REVUE DE STATISTIQUE APPLIQUÉE

SÉBASTIEN LE DIEN
JÉRÔME PAGÈS

Analyse factorielle multiple hiérarchique

Revue de statistique appliquée, tome 51, nº 2 (2003), p. 47-73
<http://www.numdam.org/item?id=RSA_2003__51_2_47_0>

© Société française de statistique, 2003, tous droits réservés.

L'accès aux archives de la revue « Revue de statistique appliquée » (http://www.sfds.asso.fr/publicat/rsa.htm) implique l'accord avec les conditions générales d'utilisation (http://www.numdam.org/conditions). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM
Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
http://www.numdam.org/

Rev. Statistique Appliquée, 2003, **LI** (2), 47-73

ANALYSE FACTORIELLE MULTIPLE HIÉRARCHIQUE

Sébastien LE DIEN, Jérôme PAGÈS

Laboratoire de mathématiques appliquées
Pôle agronomique de Rennes
65 rue de Saint Brieuc
CS 84215 - 35042 Rennes cedex, France

RÉSUMÉ

Il n'est pas rare, dans l'analyse de tableaux de type individus × variables, de rencontrer une structure en partition sur les variables. On peut citer, pour exemple, les questionnaires structurés en thèmes. Plusieurs méthodes existent, qui permettent l'analyse de ce type de tableau, dont l'analyse factorielle multiple (AFM). Il arrive également que l'on veuille prendre en compte non pas une partition sur les variables mais une hiérarchie : c'est le cas des questionnaires structurés en thèmes et en sous-thèmes. Dans cet article, nous proposons une nouvelle méthode qui généralise l'AFM en prenant en compte cette structure sur les variables : l'analyse factorielle multiple hiérarchique (AFMH). L'AFMH équilibre (au sens de l'AFM) pour chaque nœud de la hiérarchie, les groupes de variables constituant ce nœud. L'AFMH fournit des sorties analogues à celles de l'AFM mais aussi des aides à l'interprétation spécifiques, telle que la représentation de chacun des nœuds d'une part et la représentation des individus décrits par chacun des nœuds, d'autre part.

Mots-clés : *Analyse factorielle multiple hiérarchique, analyse factorielle multiple, hiérarchie de variables.*

ABSTRACT

Frequently, in the analyses of individuals × variables tables, sets of variables are structured according to a partition. Several methods make the analysis of this type of data possible, as for instance Multiple Factor Analysis (MFA, Escofier and Pagès, 1998). In some cases, users want to take into account not only one partition on the variables but a hierarchy as it is often the case for questionnaires structured into topics and sub-topics. In this article, we propose a generalisation of MFA which takes into account this structure on the variables, the hierarchical multiple factor analysis (HMFA). To do so, HMFA balances the role of each group related to its node of the hierarchy, as in MFA. HMFA provides the same outputs as MFA but it provides also its own outputs such as a graphical display of the groups of variables related to the nodes of the hierarchy or a graphical display of the individuals described by each of the groups related to the nodes of the hierarchy.

Keywords : *Hierarchical multiple factor analysis, multiple factor analysis, hierarchy on sets of variables.*

1. Introduction

L'analyse factorielle de données dans lesquelles un même ensemble d'individus est décrit par plusieurs groupes de variables est un problème auquel le statisticien est fréquemment confronté : il est souvent naturel de regrouper les variables en fonction d'une typologie; dans le cas des données d'enquête, en particulier, les questionnaires sont généralement structurés en thèmes. Ce problème, aujourd'hui classique, peut-être traité par différentes méthodes, dont l'analyse procrustéenne (Gower, 1975), la méthode STATIS (Lavit, 1988) et l'analyse factorielle multiple (Escofier, Pagès, 1998).

Une difficulté s'ajoute lorsque l'on veut tenir compte d'une hiérarchie comprenant la notion de groupes et de sous-groupes de variables : pour reprendre l'exemple des questionnaires, les thèmes précités sont souvent eux-mêmes structurés en sous-thèmes. Considérer non pas une partition sur les variables mais plusieurs partitions emboîtées, dans une analyse factorielle de l'ensemble des variables, implique d'une part d'équilibrer les thèmes et les sous-thèmes et d'autre part de fournir des sorties (graphiques ou indicateurs numériques) permettant de prendre en compte dans l'interprétation la structure *a priori* sur les variables. Ce problème, bien que souvent rencontré, ne semble jamais avoir été abordé. Dans cet article, nous proposons une nouvelle méthode, l'analyse factorielle multiple hiérarchique (AFMH), qui généralise l'AFM au cas où les variables sont structurées selon une hiérarchie.

Les principes de base de l'AFMH, ses spécificités en tant qu'analyse factorielle ainsi que les types de résultats qu'elle fournit seront présentés au travers d'un exemple issu de l'analyse sensorielle.

Dans le cadre de cet article, nous considérons le cas où toutes les variables sont quantitatives et centrées-réduites. L'extension aux variables qualitatives et aux données mixtes est en cours d'étude.

2. Données utilisées pour illustrer la présentation de la méthode

96 étudiants du Pôle Agronomique de Rennes ont évalué et apprécié 6 purs jus d'orange choisis parmi les principales marques du marché (*cf.* tableau 1).

TABLEAU 1
Jus d'orange étudiés

Pampryl ambiant	Pampryl réfrigéré
Tropicana ambiant	Fruivita réfrigéré
Joker ambiant	Tropicana réfrigéré

Les variables disponibles se répartissent d'abord selon deux groupes : *jugements hédoniques* et *caractérisation.*

Les jugements hédoniques sont ceux des 96 étudiants qui ont noté chaque jus sur une échelle allant de 1 à 5, 1 signifiant que le jus n'a pas du tout été aimé et 5 indiquant qu'il a été particulièrement apprécié.

Le groupe *caractérisation* est composé de 16 variables que l'on peut subdiviser en 9 variables physico-chimiques d'une part et en 7 descripteurs sensoriels d'autre part (*cf.* figure 1) :

1. les variables physico-chimiques sont l'acidité globale, le pH brut, le pH après centrifugation du produit, le taux d'acide citrique, le taux en saccharose, le taux en glucose, le taux en fructose, le pouvoir sucrant et enfin le taux en vitamines C;
2. les variables sensorielles sont la typicité de l'odeur, les caractères sucré, amer, acide, pulpeux, l'intensité du goût et l'intensité de l'odeur. Ces variables sont les moyennes des notes attribuées par les 96 étudiants, chaque note variant sur une échelle allant de 1 à 5, où 1 signifie que l'intensité associée au descripteur est perçue comme faible et 5 signifie au contraire qu'elle est ressentie comme très élevée.

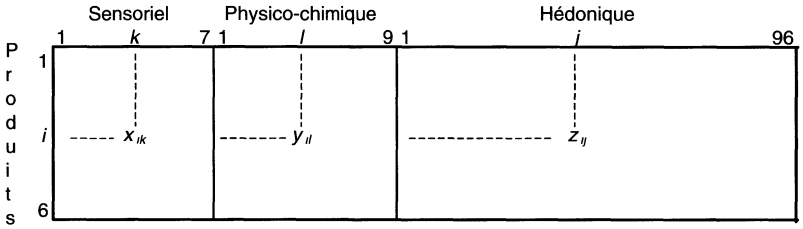

FIGURE 1
Le groupe de variables descriptives se décompose en variables physico-chimiques d'une part et en descripteurs sensoriels d'autre part

Un des enjeux de l'étude est la mise en relation des deux aspects *caractérisation* (variables physico-chimiques et sensorielles) d'une part et *jugements hédoniques* d'autre part.

Une AFM du jeu de données structuré selon ces deux aspects permet de les comparer à travers les structures qu'ils définissent sur l'ensemble des jus d'orange mais ne permet pas de distinguer les variables physico-chimiques des descripteurs sensoriels au sein de la *caractérisation.*

Par ailleurs, une AFM du jeu de données structuré selon les trois groupes physico-chimique, sensoriel et hédonique ne permet pas une comparaison directe des deux aspects *caractérisation* et *jugements hédoniques.*

L'AFMH va permettre une telle comparaison, tout en tenant compte de la structure interne du groupe des variables descriptives.

3. Notations et définitions

On note X le tableau de données individus$\times$variables, I l'ensemble des individus, D la matrice diagonale de leur poids (on désigne par p_i le poids de l'individu i; $\sum_{i \in I} p_i = 1$) et K l'ensemble des variables. Dans l'AFM, ces K variables

sont structurées suivant une partition unique composée de J groupes; en AFMH, on ne considère non plus une seule partition sur les variables mais plusieurs partitions emboîtées.

