Quadratic PLS applied to external preference mapping

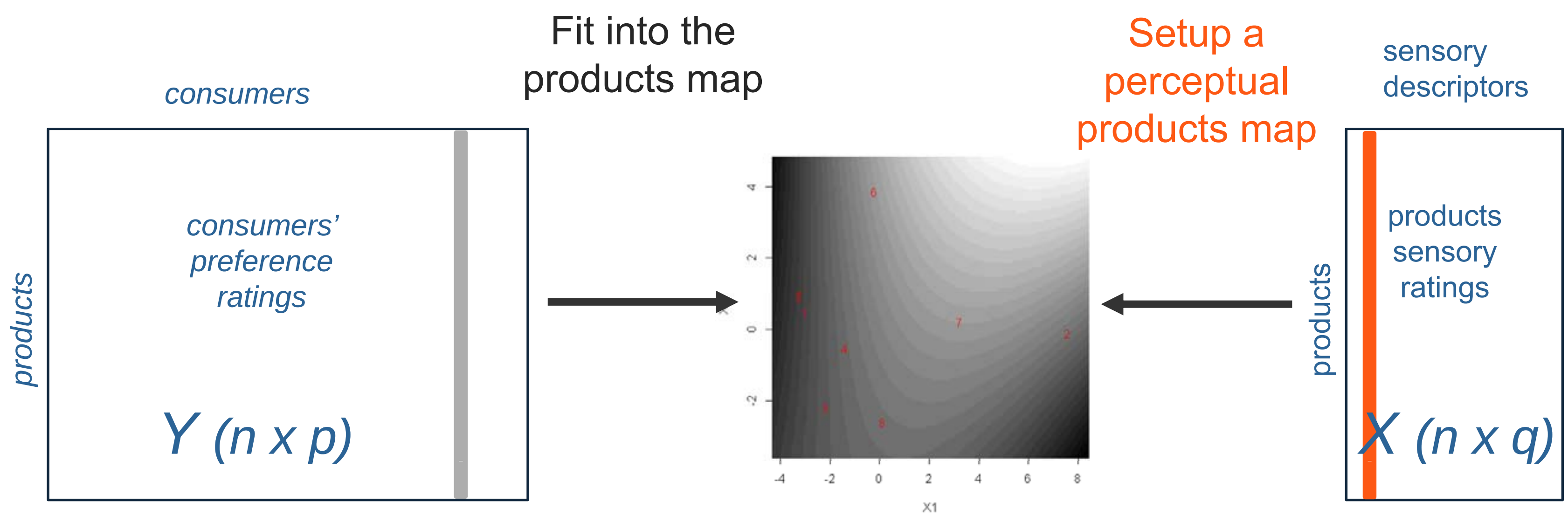
Stéphane Verdun, Véronique Cariou, El Mostafa Qannari

ONIRIS, Sensometrics and Chemometrics Laboratory,
Nantes, France

Context of Preference Mapping:

- relate data on consumers' preference to information on products
- identify attributes drivers of liking

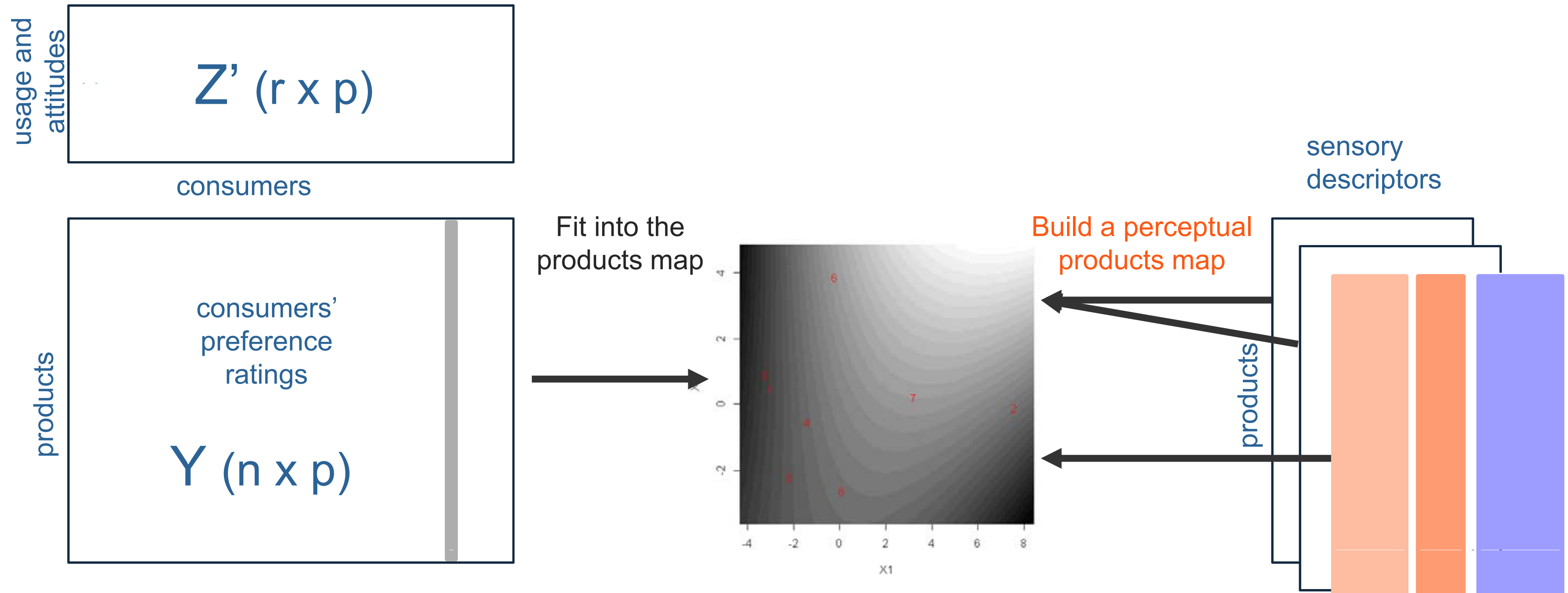
Usual implementation of external preference mapping:



