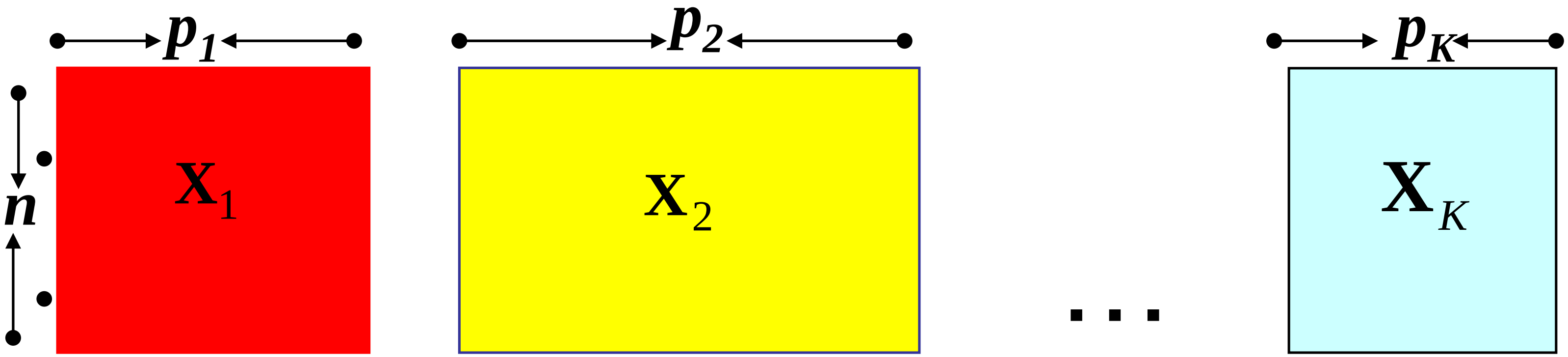
Extension of ComDim for the analysis of (K+1) datasets; Application in Sensometrics

A. El Ghaziri, V. Cariou,
D. N. Rutledge & E. M. Qannari

Introduction

- ✓ Formerly, called *Common components and specific weights analysis* (CCSWA);
- ✓ The new name is *ComDim;*
- ✓ Today, a new baby is born: *P-ComDim;*
- ✓ …More babies are on the way.

ComDim: the setting



Different types of multivariate data are measured on the same individuals.

- *Examples* :
  - Sensory analysis ➔ fixed or free choice profiling.
  - Process technology ➔ multivariate measurements are performed at different stages of the process.
  - Functional Genomics ➔ genetic data, molecular data, phenotypic data...

ComDim: how does it work?

X: → $W = XX^T$: $W = Q\Lambda Q^T$

(standardized) principal components

Variances recovered by the principal components

ComDim : how does it work?

