How to cluster the consumers taking account of sensory data

El Mostafa Qannari
E.N.I.T.I.A.A. I.N.R.A,
Nantes France

ÉCOLE NATIONALE D'INGÉNIEURS DES TECHNIQUES DES INDUSTRIES AGRICOLES ET ALIMENTAIRES

INTRODUCTION

For effective new product development, market research and quality assurance:

investigate the relationships between sensory and preference data;

Sensory characterisation of the products

judgments of consumers on the products

TWO STRATEGIES OF DATA COLLECTION

Compositional

consumers only.

Decompositional

consumers on the one hand and trained panel on the other hand.

WHICH DATA ?

- Notation on a **liking scale**.
- list of **categories** (I like very much, I like, I like moderately, I rather dislike, , I dislike much)
- **paired Comparison** data.

INTERNAL PREFERENCE MAPPING: MDPREF

	Consumer1	Consumer 2	Consumer 3	etc
Product 1	score	score	score	–·–·
Product 2	score	score	score	–·–·
Product 3	score	score	score	–·–·
Product 4	score	score	score	–·–·
etc	..	–·–·	–·–·	–·–·

PCA

However

The total variances explained by the various principal components are small

MD-PREF : example **from** *Greenhoff K. & Macfie H. J. H. (1994). Preference mapping in practice. In Measurement of food preferences, ed. H. J. H. MacFie and D. M. H. Thomson. Blackie academic & professional.*

PRODUCTS

12 beers:
1 standard beer (Std)
5 experimental beers (var1, var2, var3, var4, var5)
6 competitive beers (Brand1, Brand2, …, Brand6)

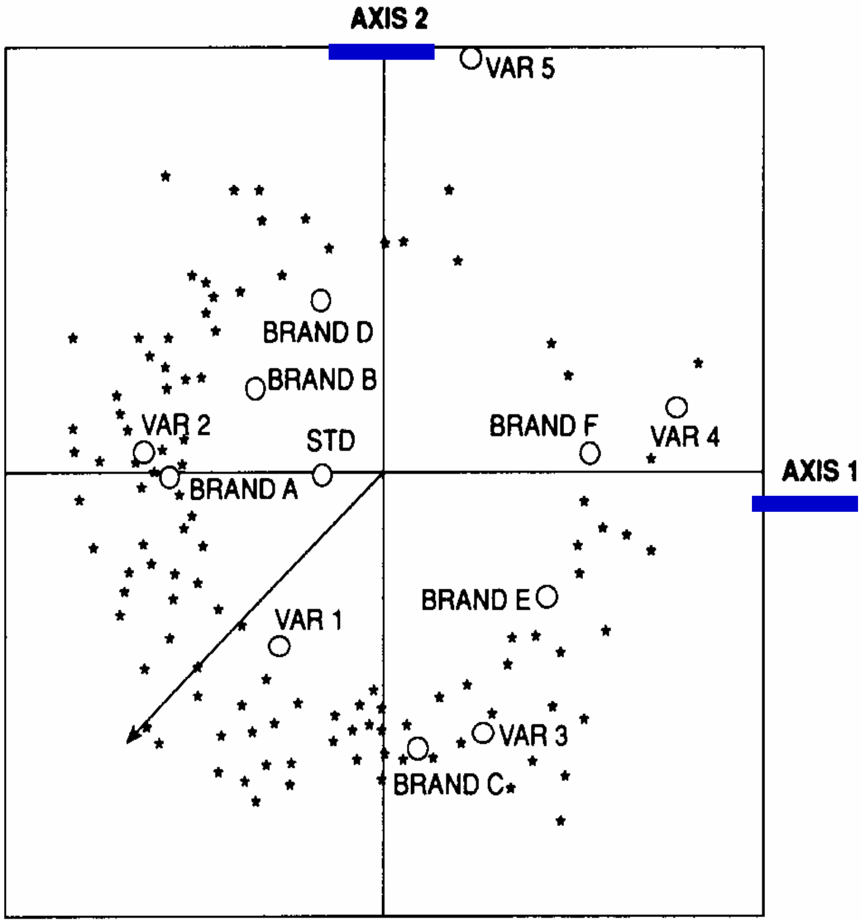
consumers

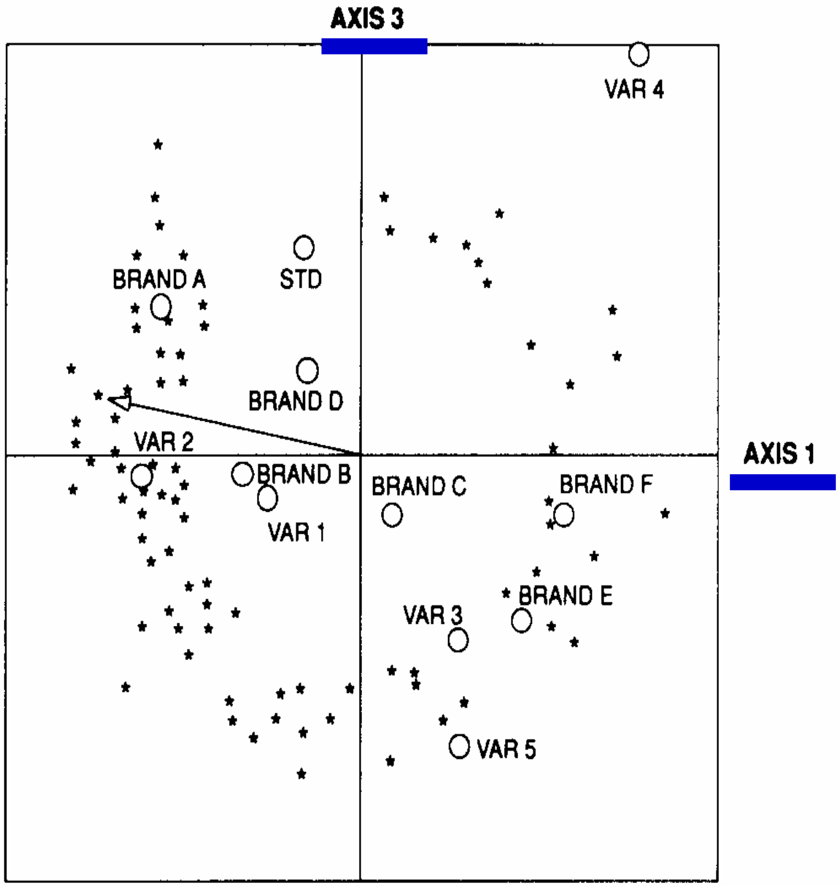
225; male, between 18 and 34 years
are accustomed to drinking beers

NOTATION

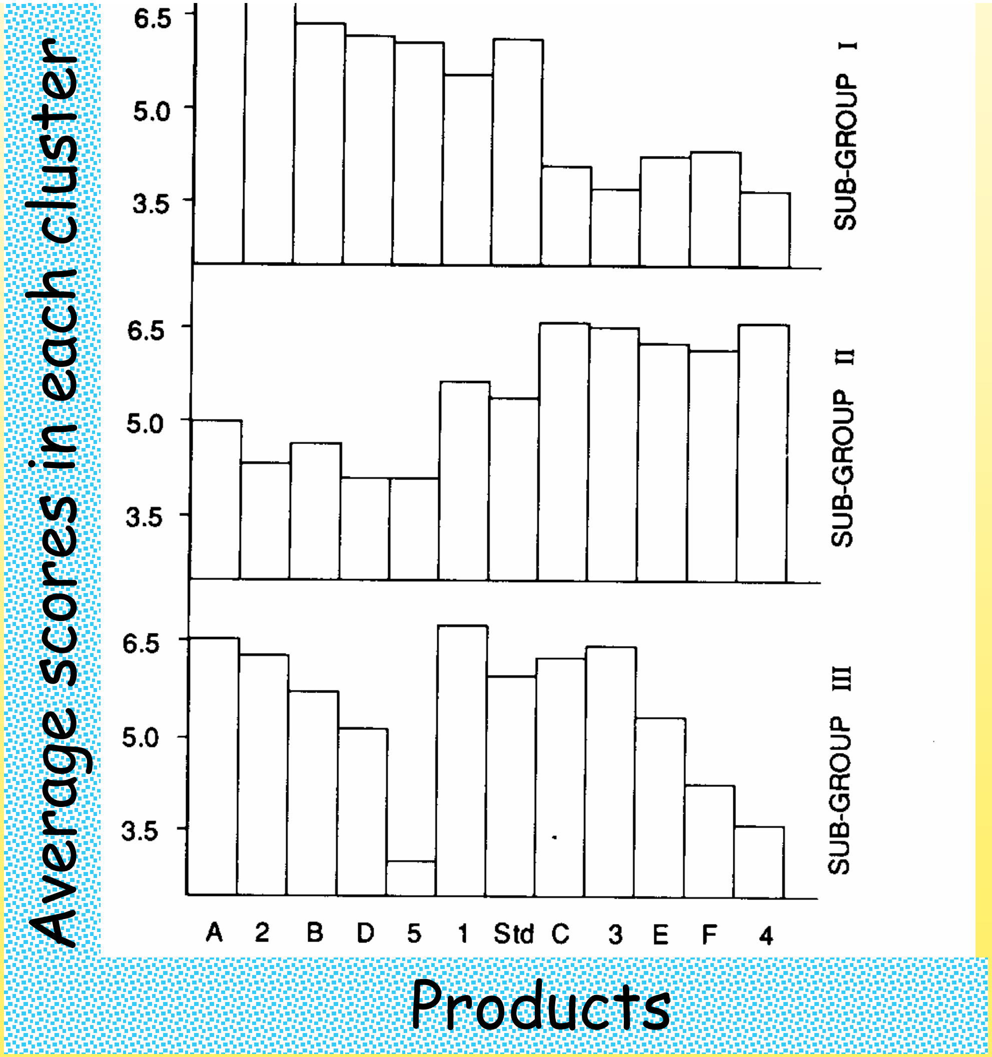
hedonic scale (1 to 9)
Two sessions of 90 minutes each

