Biom. Praxim. (1990), 30, 1-13

NEW APPROACH OF SPATIAL POINT PATTERN AS A TOOL FOR DISTINGUISHING PLANT VIRUSES

R. LAHOZ-BELTRA*, A. PEREZ DE VARGAS*, J. LOPEZ-SANCHEZ*, J.R. DIAZ-RUIZ**

*Dto. Matematica Aplicada (Seccion Biomatematica). Facultad de Biologia. Universidad Complutence. 28040-MADRID, Spain.

**UEI Fitopatologia, CIB. CSIC. Velazquez 144. 28006-MADRID, Spain.

SUMMARY

A relationship between the viruses and the spatial point pattern of the necrotic lesions produced by them is studied. By a multivariate method we found that two populations of plants inoculated by two different tobamoviruses (TMV and PMMV-S) can be distinguished by the statistical analysis of the spatial point pattern of the necrotic lesions produced by them on the surface of vegetal leaves.

1. INTRODUCTION

Many plant viruses produce localized lesions on leaves of specific hosts, upon experimental inoculation. Although the production of necrotic local lesions has been very useful in the recognition and isolation of viruses and/or virus strains, little is known about the virus-host interaction factor(s) that may influence the distribution pattern of the lesion spots on the leaves.

In the present work we have developed a new approach of spatial point pattern as a tool for diagnosis of plant viruses. We have found that two populations of plants mechanically inoculated by two different tobamoviruses (TMV and PMMV-S) (ALONSO *et al.*, 1989) can be distinguished by the discriminant analysis of the spatial point pattern of the necrotic lesions, produced by them, on the surface of leaves (Fig. 1).

The disease definition is used here in an operational sense (WOODBURY and CLIVE, 1986), regarded as a primary lesion (necrotic spots) in response to a viral infection.

Our method consists of designing a receptor. The receptor is the element which allows us to extraction the information about the design pattern, mosaic, or general motive of distribution of a set of points irregularly distributed within a region of the space.

In the receptor the input data (observed frequency distribution of necrotic spots on the surface of vegetal leaf) are subjected to a battery of tests, involving a set of aggregation measures.

The output of the receptor can be represented as a p-dimensional vector. Once, this vector for each input data in the two groups of plants is obtained, a stepwise discriminant analysis is performed.

We find a good percent correct classification in each group of plants, when the results are validated by two methods: a cross validation method and employing new cases obtained in other experiment.

This indicates the efficiency of the receptor designed detecting the pattern, the relationship between disease and spatial point pattern, and the possibility of using the aggregation measures as a computer marker in pattern recognition when the pattern is a set of points distributed within a region of space.

2. MATERIAL AND METHODS

2.1. VIRUSES AND HOSTS USED

Two tobamoviruses, the vulgare strain of tobacco mosaic virus (TMV) and the spanish strain of pepper mild mottle virus (PMMV-S) (ALONSO *et al.*, 1989) were used. *Datura stramonium* L. plants were mechanically inoculated with purified viruses (GARCIA LUQUE *et al.*, 1989) in 0.02 M phosphate buffer pH 7.0 containing 10 mg/ml of Celite, by rubbing with a forefinger over the whole upper surface of well developed leaves (50 µl/leaf).

The virus concentrations were adjusted to give 30-50 necrotic local lessions by leaf. The plants were kept in a growth chamber at 25°C with a 16 h photoperiod and a light intensity of 11.000 lux.

After 3-5 days, when the small, necrotic, pin-point spots appeared, the leaves were detached, dried between layers of filter paper and the data on the number and spatial distribution of the spots collected as described below.

2.2. TECHNIQUE OF DATA COLLECTING

The necrotic spots were counted on the surface of leaves by using a sample lattice of 483 cells in area (21 row x 23 columns), each cell of 9 mm^2. The frequencies of the number of necrotic spots/cell were then calculated for each leaf.

