Genotype × Environment Interactions and Environmental Adaptation. I Pattern Analysis — Application to Soya Bean Populations

V. E. Mungomery[A], *R. Shorter*[B] *and D. E. Byth*[C]

[A] Queensland Department of Primary Industries, Biloela, Qld. 4715.
[B] Department of Agronomy, Iowa State University, Ames, Ia, United States.
[C] Department of Agriculture, University of Queensland, St. Lucia, Qld. 4067.

Abstract

The application of numerical classificatory and ordination procedures to the analysis of adaptation of populations, cultivars, or lines is discussed. Examples given are seed yield and seed protein percentage for a population of soya bean lines evaluated across environments in south-eastern Queensland. For each line, *p* attributes (mean performance in each of *p* environments) were defined, and were considered to represent a *p*-dimensional space. The performance of the lines defined their position in this space, and the attributes were used to classify the lines into groups which differed in the magnitude and/or direction of their environmental responses. Within groups, lines exhibited variations across environments which were similar in both direction and magnitude. The relative proximity of lines in the space was obtained by ordination, and this reflected the degree of similarity of their performance across environments.

The actual environmental responses of a large number of lines may be examined by these methods, both individually and as members of groups of lines having similar responses. Unlike other analyses of adaptation, these methods require no prior assumption regarding the distribution and suitability of a particular environmental response. Consequently, the analysis is designed to allow hypothesis generation. If desired, a 'suitable' or 'ideal' response surface may be defined, but this occurs subsequent to, and may be independent of, the identification of the actual responses of the lines and groups of lines.

Introduction

Genotype × environment ($g \times e$) interactions act to confound comparisons among lines in most field situations. The importance of this has been appreciated for several decades, and testing across environments is common practice. Partitioning of variances associated with such interactions (Sprague and Federer 1951) allows more precise interpretation of experiments and predictions, but commonly results in the discard of a major portion of the variance among lines with respect to its use in selection. Restriction of the range of environments tested has been used to reduce the magnitude of $g \times e$ interactions (Horner and Frey 1957). However, the breeder is interested in comparing the responses of individual lines across environments, and such analyses are of little or no direct assistance in this area.

Comparison among lines of response across environments presents major problems, particularly in the large populations involved in most breeding programs. It is apparent that simplified procedures for comparison of responses are necessary, and it would appear that two methods may be feasible.

(*a*) The response of lines across environments may be reduced to a quantitative description, and that statistic (s) may then be analysed by conventional statistical methods.

(*b*) The effective number of comparisons necessary may be reduced by grouping lines which reveal similar responses across environments, and subsequently comparing these groups.

Several approaches to the first of these methods have been made (Yates and Cochran 1938; Plaisted and Peterson 1959; Finlay and Wilkinson 1963; Eberhart and Russell 1966; Baker 1969; Hanson 1970). However, such methods either involved basic *a priori* assumptions regarding the form of response shown by lines across environments, or simply compare the $g \times e$ interaction variance generated by individual lines. Both approaches may be suboptimal in that any response assumed may deviate substantially from the actual response of all or some lines, while estimation of the $g \times e$ interaction variance of a line allows comparisons of magnitude of environmental interactions but does not distinguish among lines on the form of the actual responses.

Numerical classificatory or pattern analysis methods have been used in botanical and other situations for many years, but to date have not been applied to the problems of varietal or line adaptation. Abou-El-Fittouh *et al.* (1969) used pattern analysis to classify locations used in cotton variety tests in the United States cotton belt, but did not apply the technique to the study of varietal adaptation. However, Perkins and Jinks (1968*b*) applied a simple binary grouping system to the interpretation of the non-linear portion of $g \times e$ interaction, depending on whether the correlations of the deviations were positive or negative. Such grouping reduced the non-linear portion of the $g \times e$ interaction, but significant deviations still existed. Perkins (1972) compared the results from a principal coordinate analysis of a group of *Nicotiana rustica* lines grown across environments with the results of adaptation analyses based on linear regression techniques. However, the principal component analysis was not shown to have general utility in the study of adaptation. Principal component analysis was also applied by Suzuki (1968) to the study of adaptability of white clover strains.

The purpose of this paper is to indicate the application of pattern analysis to the study of environmental adaptation of cultivars or lines within breeding populations. The techniques were used to examine the environmental responses of a population of soya bean lines for seed yield and seed protein percentage in south-eastern Queensland.