Soient deux partitions $\mathcal{P}_1 = \{A_1, \ldots, A_{J_1}\}$ et $\mathcal{P}_2 = \{B_1, \ldots, B_{J_2}\}$ sur un même ensemble de variables K; $\mathcal{P}_1$ est emboîtée dans $\mathcal{P}_2$, ce que l'on note $\mathcal{P}_1 \subset \mathcal{P}_2$, si tout élément de $\mathcal{P}_1$ est inclus dans un et un seul élément de $\mathcal{P}_2$:

$$\forall j_1 = 1, \ldots, J_1, \ \exists! j_2 / \ A_{j_1} \subset B_{j_2}.$$

Plus généralement, soient P partitions sur un même ensemble de variables K; une suite de partitions emboîtées à P éléments est un ensemble de partitions $\{\mathcal{P}_1, \ldots, \mathcal{P}_P\}$ noté S_P, tel que :

$$\mathcal{P}_1 \subset \ldots \subset \mathcal{P}_P.$$

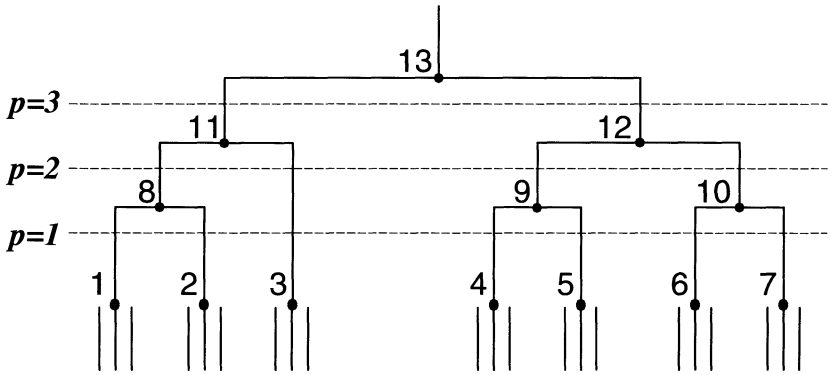

FIGURE 2

Représentation d'une suite de partitions emboîtées à trois éléments : la première (resp. deuxième; resp. troisième) coupure horizontale, à l'indice de niveau 1 (resp. 2; resp. 3), correspond à la partition $\mathcal{P}_1$ (resp. $\mathcal{P}_2$; resp.$\mathcal{P}_3$), constituée de 7 (resp. 4; resp. 2) sous-ensembles de variables (éléments terminaux de l'arbre associé à S_P). Les groupes de variables sont repérés soit par leur numéro de nœud (de 1 à 12) soit par leur numéro de groupe associé à leur partition (cf. figure 4)

Cette suite S_P, en tant que hiérarchie sur les variables de K, peut être représentée sous la forme d'un arbre hiérarchique dont les éléments terminaux correspondent chacun à une variable et dont les indices de niveaux sont entiers et correspondent chacun à une partition (*cf.* figure 2); par la suite on utilisera indifféremment les termes d'*indice de niveau* et de *niveau de partition*.

Dans la figure 2, les groupes de variables sont repérés par leur numéro de nœud, le dernier nœud correspondant à l'ensemble des variables K; il peut être utile, comme on le verra par la suite, de les repérer par leur numéro de groupe associé à leur partition.

Pour un niveau de partition donné, l'appartenance d'une variable à une classe est définie à l'aide de la fonction f_p. Par définition, f_p est l'application qui à la variable k associe l'indice du groupe auquel elle appartient au niveau de partition p.

Dans l'étude concernant les jus d'orange, un des objectifs est de les décrire au travers des deux aspects *caractérisation* et *jugements hédoniques*, en tenant compte de la distinction entre variables physico-chimiques et sensorielles au sein de la *caractérisation*. Pour cela, on considère une suite de partitions emboîtées à deux éléments S_2 (*cf.* figure 3) telle que :

1. $\mathcal{P}_1$ est la partition constituée des 3 groupes de variables physico-chimiques, sensorielles et hédoniques;
2. $\mathcal{P}_2$ est la partition constituée des 2 groupes *caractérisation* et *jugements hédoniques*, où le groupe *caractérisation* est l'union des 2 groupes physico-chimique et sensoriel de $\mathcal{P}_1$.

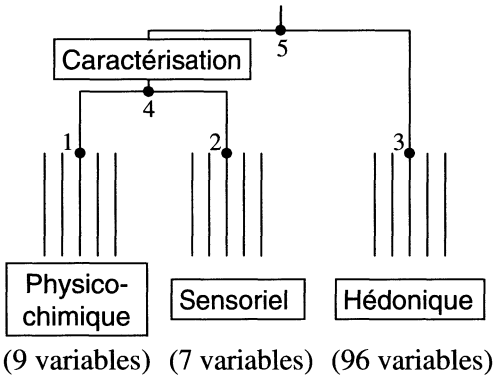

FIGURE 3

$\mathcal{P}_1$ est la partition constituée des 3 groupes physico-chimique (groupe 1, nœud 1), sensoriel (groupe 2, nœud 2) et hédonique (groupe 3, nœud 3); $\mathcal{P}_2$ est la partition constituée des 2 groupes caractérisation (groupe 1, nœud 4) et jugements hédoniques (groupe 2, nœud 3)

Si k est la variable sensorielle *caractère pulpeux* (*cf.* figure 3), alors $f_1(k) = 2$ et $f_2(k) = 1$ (le nœud 4 correspond au groupe 1 de la partition $\mathcal{P}_2$).

On note J_p l'ensemble des groupes de la partition $\mathcal{P}_p$, K_j^p l'ensemble des variables du groupe j au niveau de partition p et X_j^p le tableau associé à K_j^p (*cf.* figure 4) :

$$\forall p = 1, \ldots, P,\ K = \bigcup_{j=1}^{J_p} K_j^p.$$

Les symboles I, J, K, J_p et K_j^p désignent à la fois l'ensemble et son cardinal.

À toute suite de partitions S_P, on associe généralement deux partitions supplémentaires :

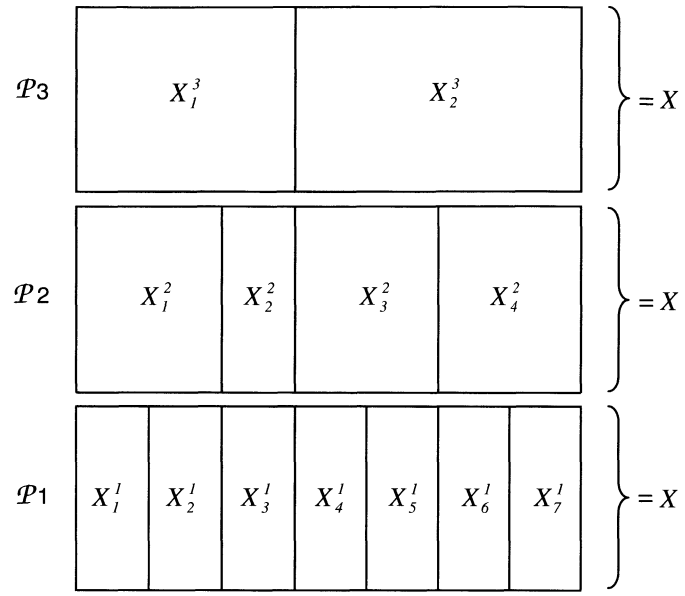

FIGURE 4
Représentation d'une suite de partitions à trois niveaux, en termes de sous-tableaux (cf. figure 2) : les groupes de variables sont repérés par leur numéro de groupe associé à leur partition

1. la partition la plus fine (niveau d'agrégation 0), où chaque variable est un groupe;
2. la partition la plus grossière (niveau d'agrégation $P+1$), où toutes les variables appartiennent à un même groupe.

4. Principe, équivalences

4.1. Principe

Prendre en compte la suite de partitions implique d'équilibrer le rôle de chaque groupe au sein de sa partition et ce à tous les niveaux de partition. Dans notre application, on veut obtenir une description des jus à travers les aspects *caractérisation* et *jugements hédoniques*, qui équilibre ces deux points de vue entre eux mais aussi, au sein de la *caractérisation*, le rôle des variables physico-chimiques d'une part et sensorielles d'autre part. En AFMH, pour obtenir un tel équilibre, on réalise une succession d'AFM à chaque nœud de l'arbre associé à la suite S_P, que l'on parcourt de l'indice de niveau le plus bas à l'indice le plus élevé. De cette façon, grâce à l'AFM, on pondère les groupes de variables issus de la décomposition d'un nœud de façon à équilibrer leur influence, tout en tenant compte de leur structure interne. L'AFMH est une analyse du tableau de données X structuré selon la hiérarchie S_P, qui peut alors se décomposer de la façon suivante :

Étape 1 : on réalise une analyse factorielle de chacun des groupes de variables associés au premier niveau de partition; on stocke les facteurs obtenus. Dans l'exemple, cela revient à réaliser une ACP de chacun des groupes de variables associés aux nœuds 1 à 7 (*cf.* figures 2 et 5).

Étape 2 : à chaque nœud de l'indice de niveau 2 (nœuds 8, 9 et 10, *cf.* figure 2), on réalise une AFM dans laquelle les groupes de variables sont ceux qui constituent le nœud (pour le nœud 8, il s'agit des groupes de variables 1 et 2). En pratique, à ce niveau, les variables d'un groupe sont remplacées par les facteurs issus de son ACP. À leur tour, les facteurs de ces AFM sont stockés : ce sont eux qui représenteront les groupes de variables associés aux nœuds dans les AFM de l'étape suivante.

(...)

Étape p : on réitère l'étape $p-1$ jusqu'à l'indice de niveau le plus élevé (*cf.* figure 5); à chaque étape, les variables sont remplacées par les facteurs issus des AFM réalisées à l'étape précédente.

Lors de la dernière étape (*i.e.* l'étape P), les principaux facteurs de variabilité de l'AFMH sont obtenus à l'issue de l'AFM des groupes de facteurs principaux associés aux nœuds de l'indice de niveau P (dans l'exemple de la figure 2, on réalise une AFM des groupes de facteurs associés aux nœuds 11 et 12). Cette étape, qui correspond donc à une ACP pondérée du tableau de données X, constitue l'analyse globale de l'AFMH.

