



HAL
open science

Approches PLS linéaires et non linéaires pour la modélisation de multi-tableaux. Théorie et applications.

Myrtille Vivien

► To cite this version:

Myrtille Vivien. Approches PLS linéaires et non linéaires pour la modélisation de multi-tableaux. Théorie et applications.. Mathématiques [math]. Migration - université en cours d'affectation, 2002. Français. NNT: . tel-00002118v1

HAL Id: tel-00002118

<https://theses.hal.science/tel-00002118v1>

Submitted on 12 Dec 2002 (v1), last revised 5 Feb 2003 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÉ MONTPELLIER I

— UFR DES SCIENCES PHARMACEUTIQUES —

N° □□□□□□□□□□

année 2002

Approches PLS linéaires et non-linéaires pour la modélisation de multi-tableaux : théorie et applications

Thèse présentée pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE MONTPELLIER I

ECOLE DOCTORALE : *Information, Structure et Systèmes*FORMATION DOCTORALE : *Biostatistique*

N° de DISCIPLINE : 42

N° CNU : 26

SPECIALITE : *Statistique*

 PAR

Myrtille VIVIEN

*soutenue publiquement le***3 décembre 2002***devant le JURY :*

M. Gilles Ducharme	Professeur, Université Montpellier II	Invité
M. Jean-François Durand	Maître de Conf., Université Montpellier II	
M. Yves Escoufier	Professeur, Université Montpellier II	
M. Georges Oppenheim	Professeur, Université de Marne la Vallée	Rapporteur
M. Robert Sabatier	Maître de Conf., Université Montpellier I	Directeur de thèse
M. Gilbert Saporta	Professeur, CNAM Paris	Rapporteur
M. Michel Tenenhaus	Professeur, HEC Paris	

Résumé

Ce travail concerne les analyses multi-tableaux de données et plus particulièrement la régression. Présentées sous forme d'un recueil, nous exposons un grand nombre de méthodes peu connues pour la majorité, pour l'analyse des multi-tableaux. Notre travail se compose de trois parties. La première partie, contenant sept chapitres, expose les méthodes de modélisation linéaire. Elles sont classées par type d'analyse (conjointe ou régression) et par domaine d'application (cube ou multi-tableaux). Nous en introduisons plusieurs nouvelles de type PLS, basées sur l'optimisation d'un critère sous contraintes par étape, pour analyser simultanément un (ACIMO) ou plusieurs (ACIMOG1,2,3) ensembles de tableaux, pour modéliser un ensemble de tableaux par un autre (ACIMO-PLS, ACIMOG-PLS1,2,3, REMUB) et appliquons l'ACIMO à l'intrastructure de STATIS pour en améliorer les trajectoires. De plus, nous proposons une généralisation de la méthode STATIS au cas de deux ensembles de tableaux (DO-ACT), aussi basée sur l'optimisation d'un critère sous contraintes. Un chapitre est consacré à ce qui touche la pratique de ces méthodes, dans lequel nous abordons la question du choix de blocs explicatifs importants dans les modèles de régression multi-tableaux, qui n'a encore jamais été abordée. Nous proposons plusieurs critères pour répondre à cela. La seconde partie, moins importante en taille, concerne la modélisation non-linéaire en régression multi-tableaux. Après une revue des méthodes de régression PLS non-linéaire pour deux tableaux, nous introduisons deux méthodes (SARMUB1 et 2) pour la régression non-linéaire des multi-tableaux, basées sur la recherche de transformations splines des prédicteurs optimales. Enfin, nos méthodes, programmées en S-Plus[®], ayant pour but principal d'être appliquées, nous proposons dans la troisième partie les résultats de plusieurs applications dans divers domaines, tels que la chimiométrie, l'analyse sensorielle, l'écologie ...

Mots-clés : régression PLS, mutli-tableaux, cubes, régression multi-blocs, B-splines, régression PLS non-linéaire, optimisation sous contraintes, méthode STATIS

Abstract

This work deals with mutliblock data analysis and especially with regression. This thesis presents, like a collection of methods, a large number of methods of multiblock data analysis, bad known for the majority of them. This thesis contains three parts. The first one, with seven chapters, presents methods for linear modelling. They are classified by kind of analysis (simultaneous or regression) and by applicability domain (multi-block or multiway). We propose several new PLS-like one's, stepwise constrained optimization based, dedied to the simultaneous analysis of one (OMCIA) or more sets (GOMCIA1,2,3) of matrices and to model a set of matrices with another one (OMCIA-PLS, GOMCIA-PLS1,2,3, MUBRE), OMCIA allowing to obtain better trajectories in the STATIS method. Moreover, we propose a generalization of the STATIS method to two multiblock tables (DO-ACT). A chapter is dedied to the practice of the multiblock methods, in which propose some criteria to choose the important predictive blocks in regression multi-block models. In the second part, we deal with the non-linear modelling. We start to give a review of non-linear PLS methods for two matrices and then propose two non-linear multiblock regression methods (MUBRAS), searching for the optimal spline tranformations of the predictors. To end, in the third part, we present several applications of our methods, programmed in S-Plus[®] language, on real data from chemometrics, sensory analysis, ecology

Keywords : PLS regression, multiblock, multiway data, multi-block regression, B-splines, nonlinear PLS regression, constrained optimization, STATIS method

Remerciements

En rentrant en DEA, je n'avais pas l'intention de faire une thèse. Le stage de recherche m'a fait réaliser que cela me plairait bien. Toute ma gratitude va à Robert Sabatier pour m'avoir donné le goût de la recherche, pour ses encouragements permanents et pour la confiance qu'il m'a accordée durant ces trois ans et demi.

Je remercie aussi Luc Maury pour m'avoir accueillie dans son laboratoire. Je tiens à dire qu'il est agréable de travailler dans ce laboratoire où il règne une ambiance joviale. Du fond de mon aquarium glacé, j'ai apprécié cette chaleur humaine même si elle n'a malheureusement pas suffi pour me réchauffer

Je le remercie à nouveau pour m'avoir aidée à transférer mon monitorat du STAPS vers la faculté de Pharmacie pour que je puisse enfin enseigner des statistiques.

Toujours, concernant mes enseignements, je n'oublie pas Mme Hélène Lubes-Niel qui m'a aiguillée vers l'ISIM alors qu'ils cherchaient un enseignant de statistiques. Merci d'avoir pensé à moi. Cela m'a bien servi.

M. Jean-François Durand a été le premier à me faire découvrir l'analyse de donnée et la régression PLS, largement citée dans ce travail. A ce titre je suis heureuse qu'il ait accepté de faire partie de mon jury. Sa lecture pointilleuse m'a permis d'améliorer mon manuscrit et de me remettre les idées en ordre. Merci.

Je remercie M. Gilbert Saporta pour avoir accepté de rapporter ce travail et MM. Michel Tenenhaus et Yves Escoufier pour avoir accepté d'être là, dans mon jury. Merci à M. Georges Oppenheim pour avoir accepté d'être rapporteur et pour tout l'intérêt qu'il a manifesté à l'égard de mon travail.

M. Gilles Ducharme, malgré notre sollicitation tardive, a accepté d'être membre invité de mon jury. Je lui en suis reconnaissante.

Je remercie aussi Mme Pascale Lacroix et Mme Anne Jauneau-Marvier de NESTLE France pour m'avoir prêté leurs données et m'avoir autorisée à les utiliser ici. Pour les mêmes raisons, je remercie Nabila Mazouni (CEPRALMAR), Monique Simier (IRD Montpellier), et MM. Bernard Vidal et Jean-Paul Biesse d'ALTADIS, en espérant que nos collaborations ne s'arrêtent pas là.

Une pensée va bien sûr vers tous mes ami(e)s qui ont su me distraire, me remonter le moral et avec qui j'apprécie passer une partie de mon temps libre. Je n'oublie pas ici les copines de la danse ainsi que les deux profs, Carole E. et Murielle B., dont les cours vont me manquer profondément quand je ne pourrai plus les suivre. J'en profite ici pour les remercier pour leurs cours dans lesquels je me retrouve et que j'ai suivis avec assiduité pendant toutes ses années à la fac, en particulier pendant mes trois années de thèse. Et un très grand MERCI à tous ceux qui vont mettre la main à la pâte pour mon pot

Pour finir, William, je sais que cela n'a pas toujours été facile de me supporter au quotidien surtout ces derniers mois. Cela devrait bientôt aller mieux. Merci d'être là, tout simplement.

Begin at the beginning and go on till you come to the end, then stop

Lewis Carroll

Table des matières

Notations	viii
Acronymes	x
Introduction générale	1
I Des modèles linéaires ...	7
Introduction	9
1 Analyse d'un cube ou Analyse 3 modes	11
1.1 Le modèle TUCKER3	12
1.1.1 Méthode 1	13
1.1.2 Méthode 2	13
1.1.3 TUCKALS3	14
1.1.4 Commentaires	15
1.2 Le modèle PARAFAC1/CANDECOMP	17
1.2.1 Commentaires	18
1.3 Le modèle PARATUCK2	19
2 Analyse de la Co-inertie de tableaux	23
2.1 STATIS et STATIS dual	23
2.1.1 Etude de l'interstructure	24
2.1.2 Le compromis	26
2.1.3 Etude de l'intrastructure	27
2.2 Analyse de CO-inertie Multiple: ACOM	28
2.2.1 Définitions et solutions	28
2.2.2 Commentaires	30
2.3 Analyse de concordance: CONCORD ou ACO	31
2.3.1 Définitions et solutions	32
2.3.2 Commentaires et cas particuliers	34
2.4 Analyse de Co-Inertie Multiple Orthogonale: ACIMO	35
2.4.1 La méthode	35
2.4.2 Commentaires et cas particuliers	38
2.5 Application de l'ACIMO à l'intrastructure de STATIS	39

2.5.1	STATIS et ACIMO	39
2.5.2	STATIS dual et ACIMO	41
3	Analyse Conjointe de plusieurs cubes ou multi-tableaux	43
3.1	Multiway Multiblock Covariates Component models: MWMBcovC	43
3.2	STATIs et Co-inertie: STATICO	46
3.2.1	Analyse de Co-Inertie et Analyse Triadique Partielle	46
3.2.2	STATICO: STATIs et CO-inertie	47
3.3	DOuble Analyse Conjointe de Tableaux: DO-ACT	48
3.3.1	La méthode	48
3.3.2	Interstructure - compromis - infrastructure	50
3.3.3	Cas particuliers	51
3.4	Analyse de Co-Inertie Multiple Orthogonale Généralisée: ACIMOG	51
3.4.1	Problème	51
3.4.2	Solutions d'ordre 1	52
3.4.3	L'algorithme	53
3.4.4	Les solutions d'ordre \mathbf{a} et propriétés	55
3.4.5	Formules de reconstitution	57
3.4.6	Commentaires	59
4	Régression cubique ou multiway regression	61
4.1	Multiway PLS: MPLS	61
4.2	Linear Three-way Decomposition: LTD	64
4.3	Multilinear PLS: NPLS	66
4.4	Régression sur les facteurs de PARAFAC	71
4.5	Multiway Covariates Regression Models	72
5	Régression multi-blocs	77
5.1	Approche PLS: APLS	78
5.2	Multi-Block Regression: MBPLS	82
5.2.1	Cas de \mathbf{K} tableaux explicatifs et \mathbf{N} tableaux réponses	82
5.2.2	Cas de \mathbf{K} tableaux explicatifs et un tableau réponse ($\mathbf{N} = 1$)	85
5.3	PLS hiérarchique: HPLS	86
5.4	Serial PLS: SPLS	90
5.5	Analyse de Co-Inertie Multiple Orthogonale PLS: ACIMO-PLS	92
5.5.1	Lien avec la régression PLS	93
5.5.2	Modélisation des tableaux $\{Y_n\}_n$	94
5.5.3	Commentaires	95
5.6	Analyse de Co-Inertie Multiple Orthogonale Généralisée PLS: ACIMOG-PLS	96
5.6.1	Définitions et solutions	96
5.6.2	Les modèles	98
5.6.3	Prédiction	99
5.6.4	Interprétation des modèles	108
5.6.5	Commentaires	109
5.6.6	A propos de la déflation et variantes	111
5.7	REGression Multi Blocs: REMUB	114

5.7.1	La méthode	114
5.7.2	L'algorithme	116
5.7.3	Propriétés et modèles	119
5.7.4	Cas particuliers	120
6	Autour de la pratique	123
6.1	Prétraitement des données	123
6.2	Validation des modèles et choix du nombre de composantes	126
6.2.1	Critère d'ajustement	126
6.2.2	Validation Croisée	127
6.3	Sélection	130
6.3.1	Sélection de variables en régression PLS	130
6.3.2	Sélection de blocs en ACIMOG-PLS et REMUB	132
6.3.2.1	Critère basé sur les coefficients du modèle	132
6.3.2.2	Critère: Bloc Importance in Prediction	135
6.3.2.3	Utilisation du coefficient RV	136
6.3.2.4	Test de permutations	138
7	Résumés et comparaisons	141
7.1	Relation entre MBPLS-REMUB-PLS-ACIMO	141
7.1.1	REMUB-MBPLS	141
7.1.2	MBPLS-ACIMO	142
7.1.3	MBPLS-PLS	143
7.2	Résumé en tableaux	145
7.2.1	Commentaires	145
7.2.2	Quelle méthode choisir?	149
II	... Aux modèles non linéaires	151
	Introduction	153
1	Modèles non linéaires PLS en régression: cas de deux tableaux	155
1.1	PLS Quadratique	156
1.2	Non Linear PLS: NLPLS	158
1.3	SPLine PLS: SPLPLS	159
1.4	Continuum Non Linear PLS: CNLPLS	161
1.5	Implicit Non Linear latent variable Regression: INLR	163
1.6	Splines Additives en régression PLS	164
1.7	Régression PLS Spline	167
1.8	GIFI PLS	168
1.9	Discussion	169
2	Modélisation non linéaire de plusieurs tableaux	171
2.1	Splines Additives en Régression Multi-Bloc: SARMUB	172
2.1.1	La première étape, $a = 1$	173
2.1.1.1	L'algorithme	175

2.1.1.2	Monotonicit� et convergence	176
2.1.1.3	Passage � l'�tape suivante	177
2.1.2	SARMUB1, �tapes $a > 1$	178
2.1.2.1	Les solutions	178
2.1.2.2	L'algorithme	181
2.1.3	SARMUB2, �tapes $a > 1$	184
2.1.3.1	Les solutions	185
2.1.3.2	Probl�mes rencontr�s	186
2.1.3.3	Proposition d'un algorithme	188
2.1.4	Nature des solutions d'ordre a : la matrice Hessienne	189
2.1.5	Propri�t�s et mod�les	191
2.1.6	Commentaires	193
2.2	Autres approches	194
 III Applications		197
Introduction		199
1	Biom�trie: ACIMO et STATIS	201
1.1	STATIS et l'ACIMO	202
1.2	STATIS dual et ACIMO	206
1.3	Conclusion	207
2	Environnement: ACIMO-PLS et ACIMO-PLSS	211
2.1	Pr�sentation des donn�es	211
2.2	ACIMO-PLS	211
2.3	L'ACIMO-PLSS	218
3	Ecologie: ACIMOG-PLS1	223
3.1	Les donn�es	223
3.2	Application de l'ACIMOG-PLS1	224
4	Chimiom�trie: ACIMOG-PLS, HPLS, MBPLS	233
4.1	R�sultats num�riques	234
4.2	R�sultats graphiques de l'ACIMOG-PLS	236
4.3	Importance des tableaux	239
4.4	Sans le NIRS	239
4.5	Pour conclure	239
5	Analyse Sensorielle: ACIMOG-PLS2 et 3, REMUB, ACIMOG, DO-ACT	243
5.1	Pr�sentation des donn�es	243
5.2	Situation 1: explication des r�ponses des consommateurs par les �valuations des experts	246
5.2.1	Pr�traitement	247
5.2.2	Application de l'ACIMOG-PLS 2 et 3	247
5.2.3	Application de REMUB	254

5.3	Situation 2: analyse de la proximité des panels de consommateurs initiés et non initiés: ACIMOG, DO-ACT	254
5.3.1	ACIMOG	254
5.3.2	DO-ACT	257
5.4	Conclusion	262
Discussion et conclusion		267
Annexes		273
A La régression PLS		273
A.1	PLS: méthode algorithmique	273
A.2	PLS: une optimisation	274
A.3	Commentaires	275
B Rappels mathématiques		277
B.1	La méthode des multiplicateurs de Lagrange	277
B.2	Les splines de régression	278
C A propos des représentations graphiques ou la perte de la dualité		281
C.1	En régression PLS	282
C.2	En analyse multi-tableaux	283
C.3	Pour conclure	284
Bibliographie		285

Notations

- x, y, \dots , des vecteurs colonne.
- X, Y, \dots des matrices.
- X' tranposé de X .
- si $X = [x_1 \dots x_J]$, $l \times J$, $X[k]$ est le $k^{\text{ième}}$ bloc ligne de X dont la dimension dépend du contexte. $x_j(k)$ est l'élément de la colonne j et de la ligne k de X .
- si x est de longueur J , $x(j)$ désigne son $j^{\text{ième}}$ élément et $x[k]$ son $k^{\text{ième}}$ bloc.
- $\underline{X}, \underline{Y}$ cubes, multi-tableaux réguliers d'ordre 3. Le terme «cube» est employé par abus de langage et simplicité. Ces tableaux à trois entrées sont en fait des parallélépipèdes rectangles.
- $l \times J \times K$ dimension du cube \underline{X} , variables explicatives.
- i, j, k les indices courants.
- $l \times M \times N$ dimension du cube \underline{Y} , variables réponses.
- i, m, n les indices courants.
- $X_i, i = 1, \dots, l$ matrices $J \times K$ dites «tranches horizontales» du cube \underline{X} , ce sont les matrices indicées suivant le premier mode.
- $X_j, j = 1, \dots, J$ matrices $K \times l$ dites «tranches latérales» du cube \underline{X} , ce sont les matrices indicées suivant le second mode.
- $X_k, k = 1, \dots, K$, les matrices $l \times J$ indicées suivant le troisième mode, dites «tranches frontales» du cube \underline{X} .
- x_{ijk} est l'élément de \underline{X} à l'intersection de la ligne i et de la colonne j de la $k^{\text{ième}}$ tranche frontale de \underline{X} .
- $\mathbf{X}_l = [X_1 \dots X_K]$, matrice $l \times JK$ correspondant au cube \underline{X} déplié le long du premier mode. \mathbf{X}_l est souvent noté X pour simplifier lorsqu'il n'y a pas d'ambiguïté.
- $\mathbf{X}_J = [X_1 \dots X_l]$, matrice $J \times Kl$ correspondant au cube \underline{X} déplié le long du second mode.
- $\mathbf{X}_K = [X_1 \dots X_J]$, matrice $K \times lJ$ correspondant au cube \underline{X} déplié le long du troisième mode.
- $X_k, k = 1, \dots, K$, dans le cas des analyses non-cubiques, est un ensemble de K tableaux (le plus souvent explicatifs) de taille $l \times J_k$. Si tous les J_k sont égaux, on les notera J . Si l'on est en présence d'un seul tableau (explicatif) ($K = 1$), on le notera simplement par X . $X = [X_1 \dots X_K]$ est le tableau conjoint des X_k .
- $Y_n, n = 1, \dots, N$, dans le cas des analyses non-cubiques, est un ensemble de N tableaux (en général à expliquer) de taille $l \times M_n$. Si tous les M_n sont égaux, on les notera M . Et si l'on est en présence d'un seul tableau réponse, on le notera simplement Y . $Y = [Y_1 \dots Y_N]$ est le tableau conjoint des Y_n .
- $Diag(A_i)$ est la matrice diagonale par blocs dont le bloc i est la matrice A_i . Cette notation est aussi valable dans le cas d'une suite de vecteurs.

- $diag(x)$ est la matrice diagonale dont la diagonale est composée des éléments du vecteur x .
- $diag(X)$ est le vecteur composé de la diagonale de la matrice X .
- D est la métrique des poids des individus, en général c'est la métrique uniforme $\frac{1}{I}Id_I$.
- Q_k^x , $k = 1, \dots, K$ et Q_n^y , $n = 1, \dots, N$ sont respectivement les métriques dans les espaces \mathbb{R}^{J_k} et \mathbb{R}^{M_n} . En général, ce sont les métriques identité.
- $W_k^x = X_k Q_k^x X_k'$ opérateur des produits scalaires entre les individus du tableau X_k .
- $W_n^y = Y_n Q_n^y Y_n'$ opérateur des produits scalaires entre les individus du tableau Y_n .
- $V_k^x = X_k' D X_k$ la matrice des produits scalaires entre les variables de X_k . Si le tableau X_k est D -centré, c'est la matrice de variance-covariance, et s'il est D -réduit, c'est la matrice des corrélations.
- $V_n^y = Y_n' D Y_n$, *idem* pour Y_n .
- $P_X = X(X' D X)^+ X' D$ projecteur D -orthogonal sur l'espace engendré par les colonnes de X . X peut être un vecteur.
- $P_X^\perp = Id - P_X$.
- A le nombre final de composantes dans les modèles.
- a l'indice courant.
- \otimes est le plus souvent le produit de Kronecker, quelques fois note aussi le produit tensoriel; cela sera précisé dans le contexte.
- \odot produit de Khatri-Rao appelé aussi produit de Kronecker colonne par colonne.
- $*$ produit de Hadamard, produit terme à terme.
- $var(x)$ D -variance du vecteur x . En général, c'est la variance usuelle.
- $var(X)$ variance usuelle de la matrice X .
- $tr(X)$ trace de la matrice carrée X .
- $cov(x, y)$ D -covariance entre les vecteurs x et y .
- $r(x, y) = cov(x, y) / \sqrt{var(x)var(y)}$ D -corrélation entre les vecteurs x et y .
 - $cov(X, Y)$ matrice des covariances entre les colonnes de X et les colonnes de Y .
 - $\|x\|^2 = x'x$ carré de la norme de x .
 - $\|x\|_M^2 = x' M x$ carré de la norme de x au sens de la métrique M .
 - $\|X\|^2 = tr(X' X)$ carré de la norme de Frobenius de X . Si X est centrée en colonne, $\|X\|^2 = var(X)$.
 - $\|X\|_M^2 = tr(X' M X)$ carré de la norme au sens de M de X .
 - $In(X)$ inertie du triplet (X, Q, D) par rapport à l'origine. $In(X) = tr(X' D X Q)$ avec X centrée en colonne pour D .
 - $covV(X, Y) = tr(X Q_x X' D Y Q_y Y' D)$ covariance vectorielle entre les deux triplets (X, Q_x, D) et (Y, Q_y, D) .
 - $varV(X) = tr(X Q_x X' D X Q_x X' D)$ variance vectorielle du triplet (X, Q_x, D) .
 - $RV(X, Y) = covV(X, Y) / \sqrt{varV(X)varV(Y)}$ corrélation vectorielle entre les triplets (X, Q_x, D) et (Y, Q_y, D) .
 - $\|X\|_{H,S}^2 = tr(X Q_x X' D X Q_x X' D) = varV(X)$ carré de la norme de Hilbert-Schmidt de (X, Q_x, D) .

On suppose que tous les tableaux sont D -centrés. Si cette hypothèse est omise, cela sera précisé dans le contexte.