MD-PREF : example **from** *Greenhoff K. & Macfie H. J. H. (1994). Preference mapping in practice. In Measurement of food preferences, ed. H. J. H. MacFie and D. M. H. Thomson. Blackie academic & professional.*





MD-PREF : example ***from*** *Greenhoff K. & Macfie H. J. H. (1994). Preference mapping in practice. In Measurement of food preferences, ed. H. J. H. MacFie and D. M. H. Thomson. Blackie academic & professional.*



Automatic clustering of the consumers (from Howard Moskowitz; Food Quality and Preference; in press)

They are mainly **three ways** of clustering the consumers.

1♦ Clustering by how they describe themselves (Socio-demographic variables and behavioral attitudes).

2♦ Clustering people on the basis of the pattern of their acceptance ratings.

3♦ Clustering people on the basis of their patterns of liking versus sensory/instrumental measures.

Automatic clustering of the consumers (from Howard Moskowitz; Food Quality and Preference; in press)

1♦ Clustering by how they describe themselves (Socio-demographic variables and behavioral attitudes).

Investigation of the relationships between preference on the one hand and usage, attitude & behavior towards the products, on the other hand.

★ Perform multiple correspondence analysis (MCA) on the qualitative variables and run a cluster analysis on the MCA scores.

★ ANOVA on preference ratings by considering the clusters as categories.

Automatic clustering of the consumers (from Howard Moskowitz: Food Quality and Preference; in press)

2 ♦ Clustering people on the basis of the pattern of their acceptance ratings.

Remarks:

★ Which strategy for clustering is better?

* clustering on the basis of the acceptance ratings?

* or clustering on the basis of the factor scores emerging from principal components analysis of these acceptance ratings.

★ Cross-tabulate these clusters with Socio-demographic, usage and attitude variables and perform a khi-squared test to investigate relationships between preference and consumers characteristics.

Automatic clustering of the consumers (from Howard Moskowitz; Food Quality and Preference; in press)

3 ♦ Clustering people on the basis of their patterns of liking versus sensory/instrumental measures.

This is most important particularly for new products development : ***the underlying directions of preference are expressed in terms of sensory/instrumental data***. This makes it possible to optimize the products in order to meet consumers' expectations.

This is the main focus of the talk!

Remark regarding the relationships between preference, sensory and consumers' backgrounds data

L-PLS strategy of analysis is one of the pioneer work to relate these data in one go!

Martens, H., Anderssen, E., Flatberg, A., Gidskehaug, L.H., Høy, M., Westad, F., Thybo, A., Martens, M.: *Regression of a data matrix on descriptors of both its rows and of its columns via latent variables: L-PLSR*. Computational Statistics & Data Analysis (2005) 48, 103-123.

EXTERNAL PREFERENCE MAPPING PREF-MAP

TO RELATE PREFERENCE DATA TO SENSORY DATA

	Attr. 1	Attr. 2	Attr. . senso 3	etc.	Consu 1	Consu 2	Consu 3	etc
Product 1								
Product 2								
Product 3								
Product 4								
etc	..							

Regression of the data of each consumer upon sensory variables

PREF-MAP : example **from** *Greenhoff K. & Macfie H. J. H. (1994). Preference mapping in practice. In Measurement of food preferences, ed. H. J. H. MacFie and D. M. H. Thomson. Blackie academic & professional.*

OBJECTIVE:

In a culinary preparation, the aim is to replace the meat by a substitute of meat.

PRODUCTS:

4 standard products: A, B, C, D.

+ 10 alternatives: a, b, c, d, e, f, g, h, i, j

SENSORY VARIABLES

9 texture attributes

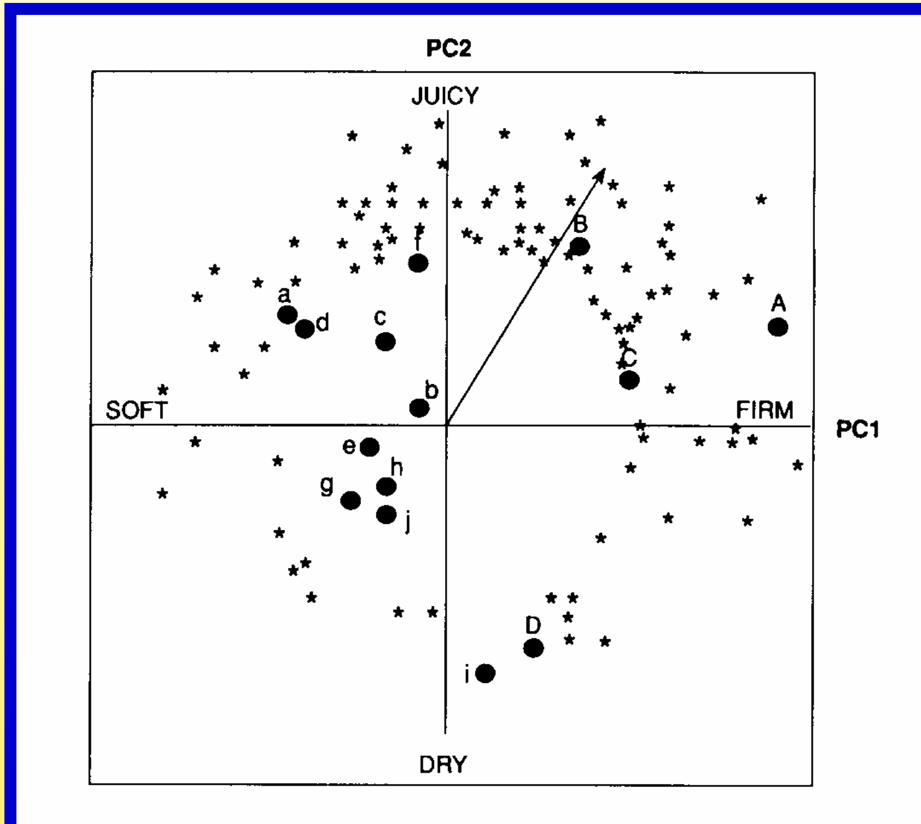
trained panel.

consumers:

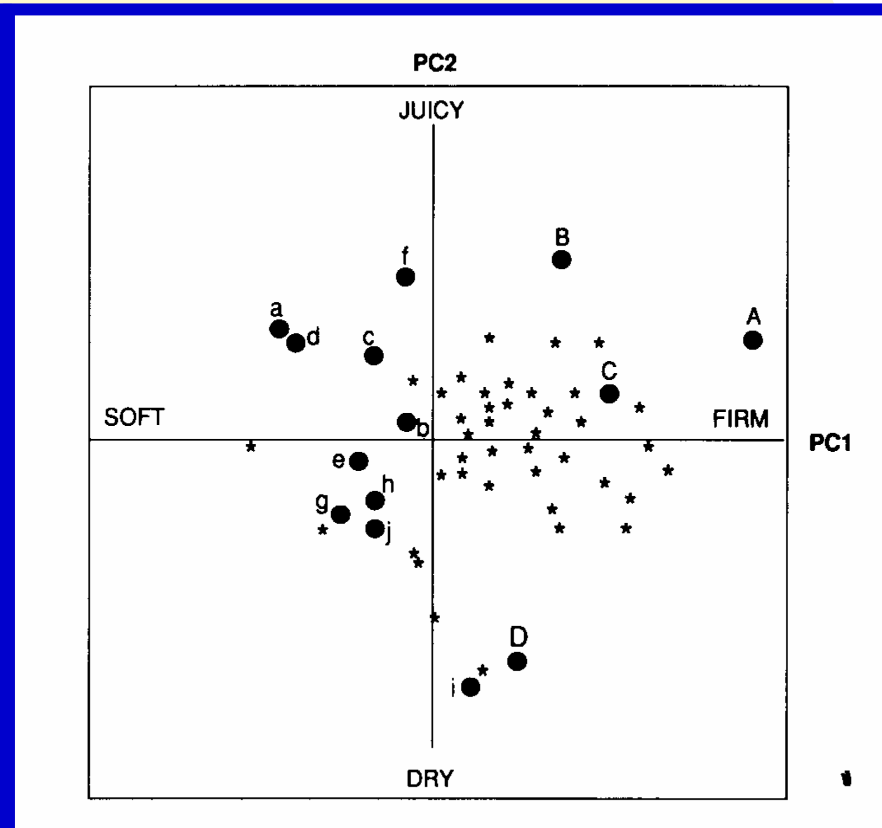
194 subjects recruited in two areas after selection

NOTATION on a scale of 9 points

PREF-MAP : example from Greenhoff K. & Macfie H. J. H. (1994). Preference mapping in practice. In Measurement of food preferences, ed. H. J. H. MacFie and D. M. H. Thomson. Blackie academic & professional.