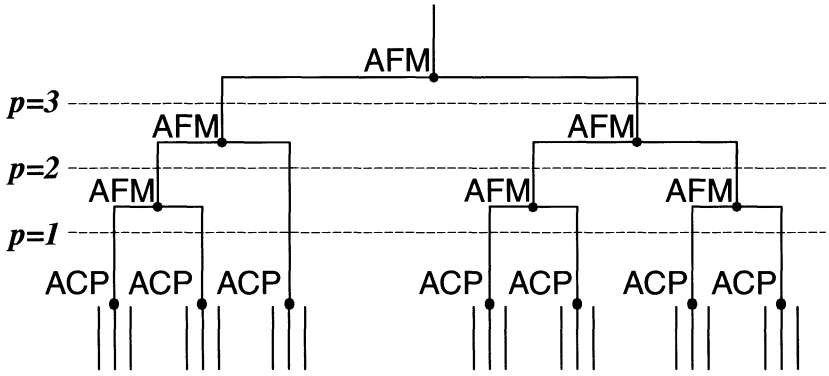

FIGURE 5

*Analyses factorielles réalisées au cours de l'*AFMH *(cf. figure 2); à l'issue de la dernière* AFM*, réalisée sur les groupes de facteurs principaux associés au dernier niveau de partition, on obtient les principaux facteurs de variabilité de l'*AFMH

On utilise pour l'analyse globale de l'AFMH les notations usuelles de l'ACP. On note u_s l'axe principal de rang s du nuage des individus et F_s le facteur associé (*i.e.* le vecteur des coordonnées des individus sur u_s); on note ν_s l'axe principal de rang s du nuage des variables et G_s le vecteur des coordonnées des variables sur ν_s. Enfin, on note λ_s la s^e valeur propre.

Le calcul des principaux facteurs de variabilité de l'AFMH passe par la réalisation d'ACP (pour les nœuds correspondant à $\mathcal{P}_1$) et d'AFM (pour les autres nœuds) séparées. Les valeurs propres de ces analyses permettent de se faire une idée précise des structures d'inertie induites par chacun des groupes de variables (*cf.* tableau 2).

TABLEAU 2
3 premières valeurs propres des analyses partielles

	Groupe		Val. propres			% inertie			% cumulés		
	$\mathcal{P}_1$	$\mathcal{P}_2$	F1	F2	F3	F1	F2	F3	F1	F2	F3
Variables physico-chimiques	1	.	6.213	1.410	1.046	69.0	15.7	11.6	69.0	84.7	96.3
Descripteurs sensoriels	2	.	4.744	1.333	0.820	67.8	19.0	11.7	67.8	86.8	98.5
Aspect caractérisation	.	1	1.791	0.476	0.294	61.2	16.3	10.0	61.2	77.5	87.5
Aspect hédonique	3	2	34.028	19.369	15.892	35.4	20.2	16.6	35.4	55.6	72.2

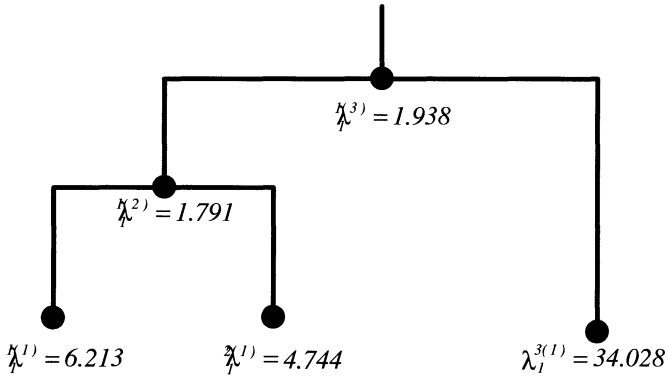

FIGURE 6
Premières valeurs propres issues des analyses partielles;
$\lambda_1^{1(3)}$ *correspond à la* 1e *valeur propre de l'*AFMH *du tableau de données structurées selon la figure 3 : l'*AFMH *est présentée comme une* ACP *pondérée de l'ensemble des variables, autrement dit comme une* ACP *pondérée des variables regroupées suivant la partition la plus grossière, où toutes les variables appartiennent à un même groupe (niveau d'agrégation* $P + 1 = 3$*, d'où la notation)*

Dans l'exemple, d'après le tableau 2 et la figure 6 :

1. Les analyses séparées des groupes de variables physico-chimiques et sensorielles mettent en évidence pour chaque groupe une direction d'inertie prépondérante sans pour autant que l'on puisse considérer ces groupes comme unidimensionnels.
2. Le groupe des jugements hédoniques est nettement plus multidimensionnel que les deux autres, ce qui est attendu étant donné son nombre relativement important de variables et leur nature (les jugements hédoniques pouvant être très divers).
3. Les fortes différences entre les premières valeurs propres des analyses séparées rendent indispensable la pondération des groupes au sein d'une analyse globale.

4.2. Équivalence avec une ACP

L'AFM étant une ACP pondérée, l'analyse globale de l'AFMH peut être présentée comme une ACP dans laquelle les variables sont pondérées de telle sorte que le rôle de chaque groupe soit équilibré au sein de sa partition, et ce à tous les niveaux de partition. Dans l'algorithme présenté, le fait de travailler à partir des facteurs principaux non normés assure l'équilibre entre les groupes de variables. Il reste maintenant à expliciter le calcul des poids des variables de base à chaque étape.

Rappelons que l'AFM pondère les variables d'un même groupe j par $1/\lambda_1^j$, où, suivant les notations de l'AFM, λ_1^j désigne la première valeur propre de l'analyse factorielle du groupe j.

La détermination des poids des variables en AFMH se déroule en P étapes, de la façon suivante :

Soit $m_0(k)$ le poids initial accordé à la variable k; généralement toutes les variables ont un poids initial égal à 1.

Étape 1 : $\forall j \in \{1, \ldots, J_1\}$ on appelle $\lambda_1^{j(1)}$ la première valeur propre issue de l'ACP du groupe j de $\mathcal{P}_1$. Cette valeur propre peut aussi être notée $\lambda_1^{f_1(k)}$ si la variable k appartient au groupe j de $\mathcal{P}_1$. Pour tout $k \in \{1, \ldots, K\}$, on calcule :

$$m_1(k) = m_0(k) \times \frac{1}{\lambda_1^{f_1(k)}}.$$

$$(\ldots)$$

Étape p : $\forall j \in \{1, \ldots, J_p\}$ on appelle $\lambda_1^{j(p)}$ la première valeur propre issue de l'ACP pondérée du groupe j de $\mathcal{P}_p$ par l'application m_{p-1}. Cette valeur propre peut aussi être notée $\lambda_1^{f_p(k)}$ si la variable k appartient au groupe j de $\mathcal{P}_p$. Pour tout $k \in \{1, \ldots, K\}$, on calcule :

$$m_p(k) = m_{p-1}(k) \times \frac{1}{\lambda_1^{f_p(k)}} = m_0(k) \times \prod_{h=1}^{p} \frac{1}{\lambda_1^{f_h(k)}}.$$

(...)

Étape P : Le poids $m(k)$ de la variable k dans l'AFMH est donnée par la formule suivante :

$$m(k) = m_P(k) = m_0(k) \times \prod_{h=1}^{P} \frac{1}{\lambda_1^{f_h(k)}}.$$

À un niveau de partition p donné et pour un groupe de variables j donné ($j = 1, \ldots, J_p$), le rôle de chacun des sous-groupes constituant le groupe j au niveau de partition $p-1$ est équilibré : aucun d'eux ne peut être prépondérant dans la première direction d'inertie du nuage décrit par les variables du groupe j. Le rôle de chacun des groupes est ainsi équilibré par nœud et ce quel que soit le niveau de partition.

À la fonction de pondération m on associe $M \in \mathcal{M}(K \times K)$ ($\mathcal{M}(K \times K')$ désigne l'ensemble des matrices de dimension $K \times K'$), la matrice diagonale des poids des variables de l'AFMH, définie par :

$$\forall k = 1, \ldots, K,\ M(k,k) = m(k).$$

À la fonction m_p on associe $M_j^p \in \mathcal{M}(K_j^p \times K_j^p)$, la matrice diagonale des poids des variables du groupe j au niveau de partition p, définie par :

$$\forall k = 1, \ldots, K_j^p,\ M_j^p(k,k) = m_p(k).$$

Remarques.

1. Dans notre programme, les calculs se font systématiquement à partir du tableau des données centrées réduites. Le fait d'affecter le poids initial 1 à chacune des variables revient à accorder, dans chaque groupe de $\mathcal{P}_1$, la même importance à chaque variable, quelle que soit sa dispersion. Si l'on désire privilégier au sein d'un même groupe de $\mathcal{P}_1$ les variables les plus dispersées, on accorde à chacune, comme classiquement, un poids initial égal à sa variance, soit $m_0(k) = s_k^2$, où s_k^2 désigne la variance de la variable k.

2. L'AFM est une AFMH où l'on ne considère qu'un seul niveau de partition ($P = 1$).

3. Le calcul de la pondération des variables de base tient compte du cas particulier où un même groupe de variables appartient à deux partitions successives $\mathcal{P}_p$ et $\mathcal{P}_{p+1}$: dans ce cas, au niveau $p+1$, la valeur de la première valeur propre associée à l'analyse factorielle du groupe est égale à 1.

4. Le calcul des pondérations des variables se révèle être plus simple en raisonnant par partition même si le raisonnement par nœud paraît plus intuitif pour d'autres aspects de l'AFMH.

4.3. Équivalence avec une AFM

La présentation de l'analyse globale du tableau de données X associé à S_P par l'AFMH, comme une ACP pondérée, ne doit pas faire oublier que l'AFMH est aussi une AFM des groupes de variables associés aux nœuds de l'indice de niveau P. Dans le cas des jus d'orange, cette AFM permet de répondre au problème de la comparaison des aspects *caractérisation* et *jugements hédoniques*; elle est essentielle à l'interprétation des résultats, car c'est à ce niveau que la structure sur les variables est la plus générale et que l'on aborde le problème dans son aspect le plus global. En AFMH, l'interprétation des résultats se fait de façon «descendante» , en partant du niveau le plus synthétique.

