Computational Statistics & Data Analysis 18 (1994) 121–140
North-Holland

Multiple factor analysis (AFMULT package)

B. Escofier

IRISA, avenue du Général Leclerc, 35042 Rennes Cedex, France
and IUT, rue Montaigne, 56008 Vannes Cedex, France

J. Pagès

ENSA, 65, rue de Saint-Brieuc, 35042 Rennes Cedex, France

Abstract: Multiple Factor Analysis (MFA) studies several groups of variables (numerical and/or categorical) defined on the same set of individuals. MFA approaches this kind of data according to many points of view already used in others methods as: factor analysis in which groups of variables are weighted, canonical analysis, Procrustes analysis, STATIS, INDSCAL. In MFA, these points of view are considered in a unique framework. This paper presents the different outputs provided by MFA and an example about sensory analysis of wines.

Keywords: Multiple factor analysis; Canonical analysis; Groups of variables; Factor analysis; Principal components analysis; Multiple correspondence analysis.

1. Introduction to the method

1.1. Introduction–summary

Multiple Factor Analysis (MFA) studies several groups of variables defined on the same set of individuals.

The core of the method is a factor analysis applied to the whole set of variables in which each group of variables is weighted. This point of view leads to a representation of individuals and variables, as in any factor analysis.

Owing to the weighting, this factor analysis can be interpreted as a canonical analysis. This point of view induces a display in which representations of the set of individuals associated to each group of variables are superposed (these displays are akin to that of procrustes analysis).

Correspondence to: B. Escofier, IRISA, Avenue du Général Leclerc, 35042 Rennes Cedex, France.

SSDI 0167-9473(92)00002-V

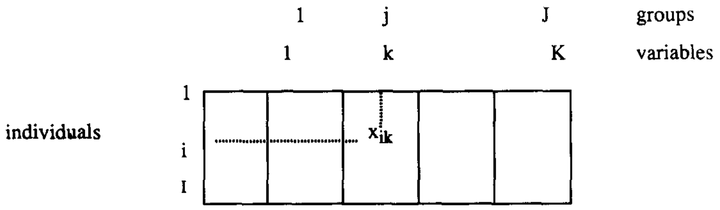

Fig. 1. Data set.

This factor analysis induces a global analysis of groups of variables in which each group is represented by the scalar product matrix it defines on the set of individuals. This point of view leads to a global representation of groups and to an estimation of INDSCAL model parameters.

After describing data and notations, we present different aspects of MFA. For each one, we set the problem, and indicate the MFA solution.

A complete presentation of MFA is included in Escofier and Pagès (1990).

1.2. Data

MFA analyses data tables in which a set of individuals (in rows) is described by several groups of variables (in columns), i.e., tables as shown in Figure 1.

Variables may be of two types: numerical (= quantitative) or categorical (= qualitative). Each categorical variable is to be coded by the set of indicator variables of its categories. Thus, later on, the word 'variables' designates either a quantitative variable or a indicator variable. Variables belonging to the same group should be of the same type.

Individuals and numerical variables may be weighted by the user.

Group of variables can get the status 'active' or 'illustrative' (an 'illustrative' – or supplementary – group does not influence factors).

Missing data are allowed in the case of categorical variables. Concretely, an individual which does not have any category for the variable k, has 0 for all the indicator variables associated to variable k.

1.3. Notation

We consider an array without supplementary elements.

I	number of individuals,
K	number of variables,
J	number of groups,
K_j	number of variables belonging to the group j,
X	data matrix (dimension: I, K),
X_j	data matrix restricted to group j (dimension I, K_j),
M	variables weights matrix (diagonal; dimension K, K),
D	individuals weights matrix (diagonal; dimension I, I),
$W_j = X_j M X_j'$	scalar products matrix associated to group j.

All elements (individuals, variables, groups) are represented in Euclidean spaces. These spaces are named according to the objects they include:

R^K	individuals space (defined by all the variables),
R^{K_j}	individuals space defined by group j variables,
R^I	variables space,
E_j	sub-space of R^I spanned by variables of group j,
$R^{I \times I}$	groups space (a group is represented by $W_j D$: dimension I, I).

1.4. General problem and weighting of variables groups

1.4.1. Problem

The problem can be roughly decomposed through three items, each one corresponding to a point of view:
(a) typology of individuals described by the whole set of variables,
(b) overview of relationships between variables,
(c) comparison of variables groups.

Items (a) and (b) are classic in factor analysis (Principal Components Analysis: PCA; Multiple Correspondence Analysis: MCA). Item (c) overlaps several objectives described afterwards.

Whatever the point of view, weighting of variables groups is necessary to make the influence of each group comparable in a global analysis. Concretely, according to the factor analysis point of view, we want to avoid the possibility of a single group having a dominant influence on the first factor (nothing can be required for further factors because a multidimensional group will always influence more factors than an unidimensional one).

1.4.2. Simultaneous analysis of numerical and categorial variables

This problem seems independent of the previous one but can be solved also by a weighting of variables. This follows from the following property.