X: → $W = XX^T$: $W = Q\Lambda Q^T$

(standardized) principal components

Variances recovered by the principal components

With several X_k

$W_k = X_k X_k^T$: nxn matrix of scalar products between individuals

$$W_k \cong Q \Lambda_k Q^T$$

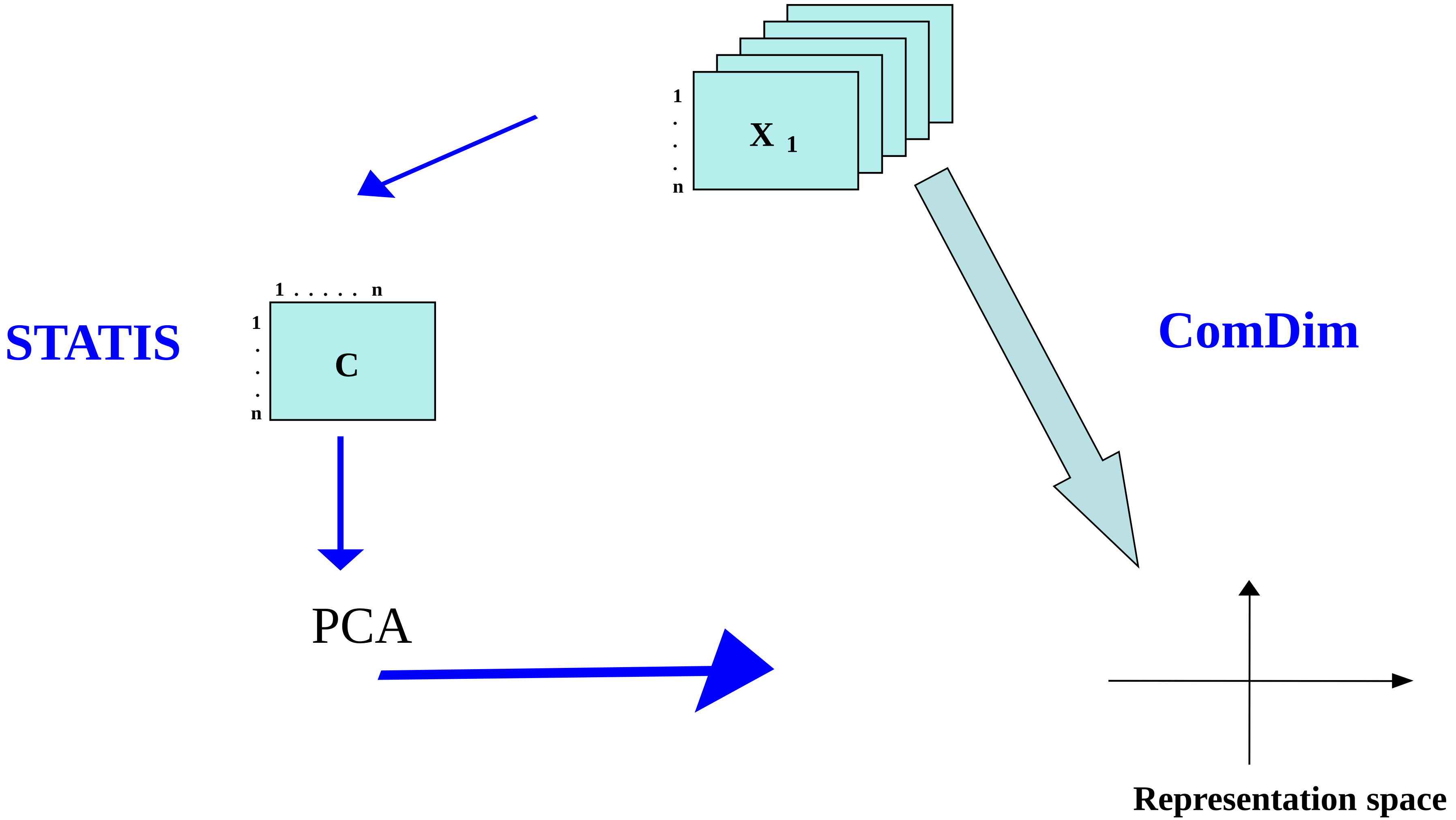
Common Principal Components

Diagonal matrix : the saliences : weights, total variances recovered by the various common components for dataset X_k.

