

Use of Categorical Information and Correspondence Analysis in Plant Disease Epidemiology

S. SAVARY

*ORSTOM Visiting Scientist at IRRI, PO Box 933, 1099 Manila,
Philippines*

L. V. MADDEN

*Department of Plant Pathology, Ohio Agricultural Research and
Development Center, The Ohio State University, Wooster, Ohio
44691-4096, USA*

J. C. ZADOKS

*Department of Phytopathology, Wageningen Agricultural Univer-
sity, PO Box 8025, 6700EE Wageningen, The Netherlands*

and H. W. KLEIN-GEBBINCK

*IRRI, Department of Plant Pathology, PO Box 933, 1099 Manila,
Philippines*

I. Introduction	214
II. Methodology	215
A. Qualitative and Quantitative Attributes of Pathosystems	215
B. Categorization of Quantitative Information	216
C. Contingency Table and Chi-square Tests	218
D. Correspondence Analysis: Procedure	220
III. Examples	224
A. Epidemic Trends in Groundnut Diseases	224
B. Components of Rice Tungro Epidemics	230
C. Relations Between Production Level and Yield Losses	233

Advances in Botanical Research Vol. 21
incorporating Advances in Plant Pathology
ISBN 0-12-005921-5

Copyright © 1995 Academic Press Limited
All rights of reproduction in any form reserved

ORSTOM Documentation



010001023

O.R.S.T.O.M. Fonds Documentaire

N° : 43564

Cote : B ex 1

IV. Overview	235
A. Combination and Link with Other Techniques	235
B. Usefulness of the Approach	236
C. The Need to Exploit Categorical Information in Plant Pathology	237
References	238

I. INTRODUCTION

It has long been recognized that a realistic approach to crop protection should consider the various pests, including diseases, that affect a crop (Padwick, 1956). Cropping practices represent major interactions within the disease tetrahedron (Zadoks and Schein, 1979) and, therefore, have to be considered when analyzing pathosystems. Systems analysis can provide an adequate set of concepts and tools to study the components of pathosystems and their interactions (Teng and Bowen, 1985). Since the number of components is large and their interactions are complex, a rationale is often needed to delineate the limits of the system to be addressed. A survey may provide the necessary overview of the pathosystem; adequate methods for analyzing survey data can produce preliminary information on its behaviour including major interactions. In this context, surveys can be considered as part of a systems approach.

Epidemiologists are frequently confronted with large data sets representing information on the characterization, the dynamics, or the behavior of a pathosystem. Examples are a survey data set, including information on disease intensity, cropping practices, and environmental data, and a germplasm database, where cultivars are represented by reaction types, disease intensities in field experiments, and quantitative measurements on the disease cycle. A range of methods has been developed, or adapted, by plant pathologists to analyze such data sets. The objectives of the analyses are as diverse as the methods, the first being to compact the information in such a way that interpretations can be made, hypotheses presented, and new experiments conducted.

A holistic description of a pathosystem should ideally involve characteristics as different as disease intensity over time and space, crop management, soil type and weather data, and socioeconomic information. Appropriate methods are needed by which the various facets of the available information can be explored. This paper presents an approach to mobilize and exploit information that is diverse in nature, precision, and accuracy.

We shall attempt to minimize the use of technical jargon. The reader needs only to know a few terms (Porkess, 1988): *class* (a collection of individuals, e.g. plots, cultivars, or sites, that share a common characteristic); *categorization* (the process of grouping individuals into classes according to numerical

boundaries); *categorical analysis* (an analysis that addresses categories and/or qualitative data); *contingency table* (a table that shows a bivariate frequency distribution, using quantitative or qualitative classification). The term *correspondence analysis* will be introduced later in the text.

This paper is a practitioner's view of a set of statistical tools applied to phytopathology. Because these tools allow one to handle information in a fresh and encompassing way, their use by plant pathologists can offer new avenues for research. We have chosen to emphasize a particular, relatively simple analytical strategy, which is described in a hypothetical example. This strategy is then applied to a set of three very different, actual examples. These examples are addressed with dissimilar perspectives, and this may require some inflection in the details of the methodology. However, the set covers such a large range of issues that we feel this demonstrates the general value of the methodology. In a final section, an overview is offered where its application domain, and the perspectives of combinations with other techniques, are briefly discussed.

II. METHODOLOGY

A. QUALITATIVE AND QUANTITATIVE ATTRIBUTES OF PATHOSYSTEMS

Plant pathologists have been concerned primarily with quantifying epidemiological characteristics of pathosystems (Zadoks, 1978). This effort is illustrated by the attention paid to the quality of disease measurement (Large, 1966; James, 1974; Daamen, 1986a, b; Kranz, 1988; Campbell and Madden, 1990). The ideal disease assessment method should be both precise and accurate (Nutter *et al.*, 1991), as well as reproducible and unbiased. However, many characteristics of pathosystems, such as cultivars and the physiological races of the pathogen, are qualitative in nature. Other examples are the cropping season, the previous crop, or the soil type. Methods are therefore needed to consider these qualitative and quantitative attributes of pathosystems simultaneously.

A hypothetical germplasm database may, for instance, include the following information: field assessments using a 5-point grading scale based on disease severity and symptom pattern (IRRI, 1988, modified), disease reaction types (Chester, 1946; Zadoks and Schein, 1979; Savary *et al.*, 1989), with four categories, and quantitative measurements of the infection efficiency, the latent period, and the sporulation intensity (Table I). A plant pathologist may wish to analyze the relation of semiquantitative field assessments with quantitative measurements of monocyclic processes, and then check how the various categories of reaction types correspond to the monocyclic components and epidemic levels.

TABLE I
A hypothetical database on disease reaction of a germplasm collection: list of variables

Symbol	Variable ^a	Measurement	Unit
F	Field reaction	Rating given to a variety using a combination of quantitative and qualitative factors, in field experiments, using a grading scale, e.g.: 0: no disease 1: less than 1% severity (apical lesions) 2: 1-5% severity (apical and some marginal lesions) 3: 6-25% severity (apical and marginal lesions) 4: 51-100% severity (apical and marginal lesions)	None
I	Infection efficiency	Average proportion of deposited spores that produce lesions; in glasshouse experiments	Lesion/spore
L	Latent period	Average delay from inoculation to first sporulation in a population of lesions; in glasshouse experiments	Hours
S	Sporulation intensity	Average spore production in a population of lesions; in glasshouse experiments	Spore/lesion
R	Reaction type	Ranking of the varieties using a typological scale consisting of standardized classes characterizing a given level of host-pathogen compatibility, e.g.: - highly susceptible (HS) - susceptible (S) - resistant (R) - highly resistant (HR)	None

^a In the example, data are assumed to be available for 78 cultivars.

B. CATEGORIZATION OF QUANTITATIVE INFORMATION

Statistical methods applicable to large, complex data sets usually imply a set of conditions, such as a linear relationship between the dependent and the independent variables for multiple linear regression analysis, and a series of prerequisites, among which is the homoscedasticity of the error term of the regression model (Butt and Royle, 1974; Teng and Gaunt, 1980; Campbell and Madden, 1990). Disease or pest variables usually do not comply with these prerequisites (McCool *et al.*, 1986), and data transformation is often needed.

Another reason for transforming pest data is to represent mechanisms of damage they induce in the crop (Teng and Gaunt, 1980). Although appropriate transformation of data may, or may not (McCool *et al.*, 1986), achieve these objectives, the result is often a reduction of the overall clarity of the model (Neter and Wassermann, 1974). Relatively complex transformations may, however, be considered in the case of simple systems, where a few independent variables are involved (Madden *et al.*, 1981; Savary and Zadoks, 1992a).

Methods that allow the simultaneous handling of the two types of attributes — quantitative and qualitative — and that do not imply *a priori* assumptions on the variables, such as a linear relationship, are thus desirable. One way is to make quantitative variables compatible with qualitative variables, and encode them into classes, i.e. define quantitative boundaries of classes, and encode the values of the quantitative variables according to these boundaries. This encoding process allows the investigator: (a) to define the boundaries such that they represent the (maximum possible) error made in the measurement of each variable (variables with low accuracy would be represented by a few, broad classes, while variables with high accuracy would be represented by a larger number of classes); and (b) to link the definition of classes with key-values, thresholds, or any information that might be available beforehand. The process of converting quantitative data into coded data is flexible, different options being available depending on the variable at hand. There is no statistical restriction for this process. The further analysis of the resulting coded data by means of contingency tables and chi-square tests, however, depends on the class-filling, and therefore on the number of classes relative to the size of the sample.

