Extension of *ComDim* for the analysis of (K+1) datasets; application in sensometrics

Angélina El Ghaziri[1], Véronique Cariou[1], Douglas N. Rutledge[2] & El Mostafa Qannari[1]

[1]LUNAM université, ONIRIS, Unité de Sensométrie et Chimiométrie, Nantes F-44322, France
INRA, Nantes, France
E-mail: angelina.el-ghaziri@oniris-nantes.fr; elmostafa.qannari@oniris-nantes.fr
[2] UMR Ingénierie Procédés Aliments, AgroParisTech, Inra, Université Paris-Saclay, 91300, Massy, France.

Abstract
ComDim is a multiblock data analysis, initially called Common Components and Specific Weights Analysis (*CCSWA*). Its objective is to determine common components of *K* datasets as well as specific weights for each dataset highlighting its contribution to the determination of these common components. Properties related to this method of analysis can be found in Hanafi & Qannari (2008). In this presentation we introduce an extension of *ComDim* to the analysis of (*K+1*) datasets. More precisely, we seek common components that highlight the relationships between a given dataset to *K* other datasets.
Keywords: *ComDim*, multi-block data analysis, *Multiblock PLS*.

Résumé en Français
ComDim est une méthode d’analyse de données multi-tableaux, initialement connue sous l’appellation *ACCPS* (Analyse en composantes communes et poids spécifiques). Elle vise à déterminer les composantes sous-jacentes aux *K* tableaux à analyser (composantes communes) et, associé à chaque tableau et chaque dimension, elle exhibe un poids qui reflète l’importance de la composante pour le tableau considéré. Nous nous intéressons ici à l’extension de *ComDim* à l’analyse de *K+1* tableaux. De manière précise, nous cherchons à explorer les relations entre un tableau Y et K tableaux $X_1, \dots, X_K$. Cette exploration est effectuée sur la base de composantes communes aux tableaux $X_1, \dots, X_K$ qui présentent un lien optimal avec Y. Des poids spécifiques aux différents tableaux permettent de mettre en évidence l’importance relative de chacun des tableaux dans la détermination de chacune des composantes communes. La méthode, appelée *P-ComDim*, est illustrée sur la base d’une étude de cas en évaluation sensorielle et étude des préférences. Les résultats de *P-ComDim* sont comparés à ceux de *PLS Multiblocs*.
Mots-clés: *ComDim*, tableaux multiples, *PLS multiblocs*.

1. Introduction

Common Component and Specific Weight Analysis was first introduced within the context of sensory analysis (Qannari *et al.*, 2000) and soon afterwards it was used in different contexts such as instrumental measurements (Nielsen et al., 2001), multivariate images (Courcoux *et al.*, 2002) … The acronym “*ComDim*” which stands for “Common Dimensions” was first used in Nielsen *et al.*, (2001) and in a *Matlab* program available under *SAISIR* environment (Cordella & Bertrand, 2014). Properties relating *ComDim* to other methods of analysis were shown (Hanafi *et al.*, 2006; 2010). Moreover, an extension of *ComDim*, called *AComDim*, for the assessment of the influence of a set of factors on multivariate data was introduced by Jouan-Rimbaud Bouveresse *et al.* (2011).

We propose a new extension of *ComDim* for the analysis of (*K+1*) datasets. In the presence of a response dataset, Y, and *K* datasets pertaining to different measurements on the same individuals, the objective is to seek common components that highlight the relationships between the *K* datasets and *Y* and determine specific weights that reflect the contributions of the various datasets to the determination of these components. Properties related to this new method will be shown and an illustration on the basis of a case study which involves sensory and preference data will be presented (Pham *et al.*, 2008).

2. From ComDim to P-ComDim

We recall that for a set of multiblock data X_1, X_2… X_K supposed to be centered, *ComDim* aims at determining common components q_1, q_2… q_n and saliences ($\lambda_r^{(k)}$) that reflect the importance of each block dataset in determining the underlying common components. These blocks and saliences are sought so as to approximate the association matrices $W_k = X_k X_k^T$ by $Q\Lambda_k Q^T = \sum_{r=1}^{n} \lambda_r^{(k)} q_r q_r^T$

This problem is solved in a sequential manner by seeking, at each stage, a common component and its associated saliences.