Pattern Analysis Methods

Pattern analysis involves the elucidation of a structure or pattern in a complex data set. Usually this has implied some form of classification and/or ordination. Classificatory methods found their initial application in the fields of ecology and numerical taxonomy (Goodall 1953; Williams and Lance 1958; Williams and Lambert 1960; Sokal and Sneath 1963). Webb *et al.* (1967*a*, 1967*b*) considered pattern analysis to be useful in elucidating the structure of tropical rain-forests, while Clifford and Goodall (1967) and Clifford *et al.* (1969) utilized numerical taxonomic procedures in the classification of the Poaceae.

Over the last 5 years, numerical analytical methods have been applied more widely in other disciplines. They have been used in the analysis of company information (Fisher *et al.* 1967) and in land surveys (Williams and Lance 1969). Variation

patterns in Australian introductions of *Glycine wightii* (Edye *et al*. 1970) and *Stylosanthes* spp. (Burt *et al*. 1971) have been obtained by numerical analysis of various morphological and agronomic characters. Classification and ordination procedures have been used in conjunction with classical statistical methods in the analysis of grazing experiments (Williams and Gillard 1971; Williams *et al*. 1971) and plant physiological studies (Gates *et al*. 1971).

For this study, the analysis involved two stages:

(i) a classificatory study to determine whether the population of lines grown across environments could be regarded as consisting of a number of partly dissociated subpopulations or groups; that is, whether a pattern existed among the lines in their response across the test environments;

(ii) an ordination study of that population to examine the relationships among individual lines in the population.

Classification

For each line, mean performance in each of the p environments was measured. It was considered that the p environments defined a p-dimensional space and this is analogous to the genotypic stability space of Hanson (1970). Each line tested was regarded as a point in this space, the coordinates of which were its performance in each of the p environments. The relative proximity of lines in this space reflected the degree of similarity of their performance across environments. The mean performances of each line in the p environments were thus regarded as p attributes, and were used to classify the lines into a number of groups. Each group contained lines which were relatively close to each other, but more distant from members of other groups in the p-dimensional space.

Williams (1971) described the major types of classificatory strategies and indicated that the hierarchical agglomerative and hierarchical divisive strategies had received considerable attention in the development of suitable computer programs. Williams (1967) emphasized that computer classifications were not unique; that is, the groups derived were not absolute entities. Consequently, small changes in a numerical model, or a change of program, could result in quite different classifications.

The data (attribute types) in the present case were quantitative (or numerical) (Lance and Williams 1967a), and therefore several of the existing strategies could have been used. A number of these were tested and the resulting classifications were examined in terms of their 'profitability' (Williams and Lance 1965) or 'utility' (Goodall 1966). A classification was regarded as useful if the derived groups of lines reflected varying environmental responses. Both weakly and intensely clustering systems gave useful results by this criterion. However, intensely clustering strategies tended to separate some lines which had similar responses, and also tended to create groups which were relatively heterogeneous with respect to the environmental responses of lines within groups. Williams (1971) indicated that a weakly clustering system is likely to be most realistic for those situations where no *a priori* knowledge of discontinuities in the population is available. From these considerations, the Canberra program CLASS, with group average sorting, was chosen. This program employs a hierarchical, agglomerative, polythetic strategy (Williams 1971) by which the most similar individuals (lines) and groups are successively fused until the complete population has been incorporated into a single group. The statistic used

in the decision function for the inter-individual distances was the unstandardized squared Euclidean distance. Thus fusion was between that pair of entities for which this statistic was a minimum. Details of the algorithms employed are given by Lance and Williams (1967*b*).

Ordination

It is impossible visually to depict the spatial distribution of lines in a space of more than three dimensions (environments). In order to display the relative positions of lines from a *p*-dimensional space in two or three dimensions, an ordination procedure described by Gower (1966, 1967) as "principal coordinate" analysis was used. By this method the matrix of inter-individual Euclidean distances generated in the previous classification procedure was efficiently reduced so that the information contained in the original matrix could be represented in a small number of orthogonal axes. The elements of each vector (axis) were regarded as the coordinates of the lines on that axis.

Table 1. **Origin and maturity of soya bean lines tested across four locations in each of two years**

Line no.	Name	Origin	Relative maturity
1–40	—	LS[A]	mid-very late (8–11)[B]
41	CPI[C] 15939 Avoyelles	Tanzania	late-mid (9)
42	CPI 15948 Hernon 49	Tanzania	late-mid (9)
43	CPI 17192 Mamloxi	Nigeria	very late (11)
44	Dorman	United States	early (5)
45	Hampton	United States	mid (8)
46	Hill	United States	early (5)
47	Jackson	United States	early-mid (7)
48	Leslie	United States	mid (8)
49	Semstar	Local cultivar	mid-late (8)
50	Wills	United States	mid (8)
51	CPI 26673	Morocco	very early (3)
52	CPI 26671	Morocco	very early (3)
53	Bragg	United States	mid (7)
54	Delmar	United States	early (4)
55	Lee	United States	early-mid (6)
56	Hood	United States	early-mid (6)
57	Ogden	United States	early-mid (6)
58	Wayne	United States	very early (3)

[A] LS, local selections from Mamloxi (CPI 17192) × Avoyelles (CPI 15939), prefixed K or X.