Acronymes

ACIMO	Analyse de Co-Inertie Multiple Orthogonale
ACIMOG	Analyse de Co-Inertie Multiple Orthogonale Généralisée
ACIMOG-PLS	Analyse de Co-Inertie Multiple Orthogonale Généralisée - Partial Least Squares
ACIMO-PLS	Analyse de Co-Inertie Multiple Orthogonale - Partial Least Squares
ACIMO-PLSS	Analyse de Co-Inertie Multiple Orthogonale - PLS Spline
ACG	Analyse Canonique Généralisée
ACO	Analyse de Concordance
ACOM	Analyse de CO-Inertie
ACP	Analyse en Composante Principale
ACT	Analyse Conjointe de Tableaux (STATIS)
AIB	Analyse Inter-Batterie de Tucker
APG	Analyse Procruste Généralisée
APLS	Approche PLS
ASPLS	Additive Spline Partial Least Squares
BIP	Bloc Importance in the Prediction
BIP _{rv}	BIP basé sur le coefficient RV
CANDECOMP	CANonical DECOMPosition
CNLPLS	Continuum Non Linear PLS
DO	Direct Orthogonalization
DO-ACT	DOuble Analyse Conjointe de Tableaux
DOSC	Direct Orthogonal Signal Correction
GIFI-PLS	GIFI-ing in Partial Least Squares
HPLS	Hierarchical PLS
INLR	Implicit Non Linear Regression
LTD	Linear Three-way Decomposition
MBPLS	Multi-Block Regression
MPCA	Multiway Pricipal Component Analysis
MPLS	Multiway PLS
MWcovR	Multiway covariates Regression model
MWMBcovC	Multiway Multi-Block covariates Component model
NIPALS	Non linear Iterative PARTial Least Squares algorithm
NIRS	Near Infra-Red Spectra (spectre proche infra-rouge)
NLPLS	Non Linear Partial Least Squares
NPLS	Multilinear PLS
O-ACIMOG	Orthogonal signal correction for ACIMOG

suite page suivante...

OSC	Orthogonal Signal Correction
O-PLS	Orthogonal Projection to Latent Structure
O2-PLS	Alternative Orthogonal Projection to Latent Structure
PARAFAC	PARAllel FACtor analysis
PARATUCK2	modèle entre PARAFAC et TUCKER2
PLS	Partial Least Squares (régression)
PLSS	régression Partial Least Squares Splines
QPLS	Quadratic Partial Least Squares régression
REMUB	REgression MUlti-Bloc
SARMUB	Splines Additives en Régression MUlti-Bloc
SPL-PLS	SPLines in Partial Least Squares
SPLS	Serial PLS
STATIS	Structuration des Tableaux à Trois Indices de la Statistique
TUCKALS	TUCKER by Alternating Least Squares
TUCKER1	modèle de Tucker avec réduction d'un seul mode
TUCKER2	modèle de Tucker avec réduction de deux modes
TUCKER3	modèle de Tucker avec réduction des trois modes
VIP	Variables Importance Selection

Introduction générale

Dans cette thèse nous proposons un regard général sur les méthodes d'analyse factorielle de multi-tableaux, c'est-à-dire sur les méthodes permettant l'analyse simultanée de plusieurs tableaux de données.

Les méthodes pour l'analyse d'un ensemble de tableaux ou d'un cube de données existent depuis plus de 35 ans. Par contre, les méthodes de régression multi-tableaux sont assez récentes. Elles sont apparues il y a environ 15 ans dans le domaine de la chimométrie. Devant le nombre croissant des variables qu'ils peuvent mesurer, les chimométriciens se trouvent confrontés à des problèmes d'interprétation et de sélection de variables. Pour éviter cela, ils ont introduit dans la littérature des algorithmes permettant de faire de la régression multi-blocs, c'est-à-dire, leur permettant d'utiliser toutes les variables mais regroupées en blocs. Ces méthodes sont applicables dans bien des disciplines. En analyse sensorielle, par exemple, on veut prédire une variable hédonique par des évaluations sensorielles effectuées par des experts. Chaque attribut sensoriel ou chaque expert peut constituer un tableau.

Il apparaît au travers des publications, que les méthodes connues et utilisées sont très variables selon les zones géographiques et les disciplines. Dans tous les domaines de recherche la quantité et le type de données mesurées croît de plus en plus vite grâce aux outils technologiques modernes. Avec les performances déjà importantes et sans cesse croissantes des ordinateurs, les praticiens sont en droit de vouloir garder le maximum d'information dans leur données en utilisant toutes les variables, plutôt que de les résumer ou d'en sélectionner, au risque de perdre de l'information. Il est de nos jours nécessaire de développer et d'utiliser des méthodes d'analyse de données adaptées à ces situations mettant en jeu de nombreuses variables, et souvent peu d'observations (individus), arrangées en plusieurs tableaux ou sous forme de tableau à trois entrées («cube»). Les méthodes d'analyse de plusieurs tableaux montrent ici tout leur intérêt.

Selon les disciplines les pratiques sont différentes. Ainsi, ce sont les psychométriciens qui ont les premiers développé des méthodes spécifiques à l'analyse d'un «cube» de données. Ces méthodes sont à différencier des méthodes d'analyse dites «multi-tableaux» ou analyse conjointe de tableaux, dont font partie les méthodes françaises telles l'Analyse Factorielle Multiple (ESCOFIER & PAGÈS, 1984, 1986), STATIS (L'HERMIER DES PLANTES, 1976), et les différentes Analyses Canoniques Généralisées (CARROLL, 1968; HORST, 1961; SABATIER, 1993). Ces méthodes semblent bien connues en France. Comme nous l'avons déjà souligné au début de cette introduction, les chimométriciens du nord de l'europe développent depuis une quinzaine d'années des méthodes de régression multi-tableaux, dites aussi multi-blocs, basées sur la régression Partial Least Squares (PLS) (WOLD, 1966b) bien connue pour son efficacité lorsque les prédicteurs sont corrélés et/ou le nombre d'observations est faible par rapport

au nombre de variables explicatives. Cette méthode est très adaptée à la chimiométrie où le nombre de variables mesurées est de plus en plus important, en particulier grâce à la technique peu coûteuse qu'est la spectrographie proche infrarouge. Les variables étant de plus en plus nombreuses les modèles de régression multi-variables deviennent impossible à interpréter et les chimiométriciens se sont ainsi penchés sur la construction de ces méthodes multi-blocs qui permettent de ranger les prédicteurs en différents groupes (blocs) qui seront alors interprétés comme un tout. En analyse sensorielle, où il se pose, par exemple, la question d'expliquer des évaluations de produits établies par des consommateurs par des évaluations de ces mêmes produits faites par des experts, ce type de méthodes reste encore méconnu en France. La régression multi-variables ordinaire est encore beaucoup appliquée mais la régression PLS commence à susciter de l'intérêt. Les méthodes de régression multi-tableaux ne semblent pas utilisées alors qu'elles pourraient être bien utiles, comme nous le verrons dans une application.

A travers cette thèse, nous voulons faire connaître les différentes familles de méthodes (analyse de co-inertie, régression, analyse de multi-tableaux ou analyse pour les «cubes» de données) en soulignant leurs avantages et inconvénients et en proposer de nouvelles. Ce travail devrait permettre aux statisticiens et praticiens de trouver une vue d'ensemble et une comparaison des méthodes récentes et/ou peu connues et d'autres complètement nouvelles, ainsi que bon nombre de références sans oublier des applications des méthodes originales à des données réelles.

Notre travail est orienté vers les méthodes multi-tableaux (ou multi-blocs) et particulièrement vers la régression multi-blocs de type PLS. Cependant, nous avons choisi d'exposer les principales méthodes cubiques afin de mettre en évidence la différence de raisonnement entre les deux types d'approches et permettre des comparaisons. Ce mémoire se présente sous la forme de trois parties. La première concerne la modélisation linéaire, c'est-à-dire les méthodes générant des modèles linéaires en les variables initiales. Le travail effectué sur les méthodes linéaires a permis d'entamer un travail très vaste concernant la régression multi-tableaux non-linéaire grâce à l'utilisation de transformations splines, ce qui constitue la seconde partie. Enfin, la dernière partie présente des applications à des données réelles. Les lecteurs non familiers avec les multi-tableaux pourront commencer par lire la troisième partie pour avoir des exemples de problèmes réels.

Les deux premières parties sont théoriques. Dans chacune d'elles, nous présentons différentes méthodes issues de la littérature et en proposons de nouvelles. Toutes ces nouvelles approches sont basées sur l'optimisation d'un critère sous contraintes, par étape, généralisant celui de la régression PLS (excepté pour DO-ACT, (§3.3)). Tout au long de cette thèse, des références à la régression PLS sont faites, celle-ci étant la base de notre travail. Nous verrons que très peu de méthodes de type PLS présentes dans la littérature sont le résultat de l'optimisation d'un critère.

Nous avons choisi de présenter notre travail comme un recueil de méthodes. Nous avons classé celles-ci par domaine d'application. Ainsi, nous distinguons les méthodes d'analyse «cubique» des méthodes d'analyse multi-blocs, et les approches visant l'analyse conjointe (simultanée) de tableaux des approches de régression ayant pour but d'expliquer un «cube» (ou un ensemble de tableaux) par un «cube» de prédicteurs (ou un ensemble de tableaux prédicteurs). Grâce à cette classification, un expérimentateur pourra accéder directement au chapitre décrivant les méthodes adaptées à son problème.

Ainsi, le premier chapitre de la première partie présente trois méthodes pour l'analyse d'un cube. Nous avons choisi de présenter les deux plus anciennes (TUCKER3 et PARAFAC) qui sont à l'origine de toutes les autres et une des plus récentes (PARATUCK2). Nous ne donnons aucune application de ces méthodes mais elles sont utiles pour la compréhension de deux méthodes présentées dans les chapitres suivants. Ce chapitre est particulier dans ce mémoire, dans la mesure où aucune méthode nouvelle n'est proposée.

Le second chapitre concerne l'analyse conjointe de tableaux, ou encore l'analyse de la co-inertie de tableaux. Il s'agit ici de chercher et d'expliquer l'éventuelle information commune à un ensemble de tableaux mesurés sur les mêmes individus. Les variables peuvent être en nombre différent selon les tableaux. Pour ce type de situation, il existe beaucoup d'approches telles que les analyses Procrustes Généralisées (GOWER, 1975), les Analyses Canoniques Généralisées (CARROLL, 1968) ou autres ACP particulières (Analyse Factorielle Multiple (ESCOFIER & PAGÈS, 1984), Analyse en Composantes Principales Généralisée (CASIN, 2001)). Nous n'avons pas choisi de parler de celles-ci qui sont déjà bien répandues dans la littérature appliquée. Par contre, nous présentons deux méthodes plus récentes (Analyse de CO-inertie Multiple, ACOM, et Analyse de la COncordance, ACO), surtout connues en écologie française au travers du logiciel ADE-4[®] (THIOULOUSE, CHESSEL, DOLÉDEC & OLIVIER, 1997), pour lesquelles nous donnons une démonstration de la nature des solutions par la méthode des multiplicateurs de Lagrange. De plus, elles utilisent le même critère, avec des contraintes différentes, que l'une des nouvelles méthodes que nous proposons, l'ACIMO, Analyse de Co-Inertie Multiple Orthogonale. L'ACOM, l'ACO et l'ACIMO sont des méthodes procédant à la recherche de composantes par étapes. L'ACO et l'ACIMO ont pour but d'étudier la proximité entre un ensemble de tableaux et un tableau supplémentaire dit de référence, et ont la première étape en commun. Les étapes suivantes, obtenues par déflations des tableaux initiaux, sont différentes. De plus, nous présentons les méthodes STATIS et STATIS dual car nous proposons d'utiliser l'ACIMO pour améliorer leur étape d'intrastructure.

Le chapitre trois regroupe des méthodes d'analyse de la co-inertie de plusieurs cubes ou multi-tableaux. Les méthodes étant si peu nombreuses, nous avons mis ensemble le cas «cubique» et le cas «multi-tableaux». La méthode cubique MultiWay Multi-Bloc covariates Component (MWMBCovC), issue de la littérature, est présentée. Celle-ci mène à la construction d'un modèle de type TUCKER ou PARAFAC par cube et d'un ensemble de super-composantes. Pour le cas de deux ensembles de tableaux mesurant les mêmes variables et non nécessairement sur les mêmes individus, nous décrivons STATICO. Puis, nous proposons deux nouvelles méthodes pour l'analyse de la co-inertie de deux ensembles de tableaux mesurés sur les mêmes individus. DOuble - Analyse Conjointe de Tableaux (DO-ACT), basée sur l'optimisation d'un critère sous contraintes, généralise en quelque sorte la méthode STATIS, encore connue sous le nom ACT. Elle présente l'avantage de se résoudre par une diagonalisation de matrice. La seconde méthode que nous proposons porte le nom d'Analyse de Co-Inertie Multiple Orthogonale Généralisée, ACIMOG. Elle étend, d'une certaine manière, l'ACIMO présentée au chapitre deux. Basée sur l'optimisation d'un critère sous contraintes à chaque étape, elle n'est pas équivalente à une diagonalisation et nécessite un algorithme itératif. Nous montrons que l'algorithme proposé est convergent. L'ACIMOG présente trois variantes, différentes à partir de la seconde étape, qui définissent les ACIMOG 1, 2 et 3. La première ACIMOG construit des ensembles de composantes orthogonales pour chaque tableaux. Les ACIMOG2 et 3 fournissent un ensemble de super-composantes orthogonales.

La régression entre cubes est abordée dans le quatrième chapitre. Nous ne proposons pas ici de nouvelle méthode. Deux des approches présentées sont basées sur des modèles de type PARAFAC ou TUCKER (régression sur composantes PARAFAC et MultiWay covariates Regression, MWcovR). Deux autres méthodes, qui s'avèrent être équivalentes, proposent des modèles trilineaires du cube prédicteur (Linear Three-way Decomposition, LTD, et Multilinear PLS, NPLS). Pour NPLS, nous donnons une démonstration des solutions, non donnée par son auteur. NPLS présente un inconvénient majeur : les composantes explicatives ne sont pas orthogonales. La dernière méthode présentée est la première méthode de régression cubique en date, Multiway PLS, MPLS. Elle est en fait équivalente à une régression PLS usuelle, et n'est donc pas réellement une méthode cubique.

Le cinquième chapitre est le plus important en volume. Ce chapitre, important à notre sens, doit être lu. Il concerne la régression linéaire multi-tableaux qui est le sujet central de notre travail. Nous présentons les méthodes existantes, qui sont toutes algorithmiques et en proposons trois nouvelles issues de l'optimisation sous contraintes d'un critère :

- L'Analyse de Co-Inertie Multiple Orthogonale - PLS (ACIMO-PLS) dont le but est de modéliser un ensemble de N tableaux de réponses par un tableau de prédicteurs. Il s'avère qu'elle peut être obtenue à partir d'une régression PLS usuelle entre deux tableaux.
- L'Analyse de Co-Inertie Multiple Orthogonale Généralisée - PLS (ACIMOG-PLS) dont le but est d'expliquer un ensemble de tableaux réponses par un ensemble de tableaux explicatifs. Elle possède trois variantes, désignées par ACIMOG-PLS1, 2, et 3. Chacune de ces variantes construit des super-composantes orthogonales ou des composantes orthogonales par tableau. L'ACIMOG-PLS présente de nombreuses propriétés similaires à celles de la régression PLS usuelle, qui sont démontrées dans la partie 5.6. Les solutions s'obtiennent par un algorithme itératif convergent.
- La REgression MUlti-Bloc (REMUB) a le même but que l'ACIMOG-PLS, mais utilise un critère différent. Les solutions s'obtiennent grâce à un algorithme itératif convergent. Cet algorithme revient à effectuer successivement une étape ACIMOG-PLS particulière et une étape PLS.

Dans ce chapitre, nous présentons l'Approche PLS, APLS, issue de la littérature, qui est une méthode ancienne mais qui semble être reconsidérée ces dernières années. Sont décrites également les méthodes de S. Wold, la PLS Hiérarchique, HPLS, et la PLS en Série, SPLS. Nous donnons une expression mathématique des coefficients des modèles de l'HPLS et nous discutons les algorithmes possibles pour SPLS dans le cas où il y a plus de deux tableaux prédicteurs. Ces deux méthodes présentent quelques imprécisions sur lesquelles nous pointons dans les parties 5.3 et 5.4. La dernière méthode issue de la littérature que nous présentons est la PLS Multi-Bloc, MBPLS. C'est une méthode très générale et beaucoup reprise ces dernières années. Pour le cas où il y a N tableaux réponses et K tableaux de prédicteurs, nous montrons qu'elle est équivalente à une succession de régressions PLS usuelles.

Il est important de noter que les méthodes présentées dans les chapitres 4 et 5 sont principalement issues de deux revues de chimométrie *Journal of Chemometrics* et *Chemometrics and Intelligent Laboratory Systems*. C'est peut-être l'une des raisons pour lesquelles elles sont peu connues des chercheurs en analyse sensorielle en France ou encore en écologie. De plus, les méthodes de régression multi-blocs sont généralement simplement basées sur des modifications de l'algorithme de la régression PLS pour l'adapter au cas de plusieurs tableaux. Ceci

engendre parfois l'équivalence de la méthode avec une régression PLS usuelle sur les tableaux conjoints (MPLS et un cas particulier de MBPLS). De plus, il paraît très régulièrement de nouvelles publications concernant la régression multi-blocs, plus particulièrement concernant des précisions, des petites modifications, des comparaisons, *etc*, au sujet de la MBPLS, ces publications étant dues à une «poignée» de chercheurs.

Le sixième et avant dernier chapitre de cette première partie est consacré à l'utilisation et la pratique des méthodes de régression multi-blocs : le pré-traitement des données, le choix du nombre de composantes, qui correspond le plus souvent au rang des modèles, et à une question non abordée dans la littérature qui est la sélection de tableaux prédicteurs «influentes» dans le modèle. Nous proposons plusieurs critères de sélection pour les méthodes ACIMOG-PLS1, 2, 3 et REMUB généralisant le critère VIP de la régression PLS ou utilisant le coefficient de corrélation vectorielle (RV). Nous rappelons dans ce chapitre ce qui est fait en régression PLS usuelle.

Pour terminer cette partie, le chapitre sept donne un résumé sous forme de tableaux des principales caractéristiques des méthodes présentées tout au long de cette partie et la démonstration de l'équivalence des méthodes MBPLS, PLS, ACIMO et REMUB lorsqu'il n'y a qu'un tableau de réponses à expliquer.

Les différents chapitres peuvent être lus dans n'importe quel ordre sauf peut-être le sixième chapitre et la première partie du septième chapitre dont la compréhension sera facilitée par la lecture du chapitre cinq concernant la régression multi-blocs.

La seconde partie de cette thèse est beaucoup moins importante en volume. Elle marque le début d'un travail sur la régression non-linéaire de type PLS de multi-tableaux. Nous n'avons pas trouvé de telles méthodes dans la littérature. Avec le développement de l'utilisation de méthodes multi-blocs, qui risque de se produire dans les années qui viennent, il est évident que des modèles non-linéaires vont être nécessaires pour certains. Le nombre de variables explicatives étant très important devant le nombre d'observations, (cas général de la régression multi-bloc), il est nécessaire de développer des modèles de type PLS. Dans cette problématique, il n'est pas vraiment envisageable d'utiliser les méthodes de modélisation linéaire basées sur des «expansions» des matrices de prédicteurs.

Ainsi, dans un premier chapitre, nous faisons une revue des méthodes de régression non-linéaire PLS pour deux tableaux, discutées en fin de chapitre.

Puis dans un second chapitre, nous proposons la méthode multi-tableaux Splines Additives en Régression Multi-Blocs, SARMUB. Elle est basée sur l'optimisation d'un critère sous contraintes à chaque étape en cherchant la meilleure transformation spline des prédicteurs, dans une base de B-splines. SARMUB est en quelque sorte une version non-linéaire de l'ACIMOG-PLS. De plus, nous proposons trois approches basées sur l'utilisation des méthodes linéaires ACIMO-PLS et ACIMOG-PLS.

Les méthodes de modélisation non-linéaires étant très peu nombreuses dans la littérature, nous ne leur avons pas consacré un chapitre. Cependant, dans l'introduction du second chapitre, on peut trouver quelques références pour des méthodes d'analyse conjointe de tableaux, dont deux très récentes datant d'août 2002.

Le développement de nouvelles approches en analyse de données n'a d'intérêt que si elles sont mises en oeuvres sur des données réelles. C'est pourquoi, la troisième partie de notre thèse regroupe cinq applications de nos méthodes multi-tableaux à des données réelles issues

de divers horizons.

Toutes les méthodes utilisées ont été programmées en langage S-Plus® pour la version 6.0 fonctionnant sous un système Linux. Certains de ces programmes sont complets et utilisables par un tiers, d'autres le sont moins. Nous en donnons une liste en introduction de cette partie.

Les premières données utilisées relèvent de la biométrie et sont issues de la littérature. Il s'agit d'étudier l'évolution des caractères morphologiques de fillettes de leur plus jeune âge à l'adolescence. Nous montrons l'utilité de l'ACIMO pour l'intrastructure de STATIS et de STATIS dual.

Dans le second chapitre nous présentons une application de l'ACIMO-PLS et l'ACIMO-PLSS à des données environnementales. Ces données nous ont été fournies par M. Simier et D. Nandris de l'IRD de Montpellier. Il s'agit d'expliquer des variables pathologiques mesurées sur des caféiers de Nouvelle Calédonie en fonction de variables de milieu.

Dans le troisième chapitre, nous utilisons l'ACIMOG-PLS1 pour expliquer l'abondance de la population de palourdes dans l'étang de Thau. Ce travail est le fruit d'une participation au Programme National - Environnement Côtier, Activité de Recherche n°4 (PNEC/ART4) au sein l'équipe de J.-P. Durbec (COM)*, N. Mazouni (CEPRALMAR), et J.-C. Gaertner (COM), R. Sabatier (UM1).

Le quatrième chapitre relève de la chimiométrie. Il s'agit d'une partie des résultats obtenus dans le cadre d'une collaboration avec J.-P. Biesse et B. Vidal du centre de recherche du groupe ALTADIS, Fleury-les-Aubrais. Nous appliquons les ACIMOG-PLS1, 2 et 3 et comparons les résultats avec ceux de la PLS usuelle, la MBPLS et HPLS.

La dernière application est la plus fournie. Il s'agit de données d'analyse sensorielle récoltées par F. Sune pour sa thèse (SUNE, 2001). Elles nous ont été gracieusement fournies par P. Lacroix, NESTLE France. Nous utilisons ici les méthodes DO-ACT et ACIMOG pour chercher une structure commune entre deux panels de consommateurs. Puis, nous utilisons les méthodes ACIMOG-PLS2, 3 et REMUB pour expliquer les réponses des consommateurs par les évaluations sensorielles d'experts.

Certaines applications ont été utilisées pour illustrer la présentation de l'ACIMOG-PLS lors de congrès. Les références seront données au fil des chapitres.

Pour conclure ce travail, nous ferons une discussion générale et donnons quelques orientations de recherche.

En annexe, nous rappelons les principales propriétés de la régression PLS usuelle pour qu'un lecteur non spécialiste puisse comprendre rapidement les références que nous faisons à PLS tout au long de cette thèse. Dans une seconde annexe, nous donnons la définition de la méthode des multiplicateurs de Lagrange que nous utilisons, puis un bref rappel sur les fonctions splines de régression qui sont utilisées dans la seconde partie. Enfin, une troisième annexe expose les «problèmes» liés aux représentations graphiques en régression PLS et en analyse multi-tableaux, dus à la perte du schéma de dualité présent en ACP.

* Centre Océanographique de Marseille

Première partie

Des modèles linéaires ...

Introduction

Cette première partie concerne la modélisation linéaire de multi-tableaux. Le terme «modélisation» fait ici référence à la fois aux méthodes factorielles permettant, par réduction de dimension, de reconstituer un ou plusieurs ensembles de tableaux de données, et aux méthodes factorielles de régression permettant d'expliquer un ensemble de tableaux par un autre. Tous les modèles présentés dans cette partie sont linéaires en les variables initiales.

Nous avons choisi de distinguer le cas où les tableaux de données sont cubiques de ceux où ils ne le sont pas. Un tableau est dit cubique lorsqu'il s'agit d'un tableau à plus de deux entrées. Par exemple, des mesures d'abondance de J différentes espèces en I lieux pendant K années. Si l'on est en présence de plusieurs tableaux mesurés sur les mêmes individus, alors on est en présence d'un ou plusieurs multi-tableaux (non cubique).