In the above-mentioned example of a germplasm database, five grades (0 to 4) of disease intensity in the field are considered (Table I). When considering this variable, the analyst may take into consideration the facts that: (a) the assessments pertain to conventional designs, i.e., contiguous plots with a few border rows, resulting in interplot interferences; (b) small plots may not be representative of full-scale commercial fields; and (c) declaring complete absence of disease might have required a detailed inspection of every plant in each plot, a procedure incompatible with large varietal trials. These facts may have affected the disease assessment, especially at lower intensities. One may decide to merge grades 0 and 1 into one class, i.e., class 1, or "very low disease" and to consider three other classes, i.e. class 2 ("low"), class 3 ("medium"), and class 4 ("high disease"), representing the previous grades 2, 3, and 4, respectively. Classes similar in size are desirable. As the database involves 78 cultivars, it is expected that disease intensity in the field, when evenly distributed, would be represented by four classes, each containing approximately 20 cultivars. In this example, they contain 18, 21, 21, and 18 cultivars, respectively.

Infection efficiency (Table I) can be categorized in three classes: low ($I \leq 0.1$; I1), medium ($0.1 < I \leq 0.2$; I2) and high ($I > 0.2$ lesion/spore; I3).

TABLE II
A hypothetical database on a disease reaction of a germplasm collection: encoding of data^a

Var.	Initial data					Coded data				
	L	S	I	F	R	L	S	I	F	R
1	6.8	3500	0.17	4	S	1	2	2	4	3
2	6.9	4200	0.08	3	R	1	2	1	3	2
3	6.8	1900	0.05	2	HR	1	1	1	2	1
4	6.9	1500	0.05	1	HR	1	1	1	1	1
5	7.1	3200	0.21	4	S	2	2	3	4	3
6	9.9	5100	0.22	3	VS	3	3	3	3	4
7	9.3	1400	0.21	2	S	3	1	3	2	3
.
.
.
78	9.5	2200	0.04	1	R	3	2	1	1	2

^a For abbreviations, see Table I.

The three classes are defined in such a way that they are represented by similar frequencies: 24, 27, and 27 cultivars in class 1, 2, and 3, respectively. Three classes are also defined for the latent period (L) and the sporulation intensity (S) using 7 and 9 days as boundary values for L, and 2000 and 5000 spores/lesion as boundary values for S.

Reaction type, being a typical qualitative variable (it is categorical, and coded), here with four classes, is not categorized, the successive levels of resistance (HR to VS) being simply coded from 1 to 4.

The process of categorization and encoding is illustrated in Table II. Its main outcome is to translate information that was diverse in its format (figures, and numerical or alphanumeric codes) into a standardized, coded format.

C. CONTINGENCY TABLE AND CHI-SQUARE TESTS

Building contingency tables is an easy way to explore the relations between paired qualitative or coded quantitative variables of a data set represented by their successive classes. The contingency tables show bivariate frequency distributions. A chi-square test can be applied to confirm the suggested pattern of the data. The null hypothesis is the independence of the distribution frequencies of the two variables. The validity of the chi-square test relies on the expected sizes of groups; it should generally only be applied to groups with a minimum size. This minimum size was fixed at 5 by Dagnélie (1973). Gibbons (1976) gave a more flexible rule: no more than 20% of the expected

TABLE III

A hypothetical database on a disease reaction of a germplasm collection: a contingency table^a

Latent period	Field reaction ^b			
	F1	F2	F3	F4
L1	3	2	7	15
L2	6	7	11	3
L3	9	12	3	0

^a Numbers represent cultivars with a particular latent period and field reaction, chi-square value: 35.8; d.f. = 6; $P < 0.0001$.

^b See Table I.

values (assuming independence) should be smaller than 5. In practice, major disequilibrium among the classes should be avoided, all classes being represented by commensurate numbers of individuals. This guideline is the only one that is essential to the next step, the analysis of a series of contingency tables by means of correspondence analysis. The way the analysis should be pursued is essentially dependent on the design of a strategy that adequately addresses the problem at hand.

In the germplasm database example, an important result of the categorization and coding of the data is a limited and consistent number (3 to 4) of classes for each variable, with classes being represented by approximately balanced numbers of cultivars. Simple contingency tables can, therefore, be assembled to analyse the relationships among variables, as, for instance, between disease intensity in the field (F) and latent period (L). The [F × L] contingency table (Table III) is a 4 × 3 matrix that shows that cultivars with very low disease (F1) usually have long latent periods (the profile is: three cultivars with short, six with medium, and nine with long latent periods), whereas cultivars with high disease (F4) usually have short latent periods (15 cultivars with short, three with medium, and none with long latent period). Profiles of relationships can also be examined row-wise: short latent periods (L1) are predominantly associated with high disease (three cultivars with very low, two with low, seven with high, and 15 with very high disease). The distribution frequencies of the two categorized attributes, F and L, are not independent as can be tested with a chi-square test, where the independence hypothesis is rejected ($P < 0.0001$).

Contingency tables can be assembled in the same way for all the variables of the database. Questions to be addressed include: (a) what is the relation between disease intensity in the field (F) and the monocyclic processes (I, L, and S); and (b) how well is R represented in terms of both disease intensity and monocyclic processes? The following strategy may be followed:

1. contingency tables are assembled using F classes as columns; in other words, disease intensity would be used as a "guide" throughout the analysis;
2. the three tables, [F × I], [F × L], and [F × S] are put together so that the table is of dimension 4 × 9; then
3. the relationships between these two types of variables, disease intensity and monocyclic components, are explored using correspondence analysis; finally,
4. the associations between F, I, L, S, and reaction types (R), are analysed using the contingency table [F × R] in the framework developed in the previous step.

D. CORRESPONDENCE ANALYSIS: PROCEDURE

Correspondence analysis (Benzécri, 1973; Hill, 1974; Greenacre, 1984) is a multivariate statistical method to represent contingency tables in pictorial and tabular form. The data handled in the analysis are classes of coded variables, as, for example, the columns and rows of the contingency tables shown in Table IV. Each class is represented by its profile, either by rows (e.g. I1 is represented by its profile in terms of disease intensities: 12, 9, 3, 0) or by columns (F1 is represented by its profile in terms of monocyclic processes: 12,

TABLE IV
A hypothetical database on a disease reaction of a germplasm collection: contingency tables arranged for the two steps of a corresponding analysis^a

Variable ^b	Field reaction			
	F1	F2	F3	F4
I1	12	9	3	0
I2	6	9	10	2
I3	0	3	8	16
L1	3	2	7	15
L2	6	7	11	3
L3	9	12	3	0
S1	12	9	3	0
S2	4	6	8	9
S3	2	6	10	9
<i>R1</i>	12	7	1	0
<i>R2</i>	3	0	11	0
<i>R3</i>	3	11	3	9
<i>R4</i>	0	3	6	9

^a See Tables I and III.

^b Variables in bold and italics are used in first and second steps of the analysis, respectively (see text for further explanation).

6, 0; 3, 6, 9; and 12, 4, 2, for I, L, and S, respectively). According to the strategy in the example, the vertical profiles of disease intensities in terms of reaction types are not yet considered; the upper part of Table IV, a 4×9 matrix, is first analysed. The associations of several variables belonging to two types (disease intensity in the field F, i.e., the result of a polycyclic process, and monocyclic attributes of the pathosystem: I, L, and S) are examined simultaneously. The procedure is similar to principal component analysis, and involves the computation of eigenvalues and eigenvectors (Benzécri, 1973; Greenacre, 1984). The sum of the eigenvalues is called the *inertia* and, with correspondence analysis, equals the chi-square statistic divided by the total number of observations. Each class contributes a fraction to the total inertia; summation of inertia over classes produces the total inertia. Coordinates for new axes are defined based on the eigenvalues, unlike principal component analysis, where the entries of the data matrix are quantitative, correspondence analysis entails computations on a data matrix of frequencies. Another difference is that classes in the columns and the rows are involved in the same way: each eigenvector is made up of weighted combinations of all the classes that have been selected for the analysis. The reason for this is that correspondence analysis is based on a chi-square distance d between classes, that can be written as (Dervin, 1988):

$$d(i, i') = \left(\sum_{j=1}^p (X_{ij}/X_{i.} - X_{i'.j}/X_{i'.})^2 / X_{.j} \right)^{1/2}$$

where: i and i' are two rows (or columns),

X_{ij} is the frequency value in the i th row and j th column (e.g.

