New Formulation for Multi-block-Partial Least Squares-Discriminant Analysis Application to metabolomics data

Véronique Cariou[1], El Mostafa Qannari[1], Mohamed Soumah[1], Marie-Cécile Alexandre-Gouabau[2] & Thomas Moyon[2]

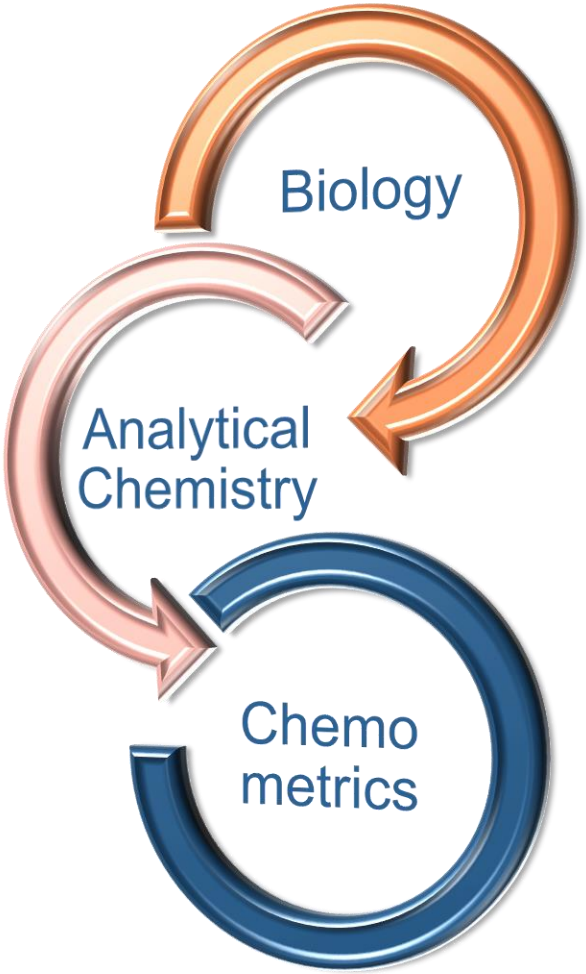
veronique.cariou@oniris-nantes.fr

[1] *Unité de Sensométrie-Chimiométrie (USC), Oniris, Nantes, France*

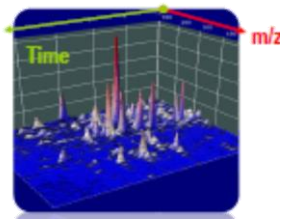
[2] *UMR PhAN, INRA, Nantes, France*

14th Symposium AgroStat, March 2016
Lausanne Suisse

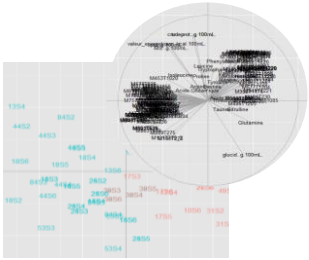
Metabolomic approach



In many life sciences applications, there is a growing interest in applying metabolomic techniques

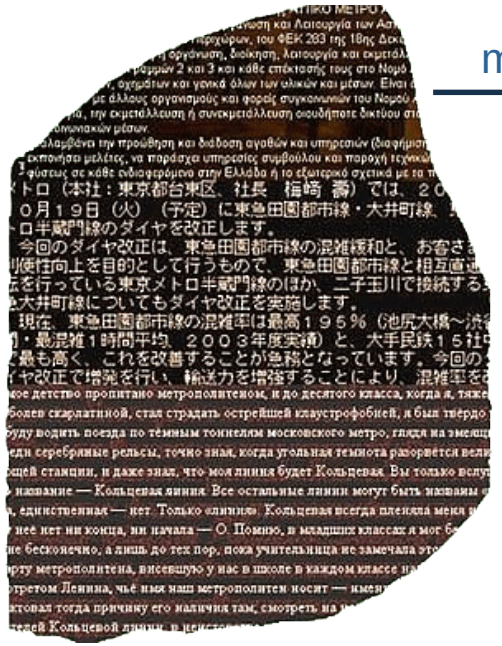


Metabolomics refers to an analytical description of biological samples by defining « metabolic » fingerprintings