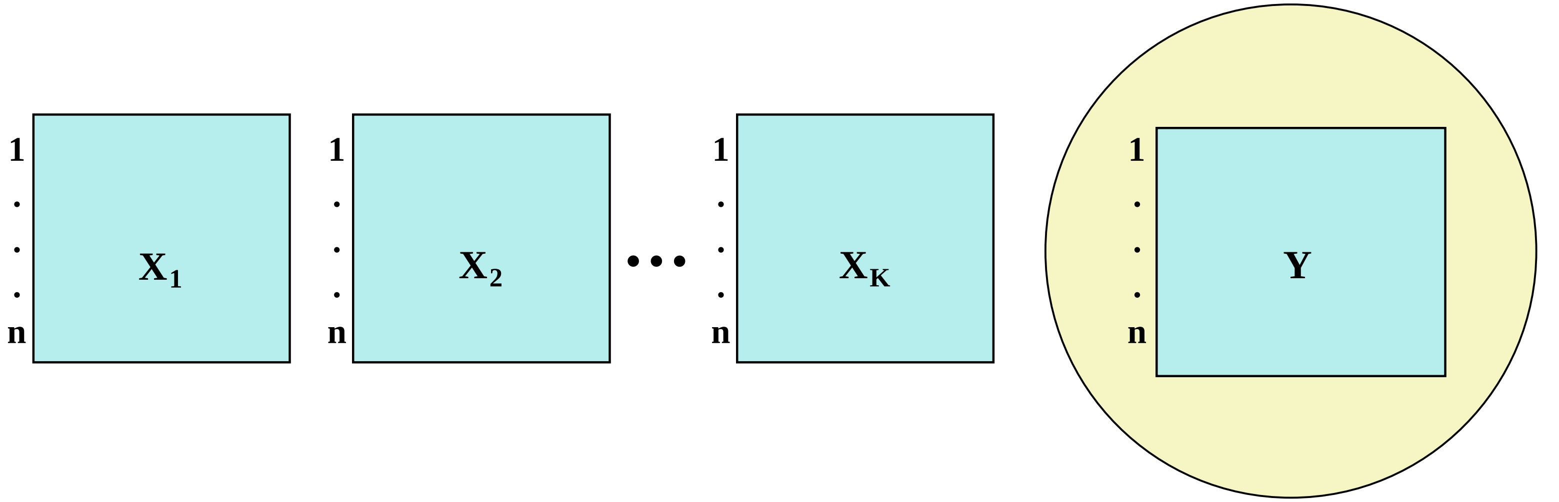
ComDim and Indscal

- **Same premises as *Indscal; but*** :
  - No degenerate solutions;
  - Simple algorithm;
  - Orthogonal dimensions;
  - Embedded spaces of representation;
  - Several properties which are useful for the interpretation;
  - Can be extended to handle various situations.

ComDim and STATIS



P-ComDim



Example :

X_k : Sensory profiling datasets (flavor, aroma…)
Y : Preference scores.

Objective : Explore the relationships between X_k and Y.

Strategy of analysis

- Consider
  - The matrices $S_k = X_k X_k^T Y Y^T$: an nxn matrix
  - Find an (approximate) common SVD:

$$S_k \cong T \Lambda_k U^T$$

Properties

The components in T are in the X_k spaces: useful to depict the products configuration.

The components in U are in the Y space, optimally linked to the components in T.

- Λ_k are the saliences which reflect the importance of each dimension for X_k <->Y

$S_k = X_k X_k^T Y Y^T$

How to find the (approximate) common SVD?

- Step by step; each step is followed by a deflation procedure.
- Four algorithms;