$X_{23} = 10$),

$X_{i.}$ is the sum of the frequencies along row (or column) i (e.g.

$X_{2.} = 27$),

$X_{.j}$ is the sum of the frequencies along column (or row) j .

This distance definition, which can be applied to rows or columns, differs from the Euclidian distance used in principal component analysis:

$$d(i, i') = \left(\sum_{j=1}^p (X_{ij} - X_{i'.j})^2 \right)^{1/2}$$

where: i and i' are two rows (or columns, depending on the type of ordination used), representing two individuals (e.g., fields, cultivars, or epidemics, *not frequencies*),

j is the index for variables (quantitative descriptors),

X_{ij} is the measure of variable j representing the i th individual.

Details on the mathematics of correspondence analysis can be found in Benzécri (1973), Greenacre (1984), and Dervin, (1988).

Table V shows the result of correspondence analysis applied to the matrix

TABLE V
Correspondence analysis: relative weights and contributions to axes

	Classes	Relative Weight	Contribution to axes					
			Axis 1			Axis 2		
			Contribution		Co-ordinate	Contribution		Co-ordinate
			To axis	Reciprocal		To axis	Reciprocal	
Columns	F1	0.231	28.7	89.7	-0.69	20.5	7.9	-0.21
	F2	0.269	13.7	88.6	-0.44	1.0	0.8	+0.04
	F3	0.269	3.4	31.7	+0.22	56.6	64.9	+0.32
	F4	0.231	54.2	95.2	+0.95	21.9	4.7	-0.21
Rows	L1	0.115	17.6	89.0	+0.77	15.9	9.9	-0.26
	L2	0.115	0.4	10.4	-0.12	25.3	78.0	+0.32
	L3	0.103	14.2	89.5	-0.73	1.2	1.0	-0.08
	S1	0.103	16.2	91.8	-0.78	10.3	7.2	-0.22
	S2	0.115	2.6	99.4	+0.29	0.0	0.1	+0.01
	S3	0.115	4.8	81.3	+0.40	8.4	17.4	+0.19
	I1	0.103	16.2	91.8	-0.78	10.3	7.2	-0.22
	I2	0.115	1.7	36.1	-0.24	24.8	63.9	+0.32
	I3	0.115	26.2	97.7	+0.93	3.9	1.8	-0.13
Additional variables	R1	0	-	79.2	-0.90	-	17.7	-0.43
	R2	0	-	0.1	+0.04	-	58.8	+0.94
	R3	0	-	6.8	+0.14	-	13.3	-0.20
	R4	0	-	97.9	+0.77	-	0.1	+0.03
Inertia accounted for by axes			86.8%			10.7%		

of Table IV. As in principal component analysis, several axes are defined (equal to the smaller dimension of the table minus 1), but only the first two axes are given. Axes one and two accounted for 86.8 and 10.7%, respectively, of the total inertia of the data set. In general, with small contingency tables such as the one used in this example, more than two axes are seldom needed to account for more than 90% of the inertia. Besides the coordinates in the newly defined axes, the classes are represented by their relative weights, contribution to each axis, and reciprocal contribution to each axis. The relative weight (or mass) of each class represents the frequency of individuals in the corresponding row (or column). For instance, $L1 = 0.115$ [$= (3 + 2 + 7 + 15)/(3 \times 78)$]. The contribution to an axis is the proportion (or percentage) of inertia of that axis which is derived from a specified class (17.6% for L1). The reciprocal contribution (or class correlation) represents the proportion of inertia of the class (row or column) that is accounted for by the specified axis. It is also the correlation between the axis and the class. For L1, 89% and 15.9% of the inertia of this class are accounted for by axes one and two, respectively (Table V). Finally, the sign of the new coordinate indicates the direction that the class deviates from the origin (i.e. marginal frequencies).

Graphs can be generated with correspondence analysis where (newly defined) coordinates of classes, rather than the frequencies, are plotted along the axes. The graph in Fig. 1A, where axis 1 is horizontal, and axis 2 vertical, illustrates the relations among classes. Proximity of points representing classes indicates correspondences that can be checked using chi-square tests. When a series of classes representing successive levels of a coded quantitative or of a semiquantitative variable that reflects a logical increase (as, for example, disease intensity, F1-F4) is considered, a path linking the successive classes can be drawn, and the movement along this path may be examined in relation with positions of other classes, paths, and axes. Fig. 1A shows the relationships among F, I, L, and S. High disease intensity (F4) is graphically associated with (close to) high infection efficiency (I3) and short latent period (L1); low disease intensity (F1) is associated with low infection efficiency (I1), long latent period (L3), and low sporulation (S1). The shapes of the paths of increasing disease intensity (F1-F4), increasing infection efficiency (I1-I3), and decreasing latent period (L3-L1) are similar, indicating strong correspondences. The relation between disease intensity and sporulation intensity (S1-S3) is not as strong. The two paths, although sharing the same movement along the direction of the horizontal axis, and starting in the same area of the graph, diverge in their later part, and do not have the same length. The graph therefore suggests that, whereas low sporulation is closely associated with low disease intensity, medium (S2) to high (S3) sporulation corresponds to similar disease intensity levels, predominantly medium (F3) and high (F4); this interpretation of the graph adequately describes the corresponding contingency table (Table IV, [F \times S]).

The next step of the analysis, the evaluation of representativeness of

reaction types, is addressed in the second part of Table V (see "Additional variables"). In the case of this additional variable, no contribution to axes is computed, since it was not involved in the computation of eigenvalues and eigenvectors. What is actually done is that the axes defined in the first step of the analysis are used as a framework, on which the new variable, represented by its four classes is superimposed. Reciprocal contributions to axes are computed, which allows an assessment of how well the additional classes fit in with the framework. This is measured by the relative contribution values: $79.2 + 17.7 = 96.9\%$ of the inertia represented by R1 is accounted for by the two first axes. Similarly, 58.9, 20.1, and 98.0% of the inertias represented by R2, R3, and R4, respectively, are accounted for. It can, therefore, be concluded that R1, R2, and R4 are mostly represented by the two first axes, whereas the representation of R3 is poor. A further inspection of computation results would indicate that 79.9% of the inertia of this reaction type is accounted for by axis 3, which in turn accounts for only 2.4% of the total information represented by the upper part of Table IV.

The four reaction types can be plotted onto the previous graph (Fig. 1B). As expected, R1 is closely associated with low infection efficiency, long latent period, low sporulation, and low disease intensity; R4 is associated with high infection efficiency, short latent period, medium to high sporulation, and high disease intensity. R2 and R3 occupy intermediate positions.

The overall conclusion of the analysis of this hypothetical example would therefore be that a close association between disease intensity and monocyclic processes, especially infection efficiency and latency period, exists. Very susceptible (R4) and highly resistant (R1) reaction types adequately predict epidemic levels; but caution is needed when considering the epidemiological effect of R2 and, especially, R3.

III. EXAMPLES

A. EPIDEMIC TRENDS IN GROUNDNUT DISEASES

A survey was conducted in Ivory Coast on the main diseases of groundnut (Savary, 1987). In West Africa, several diseases affect groundnut crops, and the aim of the survey was to assess the part taken in this multiple pathosystem by a new component, groundnut rust (*Puccinia arachidis* Speg., Savary et al., 1988). The survey was conducted during three consecutive years, yielding information on various components of the pathosystem, especially foliar diseases, in a population of 309 fields. The fields were visited at various stages in the crop development, and the analysis considered each field at each visit as a unique observation. One of the analyses performed on these data was aimed at obtaining a simplified, overall picture of a typical farmer's field, in terms of dynamics of the main diseases. Three disease variables were