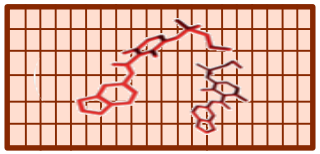
Multivariate statistical analyses are applied to characterize compounds and to differenciate samples

Metabolomics and multiblock data

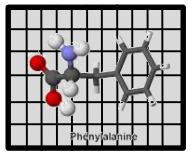


Measurements

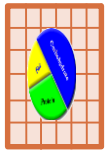
metabolomic



free amino acids



macronutrients



Multiblock data

Model

Coupling data to provide:

- ✓ a more interpretable knowledge
- ✓ a better picture of the biological system

Overview

Recall of PLS-DA

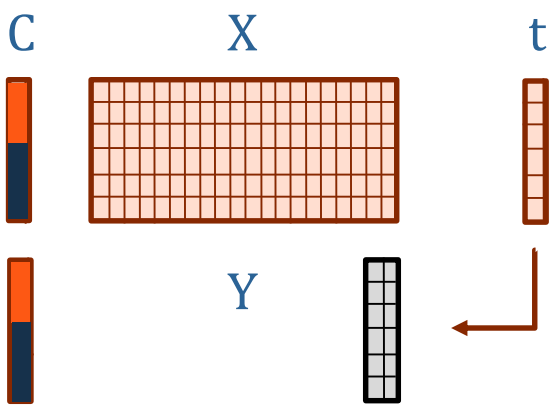
PLS-DA within the multiblock framework

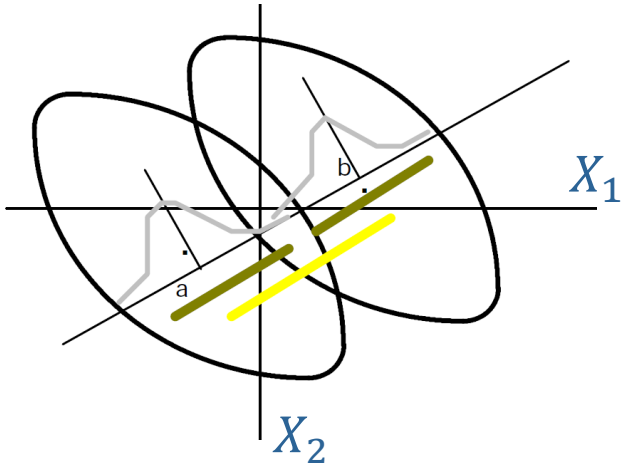
Relating the composition of breast milk with the growth of premature babies

Conclusion

Recall of PLS -DA

Determine a latent component t which separates the best the groups





PLS-DA criterion (Nocairi & al.,2005): $\max var(P_y t)$

with $\mathrm{t} = \mathrm{Xw}$ and P_y denotes the projector associated to the Y binary matrix

Component obtained by diagonalization of the between groups covariance matrix

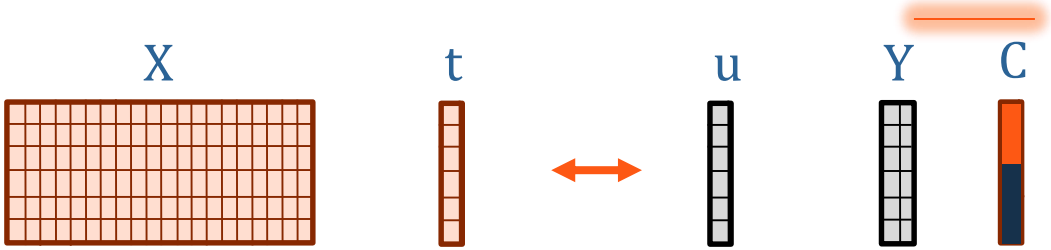
Nocairi, H., Qannari, E. M., Vigneau, E., & Bertrand, D. (2005). Discrimination on latent components with respect to patterns. Application to multicollinear data. *Computational statistics & data analysis*, *48*(1), 139-147.

