



➤ Apport de la SO-PLS à l'approche multiblocs exploratoire

Marion Brandolini-Bunlon¹ & Benoît JAILLAIS²

¹ Université Clermont Auvergne, INRAE, UNH, PFEM, MetaboHUB Clermont, Clermont-Ferrand, France.

² INRAE, TRANSFORM, Nantes, France

> Plan

1990's => Fac de chimie organique



Bon rendement / produit pur



Caractérisation **structurale** :
RMN, **MIR**, **UV-Visible**

Prémices de la chimiométrie



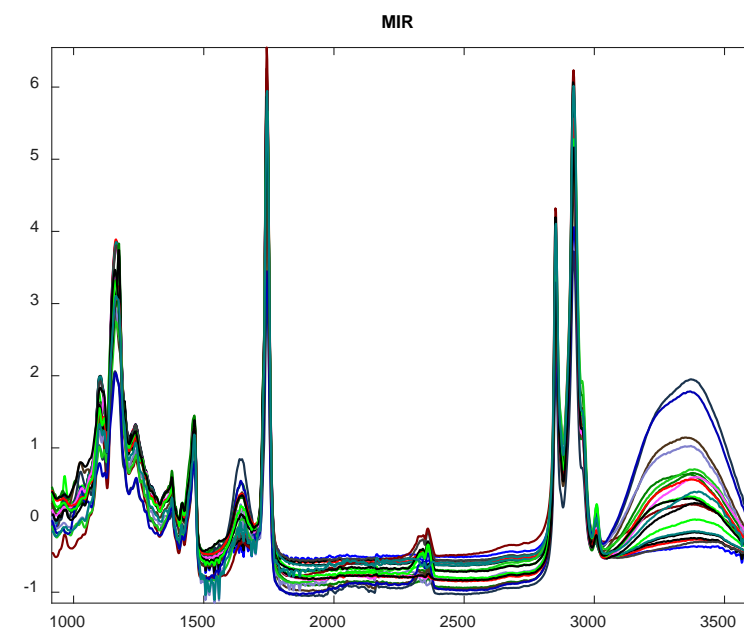
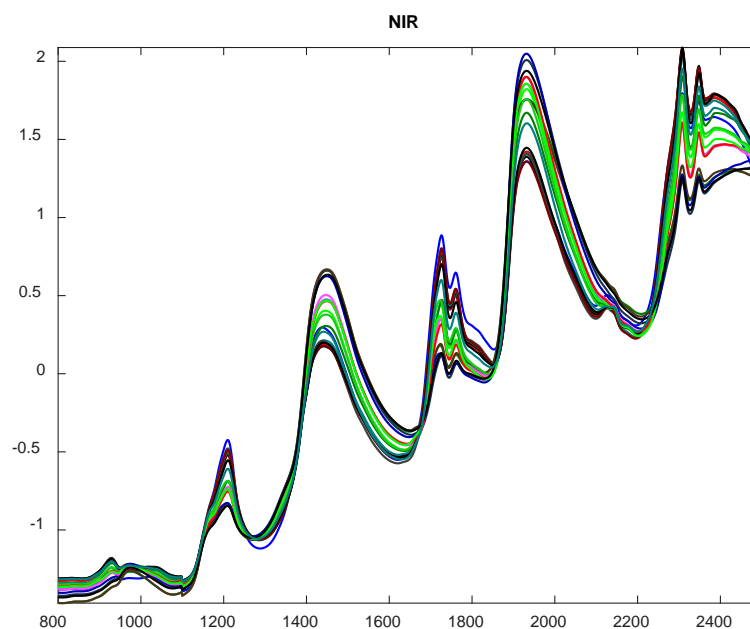
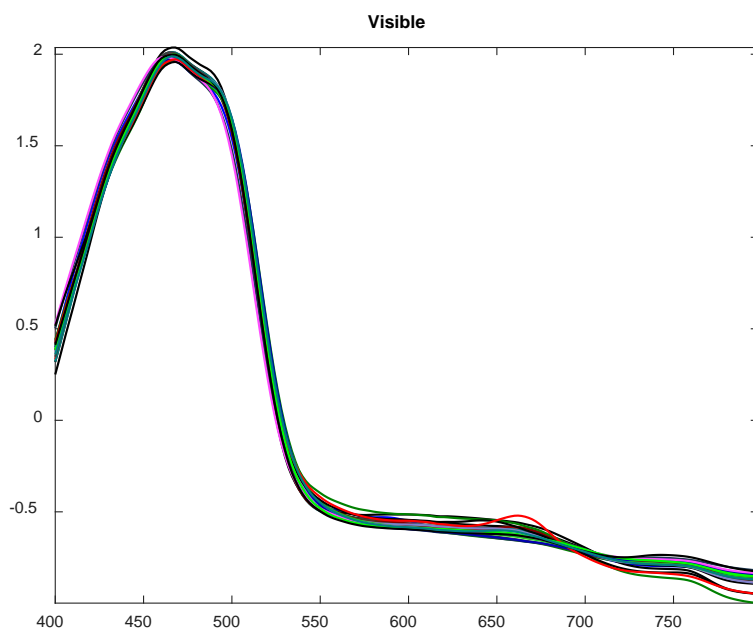
Quantification : **NIR**

Maintenant : PLSR, PLSDA, Multiblocs exploratoire, prédictif...

Utilité ? Comparaison de méthodes..

➤ Présentation des données

8 beurres (**80-82%** de MG)+ 13 margarines (**34-80%** MG)
Prétraitement SNV



> Objectifs

- Explorer la variabilité contenue dans chaque bloc => ACP
- Distinguer la variabilité spécifique à chaque bloc de celle commune à tous les blocs => MB-PCA
- Décomposer de façon supervisée la variabilité (commune vs spécifique) contenue dans les blocs => SO-PLS
- => / teneur en matière grasse
- => / type de matière grasse : animale vs végétale

