Letter to the Editor

Pattern Analysis in Epidemiological Evaluation of Cultivar Resistance

J. P. Thompson and R. G. Rees

Plant Pathology Branch, Department of Primary Industries, Queensland Wheat Research Institute, Toowoomba, Qld. 4350. Australia.
Accepted for publication 25 January 1979. 545-549.

In experiments (10,11) to evaluate responses of many wheat cultivars to epidemics of stem rust (caused by *Puccinia graminis* Pers. f. sp. *tritici* Erikss. & Henn.) and leaf rust (caused by *P. recondita* Rob. ex Desm. f. sp. *tritici* Erikss. & Henn.) several procedures were employed to analyze results. One procedure that proved to be particularly useful was pattern analysis (15). Pattern analysis comprises mathematical methods of classification and ordination which have been widely applied in taxonomy (12), ecology (8), and agricultural science (15) and which could be used to more advantage in plant pathology. In particular, our experience indicates the methods are suited to the analysis of large-scale screening experiments for slow rusting in wheat genotypes and should prove of value with similar epidemiological data for other plant diseases. This letter is written to draw attention to such possibilities by describing the operation of the method.

The field experiment that was selected to illustrate the application of pattern analysis was one of several conducted at Toowoomba, Australia, to assess slow rusting and tolerance among many wheat cultivars; detailed descriptions of the methods employed are given elsewhere (11). Briefly, 45 wheat cultivars mainly of Australian origins were tested in randomized blocks of small plots for severity (9) of rusting by *P. graminis tritici* strain 21-Anz-2,3,4,5,7 on six occasions after epidemic initiation. The logit transformation (13) was applied to the disease progress curves, with zero values replaced by 0.001 for transformation. The slope of the regression was taken as a measure of Van der Plank's apparent infection rate, *r* (13,14), and the intercept (*a*) was taken as a measure of the delay in the epidemic. The area (A) below the untransformed disease progress curve to the fifth assessment time also was calculated.

Pattern analysis as commonly employed (15) consists of the joint numerical classification and ordination of a set of "entities" on the basis of their "attributes"; eg, in a taxonomic study the entities are isolates or higher taxa and the attributes are their characteristics. Essentially both methods are tools for revealing similarities and differences among the entities; ie, the existing pattern. Numerical classification produces discrete groups of like entities such that similarities within groups are greater than between groups. Ordination does not of itself separate groups of entities but simply displays the relative geometric positions of the entities within a multidimensional space defined by the attributes. Success of ordination procedures depends upon the efficient reduction of this space to one of a few dimensions only.

For pattern analysis of this experiment the entities were the 45 cultivars and the attributes were the six occasions of rust assessment. Thus, the analyses were based on disease severities in a 45 × 6 data matrix. A polythetic, hierarchical classification of the cultivars was produced with the program MULCLAS (7) using the Canberra Metric Similarity Measure and the Group Average Fusion Strategy (6). All fusions in the classification were diagnosed with the program GROUPER (5). Principal co-ordinate analysis (4) of the similarity matrix was used to ordinate the cultivars, and the major vectors from this analysis were diagnosed with the program GOWECOR (5).

00031-949X/79/000095$03.00/0

The Canberra Metric, which is a dissimilarlity or distance measure, has certain arithmetic properties that are important in the present context. Its formula is:

$$\frac{1}{s}\sum_{k=1}^{s}\left|X_{ik}-X_{jk}\right| \;/\; (X_{ik}+X_{jk})$$

in which X_{ik} X_{jk} denote the respective values of the two entities i and j for the k'th attribute; and s represents the total number of attributes excluding those missing for either i or j.

The hierarchical classification of the cultivars (from MULCLAS) is given as a dendrogram truncated at the 20-group level (Fig. 1A). The sequence of fusions in the dendrogram shows which cultivars most resemble which other cultivars in rusting characteristics. From GROUPER analysis, the more resistant of two fusing groups was determined and the dendrogram is arranged such that the upper arm at each fusion is the more resistant. A complete dendrogram of *n* entities contains *n* − 1 groups and one can subdivide the population into any desired number of these groups. A useful degree of subdivision of the 45 cultivars is into the seven groups designated A to G in Fig. 1A.