[B] Number in parenthesis is U.S. maturity group classification or estimated equivalent.

[C] CPI, Commonwealth Plant Introduction accession number.

Application to Yield and Protein Percentage in Soya Beans

Experimental Details

Fifty-eight soya bean lines were evaluated at four locations in south-eastern Queensland in each of two years, 1970 and 1971. The locations were sited within a 100 mile radius of Brisbane, and covered a wide range of climatic and edaphic

conditions. The trials at Redland Bay, Gatton, Brookstead and Nambour were coded as 5, 6, 7 and 8 respectively for 1970, and as 9, 10, 11 and 12 for 1971.

The lines consisted of 41 local selections, 12 accessions from the United States, and others from Tanzania, Nigeria, and Morocco (Table 1). A randomized complete block design with two replications was used in each environment. A number of agronomic and chemical traits were recorded for each plot, but only the results for seed yield (kg/ha) and seed protein percentage (moisture-free basis) are considered here.

Table 2. Pooled analysis of variance of seed yield (kg/ha) and seed protein percentage for 58 soya bean lines grown at four locations in each of two years

Source of variance	Degrees of freedom	Mean squares	
		10^{-3} × yield	Protein %
Years (Y)	1	43,773	1049·03
Locations (L)	3	17,068	343·06
Y × L	3	6,218**	151·32**
Replicates in Y and L	8	274	3·40
Lines (G)	57	3,687**	64·66**
G × Y	57	263	5·56**
G × L	171	762**	5·92**
G × L × Y	171	321**	2·28
Pooled error	456	160	1·87
Population mean		2,047	40·2
Coefficient of variation		19·5	3·4

*$P < 0.05$.   **$P < 0.01$.

Pooled Analysis of Variance

Mean squares for the pooled analysis across locations and years are shown in Table 2. Line and replicate effects were considered as fixed effects, and locations and years were assumed to be random. Expected mean squares and tests of significance followed Snedecor (1962). Significant differences existed among lines for both traits. The line × location interaction was highly significant for both characters, the line × year interaction for protein percentage, and the line × location × years interaction for yield, which indicated that differences existed among lines in their response to changes of environment.

Pattern Analysis

Classification.—The delimination of the population of 58 lines into subpopulations or groups can be represented by a dendrogram. The number of groups which may be considered varies from 58 to one, depending on the level at which the hierarchy is truncated. In this study the hierarchy was truncated arbitrarily at the 10 group level; that is, fusions of individuals or groups lower in the hierarchy were not considered. Dendrograms for yield and protein percentage are presented in Fig. 1. The values on the vertical axis of each dendrogram are the unstandardized squared Euclidean distances between pairs of groups, on a group average basis. The positions of the groups along the *x*-axis of the dendrogram have no meaning, since any two groups

Table 3. Identification of lines forming the groups delimited for seed yield in Fig. 1

Group	Entry no.	Name	Group	Entry no.	Name	Group	Entry no.	Name
I	45	Hampton	VII	16	K124	X	2	K12
	48	Leslie	(contd.)	17	K130		3	K39
				18	K131		4	K44
II	49	Semstar		19	K134			(Ross)
				20	K138		5	K53
III	47	Jackson		21	K143		6	K54
	50	Wills		29	K176		7	K69
	53	Bragg		30	K186			(Daintree)
	55	Lee		36	X_2B_{49}		9	K73
	56	Hood		43	Mamloxi		10	K74
	57	Ogden						(Gilbert)
			VIII	8	K70		24	K152
IV	44	Dorman					25	K157
	46	Hill	IX	1	K8		26	K162
				14	K85		27	K164
V	54	Delmar		15	K123		32	K197
				22	K147		33	X_2L_{44}
VI	51	CPI 26673		23	K148		38	X_2S_{27}
	52	CPI 26671		28	K171		39	X_2S_{29}
	58	Wayne		31	K195		41	Avoyelles
				34	X_2L_{93}		42	Hernon
VII	11	K76		35	X_2B_7			
	12	K77		37	X_2S_{21}			
	13	K79		40	X_2S_{31}			