5. L'AFMH vue comme une analyse factorielle

Le cœur de l'AFMH repose sur une analyse factorielle pondérée du tableau de données X. Cette analyse, qui constitue l'analyse globale de l'AFMH, est une ACP pondérée lorsque toutes les variables sont quantitatives. À ce titre, l'AFMH fournit les résultats classiques de l'ACP, à savoir des représentations graphiques du nuage des individus N_I et du nuage des variables N_K ainsi que les aides à l'interprétation associées. Rappelons que les notations usuelles de l'ACP, u_s (resp. ν_s), F_s (resp. G_s) et λ_s, sont utilisées pour l'AFMH.

5.1. Nuage des variables N_K

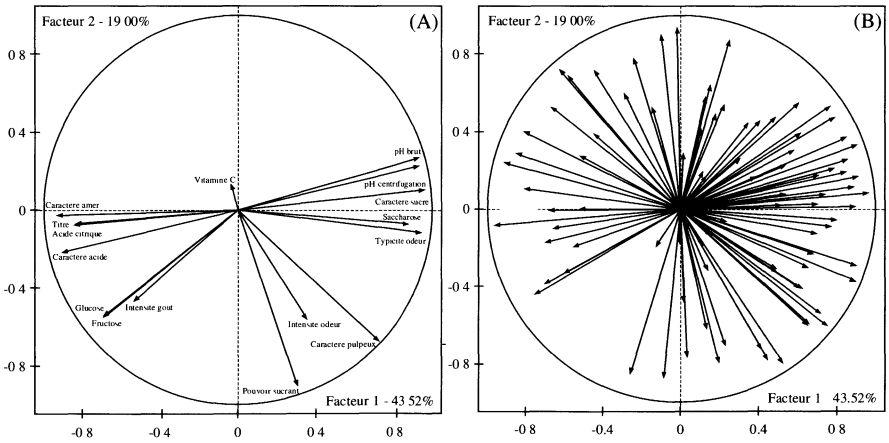

FIGURE 7
Représentation des variables, physico-chimiques et sensorielles d'une part (A) *et hédoniques d'autre part* (B), *dans l'analyse globale de l'*AFMH *(par souci de lisibilité, les variables de chacun de ces 2 groupes ont été représentées sur des graphiques différents)*

Dans l'exemple des jus d'orange, la qualité de représentation des variables physico-chimiques et sensorielles est globalement bonne (*cf.* figure 7, A) : on retrouve une opposition (*acide, amer*)/*sucré*, classique en analyse sensorielle. Cette opposition est corrélée avec le pH, ce qui apporte une validation par le sens, essentielle dans cette analyse qui ne comporte que 6 individus.

Par ailleurs, même si une légère majorité des étudiants est corrélée positivement avec le premier facteur (*cf.* figure 7, B), les jugements hédoniques individuels sont très dispersés : il n'y a pas de consensus entre les étudiants concernant leurs préférences; certains préfèrent les jus sucrés et peu acides et d'autres les produits acides et peu sucrés.

Remarque. – Sous l'action de la température et du pH, le saccharose en solution se décompose en glucose et fructose : ce phénomène d'hydrolyse se retrouve sur le premier plan factoriel avec l'opposition (*saccharose*, *pH*) et (*glucose*, *fructose*).

5.2. Nuage des individus N_I

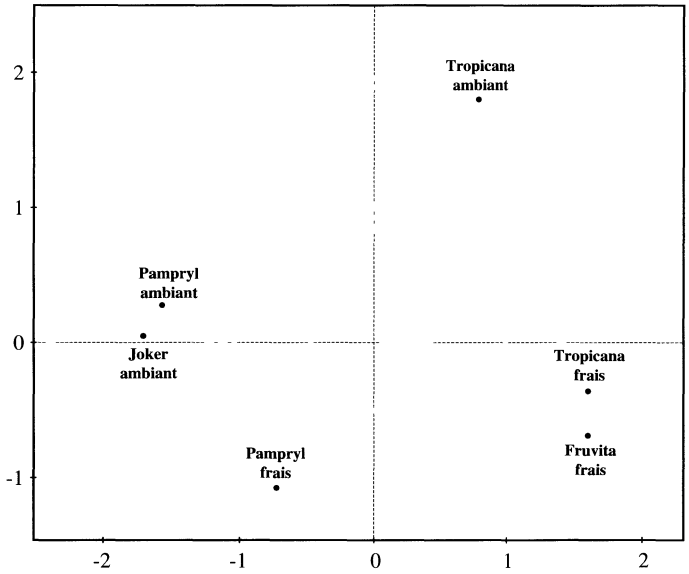

FIGURE 8
*Représentation des individus par l'*AFMH

Le premier facteur F_1 (*cf.* figure 8) oppose le Pampryl ambiant, le Joker ambiant et le Pampryl frais aux trois autres jus; ces produits ont été perçus comme plutôt acides et amers alors que les trois autres ont été perçus comme plutôt sucrés. La deuxième bissectrice, quand à elle, oppose les produits «frais» perçus comme pulpeux aux produits «ambiants» perçus comme peu pulpeux.

D'un point de vue formel, l'interprétation conjointe des graphiques des nuages des individus et des variables s'appuie sur les relations de transition.

6. Sorties spécifiques de l'AFMH

Considérer une suite de partitions emboîtées sur les variables engendre des sorties supplémentaires (graphiques, aides à l'interprétation), que nous présentons dans la partie qui suit.

6.1. Représentation superposée des nuages d'individus partiels

Principe. – Par définition, un individu partiel associé au groupe j de $\mathcal{P}_p$ est un individu décrit par les seules variables de K_j^p. Dans l'AFM, chaque individu possède autant de représentations partielles qu'il y a de groupes de variables; la représentation superposée de ces individus et de leurs représentations partielles est telle que chaque individu est au centre de gravité de ses représentations partielles.

En AFMH le problème de la représentation graphique se complique, puisque l'on considère non plus une seule mais plusieurs partitions; leur prise en compte suggère autant de représentations partielles qu'il y a de groupes de variables distincts toutes partitions confondues ou, de façon équivalente, qu'il y a de nœuds dans la suite de partitions S_P représentée sous forme d'arbre hiérarchique.

Dans l'application présentée, on peut considérer que les jus d'orange sont décrits suivant quatre groupes de variables; en effet, on dispose de notes hédoniques d'une part et de variables caractérisant les produits d'autre part, ce dernier groupe pouvant être subdivisé en deux sous-groupes composés de variables physico-chimiques d'un côté et de descripteurs sensoriels de l'autre (*cf.* figure 3). On souhaite comparer les jus suivant les aspects *caractérisation* et *jugements hédoniques*, et, au sein de la *caractérisation*, suivant les aspects sensoriels et physico-chimiques. D'où la nécessité d'un graphique dans lequel chaque individu est représenté par autant de points qu'il y a de nœuds dans la hiérarchie.

Le principe de la représentation superposée des P nuages d'individus partiels généralise celui de l'AFM : pour représenter l'ensemble des individus partiels, on les plonge dans un même espace $\mathbb{R}^K$ (espace des individus décrits par l'ensemble des variables). Pour ce faire, l'espace $\mathbb{R}^K$ est décomposé successivement en sous-espaces orthogonaux de plus en plus petits, suivant la décomposition induite par la suite S_P, parcourue de façon descendante (*i.e.* de $\mathcal{P}_P$ à $\mathcal{P}_1$) :

Étape 1 : Partant du niveau de partition P le plus général, on décompose $\mathbb{R}^K$ en J_P sous-espaces orthogonaux deux à deux et isomorphes aux espaces $\mathbb{R}^{K_j^P}$:

$$\mathbb{R}^K = \oplus_{j=1,\ldots,J_P} \mathbb{R}^{K_j^P}.$$

Étape 2 : De même que le j^e groupe de variables au niveau de partition P se décompose en sous-groupes de $\mathcal{P}_{P-1}$, $\mathbb{R}^{K_j^P}$ peut se décomposer en

$$\mathbb{R}^{K_j^P} = \oplus_{j' \in \mathcal{D}_j^P} \mathbb{R}^{K_{j'}^{P-1}},$$

où $\mathcal{D}_j^p$ désigne à la fois l'ensemble des indices des sous-groupes de $\mathcal{P}_{p-1}$ issus de la décomposition du groupe j de $\mathcal{P}_p$ et son cardinal, la formule ci-dessus correspondant au cas où $p = P$.

On en déduit une décomposition de $\mathbb{R}^K$ en somme directe de J_{P-1} sous-espaces orthogonaux deux à deux et isomorphes aux espaces $\mathbb{R}^{K_j^{P-1}}$:

$$\mathbb{R}^K = \oplus_{j=1,\ldots,J_P} \oplus_{j_1 \in \mathcal{D}_j^P} \mathbb{R}^{K_{j1}^{P-1}} = \oplus_{j=1,\ldots,J_{P-1}} \mathbb{R}^{K_j^{P-1}}.$$

$$(\ldots)$$

Étape P : Suivant le même principe, $\mathbb{R}^{K_j^2}$ se décompose en

$$\mathbb{R}^{K_j^2} = \oplus_{j' \in \mathcal{D}_j^2} \mathbb{R}^{K_{j'}^1},$$

d'où une dernière décomposition de $\mathbb{R}^K$ en somme directe de J_1 sous-espaces orthogonaux deux à deux et isomorphes aux espaces $\mathbb{R}^{K_j^1}$:

$$\mathbb{R}^K = \oplus_{j=1,\ldots,J_P} \oplus_{j_1 \in \mathcal{D}_j^P} \cdots \oplus_{j_{P-1} \in \mathcal{D}_{j_{P-2}}^2} \mathbb{R}^{K_{j_{P-1}}^1} = \oplus_{j=1,\ldots,J_1} \mathbb{R}^{K_j^1}.$$

Dans $\mathbb{R}^K$, les coordonnées du nuage des individus partiels associés au j^e groupe de $\mathcal{P}_p$ sont contenues dans la matrice de $\mathcal{M}(I, K)$ dans laquelle X_j^p est complétée par des 0 (rappelons qu'en AFMH, de même qu'en ACP, les variables sont toujours centrées). Cette représentation simultanée est artificielle et inutilisable directement, mais sert de base à une représentation simultanée obtenue par projection sur les axes principaux u_s issus de l'analyse globale de N_I par l'AFMH.