The average rusting characteristics of groups A to G is best seen from Fig. 1B. Group G is extremely fast rusting from the onset of the epidemic. By day 29, six of the seven cultivars in group G virtually had been killed by rust and reliable assessments of percentage rusting could not be obtained. The analysis is, however, relatively insensitive to such missing attributes; therefore, in practice these cultivars were grouped on the basis of the other assessment times. Groups D, E, and F also are quite susceptible, but did not rust nearly as fast as group G. In comparison with these four groups, the rust epidemic was retarded considerably in groups A, B, and C, with groups A and B exhibiting a comparatively long delay. Thus, one would look for useful sources of resistances among the cultivars of groups A, B, and C, and possibly group D.

The results of the principal co-ordinate analysis are presented in Fig. 2 as an ordination of the cultivar-points on the plane of vectors I and II. These two vectors (or dimensions) accounted for 66% of the variance and thus provide an adequate reduction of the multidimensional space. The ordination is in general agreement with the hierarchical classification and the seven groups of cultivars delimited in Fig. 1A are compared in Fig. 2. The field reaction types of the cultivars also are indicated in Fig. 2 by symbols.

GOWECOR diagnosis showed vector I to be negatively correlated with rust levels at the last two assessments whereas vector II is positively correlated with rust levels at all assessments except the last. It is apparent, however, that the distribution of the cultivar points is not linear, but curved (Fig. 2). One may regard this distribution as a rust resistance spectrum which ranges from the highly resistant cultivar Gamut in the upper right-hand quadrant, around the arc through decreasing levels of partial resistance to finally end with the extremely susceptible cultivar Morocco. Attention is drawn to cultivar Mentana by its relative displacement off this arc. Unlike all cultivars except those in groups A and B, and cultivar Kenya Governor in group C, cultivar Mentana had no rust at assessment time 1. During later assessments, Mentana noticeably differed by producing relatively smaller pustules than those on other cultivars with similar high percentages of rusting. These effects probably result from the *Sr*8 and other resistances in Mentana. The ordination of the cultivars is in general agreement with their field reaction types but does draw attention to exceptions; eg, cultivar Bordan has far lower percentage rusting of the flag leaf sheath than have many other cultivars with a similar, susceptible, field reaction type.

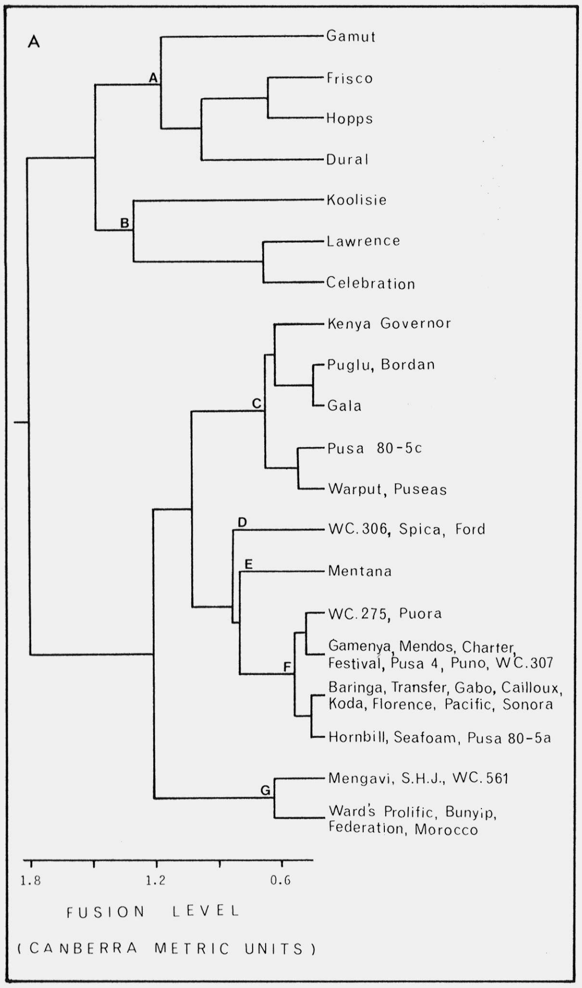
In Table 1, the cultivars are listed in decreasing order of rust resistance as determined from the hierarchical classification and GROUPER analysis, for comparison with parameters derived from the disease progress curves. The relative order of rust resistance deduced from the classification is seen to be quite acceptable as gauged by the other measures. The order is closely reflected by the area under the disease progress curve and by the logit-line intercept (*a*). It is not closely reflected by the apparent infection rate (*r*) although some local modification of the order is reflected by *r*, eg, the members of the classificatory group D would be placed higher on the basis of the logit line intercept, but infection rates and (to a lesser extent) areas under the curve, would rank them even lower if such parameters were considered independently. We have in fact examined a scatter diagram of the cultivar-points on Cartesian co-ordinates of logit-line intercept (*a*) and infection rate (*r*). It was possible to delimit as extreme groups the cultivars of groups A plus B and those of group G, but it was impossible to subdivide the other 31 cultivars into groups on this basis.