ACP par blocs

centrage



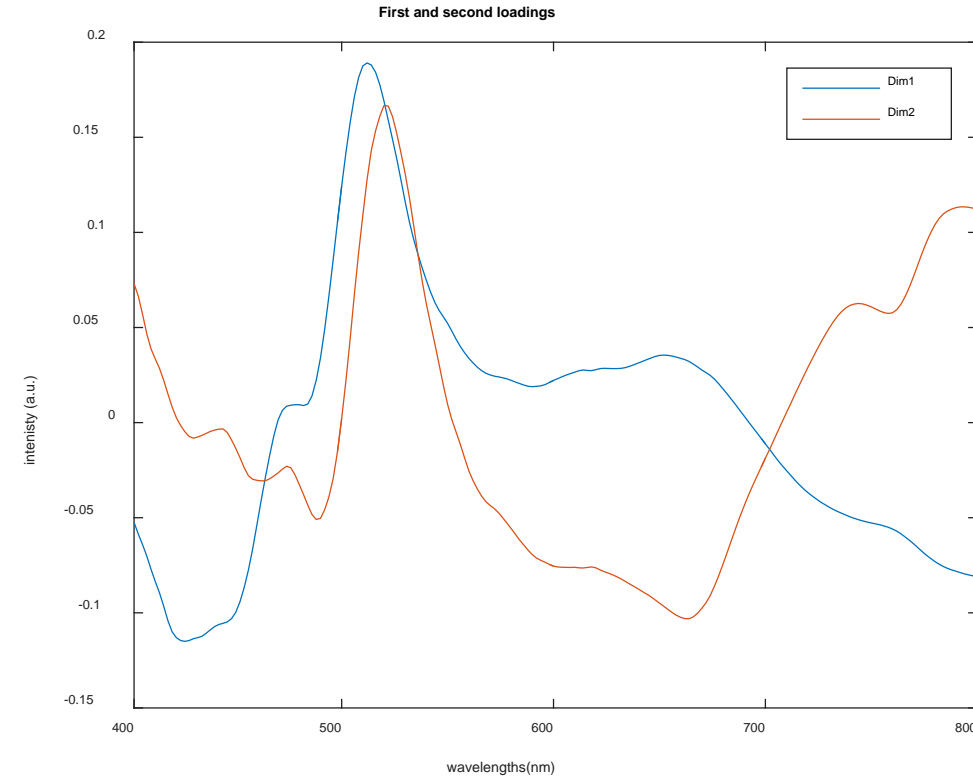
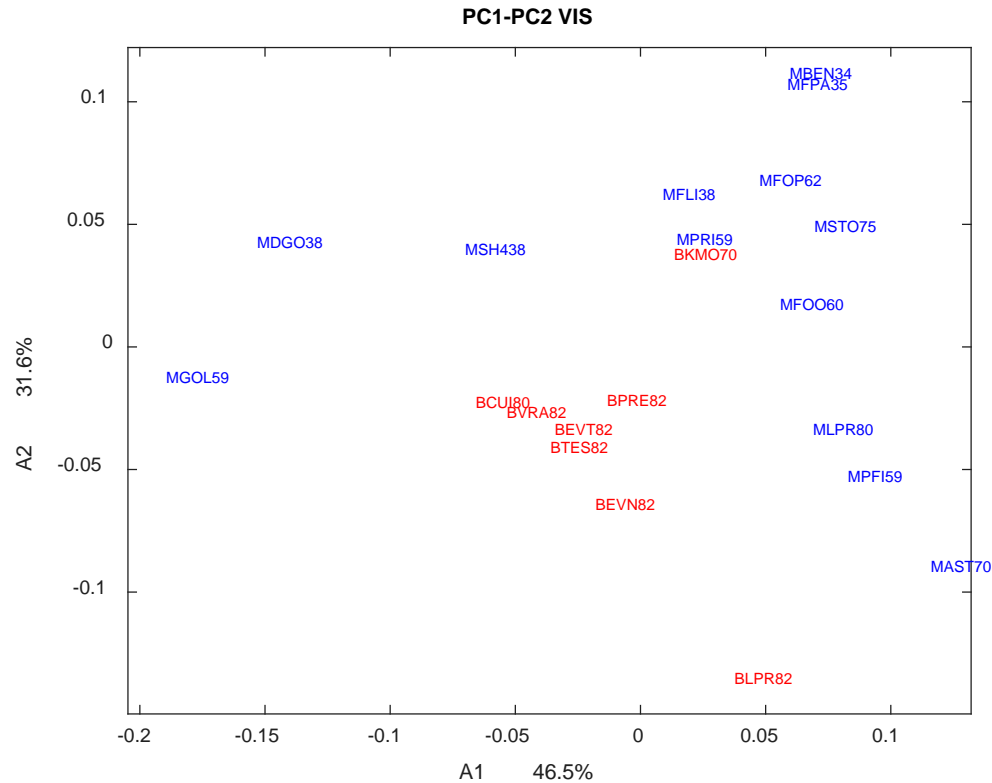
RÉPUBLIQUE
FRANÇAISE

*Liberté
Égalité
Fraternité*

INRAE

➤ ACP bloc visible

AG avec doubles liaisons conjuguées

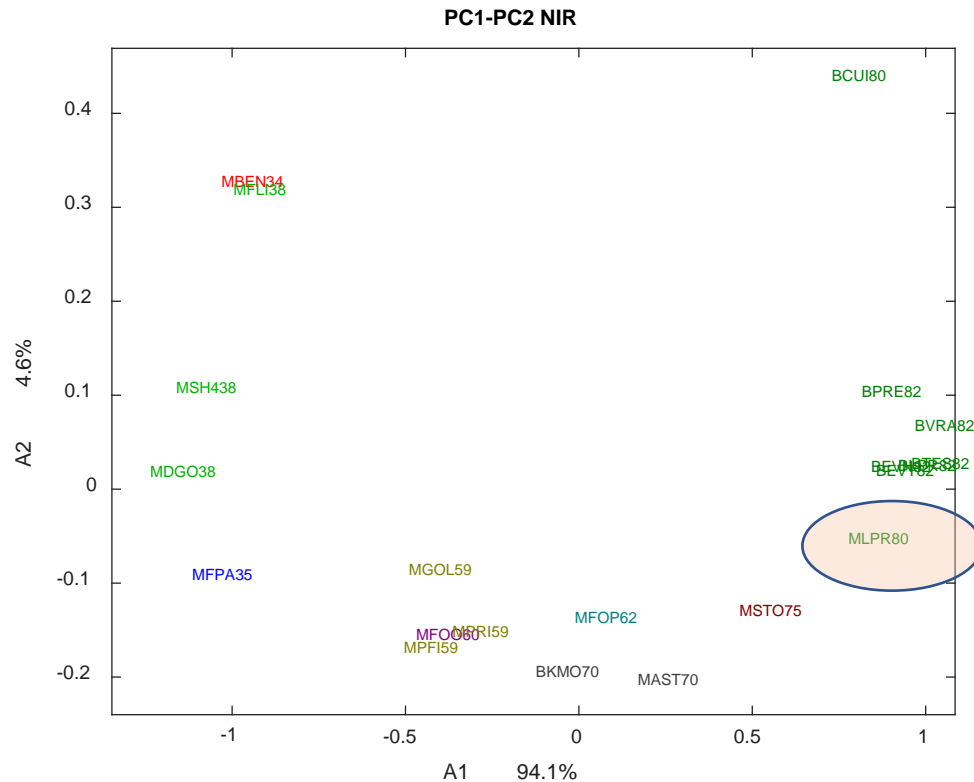


Séparation « légère » des échantillons en fonction de la catégorie beurre/margarine

Loadings liés à la variation de couleur dans le jaune sur PC1 beta-carotène (B)??

Loadings liés à la variation de couleur dans le jaune sur PC2 : additif colorant jaune (M)

➤ ACP bloc NIR



Teneur en eau

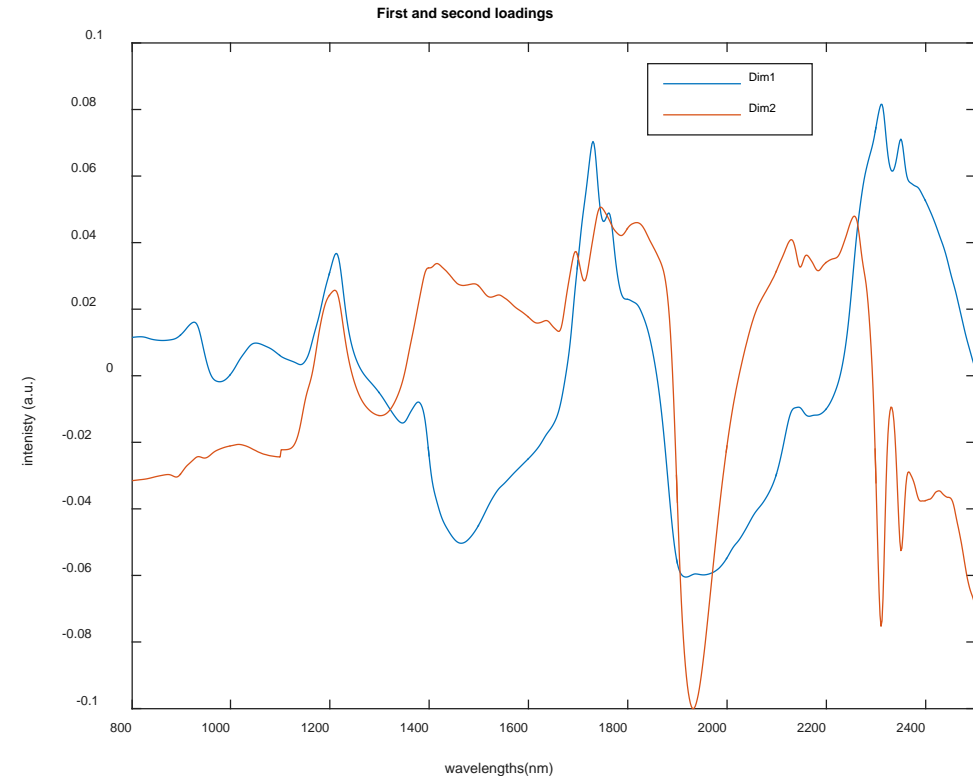
Teneur en matière grasse (MG)