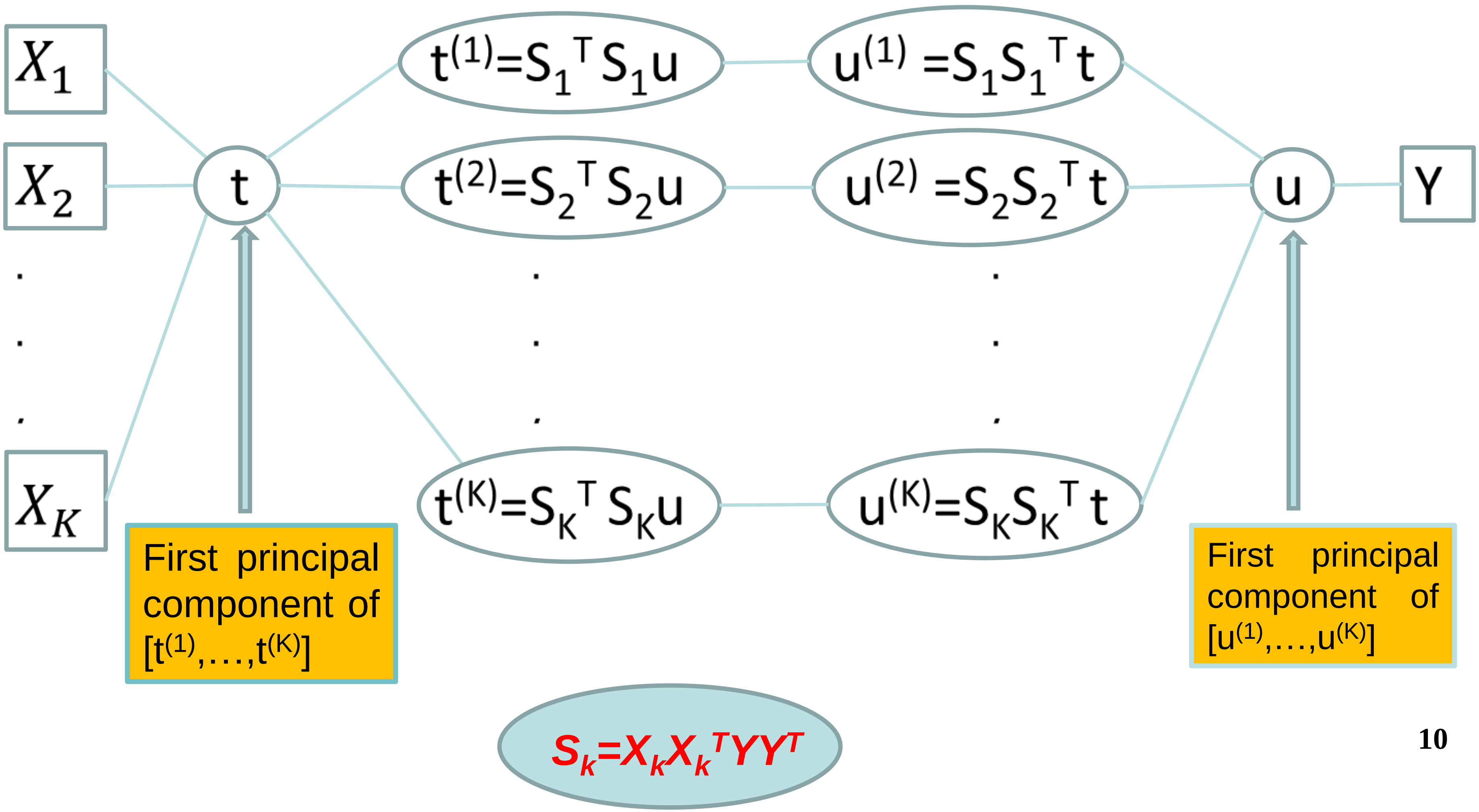


Illustration: relating sensory to preference data

The study concerns eight American dry-cured ham products labeled 1SHORT, 2SHORT, 3SHORT, 4SHORT, 5SHORT, 1LONG, 2LONG and 3LONG.

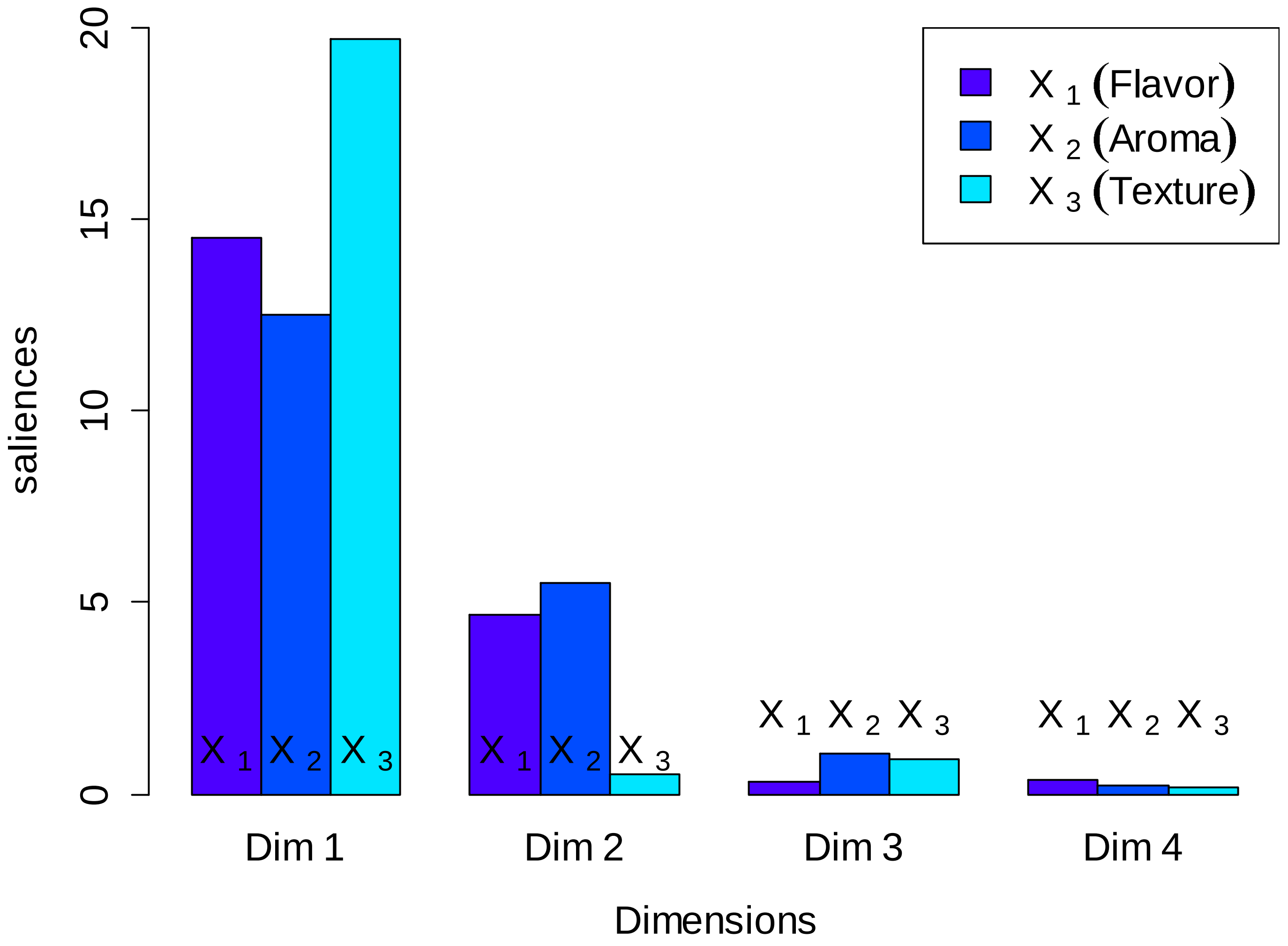
Pham, A. J.; et al (2008).
Relationships between sensory descriptors, consumer acceptability and volatile flavor compounds of American dry-cured ham. Meat Science.