Graphique. Pour représenter simultanément les nuages des individus et des individus partiels deux solutions sont envisageables (*cf.* figure 9).

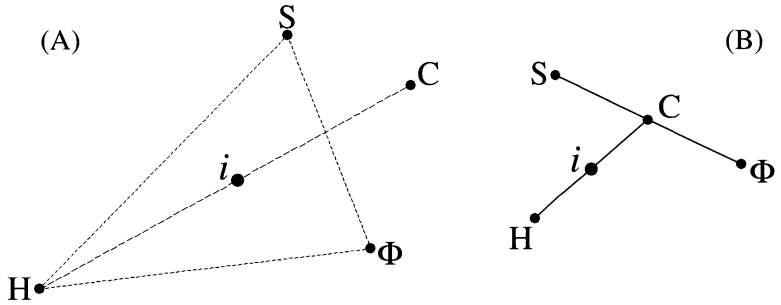

FIGURE 9

Représentation des individus partiels associés aux groupes physico-chimique (ϕ), sensoriel (S), caractérisation (C), et jugements hédoniques (H) :
(A) *l'individu i est à la fois à l'isobarycentre de $\{H, C\}$ et de $\{H, \phi, S\}$;*
(B) *i est à l'isobarycentre de $\{H, C\}$ et C est à l'isobarycentre de $\{S, \phi\}$*

Une première solution consiste à représenter les individus partiels de telle sorte que, quel que soit le niveau de partition considéré, chaque individu soit au centre de gravité de ses représentations partielles; on bénéficie alors de la propriété barycentrique de l'AFM, à chaque niveau de partition :

$$\forall i = 1, \ldots, I,\ \forall p = 1, \ldots, P,\ i \text{ est l'isobarycentre de } \{i_j^p,\ j = 1, \ldots, J_p\},$$

où i_j^p est la représentation partielle de l'individu i associée au j^e groupe de variables de $\mathcal{P}_p$ (rappelons qu'en AFM, on désigne par i^j la représentation partielle de l'individu i associée au groupe j).

Une telle représentation graphique (*cf.* figure 9, A) ne rend pas compte de l'équilibre entre les rôles joués par les variables sensorielles et physico-chimiques dans la caractérisation des jus : la représentation partielle associée à l'aspect *caractérisation* n'est pas à l'isobarycentre des représentations associées aux aspects *sensoriel* et *physico-chimique*.

La seconde possibilité consiste à représenter les individus partiels de telle sorte que :

1. chaque individu est au centre de gravité de ses représentations partielles au niveau de partition P;

2. chaque individu partiel associé à un groupe j de $\mathcal{P}_p$ ($1 < p \leqslant P$) est au centre de gravité de « ses » représentations partielles, induites par la décomposition de j en sous-groupes de $\mathcal{P}_{p-1}$.

   Autrement dit, $\forall i = 1, \ldots, I$,

1. i est l'isobarycentre de $\{i_j^P,\ j = 1, \ldots, J_P\}$;

2. $\forall p = 2, \ldots, P,\ \forall j = 1, \ldots, J_p,\ i_j^p$ est l'isobarycentre de $\{i_{j'}^{p-1},\ j' \in \mathcal{D}_j^p\}$.

Nous choisissons cette seconde solution pour laquelle un individu partiel est au centre de gravité des « ses » représentations partielles. Une telle représentation graphique (*cf.* figure 9, B) rend bien compte de l'équilibre entre les rôles joués par les sous-groupes j' de $\mathcal{P}_{p-1}$, $j' \in \mathcal{D}_j^p$, dans la constitution du groupe j de $\mathcal{P}_p$. Dans l'exemple, l'individu i est à l'isobarycentre de ses représentations partielles associées à la *caractérisation* et aux *jugements hédoniques*; C est à l'isobarycentre de S et ϕ, représentations partielles associées aux groupes *sensoriel* et *physico-chimique* : ceci traduit la notion d'équilibre inhérente à l'AFMH.

Remarques.

1. Dans certains cas, une telle représentation peut se révéler chargée; on peut alors choisir de ne représenter que certains individus, certains niveaux, ou encore l'ensemble des niveaux appartenant à une branche de l'arbre hiérarchique.

2. La propriété barycentrique de l'AFM dont hérite l'AFMH au niveau de la partition P, est due à la dilatation, selon le coefficient J_P, des nuages des individus partiels associés aux groupes de $\mathcal{P}_P$. De la même façon, pour

obtenir la représentation ci-dessus, chaque nuage d'individus partiels associés au groupe K_j^p est dilaté selon un coefficient dépendant à la fois de J_p et de $\mathcal{D}_j^p$, $j = 1, \ldots, J_p$ (le calcul de ce coefficient est explicité en annexe). Par la suite, on désigne par $N_I^{J_p} = \{i_j^p,\ i \in I,\ j = 1, \ldots, J_p\}$ le nuage des individus partiels dilaté, au niveau de partition p, qui bénéficie des propriétés exposées ci-dessus.

Exemple d'interprétation. – Au niveau de partition le plus synthétique (*cf.* figure 10, A), les représentations partielles des individus du point de vue de la *caractérisation* d'une part et des *jugements hédoniques* d'autre part sont globalement proches : l'axe 1 et à un moindre degré l'axe 2 expriment des structures communes aux deux groupes, ce qui sera quantifié par la suite.

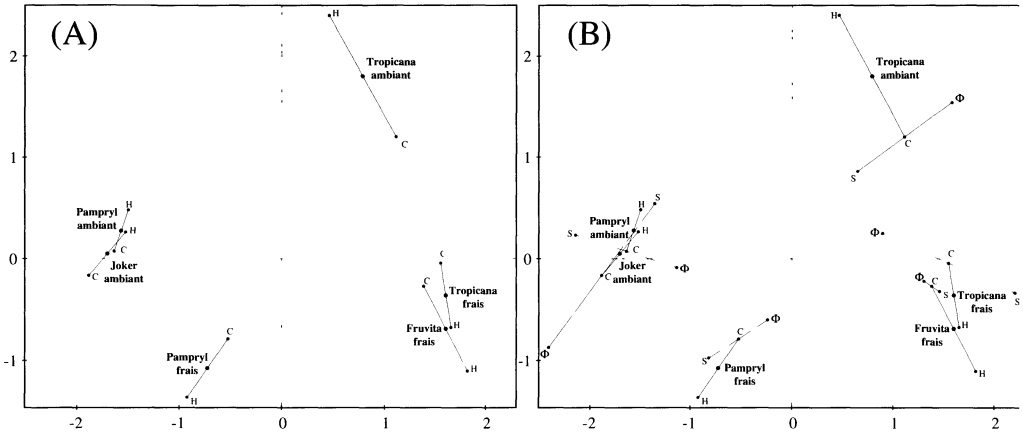

FIGURE 10

(A) *Représentation des individus partiels associés aux groupes caractérisation (C) et jugements hédoniques (H), au niveau de partition le plus synthétique.*
(B) *Représentation des individus partiels à tous les niveaux de partition*

TABLEAU 3

Extraits du tableau de données centré-réduit : quelques variables parmi les plus corrélées avec le 1er *facteur*

	Sensoriel		Physico-chimique	
	Caractère sucré	Caractère acide	Saccharose	pH
Fruivita frais	1.065	-1.000	1.317	0.929
Tropicana frais	0.933	-1.236	0.626	0.679

Au niveau de partition le plus fin (*cf.* figure 10, B), la situation est à peu près la même pour les aspects *sensoriel* et *physico-chimique*, quoique de façon un peu plus atténuée.

À titre d'exemple, on illustre, à partir des données centrées-réduites (*cf.* tableau 3), les positions respectives des individus Fruivita frais et Tropicana frais. Le Fruivita frais est également caractérisé par son aspect *sensoriel* et son aspect *physico-chimique* (les valeurs centrées-réduites sont du même ordre de grandeur en valeur absolue); le Tropicana frais est plus caractérisé par son aspect *sensoriel* que par son aspect *physico-chimique* (en valeur absolue, les valeurs centrées-réduites des variables sensorielles sont plus grandes que celles des variables physico-chimiques).

6.2. Aides à l'interprétation à la représentation superposée

La représentation superposée est souvent complexe et nécessite des indicateurs synthétiques : nous en présentons quelques-uns, fondés sur l'inertie.

Pondération. – Pour calculer ces indicateurs, on utilise une fonction de pondération qui traduit la notion d'équilibre entre les sous-groupes de variables «issus» d'un même nœud.

Cette pondération est telle que, quel que soit $i \in I$, les individus partiels «issus» d'un même nœud (*i.e.* associés aux sous-groupes de variables d'un même nœud) ont un poids identique; la somme de leur poids est égale au poids de l'individu décrit par le groupe de variables associé au nœud dont ils sont issus.