The experiments reported by Rees et al (10,11) were designed and conducted with the view to analyzing the data by methods outlined by Van der Plank (13,14) for selecting cultivars with low apparent infection rates. However, apparent infection rate (*r*) proved to be of limited interpretative value in these experiments. The logit-line intercept (*a*) seemed more useful, but uncertainty surrounded the application of linearized curve fitting procedures to these data. The area under the disease progress curve offered an alternative method, with the advantages of providing a single integrating measure of the duration and intensity of an epidemic for comparing cultivars, and without the assumption that the one form of equation is the appropriate response curve to fit to all cultivars. However, it is apparent that similar areas can lie under curves of quite different shapes and that undue emphasis may be placed on high values late in an epidemic.

The above considerations led us to examine the data from a different viewpoint by the use of pattern analysis. The objective in this application was, without assuming a definite model for a disease progress curve under these experimental conditions, to group those cultivars that most resemble one another in terms of the actual information available; ie, the severities of rusting at each of the various assessment times. By using numerical methods such grouping can be undertaken in a defined, repeatable manner.

The initial results of pattern analysis were intuitively appealing. The groups that were formed agreed well with our visual impressions of the responses of the cultivars to rust when sequentially examining them in the field and when collating and examining the data by methods such as area below disease progress curves and logit-line intercepts.

The patterns extracted and simply displayed by diagrams such as Fig. 1 and 2 are an obvious aid in considering the results. Thus