Preference mapping

Extension to:

- sensory descriptors as a single consensus data table vs multiple data blocks vs multiway data
- information on consumers with a L-shaped data structure

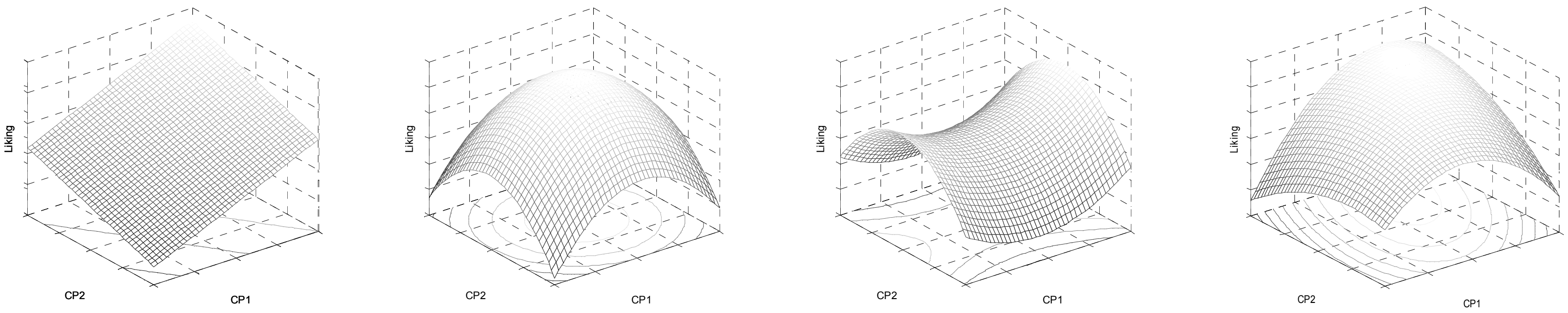


See Dijksterhuis & al. (2005) ; Lengard & al. (2006) ; Endrizzi & al. (2010) ; Mage & al. (2012).

Preference mapping

Adapt model to fit:

- vectorial vs polynomial models



Setting up of the products perceptual map:

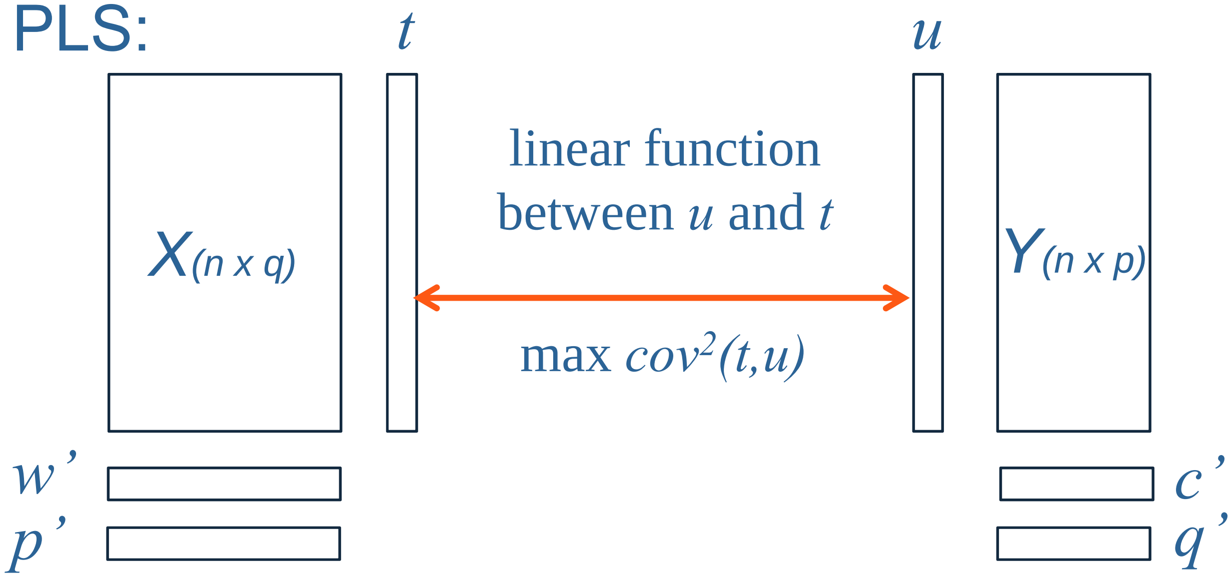
- use of PCA vs PLS

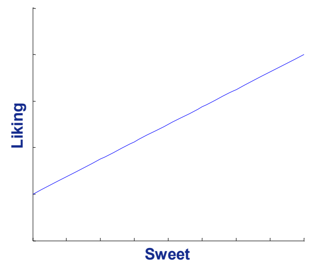
Plan

- Models of Quadratic PLS
- Extension of Hoskuldsson's approach to several Y : QPLS2
- α - Regularization of QPLS2
- Application to Preference Mapping : The Coffee dataset
- Conclusion

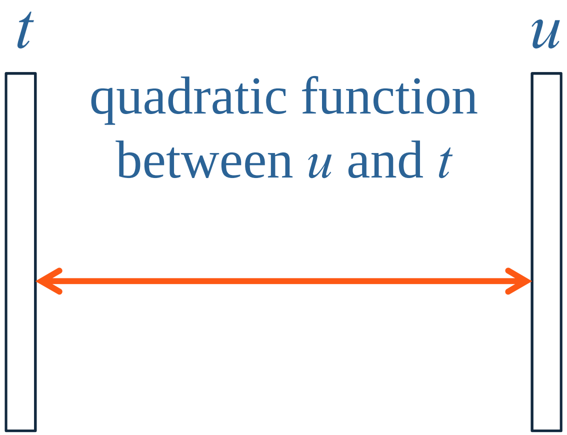
Models of Quadratic PLS

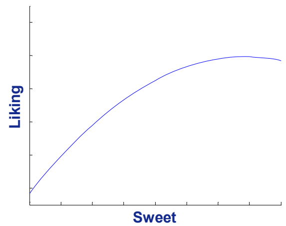
From PLS:





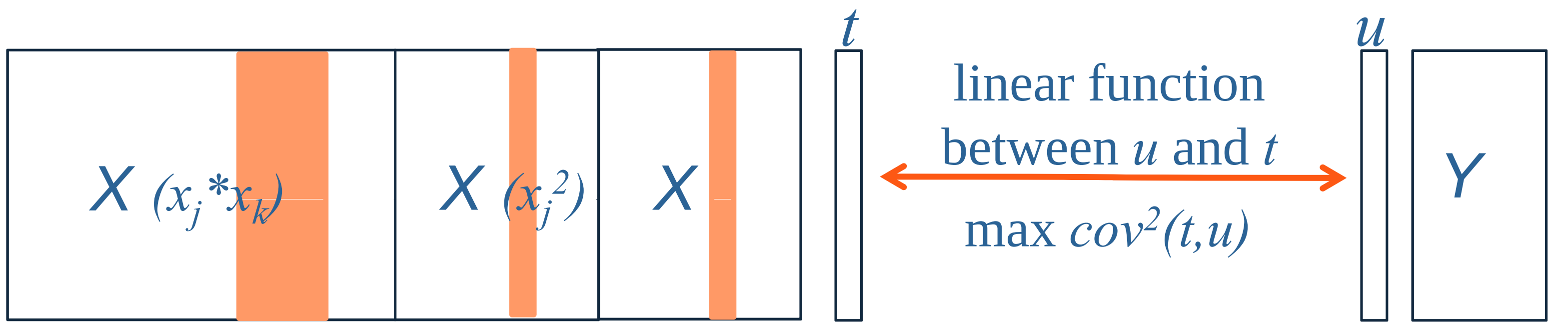
To quadratic PLS:



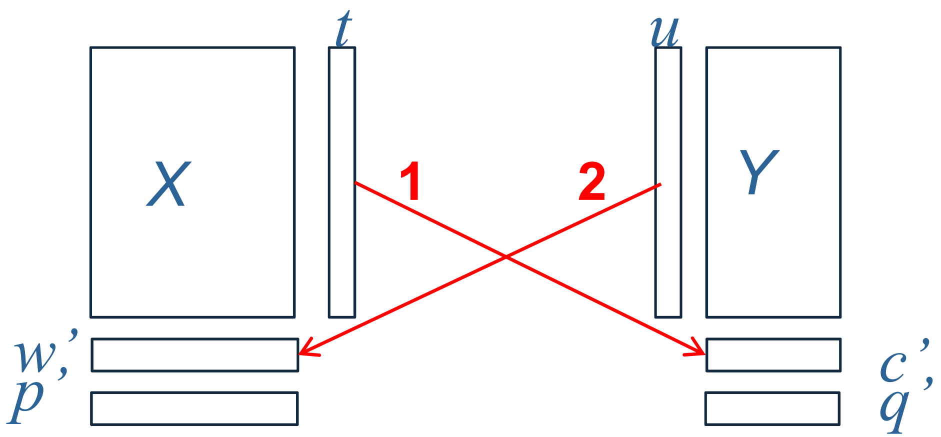


Models of Quadratic PLS

Gnanadesikan (1977), Wold & al. (1984), …



Wold & al. (1989), Baffi, Martin and Morris (1999)



1. **Estimate *c* by a quadratic regression of u on t**
2. **Update *w* by a Newton-Raphson-like linearization of the quadratic inner relation estimated by linear PLS**

Models of Quadratic PLS : Höskuldsson's approach

Höskuldsson (1992) extends the PLS1 criterion with the quadratic and interactions terms.

At step h, he seeks to maximize:

$$\operatorname{cov}^2(y,t_h)+\operatorname{cov}^2(y,t_h^2)+\operatorname{cov}^2(y,t_h t_1)+\ldots+\operatorname{cov}^2(y,t_h t_{h-1})$$

given $t_1, \ldots, t_{h-1}$ already known, and $t_h=X_{h-1}w_h$ with $\|w_h\|=1$.

On the basis of this criterion, Verdun & al. (2012) have proposed a revision of the original algorithm in order to guarantee convergence and optimality.

This latter one is extended in the case of QPLS2.

Plan

- Models of Quadratic PLS
- Extension of Hoskuldsson's approach to several Y : QPLS2
- α - Regularization of QPLS2
- Application to Preference Mapping : The Coffee dataset
- Conclusion

Extension of Hoskuldsson's approach to several Y: QPLS2

The criterion can be modified to handle the case where Y has several variables.

At a step *h*, the algorithm seeks a component $u_h = Y_{h-1} c_h$ and a component $t_h = X_{h-1} w_h$ that maximise:

$$\operatorname{cov}^2(u_h, t_h) + \operatorname{cov}^2(u_h, t_h^2) + \operatorname{cov}^2(u_h, t_h t_1) + \ldots + \operatorname{cov}^2(u_h, t_h t_{h-1})$$

Extension of Hoskuldsson's approach to several Y: some limitations

The criterion uses the covariance between u on the one hand, and t and t^2 on the other hand. These variables can be at very different scale levels.

If t^2 variance is large compared to the variance of t, the objective criterion will be dominated by t^2.

Plan

- Models of Quadratic PLS
- Extension of Hoskuldsson’s approach to several Y : QPLS2
- α - Regularization of QPLS2
- Application to Preference Mapping : The Coffee dataset
- Conclusion

α - Regularization of QPLS2

A parameter α is introduced to balance the linear and the quadratic terms, $0 \leq \alpha \leq 1$:

$$\alpha \operatorname{cov}^2(u_h, t_h) + (1-\alpha)\operatorname{cov}^2(u_h, t_h^2)$$

α is set to the value that leads to the highest R_Y^2 coefficient of the quadratic model.

During the deflation step, Y is deflated using t and t^2.