On the condition that indicator variable k is weighted by the proportion of individuals which do not possess the character k, PCA applied to a in indicator matrix is equivalent to MCA. That equivalence allows to apply the same technique (PCA) to numerical or categorical variables.

In order to take the two kinds of variables into account simultaneously, we have to balance their influence, that is to say their inertia. This is an important aspect of the weighting of groups of variables.

1.4.3. Solution

Before the weighting of groups, let us denote:

– $a(k)$ the weight of the variable k. Generally, for quantitative variables, $a(k) = 1$ for all k. Concerning qualitative variables, $a(k)$ is the proportion of individuals which do not possess category k.

– $\lambda(j, 1)$ the first eigen value of factor analysis applied to the single group j (this factor analysis is PCA in the case of a quantitative group and MCA in the case of a qualitative group).

The weighting of groups consists of attributing the weight $a(k)/\lambda(j, 1)$ to each variable of the group j. By this way, the first eigen value of factor analysis applied to the single group j becomes 1. Therefore, groups are balanced in the following sense: in any direction, maximum inertia of the sub-cloud associated to one group is 1. Thus, in a global factor analysis, it is impossible for a single group to give rise to the first factor.

Of course, group contributions to global analysis are not similar: an unidimensional group cannot exert an important influence on more than one factor; a multidimensional group will influence several factors.

This weighting is a specific characteristic of MFA; it induces many properties described later.

1.5. Representation of individuals and variables

1.5.1. Problem

These representations correspond to the classic aim of factor analysis, that is to say:

- typology of individuals,
- typology of variables,
- links between the two typologies.

1.5.2. Solution

MFA is a factor analysis applied to the array including all groups of variables. Roughly, the behaviour of the method is equivalent to PCA (concerning quantitative variables) or to MCA (concerning qualitative variables).

The use of weights $1/\lambda(j, 1)$ balances inertia between the different groups and thus balances their influences. Taking account of the PCA–MCA equivalence previously mentioned, this weighting allows applications in which some groups are quantitative and other qualitative.

1.6. Setting up common factors

1.6.1. Problem

Let us recall, schematically, that a group of variables forms a multidimensional structure. In this sense, a common factor is a dimension common to these structures.

Search for factors which are comon to several groups of variables is a problem frequently encountered in data analysis. It refers to canonical analysis. In the case of more than two groups, the most interesting generalization is, doubtless, due to Carroll (1968).

Carroll measures the relationship between a factor and a group of variables through the multiple correlation coefficient. In case of a group including strongly related variables, this measure does not give satisfaction because it does not consider relationships between the variables and lacks for stability (the sub-space spanned by the variables of the group is itself unstable).

The problem is to propose a method which follows the general principle of Carroll but uses a more stable measure of relationship between a variable and a group.

1.6.2. Solution

MFA can be considered like a canonical analysis in the sense of Carroll since it follows the principle:
(a) setting up general variables, each one related to all the groups,
(b) for each general variable z, search for the canonical variable in each group j, that is to say, linear combination of group j variables, related to z.

The main differences between the two methods are:

(a) to measure the relationship between variable z and group j, MFA uses $\mathscr{L}(z, j)$, projected inertia of group j variables along the direction defined by z. This measure possesses some interesting properties:

- Taking the weighting of variables into account: $0 \leq \mathscr{L}(z, j) \leq 1$; $\mathscr{L}(z, j) = 0$ if z is orthogonal to the sub-space spanned by group j variables; $\mathscr{L}(z, j) = 1$ if z is confounded with the first principal component of group j. If all variables of group j are orthogonal to one another, $\mathscr{L}(z, j)$ is confounded with multiple correlation coefficient.
- In MFA, the general variable z, related to all the groups, satisfies the criterion: $\Sigma_j \mathscr{L}(z, j)$ maximum. This criterion leads to the first principal component of X.
- Owing to weighting of variables, inertia of the variables of the same group can be interpreted as a relationship measure. This allows to interpret the same criterion from a PCA point of view or from a canonical analysis point of view.

(b) In Carroll's method, a canonical variable of group j, associated to the general variable z, is $P_j(z)$, projection of z on the sub-space E_j generated by j. About this point, MFA calculates $W_j D(z)$; compared to $P_j(z)$, $W_j D(z)$ extracts more inertia, and leads to a representation of individuals more easily interpretable. This point appears more precisely in Section 1.7.2. If all the variables of group j are orthogonal to one another, P_j is equal to $W_j D$ and the two ways are equivalent.

1.7. Superposed representation

1.7.1. Problem

Each group defines a structure on the individuals set. A structure defined by group j is expressed by the shape of cloud N_I^j which represents an individuals set in PCA of X_j (N_I^j belongs to R^{K_j}).

In order to compare clouds N_I^j one to another, we need a superposed representation of N_I^j which sets up the structure common to the different clouds.

Classically, this objective refers to generalised Procrustes analysis (Gower, 1975). In fact this objective is closely related to canonical analysis. The two methods express the same objective, one by the way of variables and the other by the way of individuals.

This duality does not appear in the classic approach of canonical analysis or Procrustes analysis.