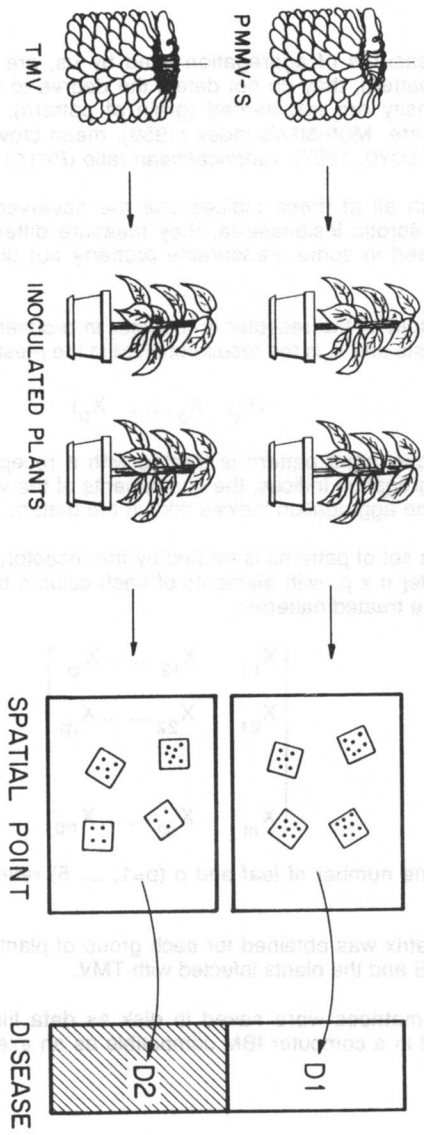

Figure 1: Diagramatic representation of the method used for the diagnosis of plant viruses (see text)

3. STATISTICAL MODEL

3.1. RECEPTOR

The measures of aggregation used by us, are all measures of the intensity of pattern; they do not detect the degree to which the patches of different density are fragmented (grain of pattern). These measures of aggregation are: MORISITA's index (1959), mean crowding (LLOYD, 1967), patchiness (LLOYD, 1967), variance/mean ratio (PIELOU, 1977) and W index.

Although all of these indices use the observed frequencies of the number of necrotic lesions/area, they measure different properties since they are based in some measurable property but different of the spatial pattern.

The output of the receptor is a vector in p-dimension, where p is the number of tests and x, is the result obtained in the i-test:

$$(X_1, X_2, \ldots, X_p)$$

If a spatial point pattern is treated with a receptor whose battery of tests are aggregation indices, the components of the vector will then be the values that the aggregation indices take in the pattern.

When a set of patterns is treated by the receptor, the result is a matrix A(n,p) of order n x p, with elements of each column being the aggregation indices for the treated patterns:

$$\begin{bmatrix} X_{11} & X_{12} \ldots \ldots X_{1p} \\ X_{21} & X_{22} \ldots \ldots X_{2p} \\ \\ X_{m1} & X_{n2} \ldots \ldots X_{np} \end{bmatrix}$$

where n is the number of leaf and p (p=1, ..., 5) refers to the aggregation index used.

This matrix was obtained for each group of plants, the plants infected with PMMV-S and the plants infected with TMV.

These matrices were saved in disk as data file. The receptor was implemented in a computer IBM compatible as an executable file (.EXE) in MS-DOS.

3.2. AGGREGATION INDICES

3.2.1. Morisita's index

Suppose a set of N points resides in a space that has been divided into cells of equal area, where xi (i=1, ..., Q) is the number of points in the ith cell. Then MORISITA's index (1959) is defined as:

$$I = \frac{Q\sum xi \quad (xi - 1)}{N(N - 1)}$$

Values of I greater than 1 indicate clumping, and those smaller than 1 indicate a hyperdispersion.

For a discussion and properties of this index see PIELOU (1977), SAKAI and ODEN (1983).

3.2.2. Mean crowding

LLOYD (1967) defines mean crowding as the mean number per point of other points (co-occupants) in a cell. If x, is the number of points in cell i, Q the number of cells of equal area and N the total number of points, then the mean crowding is defined as:

$$m = \frac{1}{N}\sum xi\,(xi - 1)$$

3.2.3. Index of patchiness

Since mean crowding is dependent on the degree of clumping and the points density, the index of patchiness is derived from the preceding index with the property that the points density effect is out of the measurement. LLOYD (1967) defined this index as:

$$c = \frac{m}{\lambda}$$

where m is the mean crowding and λ the mean number of points per cell.

3.2.4. Variance/Mean ratio

Under the null hypothesis of randomly distributed points, the variable x defined as "number of points per cell" has a Poisson distribution, with var(X)=E(X).