Table 4. Identification of lines forming the groups delimited for seed protein percentage in Fig. 1

Group	Entry no.	Name	Group	Entry no.	Name	Group	Entry no.	Name
I	13	K79	V	11	K76	VII	10	K74
	23	K148		16	K124	(contd.)		(Gilbert)
	30	K186		21	K143		14	K85
	36	X_2B_{49}		31	K195		25	K157
				34	X_2L_{93}		28	K171
II	17	K130		35	X_2B_7		33	X_2L_{44}
	18	K131		38	X_2S_{27}		48	Leslie
	20	K138		39	X_2S_{29}			
	22	K147		41	Avoyelles	VIII	5	K53
	37	X_2S_{21}		43	Mamloxi		45	Hampton
							46	Hill
III	12	K77	VI	50	Wills		49	Semstar
	19	K134		55	Lee		54	Delmar
	32	K197						
	40	X_2S_{31}	VII	1	K8	IX	44	Dorman
	42	Hernon		3	K39		47	Jackson
				4	K44		52	CPI 26671
IV	2	K12			(Ross)		51	CPI 26673
	15	K123		6	K54		53	Bragg
	24	K152		7	K69		56	Hood
	26	K162			(Daintree)		57	Ogden
	27	K164		8	K70			
	29	K176		9	K73	X	58	Wayne

may be rotated about their point of fusion. The groups identified as I–X (Fig. 1) for seed yield have no necessary relationship with those for protein percentage; that is, the groups for the two characters are derived and identified independently. The composition of the groups for seed yield and protein percentage (Fig. 1) is shown in Tables 3 and 4 respectively.

The dendrograms suggested that for each character there was a major discontinuity in the population of lines for performance of individual lines over environments (Fig. 1). The mean performance for seed yield and protein percentage over all lines within each group is presented in Table 5. The major discontinuity in the classification based on yield (Fig. 1) suggested that mean yield (Table 5) was important in identifying
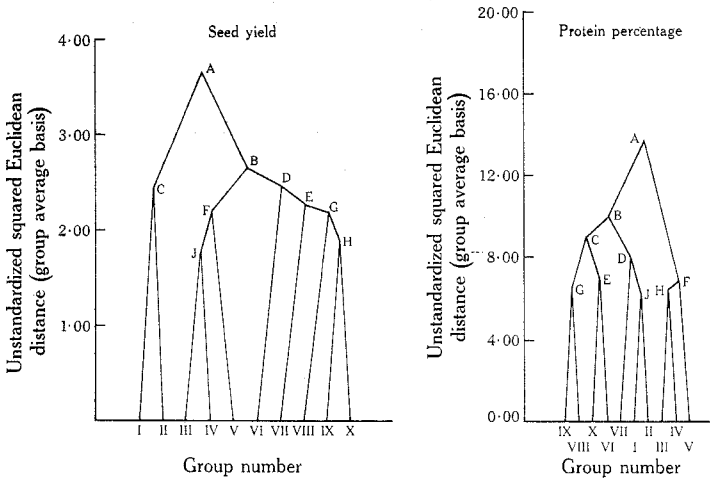
Fig. 1. Dendrograms for seed yield and seed protein percentage for 58 soya bean lines.

groups I and II. However, a second discontinuity in the yield dendrogram is not dependent solely on mean yield differences. The major discontinuity in the population for protein percentage (Fig. 1) apparently was unrelated to the mean performance of all lines within a group, averaged over environments (Table 5). However, groups I–III had higher mean protein percentages than groups IV–X.

The mean performance for each group in each environment is presented in Figs. 2 and 3 for yield and protein percentage respectively, in order to allow visual presentation of the responses of the groups across environments. The group means were obtained by using the Canberra program GROUPER (Lance et al. 1968). The environmental index on the x-axis was identical with that of Eberhart and Russell (1966); that is, the mean of all lines within each environment, expressed as a deviation from their mean over all environments. Groups which were closely related in classification (e.g. groups IX and X for yield and groups III and IV for protein percentage, Fig. 1) generally had similar responses over most environments and differed only in some environments (Figs. 2, 3). In contrast, relatively unrelated groups (Fig. 1) were more inconsistent in their responses (Figs. 2, 3) but exhibited both large differences in yield with similar responses (groups II and VI, Fig. 2) and quite different responses over most environments. These data indicated that the classificatory procedure used was effective in the delimitation of groups of lines which differed in either mean performance or environmental response or both.