1.7.2. Solution

MFA considers clouds (N_I^j; $j = 1, J$) inside the space R^K defined by all the variables (R^K can be considered as the direct sum of R^{K_j}). In that space, which contains J images for each individual i, we are looking for a representation of all the N_I^j such as:

(a) this representation must be a projection upon a sub-space,
(b) clouds N_I^j must be well represented (high projected inertia),
(c) the J points representing the same individual must be close to one another.

Let us denote i^* the centre of gravity (= centroid) of the J points representing the same individual i (in R^K). MFA searches a sub-space which maximizes projected inertia of the J points i^*. That criterion carries out a compromise between items (b) and (c).

This criterion leads to the sub-space issued from the previous PCA. Consequently, inside MFA, the Procrustes point of view consists of projecting the N_I^j upon factorial axes (clouds N_I^j appear as 'illustrative' elements).

Such a projection of N_I^j is equivalent to the canonical variables ($W_j Dz$) calculated by MFA. MFA appears as a method giving a complete solution to the dual objectives of Procustes analysis and canonical analysis.

1.8. Global representation of groups

1.8.1. Problem

In order to get an global comparison of groups, we need a display in which each group is represented by one point. This kind of representation was introduced for the first time in STATIS (Escoufier, 1980).

The operator $W_j D$ is classically used to represent the group j: $W_j D$ includes the whole structure of individuals defined by group j. $W_j D$ belongs to the space of $I \times I$ dimensions. Global comparison of groups consists of studying cloud: ($W_j D$; $j = 1, j$).

This study is worked out inside STATIS. But, in this method, the main objective of that study is building a structure common to all the groups. Hence, the $W_j D$ are projected onto a sub-space, but the directions of that sub-space are quite impossible to interpret because they are not clearly reliable to variables.

Studying the cloud ($W_j D$; $j = 1, J$) we want to single out a display of groups. This display must be a projection onto interpretable directions.

1.8.2. Solution

In the space of $I \times I$ dimension, MFA searches a sequence of dimensions such as each one:

– is associate to a single direction of the variables, space R^I. That constraint necessarily reduces the goodness of fit but ensures the interpretability of the dimensions.

– maximizes, with usual orthogonality conditions, sum of projections (and not sum of squares). This criterion has a meaning because projection of W_jD onto such a dimension is always positive. It possesses the disadvantage, very unpleasant from a theoretical point of view, of being satisfied only by dimensions and not by sub-spaces. But this disadvantage is the price to pay in order to obtain properties, analogous to duality relationships in factor analysis, which ensure coherence with previous points of view.

Those properties are the following:

– The s-order axe found in the space $R^{I \times I}$ is the scalar product matrix associated to the s-order principal component of X (which can be also interpreted as a general variable of a canonical analysis) found in R^I. Hence, these directions have the same interpretation. Denoting z a principal component (in R^I), the corresponding axe in $R^{I \times I}$ is $zz'D$.

– The coordinate of W_jD with respect to the s-order axe (in $R^{I \times I}$) is equal to projected inertia of group j variables along the direction defined by the s-order principal component in R^I. Thus, it is equal to the relationship measure $\mathscr{L}(z, j)$ used in the canonical analysis point of view. So, a proximity between two groups along direction s indicates that the common factor s has the same importance in the two groups.

1.9. INDSCAL model

1.9.1. Problem

The usual algorithm for computing INDSCAL parameters does not ensure basic properties, that is to say, mainly: convergence, positive weights, orthogonality of dimensions (Carroll, 1981).

On the other hand, the INDSCAL model applied to several groups of variables is closely related to the other points of view, and usual programs do not care about that. The relationship between INDSCAL model and canonical analysis can be summarized with two points:

– dimensions of INDSCAL model are common factors (general variables in Carroll's canonical analysis).

– INDSCAL weights are used to express general variables in each group. In other words, INDSCAL model can be viewed as a canonical analysis in which the s-order canonical variable has to respect the constraint to be proportional to the s-order general variable.

Finally, we need a INDSCAL parameters estimation

- related to the other points of view,
- without technical problems (convergence...).

1.9.2. Solution

In MFA, the problem of global representation of groups corresponds to the INDSCAL model (in the INDSCAL model, individual weights are equal to 1 and matrix D, equivalent to identity, does not appear: in that sense MFA provides a solution to a generalization of the INDSCAL model). As a matter of fact, projection of each W_j onto a sub-space spanned by S vectors, each one being associated to a single vector of R^I, can be written as a decomposition identical to the scalar product form of the INDSCAL model: $W_j = \Sigma q(s, j) z_s z_s'$ with

– $z_s z_s'$, direction of $R^{I \times I}$, is associated to z_s, direction of R^I which is an unidimensional representation of individuals,

– $q(s, j)$, projection of W_j onto the s-order dimension, is the weight of dimension s for the group j.

The INDSCAL parameters estimation provided by MFA does not satisfy the same criterion as usual techniques: MFA maximizes sum of weights and not sum of squares.

We previously discussed advantages and drawbacks of the criterion. From an INDSCAL point of view, it seems that the technical advantages (for example getting weights always included in (0, 1) thus making their comparison easier) and relations to the other points of view are more important and useful than a theoretical property of the criterion.