Equivalence between PLS –DA and PLS2

Formulation of the PLS-DA criterion: $\max var(P_y t)$

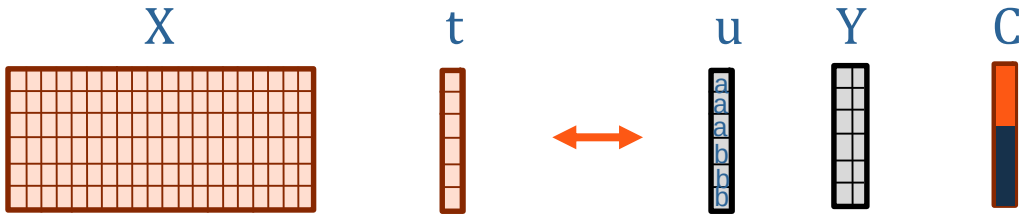
Recall of PLS2 criterion:

$$\max cov^2(t,u) = cor^2(t,u) \times var(t) \times var(u)$$



In the case of dichotomous variables, $var(u)$ should be neglected

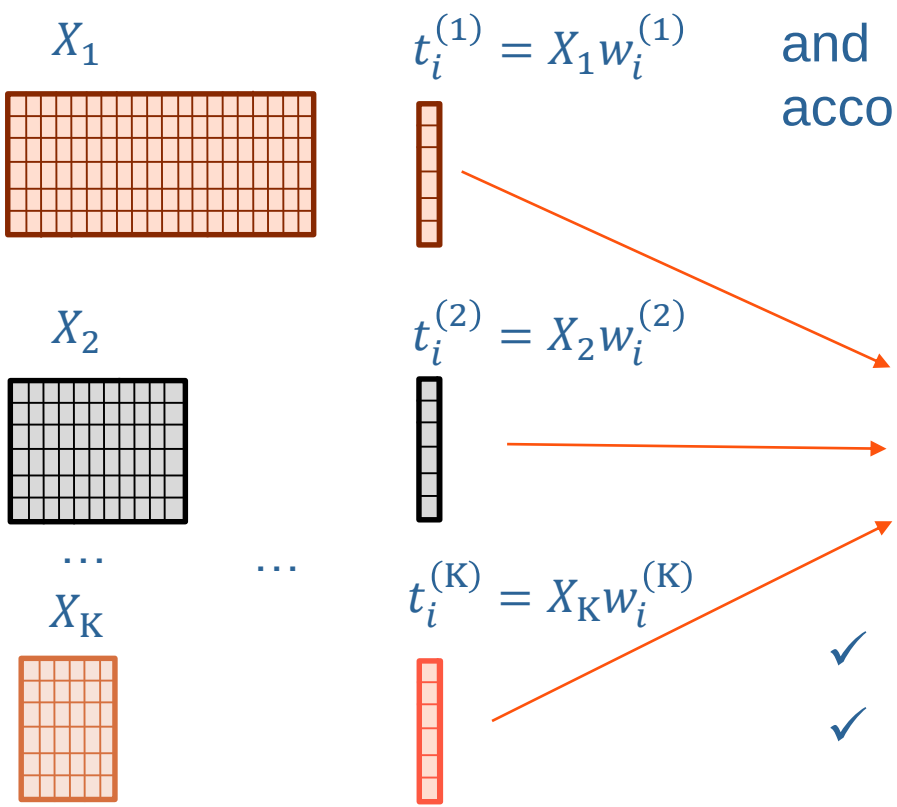
This leads to consider $Y^* = Y(Y^T Y)^{-1/2}$ instead of Y to account for the frequencies of the categories



Barker, M., & Rayens, W. (2003). Partial least squares for discrimination. *Journal of chemometrics*, *17*(3), 166-173.