- **Sensory evaluation**:
  - X_1: Flavor (11 attributes),
  - X_2: aroma (8 attributes),
  - X_3: texture (6 attributes).
- **Preference data**

Consumer acceptability of the eight dry-cured hams was evaluated by a panel of 71 consumers who were segmented in six groups

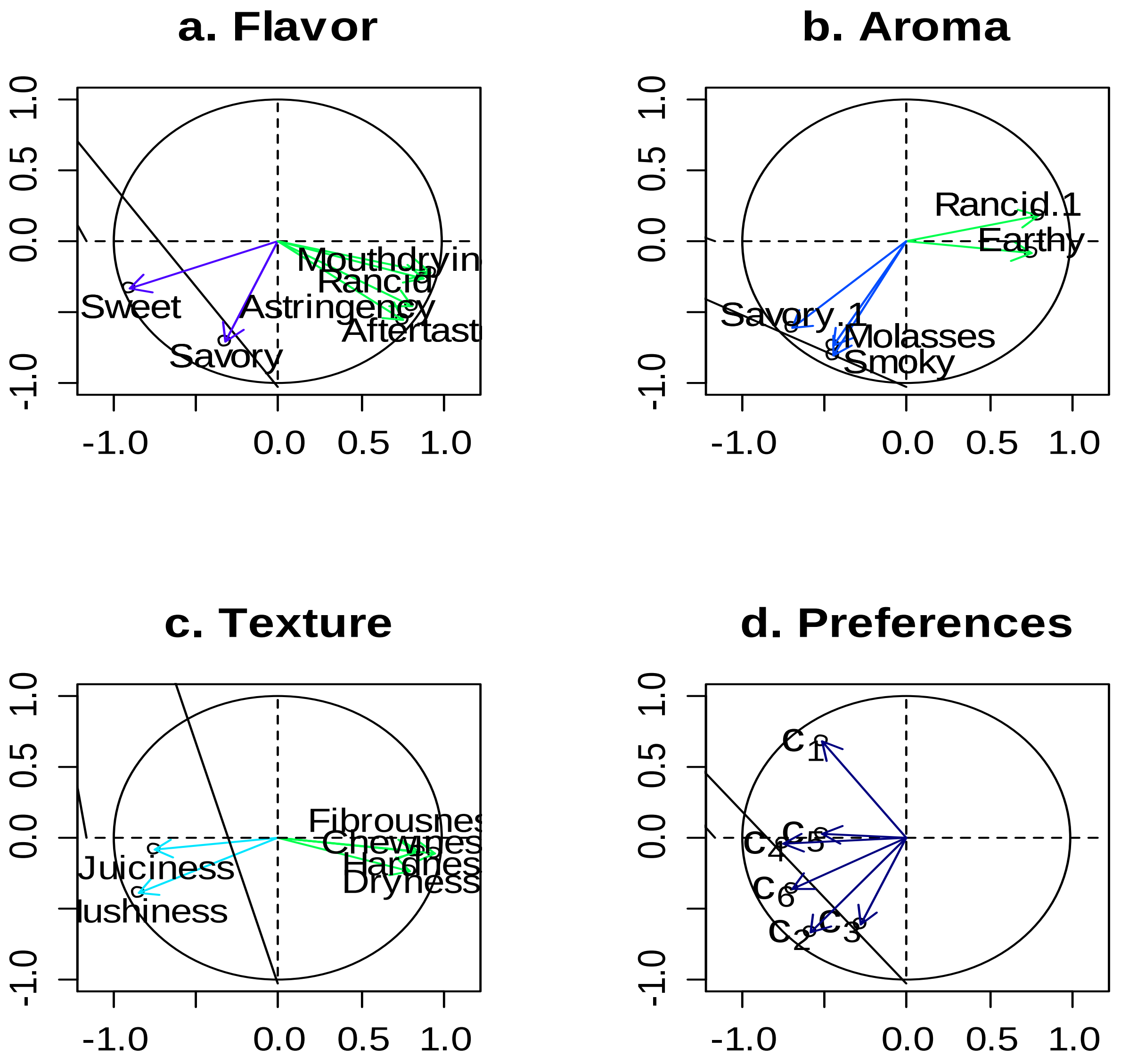
$Y = [c_1, c_2, c_3, c_4, c_5, c_6]$: average scores in the six segments