2. Characteristics of the computer program (AFMULT)

2.1. Description of the algorithm

The program first performs separate PCA for each group. The categorical variables are represented by their indicator variables. These indicator variables are weighted in such a way that their analysis by PCA is equivalent to a MCA. Thus, in each group, MFA works as PCA with numerical variables and as MCA with categorical variables.

These analyses are useful to calculate the variables weights that permit a balance of groups in an overall analysis. They also permit the addition of the first factors of each group as supplementary variables.

The overall analysis is a PCA applied to the entire table in which each column of the group j is weighted by the inverse of the first eigen value of separate PCA of group j. The PCA provides, in addition, displays and aids to the interpretation that are specific to the structure of groups of variables.

Remark. The principal calculation of AFMULT researches eigen values of a symetrical matrix. Good algorithms are easy to find. AFMULT uses the Householder method first and then the QL implicit method.

2.2. Characteristics of the program

AFMULT is written in FORTRAN. The PC version requires 512 K memory.

Performing AFMULT requires a file which includes parameters introduced by keywords in free format. For example, NF = 5 means that the program will calculate 5 factors.

The data set must be a ASCII file. Each record represents an individual described by its name and its values for variables. Variables can be numerical or indicator variables. Rank orders of variables are independent of the structure of groups.

The number of groups of variables have no theoretical limit. The practical limit only depends on the number of variables of the active groups (for the micro-computer adaptation, the limit is 150 active variables).

2.3. Outputs

The program supplies all the results previously described, that is,

(a) classic results of PCA and MCA:
- displays of the individuals, of the numerical variables, of categories, of categorical variables,
- usual aids to the interpretation: quality of representation, contributions of lines and rows...

(b) results that are specific to the multiway structure:
- displays in which groups of variables are each represented by one point,
- estimation of the parameters of the INDSCAL model,
- the goodness of fit of the INDSCAL model from a global viewpoint, for each factor and each group,
- the importance of the common factors in each group,
- a simultaneous representation of the J clouds of individuals associated to each group,
- aids to the interpretation, in order to evaluate the quality of the displays both from a global viewpoint and point by point.

2.4. Availability of the program

AFMULT is principally issued as part of the LADDAD package. The integration of AFMULT inside the LADDAD package is due to M.O. Lebeaux. It is also possible to acquire the program separately. It is available from: ADDAD, 22 rue Charcot, F75013 Paris.

The cost of the whole LADDAD package is 18000 FF and 4000 FF for AFMULT acquired separately. ADDAD allows universities special conditions.

3. An example: Some red wines of Loire Valley

3.1. Presentation of the example

3.1.1. Introduction

This example comes from C. Asselin and R. Morlat (INRA Angers, France) who study the influence of soil upon quality of wines in the Loire Valley.

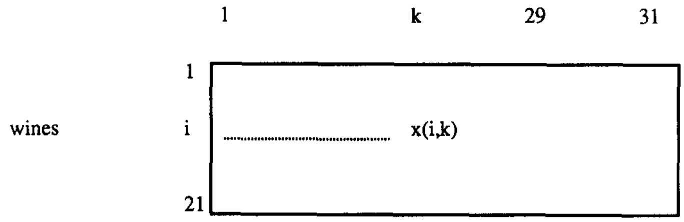

Fig. 2. Data set.

Initially, data were the following: 36 subjects judged 21 wines through a list of 29 questions (= variables). The questions are characteristics of wines: the subject have to estimate the intensity of these characteristics by the way of a scale including 5 ordered modalities (null or very weak, weak, middle, strong, very strong) coded from 1 to 5.

From these data, a matrix has been constructed (see Figure 2): for each wine and each variable, we computed the average of the intensities given by all the subjects.

Furthermore, we added 2 variables which describe origin of the wines: 'appellation' (the micro-region from which the wine comes from i.e. Saumur, Chinon, Bourgueil) and type of soil (reference, soil 2, soil 3, soil 4). One of the hypotheses of the agronomists is that the type of soil 'reference' provides the best wines. The 21 wines do not differ by other character than soil: they proceed from the same wine-plant and the same way of cultivation and wine-making.

$1 \leq k \leq 29$	: sensory characteristics,
$x(i, k)$	: average of the intensities of the characteristic k, given by the 36 subjects to the wine i,
$K = 30, 31$	: origin (appellation, type of soil),
$x\ (i,\ k)$	: code of the category (of the variable k) to which the wine i belongs,
remark	: the wines named T1 and T2 are, in fact, the same wine proposed twice (T1 at the beginning of the test; T2 at the end of the test).

3.1.2. List of sensory characteristics

Wines characteristics are ordered according to the classic phases of tasting: these phases constitute groups of variables.

- *Group 1: olfaction at rest.*
  Intensity of aroma,
  quality of aroma,
  fruit aroma,
  flower aroma,
  spice aroma.
- *Group 2: vision.*
  Intensity,
  colour (from orange to purple),
  surface (legs and tears on the glass).

