

# EXTENSIONS DE LA MÉTHODE *STATIS*. APPLICATION EN ÉVALUATION SENSORIELLE.

El Mostafa Qannari <sup>(1)</sup>

Mohamed Hanafi <sup>(1)</sup>

Pascal Schlich <sup>(2)</sup>

<sup>(1)</sup>ENITIAA / INRA Unité de Sensométrie et Chimiométrie  
Rue de la Géraudière BP 82 225  
44322 NANTES cedex 03- France  
{qannari, hanafi}@enitiaa-nantes.fr

<sup>(2)</sup> Centre Européen des Sciences du Goût  
Laboratoire d'Interface Recherche-Industrie-Sensométrie  
15, rue Hugues Picardet  
21000 Dijon - France  
schlich@cesg.cnrs.fr

## **Résumé**

*La méthode Statist est bien adaptée pour l'analyse des données de profils sensoriels avec un vocabulaire fixé ou libre. A partir d'une nouvelle formulation de cette méthode, nous proposons deux extensions et illustrons leur intérêt pratique sur la base d'études de cas en évaluation sensorielle. La première extension consiste à déterminer un tableau compromis de dimension fixée. La deuxième extension consiste à déterminer un espace de représentation de dimension fixée par l'utilisateur et contraindre cet espace à être lié de manière optimale à des données externes. Finalement, la détermination de ces espaces peut être réalisée de manière séquentielle sur le nombre de dimensions avec des contraintes d'orthogonalité.*

*Mots clefs : STATIS, analyse sensorielle.*

## **Abstract**

*Statist method is advocated for the analysis of both conventional and free-choice profiling data. A new presentation of STATIS is given which makes it possible to extend the method in two directions which have valuable implications with profiling sensory data. The first extension looks for a group average configuration having a fixed number of dimensions. The second extension looks for a group average configuration having underlying dimensions constrained to be linear combinations of external data. Finally, a sequential determination of successive orthogonal group average configurations is proposed.*

*Keywords: sensory analysis, STATIS method.*

## 1. INTRODUCTION

En analyse sensorielle, l'épreuve de profil sensoriel consiste à demander à un groupe de sujets entraînés d'attribuer des notes d'intensité aux différents produits de l'étude pour chacun des descripteurs (variables statistiques) qui constituent ce profil sensoriel. Lorsque la liste des descripteurs est imposée aux sujets, on parle de profil conventionnel ; alors que, lorsque chaque sujet est libre de noter les descripteurs de son choix, on parle de profil libre-choix. Dans les deux cas, on cherche à définir une configuration

multidimensionnelle des produits qui réalise un compromis entre les configurations individuelles. Les premières composantes principales de ce compromis sont utilisées pour dresser une carte sensorielle des produits utile pour synthétiser l'essentiel des différences entre les produits sur lesquelles les sujets s'accordent, mais aussi pour relier les aspects sensoriels aux caractéristiques physico-chimiques des produits ainsi qu'aux préférences des consommateurs pour ces produits.

La méthode STATIS (Lavit, 1976) est particulièrement bien adaptée à cette problématique (Schlich, 1996 ; Qannari et al., 1997) puisque le compromis qu'elle propose est une moyenne des configurations individuelles pondérée par l'accord de chaque sujet avec le groupe. A partir d'une nouvelle formulation de la méthode STATIS, nous proposons deux extensions de cette technique et nous illustrons leur intérêt pratique pour les données de profil sensoriel.

La première extension consiste à contraindre le compromis d'être de rang fixé inférieur au rang des configurations individuelles. La justification est que bien souvent ce ne sont que les premières composantes de ce compromis qui seront interprétables et interprétées. En restreignant la recherche de compromis à un sous-espace utile, nous espérons ainsi réduire la prise en compte de bruit dans la définition du compromis..

La deuxième extension consiste, comme dans la première extension, à déterminer un espace de représentation de dimension fixée, mais tout en contraignant celui-ci à être lié de manière optimale à des données externes.