The strategy of analysis that we shall refer to as *P-ComDim* (*i.e.* Predictive *ComDim*) tackles the situation where we aim at predicting a dataset Y from K datasets X_1, X_2,…, X_k. It follows the same pattern of analysis as *ComDim* by replacing the matrices $W_k = X_k X_k^T$ by $T_k = X_k X_k^T Y Y^T$. Obviously, these latter matrices reflect the relationships of X_k (*k=1 …K*) with Y.

In the first stage of *P-ComDim*, we seek two components t and u, supposed to be of length 1, so as to minimize the quantity:

$$\sum_{k=1}^{K} ||T_k - \lambda^{(k)} t u^T||^2$$

In this expression, both components t and u are introduced to account for the fact that, unlike $W_k = X_k X_k^T$ the matrices $T_k = X_k X_k^T Y Y^T$ are not symmetric.

Four algorithms can be proposed to solve the previous minimization problem ranging from a non-iterative algorithm to a *NIPALS*-like algorithm.

It is worth noting that some of the saliences $\lambda^{(k)}$are likely to be negative to account for the orientation of the variables in X_k in comparison to t and u. Therefore, we propose to assess the contribution of X_k to the determination of t and u by $\left(\lambda^{(k)}\right)^2$.

For the determination of common components of higher order, we propose to deflate the datasets X_k (*k=1… K*) and Y with respect to the latent variables, t, determined at the previous stages.

The strategy of analysis is illustrated using datasets from a consumer study. The aim is to explain the preference scores (dataset Y) by three datasets: X_1 (flavor), X_2 (aroma), X_3 (texture). These data are taken from Pham *et al.* (2008).

The outcomes of *P-ComDim* analysis are compared to those of *Multiblock-PLS* (Wangen *et al.*, 1989; Westerhius *et al.*, 1998)

References

Cordella, C. B. Y. & Bertrand, D. (2014). SAISIR: A new general chemometric toolbox. *Trends in Analytical Chemistry*, 54, 75-82.

Courcoux, P., Devaux, M.-F., & Bouchet, B. (2002). Simultaneaous decomposition of multivariate images using three-way data analysis: application to the comparison of cereal grains by confocal laser scanning microscopy. *Chemometrics and Intelligent Laboratory Systems*, 62, 103-113.

Hanafi, M., Kohler, A., & Qannari, E.M. (2010). Shedding new light on Hierarchical Principal Component Analysis. *Journal of Chemometrics,* 24(11-12): 703-709.

Hanafi, M., Mazerolles, G., Dufour, E., & Qannari, E.M. (2006). Common components and specific weight analysis and multiple co-inertia analysis applied to the coupling of several measurement techniques. *Journal of Chemometrics,* 20(5): 172-183.

Hanafi, M. & Qannari, E.M. (2008). Nouvelles propriétés de l'analyse en composantes communes et poids spécifiques. *Journal de la société française de statistique & revue de statistique appliquée*, 149,75-97

Jouan-Rimbaud Bouveresse, D., Climaco Pinto, R., Schmidtke, L. M., Locquet, N., & Rutledge, D. N. (2011). Identification of significant factors by an extension of ANOVA-PCA based on multi-block analysis. *Chemometrics and Intelligent Laboratory Systems,* 106(2): 173-182.

Nielsen, J.P., Bertrand, D., Micklander, E., Courcoux, P., & Muck, L.(2001). Study of NIR spectra, particle size distributions and chemical parameters of wheat flours: a multi-way approach. *Journal of Near Infrared Spectroscopy,* 9, 275-285.

Pham, A. J., Schilling, M. W., Mikel, W.B., Williams, J.B., Martin, J.M., & Coggins, P.C. (2008). Relationships between sensory descriptors, consumer acceptability and volatile flavor compounds of American dry-cured ham. *Meat Science*, 80(3): 728-737.

Qannari, E.M., Wakeling, I., Courcoux, P., & MacFie, H.J.H. (2000). Defining the underlying sensory dimensions. *Food Quality and Preference*, 12, 365-368.

Wangen, L.E., & Kowalski, B.R. (1989). A multiblock partial least squares algorithm for investigating complex chemical systems. Journal of chemometrics, 3, 3-20.

Westerhius, J.A., Kourti, T., & MacGregor, J.F. (1998). Analysis of multiblock hierarchical PCA and PLS models. Journal of Chemometrics, 12, 301-321.