VECTOR MODEL
(48.5 %: 94 consumers)



IDEAL POINT
(21 %: 41 consumers)

A better alternative is to use PLS2 as outlined in the following example

« Orange jus »

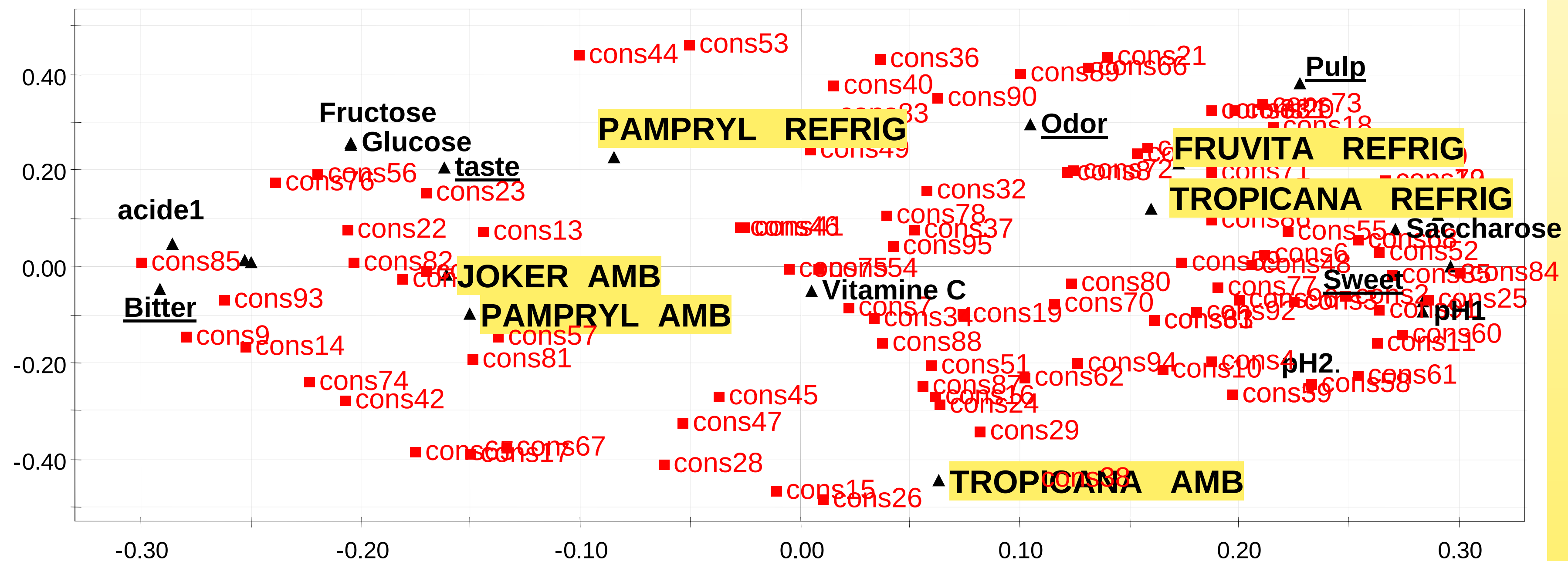
Michel Tenenhaus

Six orange jus

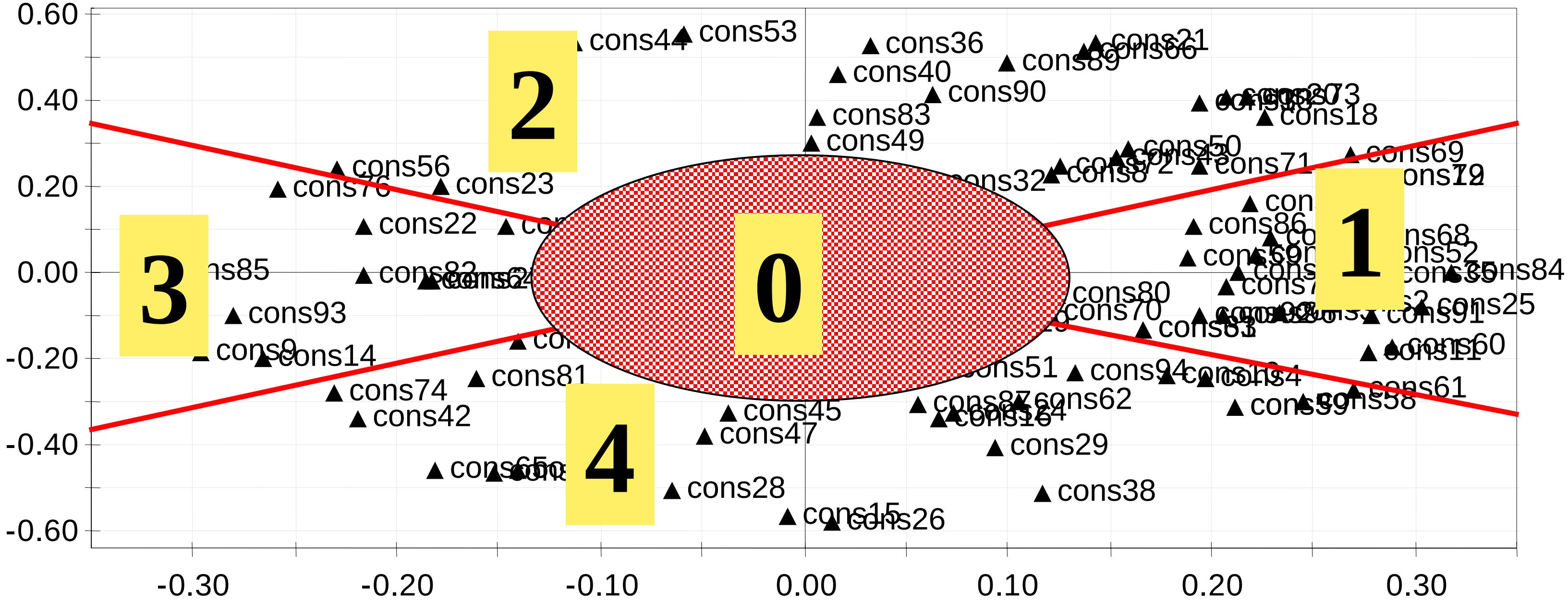
X: Physical/chemical and sensory variables

Y: preference scores of 96 consumers on 9 points liking scale.