Formellement, si $p(i_j^p)$ est le poids de l'individu partiel i associé au groupe de variables K_j^p, alors :

$$\forall i \in I,\ \forall j \in \mathcal{D}_{j'}^{p+1},\ p(i_j^p) = \frac{p(i_{j'}^{p+1})}{\mathcal{D}_{j'}^{p+1}}.$$

Rappelons que $\mathcal{D}_{j'}^{p+1}$ désigne à la fois l'ensemble des indices des sous-groupes de $\mathcal{P}_p$ issus de la décomposition du groupe j' de $\mathcal{P}_{p+1}$ et son cardinal. De plus,

$$\forall i \in I,\ \forall p = 1, \ldots, P,\ \sum_{j \in \mathcal{P}_p} p(i_j^p) = p_i.$$

En effet, par construction, $\forall i \in I,\ \forall p = 1, \ldots, P$,

$$\begin{aligned}\sum_{j \in \mathcal{P}_p} p(i_j^p) &= \sum_{j' \in \mathcal{P}_{p+1}} \sum_{j \in \mathcal{D}_{j'}^{p+1}} p(i_j^p) \\ &= \sum_{j' \in \mathcal{P}_{p+1}} p(i_{j'}^{p+1}),\end{aligned}$$

cette somme se réduisant pour $\mathcal{P}_{P+1}$ à un terme unique qui est le poids p_i de l'individu i : la partition $\mathcal{P}_{P+1}$ correspond au cas où les variables appartiennent toutes à un même groupe. Dans le cas particulier du nuage de points partiels $N_I^{J_P}$, i_1^{P+1} désigne l'individu i et son coefficient de pondération $p(i_1^{P+1})$ est alors égal à p_i.

De cette pondération, découlent les deux propriétés suivantes :

1. Quel que soit le niveau de partition p, l'individu i est au centre de gravité de ses représentations partielles associées aux groupes de $\mathcal{P}_p$, ainsi pondérées (ce choix est en phase avec la première représentation superposée proposée dans le paragraphe précédent).

2. Quel que soit le niveau de partition p, les nuages des individus partiels, $N_I^{J_p} = \{i_j^p,\ i \in I,\ j \in J_p\}$, sont centrés (ce qui découle directement de la propriété précédente).

Dans l'exemple des jus d'orange, chaque individu est affecté d'un poids de $1/I = 1/6$. Le poids des individus partiels associés aux aspects *jugements hédoniques* et *caractérisation* vaut respectivement $1/12 = 1/2 \times 1/6$ (poids identiques; $\forall i,\ p_{héd} + p_{car} = p_i = 1/6$); le poids des individus partiels associés aux aspects *sensoriels* et *physico-chimiques* vaut respectivement 1/24 (poids identiques; $\forall i,\ p_{sen} + p_{phy} = p_{car} = 1/12$).

Décomposition de l'inertie. – À tous les niveaux de partition p, l'AFMH calcule l'inertie du nuage des points partiels $N_I^{J_p}$, notée $I_{tot}(p)$. Un tel nuage de points peut se décomposer en $I \times J_{p+1}$ sous-groupes, de la façon suivante :

- dans un premier temps, on regroupe tous les points associés à un même individu i;
- dans un second temps, on décompose chacun des I groupes en J_{p+1} sous-ensembles de la forme $\{i_j^p,\ j \in \mathcal{D}_{j'}^{p+1}\},\ i \in I,\ j' \in \mathcal{P}_{p+1}$.

Quel que soit $p = 1, \ldots, P$, on note $\mathcal{P}(N_I^{J_p})$ la partition du nuage des individus partiels $N_I^{J_p}$ décrite ci-dessus. Par définition,

$$I_{tot}(p) = \sum_{i \in I} \sum_{j \in \mathcal{P}_p} p(i_j^p) d^2(O, i_j^p) \ ;$$

d'après le théorème de Huygens appliqué au nuage de points $N_I^{J_p}$ partitionné selon $\mathcal{P}(N_I^{J_p})$, $I_{tot}(p)$ se décompose de la manière suivante :

$$\begin{aligned}
I_{tot}(p) &= I_{inter}(p) + I_{intra}(p) \\
&= \sum_{i \in I} \sum_{j' \in \mathcal{P}_{p+1}} \Big(\sum_{j \in \mathcal{D}_{j'}^{p+1}} p(i_j^p) \Big) d^2(O, i_{j'}^{p+1}) \\
&\qquad\qquad + \sum_{i \in I} \sum_{j' \in \mathcal{P}_{p+1}} \sum_{j \in \mathcal{D}_{j'}^{p+1}} p(i_j^p) d^2(i_j^p, i_{j'}^{p+1}) \\
&= \sum_{i \in I} \sum_{j' \in \mathcal{P}_{p+1}} p(i_{j'}^{p+1}) d^2(O, i_{j'}^{p+1}) + \sum_{i \in I} \sum_{j' \in \mathcal{P}_{p+1}} \sum_{j \in \mathcal{D}_{j'}^{p+1}} p(i_j^p) d^2(i_j^p, i_{j'}^{p+1}) \\
&= I_{tot}(p+1) + I_{intra}(p),
\end{aligned}$$

où $I_{intra}(p)$ est l'inertie intra-classes du nuage des points partiels $N_I^{J_p}$, au niveau de partition p, pondérés par $p(i_j^p)$, regroupés suivant la partition $\mathcal{P}(N_I^{J_p})$.

Par récurrence, on a la formule de décomposition suivante :

$$I_{tot}(1) = I_{tot}(N_I) + \sum_{p=1}^{P} I_{intra}(p),$$

où $I_{tot}(N_I) = I_{tot}(P+1) = \sum_{i \in I} p_i d^2(O,i)$ désigne l'inertie du nuage des individus N_I calculée par rapport à son centre de gravité.

Remarque. – Au niveau de partition $\mathcal{P}_{P+1}$, i_1^{P+1} désigne l'individu i (son coefficient de pondération $p(i_1^{P+1})$ est alors égal à p_i) et le théorème de Huygens s'écrit :

$$\begin{aligned} I_{tot}(P) &= \sum_{i \in I} p_i d^2(O,i) + \sum_{i \in I} \sum_{j \in \mathcal{P}_P} \frac{p_i}{J_P} d^2(i_j^P, i) \\ &= I_{tot}(N_I) + I_{intra}(P). \end{aligned}$$

Les calculs précédents, de même que la formule de décomposition, restent valables en projection sur un axe. Si $I_{tot/u_s}(p)$ désigne l'inertie du nuage des points partiels $N_I^{J_p}$ calculée le long de l'axe principal de rang s, si $d^2_{u_s}(A,B)$ désigne le carré de la distance entre les points A et B sur l'axe principal u_s, alors :

$$I_{tot/u_s}(p) = \sum_{i \in I} \sum_{j \in \mathcal{P}_p} p(i_j^p) d^2_{u_s}(O, i_j^p).$$

De même, si $I_{intra/u_s}(p)$ désigne l'inertie intra-classes du nuage des points partiels $N_I^{J_p}$ calculée le long de l'axe u_s, alors :

$$I_{tot/u_s}(1) = I_{tot/u_s}(N_I) + \sum_{p=1}^{P} I_{intra/u_s}(p).$$

Analogie avec l'analyse de variance. – Soit $F_s(i_j^p)$ la coordonnée sur u_s de la représentation partielle de l'individu i associée à K_j^p. Une analyse de variance des $F_s(i_j^1)$, pondérée par $p(i_j^1)$, permet de retrouver la décomposition de l'inertie de $N_I^{J_1}$ ci-dessus. Pour cela, on considère le modèle d'analyse de variance hiérarchique suivant :

$$F_s(i_j^1) = i + \mathcal{P}_P[i] + \mathcal{P}_{P-1}[\mathcal{P}_P[i]] + \ldots + \mathcal{P}_1[\ldots \mathcal{P}_P[i]],$$

où i est le facteur *individu*, $\mathcal{P}_p$ est la facteur *partition* associé au niveau de partition p, et $B[A]$ désigne l'effet du facteur B hiérarchisé dans A.

Les différentes sommes des carrés des écarts associées aux facteurs introduits dans le modèle correspondent chacune à un élément de la décomposition de $I_{tot/u_s}(1)$ ci-dessus (*cf.* tableau 4).

TABLEAU 4
Correspondance entre la décomposition de l'inertie en AFMH
et la décomposition des carrés des écarts en analyse de variance

Inertie	Facteurs du modèle d'A.V.
$I_{tot/u_s}(1)$	$F_s(i_j^1)$
$I_{tot/u_s}(N_I)$	i
$I_{intra/u_s}(P)$	$\mathcal{P}_P[i]$
$\vdots$	$\vdots$
$I_{intra/u_s}(p)$	$\mathcal{P}_p[\ldots\mathcal{P}_P[i]]$
$\vdots$	$\vdots$
$I_{intra/u_s}(1)$	$\mathcal{P}_1[\ldots\mathcal{P}_P[i]]$

Exemple. L'axe 1 met en évidence une étroite ressemblance entre les nuages associés aux représentations partielles, tous niveaux de partition confondus, comme l'indique le rapport

$$\frac{I_{tot/u_s}(N_I)}{I_{tot/u_s}(1)} = 0.9304,$$

élevé (*cf.* tableau 5).