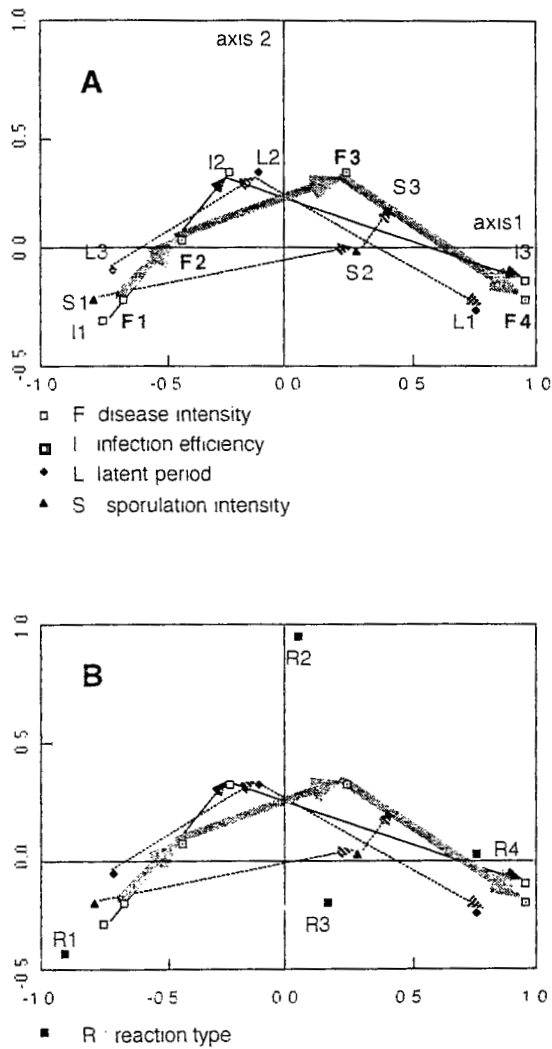


Fig. 1. A correspondence analysis of a theoretical germplasm database on disease reaction. (A) First step of the analysis, involving disease intensity (F), infection efficiency (I), latent period (L), and sporulation intensity (S). The graph is drawn using the first (horizontal) and second (vertical) axes of the analysis. The successive classes are indicated and linked into paths. (B) Second step of the analysis. Four reaction types (R) were superimposed on the graph.

TABLE VI
Contingency tables and relationships among rust, late leaf spot and early leaf spot severities, and peanut development stages

Variables	Rust severity				
	R0	R1	R2	R3	R4
P0	53	8	1	0	0
P1	14	21	11	16	2
P2	8	13	3	13	1
P3	1	10	8	36	13
P4	2	2	9	32	33
A0	26	4	0	9	3
A1	33	22	9	19	6
A2	6	4	5	13	11
A3	9	12	6	36	20
A4	3	12	12	20	9
DV1	44	6	0	2	0
DV2	24	20	5	7	2
DV3	5	13	11	25	5
DV4	3	10	10	33	11
DV5	1	5	6	30	31

^a R0-R4: classes of rust severities; P0-P4: classes of late leaf spot severities; A0-A4: classes of early leaf spot severities; DV1-DV5: crop development stages. Entries are number of fields. Each contingency table shows the bivariate frequency distribution of the sample (309 fields) for each pair of variables.

considered: rust, early leaf spot (*Cercospora arachidicola* Hori), and late leaf spot (*Phaeoisariopsis personata* (Berk. & Curt.) Deighton) severities. An additional variable, the crop development stage, was incorporated into the analysis in order to provide a physiological time reference. The frequency distributions of all three diseases were strongly asymmetrical, a large number of fields (especially those at early crop development) being unaffected, and a small number showing high severities.

Severity variables for the three diseases were encoded by use of the following boundaries: absent [0%]; very low [0, 0.01%]; low [0.01, 1%]; medium [1, 20%]; and high [20, 100%]. For each disease, five severity classes were thus defined, labeled R0, R1, R2, R3, and R4 for rust, P0, P1, P2, P3, and P4 for late leaf spot, and A0, A1, A2, A3, and A4 for early leaf spot. Using these boundaries, we represented all classes by an adequate number of fields (30-100). Five development stages were considered: seedling to fourth tetrafoliate (DV1), flowering to beginning peg (DV2), beginning pod to beginning seed (DV3), full pod to full seed (DV4), and harvest maturity (DV5).

Three contingency tables were built (Table VI), by use of the quantitative coded variables and the qualitative variable: [R × P], [R × A], and [R × DV]. Examination of the contingency tables provides indications on the

relations among diseases, and between rust and physiological time. In the first contingency table, most of the fields were distributed along the first diagonal: there was an overall increase of late leaf spot with increasing rust severity. In the second table, most of the fields at low rust and early leaf spot levels were also concentrated along or adjacent to the first diagonal: R0 corresponded to A0 (26 fields) or A1 (33 fields); R1 corresponded to A1 (22 fields). But very high rust (R4) did not correspond to high early leaf spot (A4, 9 fields), but to medium early leaf spot (A3, 20 fields). The third table, [R × DV], represents the increase of rust with the increasing development of the crop: most of the fields were distributed, again, along the first diagonal. It is worth noting, however, that the highest frequency (mode) of rust severity at DV1 (44 fields), as well as at DV2 (24), was R0; at DV3, the mode reached R3 (25), where it remained at DV4 (33); and at DV5, the mode was at the highest rust severity, R5 (31 fields). This third table indicates that the increase of rust over development stages involves two lag phases, at epidemic onset, and at the end of the epidemic. All three tables can be submitted to chi-square tests (Table VI), where independence of the distributions of the four variables was rejected ($P < 0.0001$).

A correspondence analysis was conducted using the disease variables to generate a system of axes, and the development of the crop was superimposed on these axes. In other words, a framework representing interaction among diseases was first computed, onto which a scale of development stages was projected.

A series of axes was generated (Table VII), but only the first two axes are given, as they represented a total of 95.5% of the inertia of the analysed contingency table. The first axis represents a contrast between absence of rust (R0, in the negative direction, with a very strong contribution) and R3 and R4 (in the positive direction). It also represents a contrast between absence of late leaf spot (P0, in the negative direction) with P3 and P4. As for early leaf spot, it primarily accounts for an opposition between no or low disease severity (A0 and, to a lesser degree, A1), and other disease levels. Axis 1 essentially represents a contrast between unaffected or little diseased fields, and higher diseased fields. Axis 2 represents contrasts between low rust and late leaf spot levels (R1, P1, and P2), and high disease levels (R4 and P4); the contributions to this axis of early leaf spot (A) variables are small. As a summary, the first axis can be interpreted as representing appearance of diseases, and the second axis, intensification of diseases (at least, rust and late leaf spot) in the fields. The reciprocal contributions for the variables that were involved in axis generation are high; the proportion of inertia of R4 accounted for by the two first axes is, for instance: $64.5 + 33.0 = 97.5\%$. The large proportion of total inertia accounted for by axes reflects these large reciprocal contributions. It is worth noting that the reciprocal contributions for the categories of development stage (which were not involved in axis generation, and thus have no weight nor contributions to axes) were also large: the resulting axes account,

TABLE VII
Peanut disease severities: relative weights and contributions to axes

	Classes	Relative Weight	Contribution to axes					
			Axis 1			Axis 2		
			Contribution		Co-ordinate	Contribution		Co-ordinate
			To axis	Reciprocal		To axis	Reciprocal	
Columns	R0	0.249	64.3	96.3	-1.03	9.5	3.6	+0.20
	R1	0.175	0.7	5.3	-0.13	45.0	91.2	-0.52
	R2	0.104	2.8	36.1	+0.34	9.3	30.0	-0.31
	R3	0.314	14.3	87.2	+0.43	0.1	0.1	-0.02
	R4	0.159	17.9	64.5	+0.68	36.1	33.0	+0.49
Rows	P0	0.100	47.3	95.2	-1.39	9.0	4.6	+0.31
	P1	0.104	0.4	5.1	-0.12	26.4	90.1	-0.52
	P2	0.061	0.2	3.9	-0.11	13.0	77.5	-0.47
	P3	0.110	8.7	84.9	+0.57	0.6	1.5	-0.08
	P4	0.125	18.1	70.4	+0.77	27.8	27.6	+0.48
	A0	0.068	10.4	81.8	-0.79	7.0	14.0	+0.33
	A1	0.144	5.0	81.0	-0.37	4.1	16.9	-0.17
	A2	0.063	1.6	62.1	+0.33	2.9	28.3	+0.22
	A3	0.134	4.8	80.7	+0.38	1.5	6.6	+0.11
	A4	0.091	3.5	54.7	+0.40	7.2	28.9	-0.29
Additional variables	DV1	0	-	94.1	-1.36	-	5.8	+0.34
	DV2	0	-	65.6	-0.57	-	22.5	-0.33
	DV3	0	-	33.3	+0.29	-	52.7	-0.37
	DV4	0	-	79.7	+0.48	-	5.6	-0.13
	DV5	0	-	70.6	+0.77	-	25.4	+0.44
Inertia accounted for by axes			76.1%			19.4%		

Rust severity (R0-R4), and late (P0-P4) and early (A0-A4) leaf spot severity were involved in generation of axes. Development stage was used as an additional variable.