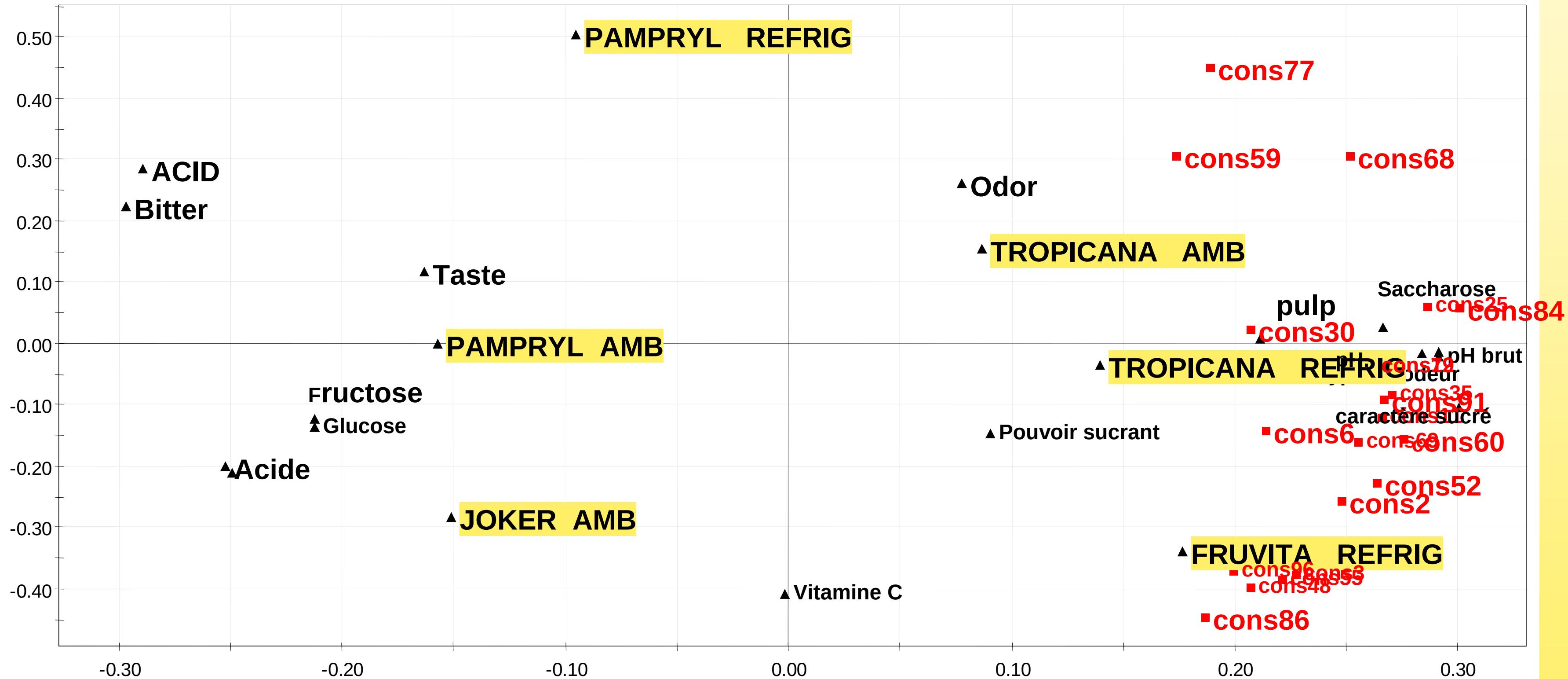
Representation of the products, sensory/chemical variables and consumers: first two PLS regression components



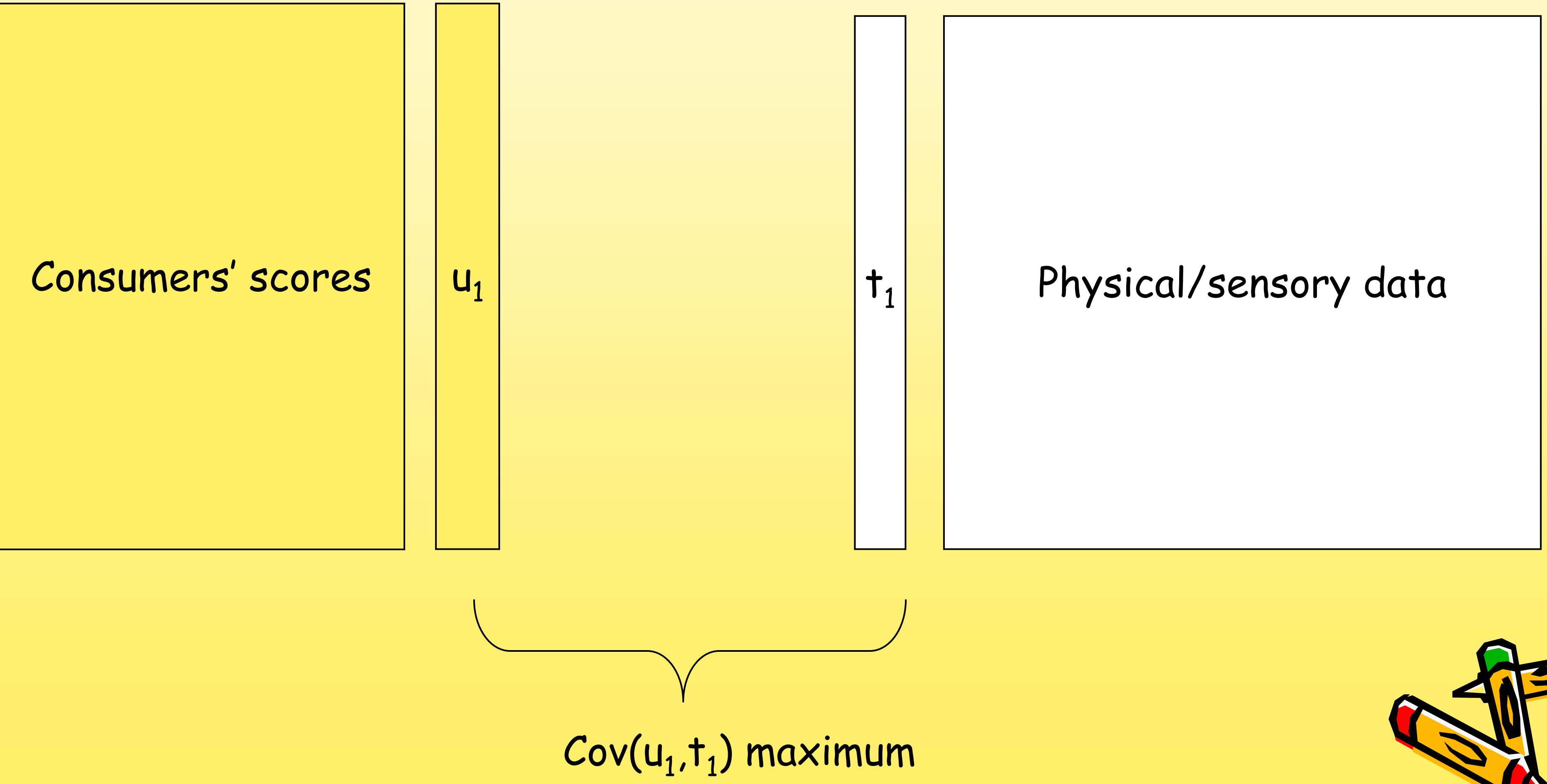
Visual cluster analysis of the consumers



PLS regression : cluster 1



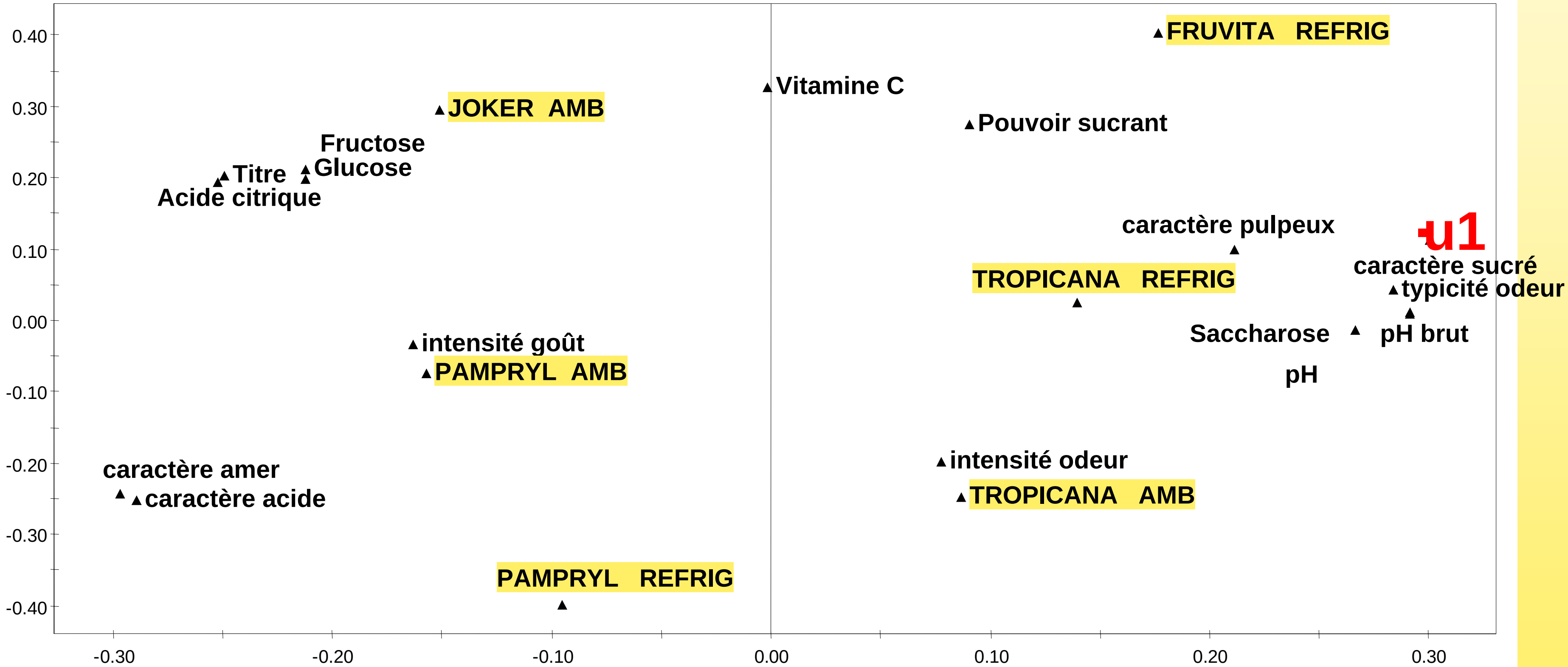
Latent directions of preferences!