En France, les termes employés sont «analyse de co-inertie (multiple)» ou encore «analyse conjointe» de tableaux. Dans la littérature anglophone, on distingue le cas de tableaux parfaitement cubiques, en parlant de «multiway analysis», du cas d'un multi-tableau non cubique, désigné par «multiblock component analysis». Pourquoi «component»? Parce que les méthodes développées dans ces situations sont des méthodes générant des composantes, combinaisons linéaires des variables initiales, étape après étape. Les approches étant différentes selon le caractère cubique ou non des données, les méthodes font l'objet d'un chapitre différent (chapitres 1 et 2). Le chapitre 3 aborde le cas de l'analyse simultanée de plusieurs cubes et multi-tableaux.

Les chapitres 4 et 5 sont consacrés à la régression cubique ou «multiway regression» et multi-tableaux ou encore «multiblock regression».

Cette partie est constituée d'une importante revue de méthodes, les plus importantes. Dans le cas multi-blocs, nous avons trouvé intéressant de récapituler les méthodes existantes afin de mieux les comparer et essayer de mettre au clair leurs algorithmes qui ne sont pas toujours très explicites dans la littérature ou qui varient selon les auteurs. En effet, nous proposerons parfois des démonstrations omises par leur(s) auteur(s).

De plus, notre travail concernant plus particulièrement les méthodes multi-tableaux non cubiques, les chapitres 1 et 4 sont essentiellement constitués d'une revue de méthodes existantes, dans lesquels nous ne proposons pas de méthode nouvelle. Toutefois, nous donnons une démonstration pour l'obtention des solutions de la méthode de régression PLS multilinéaire (chap. 4.3) non fournie par son auteur.

Le chapitre 3 contient deux méthodes originales DOuble - ACT[†], qui est une sorte de généralisation de la méthode STATIS au cas de deux multi-tableaux dont on veut mesurer la proximité éventuelle, et l'Analyse de Co-Inertie Multiple Orthogonale Généralisée, dont le but

[†] Soumis pour publication : Vivien, M. & Sabatier, R. A generalization of STATIS-ACT strategy: DO-ACT for two multiblocks tables.

est similaire, mais très différente par sa mise en oeuvre.

Les chapitres 2 et 5 contiennent quant à eux une présentation des méthodes les plus utilisées en analyse multi-tableaux pour la régression ou non. Les méthodes du chapitre 2 sont toutes françaises. Nous y proposons une méthode originale, l'Analyse de Co-Inertie Multiple Orthogonale permettant de chercher la structure commune entre un multi-tableau et un tableau de référence. Nous avons appliqué cette méthode à l'intrastructure[‡] de STATIS afin d'obtenir des trajectoires optimales suivant un critère (2.5). Dans le chapitre 5, les méthodes de régression multi-tableaux sont essentiellement issues de la chimiométrie du nord de l'europe. Les méthodes utilisées diffèrent des techniques françaises en ce qu'elles sont principalement basées sur la modification de l'algorithme de la régression PLS. Nous proposons l'Analyse de Co-Inertie Multiple Orthogonale - PLS, l'ACIMOG-PLS et REgression MUlti-Blocs qui sont basées sur l'optimisation sous contraintes d'un critère de covariance. L'ACIMO-PLS a été présentée par communication orale à un congrès (VIVIEN & SABATIER, 2000) et a fait l'objet d'une publication (VIVIEN & SABATIER, 2001c). L'ACIMOG-PLS a été présentée à plusieurs congrès (VIVIEN & SABATIER, 2001a,b, 2002) et a été soumise pour publication[§]. De plus, nous l'avons appliquée à des données réelles en collaboration avec des organismes de recherches et industriels tels que l'IFREMER dans le cadre du PNEC, NESTLE France et ALTADIS. Ces applications sont présentées dans la troisième partie aux chapitres 3, 4 et 5. Parmi ces nouvelles méthodes, nous avons surtout insisté et développé celles concernant la régression : en effet, cette situation est de plus en plus rencontrée dans la pratique. Pour cela, le chapitre 6 propose un certain nombre d'éléments nouveaux concernant l'utilisation des méthodes multi-tableaux et plus particulièrement nos méthodes de régression multi-tableaux : une adaptation de la validation croisée pour le choix du nombre d'étapes à retenir, et des critères de sélection des blocs influents dans les modèles.

Enfin, nous concluons par le chapitre 7, dans lequel nous donnons un résumé des principales caractéristiques des méthodes exposées tout au long de cette partie sous forme de tableaux et comparons certaines méthodes vues au cours de cette partie.

Les chapitres 1 à 5 étant constitués des descriptions théoriques de méthodes, ils peuvent être lus dans n'importe quel ordre, bien qu'il y ait parfois des renvois entre eux. Par contre, la lecture des chapitres 6 et 7 nécessite celle du chapitre 5.

Le lecteur peu familier avec la régression PLS, qui est essentielle dans ce travail, trouvera en annexe A les définitions et principales propriétés. L'ouvrage de TENENHAUS (1998) est également une importante référence française.

[‡] Soumis pour publication : Vivien, M. & Sabatier, R. OMCIA: a solution to obtain optimal trajectories in (Dual) STATIS.

[§] Vivien, M. & Sabatier, R. Generalized Orthogonal Multiple Co-Inertia Analysis (-PLS) : new component and regression multiblock methods.

Chapitre 1

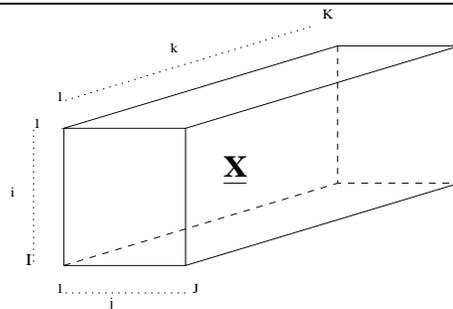
Analyse d'un cube ou Analyse 3 modes

Dans ce premier chapitre, nous faisons une description des méthodes d'analyse cubique parmi celles des plus connues. Ce n'est en aucun cas une liste exhaustive. Ces méthodes sont aussi connues sous les noms «ACP 3-modes» et «analyses multimodes» dans la littérature française et «multiway analysis» dans la littérature anglophone. Elles ont pour but de modéliser un cube de données, noté $\underline{\mathbf{X}}$. Le terme «cube», rencontré dans la littérature française, est utilisé par abus de langage et simplicité. Ce qui est appelé «cube» est en réalité une matrice de données à trois indices, c'est-à-dire un tableau à trois entrées (Fig. 1.1). Pour être tout à fait exact, il faudrait parler de parallélépipède rectangle.

Les trois dimensions (directions) du cube sont appelées les «trois modes» du cube. Le premier mode, la hauteur du cube, indice les lignes (les individus). Il est de dimension I . Le second mode, la largeur du cube, indice les colonnes (les variables). Il est de dimension J . Enfin, le troisième mode fait référence à la profondeur du cube, il est de dimension K .

Les méthodes cubiques cherchent à modéliser un tableau à trois entrées par un cube de rang plus petit. En fait, l'idée est la même que dans les méthodes d'analyse factorielle d'un tableau telles que l'Analyse en Composantes Principales (ACP). On va chercher des facteurs associés à chacun des trois modes, dont le produit tensoriel va donner une approximation du cube initial. Le tableau 1.1 précise les notations utilisées. Les modèles les plus anciens, tels que le modèle TUCKER3 (section 1.1), utilisent un cube appelé «matrice noyau», noté $\underline{\mathbf{G}}$ et de dimension

FIG. 1.1 – *Représentation d'un tableau à trois entrées*



TAB. 1.1 – Les notations utilisées

	mode 1	mode 2	mode 3
dimension de \underline{X}	I	J	K
indices courants de \underline{X}	i	j	k
nombres de facteurs cherchés	P	Q	R
indices courants	p	q	r
facteurs	$A = \{a_{i,p}\}_{i,p}$	$B = \{b_{j,q}\}_{j,q}$	$C = \{c_{k,r}\}_{k,r}$

$P \times Q \times R$, qu'il faut déterminer. Les modèles obtenus sont trinéaires en les facteurs et sont déterminés grâce à la minimisation de la somme des carrés des écarts du modèle, pour la plupart. Nous allons présenter les modèles et les algorithmes nécessaires à leur résolution pour les méthodes TUCKER3, PARAFAC/CANDECOMP et PARATUCK2. Les deux premiers modèles sont les plus anciens et sont très répandus. De plus, ils ont engendré un ensemble de modèles dérivés. Ceux-ci sont souvent vus comme des cas particuliers contraints des modèles TUCKER3 ou PARAFAC (ORTCP-A, ORTCP-B, ORTCP), d'autres en sont les versions développées pour les matrices de similarités ou de covariances, c'est-à-dire, plus généralement pour les matrices symétriques (INDSCAL, PARAFAC2, INDORT, TUCKER2, three-mode scaling). On peut trouver une description de ces méthodes dans les ouvrages (LAW, SNYDER, HATTIE & McDONALDS, 1984; KROONENBERG, 1983) et les relations entre elles dans (KIERS, 1988, 1991). La méthode PARATUCK2 est une méthode plus récente possédant l'unicité de PARAFAC et la flexibilité comme les modèles TUCKER2. Dans (KROONENBERG, 1992), on peut trouver une revue bibliographique concernant l'analyse cubique, complète jusqu'à cette date.

KIERS (2000b) propose des conventions de notations à partir de celles qui ont été préalablement utilisées dans la littérature. Les notations que nous utilisons correspondent à celles-ci. La même année, KIERS (2000a) propose aussi des moyens de représentations graphiques pour les modèles TUCKER3 et PARAFAC/CANDECOMP. Plus récemment, HARSHMAN & HONG (2002) résument les notations et les manières d'écrire les modèles de certaines méthodes cubiques, dont TUCKER3 et PARAFAC, à la façon «par tranche» ou «déplié». Nous essaierons de donner ces écritures dans ce chapitre.

1.1 Le modèle TUCKER3

Comme son nom l'indique, ce modèle, qui est le plus ancien, a été introduit par (TUCKER, 1963, 1964, 1966). Ici, les trois modes sont traités de la même façon. Pour cette raison, ce modèle est dit symétrique. Le chiffre 3 dans son nom rappelle que les trois modes sont réduits.

En désignant par x_{ijk} le terme du cube \underline{X} à l'intersection de la $i^{\text{ième}}$ ligne, de la $j^{\text{ième}}$ colonne et de la $k^{\text{ième}}$ tranche frontale, Tucker propose d'établir un modèle de la forme

$$\hat{x}_{ijk} = \sum_{p=1}^P \sum_{q=1}^Q \sum_{r=1}^R a_{i,p} b_{j,q} c_{k,r} g_{pqr} \quad \forall i, j, k$$

où le cube \underline{G} est défini par g_{pqr} , avec A, B, C orthonormales en colonnes. Pour en déterminer

les solutions il propose deux algorithmes (Méthode 1 et Méthode 2, §1.1.1 et §1.1.2). KROONENBERG & DE LEEUW (1980) présentent une écriture matricielle de ce modèle et proposent de minimiser la fonction de perte f (1.1) par un algorithme, TUCKALS3 (§1.1.3), de type Moindres Carrés Alternés (ALS, Alternating Least Squares) pour déterminer les solutions A , B , C orthonormées et \underline{G} .

$$f(A, B, C, \mathbf{G}_P) = \|\mathbf{X}_I - \widehat{\mathbf{X}}_I\|^2 = \|\mathbf{X}_I - A\mathbf{G}_P(B' \otimes C')\|^2, \quad (1.1)$$

où \otimes désigne le produit de Kronecker à droite, i.e. $A \otimes B = \{a_{ij}B\}_{i,j}$, \mathbf{G}_P est la matrice de dimension $P \times QR$ constituée des R tranches frontales juxtaposées de la matrice noyau \underline{G} recherchée et $\mathbf{X}_I = [X_1 \ \cdots \ X_K]$.

1.1.1 Méthode 1

Cet algorithme est le premier qui a été proposé par TUCKER (1966). Il consiste à déterminer les trois matrices A , B , et C indépendamment les unes des autres au moyen de trois ACP sur les trois tableaux suivants représentant le cube déplié :

- $\mathbf{X}_J = [X_1 \ \cdots \ X_I]$, $J \times KI$ où X_i est la $i^{\text{ième}}$ tranche horizontale de \underline{X} , de dimension $J \times K$,
- $\mathbf{X}_I = [X_1 \ \cdots \ X_K]$, $I \times JK$ où X_k est la $k^{\text{ième}}$ tranche frontale de \underline{X} , de dimension $I \times J$,
- $\mathbf{X}_K = [X_1 \ \cdots \ X_J]$, $K \times IJ$ où X_j est la $j^{\text{ième}}$ tranche latérale de \underline{X} , de dimension $K \times I$.

étape 1 : calcul de A

A est la matrice des P vecteurs propres normés associés aux P valeurs propres non nulles Λ_A de $\mathbf{X}_I\mathbf{X}'_I$: $\mathbf{X}_I\mathbf{X}'_IA = A\Lambda_A$.

étape 2 : calcul de B

B est la matrice des Q vecteurs propres normés associés aux Q valeurs propres non nulles Λ_B de $\mathbf{X}_J\mathbf{X}'_J$: $\mathbf{X}_J\mathbf{X}'_JB = B\Lambda_B$.

étape 3 : calcul de C

C est la matrice des R vecteurs propres normés associés aux R valeurs propres non nulles Λ_C de $\mathbf{X}_K\mathbf{X}'_K$: $\mathbf{X}_K\mathbf{X}'_KC = C\Lambda_C$.

étape 4 : calcul de \underline{G}

$$\forall p, q, r \quad g_{pqr} = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K a_{ip} b_{jq} c_{kr} x_{ijk}.$$

■

1.1.2 Méthode 2

Cet algorithme, encore proposé par TUCKER (1966), lie les trois problèmes aux valeurs propres pour le calcul de A , B et C .

étape 1 : calcul de A

A est la matrice des P vecteurs propres normés associés aux P valeurs propres non nulles Λ_A de $\mathbf{X}_I\mathbf{X}'_I$: $\mathbf{X}_I\mathbf{X}'_IA = A\Lambda_A$.

Remarque : même étape que l'étape 1 de la Méthode 1.

étape 2 : calcul de B

Soit $Z^1 = A'X_1 = [A'X_1 \ \cdots \ A'X_K]$, de dimension $P \times JK$. Z^1 définit un cube \underline{Z}^1 de dimension $P \times J \times K$ dont la $k^{\text{ième}}$ tranche frontale est $A'X_k$. Soit $Z_J^1 = [Z_1^1 \ \cdots \ Z_P^1]$, $J \times PK$, matrice conjointe des tranches horizontales de \underline{Z}^1 . B est la matrice des Q vecteurs propres normés associés aux Q valeurs propres non nulles Λ_B^Z de $Z_J^1 Z_J^{1'}$: $Z_J^1 Z_J^{1'} B = B \Lambda_B^Z$.

étape 3 : calcul de C

Soit $Z^2 = B'Z_Q^1 = [B'Z_1^1 \ \cdots \ B'Z_P^1]$, de dimension $Q \times PK$. Z^2 définit un cube \underline{Z}^2 de dimension $P \times Q \times K$ dont la $p^{\text{ième}}$ tranche horizontale est $B'Z_p^1$, de dimension $Q \times K$. Soit $Z_K^2 = [Z_1^2 \ \cdots \ Z_Q^2]$, juxtaposition des tranches latérales, $K \times QP$ de \underline{Z}^2 . C est la matrice des R vecteurs propres normés associés aux R valeurs propres non nulles Λ_C^Z de $Z_K^2 Z_K^{2'}$: $Z_K^2 Z_K^{2'} C = C \Lambda_C^Z$.

étape 4 : calcul de \underline{G}

Soit $Z^3 = Z_K^{2'} C$, de dimension $PQ \times R$.

\underline{G} , $P \times Q \times R$, est le cube défini par Z^3 , i.e. cube dont la $q^{\text{ième}}$ tranche latérale est $Z_q^{2'} C$, de dimension $P \times R$.

—■—

Cette méthode présente un inconvénient : l'erreur des approximations successives de \underline{G} , se transmet d'une étape à l'autre. L'élimination des plus petites valeurs propres à chaque étape modifie sensiblement \underline{G} par rapport à la méthode 1. D'un autre côté, cette méthode nécessite moins de calculs aux étapes 2 et 3.

1.1.3 TUCKALS3

Cet algorithme est celui qui est actuellement utilisé pour obtenir le modèle TUCKER3. Il a été proposé par KROONENBERG & DE LEEUW (1980). Les paramètres A , B , C et \underline{G} sont estimés par les moindres carrés alternés. Les auteurs commencent par montrer que le cube \underline{G} solution du problème de minimisation de f (1.1), peut s'exprimer uniquement suivant les termes A , B , C et X_1 : $G_P = A'X_1(BB' \otimes CC')$, où G_P et X_1 désignent la juxtaposition des tranches frontales de \underline{G} et \underline{X} . Ils ramènent ainsi la minimisation de f à la minimisation d'une nouvelle fonction g en A , B et C sous les contraintes d'orthonormalité colonne de ces matrices :

$$g(A,B,C) = \|X_1 - AA'X_1(BB' \otimes CC')\|^2. \quad (1.2)$$

Puis, ils montrent que l'on peut se ramener au problème de maximisation de p sous les mêmes contraintes :

$$p(A,B,C) = tr(A'X_1(BB' \otimes CC')X_1 A) \quad (1.3)$$

$$= tr(X_1 X_1') - g(A,B,C). \quad (1.4)$$

Cette maximisation définit l'algorithme TUCKALS3. La publication (KROONENBERG & DE LEEUW, 1980) contient les démonstrations de l'existence et de la nature de la solution et de la convergence de l'algorithme vers le maximum (global) de p . La première version de TUCKALS3 (KROONENBERG & DE LEEUW, 1980) revient à résoudre trois problèmes aux valeurs propres simultanément et itérativement, en utilisant seulement la première étape de la méthode de Bauer-Rutishauser (BR) (RUTISHAUSER, 1969; SCHWARTZ, RUTISHAUSER &

STIEFEL, 1968). KROONENBERG, TEN BERGE, BROUWER & KIERS (1989) proposent d'utiliser la méthode d'orthogonalisation de Gram-Schmidt (GR) au lieu de la méthode BR afin de diminuer le temps de calcul et de convergence de l'algorithme. KIERS, KROONENBERG & TEN BERGE (1992) améliorent encore la vitesse de l'algorithme en proposant une ré-organisation des calculs. Enfin, (ANDERSSON & BRO, 1998) proposent plusieurs façons d'accroître la rapidité de l'algorithme ainsi que sa précision, en utilisant une étape NIPALS plutôt que BR ou GS, ou en répétant plusieurs fois les mises à jour BR et GS dans une étape.

L'algorithme que nous présentons ici, résume ceux de (KROONENBERG & DE LEEUW, 1980) et (KROONENBERG et al., 1989).

L'indice a est le compteur d'une itération.

étape 0a : initialisation

Utiliser la méthode 1 (§1.1.1) pour initialiser A , B et C . Ces initialisations sont notées A_0 , B_0 et C_0 .

étape 0b

$$a = a + 1$$

étape 1 : calcul de A_a

Soit $P_a = \mathbf{X}_J(B_a B_a' \otimes C_a C_a') \mathbf{X}_J'$, où \mathbf{X}_J est de taille $J \times Kl$.

1 étape BR : $A_{a+1} = P_a A_a (A_a' P_a A_a)^{-1/2}$.

ou 1 étape GS : $A_{a+1} = GS(P_a A_a)$.

étape 2 : calcul de B_a

Soit $Q_a = \mathbf{X}_I(C_a C_a' \otimes A_a A_a') \mathbf{X}_I'$, où \mathbf{X}_I est de taille $l \times JK$.

1 étape BR : $B_{a+1} = Q_a B_a (B_a' Q_a B_a)^{-1/2}$.

ou 1 étape GS : $B_{a+1} = GS(Q_a B_a)$.

étape 3 : calcul de C_a

$R_a = \mathbf{X}_K(A_a A_a' \otimes B_a B_a') \mathbf{X}_K'$, où \mathbf{X}_K est de taille $K \times lJ$.

1 étape BR : $C_{a+1} = R_a C_a (C_a' R_a C_a)^{-1/2}$.

ou 1 étape GS : $C_{a+1} = GS(R_a C_a)$.

étape 4 : test de la convergence

Si $p(A_{a+1}, B_{a+1}, C_{a+1}) - p(A_a, B_a, C_a) < \epsilon$, $\|A_{a+1} - A_a\| < \epsilon$, $\|B_{a+1} - B_a\| < \epsilon$, $\|C_{a+1} - C_a\| < \epsilon$ alors on passe à l'étape suivante.

Sinon, on recommence à l'étape 0b.

étape calcul de \underline{G}

$$A = A_a, B = B_a, C = C_a.$$

$$\mathbf{G}_P = A' \mathbf{X}_I (C \otimes B).$$

■

1.1.4 Commentaires

Les matrices A , B , C sont contraintes à être orthonormales en colonne sans perte de généralité (KROONENBERG & DE LEEUW, 1980). Cela permet de simplifier l'algorithme de résolution du problème.

Les facteurs ne sont pas emboîtés : a_1 n'est *a priori* pas le même que l'on ait $P = 2$ ou $P = 3$, par exemple.

Les solutions (les facteurs) ne pas sont uniques, dans le sens où il est possible de faire des rotations des facteurs A , B et C sans changer l'ajustement du modèle $\widehat{\mathbf{X}}_1 = \mathbf{A}\mathbf{G}_P(\mathbf{B}' \times \mathbf{C}')$.

La matrice noyau $\underline{\mathbf{G}}$ indique à quel point sont proches les différents facteurs des différents modes. (Pour plus d'explication sur son interprétation, voir (KROONENBERG, 1984)). Chaque élément g_{pqr} est associé à un triplet de facteurs (a_p, b_q, c_r) et mesure la part relative de la structure ternaire symbolisée par le triplet dans la reconstitution de $\underline{\mathbf{X}}$.

Les rotations des facteurs ont pour but de permettre une interprétation plus facile des résultats. Or, ce qui est le plus difficile à interpréter, c'est la matrice noyau. En général, on préfère «simplifier» la matrice noyau (*i.e.* faire en sorte qu'il y ait un maximum de 1 et de 0 dans la matrice noyau) plutôt qu'avoir des composantes orthogonales (KROONENBERG, 1983, 1984).

L'algorithme TUCKALS3 est préférable car il possède la propriété d'optimalité de minimisation de l'erreur que n'ont pas les méthodes de TUCKER (1966).

TEN BERGE, DE LEEUW & KROONENBERG (1987) ont proposé une version de TUCKALS3 où les contraintes d'orthogonalité des composantes sont supprimées.

Dans le cas d'une décomposition complète de $\underline{\mathbf{X}}$ qui correspond au cas où $I = P$, $J = Q$, $K = R$, le tableau est entièrement reconstitué. Si le nombre de valeurs propres conservées est inférieur à la dimension du mode correspondant, ne serait-ce que sur un des trois modes, alors il y a une erreur d'estimation. Lorsque la décomposition est complète, les estimations des valeurs g_{pqr} de la matrice noyau sont du type Moindres Carrés.

On a la relation $p(A,B,C) = \|\mathbf{G}_P\|^2 = tr(\mathbf{G}_P\mathbf{G}'_P)$ à l'optimum : la norme de \mathbf{G}_P est maximisée.

FRANC (1992) a traduit le modèle TUCKER3 en écriture tensorielle et inclut sa résolution dans des problèmes d'optimisation tensorielle sous contraintes plus vastes. Ainsi le modèle TUCKER3 peut s'écrire

$$\widehat{\mathbf{X}} = \sum_{p=1}^P \sum_{q=1}^Q \sum_{r=1}^R g_{pqr} a_p \otimes b_q \otimes c_r.$$

\otimes désigne ici le produit tensoriel. Il propose aussi des modèles avec métriques dans les espaces \mathbb{R}^I , \mathbb{R}^J , \mathbb{R}^K et des généralisations à des modèles pour des tableaux à plus de trois entrées. De telles généralisations ont, en autres, aussi été faites et utilisées par HENRION (1994, 1998).