- *Group 3: olfaction after agitation.* (direct way: the subject smell, the wine which is in the glass: retronasal way: the subject smells the wine which is in his mouth.)
  Intensity of aroma (direct way),
  quality of aroma (direct way),
  fruit aroma,
  flower aroma,
  spice aroma,
  vegetal aroma,
  phenol aroma,
  intensity of aroma (retronasal way),
  persistence of aroma,
  quality of aroma (retronasal way).
- *Group 4: gustation.*
  First intensity,
  acidity,
  astringency,
  alcohol (burning sensation),
  balance (between acidity, astringency and alcohol),
  velvety,
  bitterness,
  intensity (after some instants),
  harmony.
- *Group 5: general judgement.*
  Global quality,
  typicalness.
- *Group 6: origin of wines.* (Composed by two categorical variables.)
  appellation
  type of soil.

3.1.3. Problem

- Setting up a typology of wines based upon the whole testing process (without general judgement). In this typology, the influence of each of the 4 first sets of variables have to be balanced.
- Setting up common factors. What is common between the 4 ways of appreciation (olfaction at rest, vision, olfaction after agitation, gustation)?
- Comparison of factors issued of the separated analysis of each group of variables.
- Global comparison of groups. What are the groups of variables which give a similar typology of the wines?
- Comparison of typologies of wines provided by each group. If two wines are similar from one point of view (e.g. vision), are they similar from the other points of view (olfaction, gustation)?

Table 1
Eigen values of factor analysis applied to each group

	axis 1		axis 2		axis 3	
	eigen value	eigen value (percent)	eigen value (percent)	cumulated percent	eigen value (percent)	cumulated percent
1–olfaction at rest	2.24	44.8	30.3	75.2	16.3	91.5
2–vision	2.83	94.5	5.0	99.5	0.5	100.0
3–olfaction after agitation	4.70	47.0	24.8	71.8	10.5	82.3
4–gustation	5.64	62.7	19.9	82.6	7.5	90.1
6–general judgement	1.85	92.5	7.5	100.0		
6–origin	1.45	29.0	25.6	54.6	20.0	74.6

3.2. Comments about the output

AFMULT was applied to the whole set of data: groups 5 and 6 were introduced as supplementary elements: they do not contribute to the construction of axes.

The printout of AFMULT is voluminous, and cannot be reproduced in this text. We extract main results about each theme and show how these results take place in an interpretation.

3.2.1. Eigen valus of separate analyses (cf. Sections 1.4 and 2.1)

AFMULT begins with a separate factor analysis of each group. In the case of numerical variables (group 1 to 5), this analysis is PCA; in the case of categorical variables (group 6), this analysis is MCA.

Eigen values of separate analysis (Table 1) induce several comments.

Groups 1 and 5 are quite unidimensional. The others sensory groups have two (group 4) or three (group 1 and 3) important dimensions.

Group 6 is particular, since it is composed of two categorical variables. If the design of experiments would be perfect, variables 'appellation' and 'type of soil' would be independant; thus, MCA applied to this two variables would provide 5 equal eigen values. It is not the case here and the different types of soil are not equally distributed in each appellation. In fact this dependence is mainly due to the wine twice proposed (noted T1 and T2), which is the only wine belonging to the type of soil 4.

In the overall analysis, each group is weighted with the inverse of the first eigen value of its separate analysis (cf. Section 1.4.3). One meaning of this weighting is that the first principal components of each group have an equal a priori influence in the overall analysis (MFA can be viewed as working with variables or with principal components (p.c.) of separate analysis).

Table 2
Eigen values associated to the 3 first axis of MFA

	axis 1	axis 2	axis 3
eigen value	3.46	1.37	0.62
percentage	49.4	19.5	8.8

Remarks about weighting. The strongest first eigen values belong to groups 3 and 4 which contain the greatest numbers of variables. Without weighting, the first p.c. of these groups would dominate overall analysis, that is opposed to objectives.

It should be noted that weighting the groups with the inverses of the number of variables (the number of variables of group j is equal to the global inertia of this group since variables are centred and reduced) is not suitable: in such a case, the first p.c. of group 2 would dominate the global analysis.

3.2.2. Eigen values of global analysis

The eigen value associated to an axis can be interpreted as a measure of relationship between this axis and all the groups (Section 1.6.2).

The first eigen value is 3.46 (the possible maximum is equal to 4, number of groups): at least the first axis is strongly related to all the groups (see table 2).

Considering percentages of inertia, we restrict this methodological presentation to the two first axes.

3.2.3. Canonical correlation coefficients

The canonical correlation coefficient between axis s and group j indicates whether the structure defined by axis s may be induced by variables of group j (cf. Section 1.6.2.).

Axis 1 may be considered as a factor common to the four groups; axis 2 may be considered as a factor common only to three groups (see Table 3).

Axis 3 is common to groups 1 and 3. These coefficients prove its interest which was not perceived through eigen values. However it will not be commented here.

This result is essential: it shows that there are structures common to the groups (if it is not the case, it is not useful to study these groups simultaneously).