The effectiveness of the classificatory procedure was confirmed by partitioning the mean squares for lines, lines × years, lines × locations and lines × locations × years

Table 5. Mean performance for seed yield and seed protein percentage of all lines within groups,[A] averaged over eight environments

	Seed yield (kg/ha)			Protein percentage
Group	I	2970	Group I	42·7
	II	3223	II	44·0
	III	2789	III	42·5
	IV	2269	IV	40·7
	V	2154	V	41·0
	VI	1451	VI	39·8
	VII	1505	VII	38·4
	VIII	1779	VIII	37·4
	IX	1873	IX	39·4
	X	2214	X	39·0

[A] Groups from Fig. 1, with line composition from Tables 3 and 4.

into the among- and within-group variances (Table 6). The group members were those obtained in the classifications for seed yield and protein percentage (Fig. 1; Tables 3, 4). Since the groups were derived by a strategy which minimized within-
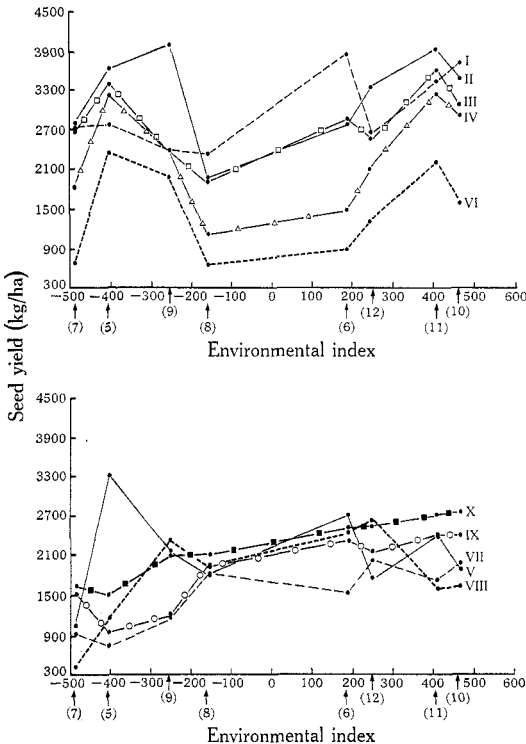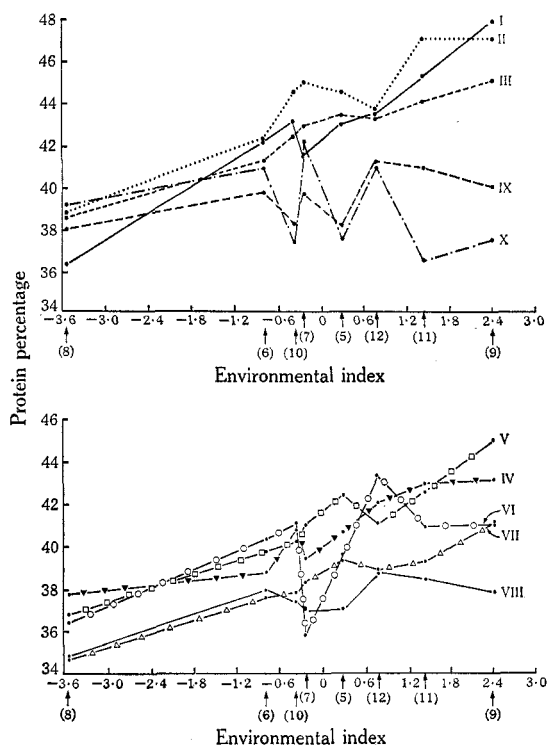

Fig. 2. Response of 10 groups of lines for seed yield across eight environments. Composition of groups is shown in Table 3.
(5)–(12), code numbers for trials (see text).

group variance, normal tests of statistical significance were not applicable. However, in all cases the within-group variances for the environmental interactions were small

in comparison with the corresponding among-group variances (Table 6). These data indicated that a similar form of response across environments existed for lines within groups, while the groups revealed substantially different responses to these environments. Similarly, the within-group mean square for lines was very small in relation to the among-group mean square for each character (Table 6). This indicated that the classification strategy delimited groups of lines which differed substantially in mean performance, and that the lines within groups had very similar means in general.

Fig. 3. Response of 10 groups of lines for seed protein percentage across eight environments. Composition of groups is shown in Table 4. (5)–(12), code numbers for trials (see text).

Ordination.—The first two vectors obtained by the ordination procedure for yield and protein percentage accounted for 78·3 and 82·6 % respectively of the information contained in the matrices of inter-individual distances. Thus a plot of vector 1 against vector 2 as Euclidean axes should provide a reasonable representation of the spatial arrangement of lines for both yield and protein percentage (Figs. 4 and 5 respectively) in the original eight-dimensional space. Group boundaries were drawn around those lines which were grouped together in the classification procedure.