Le modèle TUCKER3 a engendré plusieurs modèles dérivés : «three-mode scaling» est l'application du modèle TUCKERS3 à K matrices de similarités. Le modèle TUCKER2 (ISRAELSSON, 1969; KROONENBERG, 1994) est le modèle TUCKER3 ne réduisant pas un des modes, par exemple le premier mode. Cela signifie qu'il s'applique à un cube dont les tranches

frontales sont symétriques. Ces méthodes sont décrites dans (KROONENBERG, 1983). KIERS (1991) donne une description hiérarchique de ces modèles. Le modèle TUCKER2 s'écrit $\hat{\mathbf{X}}_l = \mathbf{H}_l(C \otimes B)'$, où \mathbf{H}_l est la juxtaposition des tranches frontales de la matrice noyau, recherchée, \mathbf{H} de dimension $l \times Q \times R$. C'est le modèle TUCKER3 où le premier mode est non réduit ou bien où $P = l$ composantes sont choisies. De la même façon, le modèle TUCKER1 (TUCKER & MESSICK, 1963) est défini par un modèle TUCKER3 où les deux modes sont entièrement reconstitués, par exemple les deux derniers, $\hat{\mathbf{X}}_l = A\mathbf{H}_p$, où \mathbf{H}_p est la juxtaposition des tranches frontales de la matrice noyau \mathbf{H} de dimension $P \times J \times K$. Il est équivalent à l'ACP du tableau \mathbf{X}_l .

1.2 Le modèle PARAFAC1/CANDECOMP

PARAFAC est l'abréviation de *PARAllel FACTor analysis*. Il a été inspiré par le principe des Parallel Proportional Profiles de (CATTELL, 1944) (cas de deux tableaux). Ce modèle a été proposé la même année par HARSHMAN (1970) et CARROLL & CHANG (1970), ces derniers l'ayant appelé CANDECOMP, CANonical DECOMPosition. Le modèle PARAFAC1/CANDECOMP est en quelques sorte une généralisation le l'ACP.

Le modèle s'écrit :

$$\hat{x}_{ijk} = \sum_{r=1}^R a_{i,r} b_{j,r} c_{k,r} \quad \forall i, j, k$$

ou matriciellement :

$$\hat{\mathbf{X}}_k = A D_k B',$$

où D_k est la matrice diagonale de la $k^{\text{ième}}$ ligne de C ; A , B et C étant les matrices de R facteurs associés à chacun des modes de \mathbf{X} . BRO (1998) propose une troisième écriture utilisant le produit matriciel de Khatri-Rao (RAO & MITRA, 1971) :

$$\hat{\mathbf{X}}_l = A(C \odot B)',$$

où \odot est produit de Khatri-Rao : si $X = [x_1 \ \cdots \ x_l]$ et $Y = [y_1 \ \cdots \ y_J]$ alors

$$X \odot Y = [x_1 \otimes y_1 \ \cdots \ x_J \otimes y_J] = [vec(y_1 x_1') \ \cdots \ vec(y_J x_J')].$$

Les facteurs A , B et C , sont déterminés par la minimisation de :

$$f(A, B, C) = \|\mathbf{X}_l - A(C \odot B)'\|^2. \quad (1.5)$$

\mathbf{X}_l est ici le tableau conjoint, $l \times JK$, des K matrices X_k de dimension $l \times J$.

La particularité de PARAFAC est qu'il n'y a pas besoin de contraintes pour déterminer les facteurs. Pour obtenir des bases des sous-espaces de dimension R dans \mathbb{R}^l , \mathbb{R}^J et \mathbb{R}^K , on peut procéder à une orthogonalisation des facteurs. L'écriture 1.5 de f permet de déterminer simplement A . f peut s'écrire de deux autres manières, permettant le calcul de B et C :

$$f(A, B, C) = \|\mathbf{X}_J - B(A \odot C)'\|^2 \quad (1.6)$$

$$= \|\mathbf{X}_K - C(B \odot A)'\|^2 \quad (1.7)$$

Il est possible de montrer que $(C \odot B)'(C \odot B) = (BB') * (CC')$ (et des expressions analogues pour $A \odot C$ et $B \odot A$) où $*$ dénote le produit de Hadamard, c'est-à-dire le produit terme à

terme.

Pour déterminer les solutions numériques, un algorithme de type ALS est utilisé :

étape 0 : initialisation

Initialiser les matrices B et C .

étape 1 : calcul de A

A est la solution de la minimisation aux moindres carrés de $f_{BC}(A) = \|X_I - A(C \odot B)'\|^2$:

$$A = X_I(C \otimes B) ((BB') * (CC'))^+.$$

étape 2 : calcul de B

B est la solution de la minimisation aux moindres carrés de $f_{AC}(B) = \|X_J - B(A \odot C)'\|^2$:

$$B = X_J(A \odot C) ((CC') * (AA'))^+.$$

étape 3 : calcul de C

C est la solution de la minimisation aux moindres carrés de $f_{AB}(C) = \|X_K - C(B \odot A)'\|^2$:

$$C = X_K(B \odot A) ((AA') * (BB'))^+.$$

étape 4 : convergence

Tant que $f(A,B,C)$ n'est pas stable, retourner à l'étape 1.

■

HARSHMAN & LUNDY (1984) proposent d'initialiser l'algorithme avec des valeurs aléatoires. Si les mêmes solutions sont trouvées suite à plusieurs initialisations différentes, on peut penser que le minimum est obtenu.

1.2.1 Commentaires

Dans la littérature, ce modèle est désigné de manière équivalente par PARAFAC ou PARAFAC1.

C'est un modèle symétrique : les trois modes sont traités de la même manière. Le même modèle sera obtenu si on permute les dimensions du cube \underline{X} .

Les facteurs ainsi obtenus ne sont pas orthogonaux. Cela rend l'interprétation plus compliquée. Il est possible d'orthogonaliser A , B et C après convergence de l'algorithme, par la méthode de Gram-Schmidt ou de d'utiliser des contraintes d'orthogonalité lors de la minimisation de f . Pour la contrainte sur A , par exemple, A est actualisé lors de l'étape 1 de l'algorithme par $X(C \odot B)' ((C \odot B)X'X(C \odot B))^{-1/2}$ (HARSHMAN, 1972a; HARSHMAN & LUNDY, 1994; BRO, 1997). A est alors orthonormalisé. HARSHMAN & LUNDY (1984) proposent d'orthonormaliser les facteurs A et B pour les situations où la méthode PARAFAC1/CANDECOMP sans contraintes mène à des solutions «dégénérées», avec de fortes corrélations entre les colonnes de A . Si on impose $A'A = Id_{\mathbb{R}}$ (*resp.* $B'B = Id_{\mathbb{R}}$, ou les deux) on obtient les modèles appelés ORTCP-A (*resp.* ORTCP-B, ORTCP), (KIERS, 1991).

Les solutions ne sont pas emboîtées. Cependant, elles sont uniques au sens où une rotation des axes engendre une perte de l'ajustement et les solutions sont entièrement déterminées sans contraintes. En vérité, les solutions sont en général uniques. On montre cette unicité à l'aide

de propriétés sur le rang des tenseurs d'ordre 3 (KRUSKAL, 1977, 1989). Il est montré que si les trois matrices de composantes A , B et C sont de plein rang colonne (*i.e.* \mathbb{R}) alors les solutions sont uniques («la décomposition est unique»). Cette condition est en général vérifiée car on prend le plus souvent $\mathbb{R} = 1, 2$ ou 3 (HARSHMAN, 1970; KRUSKAL, 1977). Mais, il existe des conditions plus souples montrant cette unicité (dues au théorème de Jennrich) (HARSHMAN, 1972b; LEURGANS, ROSS & ABEL, 1993; HARSHMAN & LUNDY, 1984).

Dans (KIERS & KRIJNEN, 1991) on peut trouver une version de l'algorithme «accélééré» pour le cas où l est beaucoup plus grand que \mathbb{JK} , ainsi que pour le cas où on impose $A'A = Id_{\mathbb{R}}$. BRO (1997, 1998) propose des techniques de compression et d'extrapolation pour accélérer l'algorithme lorsque les dimensions des différents modes sont très grandes.

Pour des explications précises en ce qui concerne l'interprétation, le prétraitement des données, et des informations générales, on peut consulter (HARSHMAN & LUNDY, 1984, 1994).

FRANC (1992) a réécrit le problème à l'aide du produit tensoriel de vecteurs. Le modèle est donné par $\hat{\underline{X}} = \sum_{r=1}^{\mathbb{R}} a_r \otimes b_r \otimes c_r$.

D'autres méthodes sont dérivées de PARAFAC. Le modèle PARAFAC2 (HARSHMAN, 1972b), est le modèle PARAFAC adapté au cas où les variables ne sont pas mesurées sur les mêmes individus. Le modèle est construit à partir des matrices de variances-covariances des tranches frontales. KIERS (1993) propose un algorithme basé sur les moindres carrés alternés pour le résoudre et KIERS, TEN BERGE & BRO (1999) proposent un algorithme modélisant directement les tranches frontales X_k , évitant ainsi de lourds calculs utilisant les matrices de variances-covariances. Le modèle s'écrit (BRO, 1998) $\hat{Y}_{\mathbb{K}} = (C' \odot C')' \text{diag}(\text{vec}(H))(A \otimes A)'$, où $Y_k = X_k X_k'$, $l \times l$. La matrice $H = B'B$, $\mathbb{R} \times \mathbb{R}$, fait partie des inconnues. Au lieu d'estimer B (composantes PARAFAC du second mode de \underline{X} , H est estimée.

Si le modèle PARAFAC2 est utilisé sur des matrices carrées non symétriques, il s'agit du modèle DEDICOM, DÉcomposition into DIRECTIONAL COMponent, dû à (HARSHMAN, 1978; HARSHMAN, GREEN, WIND & LUNDY, 1982) d'après HARSHMAN & LUNDY (1984); KIERS (1993), ce dernier proposant un algorithme. La méthode INDSCAL, INDividual difference SCALing, (CARROLL & CHANG, 1970) est l'application de CANDECOMP/PARAFAC appliqué à un ensemble de \mathbb{K} matrices de similarités. TEN BERGE & KIERS (1991b) apportent quelques clarifications sur les résultats énoncés mais non démontrés jusqu'alors, qui étaient justement l'équivalence de l'application de CANDECOMP aux matrices de similarités et du problème de minimisation de INDSCAL initial. Lorsqu'on contraint les matrices de facteurs B à être orthonormales dans INDSCAL, on obtient la méthode INDORT, ORTHogonally constrained INDSCAL (KIERS, 1991).

1.3 Le modèle PARATUCK2

Ce modèle a été proposé par (HARSHMAN & LUNDY, 1996) comme généralisation du modèle PARAFAC et TUCKER2. En effet, ce modèle garde des propriétés d'unicité du modèle PARAFAC, et garde une certaine flexibilité, comme les modèles de TUCKER2. D'où le nom PARATUCK2. La réduction se fait sur les deux premiers modes, et pas sur le troisième,

comme dans PARAFAC2 et TUCKER2.

Il a été proposé dans le but d'ajuster des données cubiques pour lesquelles le modèle PARAFAC2 n'est pas satisfaisant, car trop peu flexible. Ainsi, une matrice particulière (ci-dessous $D_k^A H D_k^B$) tient compte des liaisons entre modes d'une manière plus restrictive que la matrice noyau des modèles TUCKER2.

Jusqu'à ce que (BRO, 1998) le réalise dans sa thèse, aucune publications de l'algorithme et d'applications n'avaient été faites.

Le modèle s'écrit :

$$\hat{X}_k = A D_k^A H D_k^B B', \quad \forall k = 1, \dots, K$$

avec :

A matrice des composantes du premier mode, de taille $I \times P$,

B matrice des composantes du second mode, de taille $J \times R$,

D_k^A matrice diagonale, de taille $P \times P$,

D_k^B matrice diagonale, de taille $R \times R$,

H matrice de taille $P \times R$,

C^A matrice dont la $k^{\text{ième}}$ ligne est la diagonale de D_k^A , $K \times P$,

C^B matrice dont la $k^{\text{ième}}$ ligne est la diagonale de D_k^B , $K \times R$.

La matrice H autorise un nombre de composantes différents dans les deux modes, $P \neq R$ et la matrice $D_k^A H D_k^B$ ($P \times R$) peut être considérée comme une matrice d'interactions, matrice noyau. De ces deux points de vue, ce modèle ressemble au modèle TUCKER2.

HARSHMAN & LUNDY (1996) montrent, que grâce à la structure restrictive du «noyau» de PARATUCK2, l'unicité du modèle est conservée, comme dans les modèles PARAFAC. Ils donnent une preuve expérimentale de l'unicité pour des modèles avec $P \neq R$, mais traitent en détail le cas où $P = R$.

Ils montrent que le modèle sera la plupart du temps unique si les matrices de composantes A et B du modèle sont de plein rang et que les dimensions du tableau ne sont pas trop petites.

BRO (1998) montre que ce modèle peut s'écrire à l'aide du produit de Khatri-Rao :

$$\hat{X}_K = ((C^B)' \odot (C^A)')' \text{diag}(\text{vec}(H))(B \otimes A)',$$

où $\text{vec}(H)$ est la vectorialisation de H et $\text{diag}(\text{vec}(H))$ la matrice diagonale dont la diagonale est $\text{vec}(H)$. \otimes désigne le produit de Kronecker.

BRO (1998) propose un algorithme aux moindres carrés alternés. La fonction de perte f suivante est minimisée par rapport aux matrices de paramètres A , $I \times P$, B , $J \times R$, H , $P \times R$ (matrice «d'interactions»), et D_k^A et D_k^B diagonales de taille respectives $P \times P$ et $R \times R$:

$$f(A, B, H, D_k^A, D_k^B) = \sum_{k=1}^K \|X_k - A D_k^A H D_k^B B'\|^2.$$

étape 0 : initialisation

A , $I \times P$, matrice des P premières composantes principales de l'ACP de X_I .

(C^A) , $K \times P$, matrice des P premières composantes principales de l'ACP de X_K , et

C^B , $K \times R$, les S premières composantes de principales de X_K .

B , $J \times R$, matrice des R composantes principales de X_J .

H , $P \times R$, est prise aléatoirement si $P \neq Q$ sinon $H = Id_R$.

étape 1 : mise à jour de A

$$A = \mathbf{x}_1(F_A^+)' \text{ où } F_A = [D_1^A H D_1^B B' \quad \dots \quad D_K^A H D_K^B B'].$$

étape 2 : mise à jour de C^A

$$\text{diag}(D_k^A) = ((F_k \odot A)^+ \text{vec}(X_k))', \text{ où } \text{diag}(D_k^A) \text{ est le vecteur contenant la diagonale de } D_k^A \text{ c'est-à-dire la } k^{\text{ième}} \text{ ligne de } C^A, \text{ et } F_k = B D_k^B H'.$$

étape 3 : mise à jour de H

$$H = Z^+ \text{vec}(\mathbf{x}_1) \text{ où } Z = \begin{bmatrix} B D_1^B \otimes A D_1^A \\ \vdots \\ B D_K^B \otimes A D_K^A \end{bmatrix}.$$

étape 4 : mise à jour de B

$$B = [X_1' \quad \dots \quad X_K'] (F_B^+)' \text{ où } F_B = \begin{bmatrix} A D_1^A H D_1^B \\ \vdots \\ A D_K^A H D_K^B \end{bmatrix}.$$

étape 5 : mise à jour de C^B

$$\text{diag}(D_k^B) = ((B \odot Z_k)^+ \text{vec}(X_k))', \text{ où } Z_k = A D_k^A H \text{ et } \text{diag}(D_k^B) \text{ la } k^{\text{ième}} \text{ ligne de } C^B.$$

étape 6 : test de la convergence

Si f a convergé, s'arrêter, sinon repartir à l'étape 1.

—■—

Chapitre 2

Analyse de la Co-inertie de tableaux

Dans le chapitre précédent nous avons décrit les méthodes de base pour l'analyse d'un cube. En pratique, il arrive que les données ne soient pas «cubiques». Pour analyser de tels jeux de données, il faut utiliser des méthodes d'analyse conjointe de tableaux.

Les méthodes sont nombreuses pour cette situation. Les méthodes procustes généralisées (GOWER, 1975; QANNARI, COURCOUX, LEJEUNE & MAYSTRE, 1997), dont nous ne parlerons pas ici, recherchent les meilleures rotations (ou réflexions) permettant de rapprocher les différents tableaux. Les diverses généralisations de l'analyse canonique (CARROLL, 1968; HORST, 1961), voir aussi (SABATIER, 1993), sont basées sur des critères de corrélation dont les limites sont connues. Citons aussi les différentes ACP généralisées (CASIN, 1996, 2001) ou particulières (ESCOFIER & PAGÈS, 1984, 1986).

Nous avons choisi de présenter deux méthodes, l'Analyse de la Concordance, (ACO) et l'Analyse de CO-inertie Multiple (ACOM) qui sont basées sur l'optimisation d'un critère de covariances au carré, rappelant celui de la régression PLS. Sur ce même type de critère nous avons développé l'Analyse de Co-Inertie Multiple Orthogonale, ACIMO. L'ACIMO doit son nom à l'orthogonalité de ses composantes, contrairement à l'ACOM et à l'ACO. La méthode STATIS (dual) est aussi présente dans ce chapitre car nous avons utilisé l'ACIMO pour améliorer l'étape de l'intrastructure de STATIS (dual).

2.1 STATIS et STATIS dual

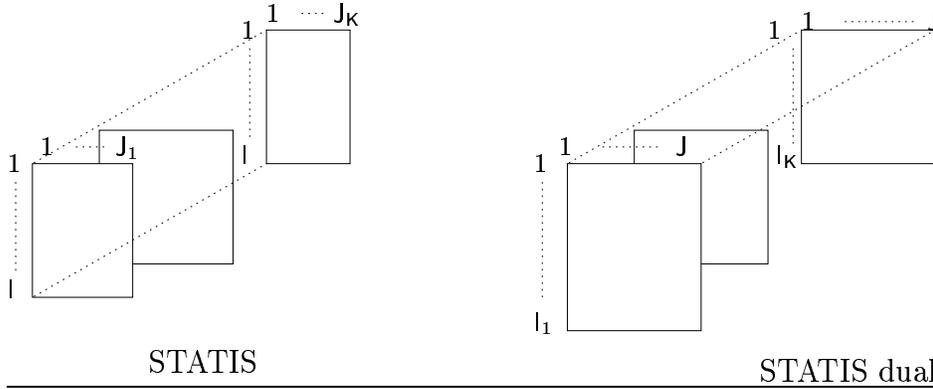
La méthode STATIS, Structuration de Tableaux A Trois Indices de la Statistique, à été introduite par L'HERMIER DES PLANTES (1976), puis développée par plusieurs auteurs, en particulier par LAVIT (1988). Elle est aussi connue sous le nom de ACT, Analyse Conjointe de Tableaux. L'idée essentielle est la recherche d'une structure commune entre K tableaux $\{X_k\}_k$ mesurés sur les mêmes l individus.

STATIS dual est une méthode analogue à STATIS adaptée au cas où les K tableaux sont appariés en colonnes, c'est-à-dire, qu'ils mesurent les mêmes J variables. Dans ce cas les individus ne sont pas forcément les mêmes (Fig. 2.1).

Les deux méthodes sont utilisables si les données sont cubiques, mais ne fournissent pas *a priori* les mêmes résultats.

A chaque tableau X_k est associé une métrique Q_k^x dans l'espace des individus, et une métrique D_k dans l'espace des variables. Lorsque la dimension commune aux K tableaux est le nombre

FIG. 2.1 – *Forme des données pour lesquelles STATIS et STATIS dual sont applicables*



l d'observations, $D_k = D, l \times l$, pour tout k . En général D est choisie comme étant la métrique uniforme.

Lorsque la dimension commune est le nombre J de variables, $Q_k^x = Id_J, J \times J$.

De plus, les tableaux sont D -centrés (ou D_k -centrés) en général, et tout au long de cette section.

Les tableaux peuvent être pondérés par un poids positif π_k . Nous notons Π la matrice diagonale de ces poids. STATIS et STATIS dual se déroulent toutes deux selon trois mêmes étapes, qui sont :

L'interstructure : un objet représentatif de chaque tableau X_k est choisi. Puis les relations entre les différents objets sont étudiées : c'est une comparaison globale de la structure des K tableaux de données.

Le compromis : c'est la recherche d'un objet résumant les K tableaux. Il devra être représentatif des tableaux selon un critère. Il est de même nature que les objets représentatifs.

L'intrastructure : c'est l'étude des structures internes des tableaux.

2.1.1 Etude de l'interstructure

Cette étape nécessite le choix d'un objet représentatif de chaque tableau et d'un produit scalaire afin de définir une distance entre eux.

STATIS :

Lorsque tous les tableaux ont le même nombre l d'individus, les objets utilisés sont les opérateurs $W_k^x = X_k Q_k^x X_k'$, qui peuvent être normés si besoin est : c'est la matrice des produits scalaires entre les individus du tableau k . Il représente un tableau des proximités entre les individus.

La distance entre deux de ces objets est définie à partir du produit scalaire de Hilbert-Schmidt :

$$(W_k^x | W_l^x)_{H.S} = tr(W_k^x D W_l^x D) \quad k, l = 1 \dots K$$

$$\|W_k^x\|_{H.S}^2 = tr(W_k^x D W_k^x D) = \sum_{i=1}^l \lambda_i^2(W_k^x D) \quad \text{pour } k = 1 \dots K$$

$$d^2(W_k^x, W_l^x) = \|W_k^x - W_l^x\|_{H.S}^2 \quad \text{pour } k, l = 1 \dots K,$$

où $\lambda_i(W_k^x D)$ désigne la $i^{\text{ième}}$ valeur propre de $W_k^x D$.

Une image euclidienne des tableaux est alors construite grâce à ces distances.

Si \mathcal{C} , $\mathbf{K} \times \mathbf{K}$, est la matrice des produits scalaires entre objets définie plus haut, la diagonalisation de $\mathcal{C}\Pi$ donne les axes permettant de construire cette image :

$$\mathcal{C}\Pi = P\Sigma P', \text{ avec } P'P = Id_r$$

où P , $\mathbf{K} \times r$, est la matrice dont les colonnes sont les vecteurs propres normés $\{p_a\}_{a=1,\dots,r}$ de \mathcal{C} , où r est le rang de \mathcal{C} , et Σ , $r \times r$, est la matrice diagonale de ses valeurs propres σ_a .

Si $\{M_k, k = 1, \dots, \mathbf{K}\}$ est le nuage des points-tableaux, les coordonnées des points M_k sur l'axe a sont les composantes du vecteur $\sqrt{\sigma_a}p_a = \frac{1}{\sqrt{\sigma_a}}\mathcal{C}\Pi p_a$ et le cosinus de l'angle entre les vecteurs OM_k et OM_l représente l'approximation du produit scalaire normé, (*i.e.* le coefficient RV) entre les opérateurs W_k^x et W_l^x .

Rappel: Le coefficient RV , corrélation vectorielle entre deux études (X_k, Q_k^x, D) et (X_l, Q_l^x, D) , introduit par ROBERT & ESCOUFIER (1976), mesure la proximité entre deux opérateurs au sens de Hilbert-Schmidt :

$$RV(X_k, X_l) = \left(\frac{W_k^x}{\|W_k^x\|_{H.S.}} \middle| \frac{W_l^x}{\|W_l^x\|_{H.S.}} \right)_{H.S.}.$$

Si $RV(X_k, X_l) = 0$, alors toutes les variables de X_k ont une covariance nulle avec celles de X_l .

Si $RV(X_k, X_l) = 1$ les études sont équivalentes, dans le sens où l'image euclidienne de l'une est proportionnelle à celle de l'autre.

Remarque: Si l'on travaille avec $\{W_k^x / \|W_k^x\|_{H.S.}\}_k$, la matrice de l'interstructure \mathcal{C} est la matrice des coefficients RV .