Table 3
Canonical correlation coefficients between active groups and the 3 first factors

groups	axis 1	axis 2	axis 3
olfaction at rest	0.89	0.96	0.89
vision	0.93	0.22	0.16
olfaction after agitation	0.97	0.89	0.90
gustation	0.95	0.87	0.30

Table 4
Inertia of the variables of each group along the two first factors of MFA (maximum = 1; cf. Section 1.4.3)

	axis 1	axis 2
olfaction at rest	0.78	0.62
vision	0.85	0.04
olfaction after agitation	0.92	0.47
gustation	0.90	0.24
Σ (= eigen value)	3.46	1.37
global judgement	0.62	0.25
origin	0.30	0.64

3.2.4. Relationship measure between axis and groups

The inertia of variables belonging to the same group j, projected onto axis s, can be interpreted as a relationship measure between group j and axis s (see Table 4). This measure completes the canonical correlation coefficients: it indicates whether structure induced by axis s corresponds, or not, to a high inertia dimension of the group j (cf. Section 1.6.2.).

The first axis is highly related to the four active groups. The structure expressed by this axis, is not only present in the four groups (high canonical correlation coefficients) but, in addition, corresponds to an important structure of each group.

The relatively strong relationship between 'origin' and axis 2 suggests that categories responsible of the dependence between 'appellation' and 'type of soil' are pointed out by this axis (cf. Section 3.2.1.).

3.2.5. Contributions of individuals to axis

The contribution of individual i to axis s, is the ratio between inertia (along axis s) of individual i and inertia associated to axis s.

As in any factor analysis, contributions indicate if an axis is only due to some individuals or expresses a global structure.

Two wines have a high contribution to the first axis: 32.8% + 26.4% = 59.2%. But taking into account the two first eigen values, the suppression of these two elements does not modify the rank order of this direction. This axis can be viewed as a global structure.

Two wines have a high contribution to the second axis: 39.3% + 29.7% = 69%. The suppression of these two elements can modify the rank order of this direction. Moreover, these two points correspond to the same wine, proposed twice to judges (T1 and T2). The interpretation of this axis must be seen as the particular case of one wine.

3.2.6. Variables display

As in PCA, the coordinate of variable k with respect to axis s is the correlation coefficient between variable k and axis s.

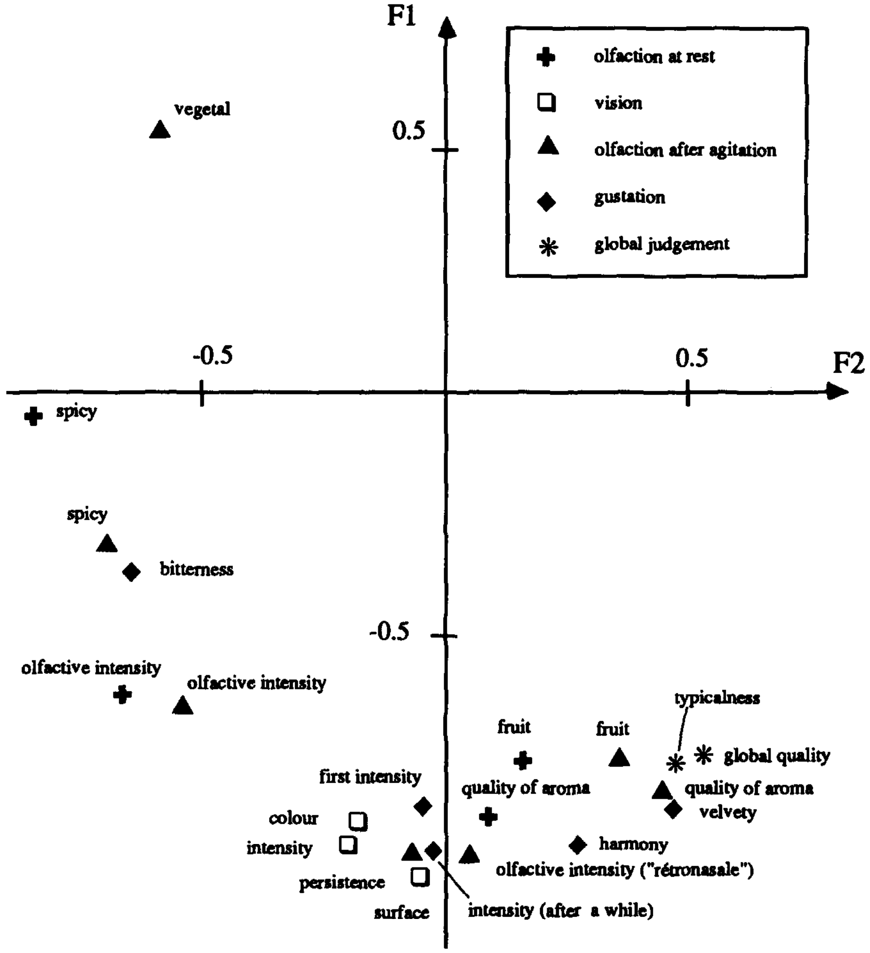

Fig. 3. Graphical display of variables in the principal plane of MFA.