PLS-DA within the multiblock framework

Extract simultaneously block components and global components which best account for the between groups variability



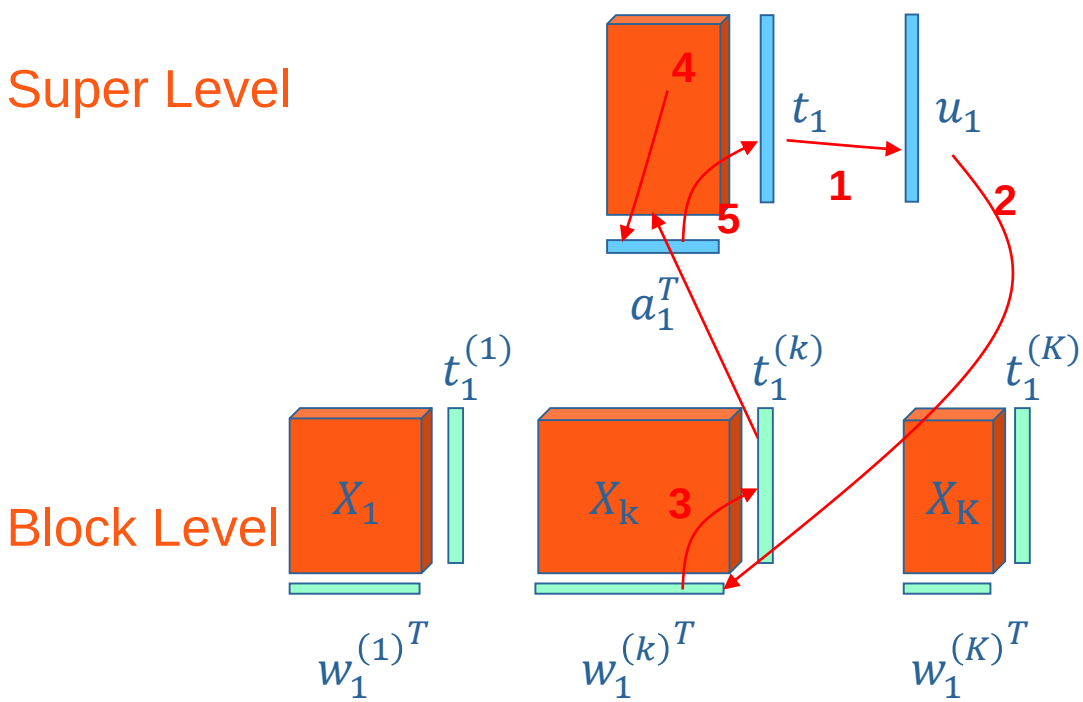
$$t_i = \sum_{k=1}^{K} a_{ik} t_i^{(k)}$$

- ✓ t_i: i^{th} global component
- ✓ $t_i^{(k)} = X_k w_i^{(k)}$: i^{th} block component associated to X_k

PLS-DA within the multiblock framework

$$\max \sum_{k=1:K} cov^2\left(t_1^{(k)}, u_1\right) \text{ with } t_1^{(k)} = X_k w_1^{(k)} \quad \left\|w_1^{(k)}\right\| = 1 \quad k = 1:K$$