Séparation des échantillons en fonction de la teneur en MG/eau

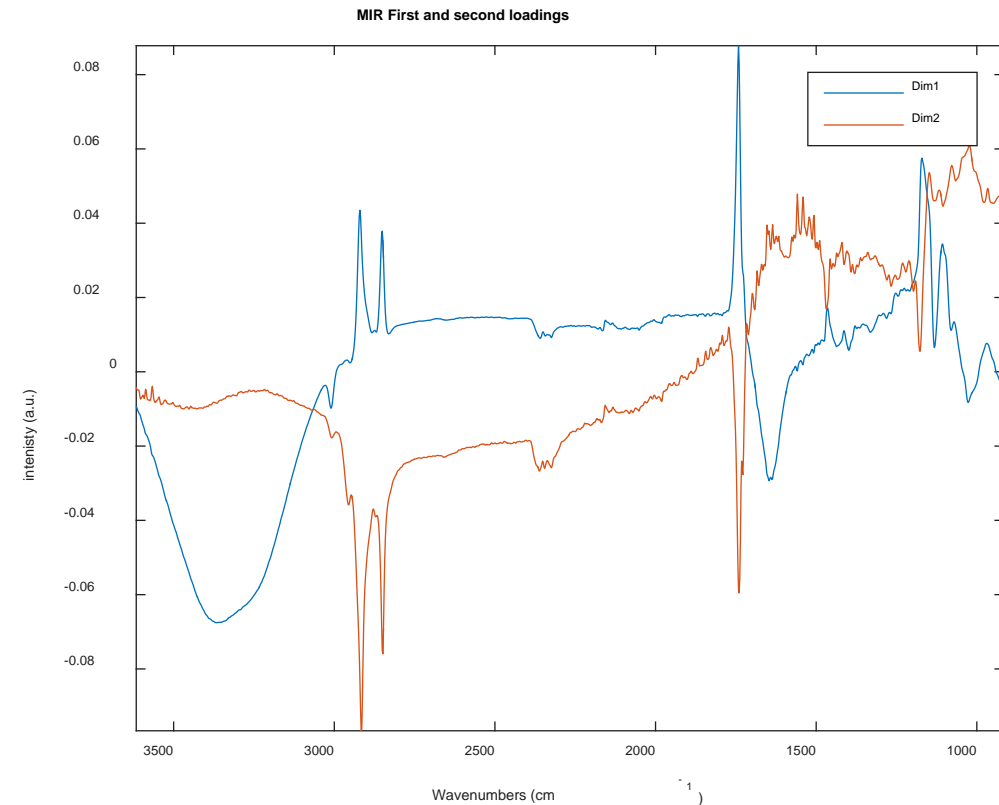
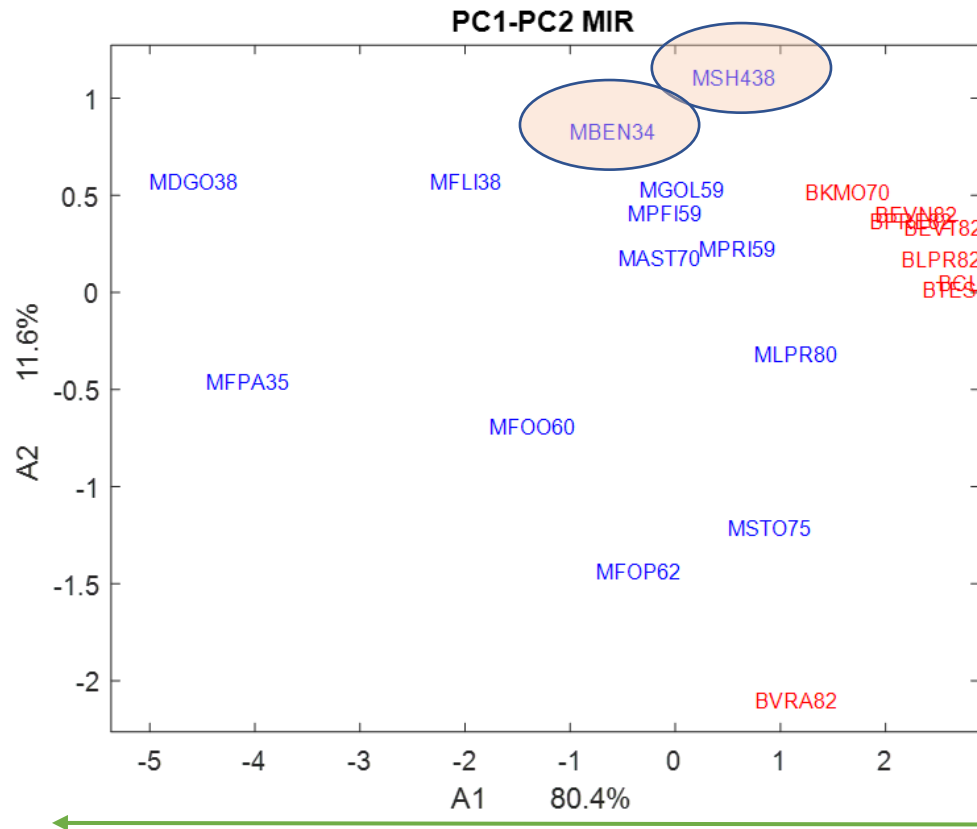
Loadings liés à la variation des bandes liées à l'eau sur PC1 : 1440, 1930 (nég)

Loadings liés à la variation des bandes liées à la matière grasse sur PC1 : 1210, 1730, 1740, 2310, 2350 (pos)

Loadings liés à la variation de l'interaction eau-protéine PC2 : 1920 ??



➤ ACP bloc MIR



Séparation « franche » des échantillons en fonction de la catégorie Beurre/margarine

Loadings liés à la variation des pics liés à l'eau sur PC1 : 3370 (nég) [et 3010 pour cis]