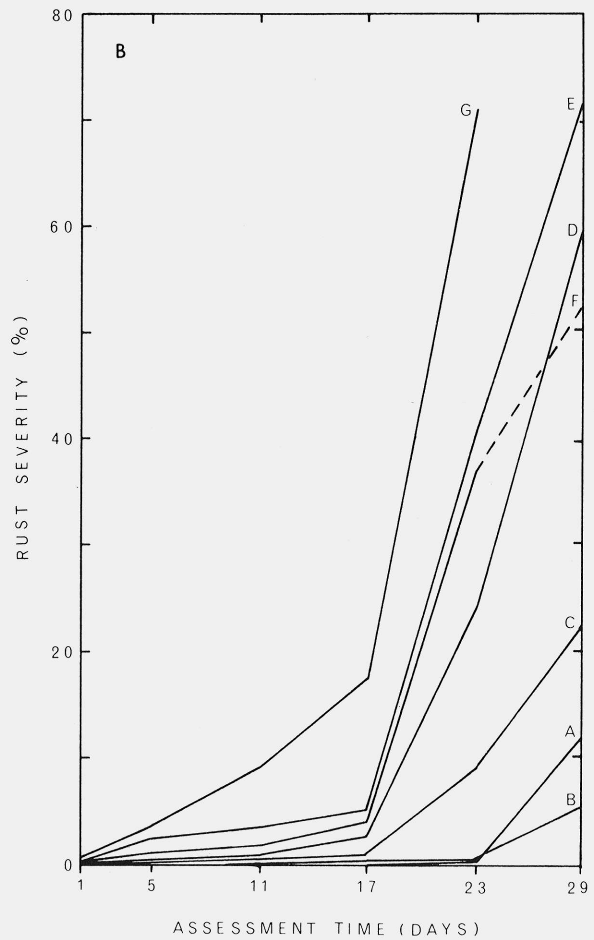


Fig. 1. A) Hierarchical classification of 45 wheat cultivars based on their responses to a field epidemic of stem rust (caused by *Puccinia graminis* f. sp. *tritici* strain 21-Anz-2,3,4,5,7) and analysis by Canberra Metric Similarity Measure and Group Average Fusion Strategy. **B)** Average progress of the stem rust epidemic for the seven major groups of wheat cultivars (groups A to G) formed by the hierarchical classification of the 45 cultivars. Missing values at day 29 for the cultivars in groups G and F (8 missing of 20) account for the incomplete curves of these groups. (Fig. 1B from Rees et al. Aust. J. Agric. Res. 30:403-419).

instead of thinking of individual cultivars one can initially think in terms of a small number of groups varying in resistance and then direct one's attention to finer comparisons of cultivars within the groups of most interest. Additional information not used for the analyses can be incorporated in the diagrams for further comparisons; eg, the field reaction types in Fig. 2. By using polythetic methods of classification the results are based on all the available information incorporated in the one similarity measure and the versatility of the methods, particularly with the diagnostic GROUPER program, allows one to examine the components of that measure whenever desired. We have now used pattern analysis on a number of sets of data comparable to the example given here. In each case interpretation has been facilitated by using pattern analysis as a complement to conventional procedures. This approach should find considerable application in the preliminary evaluation of rates of disease development on the very large numbers of genotypes tested in some plant breeding programs.

An epidemic is the progress of a disease in both time and space. In the experiment discussed above only temporal aspects of the disease were considered. We have also used pattern analysis in a pilot experiment to compare 30 cultivars which were assessed for rusting at various times both close to an initial point inoculation and at various distances from that point. In such instances the patterns extracted are a product of both temporal and spatial aspects of the epidemics. The main divisions into groups tended to be on the overall severity of rusting in both time and space, but some subgroups were delimited because increased rusting with time close to the point inoculation was not reflected in a comparable spatial spread of the disease.

Another combination investigated has been the responses of the one set of cultivars to more than one strain of the same rust species. Once again the main divisions were on overall severity of rusting, but cultivars that responded differentially to the rust strains were subdivided from those that responded similarly.

Several options are available in classification and ordination procedures. Kranz (2,3) drew attention to the use of classification in comparative epidemiology and described the classification of 40 different host-pathogen combinations. The attributes used in the classification were various parameters of bilateral progress curves for natural epidemics arising at various times throughout the year.