$$\max var\left(P_y t_1\right)$$



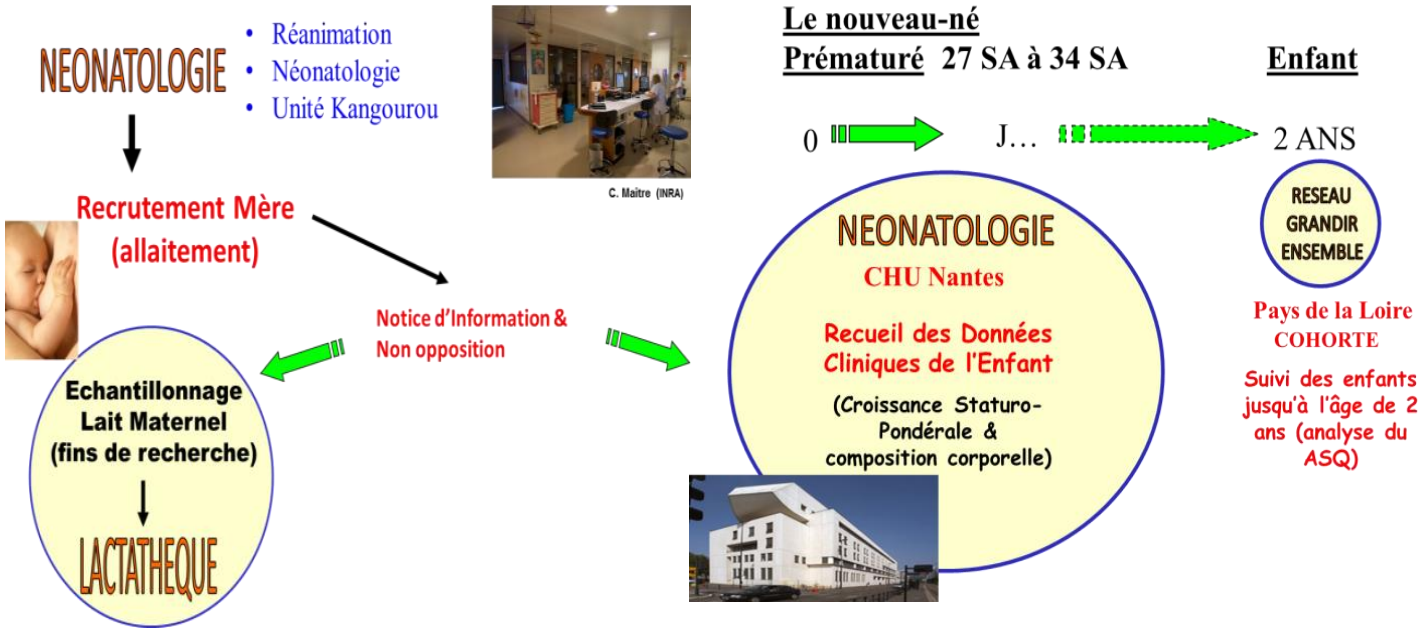
1. $u_1 = P_y t_1$
2. $w_1^{(k)} = \frac{X_k^T P_Y t_1}{\|X_k^T P_Y t_1\|}$
3. $t_1^{(k)} = X_1 w_1^{(k)}$
4. $a_{1k} = cov\left(t_1^{(k)}, u_1\right)$
   $\|a_1\| = 1$
5. $t_1 = \sum_{k=1}^{K} a_{1k} t_1^{(k)}$

subsequent components determined in the same way after a deflation step

Relating the composition of breast milk with the growth of premature babies

LACTACOL project French Clinical Trial N° NCT01493063

A better understanding of the impact of perinatal nutrition on the growth of premature babies

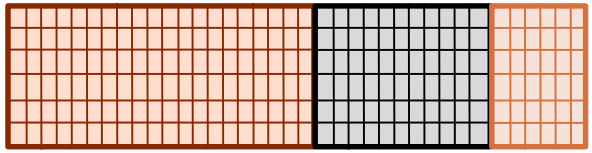


Relating metabolomics, free amino-acids and nutritional blocks with the growth

- 18 pre-term newborns (28 - 34 weeks of amenorrhea) classified into two categories of postnatal growth:
  - ✓ expected growth (n_1=9 newborns)
  - ✓ suboptimal growth (n_2=9 newborns)