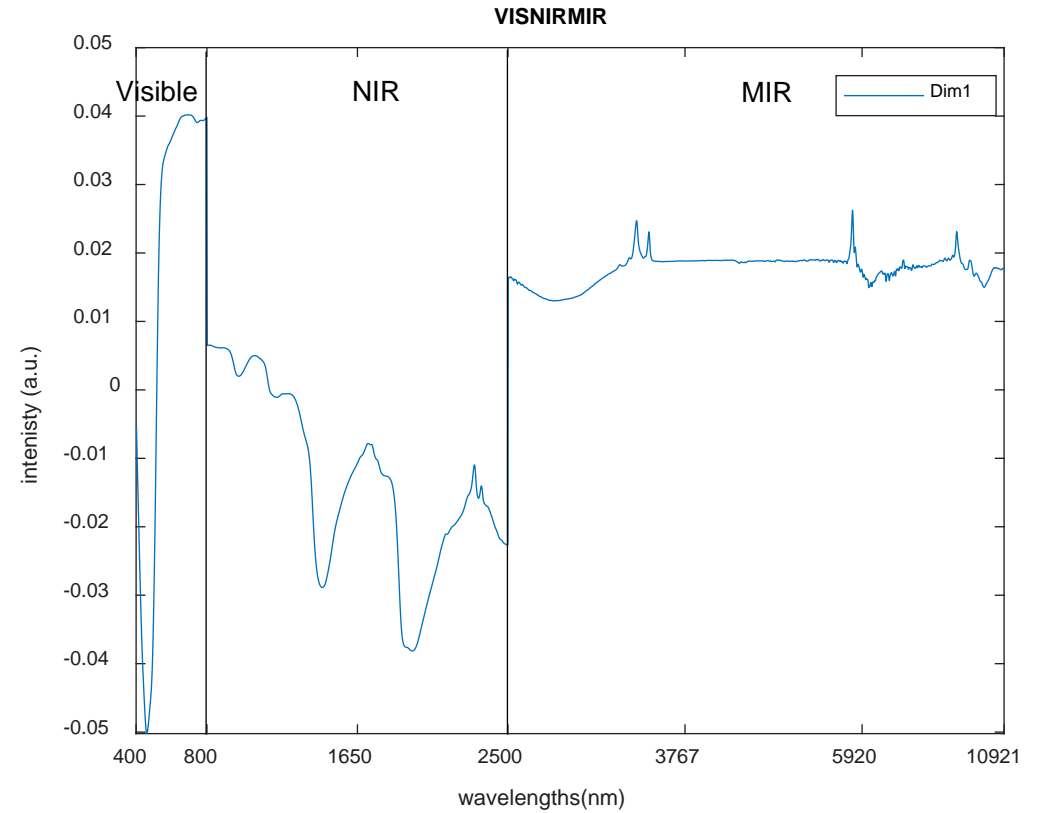
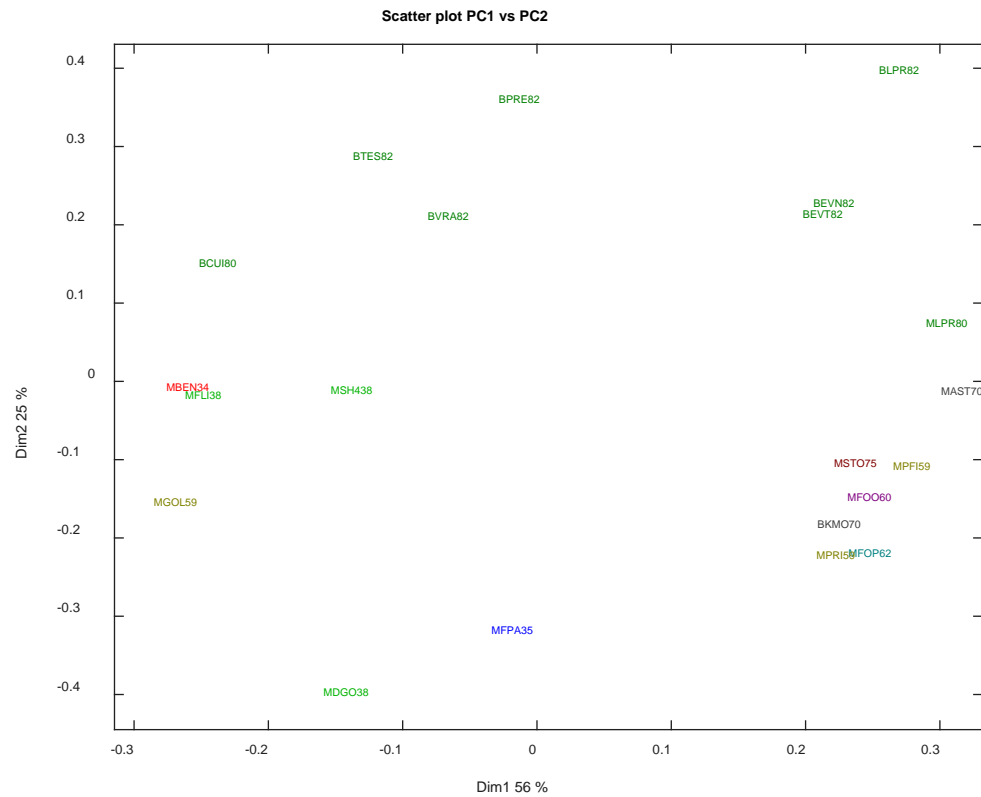
Loadings liés à la variation des pics liés à la présence de double liaisons « trans » sur PC1 : 2851, 2920 (pos)

Loadings liés à la variation des pics liés à la présence de double liaisons « cis » sur PC2 : 2853, 2916 (neg)

ACP multi-blocs (MB-PCA)

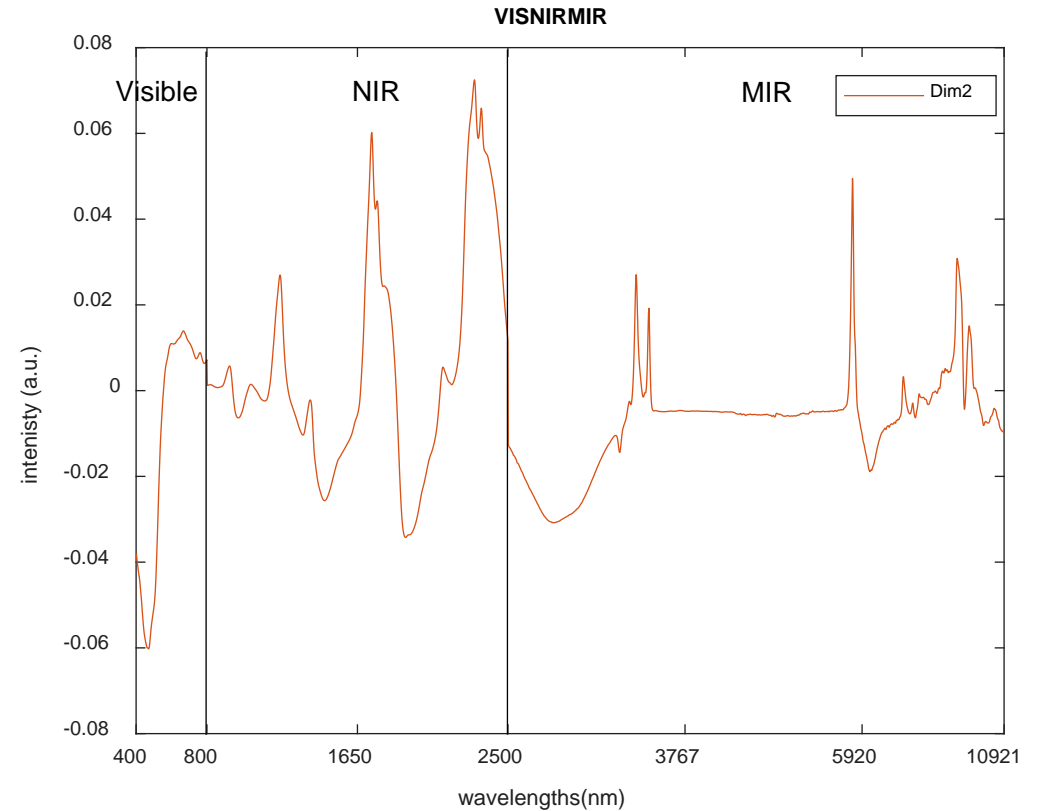
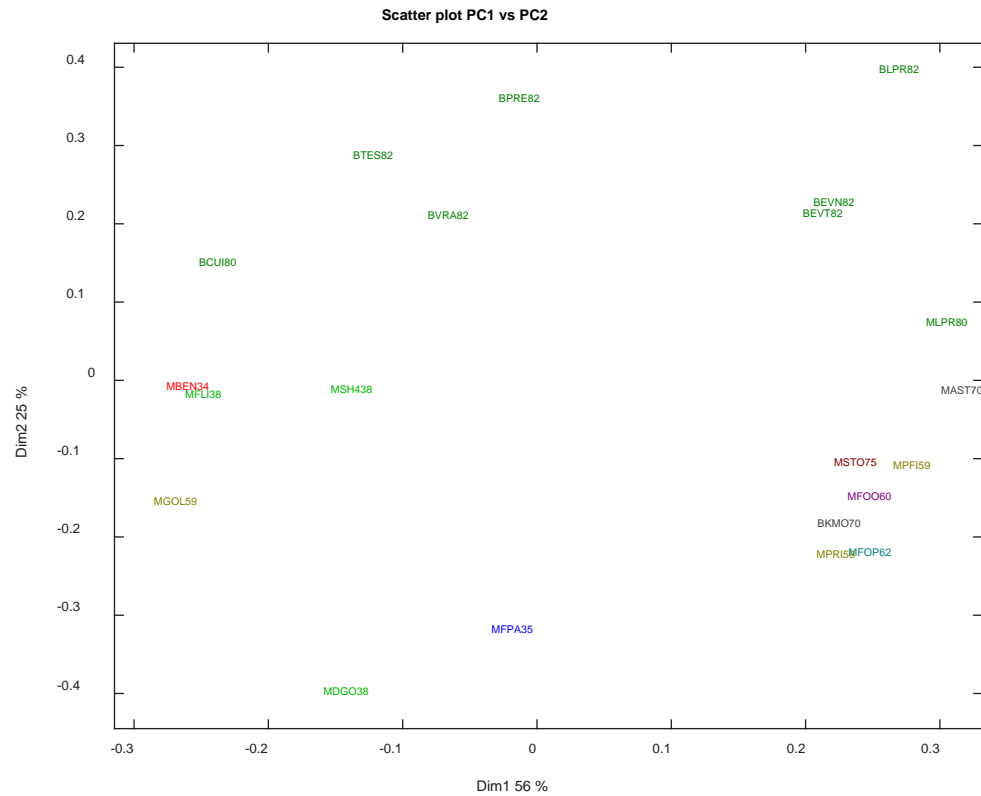
centrage