TABLEAU 5
Décomposition de l'inertie associée à chacun des 5 axes factoriels.
Remarquons que les valeurs prises par $I_{tot/u_s}(N_I)$ sur ces axes
*correspondent aux valeurs propres de l'*AFMH

		Valeurs brutes					Valeurs en %				
Inertie	S.C.E.	Axe 1	Axe 2	Axe 3	Axe 4	Axe 5	Axe 1	Axe 2	Axe 3	Axe 4	Axe 5
$I_{tot/u_s}(1)$	$F_s(i_j^1)$	2.083	1.043	1.061	0.900	0.717	100	100	100	100	100
$I_{tot/u_s}(N_I)$	i	1.938	0.846	0.650	0.513	0.507	93.04	81.11	61.26	57	70.71
$I_{intra/u_s}(2)$	$\mathcal{P}_2[i]$	0.039	0.134	0.252	0.169	0.166	1.87	12.85	23.75	18.77	23.15
$I_{intra/u_s}(1)$	$\mathcal{P}_1[\mathcal{P}_2[i]]$	0.106	0.063	0.159	0.218	0.044	5.09	6.04	14.99	24.23	6.14

La faible différence d'inertie entre les nuages N_I et $N_I^{J_1}$ est due plus au premier niveau de partition qu'au second : l'hétérogénéité est plus importante entre les

aspects *sensoriel* et *physico-chimique* ($I_{intra/u_s}(1) = 0.106$), qu'entre les aspects *caractérisation* et *jugements hédoniques* ($I_{intra/u_s}(2) = 0.039$), dont les points partiels associés à un même individu sont pratiquement superposés.

Sur le deuxième axe factoriel, la structure commune entre les groupes est moins nette mais semble digne d'intérêt au sens où le rapport

$$\frac{I_{tot/u_s}(N_I)}{I_{tot/u_s}(1)} = 0.8111$$

est sensiblement plus élevé que pour les axes suivants. En termes d'inertie, la variabilité intra-classes, pour ce deuxième axe, se réfère surtout à la partition $\mathcal{P}_2$ ($I_{intra/u_s}(2) = 12.85\%$); rappelons cependant que les individus partiels associés à cette partition ont un poids deux fois plus élevé que celui des individus partiels associés aux aspects *sensoriel* et *physico-chimique*.

6.3. Variables canoniques

Principe. – Soit $F_s^{j(p)}$ le vecteur des coordonnées des individus partiels associés au groupe j de $\mathcal{P}_p$, sur l'axe de rang s; ce vecteur définit une nouvelle fonction sur les individus, obtenue en projetant le nuage partiel associé au groupe j de $\mathcal{P}_p$ sur u_s, appelée variable canonique.

TABLEAU 6
Corrélations entre les variables canoniques et les variables générales

Niveau de partition 2					
Aspects	Axe 1	Axe 2	Axe 3	Axe 4	Axe 5
Caractérisation	0.990	0.971	0.803	0.961	0.967
Jugements hédoniques	0.990	0.993	0.970	0.997	0.997
Niveau de partition 1					
Aspects	Axe 1	Axe 2	Axe 3	Axe 4	Axe 5
Physico-chimique	0.910	0.798	0.637	0.449	0.543
Sensoriel	0.973	0.896	0.477	0.408	0.848
Jugements hédoniques	0.990	0.993	0.970	0.997	0.997

Aides à l'interprétation, exemple. – Les corrélations entre variables canoniques et variables générales (facteurs de l'analyse globale) sont calculées à tous les niveaux de partition. Quel que soit le niveau de partition, ces coefficients de corrélation se commentent comme ceux d'une AFM.

Le coefficient de corrélation linéaire entre la variable canonique $F_s^{j(p)}$ et le facteur F_s de l'AFMH montre dans quelle mesure ce dernier peut être considéré comme une structure du groupe j de $\mathcal{P}_p$.

En l'absence d'indicateurs statistiques qui permettraient de sélectionner les corrélations significatives, les coefficients relatifs aux axes de rang élevé servent empiriquement de base de comparaison.

Dans l'exemple, les 5 premiers facteurs de l'AFMH sont communs aux deux groupes *caractérisation* et *jugements hédoniques* de $\mathcal{P}_2$. Notons que ces deux groupes comportent chacun un nombre de variables grand (17 pour le premier et 96 pour le second) relativement à la taille de l'espace des variables (de dimension 5 seulement); les sous-espaces engendrés par ces groupes recouvrent donc tout cet espace, ce qui explique les valeurs élevées (et attendues) des corrélations (*cf.* tableau 6).

Remarque. – Ce nombre de variables, grand relativement au nombre d'individus, est très courant en analyse sensorielle : le nombre de produits mis en jeu est toujours inférieur au nombre de descripteurs et très inférieur au nombre de jugements hédoniques. Il impose une grande prudence dans l'interprétation des relations entre variables.

Au niveau de partition $\mathcal{P}_1$, les coefficients de corrélations suggèrent une structure commune entre les aspects *sensoriel* et *physico-chimique* pour les deux premiers axes.

6.4. Groupes de variables

Principe et graphique. – De même que la prise en compte de plusieurs partitions amène à autant de représentations partielles qu'il y a de groupes de variables distincts, l'AFMH fournit une représentation de l'ensemble des groupes, toutes partitions confondues.

Dans l'exemple des jus d'orange, l'AFMH, en tant qu'AFM au niveau de partition P, fournit une représentation des groupes au niveau de partition le plus synthétique; elle fournit également une représentation des groupes *physico-chimique* et *sensoriel* issus de la décomposition du groupe *caractérisation*.

Le principe de cette représentation repose sur celui de l'AFM : les groupes sont représentés dans l'espace $\mathbb{R}^{I^2}$ par leur matrice des produits scalaires; leurs coordonnées, sur les vecteurs de $\mathbb{R}^{I^2}$ induits par les composantes principales de N_K dans $\mathbb{R}^I$, sont obtenues par projection.

La pondération des variables dans le calcul des produits scalaires entre individus décrits par un même groupe peut être envisagée de deux façons.

La première consiste à utiliser les poids «finaux» associés à la fonction de pondération m. Dans ce cas, la projection du j^e groupe de $\mathcal{P}_p$ sur le vecteur induit par la composante principale ν_s s'interprète comme la contribution du groupe à l'inertie de la composante.

La seconde consiste à utiliser les poids «instantanés» associés aux fonctions m_p, calculés à chaque étape de l'algorithme présenté dans la partie 4.2. Dans ce cas,

la coordonnée du groupe j de $\mathcal{P}_p$ sur le vecteur induit par la composante principale ν_s s'écrit

$$< W_j^p D, \nu_s \nu_s' D >= \sum_{k \in K_j^p} m_p(k) r^2(k, \nu_s),$$

où $W_j^p = X_j^p M_j^p (X_j^p)'$ désigne la matrice des produits scalaires associée aux variables du groupe K_j^p pondérées par m_p et $r^2(k, \nu_s)$ le carré du coefficient de corrélation entre la variable k et la composante principale ν_s. Le second membre de la formule ci-dessus correspond à la mesure de liaison $\mathcal{L}_g$ classiquement utilisée en AFM pour traduire la liaison entre une variable (ici ν_s) et un groupe de variables (ici K_j^p), la fonction de pondération m_p garantissant que la plus grande valeur propre issue de l'analyse factorielle des variables de K_j^p est égale 1.

Cette seconde solution est choisie pour l'AFMH. Il en résulte la relation suivante entre la matrice des produits scalaires associée aux variables du groupe K_j^p, $p \geqslant 2$, et celles associées aux sous-groupes j' de $\mathcal{P}_{p-1}$ qui constituent ce dernier :

$$W_j^p = \frac{1}{\lambda_1^{j(p)}} \sum_{j' \in \mathcal{D}_j^p} W_{j'}^{p-1}.$$

De cette égalité, on déduit la relation entre les coordonnées d'un groupe j de $\mathcal{P}_p$, $p \geqslant 2$, et celles des sous-groupes j' de $\mathcal{P}_{p-1}$ qui le constituent :

$$< W_j^p D, \nu_s \nu_s' D >= \frac{1}{\lambda_1^{j(p)}} \sum_{j' \in \mathcal{D}_j^p} < W_{j'}^{p-1} D, \nu_s \nu_s' D > .$$

Géométriquement, dans l'espace des groupes $\mathbb{R}^{I^2}$, le groupe j de $\mathcal{P}_p$ est à l'intérieur du cône de sommet O et dont la base est engendrée par les points j' de $\mathcal{P}_{p-1}$ appartenant à $\mathcal{D}_j^p$.

Remarques.

1. Au niveau de partition P, les propriétés de l'AFM s'appliquent : en tant qu'inertie projetée du groupe, la coordonnée mesure l'importance du groupe dans la détermination de la composante; cette coordonnée vaut 1 lorsque l'axe de l'analyse globale coïncide avec la première composante principale du groupe.
2. Soient K_1^p et K_2^p deux groupes identiques de $\mathcal{P}_p$, $K_1^{p+1} = K_1^p \cup K_2^p$ le groupe de $\mathcal{P}_{p+1}$ issu de leur union, alors les représentations respectives de ces trois groupes dans $\mathbb{R}^{I^2}$ sont confondues.
3. Dans $\mathbb{R}^{I^2}$, le groupe j de $\mathcal{P}_p$ est situé sur l'axe passant par l'isobarycentre des groupes j' de $\mathcal{P}_{p-1}$ qui le constituent et par l'origine. Notons que dans cet espace, le poids d'un groupe n'est pas égal à la somme des poids des variables qui le composent et que la propriété géométrique énoncée précédemment suppose implicitement que les groupes d'un même nœud ont un même poids.