Un théorème (voir LAVIT, ESCOUFIER, SABATIER & TRAISSAC, 1994) dit que la matrice \mathcal{C} admet un premier vecteur propre dont toutes les composantes sont de même signe. Cela signifie que les points-tableaux ne se distinguent vraiment que sur l'axe 2.

Par la suite, nous désignerons par $p_1 = (\alpha_1, \dots, \alpha_{\mathbf{K}})'$ le vecteur propre associé à σ_1 dont toutes les coordonnées sont choisies positives. p_1 intervient dans la construction du compromis.

STATIS dual :

Lorsque tous les tableaux sont appariés en colonnes, les objets représentatifs sont les opérateurs $V_k^x = X_k' D X_k$, matrices de produits scalaires entre les variables, qui permettent de mesurer la proximité entre les variables d'un même tableau. Lorsque les tableaux sont centrés et réduits selon la métrique D , V_k^x est la matrice de corrélation entre les variables du tableau X_k .

Comme dans STATIS, la matrice des produits scalaires S , au sens de Hilbert-Schmidt (défini précédemment), des objets V_k^x , normés ou non, va être diagonalisée pour fournir une image euclidienne des tableaux :

$$S = (tr(V_k^x V_{k'}^x))_{k,k'}, \quad \forall k, k';$$

ou

$$S = \left(tr \left(\frac{V_k^x}{\|V_k^x\|_{H.S.}} \frac{V_{k'}^x}{\|V_{k'}^x\|_{H.S.}} \right) \right)_{k,k'}, \quad \forall k, k'.$$

L'image euclidienne des \mathbf{K} tableaux est appelée plan de l'interstructure. Les X_k sont identifiés par les points $N_1, \dots, N_{\mathbf{K}}$, dont les coordonnées sur l'axe a sont données par les composantes

du vecteur

$$\sqrt{\lambda_a^{sd}} \gamma_a^{sd} = S\Pi \gamma_a^{sd} / \sqrt{\lambda_a^{sd}},$$

où λ_a^{sd} est la $a^{\text{ième}}$ valeur propre de $S\Pi$ et γ_a^{sd} son vecteur propre normé à l'unité associé.

Comme pour p_1 , premier vecteur propre de \mathcal{C} , γ_1^{sd} est choisi avec toutes ses coordonnées positives.

Remarque : L'étude de l'interstructure permet de mettre en évidence les ressemblances entre les tableaux sans pour autant les expliquer.

2.1.2 Le compromis

L'idée des méthodes STATIS et STATIS dual est la suivante : si des ressemblances ont été mises en évidence lors de l'étude de l'interstructure, il est légitime de définir un compromis de la même nature que les objets représentatifs qui soit un bon «résumé» de l'ensemble des objets représentatifs au sens d'un critère.

STATIS :

Le compromis W est une somme pondérée des W_k^x (ou des $W_k^x / \|W_k^x\|_{H.S}$) définie de la manière suivante :

Définition 2.1 *Le compromis W est donné par :*

- $W = \alpha \sum_{k=1}^K \pi_k \alpha_k W_k^x$ avec $\alpha = \sum_k \pi_k \|W_k^x\|_{H.S} / \sqrt{\sigma_1}$, dans le cas d'objets non normés,
- $W = \alpha \sum_{k=1}^K \pi_k \alpha_k W_k^x / \|W_k^x\|_{H.S}$ avec $\alpha = 1 / \sqrt{\sigma_1}$, dans le cas d'objets normés.

On a $\|W\|_{H.S} = \alpha \sigma_1$.

α est appelé coefficient normalisateur. Il permet de ramener le compromis W à la même échelle que les objets utilisés. Ce compromis présente des propriétés d'optimalité au sens de certains critères :

Propriété 2.1 *W est un objet de même nature que les objets W_k^x (resp. $W_k^x / \|W_k^x\|_{H.S}$) et W est l'objet le plus corrélé, au sens de Hilbert-Schmidt, avec les objets W_k^x (resp. $W_k^x / \|W_k^x\|_{H.S}$), $k = 1, \dots, K$.*

Cela signifie que le compromis est la combinaison linéaire $\sum_{k=1}^K \beta_l W_l^x$ la plus proche de l'ensemble des opérateurs W_k^x , ce qui se traduit par la maximisation de

$$\left(\sum_{k=1}^K \pi_k \left(\sum_{l=1}^K \beta_l W_l^x D | W_k^x \right)_{H.S} \right)^2 / (\beta' \Pi^{-1} \beta),$$

où $\beta = (\beta_1, \dots, \beta_K)'$. On obtient de cette optimisation $\beta = \alpha \Pi p_1$. La preuve de cette propriété est disponible dans (LAVIT et al., 1994; LAVIT, 1988, pp93-96).

W étant de même nature que les objets W_k^x , il peut être vu comme la matrice des produits scalaires entre les individus du tableau $Y_c = [\sqrt{\alpha} \sqrt{\pi_1 \alpha_1} X_1 | \dots | \sqrt{\alpha} \sqrt{\pi_K \alpha_K} X_K]$, $l \times \sum_k J_k$, ou si les objets sont normés, $Y_c = [\sqrt{\alpha} \sqrt{\pi_1 \alpha_1} X_1 / \sqrt{\|W_1^x\|_{H.S}} | \dots | \sqrt{\alpha} \sqrt{\pi_K \alpha_K} X_K / \sqrt{\|W_K^x\|_{H.S}}]$. Y_c est le tableau-compromis.

Remarque : W est situé sur le premier axe de l'interstructure à la distance $\|W\|_{H.S} = \alpha\sqrt{\sigma_1}$ de l'origine (LAVIT, 1988, pp93-94).

STATIS dual :

Le compromis sera une matrice V de taille $J \times J$ représentant la matrice de variance-covariance (ou de corrélation si les études sont réduites), d'une certaine matrice, notée X_c , de taille $\sum_{k=1}^K I_k \times J$.

Définition 2.2 Pour des objets non normés, $V = \sum_{k=1}^K \frac{\pi_k \alpha_k}{\sum_k \pi_k \alpha_k} V_k^x$, où α_k est la coordonnée k de γ_1^{sd} et avec $\beta_k = \pi_k \alpha_k / \sum_k \pi_k \alpha_k$, pour $k = 1, \dots, K$,

$$X_c = \begin{bmatrix} \sqrt{\beta_1} X_1 \\ \vdots \\ \sqrt{\beta_K} X_K \end{bmatrix}.$$

Si les objets V_k^x sont normés, alors $V = \sum_{k=1}^K \frac{\pi_k \alpha_k}{\sum_k \pi_k \alpha_k} \frac{V_k^x}{\|V_k^x\|_{H.S.}}$ et

$$X_c = \begin{bmatrix} \sqrt{\beta_1} X_1 / \sqrt{\|V_1^x\|_{H.S.}} \\ \vdots \\ \sqrt{\beta_K} X_K / \sqrt{\|V_K^x\|_{H.S.}} \end{bmatrix}.$$

La somme des coefficients β_k est égale à 1.

Dans STATIS dual, il est préférable de centrer et réduire les tableaux X_k et de faire les calculs sur les objets V_k^x non réduits. Ainsi, les objets V_k et le compromis V représentent des matrices de corrélations, ce qui est plus facile à interpréter.

2.1.3 Etude de l'intrastructure

L'étape de l'intrastructure consiste à représenter les individus (ou variables) compromis et les individus (ou variables) afin de les comparer. Ceci prend encore plus de sens lorsqu'il y a une évolution temporelle le long de k .

STATIS :

Dans un premier temps, on fait une représentation des individus du tableau-compromis. Une image euclidienne des individus est obtenue en faisant l'ACP de Y_c avec la métrique diagonale par blocs des métriques Q_k^x . En notant λ_{cl} la $l^{\text{ième}}$ valeur propre de WD et c_l le $l^{\text{ième}}$ vecteur propre de norme unité, les coordonnées des individus-compromis sur l'axe l sont données par :

$$\sqrt{\lambda_{cl}} c_l = WDc_l / \sqrt{\lambda_{cl}}.$$

Ensuite, on définit des trajectoires* des individus en utilisant une technique de projection en supplémentaire les individus des objets W_k^x (resp. $W_k^x / \|W_k^x\|_{H.S.}$). Les coordonnées des l individus du tableau X_k sur l'axe l sont données par :

$$W_k^x Dc_l / \sqrt{\lambda_{cl}} \quad (\text{resp. } W_k^x Dc_l / \sqrt{\lambda_{cl} \|W_k^x\|_{H.S.}}).$$

* Le terme «trajectoires» fait référence au temps, ce qui n'est bien sûr pas toujours le cas.

STATIS dual :

Comme dans STATIS, l'image euclidienne des variables-compromis est obtenue en faisant l'ACP de X_c avec la métrique diagonale par bloc des métriques D_k . En notant λ_{cl}^{sd} la $l^{\text{ième}}$ valeur propre de $VQ = V$ et c_l^{sd} le $l^{\text{ième}}$ vecteur propre de norme unité associé, les coordonnées des variables-compromis sur l'axe l sont données par :

$$\sqrt{\lambda_{cl}^{sd}} c_l^{sd} = V c_l^{sd} / \sqrt{\lambda_{cl}^{sd}}.$$

Les trajectoires des variables sont obtenues en représentant en supplémentaire les variables des objets V_k^x (ou $V_k^x / \|V_k^x\|_{H.S.}$). Les coordonnées des J variables du tableau X_k sur l'axe l sont données par :

$$V_k^x c_l^{sd} / \sqrt{\lambda_{cl}^{sd}} \quad (\text{resp. } V_k^x c_l^{sd} / \sqrt{\lambda_{cl}^{sd} \|V_k^x\|_{H.S.}}).$$

Cette représentation des trajectoires n'est pas forcément optimale. Les individus des tableaux ne sont pas projetés sur les plans des individus compromis, et ne sont pas non plus donnés par une représentation euclidienne.

Le compromis et les trajectoires décrits ici, sont tels que les a définis LAVIT (1988, pp 117,120). il est possible d'utiliser des compromis et des trajectoires différents. Notamment, CHESSEL & HANAFI (1996) n'utilisent pas les mêmes définitions, à travers l'utilisation du logiciel ADE-4[†]. Un bref parcours des trajectoires «existantes» et les différents objets représentatifs des tableaux utilisables sont donnés par GLAÇON (1981).

2.2 Analyse de CO-inertie Multiple : ACOM

L'Analyse de CO-inertie Multiple (ACOM) a été proposée par CHESSEL & HANAFI (1996) comme généralisation à K tableaux de l'analyse de co-inertie.

Elle s'applique à K triplets $\{(X_k, Q_k^x, D)\}_k$, i.e. un ensemble de K tableaux munis de métriques, éventuellement pondérés par un poids π_k ($0 \leq \pi_k \leq 1$, $\sum_k \pi_k = 1$) mesurés sur les mêmes l individus et réalise K analyses dans K espaces séparés, \mathbb{R}^{J_k} .

2.2.1 Définitions et solutions

Définition 2.3 *L'ACOM d'ordre 1 (ou de rang 1) d'un ensemble de tableaux (X_1, \dots, X_K) est la recherche de K vecteurs $a_{k,1}$, Q_k^x -normés, et d'une variable D -normée de \mathbb{R}^l , v_1 , dite auxiliaire, qui maximisent le critère :*

$$C(\{a_k\}, v) = \sum_{k=1}^K \pi_k \text{cov}^2(X_k Q_k^x a_k, v).$$

On appelle ACOM d'un ensemble de tableau, d'ordre a , la recherche de K vecteurs $a_{k,a}$ et d'un vecteur v_a maximisant le critère sous les contraintes :

$$(v_a | v_j)_D = 0, 1 \leq j \leq a - 1, \text{ et } (a_{k,a} | a_{k,j})_{Q_k^x} = 0, 1 \leq j \leq a - 1, \forall k.$$

[†] Logiciel gratuit téléchargeable à l'adresse <http://pbil.univ-lyon1.fr/ADE-4/ADE-4F.html>.

Propriété 2.2 Les solutions d'ordre 1 vérifient :

$$\begin{cases} a_{k,1} = X'_k Dv_1 / \|X'_k Dv_1\|_{Q_k^x} \\ \sum_{k=1}^K \pi_k W_k^x Dv_1 = \lambda_1 v_1. \end{cases}$$

v_1 est le vecteur propre normé associé à la plus grande valeur propre de la somme des opérateurs d'inertie.

Pour chaque tableau, on a : $t_{k,1} = X_k Q_k^x a_{k,1} = W_k^x Dv_1 / \|X'_k Dv_1\|_{Q_k^x}$.

Démonstration :

La démonstration que nous proposons n'est pas celle des auteurs (CHESSEL & HANAFI, 1996). Elle nous paraît plus directe et plus simple.

La méthode des multiplicateurs de Lagrange, avec le Lagrangien

$$\mathcal{L}(\{a_k\}, v) = \sum_{k=1}^K \pi_k (v' D X_k Q_k^x a_k)^2 - \lambda (v' D v - 1) - \sum_{k=1}^K \mu_k (a'_k Q_k^x a_k - 1)$$

où λ et $\{\mu_k\}_k$ sont les multiplicateurs associés aux contraintes, mène aux équations normales :

$$\frac{1}{2} \nabla_v \mathcal{L} = \sum_{k=1}^K \pi_k (v' D X_k Q_k^x a_k) X'_k D X_k Q_k^x a_k - \lambda D v = 0 \quad (2.1)$$

$$\frac{1}{2} \nabla_{a_k} \mathcal{L} = \pi_k (v' D X_k Q_k^x a_k) Q_k^x X'_k D v - \mu_k Q_k^x a_k = 0 \quad k = 1, \dots, K \quad (2.2)$$

$$a'_k Q_k^x a_k - 1 = 0 \quad k = 1, \dots, K \quad (2.3)$$

$$v' D v - 1 = 0. \quad (2.4)$$

En multipliant à gauche l'équation (2.1) par v' et les équations (2.2) par a'_k , on trouve :

$$\lambda = \sum_{k=1}^K \mu_k, \text{ optimum du critère,} \quad (2.5)$$

$$\mu_k = \pi_k \text{cov}^2(v, X_k Q_k^x a_k) \quad k = 1, \dots, K.$$

En utilisant ces résultats dans les deux premières équations, on obtient :

$$a_k = \frac{\sqrt{\pi_k}}{\sqrt{\mu_k}} X'_k D v \quad (2.6)$$

$$\sum_{k=1}^K \pi_k W_k D v = \lambda v. \quad (2.7)$$

a_k étant de norme unité, $\sqrt{\pi_k} \|X'_k D v\|_{Q_k^x} = \sqrt{\mu_k}$.

■

Note : Afin d'alléger les notations, nous n'utiliserons plus de pondérations π_k dans la suite.

En notant $P_{a_k, a-1}^\perp = Id_{J_k} - a_{k, a-1} a'_{k, a-1} Q_k^x$ le projecteur Q_k^x -orthogonal sur l'espace engendré par $a_{k, a-1}$, les solutions d'ordre a sont déterminées grâce à la propriété suivante :

Propriété 2.3 *Les solutions d'ordre a , $a \geq 2$, sont les solutions de l'ACOM d'ordre 1 des K triplets $\{(X_k^{(a-1)}, Q_k^x, D)\}_{k=1, \dots, K}$, avec $X_k^{(a-1)} = X_k^{(a-2)} P_{a_k, a-1}^{\perp'}$ et $X_k^{(0)} = X_k$, $k = 1, \dots, K$.*

Démonstration :

(CHESSEL & HANAFI, 1996, pp 43-45)

■

2.2.2 Commentaires

Ici, les auteurs travaillent dans les espaces \mathbb{R}^{J_k} . Ce sont les individus de chaque tableau qui sont projetés à chaque étape sur $a_{k, a-1}^\perp$. Généralement, on a plutôt l'habitude de travailler dans l'espace \mathbb{R}^1 , afin d'étudier les variables. Avec l'ACOM, le but est de découper l'inertie des K nuages d'individus dans \mathbb{R}^{J_k} . Les cartes des individus sont alors optimales, pas celles des variables.

L'ACOM fournit, pour chaque triplet, un système d'axes de co-inertie Q_k^x -orthonormé $\{a_{k, a}\}_{a=1, \dots, A}$ qui peut être comparé aux axes d'inertie définis par une ACP sur chacun des triplets.

De plus, les variables auxiliaires de norme unité $\{v_a\}_a$, qui jouent le rôle de compromis, sont D -orthogonales entre elles, par contre les composantes de chaque tableaux ne le sont pas. Les cartes des individus de chaque espace sont obtenus en les projetant sur le plan dont les axes $a_{k, a}$ et $a_{k, \ell}$ forment un repère orthonormé. Les coordonnées de ces points sont données par les composantes $t_{k, a}$ et $t_{k, \ell}$. K cartes des variables de chaque groupe sont obtenues en projetant chaque variable dans le repère orthonormé (v_a, v_ℓ) de \mathbb{R}^1 . Les coordonnées des points projetés sont données par $(\|X_k' D v_a\|_{Q_k^x} a_{k, a}, \|X_k' D v_\ell\|_{Q_k^x} a_{k, \ell})$.

Cette méthode peut être vue comme une amélioration de l'Analyse Canonique Généralisée de Carroll (CARROLL, 1968) dont on peut trouver une description dans (SAPORTA, 1990, page 192), puisqu'ici, les composantes $t_{k, a}$ sont aussi explicatives que possible de leur propre groupe de variables.

Le programme de l'ACOM implémenté dans ADE-4[®] † utilise une autre façon de calculer les solutions d'ordre 1, basée sur la décomposition en valeurs singulières de la matrice $Y = D^{1/2} [\sqrt{\pi_1} X_1 | \dots | \sqrt{\pi_K} X_K] Q^{1/2}$, où $Q = \text{Diag}(Q_k^x)$, matrice diagonale par bloc. Soient v la variable auxiliaire de l'ACOM de rang 1 recherchée, a_k les axes de co-inertie de l'ACOM de rang 1 recherchés, λ le carré de la première valeur singulière, u , $(\sum_{k=1}^K J_k \times 1)$, et c , (1×1) , les premiers vecteurs de norme unité issus de la décomposition en valeurs singulières

† <http://pbil.univ-lyon1.fr/ADE-4/ADE-4F.html>

de Y , ce qui se traduit mathématiquement par :

$$\begin{aligned} YY'c &= \lambda c \\ c'c &= 1 \\ Y'Yu &= \lambda u \\ u'u &= 1, \text{ et} \\ c &= \frac{1}{\sqrt{\lambda}}Yu \\ u &= \frac{1}{\sqrt{\lambda}}Y'c. \end{aligned}$$

u est formé de K sous-vecteurs $u[k]$ de longueur J_k .

Alors $v = D^{-1/2}c$ et $a_k = Q_k^{x^{-1/2}} \frac{u[k]}{\|u[k]\|}$.

En effet, v est norme D -unité: $\|v\|_D = c'D^{-1/2}DD^{-1/2}c = c'c = 1$ et comme $YY' = D^{1/2} \sum_{k=1}^K \pi_k W_k^x D^{1/2}$, il vient :

$$\begin{aligned} \lambda D^{-1/2}c &= \sum_{k=1}^K \pi_k W_k^x D D^{-1/2}c \\ \lambda v &= \sum_{k=1}^K \pi_k W_k^x D v. \end{aligned}$$

v est donc bien le premier vecteur propre de norme unité de $\sum_{k=1}^K \pi_k W_k^x$.

De plus, a_k est de Q_k^x -norme unité :

$$\|a_k\|_{Q_k^x} = \frac{u[k]'}{\|u[k]\|} Q_k^{x-1/2} Q_k^x Q_k^{x-1/2} \frac{u[k]}{\|u[k]\|} = \frac{u[k]'u[k]}{\|u[k]\|^2} = 1,$$

et comme $u = \frac{1}{\sqrt{\lambda}}Y'c$,

$$\begin{aligned} u[k] &= \frac{\sqrt{\pi_k}}{\sqrt{\lambda}} Q_k^{x1/2} X_k' D^{1/2} c \\ u[k]/\|u[k]\| &= Q_k^{x1/2} X_k' D^{1/2} c / \|X_k' D^{1/2} c\|_{Q_k^x} \\ &= Q_k^{x1/2} X_k' D v / \|X_k' D v\|_{Q_k^x}. \end{aligned}$$

D'où $Q_k^{x-1/2} \frac{u[k]}{\|u[k]\|} = X_k' D v / \|X_k' D v\|_{Q_k^x} = a_k$, qui est bien la solution recherchée.

2.3 Analyse de concordance : CONCORD ou ACO

Tandis que l'ACOM recherche des ressemblances entre K triplets (X_k, Q_k^x, D) à l'aide d'une variable auxiliaire, l'Analyse de la Concordance (LAFOSSE & HANAFI, 1997), CONCORD notée aussi ACO, suppose que l'on dispose d'un triplet supplémentaire (Y, Q^y, D) , mesuré sur les mêmes l individus que les K tableaux, et auquel ces derniers vont être comparés. Y est, en ce sens, un tableau de référence. Il est possible de pondérer les K études (X_k, Q_k^x, D) comme cela est fait dans l'ACOM (§2.2), mais pour plus de lisibilité, nous ne le ferons pas. (Il suffit de considérer que le poids π_k est compris dans la métrique Q_k^x pour chacun des tableaux).

2.3.1 Définitions et solutions

Cette méthode a été proposée par LAFOSSE & HANAFI (1997) comme généralisation à plus de deux tableaux de l'Analyse Inter-Batterie Non-Symétrique, AIBNS, (TENENHAUS, 1995). De ce point de vue, l'analyse de la concordance d'ordre 1 de K tableaux est très simple à définir. Rappelons que l'AIBNS des deux triplets (X, Q^x, D) et (Y, Q^y, D) est la recherche des axes a et b , respectivement Q^x et Q^y normés à l'unité et associés à X et Y . b est obtenu par la maximisation de $b'Q^yY'DXQ^xX'DYQ^yb$ sous la contrainte $\|b\|_{Q^y} = 1$ puis a par la formule de transition $a = X'DYQ^yb/\|X'DYQ^yb\|_{Q^x}$.

Définition 2.4 *L'ACO d'ordre 1 est l'Analyse Inter-Batterie Non-Symétrique d'ordre 1 du couple (X, Y) où X est le tableau conjoint des K tableaux X_k . Ainsi, b_1 , le premier axe associé à Y est le premier vecteur propre normé à 1 de $V_{YX}Q^yV_{XY}Q^y$ et $a_{k,1}$, normé à 1, qui est l'axe associé à X_k est donné par $a_{k,1} = V_{kY}Q^yb_1/cov(X_kQ_k^xa_{k,1}, YQ^yb_1)$, où $V_{XY} = X'DY = X'_{YX}$, $V_{k,Y} = X'_kDY = V'_{Yk}$ et $Q^x = \text{Diag}(Q_k^x)$.*

LAFOSSE & HANAFI (1997) montrent l'équivalence de ce problème avec le suivant, que nous utiliserons comme définition de l'Analyse de COncordance (ACO) :

Définition 2.5 *L'ACO d'ordre a est la recherche de $(K+1)$ -uples : $(YQ^yb_a, X_1Q_1^xa_{1,a}, \dots, X_KQ_K^xa_{K,a})$ maximisant le critère :*

$$C_1(\{a_k\}, b) = \sum_{k=1}^K cov^2(YQ^yb, X_kQ_k^xa_k)$$

sous les contraintes d'orthogonalité, de Q_k^x -norme unité pour les $\{a_{k,\ell}\}_{\ell=1,\dots,a}$ et de Q^y -norme unité pour $\{b_a\}_a$.

Nous proposons de montrer par la méthode des multiplicateurs de Lagrange que les solutions d'ordre 1 de ce problème sont bien égales à celle de la définition 2.4. Les auteurs ne le font pas de cette manière.