➤ Scores MB-PCA



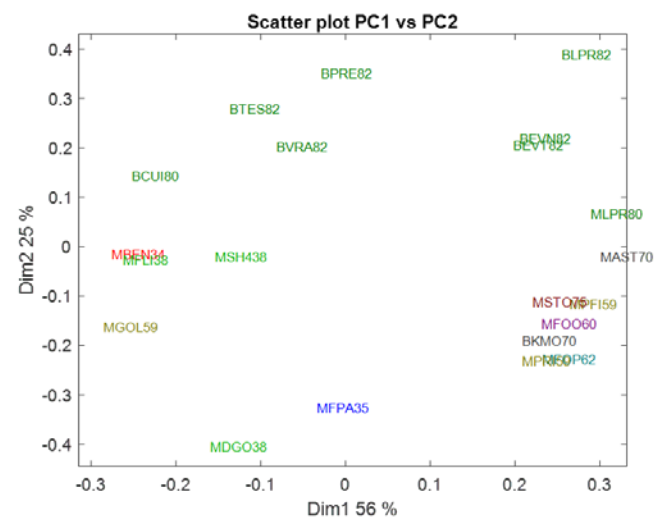
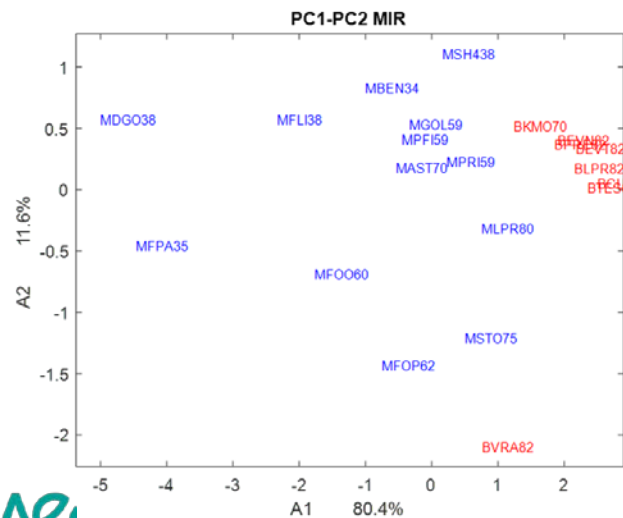
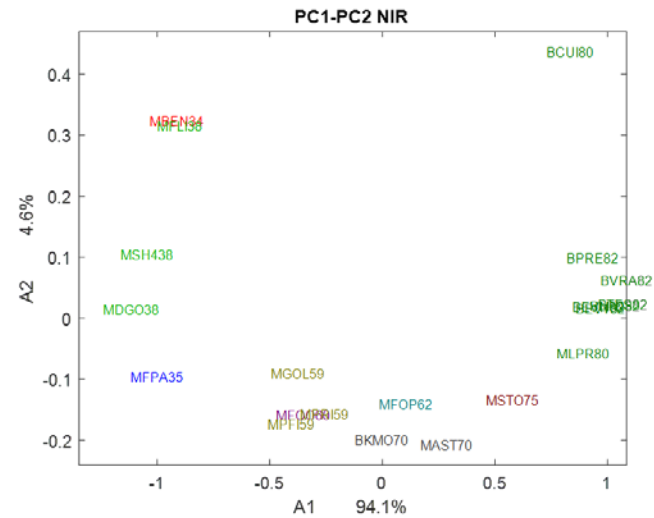
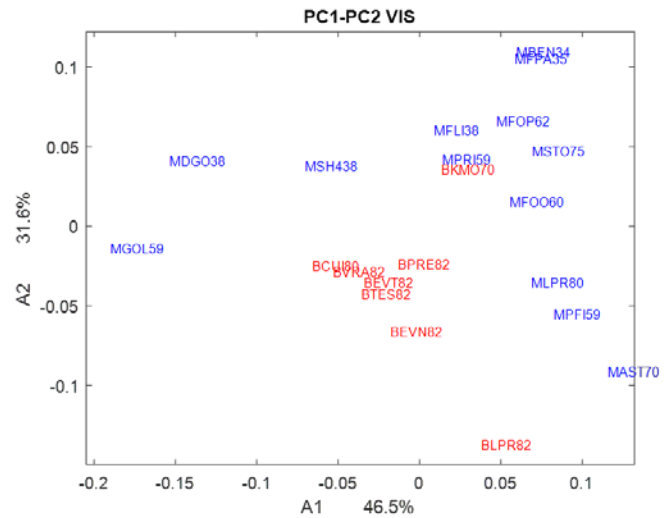
Axe 1 : Séparation pas claire : combinaison des bandes liées à l'eau (NIR) et à doubles liaisons cis (MIR)?

➤ Scores MB-PCA



Axe 2 : Bonne séparation des échantillons en fonction de la catégorie Beurre/margarine et teneur MG :
opposition lipides≠eau en NIR + doubles liaisons « trans » MIR

➤ Scores mono- vs multi- blocs



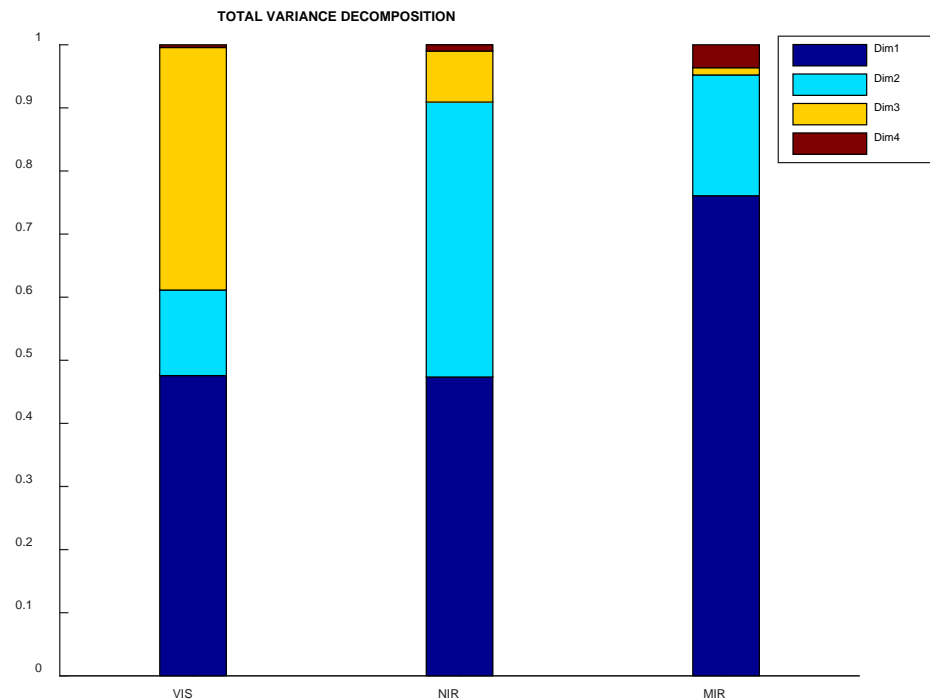
Ajustement de la direction des composantes :