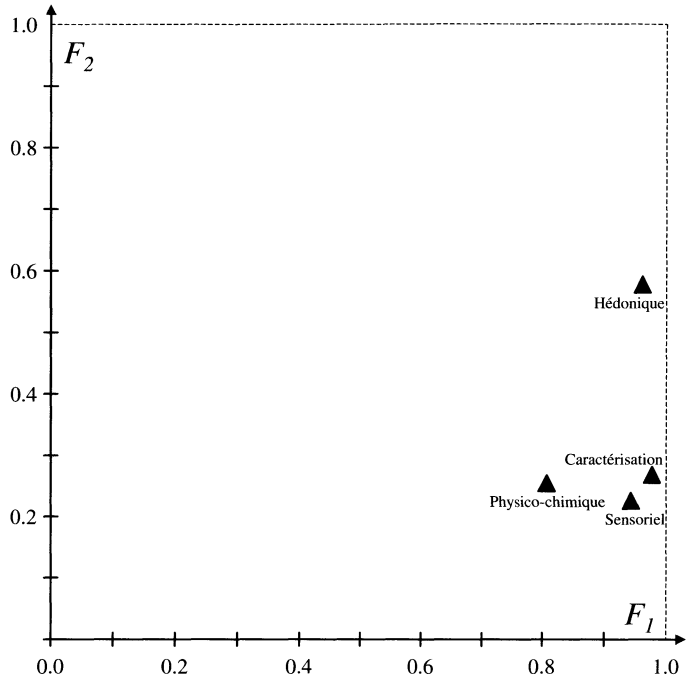

FIGURE 11
Représentation des groupes de variables :
la coordonnée d'un groupe s'interprète comme la mesure de liaison $\mathcal{L}_g$

Aides à l'interprétation, exemple. – En complément au graphique, on calcule, pour chaque axe, le rapport *Inertie projetée/Inertie totale*, indicateur classique de qualité de représentation, qui, dans le cas d'un groupe, s'interprète comme un cosinus carré (*cf.* tableau 7).

Dans l'exemple, le premier facteur F_1 correspond à une direction d'inertie importante de l'ensemble des groupes, toutes partitions confondues : sur cet axe, les coordonnées des groupes sont toutes supérieures à 0.807.

Le deuxième axe est principalement lié (au sens de la mesure de liaison $\mathcal{L}_g$) au groupe *jugements hédonique*, ce qui est à relier au fait que ce dernier est plus multidimensionnel que les trois autres groupes.

Soulignons que la qualité de représentation des groupes sur le premier plan factoriel est relativement bonne; le graphique suggère une plus grande ressemblance entre les deux aspects de la caractérisation qu'entre la caractérisation et les jugements hédoniques. Ceci est en accord avec le choix des variables physico-chimiques, sensées traduire les perceptions sensorielles, alors qu'un très grande diversité des jugements hédoniques est toujours possible.

TABLEAU 7
Coordonnées et qualité de représentation des groupes, toutes partitions confondues

Niveau de partition 2	Coordonnées			Cosinus carrés		
Aspects	F_1	F_2	F_3	F_1	F_2	Plan (1,2)
Caractérisation	0.977	0.268	0.163	0.855	0.064	0.919
Jugements hédoniques	0.961	0.578	0.487	0.499	0.312	0.811
Niveau de partition 1	Coordonnées			Cosinus carrés		
Aspects	F_1	F_2	F_3	F_1	F_2	Plan (1,2)
Physico-chimique	0.807	0.255	0.233	0.602	0.059	0.661
Sensoriel	0.943	0.226	0.059	0.801	0.046	0.847
Jugements hédoniques	0.961	0.578	0.487	0.499	0.312	0.811

Remarques.

1. Comme attendu, les deux groupes de $\mathcal{P}_2$ contribuent presque également à la formation du premier facteur.
2. Bien que présent à la fois dans les partitions $\mathcal{P}_1$ et $\mathcal{P}_2$, le groupe *jugements hédoniques* n'est représenté qu'une seule fois : d'une partition à l'autre, ses coordonnées sur les composantes principales restent identiques.

6.5. Facteurs des analyses séparées

Principe et graphique. – Les facteurs normés issus des analyses factorielles réalisées à chaque nœud sont projetés, comme des variables illustratives, sur les facteurs issus de l'AFMH; leurs coordonnées s'interprètent comme des coefficients de corrélation.

Aides à l'interprétation, exemple. – Parmi les aides à l'interprétation, citons la qualité de représentation des axes partiels actifs, qui permet de comparer la représentation optimale d'un groupe issue de son analyse séparée à celle obtenue par l'AFMH; citons également les contributions des facteurs des analyses séparées à la formation des facteurs de l'AFMH.

Dans l'exemple, la première (resp. la deuxième) composante de l'AFMH est très liée à la première (resp. la deuxième) composante principale de chaque groupe, toutes partitions confondues.

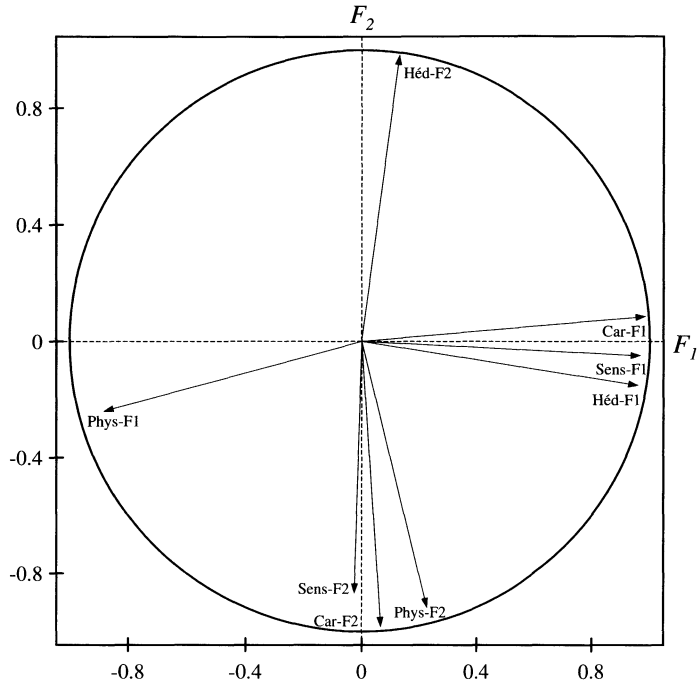

FIGURE 12
F1 *(resp.* F2*) correspond au premier (resp. deuxième) facteur de l'analyse séparée des groupes physico-chimique (Phys), sensoriel (Sens), caractérisation (Car) et hédonique (Héd)*

7. Références

[1] Carroll, J.D. (1968). A generalization of canonical correlation analysis to three or more sets of variables. *Proc.* 76th *Conv. Am. Psychol. Assoc.*, 227–228.

[2] Escofier, B., Pagès, J. (1998). *Analyses factorielles simples et multiples.* 3e édition. Dunod. Paris.

[3] Gower, J. C. (1975). Generalized Procrustes Analysis. *Psychometrika. 40(1)*, 33–51.

[4] Lavit, C. (1988). *Analyse conjointe de tableaux quantitatifs.* Édition Masson.

Annexe : calcul du coefficient de dilatation utilisé dans la représentation des individus partiels

Soit $\tilde{X}_j^p$ la restriction dans $\mathbb{R}^K$ de X aux variables de K_j^p ; $\tilde{X}_j^p$ est la matrice de $\mathcal{M}(I, K)$ dans laquelle X_j^p est complétée par des 0. Par construction, $\tilde{X}_j^p = XI_j^p$,

où I_j^p est la matrice diagonale de $\mathcal{M}(K,K)$ définie par :

$$I_j^p = \begin{cases} I_j^p(k,k) = 1 & \text{si } k \in K_j^p, \\ I_j^p(k,h) = 0 & \text{sinon.} \end{cases}$$

Pour bénéficier des propriétés barycentriques exposées dans la partie 6.1, on dilate les nuages des individus partiels par un coefficient noté α_j^p, calculé en remarquant que :

1. au niveau de partition P,

$$\sum_{j=1}^{J_P} \tilde{X}_j^P = X \Longleftrightarrow \sum_{j=1}^{J_P} 1(J_P \tilde{X}_j^P) = (\sum_{j=1}^{J_P} 1)X; \tag{1}$$

2. $\forall\ 1 < p \leqslant P$,

$$\sum_{j' \in \mathcal{D}_j^p} \tilde{X}_{j'}^{p-1} = \tilde{X}_j^p \Longleftrightarrow \sum_{j' \in \mathcal{D}_j^p} 1(\mathcal{D}_j^p \tilde{X}_{j'}^{p-1}) = (\sum_{j' \in \mathcal{D}_j^p} 1)\tilde{X}_j^p. \tag{2}$$

D'après (1), pour que chaque individu soit au centre de gravité de ses représentations partielles au niveau de partition P, les nuages des individus partiels associés aux groupes de $\mathcal{P}_P$ doivent être dilatées selon le coefficient J_P.

D'après (2), pour que chaque individu i_j^p soit au centre de gravité de ses représentations partielles, les nuages des individus partiels doivent être dilatés selon le coefficient $\mathcal{D}_j^p$.

Si $N_I^{j(p)}$ désigne le nuage des individus partiels décrits par K_j^p, les coordonnées de $N_I^{j'(P-1)}$ se déduisent de (1) et (2) en multipliant $\tilde{X}_{j'}^{P-1}$ par $J_P \mathcal{D}_j^P$, où j est le groupe $\mathcal{P}_P$ qui contient j' de $\mathcal{P}_{P-1}$.

Ce calcul se généralise aux autres niveaux de partition et l'on montre que les coordonnées de $N_I^{j(p)}$ s'obtiennent en multipliant $\tilde{X}_j^p$ par α_j^p, défini par

$$\alpha_j^p = \begin{cases} J_P \text{ si } p = P, \\ J_P \prod_{h=p+1}^{P} \mathcal{D}_{g_p^h(j)}^h \text{ sinon,} \end{cases}$$

où $g_p^h(j)$ est le numéro du groupe de $\mathcal{P}_h$ qui contient le groupe j de $\mathcal{P}_p$.