- Representative breast milk sample of the last 24 hours, each week

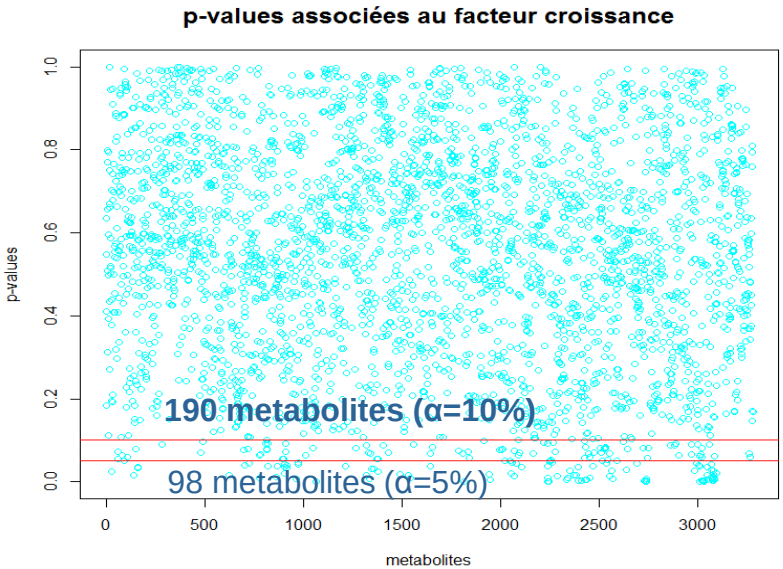
- Lipidomic fraction, extracted using Bligh-Dyer method, analyzed through metabolomics analytical tools (namely LCHR-MS)

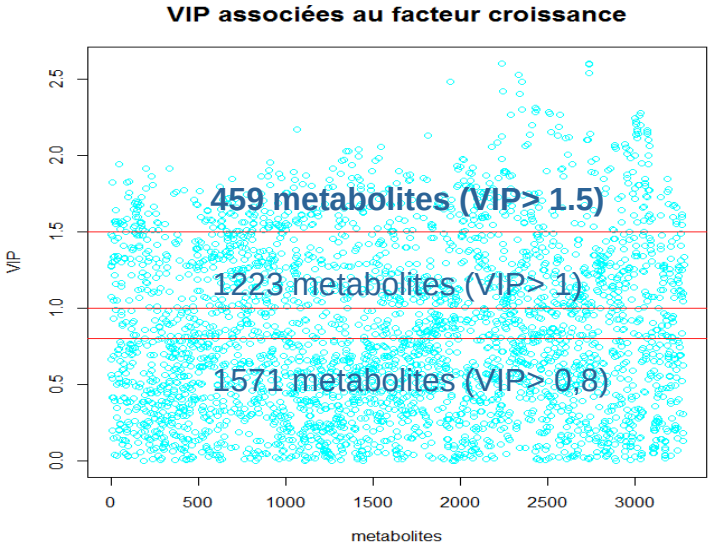


3280 metabolites | 21 free amino acids | 6 macronutrients

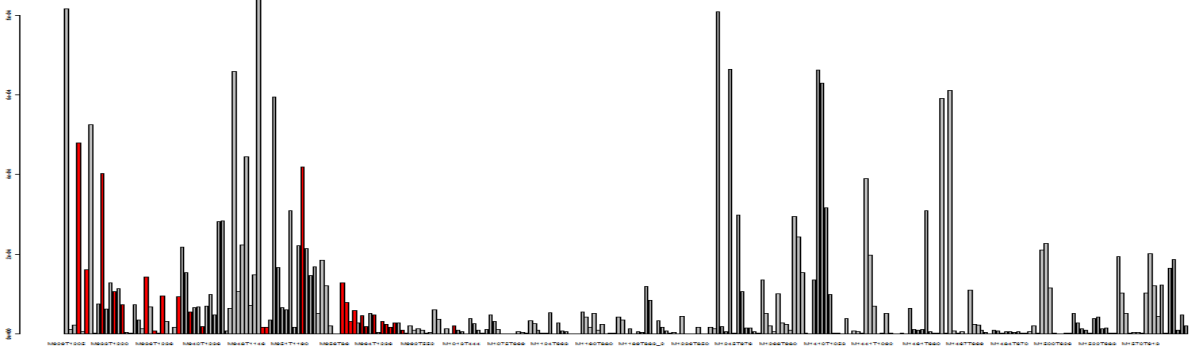
Variable selection on the metabolomics block