-> Dim 1 : mélange des 3 domaines spectraux

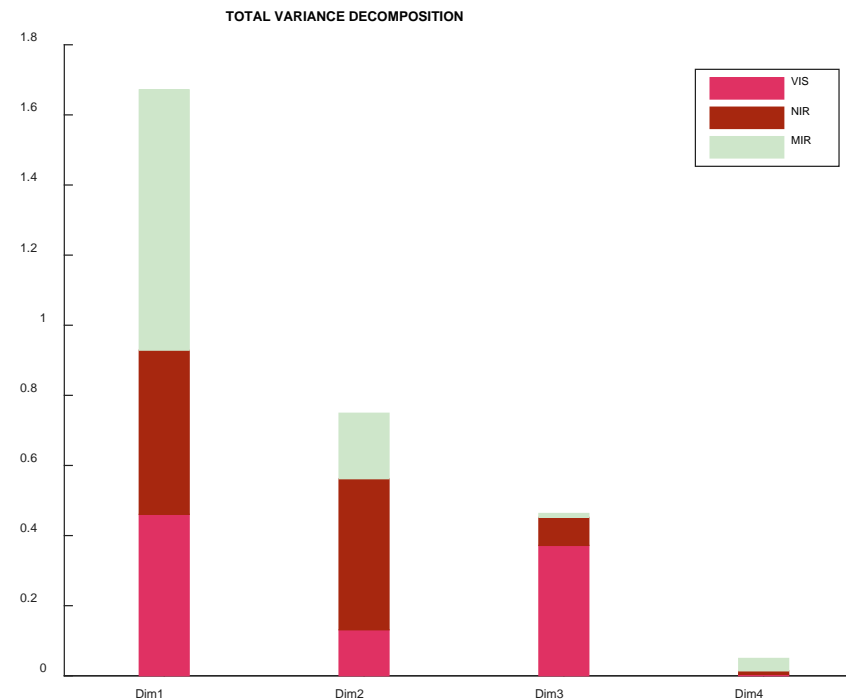
-> Dim 2 : Perte de précision des composantes communes (% MG) par rapport au composantes spécifiques

➤ Décomposition de la variabilité

Par bloc



Par dimension



SO-PLS vs MG

Données SNV

Package Matlab : Federico Marini

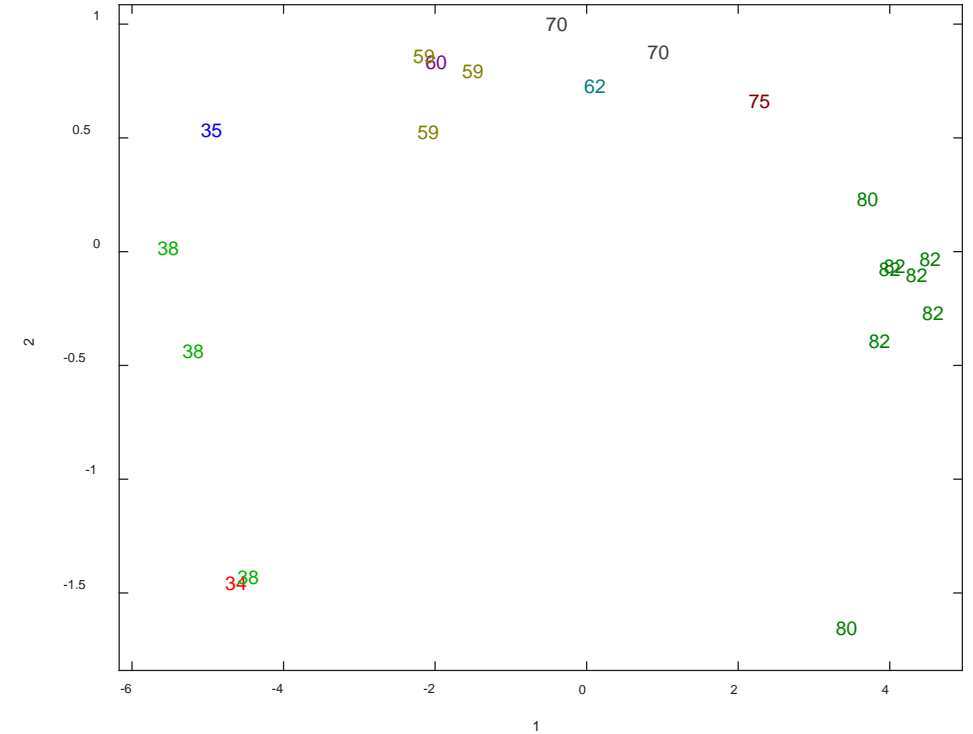
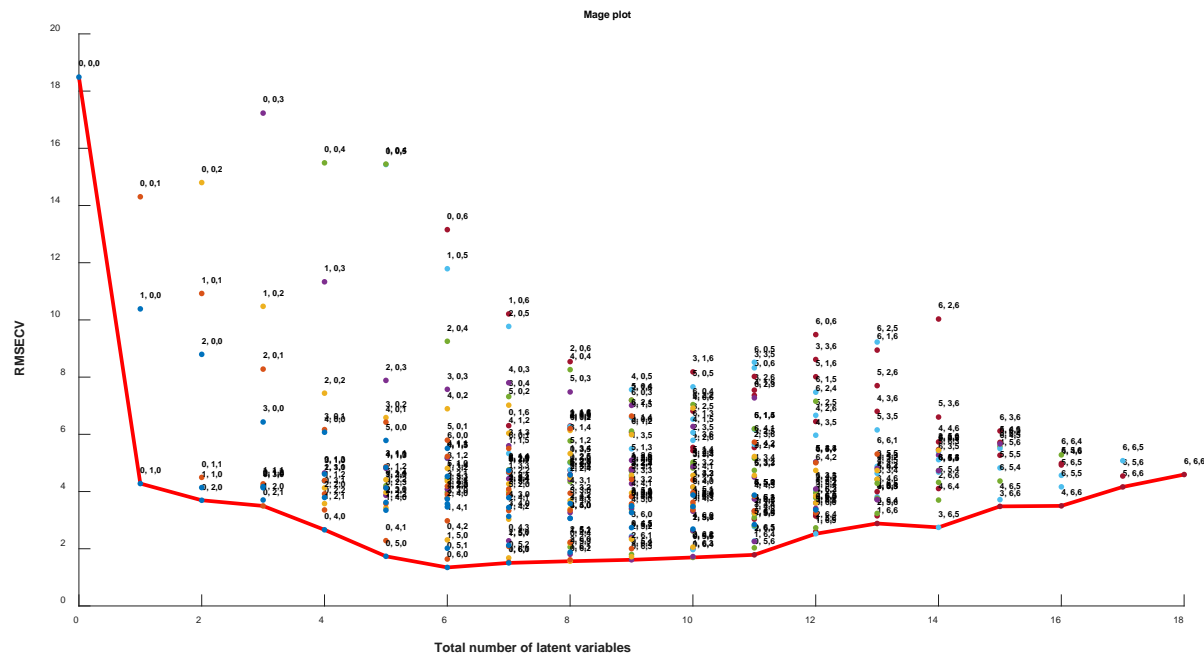
Package « R » Rchemo Marion Brandolini-Bunlon & Matthieu Lesnoff

> MIR + NIR + VIS

- Pretraitement = centrage 'mean'
- Model.CV :
 - Leave-one-out
 - LVopt: [0 6 0] [MIR NIR VIS]
 - r2opt: 0.9942
 - rmsecvopt: 1.3460