As shown in Figure 3, variables which are highly correlated to axis 1 are:

Olfaction at rest	: quality of aroma, fruit aroma,
Vision	: surface (legs and tears), intensity, colour (purple),
Olfaction after agitation	: persistence of aroma, intensity of aroma (retro - nasal way), quality of aroma (direct and retronasal way),
Gustation	: intensity after some instants, harmony, first intensity, velvety,
General judgment	: global quality, typicalness.

This first axis expresses concepts often involved in the words strength and harmony, which have positive connotations. Usually, these two words are not synonymous but, with regard to these wines, are related.

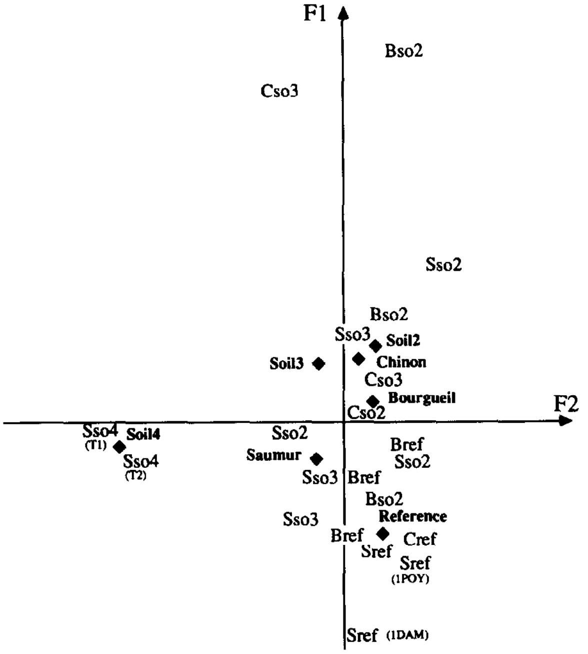

Fig. 4. Graphical display of wines in the principal plane of MFA. Each wine is represented by the first letter of its appellation (S = Saumur; B = Bourgueil; C = Chinon) followed by its type of soil (ref = reference; so2 = type of soil 2...). Four wines have also their name (1DAM, 1POY, T1, T2). Besides, the centre of gravity of each category is displayed.

Variables, the most correlated to the second axis, are:

Olfaction at rest	: spice aroma, intensity of aroma,
Olfaction after agitation	: spice aroma, vegetal aroma, intensity of aroma,
Gustation	: bitterness.

This second axis is mainly due to one wine (twice proposed) which has a olfactive characteristic associated to words 'vegetal' or 'spice'. It must be noted that the aromas 'underwood', 'mushroom', were spontaneously written on questionnaires concerning this wine. These data confirm the interpretation of this axis as an olfactive peculiarity.

3.2.7. Display of individuals and categories

The individuals display given by AFM, can be interpreted as those of PCA.

Thus, as it is shown figure 4, wine 1 DAM was considered as the strongest and the most harmonious one. At the opposite, wines Cmi3 et Bmi2, having the highest coordinates, were perceived neither strong nor harmonious.

A category of a categorical variable is represented by means of the centre of gravity of individuals which possess this category.

The categories 'Saumur', 'Chinon' and 'Bourgueil' are near the origin of axis: appellation is not related to the main factors of variability of these wines.

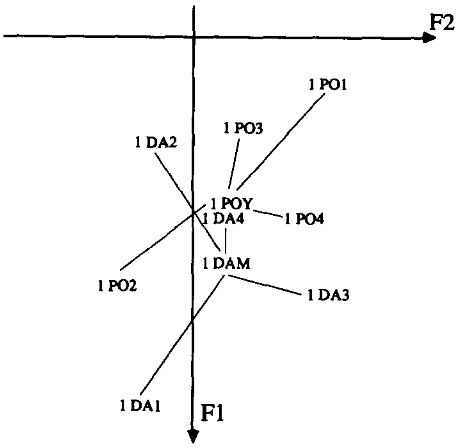

Fig. 5. Superposed display of the wines 1DAM and 1POY, defined with respect to each of the four active groups of variables and the whole set of variables as well. 1DAM, 1POY: wines perceived with respect to the whole set of variables; 1DA1, 1PO1: wines perceivedc with respect to the variables of group 1.

The category 'reference soil', is far with respect to axis 1: this type of soil has a priori, the best agronomical potential. This hypothesis is confirmed by the place of this point.

Remark. In MFA, numerical and categorical variables can be introduced simultaneously as active ones. This possibility is not used here.

3.2.8. Superposed display (cf. Section 1.7)

To each group j corresponds a representation of individuals noted N_I^j. MFA provides a display of each N_I^j (for active groups only), superposed upon the previous individuals display. In this representation each wine appears by means of five poins: one point for each active group and one point from the previous display (Figure 4); these last points are confounded with the centre of gravity of the four first ones.

Table 5
Some values from the variables belonging to the group olfaction at rest, which are the most related to axis 1

	Quality of aroma	Fruit aroma
maximum	3.429	3.154
1 DAM	3.429	3.154
1 POY	3.107	2.731
mean	3.046	2.714

Table 6
Some values from the variables belonging to the group gustation, the most related to axis 1

	Velvety	First intensity	Harmony
maximum	3.286	3.676	3.786
1 POY	3.231	3.667	3.786
1 DAM	3.036	3.643	3.643
mean	2.674	3.166	3.148

Figure 5 is extracted from the superposed display. For simpleness, it is limited to two wines: 1 DAM and 1 POY (the strongest and the most harmonious ones).