Synthetic direction of preference

Jus d'orange	u_1
pampryl ambiant	-3.49159
tropicana ambiant	1.92107
fruvita réfrigéré	3.93017
joker ambiant	-3.36021
tropicana réfrigéré	3.10965
pampryl réfrigéré	-2.10908

Cluster1 : PLS Regression of u1 upon X



A summing up

When it comes to clustering the consumers (Y) taking account of external data (sensory, physical/chemical, …).

→ Projection of Y on X and perform cluster analysis on the projected data.

OR

→ Perform PLS2(Y,X) and consider u_1, u_2, … as latent directions of preferences around which revolve clusters of consumers.

2-in-1: Clustering of variables around latent components : CLV

- **Vigneau E., Qannari E. M., Punter P. H., Knoops S. (2001)**
 Segmentation of a panel of consumers using clustering of variables around latent directions of preference.
 Food Quality and Preference, 12 (5-7), 359-363.

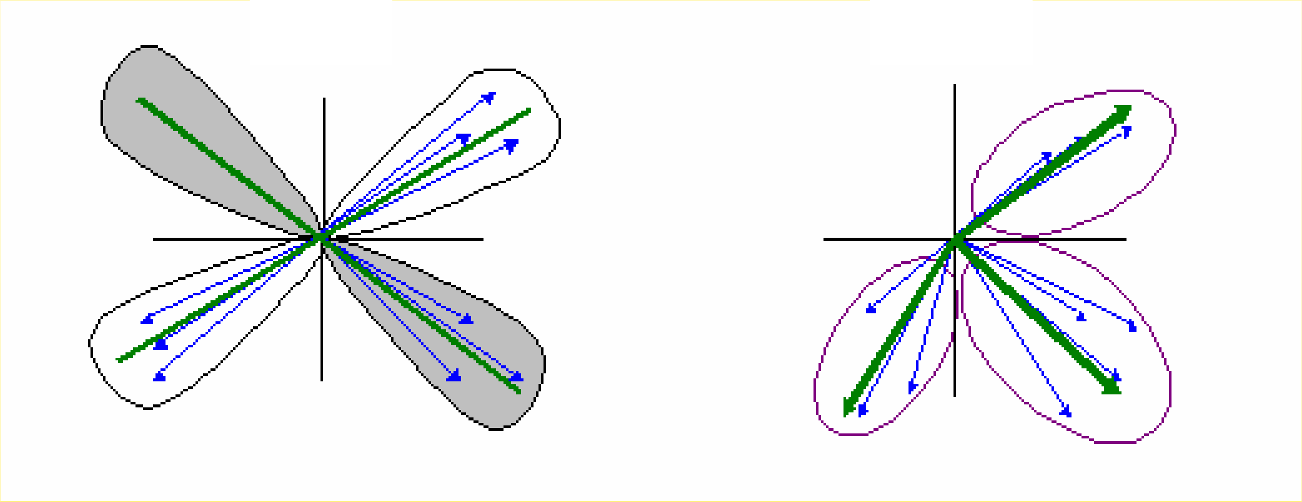
- **Vigneau E., Qannari E. M. (2002)**
Segmentation of consumers taking account of external data. A clustering of variables approach. *Food Quality and preference, 13, 515-521.*

- **Vigneau, E. and Qannari, E.M. (2003):**
 Clustering of variables around latent components,
 Communications in Statistics - Simulation and Computation, **32 (4)**, 1131-1150

- **Vigneau, E., Qannari, E.M., Sahmer, K. and Ladiray, D. (2006)**
 Classification de variables autour de composantes latentes,
 Revue de Statistique Appliquée, **LIV (1)**, 27-45

Overview of CLV

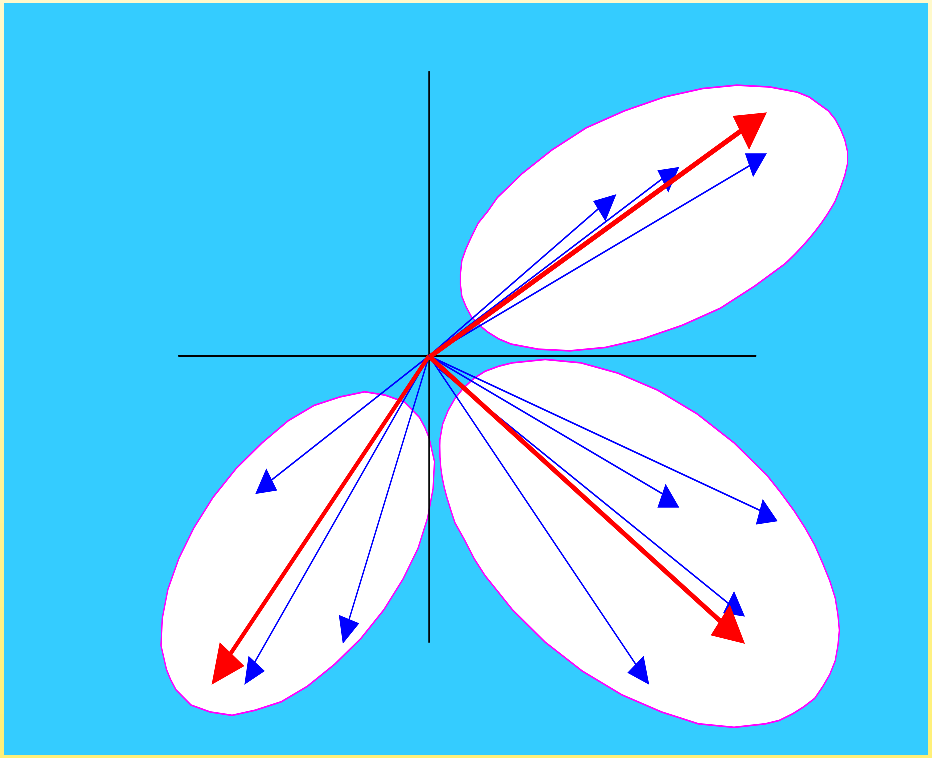
- A method of clustering of variables
- Each group G_k is represented by a latent component c_k
- Two options:



Moreover

Possibility to take account of external data

Application to the clustering of a panel of consumers.



negative correlations show disagreement among variables

Algorithm

Hierarchical algorithm

↓

Choice of the number of clusters

↓

Partitioning algorithm

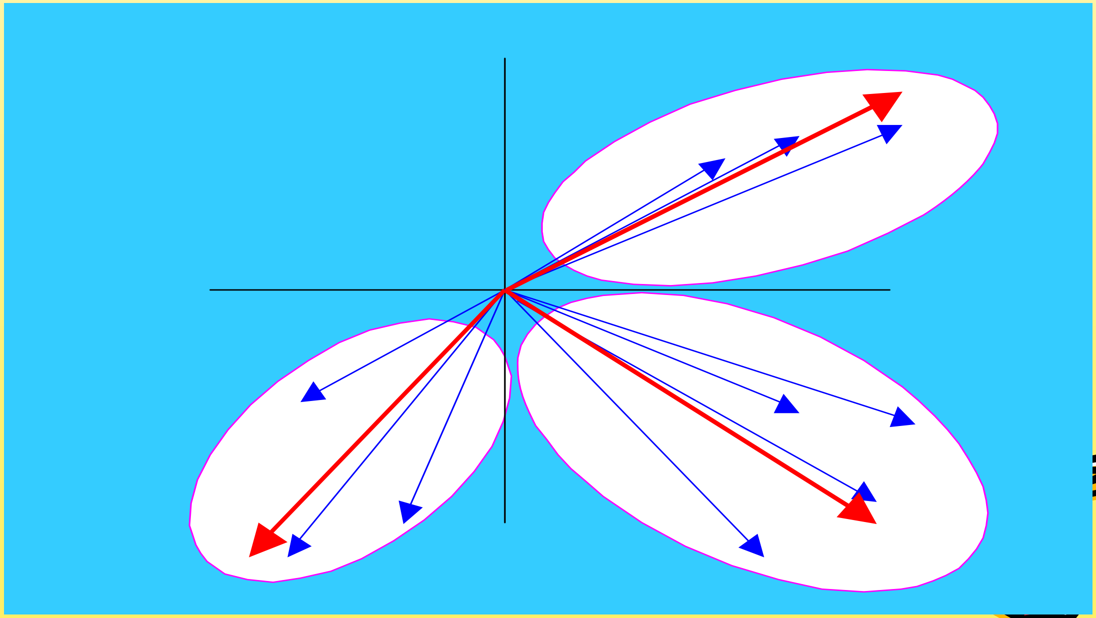
Clustering of consumers

y_1, y_2, ...y_p : the p variables (consumers) to be clustered

find K groups of variables G_1, G_2, ..., G_K
and K latent components C_1, C_2, ..., C_K such that