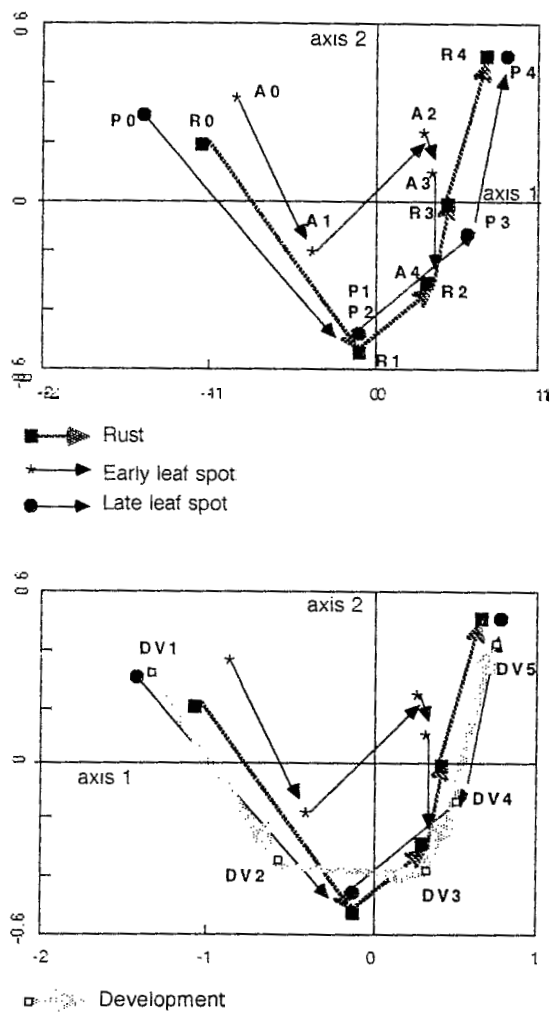


Fig. 2. A correspondence analysis of rust, late leaf spot, and early leaf spot of groundnut in farmer's fields in the Ivory Coast. Axes 1 and 2 are represented horizontally and vertically, respectively. The axes are defined using classes of rust (R), early (A) and late (P) leaf spot intensities. The successive development stages of the crop (DV) were superimposed on the graph.

to a very large extent, for the development of the crop.

Fig. 2 illustrates the associations indicated by the two axes. The graph indicates a very strong association between the paths of increasing rust and late leaf spot severities: the two paths show the same movement in the same direction. The association between the increase of rust and late leaf spot is also

strongly corresponding to the development of the crop. The path representing early leaf spot initially corresponds to the paths of rust and late leaf spot (the movement from A0 to A1, and A2 parallels that of R0 to R3, and that of P0 to P3), but then shows a different movement (the increase from A2 to A3 and A4 corresponds to a direction opposite to the increase of rust from R3 to R4, and P3 to P4). This second phase of increase in early leaf spot intensity is also opposed in direction to the development of the crop (DV4 to DV5). As a result, maximum early leaf spot corresponds to beginning pod–beginning seed (DV3), not to harvest maturity. Therefore the graph indicates that the increase in rust and late leaf spot intensity corresponds to the overall increase in physiological age of the crop, whereas change in early leaf spot intensity consists of two phases: an initial increase, and a further decrease when the crop approaches maturity.

Additional details in the description of variation in rust intensity over groundnut development are also indicated by the graph. R0 is close to DV1 (seedling to fourth tetrafoliate), and R1 is positioned between DV2 (flowering to first tetrafoliate) and DV3 (beginning pod to beginning seed); the delay to reach R1 is long. R2 is very close to DV3; the delay between R1 and R2 is, in contrast, short. R3 is close to DV4 (full pod to full seed), indicating an increased delay in disease increase from R2 to R3, which seems equivalent to the delay between R3 and R4. Based on this development scale, the progress of groundnut rust appears sigmoid: slow at the early crop development, then fast, and finally slower. The increase of late leaf spot is very similar except for the proximity of P1 and P2, suggesting an extremely fast disease increase in this phase of disease intensification. It is worth noting that A1 is very close to DV2, i.e., corresponds to an earlier development stage than P1 or R1. As a conclusion, this graph summarizes the appearance and further intensification of three groundnut diseases in an average farmer's field in Ivory Coast as follows: early leaf spot appears first, followed by rust and late leaf spot; the latter diseases have regular, sigmoidal progress throughout the cropping season, whereas early leaf spot reaches its maximum severity at V3 (beginning pod to beginning seed), after which severity decreases.

B. COMPONENTS OF RICE TUNGRO EPIDEMICS

Tungro is a major virus disease of rice in south and south-east Asia (Sogawa, 1976). The pathosystem involves several components: the spherical and bacilli-form virus particles, vectors (among which *Nephotettix virescens* Distant plays a pre-eminent role), and the rice host plant (Ou, 1987). In some areas in south-east Asia, the disease appears to be present in every cropping season, whereas in some others, sporadic, and sometimes most damaging outbreaks occur. One major challenge in the understanding of rice tungro epidemics lies in the explanation of the polyetic processes which account for the endemicity of the

disease in some areas, and for irregular, explosive epidemics in some others. Analyses on components of the tungro pathosystem were carried out from historical survey data conducted in the Philippines (Savary *et al.*, 1993).

The data were collected by the Philippines Department of Agriculture at several sites. In one province of the southern Philippines, North Cotabato, 130 fields were visited at the tillering phase during nine consecutive cropping seasons. The data included: cropping season (CS), planting date (PD), tungro incidence (IN), total vector population (V), and proportion of viruliferous vectors (VV). The disease is known to be endemic to the site. During the observation period, the disease was present in all years and cropping seasons, and had an overall disease prevalence (percentage of fields with infection) of 84.6%.

An analysis of these data was conducted to provide an overall description of the relationships between some components of the pathosystem, and assess the possibility of adequately characterizing tungro epidemics with a small set of variables that might be useful in tungro management.

Three quantitative variables, IN, V, and VV, were measured with low precision. Incidence (IN) was defined as the proportion of field area affected, and was assessed using an approximate grading scale with higher detail at low disease level (i.e., 0, 1, 3, 5, 10, 15, 20, 30, ... 100% field area affected). Disease incidence was encoded into disease classes: absence of disease (IN = 0: IN0), low incidence ($0 < IN \leq 2\%$: IN1), average ($2 < IN \leq 5\%$: IN2), and high incidence ($5 < IN \leq 100\%$: IN3). Estimates of the vector population (V) were obtained by counting individuals caught in ten sweepnet strokes above the canopy. Total vector counts were encoded as: low ($0 \leq V \leq 5$: V1), medium ($5 < V \leq 10$: V2), high ($10 < V \leq 15$: V3), and very high ($V > 15$ vectors: V4). The proportion of viruliferous vectors (VV) was encoded using four classes: none (VV = 0: VV0), low ($0 < VV \leq 5$: VV1), medium ($5 < VV \leq 15$: VV2), and high (VV > 15% of the vectors inducing a positive reaction in the transmission test: VV3). The planting dates were classified from very early (PD1) to very late (PD5) in each cropping season, and the two cropping seasons were categorized as rainy (RS) or dry (DS).

Using these five categorized variables, we built four contingency tables: IN \times V, IN \times VV, IN \times PD, and IN \times CS. From these contingency tables, chi-square values were computed, which indicated a significant association ($P < 0.05$) between variables in all cases. For instance, the IN \times PD contingency table indicated lower disease incidence in off-season plantings and the hypothesis of independence of the two variables was rejected ($P < 0.01$) on the basis of the chi-square value (58.7).

Tungro incidence may be seen as the outcome of the interactions among V, VV, PD, and CS. Thus, the four contingency tables were consolidated, using the classes of the variable to be explained, IN, as columns, and the different classes of the other variables as rows. This set of contingency tables was submitted to correspondence analysis.