Use of a mixed model and PLS-DA with the VIP criterion

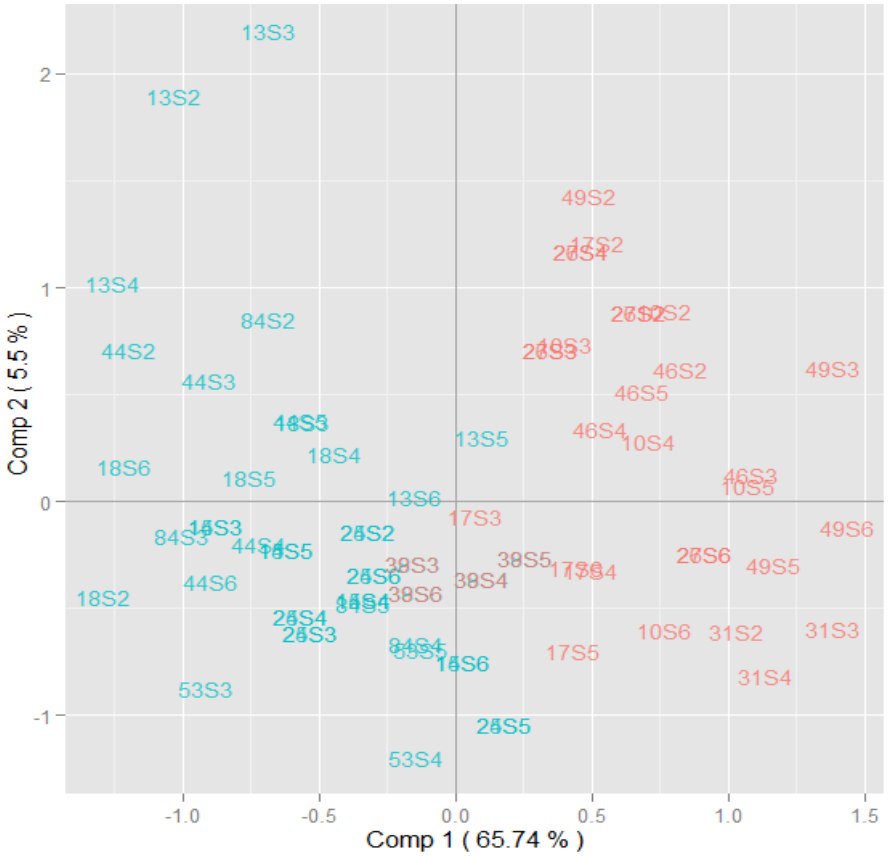


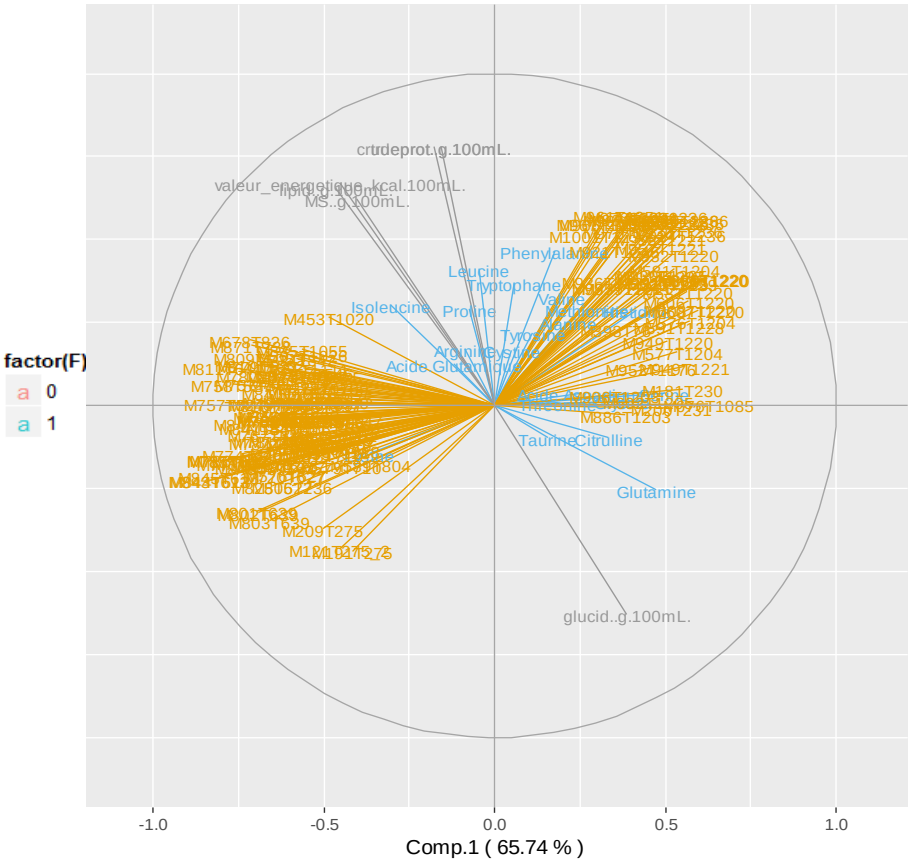


141 metabolites common to the two selection procedures



MB PLS-DA applied on metabolomics, free amino acids and nutritional blocks





MB PLS-DA applied on metabolomics, free amino acids and nutritional blocks

Block weights a

	Comp 1	Comp 2
Metabo	0.74	0.23
Acides A	0.14	0.22
MIRIS	0.12	0.55

Global components correlation ratio

	Comp 1	Comp 2
RG	0.66	0.06

Block components correlation ratio

	Comp 1	Comp 2
Metabo	0.62	0.06
Acides A	0.21	0.04
MIRIS	0.08	0.03

Best separation with the metabolomic block

MB PLS-DA provides:

- ✓ correlation ratio both at the global and block levels
- ✓ Factorial representations associated to each block

Conclusion

MB PLS-DA:

- a synthesis of PLS-DA performed separately on each data block
- determination of weights associated to each block reflecting its contribution to the separation of the groups

Biological interpretation:

- Among the 141 metabolites selected, 26 have been annotated. They correspond to different categories of lipids based on their chemical structure

More work is needed regarding the cross-validation step

References