Figure 5 suggests interpretations as the following:

- From the point of view of olfaction at rest, wine 1 DAM has been perceived as the strongest and the harmonious one, and wine 1 POY has been perceived as mean.
- On the other hand, from the point of view of gustation, 1 POY has been perceived as the strongest and the most harmonious one. But the difference between 1 POY and 1 DAM is smaller than before.

This information can be corroborated directly from data (Tables 5 and 6).

3.2.9. Global display of groups (cf. Section 1.8)

In this output (Figure 6), each group is represented by one point. Two kind of interpretations can be made.

A: aid to the interpretation of the global PCA. The coordinate of the group *j* with respect to axis *s* is the inertia of the variables of the group *j* along axis *s*.

According to this point of view, this display gives an illustration of Table 4.

Thus the high coordinate of the four active groups with respect to the first factor (strength and harmony) shows that this factor corresponds to an important direction of each group of variables.

The position of group 6 on axis 1 shows that origin of wines is weakly related to this factor.

The second factor is mainly due to olfaction (groups 1 and 3) and weakly to gustation. It is related to origin of wines (let us recall that the wine noted T1 or T2 is the only one which posseses category 'type of soil: 4').

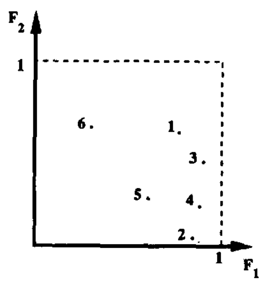

Fig. 6. Graphical display of groups in the principal plane of MFA.

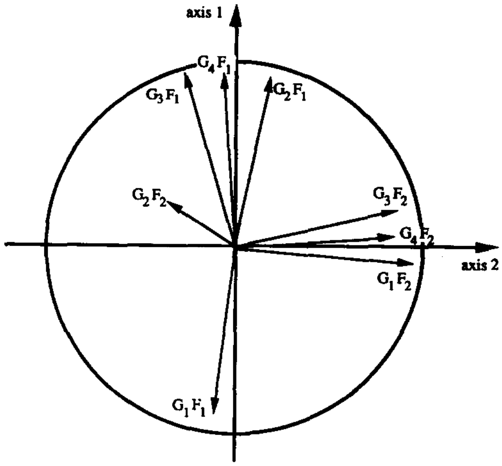

Fig. 7. Graphical display of the two first principal components of each group, in the principal plane of MFA (G1F2 = second component of group 1).

B: optimum display of groups. Figure 6 can be viewed as an orthogonal projection of the cloud of groups. In this cloud, two groups are close to one another if they induce the same structure upon individuals.

Group 6 is far from the others: origin of wines is weakly related to their main sensory characteristics.

The closest groups are olfaction after agitation and gustation; these characters are evaluated quite at the same moment, especially characters of retronasal olfaction and gustation are perceived simultaneously.

This display is precious when groups are numerous.

3.2.10. Display of principal components of each groups (cf. Section 2.1)

AFMULT considers the principal components of each group as supplementary variables.

Figure 7 shows the representation of these components upon the two first axis of MFA. This display can be superposed onto Figure 3.

The first principal component of each group is highly correlated with the first factor of MFA. We already saw (cf. Section 3.2.4) that the four groups have a common direction associated with an important inertia; we see now that this common factor is, for the four groups, close to the direction associated with the most important inertia.

The second factor of MFA is highly correlated with the second principal component of groups 1, 3 and 4.

We already saw that this factor was common to these three groups; this new results gives information about the importance of this factor in each group.

3.3. Conclusion

There is a structure common to the four groups of variables. This structure opposes wines strong and harmonious from the four points of view to wines neither strong nor harmonious. This common structure is also the main structure inside each group.

An other structure is common to three groups: the peculiar case of the wine noted T1 or T2 which presents a bad aroma perceived during olfaction and gustation but which does not influence vision.

Appellation is not related to these structures; the type of soil 'reference' characterized by the first structure, has the best potential.

This exemple shows that the main interest of MFA is an overall approach which includes all the aspects of the study of several groups of variables.

References

Carroll, J.D., A generalization of canonical correlation analysis to three or more sets of variables, in: Proceedings of the 76th Annual Convention of the American Psychological Association, (1968) 227–228.

Carroll, J.D., 1981, INDSCAL, in: Schiffman, Reynolds and Young (Eds.), *Introduction to Multidimensional Scaling* (Academic Press, New-York, 1981) pp. 371–389.

Escoufier, Y., L'analyse conjointe de plusieurs matrices, in: *Biométrie et temps* **59** (76 édité par la Société Française de Biométrie, 1980).

Escofier, B. and J. Pagès, *Analyses Factorielles Simples et Multiples: Objectifs, Méthodes, Interprétations*; (Dunod, Paris, 1990). p. 267.

Gower, J.C., Generalised procrustes analysis, *Psychometrika*, **40** (1) (1975) 33–51.