$$\boxed{S = \sum_{G1} Cov(y_j, C_1) + \sum_{G2} Cov(y_j, C_2) + \cdots \sum_{GK} Cov(y_j, C_K)}$$

is maximized



Clustering of consumers taking account of external data

$y_1, y_2, \ldots y_p$: the p variables (consumers) to be clustered

find <u>K groups</u> of variables $G_1, G_2, \ldots, G_K$

and <u>K latent</u> <u>components</u> $C_1, C_2, \ldots, C_K$ such that

$$\boxed{S = \sum_{G1} Cov(y_j, C_1) + \sum_{G2} Cov(y_j, C_2) + \cdots \sum_{GK} Cov(y_j, C_K)}$$

<u>is maximized</u>

When external data (X) are available, we impose the constraints:

$$\boxed{C_k = X\ a_k} \quad \text{and} \quad a_k{}^T a_k = 1$$

Latent variables in the clusters

For a given partition :

* <u>No external data</u>

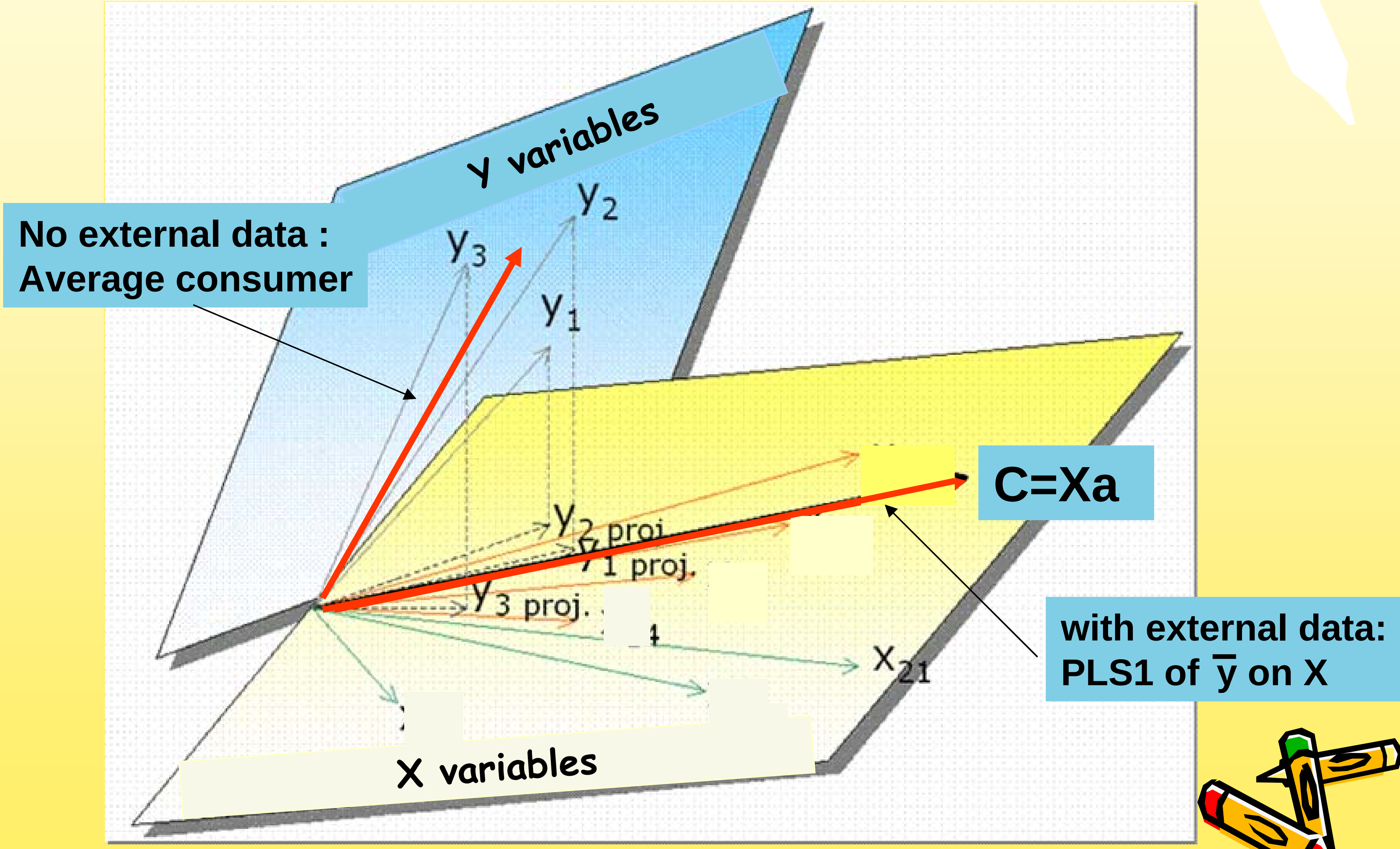
$$C_k \;\; proportional \;\; to \;\; \bar{y}_k$$

* <u>External data (X)</u>

$$C_k = X\, a_k$$

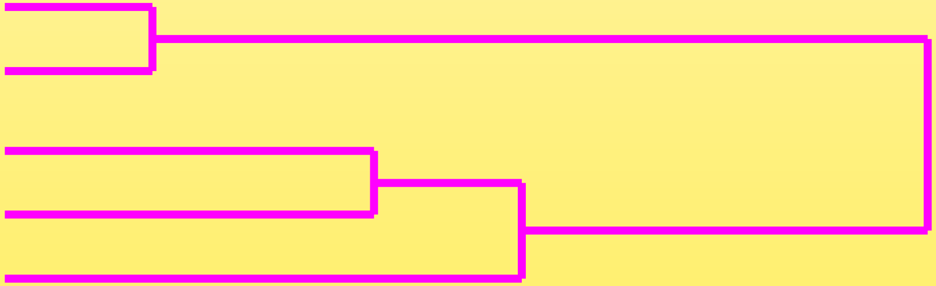
$$a_k \;\; proportional \; to \; X^T \, \bar{y}_k$$

Latent variables in each cluster



Algorithm : Hierarchical algorithm

- *Help in choosing the appropriate number of clusters.*
- *Determination of the initial partition to be used in the partitioning algorithm.*



Algorithm : Hierarchical algorithm

At the beginning, each variable (consumer) forms a cluster by itself

$$S_0 = \sum_{j=1}^{p} Std(y_j) \quad (no\ \ external\ \ data\ )$$

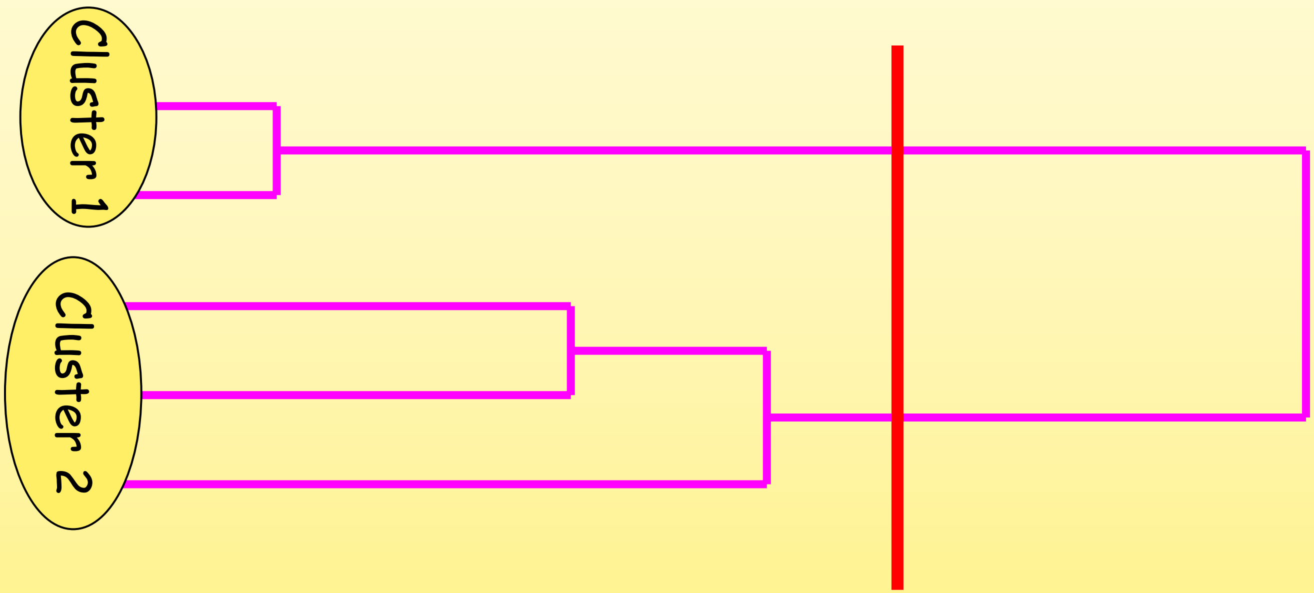
Or

$$S_0 = \sum_{j=1}^{p} \sqrt{y_j^T XX^T y_j} \quad (with\ \ X\ \ as\ \ external\ \ data\ )$$

At each stage *i*, merge the two clusters A and B for which

the decrease of $\Delta = S_{i-1} - S_i$ is the smallest.