If we have patches of points then var(X)>E(X), whereas for more regularly spaced points we would expect var(X)<E(X). These properties suggests (PIELOU, 1977; PERRY and MEAD, 1979) the use of the variance/mean - ratio as an index to measure pattern:

$$\frac{var(X)}{E(X)}$$

with

$$var(x) = \Sigma \frac{(xi - \bar{x})^2}{Q - 1}$$

and

$$E(X) = \Sigma \frac{xi}{Q}$$

where xi is the number of points in the ith cell and Q the number of cells of equal area.

3.2.5. Index W

This index has been defined by us in this study as the product of the variance/mean ratio and (N-1), where N is the total number of points:

$$W = (N - 1)\frac{var(X)}{E(X)}$$

3.3. STEPWISE DISCRIMINANT ANALYSIS

The estimation of classification functions was carried out by the cross validation method using the BMDP7M program (DIXON, 1985).

The cross validation method allows us to subdivide randomly the cases in two separate groups. Once the two subsets are obtained, the cases of the first group (G1, training set) are then used to estimate the classification functions and the cases or the second group (G2, cross validation set) are classified according to the functions estimated with the first group (Fig. 2).

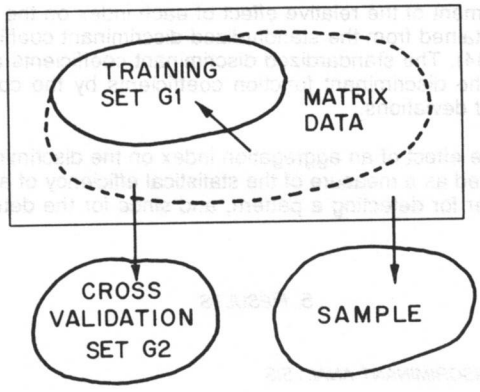

Figure 2: Estimation of classification functions and validation by two methods: (a) cross validation method and (b) employing a sample obtained in a second experiment.

The matrix data A(m,n,p) in the discriminant analysis is the cases (X_{1np}) belonging to the plants infected with PMMV-S plus the cases (X_{2np}) where the plants were infected with TMV:

$$\begin{bmatrix} X_{111} & X_{112} \cdots \cdots X_{11p} \\ X_{1n1} & X_{1n2} \cdots \cdots X_{1np} \\ X_{211} & X_{212} \cdots \cdots X_{21p} \\ \\ X_{2n1} & X_{2n2} \cdots \cdots X_{2np} \end{bmatrix}$$

where m (m=1, 2) is the group of plants infected with different virus (PMMV-S, TMV), n is the number of leaf and p (p=1, ... , 5) refers to the aggregation index.

New cases (sample size 36) obtained in other experiment with the same plants and viruses, were classified using the classification functions (Fig. 2). By observing the proportion of correct classification for the second group (G2) and the sample of new cases, we obtained the empirical and real measures for the success of the discrimination, respectively.

4. STATISTICAL EVALUATION OF THE AGGREGATION INDEXES AS COMPUTER MARKERS IN THE DETECTION PATTERN

An assessment of the relative effect of each index on the discriminant function was obtained from the standardized discriminant coefficients (AFIFI and CLARK, 1984). The standardized discriminant coefficients are obtained by multiplying the discriminant function coefficients by the corresponding pooled standard deviations.

The relative effect of an aggregation index on the discriminant function can be considered as a measure of the statistical efficiency of an index as a computer marker for detecting a pattern, and since for the detection of the disease.

5. RESULTS

5.1. STEPWISE DISCRIMINANT ANALYSIS

The classification functions estimated using the cases of the first group (G1) are:

$$F(\text{PMMV-S}) = 1.217\,(\text{MORIS}) + 0.075\,(\text{MECRO}) + 0.012\,(\text{W}) - 3.879$$
$$F(\text{TMV}) = 1.000\,(\text{MORIS}) + 5.447\,(\text{MECRO}) + 0.002\,(\text{W}) - 5.730$$

and the linear discriminant function is:

$$F = 0.216\,(\text{MORIS}) - 5.372\,(\text{MECRO}) + 0.009\,(\text{W}) - 1.851 = 0$$

where MORIS, MECRO, and W are the Morisita, mean crowding and W indices, respectively.

Applying the above classification functions to the cases belonging to groups G2 and sample of new cases we obtained a good correct percent classification in each group of viruses (Fig. 3).

5.2. CONCLUDING REMARKS

The results obtained allows us to conclude:

a) The designed receptor is efficient in the detection pattern.
b) The possibility of using the aggregation indexes as computer marker in pattern recognition.
c) The relation between diseases - spatial point pattern, and the possibility of using the spatial point pattern as an indicator in the diagnosis of diseases with symptoms characterized by the apparition of spots.