Two main axes, accounting for 72.2 and 25.31% of total inertia, were

found. The first (horizontal) axis opposes absence of tungro (IN0) to presence (IN1, IN2, and IN3), whereas the second (vertical) axis opposes low (IN1) to high (IN3) incidence. As in the previous example, the two axes may therefore be interpreted as representing disease appearance and intensification. The axes, of course, also incorporate contribution of the row-variables. On the first axis, absence of tungro is associated with absence of viruliferous vectors and off-season (PD1 and PD5) plantings. The second axis involves a contrast between cropping seasons, and a strong contribution of early-intermediate plantings, associated with a very high incidence of tungro.

Fig. 3 shows the overall association among variables: absence of tungro (IN0, on right hand-side) is associated with absence of viruliferous vector and off-season plantings (PD1 and PD5); low tungro (bottom) is associated with late plantings (PD4), medium to high population of vector (V2 and V3), low to medium proportion of viruliferous vectors (VV1 and VV2), and dry season (DS). High tungro incidence (top) is associated with early or intermediate plantings (PD2 and PD3), large population of vector (V4), very high proportion of viruliferous vectors (VV3), and rainy season (RS). The path of increasing tungro incidence corresponds to that of increasing proportion of viruliferous vectors in a stepwise pattern. Whereas the path of increasing tungro incidence can be projected on both axes with two phases, disease appearance (IN0-IN1, horizontal axis) and disease intensification (IN1-IN2-IN3, vertical axis), the variation in total vector population (V) is almost entirely associated with axis 2. This suggests that the increase in vector population is primarily associated

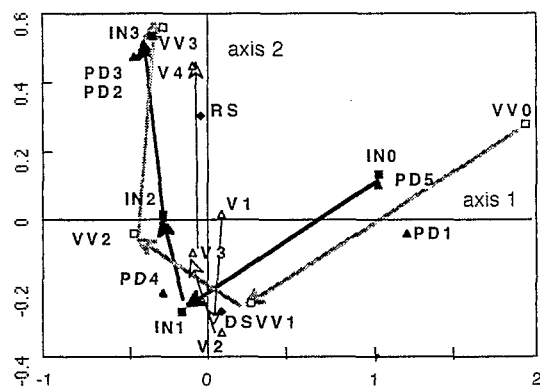


Fig. 3. Characterization of rice tungro epidemics in one 'hot spot' in the Philippines. The graph represents data from a total of 130 fields visited during 9 consecutive cropping seasons at the tillering stage. Intensity of tungro (IN), total vector population (V), proportion of viruliferous vectors (VV), planting period (PD), and cropping season (RS or DS) were considered in the analysis. The classes that can be described by a natural increase (e.g. IN0 to IN3) were linked to show paths.

with disease intensification, not disease appearance. On the other hand, there is a close, stepwise correspondence between increasing proportion of viruliferous vectors, and tungro incidence. Hence, the analysis indicates that any increment of the proportion of viruliferous vectors is associated with a corresponding disease increase. This analysis was compared to those representing different endemic or non-endemic sites in the Philippines (Savary *et al.*, 1993), and was interpreted as a typical endemic situation.

C. RELATIONS BETWEEN PRODUCTION LEVEL AND YIELD LOSSES

A database for the pathosystem of groundnut was established from a series of six independent experiments (Savary and Zadoks, 1992a). Each experiment involved one main factor with three levels and three blocks. The main factors were water management, weed control, fertilizer input (two experiments), seeding rate, and cultivar potential yield. In addition, different disease treatments were used. Treatments were established within each block, representing combinations of manipulated levels of rust (low or high) and leafspot (low or high), with one control plot (C; no disease). The two diseases were independently manipulated by means of inoculations and sprays with a protectant fungicide. Each experiment included one reference block where inputs were set to default values, representing the production level of an average farmer's field: suboptimal water management and weed control, no fertilizer, low seeding rate, and cultivar with low potential yield. In each experiment, three types of yield were obtained. Actual yield (Y) was obtained from plots with different combinations of diseases and the main (input) factor. Attainable yield (Y_a) was obtained from the control plots for each level of input factor. The attainable yield from the reference block was the reference yield (Y_{ref}) of an experiment.

Because the experiments were conducted over different seasons, and in a range of soil fertility, a large variation in reference yield was observed. The variation of actual yield was attributable to the superimposed effect of diseases. The overall relationships between Y_{ref} , Y , and presence or absence of diseases (D) was studied. The variable to be explained was Y , and two explanatory variables were considered: Y_{ref} and D . Only two categories of plots were therefore considered with respect to diseases: nearly disease-free, or affected by any of the two diseases. Variation of both yield variables was categorized in five classes, using the following boundaries: 0, 850, 1400, 2300, 3000, and 4500 kg/ha. Because plots used to estimate the reference yield were protected against diseases, no Y_{ref} values were found in the lowest class. Two contingency tables were built: $Y \times Y_{ref}$ and $Y \times D$. Both indicated a significant association among variables ($P < 0.01$), and were submitted to correspondence analysis.

The two axes shown in Fig. 4 account for 62.6% (horizontal axis) and 21.8%

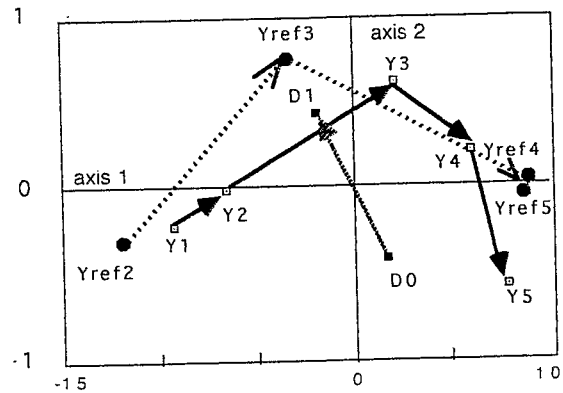


Fig. 4. Overall relationship between reference yield, actual yield, and foliar diseases from an experimental database on groundnut. Reference yield (Y_{ref}) was measured in nearly disease-free plots, and represents a variation in production level due to climate and soil. Five classes of actual yield (Y) were considered; their boundaries are the same as for Y_{ref} , where no very low yield (Y_{ref1}) was recorded. Diseases are represented by two categories only: absent ($D0$) or present ($D1$).

(vertical axis) of inertia. The sequences of classes for each variable could be delineated (e.g., from $Y1$ to $Y5$), and the resulting paths analysed. This analysis indicated two phases in actual yield (Y) progress: (a) a first phase of parallel increase of the actual yield, along with the reference yield, independent of disease levels (the vector $D0-D1$ was orthogonal (perpendicular) to the Y direction), and (b) a second phase where the increase of actual yield, still related to that of reference yield, was opposed to the occurrence of diseases, i.e., the change in Y ($Y3$ to $Y5$) was opposite to the shift from $D0$ to $D1$. The graph indicates a threshold in reference yield, above which the increase of actual yield not only depends on the improvement of the production situation, but also on the protection of the crop against the two diseases. For instance, it was only when the reference yield was higher than Y_{ref3} that the concurrent increase in actual yield was strongly influenced by (opposite to) the occurrence of diseases. In other words, the graph indicates a threshold in the improvement of the production situation, in terms of the hazard incurred by the two diseases. The threshold was represented by a class of reference yields ranging from 1400 to 2300 kg/ha (Y_{ref3}). It is worth noting that this represents the uppermost range of groundnut yields in farmers' fields in most of West Africa. The analysis indicates that progress in yield beyond this threshold cannot be foreseen without strict control of rust and leaf spot diseases. This analysis provides an overview of complex relationships in a two-disease pathosystem, that may be useful to set priorities in disease management. Further analysis of the

data indicated that the two diseases differ in their harmful effects, and had less than additive effects on yield reduction (Savary and Zadoks, 1992c).

IV. OVERVIEW

A. COMBINATION AND LINK WITH OTHER TECHNIQUES

Many multivariate methods have been used by epidemiologists, and have been reviewed by Madden (1983), Hau and Kranz (1990), and James and McCulloch (1990). One area where these methods have been particularly useful is the analysis of survey data.

Examples of holistic pest surveys, which generated a better knowledge of multiple pests systems, include the analysis of bean constraints in Colombia (Pinstrup-Andersen *et al.*, 1976), of yield-determining factors in pea in Idaho (Wiese, 1980), of wheat constraints in South Australia (Stynes, 1980) and in Germany (Kranz and Jörg, 1989). Numerous of methods were used to analyse the survey data, including regression analysis (Pinstrup-Andersen *et al.*, 1976; Wiese, 1980), canonical correlation, principal component analysis (Stynes, 1980), and path coefficient analysis (Kranz and Jörg, 1989). Techniques involved in the comparison of epidemics are similar in the methodological approach (Kranz, 1974).