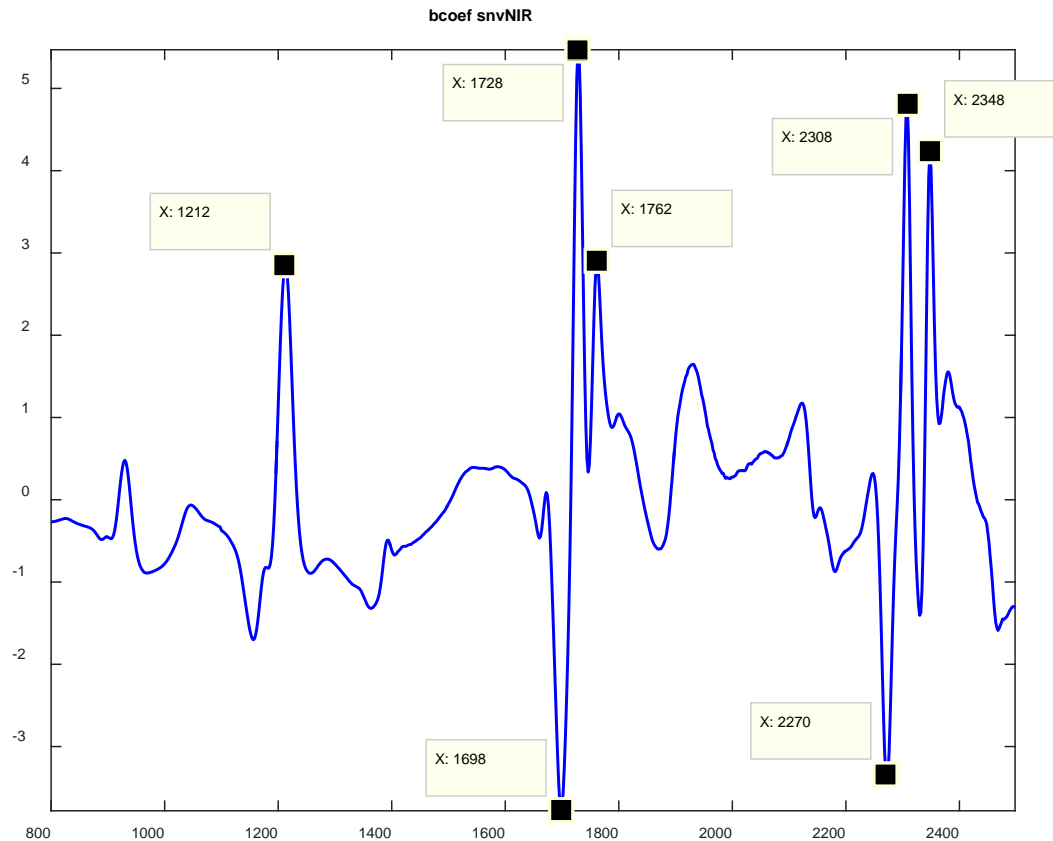
MG.d	model.CV.predYopt
ans =	ans =
80	80.5292
82	82.0847
82	82.2180
70	67.7924
82	82.3725
82	82.8480
82	82.3427
82	80.8525
70	71.6446
34	33.2823
38	37.6678
38	39.0606
60	57.5636
62	65.7025
35	37.2263
59	59.2315
80	80.0352
59	58.5618
59	59.4874
38	37.4561
75	74.1577

➤ Mageplot + Scores



Très bonne prédiction de la teneur en MG, faible erreur, séparation correcte des échantillons
6 LV optimales : toutes dans le NIR, mais 2 LV suffisent pour avoir un modèle plus parcimonieux.

➤ Conclusion teneur en MG



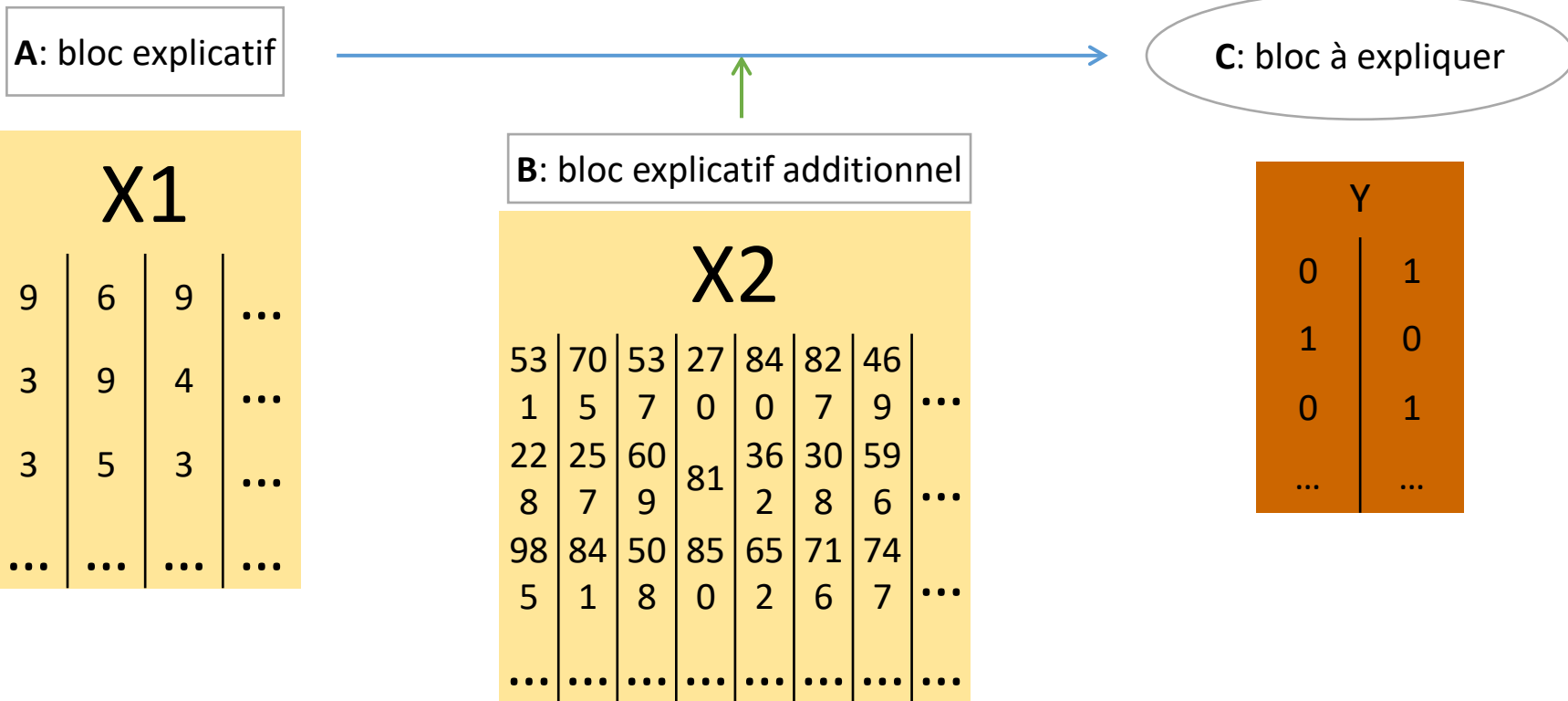
	MIR + NIR + VIS
LV opt	0 – 6 - 0
R ² opt	0.9942
RMESCV opt	1.3460

Meilleur modèle obtenu avec NIR => le plus performant pour prédire la teneur en MG,

SOPLS beurre margarine B/M

Y = codage Beurre/Margarine

Sequential and Orthogonalized-PLS (DA)



1. Régression PLS de **A** sur **C**
2. Orthogonalisation de **B** par rapport aux composantes obtenues
3. Régression PLS de **B_{orth}** sur la matrice de résidus $E_{(C \leftarrow A)}$
4. Prédiction finale de C, par addition des predictions des 2 modèles de regression

> MIR + NIR + VIS

- Pretraitement = centrage 'mean'
- Model.CV :
 - LDA sur y prédits
 - Leave-one-out
- LVopt: [3 0 0] [MIR NIR VIS]
- accuracyopt: 0.9524
- rmsecvopt: 0.3588