Plan

- Models of Quadratic PLS
- Extension of Hoskuldsson's approach to several Y : QPLS2
- α - Regularization of QPLS2
- Application to Preference Mapping : The Coffee dataset
- Conclusion

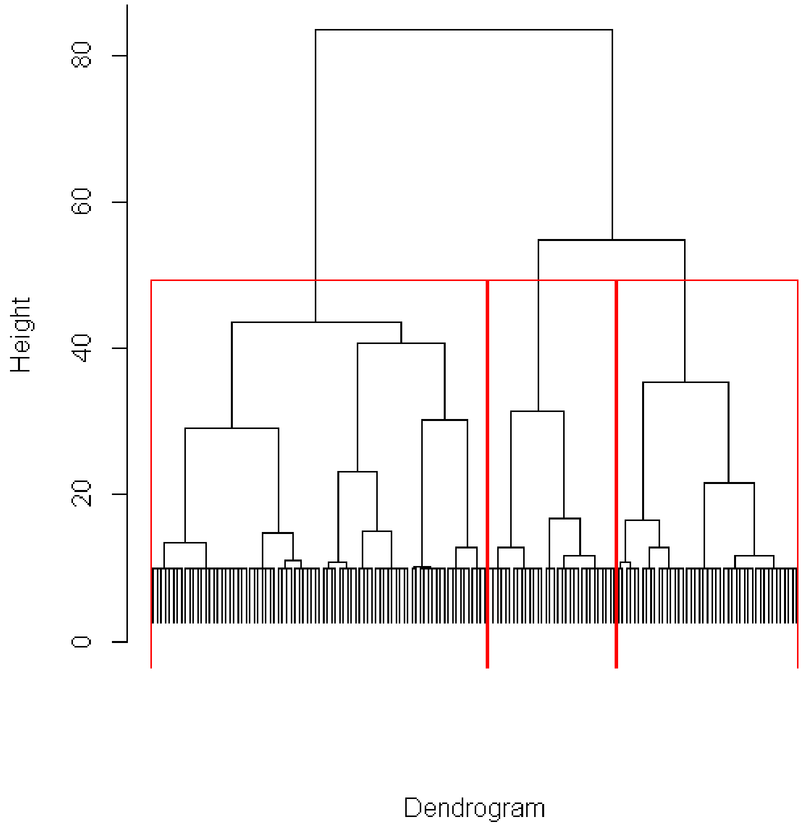
Application to Preference Mapping : The Coffee dataset (ESN, 1996)

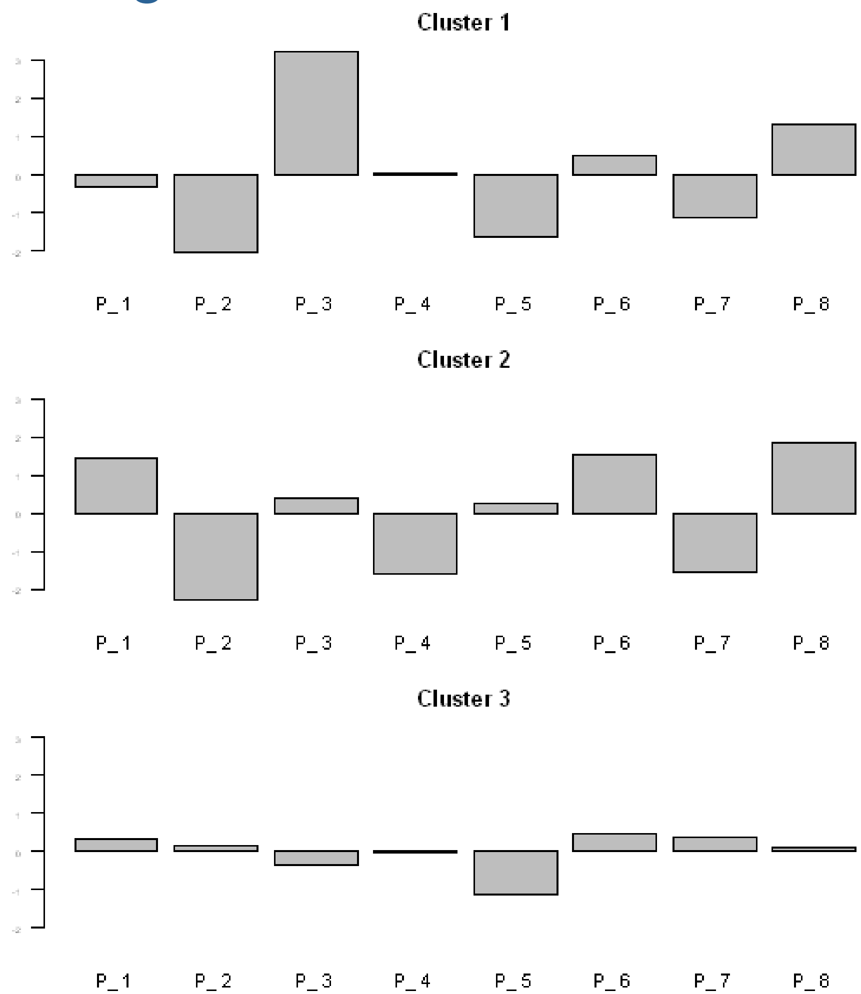
Data

- 8 coffees
- 160 french consumers who have been partitionned in 3 clusters
- Quantitative descriptive analysis: 23 sensory descriptors
    - Smell descriptors : chocolate, intensity, moisty, sweet...
    - Taste descriptors : sour, chocolate, metallic...