As mentioned earlier, correspondence analysis shares common features with principal component analysis on the computational aspects. Principal component analysis was used, for instance, to characterize tomato early blight epidemics (Madden and Pennypacker, 1979). The analysis yielded three major, independent, components representing maximum disease severity, rate of disease increase, and shape of the disease progress curves. Its result was to highlight, on the basis of a sample of epidemics, the value of key epidemiological attributes for characterizing epidemics.

Because it involves the use of qualitative information, correspondence analysis, as an analytical tool, can also be related to two other techniques, discriminant analysis and multiple regression. The main aim of discriminant analysis is to characterize separate groups in a population, by means of one or several discriminant functions (Madden *et al.*, 1983; Hau and Kranz, 1990). Population densities of *Verticillium dahliae* and *Pratylenchus penetrans* at planting were, for instance, used to classify potato yields as being above or below 90% of the control (uninoculated) yield (Francl *et al.*, 1987). Multiple regression involving qualitative ("dummy") variables (Chatterjee and Price, 1977) have seldom been used in plant pathology. Daamen (1986a) used this technique to analyse wheat mildew populations, and tested the effect of years, growth stages, and cultivars. Regression can also be used to evaluate the results of correspondence analyses. In the case of yield losses of groundnut, a backward, stepwise regression analysis yielded an equation of the following form (Savary and Zadoks, 1992a):

$$Y = a + b_1 Ya \times RI + b_2 Ya \times SI + b_3 RI \times SI$$

where Y and Ya are the individual plot and attainable yields, respectively, and RI and SI are log-transformed areas under disease progress curves for rust and leaf spot, respectively. This equation describes yield variation as a response surface to interactions between attainable yield and disease injuries. Such a relationship is in agreement with the conclusion of the third example of correspondence analysis — the reference yield being the attainable yield associated with given climatic and soil conditions at a reference level of inputs. Further examination of the relationship indicated a less than additive effect of the two diseases on yield, which was confirmed by correspondence analysis on the same data set (Savary and Zadoks, 1992c), and independent factorial experiments (Savary and Zadoks, 1992b).

Another multivariate tool that can be associated with correspondence analysis is cluster analysis. Clusters can be considered as qualitative variables, among which relationships can be analysed. For instance, eight patterns of cropping practices, and eight pest (insects, diseases, and weeds) profiles were identified from the data resulting from a survey on rice crops in Central Luzon, Philippines (Elazegui *et al.*, 1990), using two separate cluster analyses. These patterns of cropping practices and types of pest profiles were then related to categorized levels of yields (Savary *et al.*, 1994).

B. USEFULNESS OF THE APPROACH

1. Precision of the data

Categorization of quantitative information representing components of a pathosystem is one means to take into account the level of precision of the data. This is particularly well exemplified in our second example on tungro epidemics. When numerous, precise data are available, a range of categories can be devised; when data with only low precision are available, a few, broad categories can be defined. This flexibility in handling data may be a critical advantage to enable the analysis of large, extremely valuable, sometimes heterogeneous data sets, such as those generated by plant protection or extension services.

2. Normalization

Categorization is also a means to normalize the variables (a) along their range, since each category should contain equivalent (commensurate) numbers of individuals (e.g., fields or epidemics), and (b) across data sets (e.g., surveys in different areas or years), so that comparisons are facilitated. Our third example on crop losses illustrates how a series of experiments, each of them representing a valuable fraction of the information needed for an overall interpretation, can be combined together without loss of vital information such as a variation in production situation, represented by the attainable yield.

3. *Relations among variables*

One strong advantage of dealing with categorized quantitative data is the rapid detection of complex, non-linear relations among several variables. This is best shown in our first example on groundnut diseases: all three disease variables are correlated with crop development, but with differing patterns, including lag phases, and disease decline. This can be detected with conventional non-linear regressions (Savary, 1987), at the cost of the choice of specific models. Correspondence analysis allows one to forward and test similar hypotheses, without specifying a functional relationship between development stage and disease intensity, thus providing an overall and neutral framework of relationships. These may further be explored more explicitly using regression techniques.

4. *Thresholds*

Finally, in the process of devising the boundaries and the categories, it is possible to introduce key or threshold values, the relevance of which might be revealed in the analyses. A domain where this approach might be particularly relevant is the development of strategies for tropical crop protection. Damage functions (Zadoks, 1985) are needed to define thresholds, which in turn are necessary to outline pest management strategies. These thresholds vary with production situations (De Wit and Penning de Vries, 1982) and are influenced by interactions between pests (Zadoks, 1985; Johnson *et al.*, 1986; Savary and Zadoks, 1992a, b). An approach that consists of quantifying thresholds that are specific to each separate pest, in a range of production situations, will often be impracticable. Besides, yield estimates in the field are usually difficult, and involve high errors (Poate, 1988). So are yield loss estimates. Farmers often have a good perception of what the yield of their crop may be, or what the desirable yield of a crop is; they also have their own classification of yields, from very poor, over regular, to very good. Similarly, the farmer's input to the crop may be categorized as low, medium, or high. Such classifications are particularly relevant and convenient in the context of tropical small-scale farming. Decision-making in agriculture, either strategic (e.g., the choice of a cultivar), or tactical (e.g., to spray, not to spray, or to wait) is, by nature, categorical. Analogously, levels of hazard due to pests could be defined that would be conceptually compatible with decision-making.

C. THE NEED TO EXPLOIT CATEGORICAL INFORMATION IN PLANT PATHOLOGY

Correspondence analysis appears to be a robust technique that basically requires a conceptualization of a framework of relationships among components in the studied system. One might see it as just another multivariate tool where these components are represented in the form of classes, frequencies of classes, and then related in contingency tables. There are, however,

solid reasons to consider correspondence analysis (and methods for analysing qualitative and categorical data in general) if plant pathologists are taking seriously the challenge of analysing the interactions that are at work within the disease tetrahedron. These interactions involve variables that are heterogeneous, and the approach outlined here is one means to analyze them.

In many pathosystems, patterns of cropping practices are driving variables of epidemics; whether a farmer's field belongs to a cluster representing intensive or extensive agriculture is a key to explain levels of a disease, or disease combinations (Savary *et al.*, 1994). The pattern of cropping practice is, by essence, a qualitative attribute. Its introduction into analyses provides scope for a novel approach of the data, so that new hypotheses can be constructed.

REFERENCES

- Benzécri, J. P. (1973). "L'Analyse des Données". Vol. 2. L'Analyse des Correspondances. Dunod, Paris.
- Butt, D. J. and Royle, D. J. (1974). Multiple regression analysis in the epidemiology of plant disease. In "Epidemics of Plant Diseases" (J. Kranz, ed.), pp. 78-114. Springer Verlag, Berlin, Heidelberg, New York.
- Campbell, L. C. and Madden, L. V. (1990). "Introduction to Plant Disease Epidemiology". John Wiley, New York.
- Chatterjee, S. and Price, B. (1977). "Regression Analysis by Example", Chapter 4, pp. 74-100. Qualitative variables as regressors. John Wiley, New York.
- Chester, K. S. (1946). "The Nature and Prevention of the Cereal Rusts as Exemplified in the Leaf Rust of Wheat". The Chronica Botanica Company, Waltham, MA.
- Daamen, R. A. (1986a). Measures of disease intensity in powdery mildew (*Erysiphe graminis*) in winter wheat. 1. Errors in estimating pustule number. *Netherlands Journal of Plant Pathology* 92, 197-206.
- Daamen, R. A. (1986b). Measures of disease intensity in powdery mildew (*Erysiphe graminis*) in winter wheat. 2. Relationships and errors of estimation of pustule number, incidence and severity. *Netherlands Journal of Plant Pathology* 92, 207-222.
- Dagnélie, P. (1973). "Théorie et Méthodes Statistiques. Applications Agronomiques", Vol. 2. Presses Agronomiques de Gembloux.
- Dervin, C. (1988). "Comment Interpréter les Résultats d'une Analyse Factorielle des Correspondances?" Institut Technique des Céréales et Fourrages, Paris.
- De Wit, C. T. and Penning de Vries, W. W. T. (1982). L'analyse des systèmes de production primaire. In "La Productivité des Pâturages Sahéliens" (W. W. T. Penning de Vries and M. A. Djiteye, eds), pp. 275-283. Agricultural Research Report 918, Pudoc, Wageningen.
- Elazegui, F. A., Soriano, J., Bandong, J., Estorninos, L., Jonson, I., Teng, P. S., Shepard, B. M., Litsinger, J. A., Moody, K. and Hibino, H. (1990). Methodology used in the IRRI integrated pest survey. In "Crop Loss Assessment in Rice", pp. 241-271. International Rice Research Institute, Los Baños, Philippines.
- Francl, L. J., Madden, L. V., Rowe, R. C. and Riedel, R. M. (1987). Potato yield loss prediction and discrimination using preplant population densities of *Verticillium dahliae* and *Pratylenchus penetrans*. *Phytopathology* 77, 579-584.