Algorithms : Partitioning algorithm



The partitioning algorithm acts as a *consolidation* of the solution obtained by the hierarchical algorithm :
→Improves the solution in terms of the criterion to be maximized

Algorithms : Partitioning algorithm

Iteratively, until convergence :

- for k=1,..., K, define

$$C_k = X a_k$$

$$a_k \ \text{proportional to} \ X^T \bar{y}_k$$

- Variables are allowed to change memberships considering their covariance coefficients with the latent components of the different clusters.

- The latent variables are updated.

Panel segmentation in acceptability studies
An example (1/4)

<u>*Source*</u> : ESN Coffee experiment (1996)

- Overall liking scores of
- 160 French consumers
- for 8 coffees

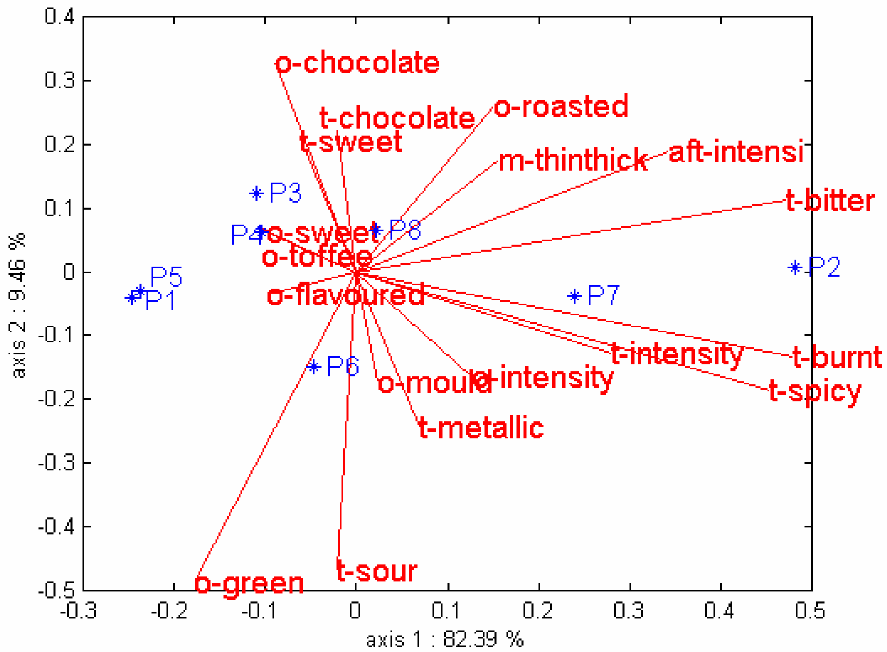
Y

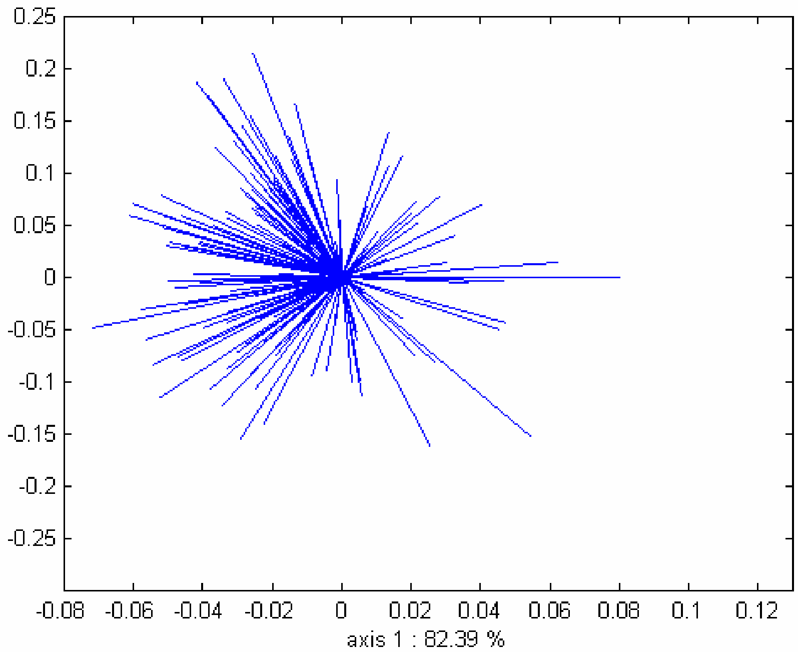
- 18 sensory attributes

X

Panel segmentation in acceptability studies
An example (2/4)

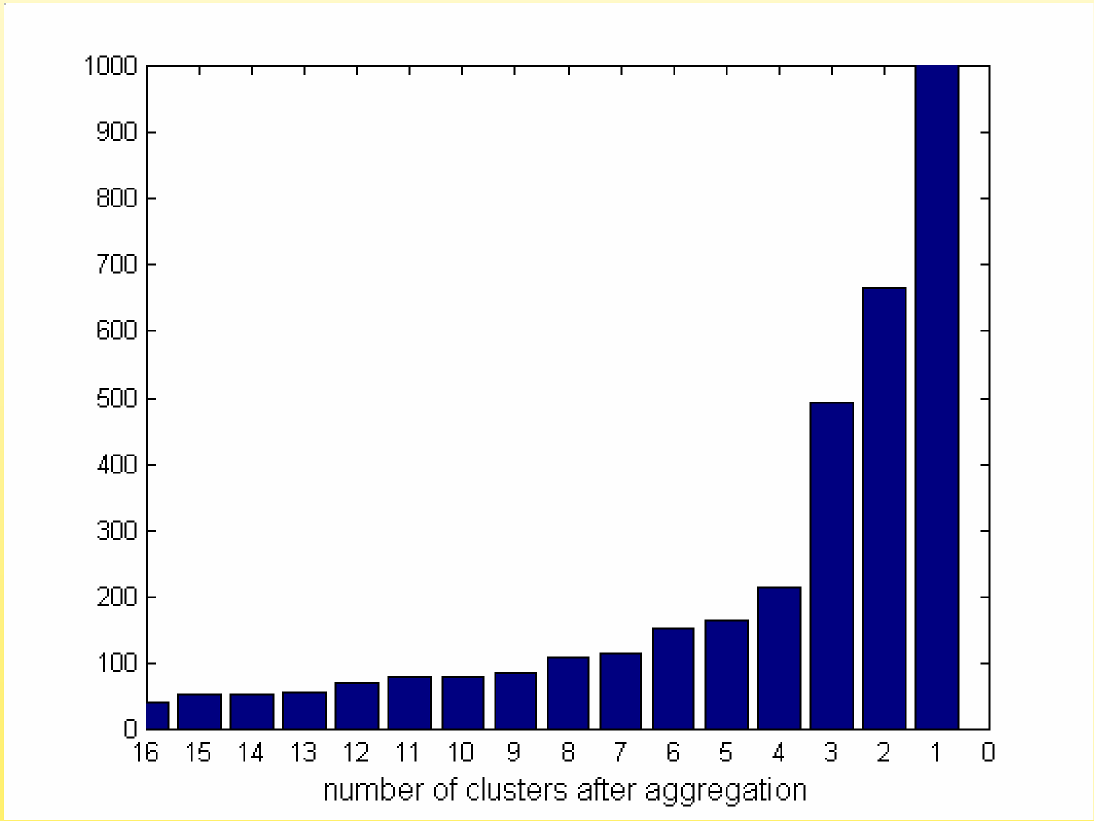
Two main displays resulting from PrefMap (with vector model)





Panel segmentation in acceptability studies
An example (3/4)