Acar, E., Rasmussen, M.A., Savorani, F., Næs, T. & Bro, R. (2013). Understanding data fusion within the framework of coupled matrix and tensor factorizations. *Chemometrics and Intelligent Laboratory Systems*, 129, 53-63

Barker, M., & Rayens, W. (2003). Partial least squares for discrimination. *Journal of chemometrics*, *17*(3), 166-173.

Biancolillo, A., Måge, I., & Næs, T. (2015). Combining SO-PLS and linear discriminant analysis for multi-block classification. *Chemometrics and Intelligent Laboratory Systems*, *141*, 58-67.

Boccard, J., & Rutledge, D. N. (2013). A consensus orthogonal partial least squares discriminant analysis (OPLS-DA) strategy for multiblock Omics data fusion. *Analytica chimica acta*, *769*, 30-39.

Boccard, J., Badoud, F., Grata, E., Ouertani, S., Hanafi, M., Mazerolles, G., Lantéri, P., Veuthey, J-L., Saugy, M. & Rudaz, S. (2011). A steroidomic approach for biomarkers discovery in doping control. *Forensic Science International*, *213*(1–3), 85-94

Nocairi, H., Qannari, E. M., Vigneau, E., & Bertrand, D. (2005). Discrimination on latent components with respect to patterns. Application to multicollinear data. *Computational statistics & data analysis*, *48*(1), 139-147.

Thank you for your attention