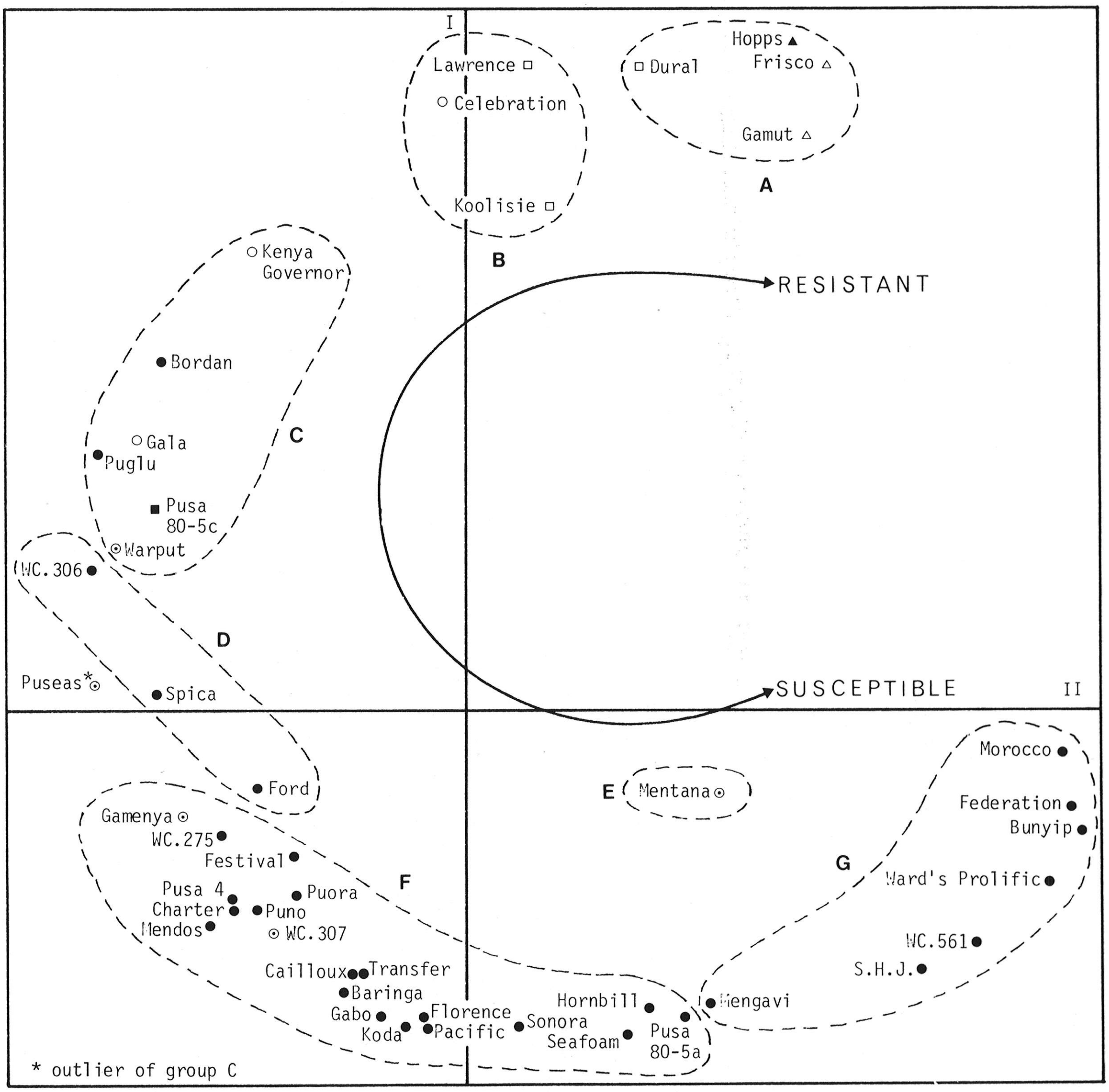


Fig. 2. Ordination of 45 wheat cultivars based on their responses to a field epidemic of stem rust (caused by *Puccinia graminis* f. sp. *tritici* strain 21-Anz-2,3,4,5,7). The ordination is a perpendicular projection of the cultivar points onto the plane of vectors I and II from a principal co-ordinate analysis. Vectors I and II are respectively negatively and positively correlated with rust severity. The seven groups of cultivars indicated are those delimited in the hierarchical classification. The field reaction type is indicated by the symbol marking the cultivar-point in the categories △ resistant, ▲ moderately resistant to resistant, □ moderately resistant, ■ moderately susceptible to moderately resistant, ○ moderately susceptible, ⊙ susceptible to moderately susceptible, ● susceptible (From Rees et al. Aust. J. Agric. Res. 30:403-419).

Our classificatory problem was of a different nature involving the one pathogen on 45 cultivars of the one annual host species, which allowed the determination of only partial unilateral curves. For the purpose of comparing cultivars it seemed preferable, and was certainly simpler, to classify and ordinate on the basis of attributes that were the actual rust assessments rather than derived parameters. This approach has provided results of useful interpretative value.