Pour chacune de ces extensions, nous pouvons également définir une démarche séquentielle qui consiste à chercher des espaces successifs de dimensions croissantes avec contrainte d'orthogonalité de chaque nouvelle dimension avec les précédentes. L'ouverture serait alors de bâtir un test de dimensionalité du compromis

## II MÉTHODES

Soit  $m$  le nombre de sujets du groupe effectuant le profil sensoriel de  $n$  produits. A chaque juge  $i$ , est associé le tableau  $X_i$  dont les lignes correspondent aux  $n$  produits évalués et les colonnes aux descripteurs sensoriels. Nous supposons dans la suite que les tableaux  $X_i$  sont centrés en colonnes et que chaque tableau  $X_i$  est multiplié par un facteur d'échelle de manière à avoir  $\|W_i\| = 1$ .

La méthode *STATIS* est basée sur les matrices individuelles  $W_i = X_i X_i^T$  des produits scalaires entre produits. Le recours aux matrices des produits scalaires au lieu des configurations elles-mêmes présente l'avantage de se dispenser de déterminer explicitement les rotations qui ajustent les configurations. Le problème de détermination d'un compromis est, par conséquent, nettement simplifié par rapport à celui de la méthode de Procruste généralisée (Gower, 1975).

Nous proposons de déterminer des scalaires positifs  $\alpha_i$  et un opérateur compromis  $W$  de sorte à minimiser :

$$(P1) \quad \begin{array}{ll} \text{Min} & \sum_{i=1}^m \|W_i - \alpha_i W\|^2 ; \\ & \{\alpha_i, W\} \end{array}$$

Nous montrons que  $W = \sum \alpha_i W_i$  avec  $\alpha_i$  la  $i^{ème}$  coordonnée du premier vecteur propre de la matrice RV dont l'élément générique est donné par :

$$RV(l, k) = RV(X_l, X_k) = \frac{\text{trace}(W_l W_k)}{\sqrt{\text{trace}(W_l W_l)} \sqrt{\text{trace}(W_k W_k)}}.$$

Le coefficient  $RV$  est un indice d'association entre les configurations  $X_i$  et  $X_j$  qui a été introduit par Robert et Escoufier (1976) et qui équivaut au coefficient de corrélation de Pearson appliqué aux opérateurs  $W_l$  et  $W_k$  dont les éléments auraient été rangés dans des vecteurs à une dimension.

Etant semi-définie positive,  $W$  peut s'écrire sous la forme :  $W = C C^T$ . Le tableau  $C$  ainsi obtenu constitue une configuration compromis de  $X_1, X_2, \dots, X_m$ . Ainsi, le problème P1 permet de retrouver la méthode *STATIS*.

Enfin, *STATIS* effectue une *ACP* du tableau  $C$  pour permettre une représentation graphique des produits qui utilise rarement plus que les deux premières composantes principales. Il apparaît donc naturel de rechercher directement un espace  $C$  de dimension fixée par l'utilisateur qui servira de base pour la représentation factorielle des produits. Cette approche a été étudiée dans le cadre de l'analyse procruste généralisée par Peay (1988).

Si l'on admet que chaque  $W_i$  contient une part d'information et une part de bruit, on peut penser que les coefficients  $\alpha_i$  de *STATIS* tiennent compte de l'accord global des sujets autant sur l'information que sur le bruit. Nous pensons que notre approche focalisera l'accord entre les sujets sur l'information.

La recherche d'un compromis de dimension fixée nous conduit à considérer le problème d'optimisation suivant :

$$(P2) \quad \text{Min} \quad \sum_{i=1}^m \|W_i - \alpha_i C \Lambda C^T\|^2 ;$$

$$\{\alpha_i, C, \Lambda\}$$

sous les contraintes  $C = [c_1, c_2, \dots, c_r]$  ( $r$  étant un nombre entier fixé),  $C^T C = I$  et  $\Lambda = \text{diag}(\lambda_j)$ .