Saliences

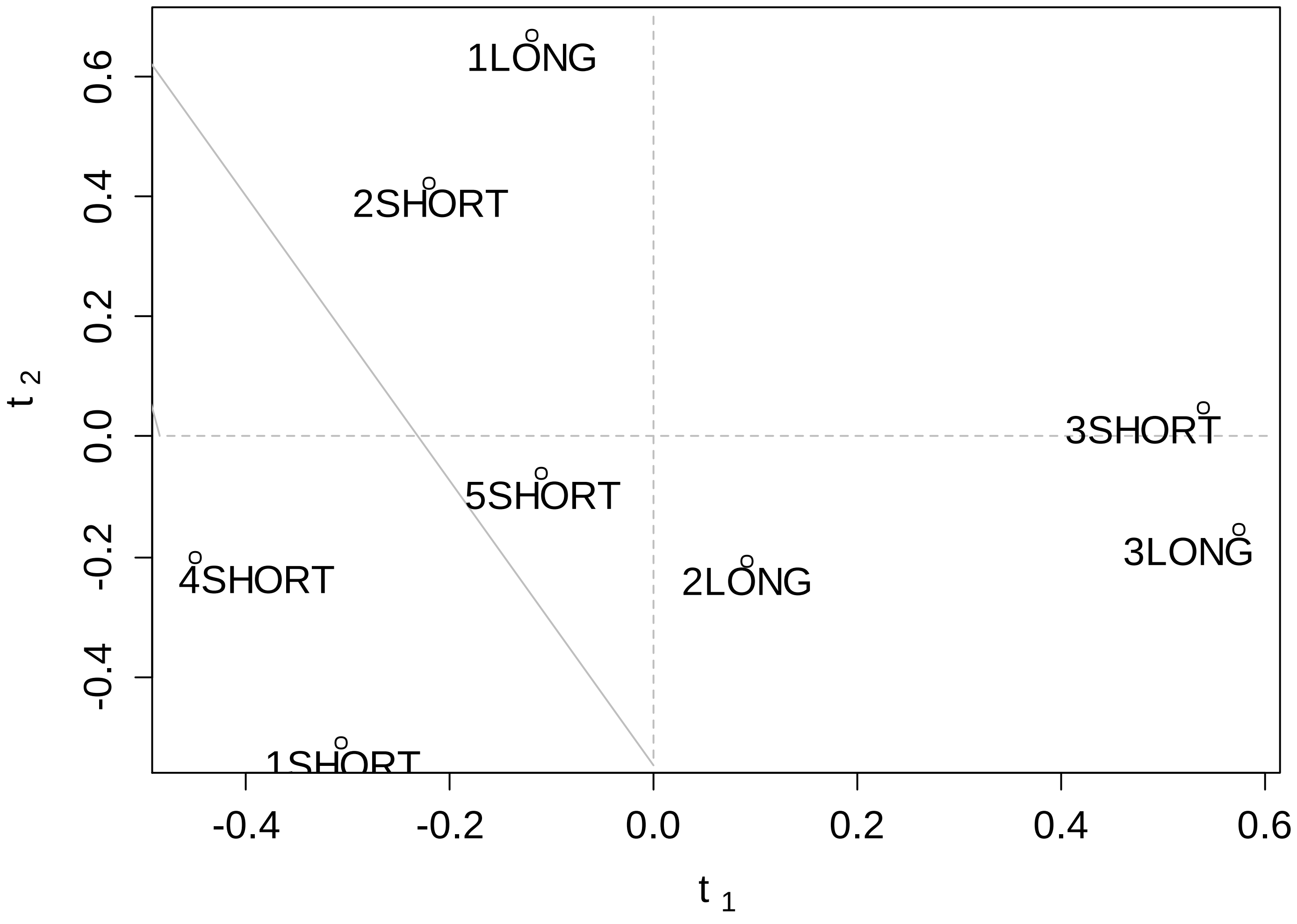


20
15
10
5
0
saliences
X 1 (Flavor)
X 2 (Aroma)
X 3 (Texture)
X 1 X 2 X 3
X 1 X 2 X 3
X 1 X 2 X 3
X 1 X 2 X 3
Dim 1
Dim 2
Dim 3
Dim 4
Dimensions