For similarity measure we have routinely employed the Canberra Metric. This metric depends solely on the pair of individuals or groups being compared and is unaffected by the values of the other members of the population. More important, it is sensitive to proportional rather than absolute differences; eg, the Canberra Metric difference between 1 and 10% rust is greater than that between 51 and 60% rust. Thus implicit in the Canberra Metric is a transformation whereby differences in rust severity between cultivars early in an epidemic are emphasized more than similar absolute differences later in an epidemic. This seems to us to be a biologically sensible transformation. It also means that classifications and ordinations based on the Canberra Metric emphasize differences among cultivars in the more resistant half of the rust resistance spectrum. Even though one is interested in the relative resistance of all cultivars under examination, most interest from the viewpoint of practical plant pathology lies with the more resistant cultivars and this emphasis is a desirable feature.

Standardized Euclidean Distance (1), which is a geometrically strict measure of distance, is another similarity measure that we have used for pattern analysis of rust epidemics. When applied to inverse sine transformations of the data for the experiment discussed in this letter, the patterns obtained were in general

TABLE 1. Wheat cultivars listed in decreasing order of resistance to a field epidemic of stem rust (caused by *Puccinia graminis tritici* strain 21-Anz-2,3,4,5,7) as determined from hierarchical classification and fusion by GROUPER analysis, for comparison of three measures from disease progress curves. Groups A to G are the same as those shown in Figs. 1 and 2. (From Rees et al. Aust. J. Agric. Res. 30:403-419)

Classificatory order of resistance		Apparent infection rate (r) (/day)	Logit-line intercept (a)	Area below curve (A)
Group	Cultivars			
A	Gamut	0.00	−11.5	0
	Frisco	0.07	−11.3	0.2
	Hopps	0.22	−11.7	0.6
	Dural	0.28	−12.4	2.3
B	Koolisie	0.12	−10.1	1.6
	Lawrence	0.20	− 9.3	4.8
	Celebration	0.28	− 9.9	11.6
C	Kenya Governor	0.36	− 9.8	39.3
	Puglu	0.27	− 8.2	50.1
	Bordan	0.27	− 8.6	31.3
	Gala	0.24	− 8.0	38.4
	Pusa 80-5c	0.16	− 6.8	25.7
	Warput	0.14	− 6.4	31.6
	Puseas	0.19	− 6.6	58.4
D	WC.306	0.31	− 8.5	52.0
	Spica	0.36	− 8.9	88.7
	Ford	0.38	− 8.8	145.1
E	Mentana	0.39	− 8.4	185.8
F	WC.275	0.31	− 8.1	86.6
	Puora	0.31	− 7.7	121.5
	Gamenya	0.25	− 7.2	86.6
	Mendos	0.28	− 7.6	93.4
	Charter	0.31	− 8.0	95.5
	Festival	0.16	− 5.7	69.2
	Pusa 4	0.24	− 7.0	80.3
	Puno	0.23	− 6.8	79.2
	WC.307	0.20	− 6.1	94.5
	Baringa	0.27	− 6.7	181.5
	Transfer	0.31	− 7.4	177.6
	Gabo	0.30	− 7.2	183.5
	Cailloux	0.36	− 8.2	163.5
	Koda	0.25	− 6.4	160.4
	Florence	0.34	− 7.5	219.5
	Pacific	0.29	− 6.7	208.3
	Sonora	0.29	− 6.7	230.3
	Hornbill	0.32	− 6.6	280.0
	Seafoam	0.31	− 6.9	232.5
	Pusa 80-5a	0.23	− 5.7	194.2
G	Mengavi	0.21	− 5.4	205.3
	S.H.J.	0.23	− 5.2	279.3
	WC.561	0.25	− 5.2	329.0
	Ward's Prolific	0.25	− 4.9	398.4
	Bunyip	0.26	− 4.8	456.4
	Federation	0.22	− 4.2	429.8
	Morocco	0.36	− 5.0	642.6
L.S.D. ($P < 0.05$)		0.127	2.37	72.36

agreement with those obtained with the Canberra Metric. The relative rankings of the cultivars were similar by both methods but with some notable exceptions. However, Euclidean Distance tends to be sensitive to outliers, and in both the hierarchical classification and the ordination there was finer subdivision among the more susceptible than among the more resistant cultivars. This relative emphasis on extremely susceptible cultivars was of little practical interest in the present context, therefore we have preferred to use the Canberra Metric. For other purposes the use of Standardized Euclidean Distance for classification and ordination of epidemiological data may be preferable.