Lines within some groups (e.g. group VI for yield, Fig. 4; groups I and III for protein percentage, Fig. 5) were in close proximity, whereas other groups were more diffuse. Two possible reasons for this were: (*a*) only the first two vectors were plotted in each case and these could not depict totally the scatter of lines in the eight-dimensional space; (*b*) numerous sharp discontinuities did not exist in the data for either character and the rather loose association of lines within some groups may have reflected this. It is possible that the identification of 10 groups in this case may have been excessive. Groups which were relatively unrelated with respect to classifi-

cation (e.g. yield groups I and VIII; protein groups II and VIII, Fig. 1) apparently occupied substantially different positions within the eight-dimensional space (Figs. 4 and 5 respectively). This further indicated that the classificatory procedure had been effective in the identification of groups of lines with different environmental responses.

Lines in seed yield groups I–V were predominantly of United States origin while those in yield groups VII–X were selected locally (Tables 1, 3). The ordination diagram (Fig. 4) suggested that these two broad subpopulations occupied substantially different positions in the eight-dimensional space, i.e. that they exhibited substantially different environmental responses. This is confirmed by the responses for groups across environments (Fig. 2).

Table 6. Partition of among- and within-group[A] variances for lines and for line × environment interactions

Source of variance		Degrees of freedom	Mean squares	
			$10^{-3} \times$ yield[B]	Protein %
Lines (G):	among groups	9	21,918	385·68
	within groups	48	268	4·47
G × year (Y):	among groups	9	751	24·05
	within groups	48	172	2·09
G × location (L):	among groups	27	3,578	22·55
	within groups	144	234	2·80
G × L × Y:	among groups	27	937	3·75
	within groups	144	205	2·00
Pooled G × E:	among groups	63	2,042	14·71
	within groups	336	213	2·36

[A] Groups from Fig. 1, with line composition from Tables 3 and 4.
[B] Kilograms per hectare.

Discussion

Both the regression approach (Finlay and Wilkinson 1963; Eberhart and Russell 1966; Perkins and Jinks 1968*a*) and the relative stability approach (Hanson 1970) to the study of varietal adaptation necessitate the assumption of some form of expected environmental response. In the regression method the response is assumed to be linear (or curvilinear if a log transformation is used), while in the relative stability method it may be linear, parabolic or any other response form specified. The utility of these methods would be affected critically by the adequacy of the definition of this expected response. For example, in soya beans evaluated in south-eastern Queensland, there is evidence that a major portion of the genotype × environment interaction is not linearly related to the environmental index for most agronomic and chemical traits (Mungomery, Shorter and Byth, unpublished data). If a plant breeder could define objectively the form of adaptation response desirable in his situation and equate it to expected stable performance, then the value of these methods would be increased substantially.

In contrast, the grouping and ordination procedures used herein examined the actual environmental responses of a large number of lines and groups of lines, and no *a priori* assumption of an 'expected' response was required. The results indicated

that such procedures were effective in delimiting groups of lines which differed in mean performance and/or their environmental responses and within which the
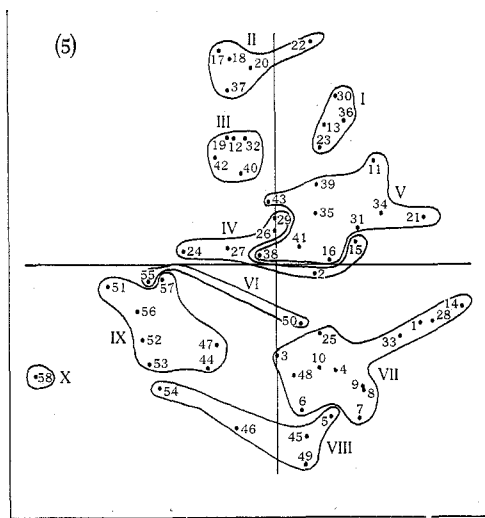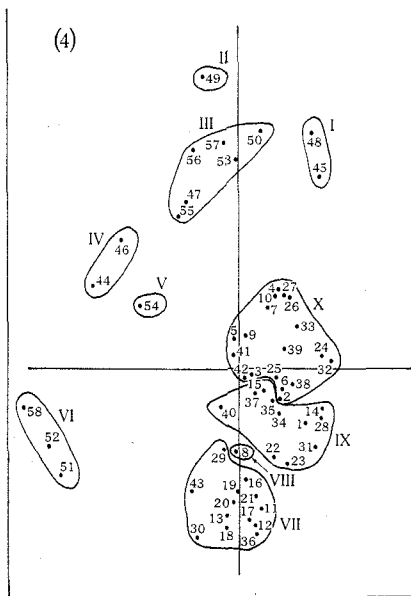

Fig. 4. Spatial arrangement of lines for seed yield in two dimensions obtained by ordination procedures.

