: 59–72.

NARSINGHANI, V. G., K. S. KANWAL & S. P. SINGH, 1978. Genetic divergence in peas. Indian J. Genet. Plant Breed. 38: 375–379.

OSLER, R. D. & J. L. CARTTER, 1954. Effect of planting date on chemical composition and growth characteristics of soybeans. Agron. J. 46: 267–270.

SACHAN, K. S. & R. J. SHARMA, 1971. Multivariate analysis of genetic divergence in tomato. Indian J. Genet. Plant Breed. 31: 86–93.

SASMAL, B., 1978. An estimate of genetic divergence in jute. Indian Agric. 22: 143–150.

SINGH, R. B. & M. P. GUPTA, 1968. Multivariate analysis of divergence in upland cotton (*Gossypium hirsutum* L.). Indian J. Genet. Plant Breed. 28: 151–157.

SINGH, S. P., H. N. SINGH & J. N. RAI, 1980. Multivariate analysis in relation to breeding system in okra (*Abelmoschus esculentus* MOENCH). Zeitschrift für Pflanzenzüchtung. 84: 57–62.

SNEATH, P. H. A., 1976. Some applications of numerical taxonomy to plant breeding. Zeitschrift für Pflanzenzüchtung 76: 19–46.

WARD, J. H., 1963. Hierarchial grouping to optimize an objective function. J. Amer. Stat. Assoc. 58: 236–244.

WHITEHOUSE, R. N. H., 1971. Canonical analysis as an aid to plant breeding. In: R. A. NILAN (Ed.). Barley Genetics II. Washington State University Press.