En utilisant le Lagrangien :

$$\mathcal{L}(\{a_k\}, b) = \sum_{k=1}^K (b'Q^yV_{Yk}Q_k^xa_k)^2 - \lambda(b'Q^yb - 1) - \sum_{k=1}^K \mu_k(a'_kQ_k^xa_k - 1)$$

où λ et $\{\mu_k\}$ sont les multiplicateurs associés aux contraintes, nous déduisons les équations normales

$$\frac{1}{2}\nabla_b\mathcal{L} = \sum_{k=1}^K (b'Q^yV_{Yk}Q_k^xa_k)Q^yV_{Yk}Q_k^xa_k - \lambda Q^yb = 0 \quad (2.8)$$

$$\frac{1}{2}\nabla_{a_k}\mathcal{L} = (b'Q^yV_{Yk}Q_k^xa_k)Q_k^xV_{kY}Q^yb - \mu_kQ_k^xa_k = 0 \quad k = 1, \dots, K \quad (2.9)$$

$$b'Q^yb - 1 = 0 \quad (2.10)$$

$$a'_kQ_k^xa_k - 1 = 0 \quad k = 1, \dots, K \quad (2.11)$$

En multipliant à gauche l'équation (2.8) par b' et à gauche les équations (2.9) par a'_k , nous obtenons :

$$\begin{aligned}\mu_k &= \text{cov}^2(YQ^y b, X_k Q_k^x a_k) \quad k = 1, \dots, K \\ \lambda &= \sum_{k=1}^K \mu_k.\end{aligned}\quad (2.12)$$

Enfin, en utilisant ces résultats dans les équations (2.9) et (2.8), les relations obtenues sont :

$$a_k = \frac{1}{\sqrt{\mu_k}} V_{kY} Q^y b \quad (2.13)$$

$$\sum_{k=1}^K V_{Yk} Q_k^x V_{kY} Q^y b = \lambda b, \quad (2.14)$$

a_k étant de norme unité, il vient $\sqrt{\mu_k} = \|V_{kY} Q^y b\|_{Q_k^x}$.

L'axe de co-inertie b_1 recherché est donc bien le premier vecteur propre de la matrice $\sum_{k=1}^K V_{Yk} Q_k^x V_{kY} Q^y$ associé à la première valeur propre λ_1 . Les axes $\{a_{k,1}\}_k$ sont ensuite déterminés par les formules de transition (2.13).

Les solutions obtenues sont bien au maximum de la fonction objectif, car les axes b et a_k sont caractérisés par λ , égal au critère, qui est choisi maximum.

Nous en déduisons le même type de relations pour les composantes $t_k = X_k Q_k^x a_k$, $k = 1, \dots, K$, et $u = YQ^y b$ associées respectivement aux tableaux X_k , $k = 1, \dots, K$, et Y :

$$YQ^y YD \left(\sum_{k=1}^K \sqrt{\mu_k} t_k \right) = \lambda u \quad (2.15)$$

$$W_k^x D u = \sqrt{\mu_k} t_k. \quad (2.16)$$

On retrouve bien les résultats de l'AIBNS (déf. 2.4) de (X, Y) car $\sum_{k=1}^K W_k^x D = X \text{Diag}(Q_k^x) X' D$.

Donc, u est le premier vecteur propre, non normé à l'unité, de $YQ^y Y' D \sum_{k=1}^K W_k^x D$.

De plus :

$$YQ^y Y' D \left(\sum_{k=1}^K W_k^x D \right) u = \lambda u \quad (2.17)$$

$$\sum_{k=1}^K W_k^x D YQ^y Y' D \left(\sum_{k=1}^K \sqrt{\mu_k} t_k \right) = \lambda \left(\sum_{k=1}^K \sqrt{\mu_k} t_k \right). \quad (2.18)$$

Ceci montre que l'ACO d'ordre 1 peut aussi être considérée comme une Analyse Inter-Batterie de Tucker (TUCKER, 1958a) d'ordre 1 de (X, Y) , dont les composantes seraient respectivement $(\sqrt{\mu_1}/\sqrt{\lambda} t_1, \dots, \sqrt{\mu_K}/\sqrt{\lambda} t_K)'$, et u .

Les solutions d'ordre supérieur sont en fait déterminées grâce à la propriété suivante :

Propriété 2.4 *A l'étape a , (ou ordre a), ($a \geq 2$), les axes recherchés sont les solutions d'ordre 1 de l'analyse de concordance des triplets: $(X_1^{(a-1)}, Q_1^x, D), \dots, (X_K^{(a-1)}, Q_K^x, D), (Y, Q^y, D)$ où $X_k^{(a-1)} = X_k^{(a-2)} P_{a_k, a-1}^{\perp'}$ et $X_k^{(0)} = X_k$ avec $P_{a_k, a-1}^{\perp} = Id_{J_k} - a_{k, a-1} a'_{k, a-1} Q_k^x$.*

Pour la démonstration, se référer à (LAFOSSE & HANAFI, 1997) et (VIVIEN, 1999, annexe C). La déflation des individus des tableaux X_k sur les axes a_k permet de décomposer de façon additive l'inertie « commune » de chacun des nuages correspondants. L'orthogonalité des axes $\{a_{k,a}\}_a$ entre eux, due à la déflation, entraîne l'orthogonalité du système de axes $\{b_a\}_a$, ce qui permet un découpage de l'inertie du tableau Y (LAFOSSE & HANAFI, 1997).

Propriété 2.5

- La composante $YQ^y b_j$ est non corrélée aux composantes $X_k Q_k^x a_{k,l}$, $l = j + 1, \dots, A$
- Les corrélations entre la composante $YQ^y b_j$ et les composantes $X_k Q_k^x a_{k,l}$, $l = 1, \dots, j - 1$ sont faibles.

On peut trouver la démonstration de cette propriété dans (LAFOSSE & HANAFI, 1997) et VIVIEN (1999).

2.3.2 Commentaires et cas particuliers

L'ACO de $K + 1$ tableaux a pour but un découpage de l'inertie du tableau de référence. C'est pour cette raison que les déflations sont faites dans \mathbb{R}^{J^k} . Ainsi, les axes $\{a_{k,a}\}_a$ (*resp.* $\{b_a\}_a$) sont orthonormés pour Q_k^x dans \mathbb{R}^{J^k} (*resp.* pour Q^y dans \mathbb{R}^M). Deux axes d'ordres différents forment un système orthonormé $(a_{k,a}, a_{k,\ell})$ de \mathbb{R}^{J^k} (*resp.* (b_a, b_ℓ) de \mathbb{R}^M) dans le quel les individus projetés de X_k (*resp.* Y) ont pour coordonnées les composantes $t_{k,a} = X_k Q_k^x a_{k,a}$ et $t_{k,\ell} = X_k Q_k^x a_{k,\ell}$ (*resp.* $u_a = YQ^y b_a$ et $u_\ell = YQ^y b_\ell$). Par contre, les composantes $\{t_{k,a}\}_a$ n'étant pas D -orthogonales, on ne peut pas représenter simplement la projection des variables dans le plan que deux d'entre elles engendrent. Les axes ne sont pas composés, *a priori*, des coordonnées des projections des variables dans ce plan, et la représentation des J_k points de coordonnées $(a_{k,a}, a_{k,\ell})$ dans un repère orthonormé est à interpréter avec précaution. De même, les corrélations (ou covariances si les tableaux ne sont pas réduits) entre les variables de X_k et les composantes $t_{k,a}$ et $t_{k,\ell}$ ne sont pas non plus les coordonnées, dans un repère orthonormé de ce plan, des variables projetées. Les propriétés de dualité de l'ACP ne sont plus valables dans le cas de K tableaux (voir l'annexe C).

Si on prend pour Y le tableau X , concaténation des K tableaux X_k , et comme métrique Q^y , la métrique diagonale par blocs des Q_k^x , la première composante $u_1 = YQ^y b_1$ de l'ACO est égale à la première variable auxiliaire v_1 de l'ACOM, à la norme près. Cette analyse est appelée Analyse de Concordance symétrique, symétrique dans le sens où cela revient à faire l'ACO de (X, X) (LAFOSSE & HANAFI, 1997). En effet, dans cette situation u_1 est le premier vecteur propre non normé de $(\sum_{k=1}^K W_k^x D)^2$. Ce qui est aussi le cas de v_1 qui est le premier vecteur propre normé de $\sum_k W_k^x D$. Ceci implique $u_1 = \|u_1\|_D v_1$, et l'égalité des composantes $t_{k,1}$ de l'ACO avec les premières composantes de l'ACOM.

Si, de plus, on utilise la métrique de Mahalanobis $Q^y = V_{Y^y}^{-1}$, (au lieu de la matrice diagonale par blocs des métriques Q_k^x), l'ACOM est exactement obtenue. En effet, dans cette situation, l'opérateur $W_Y D = YQ^y Y' D$ est égal à $P_Y = P_{[X_1 \dots X_K]}$. Or, $P_Y W_k^x D = W_k^x D$, car $Im(X_k)$ est inclus dans $Im(Y)$. D'où, le vecteur u est le premier vecteur propre associé à la plus grande valeur propre λ de $\sum_{k=1}^K W_k^x D$. u est de D -norme unité dans \mathbb{R}^1 à cause de la forme de la métrique Q^y : $\|u\|_D^2 = b' Q^y Y' D Y Q^y b = b' Q^y b = 1$. De plus,

étant donné la définition de l'ACO pour les étapes suivantes, l'ACOM est entièrement définie.

Si $Q^y = V_{YY}^{-1}$, u_1 est la première composante normée à l'unité de l'ACP sur Variables Instrumentales (RAO, 1964) du triplet $(X = [X_1 \ \dots \ X_K], Q^x = \text{Diag}(Q_k^x), D)$ par rapport à Y .

En effet, l'opérateur $W_Y D$ est égal à P_Y . u_1 est alors le vecteur propre associé à la plus grande valeur propre λ_1 de $P_Y \sum_{k=1}^K W_k^x D$. Ceci implique que u_1 appartient à $\text{Im}(Y)$. Donc u_1 est le vecteur propre associé à la plus grande valeur propre λ_1 de $P_Y \sum_{k=1}^K W_k^x D P_Y$. Enfin, $\sum_{k=1}^K W_k^x D$ est égal à l'opérateur $W_X = X Q^x X' D$ de X . D'où le résultat.

Si de plus $Q_k^x = V_{kk}^{-1}$ pour tout k , alors $W_k^x D = P_{X_k}$ et on retrouve une Analyse Canonique Généralisée d'ordre 1 maximisant le critère C_1 sous les contraintes de norme unité pour les composantes t_k et u (VIVIEN, 1999, Annexe D).

Si Y est une matrice $l \times \sum_{k=1}^K J_k$, $Q_k^x = V_{kk}^{-1}$ pour tout k et $Q^y = \text{Diag}(Q_k)$, alors la composante u_1 est la première variable auxiliaire de l'ACG de Carroll, et $t_{k,1}$ est égal, à la norme près, à la première variable canonique $c_{k,1}$ du tableau k . En effet, u_1 est le vecteur propre associé à la plus grande valeur propre λ_1 de $\left(\sum_{k=1}^K P_{X_k}\right)^2$. La matrice $\sum_{k=1}^K P_{X_k}$ étant D -auto-adjointe, u_1 est le vecteur propre de norme unité associé à la plus grande valeur propre $\sqrt{\lambda_1}$ de $\sum_{k=1}^K P_{X_k}$: il correspond à la première variable auxiliaire de l'ACG de Carroll (CARROLL, 1968; SAPORTA, 1990, page 192). Quant aux composantes $t_{k,1}$, elles sont de norme unité et égales à $\frac{P_{X_k} u_1}{\|P_{X_k} u_1\|_D}$, tandis que les variables canoniques $c_{k,1}$ ne sont pas de norme unité et sont de la forme $\frac{P_{X_k} u_1}{\sqrt{\lambda_1}}$.

Dans le cas où $Y = D^{-1}$ et $D = Q^y$, les solutions sont celles de l'ACOM. En effet, le critère à maximiser à l'étape a devient $\sum_{k=1}^K \text{cov}^2(b, X_k^{(a-1)} Q_k^x a_k)$ et b_a est le vecteur propre associé à la plus grande valeur propre de λ de $\sum_{k=1}^K W_k^{x(a-1)} D$, où $X^{(a-1)} = X_k P_{\{a_{k,1}, \dots, a_{k,a-1}\}}^{\perp}$. Si, de plus, $Q_k^x = V_{kk}^{-1}$ pour tout k , on retrouve l'ACG de Carroll d'ordre 1, $t_{k,1}$ étant égal à la première variable canonique $c_{k,1}$ à la norme près.

2.4 Analyse de Co-Inertie Multiple Orthogonale : ACIMO

Dans ce paragraphe, nous présentons une nouvelle méthode de recherche d'axes de co-inertie[§] : l'Analyse de Co-Inertie Multiple Orthogonale (ACIMO). Le terme «orthogonale» est employé ici pour faire référence à l'orthogonalité des composantes obtenues à partir de l'ACIMO. Cette orthogonalité est importante pour qu'à chaque étape, chaque composante explique une part différente (des étapes précédentes) de l'inertie du tableau qu'elle «résume».

2.4.1 La méthode

Son but est de rechercher la structure commune entre les K tableaux X_k et un tableau Y , tout en expliquant au mieux les tableaux individuellement.

[§] soumis : Vivien, M. & Sabatier, R. OMCIA : a useful method to obtain optimal trajectories in (dual) STATIS.

Pour cela, nous maximisons le critère C_1 :

$$C_1(\{a_k\}, b) = \sum_{k=1}^K \text{cov}^2(YQ^y b, X_k Q_k^x a_k),$$

sous les contraintes de norme unité pour les axes $\{a_k\}_{k=1, \dots, K}$ et b .

La méthode des multiplicateurs de Lagrange permet d'en déterminer les solutions d'ordre 1. Le Lagrangien utilisé est le suivant :

$$\mathcal{L}(\{a_k\}, b) = \sum_{k=1}^K (b' Q^y Y' D X_k Q_k^x a_k)^2 - \lambda (b' Q^y b - 1) - \sum_{k=1}^K \mu_k (a_k' Q_k^x a_k - 1)$$

où λ et $\{\mu_k\}_k$ sont les multiplicateurs associés aux contraintes.

Les équations normales en découlent :

$$\frac{1}{2} \nabla_b \mathcal{L} = \sum_{k=1}^K (b' Q^y Y' D X_k Q_k^x a_k) Q^y Y' D X_k Q_k^x a_k - \lambda Q^y b = 0 \quad (2.19)$$

$$\frac{1}{2} \nabla_{a_k} \mathcal{L} = (b' Q^y Y' D X_k Q_k^x a_k) Q_k X_k' D Y Q^y b - \mu_k Q_k^x a_k = 0 \quad (2.20)$$

$$b' Q^y b - 1 = 0 \quad (2.21)$$

$$a_k' Q_k^x a_k - 1 = 0 \quad k = 1, \dots, K \quad (2.22)$$

La multiplication à gauche de l'équation (2.19) par b' et à gauche des équations (2.20) par a_k' donnent l'expression des multiplicateurs :

$$\begin{aligned} \mu_k &= \text{cov}^2(YQ^y b, X_k Q_k^x a_k) \quad k = 1, \dots, K \\ \lambda &= \sum_{k=1}^K \mu_k. \end{aligned} \quad (2.23)$$

Enfin, en utilisant ces résultats dans l'équation (2.20) et en reportant dans (2.19), les expressions des axes sont obtenues :

$$a_k = \frac{1}{\sqrt{\mu_k}} X_k' D Y Q^y b \quad (2.24)$$

$$\sum_{k=1}^K Y' D X_k Q_k^x X_k' D Y Q^y b = \lambda b, \quad (2.25)$$

a_k étant de norme unité, $\sqrt{\mu_k} = \|X_k' D Y Q^y b\|_{Q_k^x}$: la covariance entre $u = YQ^y b$ et $t_k = X_k Q_k^x a_k$ est positive quelque soit k .

L'axe de co-inertie b_1 recherché est le premier vecteur propre associé à la première valeur propre λ_1 de la matrice $\sum_{k=1}^K Y' D X_k Q_k^x X_k' D Y Q^y$. Les axes $\{a_{k,1}\}_k$ sont ensuite déterminés par les formules de transition (2.24).

Les solutions sont bien au maximum du critère C_1 , car les axes b et a_k sont caractérisés par λ , égal au critère, qui est choisi maximum.

On en déduit le même type de relations pour les composantes t_k , et u associées respectivement aux tableaux X_k , $k = 1, \dots, K$, et Y :

$$W^y D \left(\sum_{k=1}^K \sqrt{\mu_k} t_k \right) = \lambda u \quad (2.26)$$

$$W_k^x D u = \sqrt{\mu_k} t_k. \quad (2.27)$$

u est donc le premier vecteur propre, non normé à l'unité, de $W^y D \sum_{k=1}^K W_k^x D$. Nous pouvons aussi écrire :

$$W^y D \left(\sum_{k=1}^K W_k^x D \right) u = \lambda u \quad (2.28)$$

$$\sum_{k=1}^K W_k^x D W^y D \left(\sum_{k=1}^K \sqrt{\mu_k} t_k \right) = \lambda \left(\sum_{k=1}^K \sqrt{\mu_k} t_k \right). \quad (2.29)$$

Afin, que la variation expliquée de chaque tableau soit différente à chaque étape, nous avons choisi dans les étapes suivantes de chercher des ensembles de composantes orthogonales dans \mathbb{R}^l , $\{t_{k,a}\}_a$ et $\{u_a\}_a$.

Les solutions d'ordre a sont définies de la manière suivante :

Définition 2.6 Les solutions d'ordre a , avec $2 \leq a \leq A \leq \min(\text{rg}(X_k))$, de l'ACIMO sont les solutions d'ordre 1 de l'ACIMO des triplets : $(X_1^{(a-1)}, Q_1^x, D), \dots, (X_K^{(a-1)}, Q_K^x, D), (Y^{(a-1)}, Q^y, D)$ où $X_k^{(a-1)} = P_{t_{k,a-1}}^\perp X^{(a-2)}, Y^{(a-1)} = P_{u_{a-1}}^\perp Y^{(a-2)}$, avec $X_k^{(0)} = X_k, Y^{(0)} = Y$.

Alors, les K systèmes de composantes $\{t_{k,a} = X_k^{(a-1)} Q_k^x a_{k,a}\}_{a=1, \dots, A}$ sont D -orthogonaux ainsi que le système $\{u_a = Y^{(a-1)} Q^y b_a\}_{a=1, \dots, A}$.

Propriété 2.6 Les axes $\{a_{k,a}\}_{a=1, \dots, A}, k = 1, \dots, K$ et $\{b_a\}_{a=1, \dots, A}$ forment des systèmes de vecteurs Q_k^x et Q^y orthonormés de \mathbb{R}^{J_k} et \mathbb{R}^M .

Démonstration :

Notons d'une * les éléments correspondants à l'étape présente, les autres sont d'une étape précédente [¶].

On a $u^* = Y^* Q^y b^*$ et $t_k^* = X_k^* Q_k^x a_k^*$ avec $Y^* = P_U^\perp Y$ et $X_k^* = P_{T_k}^\perp X_k$ où U (resp. T_k) est la matrice formée par juxtaposition des u_a (resp. $t_{k,a}$) des étapes précédentes. D'où $Y^* Q^y b = P_U^\perp Y Q b = P_U^\perp u = 0$ et $X_k^* Q_k^x a_k = P_{T_k}^\perp t_k = 0$.

[¶] Si l'on se place à l'étape a , on a $u^* = u_a, t_k^* = t_{k,a}, b^* = b_a, a_k^* = a_{k,a}, Y^* = Y^{(a-1)}, X_k^* = X_k^{(a-1)}$ et $V_{Y_k}^* = Y^{*'} D X_k^*$. Les éléments sans étoiles correspondent, de la même manière, à des éléments d'une étape précédente.

D'après les équations (2.8) et (2.9) on a :

$$\begin{aligned}\lambda^* b' Q b^* &= \sum_{k=1}^K \sqrt{\mu_k^*} b' Q^y V_{Y^*}^* Q_k^x a_k^* \\ &= \sum_{k=1}^K \sqrt{\mu_k^*} b' Q^y Y^{*'} D X_k^* Q_k^x a_k^* \\ &= 0\end{aligned}$$

et pour $k = 1, \dots, K$

$$\begin{aligned}\mu_k^* a_k' Q_k^x a_k^* &= \sum_{k=1}^K \sqrt{\mu_k^*} a_k' Q_k^x V_{kY^*}^* Q^y b^* \\ &= \sum_{k=1}^K \sqrt{\mu_k^*} a_k' Q_k^x X_k^{*'} D Y^* Q^y b^* \\ &= 0.\end{aligned}$$

■

2.4.2 Commentaires et cas particuliers

Les individus «vus par chacun des tableaux» et ceux «vus par Y » peuvent être représentés sur un même graphique. Cette représentation conjointe des individus dans le plan (i, j) est obtenue en croisant $\left\{ \left(\frac{t_{k,i}}{\|t_{k,i}\|_D}, \frac{t_{k,j}}{\|t_{k,j}\|_D} \right) \right\}_k$ et $\left(\frac{u_i}{\|u_i\|_D}, \frac{u_j}{\|u_j\|_D} \right)$.

Des cartes pour les variables peuvent être obtenues par les corrélations (ou les covariances) entre les variables d'un tableau et deux composantes de ce tableau. Ces corrélations sont en réalité les coordonnées des variables projetées dans le repère orthonormé de \mathbb{R}^1 défini par deux composantes. Ainsi, les corrélations $r(t_{k,1}, x_k^j)$ et $r(t_{k,2}, x_k^j)$ sont les coordonnées de la variable x_k^j de X_k projetée dans le repère $\left(\frac{t_{k,1}}{\|t_{k,1}\|_D}, \frac{t_{k,2}}{\|t_{k,2}\|_D} \right)$. De même, les variables de Y projetées dans $\left(\frac{u_1}{\|u_1\|_D}, \frac{u_2}{\|u_2\|_D} \right)$ ont pour coordonnées les corrélations $r(u_1, Y)$ et $r(u_2, Y)$. De plus, les vecteurs $\frac{\sqrt{\mu_{k,a}}}{\|u_a\|_D} a_{k,a}$ et $\frac{\sqrt{\mu_{k,\ell}}}{\|u_\ell\|_D} a_{k,\ell}$ contiennent les coordonnées des variables de X_k projetées dans le repère orthonormé $\left(\frac{u_a}{\|u_a\|_D}, \frac{u_\ell}{\|u_\ell\|_D} \right)$. En effet, $P_{u_a} X_k = \frac{u_a}{\|u_a\|_D} \frac{u_a'}{\|u_a\|_D} D X_k = \frac{u_a}{\|u_a\|_D} \frac{u_a'}{\|u_a\|_D} D (X_k - P_{u_1, \dots, u_{a-1}} X_k) = \frac{u_a}{\|u_a\|_D} \frac{u_a'}{\|u_a\|_D} D X_k^{(a-1)}$. Le critère cherche des composantes t_k et u qui soient le plus proche possible. Si ces $K + 1$ composantes sont suffisamment proches, on peut penser que les variables de Y projetées sur $\left(\frac{u_a}{\|u_a\|_D}, \frac{u_\ell}{\|u_\ell\|_D} \right)$ auront des coordonnées proches de $\lambda_a b_a = \lambda_a \sum_{k=1}^K \sqrt{\mu_{k,a}} t_{k,a}$. Ainsi, en utilisant les vecteurs b_a , on devrait obtenir une carte des variables dont la configuration (positions relatives) des points est proche de celle obtenue en utilisant les corrélations entre Y et u_a .

La variance des composantes de co-inertie $t_{k,a}$ peut être comparée avec la variance des composantes principales de même ordre du tableau k . Plus elles sont proches, meilleure est

l'explication du tableau k par l'axe de co-inertie.

L'ACIMO d'ordre 1 est équivalente à l'Analyse de la Concordance d'ordre 1 (LAFOSSE & HANAFI, 1997).

Dans le cas où $K = 1$, on retrouve l'Analyse Canonique PLS de TENENHAUS (1998).