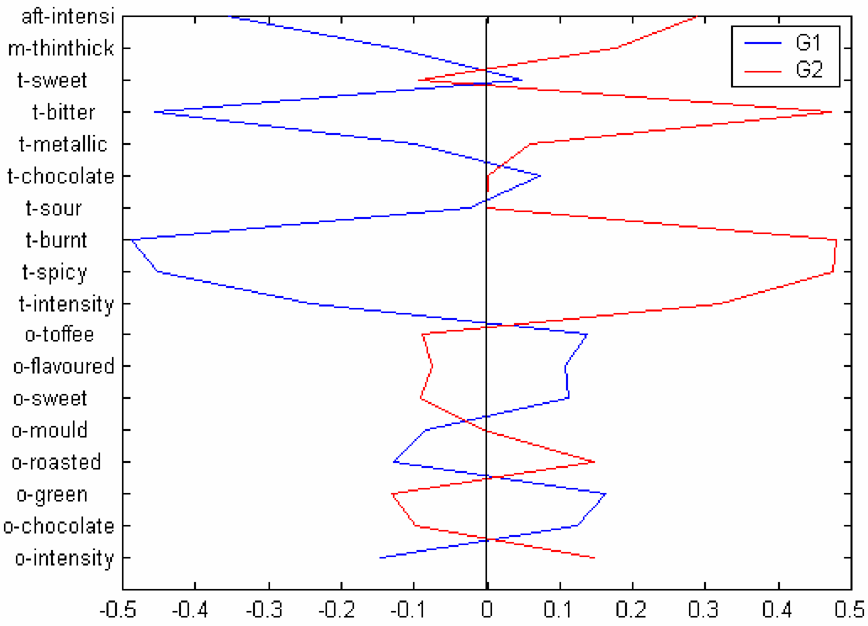
hierarchical algorithm ➔ Choice of the number of segments



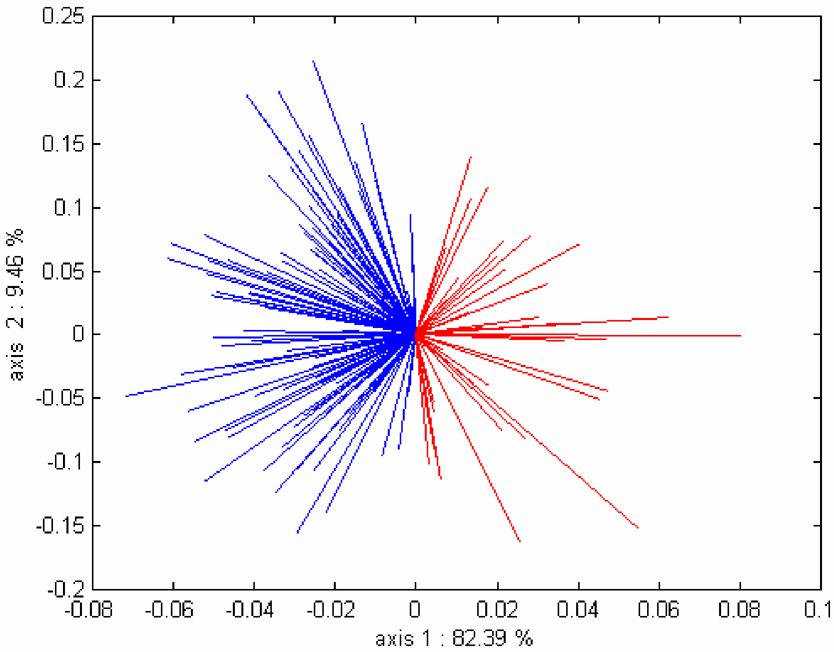
Evolution of the aggregation criterion $\Delta = S_{i-1} - S_i$

Panel segmentation in acceptability studies
An example (4/4)

- G1 : 120 consumers
- G2 : 40 consumers

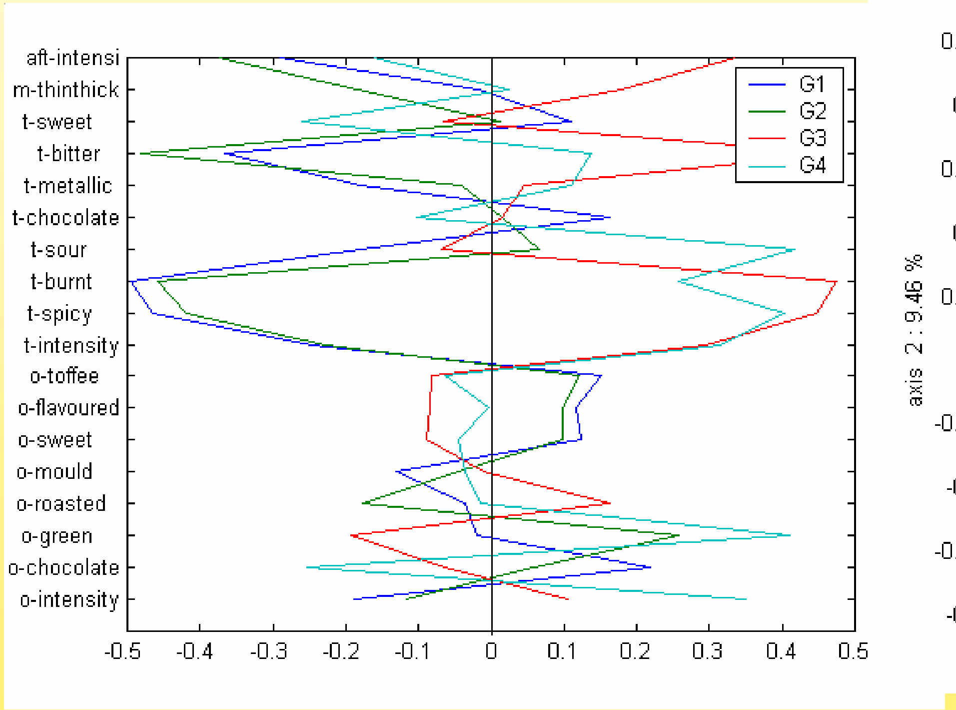


Coefficients (loadings) of the latent components associated with the segments

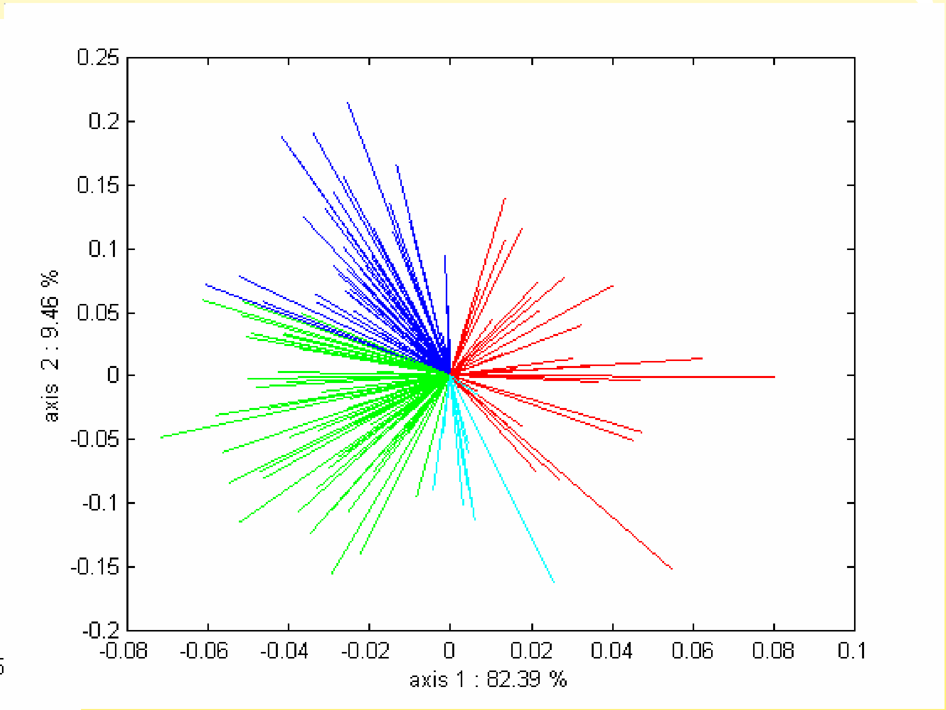


Identification of the segment on the map obtained with PrefMap

- G1 : 52 consumers
- G2 : 65 consumers
- G3 : 30 consumers
- G4 : 13 consumers



Coefficients (loadings) of the latent components associated with the segments



Identification of the segment on the map obtained with PrefMap

Conclusion

Clustering of variables around latent components under constraints

Complementarity with PrefMap method

~~One model per consumer~~
One model per segment
Direct segmentation

Further investigations :

Cluster analysis taking account of sensory and consumers’ backgrounds