Application to Preference Mapping : The Coffee dataset

Results of the Hierarchical Clustering:

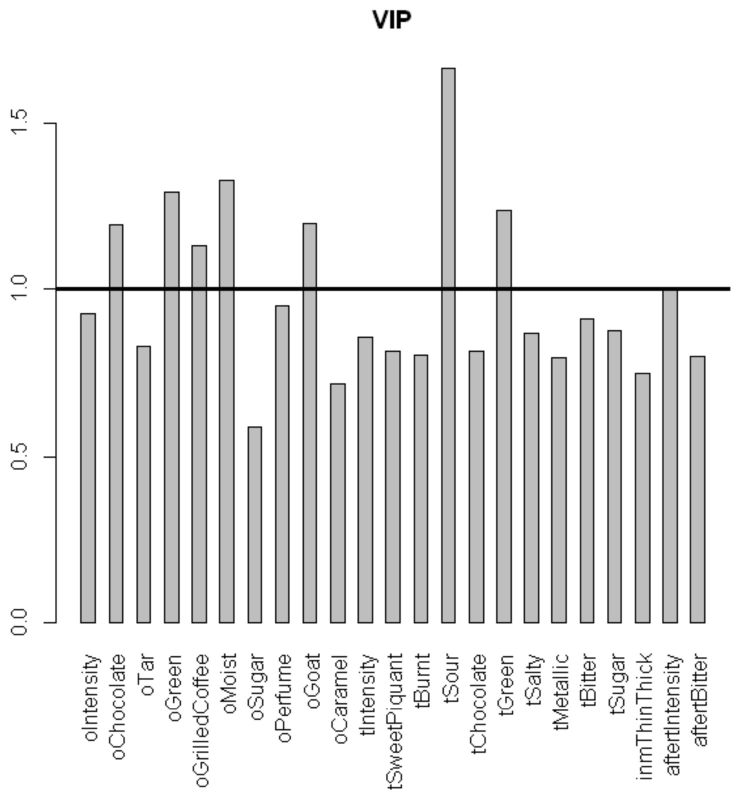


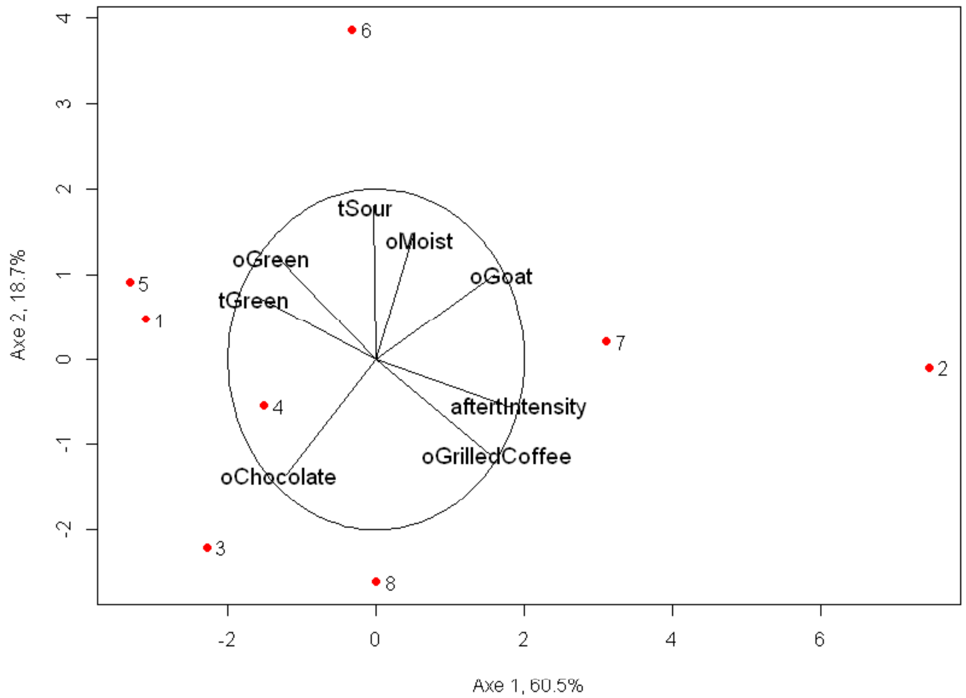


Application to Preference Mapping : The Coffee dataset

Products map obtained by QPLS:

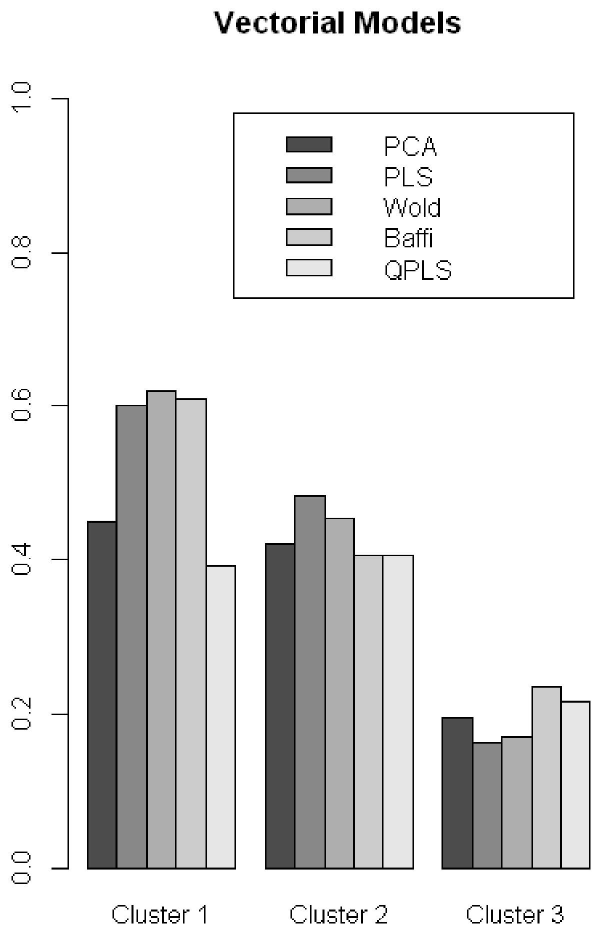
- **~ 79% of the variance of X explained by t_1 and t_2**
- **~ 80% of the variance of Y explained by the quadratic model (t_1, t_2)**