- Gibbons, J. D. (1976). "Non Parametric Methods for Quantitative Analysis". Holt, Rinehart and Winston, New York.
- Greenacre, M. J. (1984). "Theory and Applications of Correspondence Analysis". Academic Press, London.
- Hau, B. and Kranz, J. (1990). Mathematics and statistics for analysis in epidemiology. In "Epidemics of Plant Diseases. Mathematical Analysis and Modeling", 2nd edition. (J. Kranz, ed.), pp. 12-52. Springer Verlag, Berlin, Heidelberg, New York.
- Hill, M. O. (1974). Correspondence analysis: a neglected multivariate analysis. *Applied Statistics* 23, 340-354.
- IRRI (International Rice Research Institute) (1988). "Standard Evaluation System for Rice", 3rd edition. International Rice Research Institute. Los Baños, Philippines.
- James, F. C. and McCulloch, C. E. (1990). Multivariate analysis in ecology and systematics: panacea or Pandora's box? *Annual Review of Ecology Systematics* 21, 129-166.
- James, W. C. (1974). Assessment of plant diseases and losses. *Annual Review of Phytopathology* 12, 27-48.
- Johnson, K. B., Radcliffe, E. B. and Teng, P. S. (1986). Effects of interacting populations of *Alternaria solani*, *Verticillium dahliae* and the potato leafhopper (*Empoasca fabae*) on potato yield. *Phytopathology* 76, 1046-1052.
- Kranz, J. (1974). Comparison of epidemics. *Annual Review of Phytopathology* 12, 355-374.
- Kranz, J. (1988). Measuring plant disease. In "Experimental Techniques in Plant Disease Epidemiology" (J. Kranz and J. Rotem, eds), pp. 35-50. Springer Verlag, Berlin, Heidelberg, New York.
- Kranz, J. and Jörg, E. (1989). The synecological approach in plant disease epidemiology. *Review of Tropical Plant Pathology* 6, 27-38.
- Large, E. C. (1966). Measuring plant disease. *Annual Review of Phytopathology* 11, 47-57.
- Madden, L. V. (1983). Measuring and modeling crop loss at the field level. *Phytopathology* 73, 591-596.
- Madden, L. V. and Pennypacker, S. P. (1979). Principal component analysis of tomato early blight epidemics. *Phytopathologische Zeitschrift* 98, 364-369.
- Madden, L. V., Pennypacker, S. P., Antle, C. E. and Kingsolver, C. H. (1981). A loss model for crops. *Phytopathology* 71, 685-689.
- Madden, L. V., Knoke, J. K. and Louie, R. (1983). Classification and prediction of maize dwarf mosaic intensity. In "Proceedings of an International Maize Virus Disease Colloquium and Workshop" D. T. Gordon, J. K. Knoke, L. R. Naut and R. M. Ritter, eds), pp. 238-242. The Ohio State University, Ohio Agricultural Research and Development Center, Wooster.
- McCool, P. M., Younglove, T. and Musselman, R. C. (1986). Plant injury analysis: contingency tables as an alternative to analyses of invariance. *Plant Disease* 70, 357-360.
- Neter, J. and Wasserman, W. (1974). "Applied Linear Statistical Models". Richard D. Irwin, Homewood, Illinois.
- Nutter, F. W., Teng, P. S. and Shokes, F. M. (1991). Disease assessment terms and concepts. *Plant Disease* 75, 1187-1188.
- Ou, S. H. (1987). "Rice Diseases", 2nd Edition. CAB International, Farnham House, Farnham Royal, Slough.
- Padwick, G. W. (1956). Losses caused by plant diseases in the tropics. Commonwealth Mycological Institute Kew, Surrey, UK, Phytopathology Papers No. 1. 60pp.
- Pinstrup-Andersen, P., de Londono, N. and Infante, M. (1976). A suggested procedure for estimating yield and production losses in crops. *PANS* 22, 359-365.

- Poate, D. (1988). A review of methods for measuring crop production from small-holder producers. *Experimental Agriculture* **24**, 1-14.
- Porkess, R. (1988). "Dictionary of Statistics". Collins, London and Glasgow.
- Savary, S. (1987). Enquête sur les maladies fongiques de l'arachide (*Arachis hypogaea*) en Côte d'Ivoire (I). Méthodes d'enquête et étude descriptive: les conditions culturales et les principales maladies. *Netherlands Journal of Plant Pathology* **93**, 167-188.
- Savary, S. and Zadoks, J. C. (1992a). Analysis of crop loss in the multiple pathosystem groundnut-rust-leaf spot. I. Six experiments. *Crop Protection* **11**, 99-109.
- Savary, S. and Zadoks, J. C. (1992b). Analysis of crop loss in the multiple pathosystem groundnut-rust-leaf spot. II. Study of the interaction between diseases and crop intensification in factorial designs. *Crop Protection* **11**, 110-120.
- Savary, S. and Zadoks, J. C. (1992c). Analysis of crop loss in the multiple pathosystem groundnut-rust-late leaf spot. III. Correspondence analyses. *Crop Protection* **11**, 229-239.
- Savary, S., Subba Rao, P. V. and Zadoks, J. C. (1989). A scale of reaction types of groundnut to *Puccinia arachidis* Speg. *Journal of Phytopathology* **124**, 259-266.
- Savary, S., Noirot, M., Bosc, J. P. and Zadoks, J. C. (1988). Peanut rust in West Africa: a new component in a multiple pathosystem. *Plant Disease* **78**, 1001-1009.
- Savary, S., Tiongco, E., Fabellar, N. and Teng, P. S. (1993). A characterization of rice tungro epidemics from historical survey data in the Philippines. *Plant Disease* **77**, 376-382.
- Savary, S., Elazegui, F. A., Moody, K. and Teng, P. S. (1994). Characterization of rice cropping practices and multiple pest systems in the Philippines. *Agricultural Systems* **46**, 385-408.
- Sogawa, K. (1976). Rice tungro virus and its vectors in tropical Asia. *Review of Plant Protection and Research* **9**, 21-46.
- Stynes, B. A. (1980). Synoptic methodologies for crop loss assessment. In "Assessment of Losses Which Constrain Production and Crop Improvement in Agriculture and Forestry" (P. S. Teng and S. V. Krupa, eds), pp. 166-175. Miscellaneous Publication 7, University of Minnesota Agricultural Experiment Station.
- Teng, P. S. and Bowen, K. L. (1985). Disease modelling and simulation. In "The Cereal Rusts", Vol. II. (A. P. Roelfs and W. R. Bushnell, eds), pp. 435-466. Academic Press, New York.
- Teng, P. S. and Gaunt, R. E. (1980). Modelling systems of disease and yield loss in cereals. *Agricultural Systems* **6**, 131-154.
- Wiese, M. V. (1980). Comprehensive and systematic assessment of crop yield determinants. In "Assessment of Losses Which Constrain Production and Crop Improvement in Agriculture and Forestry" (P. S. Teng and S. V. Krupa, eds), pp. 262-269. Miscellaneous Publication 7, University of Minnesota Agricultural Experiment Station.
- Zadoks, J. C. (1978). Methodology of epidemiological research. In "Plant Disease: An Advanced Treatise", Vol. II (J. G. Horsfall and E. B. Cowling, eds), pp. 63-96. Academic Press, New York.
- Zadoks, J. C. (1985). On the conceptual basis of crop loss assessment: the threshold theory. *Annual Review of Phytopathology* **23**, 455-473.
- Zadoks, J. C. and Schein, R. D. (1979). "Epidemiology and Plant Disease Management". Oxford University Press, New York.