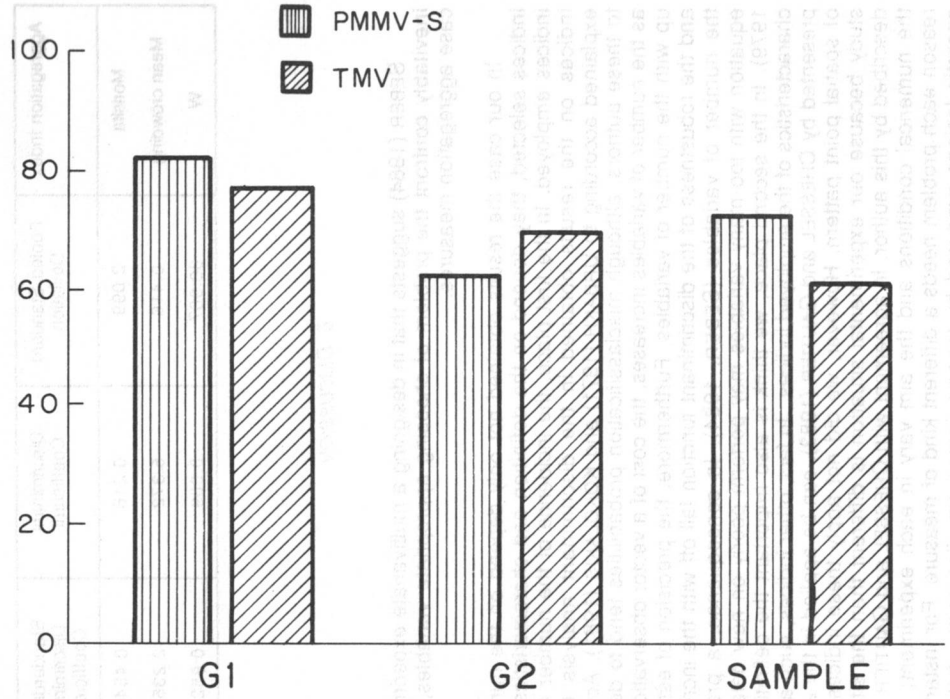

Figure 3: Percent correct classification obtained in each group of plant viruses.

Table 1 shows the standardized discriminant coefficients. The values obtained indicate that the mean crowding index has a larger effect on the discriminant function than W and Morisita indices.

Table 1: Relative effect of each aggregation index on the discriminant function obtained from the standardized discriminant coefficients.

Aggregation index	Pooled Standard Deviation	Discriminant Coefficient	Standardized Discriminant Coefficient
Morisita	2.099	0.216	0.454
Mean crowding	0.416	5.372	2.235
W	86.967	0.009	0.852

6. DISCUSSION

SEBER (1984) suggests that in designing a multivariate experiment we inevitably confront the problem of choosing appropriate variables, in this case aggregation measures.

In our case the results obtained not only depend on the number of indices selected, they depend on the definition and characteristics of the indices employed. In the first place, the incidence of the number of used indices on the results obtained in the discriminant analysis can be explained according to URBAKH (1971) and KOKOLAKIS (1981). According to these authors, although misclassification probabilities tend to decrease as the number of variables increases, the cost of a vector observation goes up with the number of variables. Furthermore, the precision of estimation and the robustness of the discriminant function fall off with the increase of the number of variables (SEBER, 1984). In consequence, a predictive equation with too many variables may perform poorly on new data (HILL, 1979). In the second place, we think is also important the definition and characteristics of the employed indices. In fact, other indices, such as those presented by CHESSEL and GAUTIER (1983), can be applied in the analysis of spatial point pattern. However, we did not apply these indices in this study because our experimental situation is different than the situations described by this author. In agreement with CHESSEL and GAUTIER (1983), the numerical conditions and the aim vary in each experiment; for this reason each problem needs a different kind of measure. For instance, the necrotic lesions produced by tobamoviruses are discernible if the virus concentration is adjusted in the laboratory to produce between 30 and 50 necrotic lesions by leaf. In this situation we did not use any index based on presence-absence sampling method, because the necrotic lesions are always discernible in such experimental conditions.

Another important fact is in the sample: the percentage for TMV is ± 60. A possible explication is that the model may be subject to prediction bias (NETER *et al.*, 1985). In this case, the predictive ability of the model for the data on which the model is based may be greater than the predictive ability for new data.