Les composantes $t_{k,a}$ sont des combinaisons linéaires des variables initiales de X_k (pour tout a et tout k) et les composantes u_a sont des combinaisons linéaires des variables initiales Y . On peut montrer que les composantes $t_{k,a}$, (*resp.* u_a) peuvent s'écrire de la forme $t_{k,a} = X_k Q_k^x \alpha_{k,a}$ (*resp.* $u_a = Y Q^y \gamma_a$), avec $\alpha_{k,a} = a_{k,a} - \sum_{l=1}^{a-1} \frac{\alpha_{k,l} \alpha'_{k,l}}{\|t_{k,l}\|_D^2} X_k' D X_k Q_k^x a_{k,a}$, (*resp.* $\gamma_a = b_a - \sum_{l=1}^{a-1} \frac{\gamma_l \gamma_l'}{\|u_l\|_D^2} Y' D Y Q^y b_a$), pour tout $a = 1, \dots, A$. La démonstration de ce résultat est similaire à la démonstration de la proposition 5.3, p 99.

2.5 Application de l'ACIMO à l'intrastructure de STATIS

L'intrastructure de STATIS et de STATIS dual n'étant pas optimale, nous proposons de représenter les trajectoires d'une autre manière: en utilisant l'ACIMO.

Dans un premier temps, nous expliquerons en détail comment utiliser l'ACIMO avec STATIS et énoncerons les propriétés relatives.

Ensuite, nous montrerons comment adapter tout ce qui précède à STATIS dual.

2.5.1 STATIS et ACIMO

Nous proposons de réaliser l'intrastructure de STATIS d'une autre manière. Rechercher des axes de co-inertie à l'aide de la méthode ACIMO appliquée aux K tableaux X_k et au tableau-compromis $Y_c = \sqrt{\alpha} [\sqrt{\pi_1 \alpha_1} X_1 | \dots | \sqrt{\pi_K \alpha_K} X_K]$

(ou bien $Y_c = \sqrt{\alpha} [\sqrt{\pi_1 \alpha_1} X_1 / \sqrt{\|W_1^x\|_{H.S}} | \dots | \sqrt{\pi_K \alpha_K} X_K / \sqrt{\|W_K^x\|_{H.S}}]$).

Alors, la représentation simultanée des individus de chacun des tableaux (et des trajectoires) et des individus-compromis est optimale au sens où les composantes t_k sont les plus proches de la composante u du tableau-compromis.

La «représentation simultanée» est obtenue par la superposition sur un même plan des quantités $t_k / \|t_k\|_D$ et $u / \|u\|_D$.

Propriété 2.7 *Si les tableaux sont pondérés uniformément ($\pi_k = \pi, \forall k$) et si les coefficients $\alpha_1, \dots, \alpha_K$ sont tous égaux à une constante a ($= 1/\sqrt{K}$), alors les composantes u_a de l'ACIMO de (X_1, \dots, X_K, Y_c) , avec $Y_c = [\sqrt{\alpha} \sqrt{\pi_1 \alpha_1} X_1 | \dots | \sqrt{\alpha} \sqrt{\pi_K \alpha_K} X_K]$, sont les composantes principales, non normées à l'unité, $u_{cl} = \sqrt{\lambda_{cl}} c_l$, du tableau-compromis Y_c .*

Démonstration :

On suppose que $\alpha_k = \beta \forall k$.

Montrons tout d'abord que la propriété est vraie à l'étape 1.

Alors, $\sum_{k=1}^K W_k^x D = W D / (\pi \alpha \beta)$.

Comme $\lambda_1 u_1 = W D \sum_{k=1}^K W_k^x D u_1$, on a $\pi \alpha \beta \lambda_1 u_1 = (W D)^2 u_1$.

u_1 est donc le premier vecteur propre de WD^2 .

Notons u_{ci} et λ_{ci} le $i^{\text{ième}}$ vecteur propre (composante principale de norme $\sqrt{\lambda_{ci}}$) et la $i^{\text{ième}}$ valeur propre de WD . On a $\lambda_{ci}^2 u_{ci} = (WD)^2 u_{ci} \forall i$.

u_{c1} est aussi le premier vecteur propre de WD^2 . On a donc les relations :

$$\begin{aligned} u_{c1} &= u_1 \\ \lambda_{c1}^2 &= \pi\alpha\beta\lambda_1. \end{aligned}$$

Remarquons qu'à chaque étape a , $a = 1, \dots, A$, $\pi\alpha\beta\lambda_a u_a = (W^{(a-1)}D)^2 u_a$.

Montrons la propriété pour l'étape a .

On suppose que la propriété est vraie jusqu'à l'étape $a - 1$.

A l'étape a , les calculs se font sur le tableau $Y^{(a-1)} = P_{u_{a-1}}^\perp Y^{(a-2)}$.

Soit U la matrice formée, en colonne, des vecteurs u_1, \dots, u_{a-1} et P_U le projecteur D -orthogonal sur l'espace engendré par ces vecteurs, $Im(U)$.

Alors, $Y^{(a-1)} = P_U^\perp Y$.

Nous allons montrer que u_a est le $a^{\text{ième}}$ vecteur propre de WD^2 .

On a :

$$W^{(a-1)}D = WD - P_U WD - WDP_U + P_U WDP_U.$$

Or, $WDU = \Lambda U$ où Λ est la matrice diagonale des valeurs propres λ_{ci} , $i = 1, \dots, a-1$, car on suppose la propriété vérifiée jusqu'à l'étape $a-1$. De plus, $U'DU$ étant la matrice de variance-covariance des $\{u_1, \dots, u_{a-1}\} = \{u_{c,1}, \dots, u_{c,a-1}\}$, on a $U'DU = \Lambda$.

D'où :

$$\begin{aligned} P_U WD &= WDP_U = P_U WDP_U = UU'D \\ W^{(a-1)}Du_a &= WDu_a - WDP_U u_a \\ &= WDu_a \text{ car } u_a \text{ est orthogonal à } u_i, i = 1, \dots, a-1 \\ (W^{(a-1)}D)^2 u_a &= (WD)^2 u_a - WDU \underbrace{U'Du_a}_0 \\ &= (WD)^2 u_a. \end{aligned}$$

On obtient

$$\pi\alpha\beta\lambda_a = (WD)^2 u_a.$$

Enfin, u_a est orthogonal à u_i , $i = 1, \dots, a-1$, ce qui implique que u_a est bien le $a^{\text{ième}}$ vecteur propre, non normé à l'unité, de WD^2 , et on a les relations :

$$\begin{aligned} \pi\alpha\beta\lambda_a &= \lambda_{ca}^2 \\ u_a &= u_{ca}. \end{aligned}$$

L'inertie du nuage compromis projeté sur les axes d'inertie est égale à l'inertie du nuage projeté sur les axes de co-inertie. ■

Dans le cas d'objets W_k^x non normés, lorsque le compromis est bien représentatif de tous les tableaux, les coefficients α_k du compromis ne sont pas trop éloignés les uns des autres. Alors les composantes principales du compromis sont «proches» des u_a .

Si le compromis n'est représentatif que de certains tableaux, on peut se demander s'il est utile de faire une étude de l'intrastructure de l'ensemble total des tableaux. Il est préférable alors d'en sélectionner un sous-groupe et d'en recommencer l'analyse.

Dans le cas d'objets normés $W_k^x/\|W_k^x\|_{H.S}$, on a une propriété analogue à la propriété 2.7 en remplaçant α_k par $\alpha_k/\|W_k^x\|_{H.S}$.

Commentaires :

Bien sûr, les coordonnées des individus de chacun des tableaux pour la représentation des trajectoires sont différentes des composantes t_k calculées ici.

Rappelons que les trajectoires proposées par LAVIT (1988) sont obtenues par $\{W_k^x Dc_a/\sqrt{\lambda_{ca}} = W_k^x Du_{ca}/\lambda_{ca}\}_{k=1,\dots,K}$ (resp. $\{W_k^x Dc_a/\sqrt{\lambda_{ca}\|W_k^x\|_{H.S}} = W_k^x Du_{ca}/(\lambda_{ca}\|W_k^x\|_{H.S})\}_{k=1,\dots,K}$) où a est le numéro de l'axe.

Ici, les trajectoires sont données par $\{t_k = W_k^x Du_c/\sqrt{\mu_{ka}}\}_{k=1,\dots,K}$ quelque soit l'objet utilisé. Si l'on considère l'égalité des coefficients $\{\alpha_k\}_k$ (resp. $\{\alpha_k/\|W_k^x\|_{H.S}\}_k$) du compromis alors elle sont données par $\{t_k = W_k^x Du_{ca}/\sqrt{\mu_{ka}}\}_{k=1,\dots,K}$. Alors, si μ_{ka} est égal à λ_{ca}^2 , les individus du tableau k auront les mêmes positions dans ces trajectoires que pour les trajectoires proposées par Lavit.

Le nombre d'étapes A ne peut dépasser le plus petit des rangs des tableaux, car on ne peut avoir plus de composantes orthogonales que le plus petit des rangs des tableaux.

2.5.2 STATIS dual et ACIMO

Les tableaux X_k sont de taille $l_k \times J$. L'ACIMO travaille sur des tableaux ayant le même nombre de lignes. L'idée est donc d'utiliser l'ACIMO sur les triplets $\{(X'_k, D_k, Id_J)\}_k$ et $(X'_c, Diag(D_k), Id_J)$, où les tableaux X_k auront été préalablement pré-traités. Les tableaux X'_k , pour tout k , et X'_c ne sont ni centrés ni réduits pour les métriques correspondantes.

La procédure est exactement la même qu'avec STATIS et la propriété 2.7 peut être adaptée :

Propriété 2.8 *Si les tableaux sont pondérés uniformément et si les coefficients α_k , $k = 1, \dots, K$, sont tous égaux à une constante β , alors les composantes u_a de l'ACIMO de X'_1, \dots, X'_K et X'_c avec les métriques correspondantes, sont les axes principaux, non normés à l'unité, du tableau-compromis X_c .*

Démonstration :

Il suffit de reprendre la démonstration de la propriété 2.7 en remplaçant les tableaux X_k par leur transposé, Y_c par X_c , D par Id_J et Q_k^x par D_k . Alors les opérateurs $W_k^x D$ et WD deviennent les opérateurs V_k^x et V . —■—

Notons que si $\alpha_k = \beta$ pour $k = 1, \dots, K$, alors nécessairement $\alpha_k = 1/\sqrt{K}$.

Les données de croissances de Sempé (SEMPÉ, PEDRON & ROY-PERNOT, 1979) sont traitées dans le chapitre 1 de la troisième partie pour illustrer tout ce qui a été dit à propos de STATIS et STATIS dual.

Chapitre 3

Analyse Conjointe de plusieurs cubes ou multi-tableaux

Les deux chapitres précédents concernaient l'analyse d'un cube et l'analyse de plusieurs tableaux. Celui-ci concerne l'analyse conjointe de plusieurs cubes ou multi-tableaux. Nous avons regroupé dans un même chapitre les techniques cubiques et multi-blocs car elles sont si peu nombreuses que l'on ne pouvait pas en faire deux chapitres.

Nous présentons deux méthodes récentes apparues depuis 1999, Multiway Multiblock Covariates Component models (MWMBcovC) et STATICO, cette dernière étant une méthode française basée sur la méthode STATIS (§2.1). MWMBcovC concerne l'analyse conjointe de plusieurs cubes ayant le premier mode en commun (les individus). Aucune application n'a été trouvée dans la littérature. STATICO permet d'étudier la structure commune à deux multi-tableaux ayant en commun les variables.

Il n'existe pas vraiment de méthodes de type PLS pour ces situations, à part la PLS Multi-Bloc (MBPLS) que nous décrivons dans le chapitre 5.2 dans le cas de la régression, situation où elle est le plus souvent utilisée en pratique.

Nous proposons ensuite deux nouvelles méthodes basées chacune sur l'optimisation d'un critère sous contraintes. Nous avons appelé la première DOuble - Analyse Conjointe de Tableaux, DO-ACT, en référence à STATIS qui est aussi connue sous le nom de ACT. Il s'agit pour DO-ACT de chercher la structure commune entre deux multi-tableaux mesurés sur les mêmes individus. Une application de cette méthode est faite dans la partie III-5.

La seconde méthode, Analyse de Co-Inertie Multiple Orthogonale Généralisée, ACIMOG, est dédiée à la recherche d'une structure commune entre deux multi-tableaux mesurés sur les mêmes individus. Cette méthode généralise en quelque sorte l'ACIMO (§2.4). Contrairement à DO-ACT, l'ACIMOG est une méthode «component-wise» à la manière de PLS.

3.1 Multiway Multiblock Covariates Component models : MWMB-covC

SMILDE, WESTERHUIS & BOQUÉ (2000) proposent un ensemble de méthodes basées sur la méthode pour deux tableaux Principal Covariates Regression (DE JONG & KIERS, 1992) pour différentes situations : analyse d'un cube, analyse conjointe de cubes, régression entre deux

cubes et régression entre plusieurs cubes. Dans le chapitre 4.5 nous présentons la méthode concernant la régression entre deux cubes. Ici, nous allons exposer la méthode permettant l'analyse conjointe de plusieurs cubes. Nous la noterons MWMBcovC, pour MultiWay Multi-Block Covariates Component.

Ainsi, le but est ici d'analyser un ensemble de L cube \underline{X}_l de dimension $I \times J_l \times K_l$. MWMBcovC, comme dans les méthodes d'analyse cubique telles que PARAFAC (§1.2) et TUCKER3 (§1.1), va calculer simultanément P_l composantes par cubes et par modes, regroupées dans T_l ($I \times P_l$), B_l ($J_l \times Q_l$) et C_l ($K_l \times R_l$). Comme le but est de faire une analyse conjointe des cubes, MWMBcovC va aussi chercher un ensemble de A super-composantes T_T résumant les composantes T_l .

L'utilisateur peut choisir de modéliser chaque cube \underline{X}_l par un modèle de type TUCKER1, TUCKER2, TUCKER3 ou PARAFAC au travers d'une matrice Z_l qui va s'exprimer en fonction d'une matrice noyau (dépliée) G_l et des composantes B_l et C_l de façon à obtenir le modèle choisi.

MWMBcovC travaille sur les cubes dépliés. Notons X_l , la matrice de taille $I \times J_l K_l$ constituée des tranches frontales de \underline{X}_l juxtaposées. De plus, notons $T = [T_1 \ \cdots \ T_L]$, $I \times (\sum_{l=1}^L P_l)$.

La méthode MWMBcovC est basée sur la minimisation de f en W_l par les moindres carrés alternés :

$$f(W_1, \dots, W_L, W) = \sum_{l=1}^L \alpha_l \|X_l - X_l W_l Z_l'\|^2 + \alpha_T \|T - T W Z_T'\|^2.$$

Les matrices W et W_l sont les matrices telles que $T_l = X_l W_l$ et $T = T W Z_T' + E_T$, E_T étant un terme d'erreur. $T W = T_T$ est une matrice de super-composantes, combinaison linéaire des composantes contenues dans T_l . Les coefficients α_l et α_T sont positifs et compris entre 0 et 1. De plus, il doivent être choisis tels que $\alpha_T + \sum_{l=1}^L \alpha_l = 1$.

Le problème est donc de minimiser les erreurs d'ajustement des X_l et de la matrice des A super-composantes T_T , l'optimisation se faisant sur les matrices W_l et W qui sont en quelques sorte des matrices de vecteurs de poids pondérant les variables dans l'expression des composantes T_l et T_T .

Afin d'établir un algorithme de type moindres carrés alternés, les auteurs montrent que la minimisation en W revient à minimiser $\alpha_T \|T - T W Z_T'\|^2$ et que la minimisation en W_l revient à minimiser $\|F_l - X_l W_l H_l\|^2$ où $F_l = \left[\sqrt{\alpha_l} X_l \quad \sum_{l' \neq l} X_{l'} W_{l'} (Id_{\sum_{i=1}^L P_i} - W Z_T') [l'] \right]$, de taille $I \times (\sum_{i=1}^L P_i + J_l K_l)$, et $H_l = \left[\sqrt{\alpha_l} Z_l' \quad \sqrt{\alpha_T} (Id_{\sum_{i=1}^L P_i} - W Z_T') [l] \right]$, de dimension $P_l \times (\sum_{i=1}^L P_i + J_l K_l)$. D'où l'algorithme ci-après.

Il faut commencer par choisir le type des modèles pour \underline{X}_l :
 si l'on désire un modèle TUCKER3, on choisira $Z_l' = G_l (C_l \otimes B_l)'$, G_l étant la juxtaposition des tranches frontales de la matrice noyau associée au modèle TUCKER3 de \underline{X}_l , \underline{G}_l , $P_l \times Q_l \times R_l$. $G_l = T_l' X_l (C_l \otimes B_l)$.
 Si l'on désire un modèle TUCKER2, on choisira $Z_l' = G_l (Id_{K_l} \otimes B_l)'$, $G_l = T_l' X_l (Id_{K_l} \otimes B_l)$. \underline{G}_l est de dimension $P_l \times Q_l K_l$.

Si l'on désire un modèle TUCKER1, on choisira $Z_l' = G_l(Id_{K_l} \otimes Id_{J_l})' = G_l$, $G_l = T_l' X_l (Id_{K_l} \otimes Id_{J_l}) = T_l' X_l$. G_l est de dimension $P_l \times J_l \times K_l$.

Si l'on désire un modèle PARAFAC, on choisira $Z_l' = I_{P_l}(C_l \otimes B_l)'$, où I_{P_l} est la juxtaposition des tranches frontales du cube identité de dimension $P_l \times P_l \times P_l$.

L'algorithme est donc :

étape 0 : initialisation de T_l , B_l , C_l , T_T

T_l : matrice des P_l premiers vecteurs singuliers à gauche de X_l .

$$W_l = X_l^+ T_l.$$

T_T : matrice des A premiers vecteurs singuliers à gauche de T .

$$W = T^+ T_T.$$

B_l et C_l : matrice des Q_l et R_l premiers vecteurs singuliers à gauche du cube \underline{X} déplié dans le sens adéquat.

étape 1 : calculer G_l

si TUCKER1 : $G_l = T_l' X_l$

si TUCKER2 : $G_l = T_l' X_l (Id_{K_l} \otimes B_l)$

si TUCKER3 : $G_l = T_l' X_l (C_l \otimes B_l)$

si PARAFAC : $G_l = I_{P_l}$

en déduire Z_l .

étape 2 : calculer f

calcul de la somme des carrés des écarts actuels :

$$f_{old} = \sum_{l=1}^L \alpha_l \|X_l - T Z_l'\|^2 + \alpha_T \|T - T_T Z_T'\|^2$$

étape 3 : pour $l = 1, \dots, L$, faire

$$W_l = X_l^+ F_l H_l^+$$

$$T_l = X_l W_l$$

B_l et C_l mis à jour par une étape TUCKALS (dans le cas de modèles de TUCKER) ou par l'actualisation de PARAFAC.

G_l mis à jour comme dans l'étape 1.

$$T = [T_1 \dots T_L]$$

$$Z_T = (T_T^+ T)'$$

étape 4 : actualisation de W

$$W = T^+ T (Z_T')^+$$

étape 5 : actualisation de T_T

$$T_T = T W$$

étape 6 : calcul de f

$$f = \sum_{l=1}^L \alpha_l \|X_l - T Z_l'\|^2 + \alpha_T \|T - T_T Z_T'\|^2$$

Si $f - f_{old} < \epsilon$, on s'arrête. Sinon, on retourne à l'étape 3.

—■—

Les auteurs disent que chaque étape de l'algorithme fait décroître f . Pour cette raison, à l'étape 6, on ne prend pas la valeur absolue de la différence entre f et f_{old} pour tester la convergence.

Si les modèles sont tous des modèles de TUCKER, les matrices T_l de composantes peuvent être choisies orthonormales sans modifier l'ajustement des modèles, comme dans TUCKER3. Par contre, si les modèles sont de type PARAFAC, ce n'est plus le cas. Les super-composantes contenues dans T_T peuvent toujours être choisies orthogonales.

Les auteurs recommandent de choisir les paramètres P_l , Q_l et R_l en ajustant les modèles séparément sur chacun des \underline{X}_l . Effectuer des validations-croisées pour les déterminer dans MWMBcovC serait beaucoup trop long. Les poids α_l doivent être choisis de façon à donner de l'importance à certains cubes en fonction de l'information obtenue à partir des modèles séparés. Par contre, le nombre de super-composantes, A , et le coefficient α_T doivent être choisis par validation-croisée.

Dans la même publication, SMILDE et al. (2000) exposent le même type d'algorithme pour les cas de regression multi-bloc (Multiblock covariates regression models, MBcovR) et de régression mettant en jeu plusieurs cubes (Multiway multiblock covariates regression models, MWMBcovR) et en donnent deux exemples d'applications dont les résultats sont comparés à la Multi-Block Regression (MBPLS) (WANGEN & KOWALSKI, 1988), (voir §5.2). SMILDE & KIERS (1999) ont auparavant développé le cas de la régression entre deux cubes, Multiway covariates regression models, MWcovR, que nous décrivons dans le chapitre 4.5.

En ce qui concerne l'analyse conjointe, on peut trouver dans (SMILDE et al., 2000) une version simplifiée de l'algorithme que nous venons de présenter pour le cas de l'analyse conjointe de plusieurs matrices, ce qu'ils appellent Multiblock covariates Components models, MBcovC, et dans (SMILDE & KIERS, 1999), la version pour l'analyse d'un seul cube, qui ressemble beaucoup à PARAFAC (§1.2). Nous avons décrit cette méthode car les techniques d'analyse de plusieurs multi-tableaux ou cubes sont rares. Mais nous pensons qu'elle est un peu compliquée et demande beaucoup de suppositions *a priori*.

3.2 STATIs et Co-inertie : STATICO

La méthode STATICO a été proposée par SIMIER, BLANC, PELLEGRIN & NANDRIS (1999) comme une méthode combinant la méthode STATIS et l'Analyse de Co-Inertie, ACI. STATICO s'applique à deux ensembles de tableaux de même longueur $\{X_k\}_k$ de dimension $l_k \times J$ et $\{Y_k\}_k$ de dimension $l_k \times M$.

Commençons par rappeler ce qu'est l'Analyse de Co-inertie et la méthode «STATIS sur tableaux» appelée encore Analyse Triadique Partielle.

3.2.1 Analyse de Co-Inertie et Analyse Triadique Partielle

L'Analyse de Co-Inertie, ACI, a été proposée par CHESSEL & MERCIER (1993). Il s'agit d'étudier la structure commune à deux tableaux X et Y mesurés sur mes mêmes l observations. A ces tableaux sont associées des métriques Q^x et Q^y dans les espaces des individus \mathbb{R}^J et \mathbb{R}^M , et une métrique D dans l'espace des variables \mathbb{R}^l .

L'ACI est l'analyse du triplet $(Y'DX, Q^x, Q^y)$. Ainsi, les axes c_a et u_a de co-inerties sont obtenus en diagonalisant les matrices symétriques $V = Q^{x1/2} X' D Y Q^y Y' D X Q^{x1/2}$ et $W = Q^{y1/2} Y' D X Q^x X' D Y Q^{y1/2}$, qui ont les mêmes valeurs propres.

L'Analyse Triadique Partielle est en fait une méthode STATIS un peu particulière. Lorsque les données \underline{X} sont cubiques, on peut utiliser la méthode de trois façons :

- STATIS sur les « WD », en considérant \underline{X} comme un ensemble de tableaux appariés en lignes,
- STATIS dual, en considérant \underline{X} comme un ensemble de tableaux appariés en colonnes
- «STATIS sur tableaux» qui est aussi appelée Analyse Triadique Partielle. La différence avec STATIS et STATIS dual est que l'on utilise directement les tableaux X_k sans prendre d'objets représentatifs. Alors l'interstructure consiste à diagonaliser la matrice \mathcal{C} , $\mathbf{K} \times \mathbf{K}$ dont le terme général $C_{k,l}$ est donné par $tr(X'_k D X'_l Q^x)$. Le premier vecteur propre de norme unité α est alors utilisé pour construire le tableau X_c compromis : $X_c = \sum_{k=1}^{\mathbf{K}} \alpha_k X_k$. α est tel que X_c muni des métriques Q^x et D soit d'inertie maximum. Ici, les coefficients α_k ne sont pas forcément tous positifs, contrairement à la méthode STATIS. L'Analyse en Composante Principale de X_c donnent des composantes et des axes permettant de faire des représentations des individus moyens et des variables moyennes, expliquant l'information commune aux \mathbf{K} tableaux X_k . Enfin, l'intrastructure consiste à projeter en supplémentaire les individus (puis les variables) de chacun des X_k sur les plans principaux du compromis.