We have indicated how pattern analysis can be used as an aid in interpreting experiments involving wheat rust epidemics. The methods should be readily applicable to similar epidemiological work with other diseases. In fact, Kranz (3) referring to classification, as achieved by pattern analysis, has stated, "Classification problems may one day attain great prominence in comparative epidemiology. As epidemiology deals with populations, it can be useful to delimit more objectively groupings of individuals, strains, races, varieties, treatments, reactions, which are more similar in epidemiological behaviour amongst themselves than compared with other groupings." We do not suggest that pattern analysis should supplant other methods for analyzing epidemiological data but that it does provide a valuable complement to other methods. Pattern analysis, by extracting and displaying the main patterns and trends in multivariate data often enables one to obtain new perspectives of the problems under consideration.

The programs used in this work are from the TAXON package of the CSIRONET computing network in Australia and we thank M. B. Dale for advice on their use.

LITERATURE CITED

1. BURR, E. J. 1968. Cluster sorting with mixed character types. I. Standardization of character values. Aust. Comput. J. 1:97-99.
2. KRANZ, J. 1974. Comparison of epidemics. Annu. Rev. Phytopathol. 12:355-374.
3. KRANZ, J. 1974. The role and scope of mathematical analysis and modeling in epidemiology. Pages 7-54 (Chapter 1) in J. Kranz, ed. Epidemics of Plant Diseases Mathematical Analysis and Modeling. Chapman and Hall, London. 170 pp.
4. GOWER, J. C. 1967. Multivariate analysis and multidimensional geometry. Statistician 17:13-28.
5. LANCE, G. N., P. W. MILNE, and W. T. WILLIAMS. 1968. Mixed-data classificatory programs. III. Diagnostic systems. Aust. Comput. J. 1:178-181.
6. LANCE, G. N., and W. T. WILLIAMS. 1967. A general theory of classificatory sorting strategies. 1. Hierarchical systems. Comput. J. 9:373-380.
7. LANCE, G. N., and W. T. WILLIAMS. 1967. Mixed-data classificatory programs. I. Agglomerative systems. Aust. Comput. J. 1:15-20.
8. ORLOCI, L. 1975. Multivariate analysis in vegetation research. Dr. W. Junk B. V. Publishers, The Hague. 276 pp.
9. PETERSON, R. F., A. B. CAMPBELL, and A. E. HANNAH. 1948. A diagrammatic scale for estimating rust intensity on leaves and stems of cereals. Can. J. Res. Sect. C. Bot. Sci. 26:496-500.
10. REES, R. G., J. P. THOMPSON, and E. A. GOWARD. 1979. Slow rusting and tolerance to rusts in wheat. II. The progress and effects of epidemics of Puccinia recondita tritici in selected wheat cultivars. Aust. J. Agric. Res. 30:421-432.
11. REES, R. G., J. P. THOMPSON, and R. J. MAYER. 1979. Slow rusting and tolerance to rusts in wheat. I. The progress and effects of epidemics of Puccinia graminis tritici in selected wheat cultivars. Aust. J. Agric. Res. 30:403-419.
12. SNEATH, P. H. A., and R. R. SOKAL. 1973. Numerical Taxonomy. The Principles and Practice of Numerical Classification. W. H. Freeman and Co., San Francisco. 573 pp.
13. VAN DER PLANK, J. E. 1963. Plant Diseases: Epidemics and Control. Academic Press, New York. 349 pp.
14. VAN DER PLANK, J. E. 1968. Disease Resistance in Plants. Academic Press, New York. 206 pp.
15. WILLIAMS, W. T. (ed.) 1976. Pattern Analysis in Agricultural Science. CSIRO: Melbourne and Elsevier, Amsterdam, Oxford and New York. 331 pp.