Fig. 5. Spatial arrangement of lines for seed protein percentage in two dimensions obtained by ordination procedures.

individual lines had a relatively homogeneous response. The procedure may be used to augment considerably the traditional genotype × environment analysis of variance approach. The information gained has numerous applications, largely relating to the generation of hypotheses, in the interpretation of adaptation responses, and in practical breeding situations. Some of these applications are as follows.

(*a*) Similarities and differences in plant character development may be analysed in lines or groups of lines which reveal similar forms of response or which show contrasting performance in specific environments.

(*b*) The influence of parentage or origin of lines on the similarity of responses may be analysed.

(*c*) For breeding purposes, it may be possible to:

   (i) Identify groups of lines which reveal complementary actual responses across environments and which may have potential as parents;

   (ii) Predict the average adaptation response of progeny derived from parents with known environmental responses.

(*d*) For selection purposes:

   (i) A group or groups of lines may be identified which show responses across environments which are considered suitable for commercial production in specific regions of known environmental and/or production characteristics;

(ii) In some circumstances it may be relevant to define as 'suitable' or 'ideal' the response of a particular contemporary cultivar, in order that lines responding similarly to that cultivar (i.e. grouped with it) may be identified conveniently and accurately for more detailed study and selection. Such a selection criterion would not negate the identification of other groups of lines which show similar responses across environments but at a higher mean level, or which show more suitable forms of response. Consequently, greater convenience, flexibility and precision of selection should be made possible by using these procedures, particularly for large breeding populations.

Aspects of the application of this technique will be discussed in subsequent publications. It is emphasized that the definition of an 'expected', or a 'suitable' or 'ideal' response is unnecessary, and that actual responses are examined. If desired, any expected, suitable or ideal response may be defined and applied. However, that definition would occur subsequent to, and independently of, the identification of the actual environmental responses of lines and groups of lines. That group or those groups of lines which approached most closely to the defined response, or which appeared most 'suitable' for that region, could then be examined in greater detail.

The stability parameters obtained from the regression and relative stability approaches reflect the average or net response of a line over all environments. Two or more lines may have similar net responses, although their performance in certain individual environments may differ considerably for any particular trait. It would be informative for a plant breeder to be able to differentiate among such lines when a large number of lines was being examined. The numerical classificatory and ordination procedures used herein facilitate such a differentiation.

References

Abou-El-Fittouh, H. A., Rawlings, J. O., and Miller, P. A. (1969). Classification of environments to control genotype by environment interactions with an application to cotton. *Crop Sci.* **9**, 135–40.

Baker, R. J. (1969). Genotype–environment interactions in yield of wheat. *Can. J. Pl. Sci.* **49**, 743–51.

Burt, R. L., Edye, L. A., Williams, W. T., Grof, B., and Nicholson, C. H. L. (1971). Numerical analysis of variation patterns in the genus *Stylosanthes* as an aid to plant introduction and assessment. *Aust. J. agric. Res.* **22**, 737–57.

Clifford, H. T., and Goodall, D. W. (1967). A numerical contribution to the classification of the Poaceae. *Aust. J. Bot.* **15**, 499–519.

Clifford, H. T., Williams, W. T., and Lance, G. N. (1969). A further numerical contribution to the classification of the Poaceae. *Aust. J. Bot.* **17**, 119–31.

Eberhart, S. A., and Russell, W. A. (1966). Stability parameters for comparing varieties. *Crop Sci.* **6**, 36–40.

Edye, L. A., Williams, W. T., and Pritchard, A. J. (1970). A numerical analysis of variation patterns in Australian introductions of *Glycine wightii* (*G. javanica*). *Aust. J. agric. Res.* **21**, 57–69.

Finlay, K. W., and Wilkinson, G. N. (1963). The analysis of adaptation in a plant breeding programme. *Aust. J. agric. Res.* **14**, 742–54.

Fisher, R. G., Williams, W. T., and Lance, G. N. (1967). An application of techniques of numerical taxonomy to company information. *Econ. Rec.* **43**, 566–87.