BM.d

model.CV.predclassopt

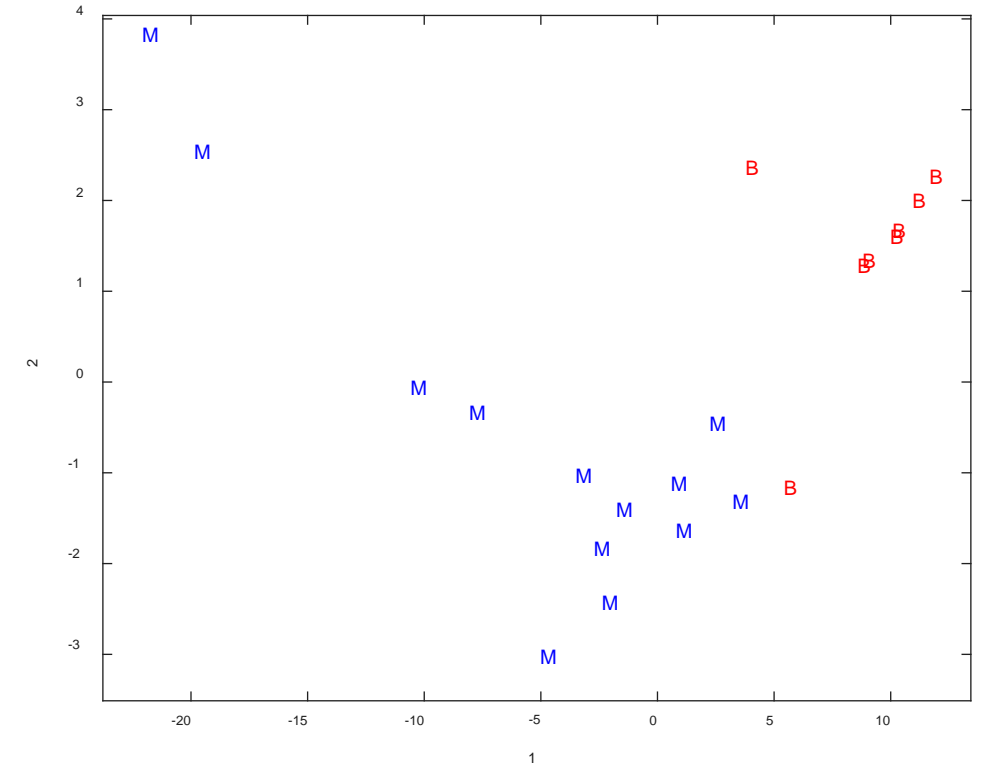
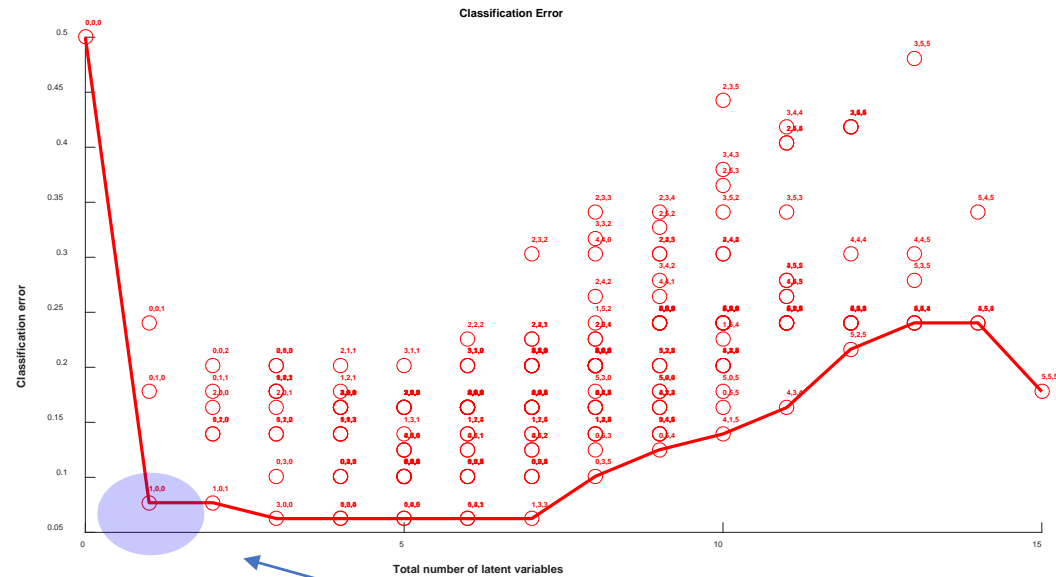
ans =

1
1
1
1
1
1
1
1
2
2
2
2
2
2
2
2
2
2
2
2
2
2
2

ans =

1
1
1
2
1
1
1
1
2
2
2
2
2
2
2
2
2
2
2
2
2
2
2

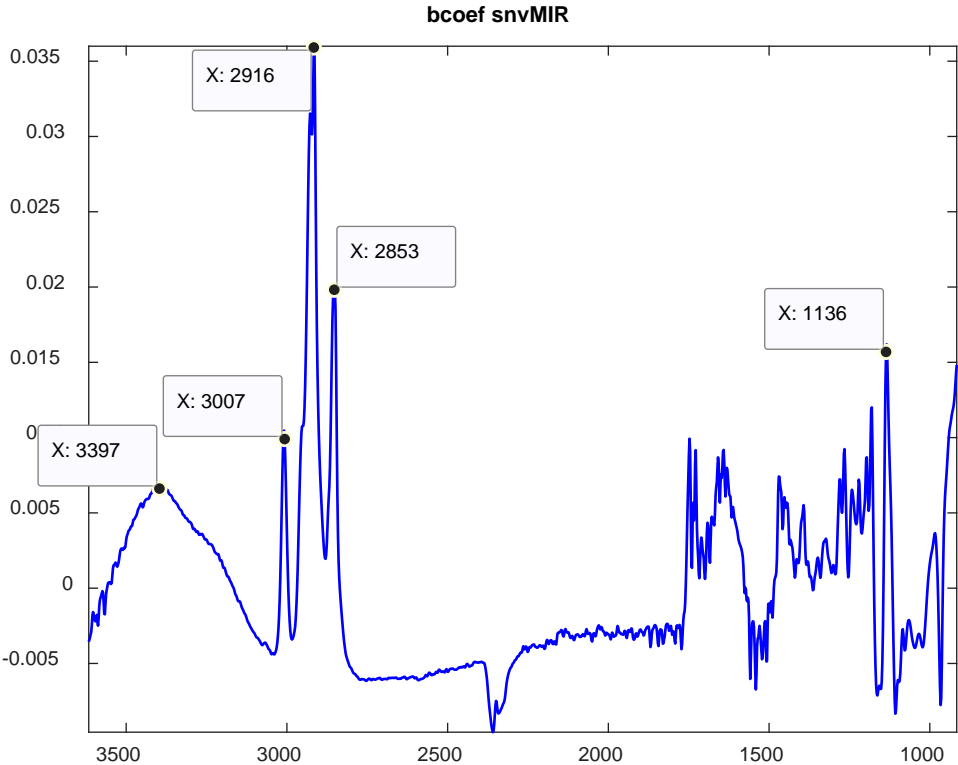
➤ Mageplot + concscores



NB : sur la base de l'erreur de classification, on aurait 1 LV pour un modèle plus parcimonieux. De fait, il n'y a que deux niveaux au facteur « typologie de produit », donc on ne devrait avoir qu'une composante.

Conclusion B vs M

	MIR + NIR + VIS
LV opt	3 – 0 - 0
accuracyopt	0.9524
RMESCV opt	0.3588



BKMO70
 Beurre irlandais à
 70% MG

BM.d	model.CV.pred
ans =	classopt
ans =	ans =
1	1
1	1
1	1
1	2
1	1
1	1
1	1
1	1
2	2
2	2
2	2
2	2
2	2
2	2
2	2
2	2
2	2
2	2
2	2
2	2
2	2
2	2
2	2
2	2
2	2

Seul le MIR sert à discriminer

> Conclusion

- Dans ces travaux avec **uniquement** des blocs **spectroscopiques**:
- « **Faible** » **apport** de la MB-PCA pour distinguer les produits en fonction de MG et typologie.
- **Fort apport** de la SO-PLS qui :
 - Permet de savoir quels sont les blocs qui portent de l'information et combien de composantes par bloc sont nécessaires,
 - Confirme le savoir du chimiste sur les méthodes à utiliser en fonction du but recherché.

Merci de votre attention