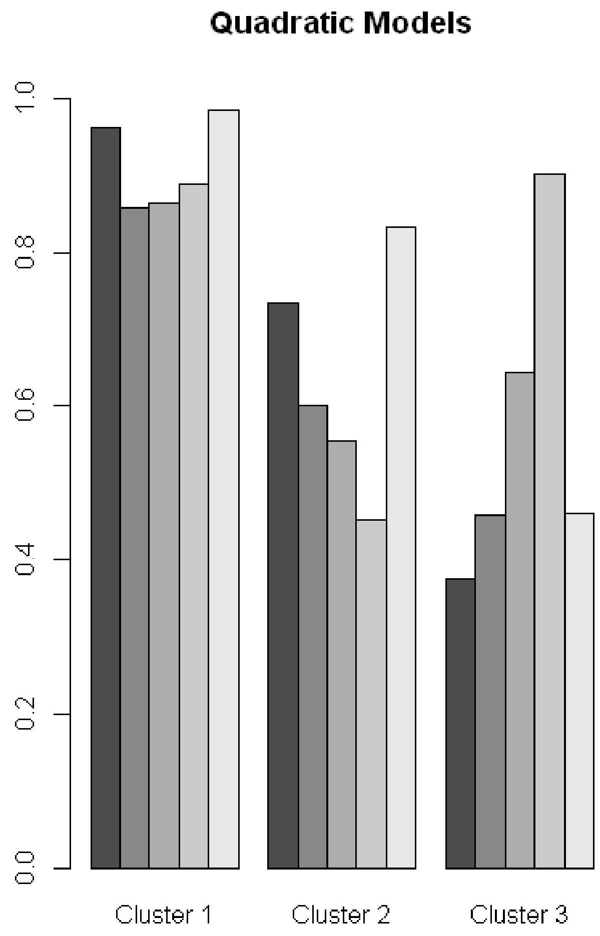




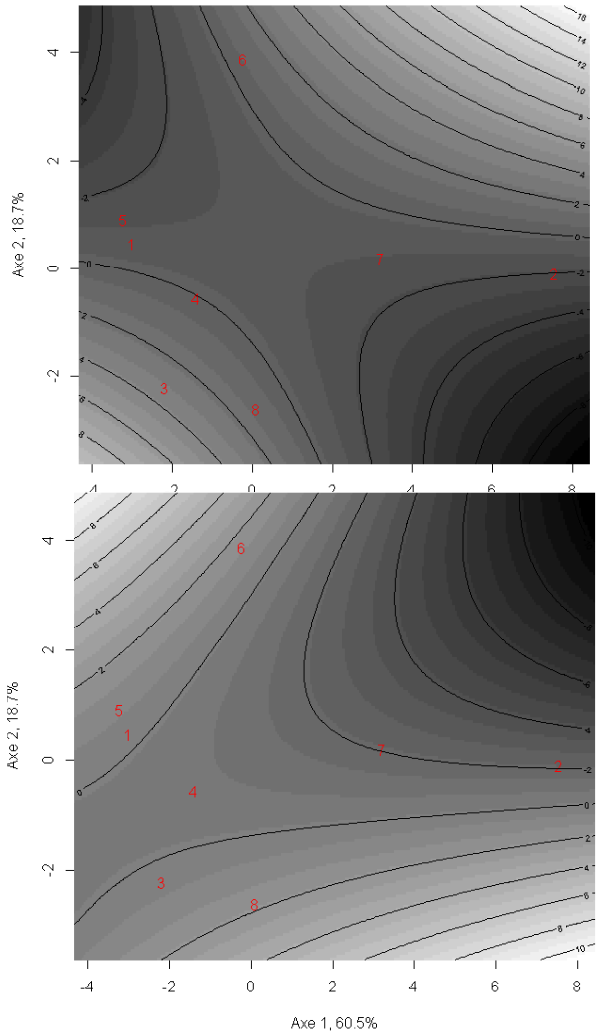
Application to Preference Mapping : The Coffee dataset

R_Y^2 explained for the three clusters





Application to Preference Mapping : The Coffee dataset – Clusters



Consumers like both smooth coffee with chocolate odors and strong coffee with intense odor

Consumers like both the green coffee with a lot of perfume and the bitter coffee with an intense aftertaste and an odor of roasted coffee

Plan

- Models of Quadratic PLS
- Extension of Hoskuldsson's approach to several Y : QPLS2
- α - Regularization of QPLS2
- Application to Preference Mapping : The Coffee dataset
- Conclusion

Conclusion

QPLS presents several advantages for preference mapping:

- the criterion is very explicit and in line with PLS regression
- the (new) algorithm is simple and convergent.

But :

- in order to enhance the interpretation of the model, there is a need to operate a model selection.

Bibliography

Baffi, G., Martin, E. B., & Morris, A. J. (1999). Non-linear projection to latent structures revisited: the quadratic PLS algorithm. *Computers and Chemical Engineering, 23*(3), 395-411.

Dijksterhuis, G., Martens, H., & Martens, M. (2005). Combined Procrustes analysis and PLSR for internal and external mapping of data from multiple sources. *Computational Statistics and Data Analysis, 48*(1), 47-62.

Endrizzi, I., Gasperi, F., Calo, D. G., & Vigneau, E. (2010) Two-step procedure for classifying consumers in a L-structured data context. *Food Quality and Preference, 21*(3), 270-277.

ESN (1996). A European sensory and consumer study. A case study on coffee. European Sensory Network, Chipping Campden, Gloucestershire.

Gnanadesikan, R. (1977). Methods for statistical data analysis of multivariate observations. Wiley, New York, 1977.

Höskuldsson, A. (1992). Quadratic PLS regression. *Journal of Chemometrics, 6*(6), 307-334.

Lengard, V. r., & Kermit, M. (2006). 3-Way and 3-block PLS regressions in consumer preference analysis. *Food Quality and Preference, 17*(3-4), 234-242.

Mage, I., Menichelli, E., & Naes, T. (2012). Preference mapping by PO-PLS: Separating common and unique information in several data blocks. *Food Quality and Preference, 24*(1), 8-16.

Verdun, S., Hanafi, M., Cariou, V. & Qannari, E. M. (2012), Quadratic PLS1 regression revisited. J. Chemometrics, 26: 384–389.

Wold, S., Kettaneh-Wold, N., & Skagerberg, B. (1989). Nonlinear PLS modeling. *Chemometrics and Intelligent Laboratory Systems,* 7(1-2), 53-65.

Wold, S., Albano, C., Dunn III, W.J., Edlund,U., Esbensen, K., Geladi, P., Helberg, S., Johansson, E., Lindberg, W. & Sjostrom, M. (1984). Multivariate data analysis in chemistry, in B.R. Kowalski (Eds), *Chemometrics, Mathematics and Statistics in Chemistry*, Reidel, Dordrecht, p. 17.

QPLS, What else ?

Thank you for your attention