Gates, C. T., Williams, W. T., and Court, R. D. (1971). Effect of droughting and chilling on maturation and chemical composition of Townsville stylo (*Stylosanthes humilis*). *Aust. J. agric. Res.* **22**, 369–81.

Gower, J. C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* **53**, 325–38.

Gower, J. C. (1967). Multivariate analysis and multi-dimensional geometry. *Statistician* **17**, 13–28.

Goodall, D. W. (1953). Objective methods for the classification of vegetation. I. The use of positive interspecific correlation. *Aust. J. Bot.* **1**, 39–63.

Goodall, D. W. (1966). Classification, probability and utility. *Nature. Lond.* **211**, 53–54.

Hanson, W. D. (1970). Genotypic stability. *Theor. appl. Genet.* **40**, 226–31.

Horner, T. W., and Frey, K. J. (1957). Methods for determining natural areas for oat varietal recommendations. *Agron. J.* **49**, 313–15.

Lance, G. N., Milne, P. W., and Williams, W. T. (1968). Mixed data classification programs. III. Diagnostic systems. *Aust. Comput. J.* **1**, 178.

Lance, G. N., and Williams, W. T. (1967*a*). Mixed data classificatory programs. I. Agglomerative systems. *Aust. Comput. J.* **1**, 15–20.

Lance, G. N., and Williams, W. T. (1967*b*). A general theory of classificatory sorting strategies. I. Hierarchical systems. *Comput. J.* **9**, 373–80.

Perkins, J. M. (1972). The principal component analysis of genotype–environmental interactions and physical measures of the environment. *Heredity, Lond.* **29**, 51–70.

Perkins, J. M., and Jinks, J. L. (1968*a*). Environmental and genotype–environmental components of variability. III. Multiple lines and crosses. *Heredity, Lond.* **23**, 339–56.

Perkins, J. M., and Jinks, J. L. (1968*b*). Environmental and genotype–environmental components of variability. IV. Non-linear interactions for multiple inbred lines. *Heredity, Lond.* **23**, 525–35.

Plaisted, R. L., and Peterson, L. C. (1959). A technique for evaluating the ability of selections to yield consistently in different locations or seasons. *Am. Potato J.* **36**, 381–5.

Sokal, R. R., and Sneath, P. H. A. (1963). 'Principles of Numerical Taxonomy.' (W. H. Freeman & Co.: San Francisco.)

Snedecor, G. W. (1962). 'Statistical Methods.' (Iowa St. Univ. Press; Ames.)

Sprague, G. F., and Federer, W. T. (1951). A comparison of variance components in corn yield trials. II. Error, year × variety, location × variety, and variety components. *Agron. J.* **43**, 535–41.

Suzuki, S. (1968). Studies on the evaluation of strain adaptability. Proc. 12th int. Congr. Genet., Tokyo, 1968, vol. 1, p. 270.

Webb, L. J., Tracey, J. G., Williams, W. T., and Lance, G. N. (1967*a*). Studies in the numerical analysis of complex rain-forest communities. I. A comparison of methods applicable to site/species data. *J. Ecol.* **55**, 171–91.

Webb, L. J., Tracey, J. G., Williams, W. T., and Lance, G. N. (1967*b*). Studies in the numerical analysis of complex rain-forest communities. II. The problem of species sampling. *J. Ecol.* **55**, 525–38.

Williams, W. T. (1967). The computer botanist. *Aust. J. Sci.* **29**, 266–71.

Williams, W. T. (1971). Principles of clustering. *A. Rev. Ecol. System.* **2**, 303–26.

Williams, W. T., and Gillard, P. (1971). Pattern analysis of a grazing experiment. *Aust. J. agric. Res.* **22**, 245–60.

Williams, W. T., Haydock, K. P., Edye, L. A., and Ritson, J. B. (1971). Analysis of a fertility trial with Droughtmaster cows. *Aust. J. agric. Res.* **22**, 979–91.

Williams, W. T., and Lambert, J. M. (1960). Multivariate methods in plant ecology. II. The use of an electronic digital computer for association analysis. *J. Ecol.* **48**, 689–710.

Williams, W. T., and Lance, G. N. (1958). Automatic subdivision of associated populations. *Nature, Lond.* **182**, 1755.

Williams, W. T., and Lance, G. N. (1965). Logic of computer-based intrinsic classifications. *Nature, Lond.* **207**, 159–61.

Williams, W. T., and Lance, G. N. (1969). Application of computer classification techniques to problems in land survey. *Bull. Inst. int. Statist.* **42**, 345–55.

Yates, F., and Cochran, W. G. (1938). The analysis of groups of experiments. *J. agric. Sci. Camb.* **28**, 556–80.