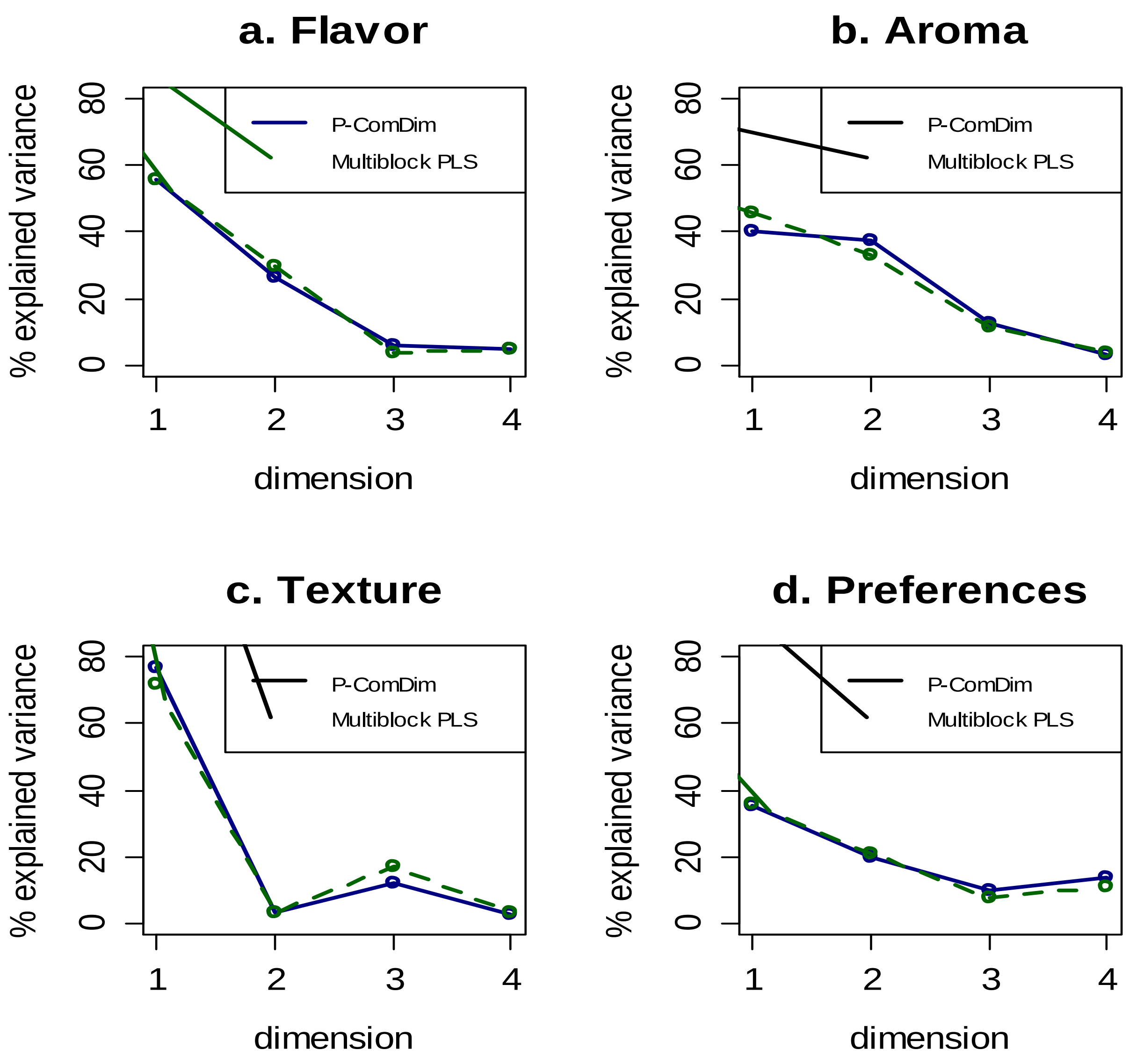
Correlation of the sensory and preference variables with t_1 and t_2