It is interesting to note that mean crowding and index of patchiness (LLOYD, 1967) have same structural relationship with MORISITA's (1959) index of dispersion under certain premises (PIELOU, 1977). The effect of such links on the results obtained in the discriminant analysis can be explained in agreement with COCHRAN (1964) in terms of the probability of misclassification. Any negative correlation between variables reduces the probability of misclassification, whereas a positive correlation increases this probability.

The results of the aggregation indices are not presented because our aim has been to study under a multivariate design the direction and degree to which each index contributes to the classification of a virus, given the spatial distribution of the symptoms. The relationship between disease and the spatial point pattern of the symptoms opens up the possibility of using the statistical analysis of spatial point pattern as a tool for diagnosis of plant viruses.

RESUME

UNE NOUVELLE METHODE D'ANALYSE DES MOTIFS SPATIAUX DE POINTS COMME OUTIL DE DETERMINATION DES VIRUS VEGETAUX

On étudie une relation entre la nature des virus et les motifs spatiaux des points de nécrose qu'ils provoquent sur les feuilles des plantes. On a notamment pu montrer que l'utilisation d'une méthode multivariée permettait de distinguer deux populations de plantes inoculées par deux types de virus du tabac (TMV et PMMV-S).

ACKNOWLEDGEMENTS

J.R. DIAZ-RUIZ research was supported by grants from CAICYT (PR84-0025), PLANICYT(AGR88-0082) and Fundacion Ramon Areces.

REFERENCES

AFIFI A.A., CLARK V. (1984). *Computer aided multivariate analysis*. Lifetime Learning Publication, California, 261 p.

ALONSO E., GARCIA-LUQUE I., AVILA-RINCON M.J., WICKE B., SERRA M.T., DIAZ-RUIZ J.R. (1989). A tobamovirus causing heavy losses in protected pepper crops in Spain. *J. Phytopathol.* 125, 67-76.

CHESSEL D., GAUTIER C. (1984). *Statistical pattern analysis of a plant population measured by geometric sampling on a limited space*. In: Knapp, O. (ed.). Sampling methods and taxon analysis in vegetation science. Dr. W. Junk Publishers, The Hague, 61-76.

COCHRAN W.G. (1964). On the performance of the linear discriminant function. *Technometrics* 6, 179-190.

DIXON W.J. (1985). *BMDP Statistical Software*. University of California Press, Berkeley, CA.

GARCIA-LUQUE I., SERRA M.T., ALONSO E., WICKE B., FERRERO M.L., DIAZ-RUIZ J.R. (1989). Characterization of an spanish strain of pepper mild mottle virus (PMMV-S) and its relationship to other tobamoviruses. *J. Phytopathol.* (in press.).

HILL M.A. (1979). Annotated computer output for regression analysis: a supplement to writeups for computer programs BMDP2R and BMDP9R. *BMDP Technical Report* n° 48, BMDP Statistical Software.

KOKOLAKIS G.E. (1981). On the expected probability of correct classification. *Biometrika* 68, 477-483.

LLOYD M. (1967). Mean crowding. *J. Anim. Ecol.* 36, 1-30.

MORISITA M. (1959). Measuring the dispersion of individuals and analysis of the distributional patterns. *Mem. Fac. Sci. Kyushu Univ. Ser.* 2, 215-235.

NETER J., WASSERMAN W., KUTNER M.H. (1985). *Applied linear statistical models*. Richar D. Irwin, Inc., Illinois, 437 p.

PERRY J.N., MEAD R. (1979). On the power of the index of dispersion test to detect spatial patterns. *Biometrics* 35, 613-622.

PIELOU E.C. (1977). *Mathematical Ecology*. Wiley, New York, 384 p.

SAKAI A.K., ODEN N.L. (1983). Spatial pattern of sex expression in silver maple (*Acer saccharinum* L.): Morisita's index and spatial autocorrelation. *Am. Nat.* 122, 489-508.

SEBER G.A.F. (1984). *Multivariate observations.* Wiley, New York, 507 p.

URBAKH V.Yu. (1971). Linear discriminant analysis: loss of disriminating power when a variate is omitted. *Biometrics* 27, 531-534.

WOODBURY M.A., CLIVE J.M. (1986). Continuous and discrete global models of disease. *Mathematical Modelling* 7, 1137-1154.