La résolution de ce problème conduit à un algorithme itératif qui se déroule de la manière suivante :

(a) choisir  $m$  valeurs positives  $\alpha_i$  ( $i = 1, \dots, m$ ) (par exemple,  $\alpha_i = 1/m$ );

(b) calculer la matrice  $W = \sum_{i=1}^m \alpha_i W_i$  ;

(c) définir  $\Lambda = \text{diag}(\lambda_j)$  à l'aide des relations :

$$\lambda_j = \frac{\mu_j}{\sum_i \alpha_i^2}$$

où  $\mu_j, j=1, \dots, r$  sont les  $r$  premières valeurs propres de la matrice  $W$  ;

(d) définir  $C$  comme étant la matrice dont les colonnes sont constitués de  $r$  vecteurs propres normés associés aux  $r$  premières valeurs propres de la matrice  $W$ ;

(e) mettre à jour les valeurs  $\alpha_i$  par :

$$\alpha_i = \frac{\text{trace}(W_i C \Lambda C^T)}{\text{trace}(\Lambda^2)}, \quad (i = 1, 2, \dots, m);$$

(f) aller à l'étape (b) jusqu'à convergence.

Nous montrons également comment ce problème peut être étendu afin de déterminer plusieurs compromis successifs permettant de compléter le premier compromis déterminé par l'algorithme décrit précédemment.

Un cas qui retiendra particulièrement notre attention est le cas où les espaces de représentation déterminés de manière séquentielle sont des espaces à une dimension. Nous retrouvons dans ce cas, l'analyse en composantes communes et poids spécifiques dont l'intérêt a été montré dans le cadre de la sensométrie (Qannari *et al.*, 2000) et la chimiométrie (Mazerolles *et al.*, 2002, Pram-Nielsen *et al.*, 2001).

Une autre adaptation de la méthode d'analyse consiste à déterminer un espace de représentation  $C$  de dimension fixée et à imposer à cet espace d'être lié de manière optimale à des données externes (données physico-chimiques, par exemple). Nous montrons que ce problème peut être résolu à l'aide d'un algorithme itératif similaire à celui décrit ci-dessus.

L'intérêt pratique de la démarche est illustré en l'appliquant à différentes études d'analyse sensorielle et en simulant des données permettant de justifier l'apport de cette démarche par rapport à celle du STATIS classique.

## Bibliographie

- [1] Gower J. C. (1975) Generalized Procrustes Analysis. *Psychometrika*, 40, 33-51.
- [2] Mazerolles G. , Devaux M.F. Dufour E., Qannari E.M., Courcoux Ph. (2002) Chemometric methods for the coupling of spectroscopic techniques and the extraction of relevant information contained in spectral data. *Chemometrics and Intelligent Laboratory Systems*, 63, 57-68.
- [3] Peay E. R. (1988) Multidimensional rotation and scaling of configurations to optimal agreement. *Psychometrika*, 53, 199-208.
- [4] Lavit C. (1976). Analyse conjointe de tableaux quantitatifs. Masson, Paris.
- [5] Qannari E. M., Courcoux Ph., Lejeune M. & Maystre O. (1997). Comparaison de trois stratégies de détermination d'un compromis en évaluation sensorielle. *Revue de statistique appliquée*. XLV(1), 61-74.
- [6] Qannari E.M., Wakeling I., Courcoux Ph., MacFie J.H. (2000). Defining the underlying sensory dimensions. *Food Quality and Preference* 11, 151-154.
- [7] Pram Nielsen J., Bertrand D., Micklander E., Courcoux P., Munck L. (2001). Study of NIR spectra, particle size distribution and chemical parameters of wheat flours a multi-way approach. *Journal of Near Infrared Spectroscopy*, 9, 275-285.
- [8] Robert P. & Escoufier Y. (1976). A unifying tool for linear multivariate statistical methods: the RV coefficient. *Applied Statistics*, 25, 257-267.
- [9] Schlich P. (1996). Defining and validating assessor compromises about product distances and attribute correlations. *In Multivariate analysis of data in sensory science*. ed. Naes and Risvik. Elsevier.