Products configuration on the basis of t_1 and t_2



Comparison with multiblock PLS



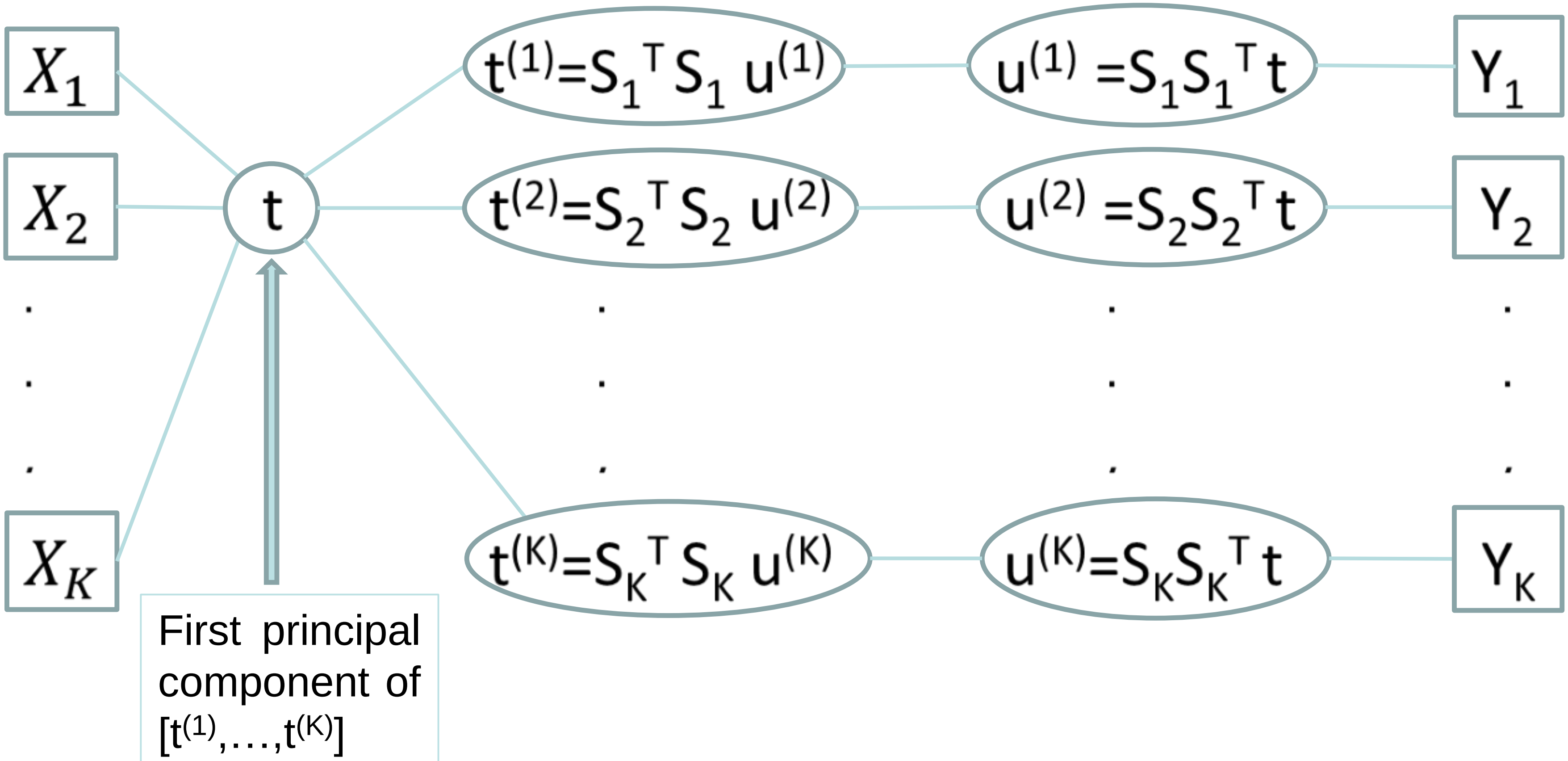
Conclusion

- *P-ComDim* : same performance as multiblock PLS in terms of prediction;
- Provides tools to better interpret the outputs
- Suggests more extensions :

  For instance :

  - X_1: Flavor (11 attributes), Y_1 : Preference of flavor
  - X_2: aroma (8 attributes), Y_2 : Preference of aroma
  - X_3: texture (6 attributes), Y_3 : Preference of texture

Conclusion



$$S_k = X_k X_k^T Y_k Y_k^T$$

Conclusion

- Extension to a path modeling framework