3.2.2 STATICO : STATIs et CO-inertie

Comme son nom le laisse entendre, STATICO sert à analyser la co-inertie de multi-tableaux. Il s'agit d'appliquer l'Analyse Triadique Partielle aux opérateurs de co-inertie $Z_k = Y'_k D X_k$ de dimension $\mathbf{M} \times \mathbf{J}$ munis des métriques Q^x dans $\mathbb{R}^{\mathbf{J}}$ et Q^y dans $\mathbb{R}^{\mathbf{M}}$.

Lors de l'étape de l'interstructure, la matrice \mathcal{C} à diagonaliser a pour terme général $tr(Z'_k Q^y Z'_l Q^x) = tr(X'_k D Y_k Q^y Y'_l D X_l Q^x)$. Les vecteurs propres de \mathcal{C} , de norme égale à la racine carrée de la valeur propre associée, donnent une représentation euclidienne des co-structures des paires (X_k, Y_k) .

Le premier vecteur propre de norme unité, α , de \mathcal{C} permet l'expression du compromis, $\mathbf{M} \times \mathbf{J}$, $Z_c = \sum_{k=1}^{\mathbf{K}} \alpha_k Z_k = \sum_{k=1}^{\mathbf{K}} \alpha_k Y'_k D X_k$. Z_c est la matrice de co-inertie $Y'_c \text{Diag}(D) X_c$ de deux tableaux compromis X_c et Y_c :

$$X_c = \begin{bmatrix} \sqrt{\alpha_1} X_1 \\ \vdots \\ \sqrt{\alpha_{\mathbf{K}}} X_{\mathbf{K}} \end{bmatrix},$$

$$Y_c = \begin{bmatrix} \sqrt{\alpha_1} Y_1 \\ \vdots \\ \sqrt{\alpha_{\mathbf{K}}} Y_{\mathbf{K}} \end{bmatrix}.$$

$\text{Diag}(D)$ est la matrice diagonale par bloc, où le $k^{\text{ième}}$ bloc vaut D .

L'analyse du compromis se fait par l'analyse de co-inertie de (Z_c, Q^y, Q^x) . Les composantes associées donnent des représentations moyennes des variables des Y_k et X_k .

L'intrastructure donne les représentations des variables pour chaque couple (X_k, Y_k) c'est-à-dire les représentations graphiques des lignes et des colonnes de Z_k . Pour ce faire, la procédure utilisée est celle utilisée dans l'intrastructure de la méthode STATIS (chapitre 2.1), à savoir une technique se rapprochant de la projection de points en supplémentaires.

3.3 DOuble Analyse Conjointe de Tableaux : DO-ACT

Nous présentons ici une nouvelle méthode que nous avons appelée DOuble Analyse Conjointe de Tableaux, DO-ACT*. Rappelons que ACT est aussi un nom désignant la méthode STATIS (chap. 2.1). DO-ACT est en quelque sorte un «double STATIS». Elle s'applique à deux ensembles de tableaux $\{X_k\}_k$ et $\{Y_n\}_n$ et permet d'étudier leur proximité au moyen de compromis.

Elle se décompose également en trois étapes. La première permet de résumer les deux ensembles de tableaux à l'aide de deux compromis, la seconde étape donne une représentation optimale des individus pour chaque ensemble, et la dernière fournit des représentations globales de chaque tableau.

3.3.1 La méthode

DO-ACT, comme la méthode STATIS, travaille avec les opérateurs des produits scalaires entre individus. On note $W_k^x D = X_k Q_k^x X_k' D$ et $W_n^y D = Y_n Q_n^y Y_n' D$ les opérateurs associés respectivement aux tableaux $\{X_k\}_k$ et $\{Y_n\}_n$. L'idée est de trouver deux opérateurs compromis $W_X D$ et $W_Y D$, résumant respectivement les ensembles $\{X_k\}_k$ et $\{Y_n\}_n$ munis des métriques $\{Q_k^x\}_k$, $\{Q_n^y\}_n$ et D des espaces \mathbb{R}^{J_k} , \mathbb{R}^{M_n} et \mathbb{R}^I . Chaque compromis est une combinaison linéaire des opérateurs correspondants aux tableaux de l'ensemble qu'il résume, en maximisant leur produit scalaire $tr(W_X D W_Y D)$.

Ainsi, DO-ACT est définie par :

Définition 3.1 *DO-ACT est la recherche des opérateurs compromis $W_X D = \rho \sum_{k=1}^K \pi_k \alpha_k W_k^x D$ et $W_Y D = \tau \sum_{n=1}^N \omega_n \beta_n W_n^y D$ maximisant le critère $C_4(\alpha, \beta) = tr(W_X D W_Y D)$ sous les contraintes $\|\alpha\|^2 = \sum_{k=1}^K \alpha_k^2 = 1$ et $\|\beta\|^2 = \sum_{n=1}^N \beta_n^2 = 1$, où ρ et τ désignent des coefficients de normalisation et $\Pi = diag(\pi_k)$ et $\Omega = diag(\omega_n)$ sont deux matrices diagonales de poids choisis a priori pour pondérer les opérateurs.*

Sans perte de généralité, nous supposons dorénavant que les coefficients ρ et τ sont égaux à 1. Si tel n'est pas le cas, on remplace dans les expressions $W_X D$ par $W_X D / \|W_X D\|_{H.S}$ et $W_Y D$ par $W_Y D / \|W_Y D\|_{H.S}$. Les solutions sont déterminées par la méthode des multiplicateurs de Lagrange.

Proposition 3.1 *Les solutions de la maximisation du critère $C_4(\alpha, \beta) = tr(W_X D W_Y D)$ sous les contraintes $\|\alpha\|^2 = \sum_{k=1}^K \alpha_k^2 = 1$ et $\|\beta\|^2 = \sum_{n=1}^N \beta_n^2 = 1$ sont solutions du problème aux valeurs propres :*

$$\Omega C' \Pi^2 C \Omega \beta = \mu^2 \beta \quad (3.1)$$

$$\Pi C \Omega^2 C' \Pi \alpha = \mu^2 \alpha \quad (3.2)$$

$$\beta' \beta = 1 \quad (3.3)$$

$$\alpha' \alpha = 1. \quad (3.4)$$

Le maximum du critère C_4 est égal à μ .

* publication soumise : Vivien, M. & Sabatier, R. A generalization of STATIS-ACT strategy : DO-ACT for two multiblocks tables.

Démonstration :

En notant C la matrice de dimension $\mathbf{K} \times \mathbf{N}$ définie par les produits scalaires entre opérateurs, $C(k,n) = tr(W_k^x DW_n^y D)$, on a :

$$\begin{aligned} C_4(\alpha,\beta) = tr(W_X DW_Y D) &= \sum_{k=1}^{\mathbf{K}} \sum_{n=1}^{\mathbf{N}} \pi_k \alpha_k \omega_n \beta_n tr(W_k^x DW_n^y D) \\ &= \alpha' \Pi C \Omega \beta \\ &= \beta' \Omega C' \Pi \alpha. \end{aligned}$$

Le problème revient donc à maximiser une forme bilinéaire sous contraintes. La méthode des multiplicateurs de Lagrange permet d'en trouver les solutions. Le Lagrangien \mathcal{L} est

$$\mathcal{L}(\alpha,\beta) = \alpha' \Pi C \Omega \beta - \frac{\mu}{2} (\|\alpha\|^2 - 1) - \frac{\nu}{2} (\|\beta\|^2 - 1).$$

Les équations normales associées sont :

$$\nabla_{\alpha} \mathcal{L} = \Pi C \Omega \beta - \mu \alpha = 0, \quad (3.5)$$

$$\nabla_{\beta} \mathcal{L} = \Omega C' \Pi \alpha - \nu \beta = 0, \quad (3.6)$$

$$\alpha' \alpha - 1 = 0, \quad (3.7)$$

$$\beta' \beta - 1 = 0. \quad (3.8)$$

En multipliant à gauche (3.5) par α' et en utilisant la contrainte (3.7), on obtient $\mu = \alpha' \Pi C \Omega \beta$. De la même manière, avec les équations (3.6) et (3.8), $\nu = \mu = \alpha' \Pi C \Omega \beta$. Ainsi, μ et ν sont égaux à l'optimum du critère C_4 .

L'équation (3.5) donne $\alpha = \Pi C \Omega \beta / \mu$. La contrainte (3.7) implique $\mu = \|\Pi C \Omega \beta\|$: μ et ν sont positifs.

En utilisant l'expression de α ci-dessus dans l'équation (3.6), (3.2) est obtenue. En procédant de même avec l'expression de β donnée par (3.6) et avec l'équation (3.5), (3.1) est obtenue.

Ainsi, μ^2 est la plus grande valeur propre de $\Omega C' \Pi^2 C \Omega$, et μ est assurément un maximum local ou global sur l'espace défini par les contraintes. —■—

En pratique, il n'est pas nécessaire de diagonaliser les deux matrices dans (3.1) et (3.2). Il suffit de diagonaliser une des deux matrices et de calculer l'autre vecteur (α ou β) en utilisant l'équation (3.5) ou (3.6).

Si les matrices de pondérations Π et Ω sont les matrices identité, les matrices à diagonaliser deviennent $C' C$ et $C C'$.

Les matrices $\Omega C' \Pi^2 C \Omega$ et $\Pi C \Omega^2 C' \Pi$ sont non-négatives car la trace d'un produit d'opérateurs est positive. Ainsi, l'application du théorème de Perron-Frobenius entraîne la positivité de tous les éléments des vecteurs α et β , premiers vecteurs propres de $\Omega C' \Pi^2 C \Omega$ et $\Pi C \Omega^2 C' \Pi$. Ceci engendre que les compromis $W_X D$ et $W_Y D$ sont semi-définis positifs.

Les équations de la Proposition 3.1 sont similaires à celles produites par l'Inter-Batterie de

Tucker. Rappelons que l'Analyse Inter-Batterie de Tucker de deux tableaux X et Y consiste en la recherche de deux ensembles de composantes $\{c_a^x = XQ^x a_a\}_a$ et $\{c_a^y = YQ^y b_a\}_a$ qui maximisent $cov(c_a^x, c_a^y)$ sous les contraintes de normes unité pour les axes a_a et b_a . Les solutions c_a^x et c_a^y sont les $a^{\text{ièmes}}$ vecteurs propres des produits d'opérateurs $W_X D W_Y D$ et $W_Y D W_X D$ associés à la même valeur propre λ_a^2 , λ_a étant l'optimum du critère. Alors que les axes forment des systèmes orthonormés pour les métriques Q_X et Q_Y , les systèmes de composantes $\{c_a^x\}_a$ et $\{c_a^y\}_a$ ne sont pas D -orthogonaux.

DO-ACT se décompose en trois étapes, comme la méthode STATIS (LAVIT, 1988), appelées *interstructure*, *compromis* et *intrastructure*.

3.3.2 Interstructure - compromis - intrastructure

Soient α_a et β_a les $a^{\text{ièmes}}$ vecteurs propres de $\Pi C \Omega^2 C' \Pi$ et $\Omega C' \Pi^2 C \Omega$ associés à μ_a^2 . Alors le graphique donné par $(\mu_a \alpha_a, \mu_{a'} \alpha_{a'})$ donne une représentation euclidienne des opérateurs $W_k^x D$ sur le plan (a, a') . De même, $(\mu_a \beta_a, \mu_{a'} \beta_{a'})$ donne une représentation euclidienne des opérateurs $W_n^y D$ sur le plan (a, a') .

Ces deux représentations définissent l'*interstructure* de DO-ACT. Cela permet d'étudier les proximités entre les X_k d'une part et les Y_n d'autres part. De plus, si $K = N$ et décrivent un même phénomène (les mêmes dates dans le temps par exemple), les deux plans peuvent être comparés.

Proposition 3.2 *Les deux matrices X_c et Y_c définies ci-dessous sont les tableaux compromis dont $W_X D$ et $W_Y D$ sont les opérateurs associés :*

$$X_c = [\sqrt{\rho\pi_1\alpha_1}X_1 \quad \dots \quad \sqrt{\rho\pi_K\alpha_K}X_K] \quad (3.9)$$

$$Y_c = [\sqrt{\tau\omega_1\beta_1}Y_1 \quad \dots \quad \sqrt{\tau\omega_N\beta_N}Y_N], \quad (3.10)$$

où $\alpha_1, \dots, \alpha_K$ et β_1, \dots, β_N sont les éléments du premier vecteur propre α de $\Pi C \Omega^2 C' \Pi$ et du premier vecteur propre β de $\Omega C' \Pi^2 C \Omega$ associés à la plus grande valeur propre μ_1^2 .

Le démonstration est sans équivoque.

Cette proposition montre que la stratégie de construction des compromis optimale de DO-ACT n'est pas l'analyse Inter-batterie de Tucker (TUCKER, 1958b) appliquée aux tableaux conjoints $X = [X_1 \dots X_K]$ et $Y = [Y_1 \dots Y_N]$. Par contre, il se peut, par l'utilisation de DO-ACT, que les coefficients de pondération $\sqrt{\rho\pi_k\alpha_k}$ et $\sqrt{\tau\omega_n\beta_n}$ soient égaux à 1, sans être choisis *a priori*.

Proposition 3.3 *Si $\|W_X D\|_{H,S}^2 = 1$ et $\|W_Y D\|_{H,S}^2 = 1$, les solutions de la proposition 3.1 sont aussi les solutions des deux problèmes d'optimisations suivants :*

- $\max_{\alpha, \beta} (RV(X_c, Y_c))$ et l'optimum est μ ,
- $\min_{\alpha, \beta} (\|W_X D - W_Y D\|_{H,S}^2)$ et l'optimum est $2(1 - \mu)$.

Démonstration :

La preuve est directe car $RV(X_c, Y_c) = tr(W_X D W_Y D)$ sous les hypothèses de la proposition et $\|W_X D - W_Y D\|^2 = 2(1 - tr(W_X D W_Y D))$. —■—

Définition 3.2 Les deux inrastructures sont données par les représentations des observations dans l'Analyse Inter-Batterie de Tucker (TUCKER, 1958b) de X_c et Y_c .

De plus, les comparaisons globales des observations vues par les $K + N$ tableaux, peuvent être réalisées en représentant les observations de tous les tableaux sur les composantes de l'analyse Inter-Batterie de Tucker.

3.3.3 Cas particuliers

Le premier cas particulier est le cas où les deux ensembles de tableaux $\{X_k\}_k$ et $\{Y_n\}_n$ sont égaux. Alors, on retrouve les solutions de la méthode STATIS.

Si $N = 1$ alors $W_X D = \sum_{k=1}^K \pi_k \text{tr}(W_k^x W^y D) W_k^x D$: l'opérateur $W_k^x D$ est pondéré par la covariance vectorielle entre X_k et Y .

Il est intéressant de comprendre la différence entre la méthode STATICO (SIMIER et al., 1999), en supposant les deux ensembles de tableaux cubiques. La première différence est que l'on ne travaille pas dans le même espace. DO-ACT travaille dans l'espace des opérateur de \mathbb{R}^1 , avec les opérateurs « WD », alors que STATICO (§3.2) travaille avec les tableaux de co-inertie $Y_k' D X_k$ qui sont de dimension $M \times P$. Le compromis de DO-ACT cherche l'information au travers de produits croisée entre les $W_k^x D$ et $W_k^y D$ alors que STATICO le fait au travers des produits croisés entre $Y_k' D X_k$.

3.4 Analyse de Co-Inertie Multiple Orthogonale Généralisée: ACIMOG

3.4.1 Problème

Afin de généraliser l'ACIMO au cas de deux suites de tableaux, X_1, \dots, X_K et Y_1, \dots, Y_N , nous avons choisi de maximiser le critère C_2 :

$$C_2(\{a_k\}, \{b_n\}) = \sum_{k=1}^K \sum_{n=1}^N \text{cov}(t_k, u_n),$$

où $t_k = X_k Q_k^x a_k$ et $u_n = Y_n Q_n^y b_n$, avec les contraintes de norme unité sur les axes a_k et les axes b_n .

La raison pour laquelle nous avons choisi de maximiser la double somme des covariances et non la double somme des covariances au carré comme dans l'ACIMO est que le critère choisi est plus simple à optimiser.

De plus, avec ce critère C_2 , nous allons généraliser en quelque sorte l'ACIMO. Cette méthode, comme toujours, va se résoudre pas à pas : à chaque étape a , nous allons calculer des ensembles d'axes $\{a_k\}_k$ et $\{b_n\}_n$ normés pour la métrique correspondante et des ensembles de composantes $\{t_k = X_k Q_k^x a_k\}_k$ et $\{u_n = Y_n Q_n^y b_n\}_n$.

Ainsi, nous allons présenter trois méthodes basées sur l'optimisation de C_2 , appelées Analyse de Co-Inertie Multiple Orthogonale Généralisée, ACIMOG.

3.4.2 Solutions d'ordre 1

Nous allons décrire pour commencer la première étape, qui est commune aux différentes généralisations.

Les objets issus de l'étape courante a seront indicés par a , sauf pour cette première étape, afin d'alléger les notations.

Pour résoudre ce problème, nous utilisons comme précédemment la méthode des multiplicateurs de Lagrange.

Le lagrangien associé est :

$$\begin{aligned} \mathcal{L}(\{a_k\}_k, \{b_n\}_n) &= \sum_{k=1}^K \sum_{n=1}^N \text{cov}(X_k Q_k^x a_k, Y_n Q_n^y b_n) \\ &\quad - \sum_{k=1}^K \frac{1}{2} \mu_k (a'_k Q_k^x a_k - 1) - \sum_{n=1}^N \frac{1}{2} \lambda_n (b'_n Q_n^y b_n - 1), \end{aligned}$$

où $\mu_k/2$, $k = 1, \dots, K$ et $\lambda_n/2$, $n = 1, \dots, N$ sont les multiplicateurs de Lagrange associés aux contraintes.

Ainsi, nous obtenons les équations normales :

$$\nabla_{a_k} \mathcal{L} = \sum_{n=1}^N Q_k^x X'_k D Y_n Q_n^y b_n - \mu_k Q_k^x a_k = 0, \quad k = 1, \dots, K \quad (3.11)$$

$$\nabla_{b_n} \mathcal{L} = \sum_{k=1}^K Q_n^y Y'_n D X_k Q_k^x a_k - \lambda_n Q_n^y b_n = 0, \quad n = 1, \dots, N \quad (3.12)$$

$$a'_k Q_k^x a_k - 1 = 0, \quad k = 1, \dots, K \quad (3.13)$$

$$b'_n Q_n^y b_n - 1 = 0, \quad n = 1, \dots, N \quad (3.14)$$

La nullité des dérivées premières de \mathcal{L} et les contraintes (3.13) et (3.14) permettent d'obtenir les expressions de μ_k , $k = 1, \dots, K$, λ_n , $n = 1, \dots, N$ et de l'optimum du critère C_2 :

$$\mu_k = \sum_{n=1}^N a'_k Q_k^x X'_k D Y_n Q_n^y b_n = \text{cov}(t_k, \sum_{n=1}^N u_n), \quad k = 1, \dots, K \quad (3.15)$$

$$\lambda_n = \sum_{k=1}^K a'_k Q_k^x X'_k D Y_n Q_n^y b_n = \text{cov}(u_n, \sum_{k=1}^K t_k), \quad n = 1, \dots, N \quad (3.16)$$

$$C_{2opt.} = \sum_{k=1}^K \sum_{n=1}^N a'_k Q_k^x X'_k D Y_n Q_n^y b_n = \sum_{n=1}^N \lambda_n = \sum_{k=1}^K \mu_k. \quad (3.17)$$

La combinaison des équations (3.11) et (3.13) puis de (3.12) et (3.14), nous donne l'expression

des axes a_k , $k = 1, \dots, K$ et de b_n , $n = 1, \dots, N$:

$$a_k = \frac{1}{\mu_k} \sum_{n=1}^N X_k' D Y_n Q_n^y b_n, \quad k = 1, \dots, K \quad (3.18)$$

$$b_n = \frac{1}{\lambda_n} \sum_{k=1}^K Y_n' D X_k Q_k^x a_k, \quad n = 1, \dots, N. \quad (3.19)$$

Remarque : Les coefficients μ_k et λ_n , pour tout k et tout n sont non nuls, sinon il n'y aurait plus de contraintes. Notons de plus, qu'ils peuvent être choisis positifs, ce que nous faisons, en raison de la norme unité des axes a_k et b_n . En effet, $\mu_k = \|X_k' D \sum_{n=1}^N u_n\|_{Q_k^x}$, pour $k = 1, \dots, K$ et $\lambda_n = \|Y_n' D \sum_{k=1}^K t_k\|_{Q_n^y}$ pour $n = 1, \dots, N$.

Le coefficient μ_k mesure d'une certaine manière la proximité entre le tableau X_k et l'ensemble des tableaux Y_n . C'est-à-dire, il mesure la proximité entre la structure du tableau X_k et une structure commune à tous les Y_n . De même, le coefficient λ_n mesure la proximité entre la structure du tableau Y_n et une structure commune à tous les X_k .

Ce problème n'a pas de solutions exactes et n'est pas non plus équivalent à une diagonalisation. Pour le résoudre, nous allons proposer un algorithme itératif.

Notons tout d'abord que ces $K + N$ dernières équations, (3.18) et (3.19), peuvent se résumer par les expressions matricielles suivantes :

$$\begin{bmatrix} \mu_1 a_1 \\ \mu_2 a_2 \\ \vdots \\ \mu_K a_K \end{bmatrix} = X' D Y \text{Diag}(Q_n^y) \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_N \end{bmatrix} \quad (3.20)$$

$$\begin{bmatrix} \lambda_1 b_1 \\ \lambda_2 b_2 \\ \vdots \\ \lambda_N b_N \end{bmatrix} = Y' D X \text{Diag}(Q_k^x) \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_K \end{bmatrix}, \quad (3.21)$$

où X et Y sont les tableaux conjoints $[X_1 \ \dots \ X_K]$ et $[Y_1 \ \dots \ Y_N]$. $\text{Diag}(Q_k^x)$ et $\text{Diag}(Q_n^y)$ sont respectivement les matrices diagonales par bloc des métriques Q_k^x et Q_n^y .

3.4.3 L'algorithme

L'algorithme que nous proposons pour déterminer les éléments recherchés est le suivant :

étape 0 : initialisation

Les tableaux X_k , $k = 1, \dots, K$ et Y_n , $n = 1, \dots, N$ sont centrés et réduits éventuellement selon les métriques D .

Nous posons b_n comme étant le premier vecteur singulier à gauche de $Y_n' D X$.

$$a_k = X_k' D Y \text{Diag}(Q_n^y) b \text{ avec } b' = [b'_1 \ \dots \ b'_N].$$

$$a_k = a_k / \|a_k\|_{Q_k^x}.$$

$$\mu_k = b' \text{Diag}(Q_n^y) Y' D X_k Q_k^x a_k.$$

$$\lambda_n = a' \text{Diag}(Q_k^x) X' D Y_n Q_n^y b_n, \text{ avec } a' = [a'_1 \ \dots \ a'_K].$$

$$C_2 \text{opt.} = \sum_k \mu_k.$$

$iter = 0$, indice des itérations.

étape 1 : calcul des axes

$$\begin{aligned}
iter &= iter + 1, \\
b_n &= Y_n' D X \text{Diag}(Q_k^x) a. \\
b_n &= b_n / \|b_n\|_{Q_n^y}. \\
a_k &= X_k' D Y \text{Diag}(Q_n^y) b. \\
a_k &= a_k / \|a_k\|_{Q_k^x}. \\
\mu_k &= b' \text{Diag}(Q_n^y) Y' D X_k Q_k^x a_k. \\
\lambda_n &= a' \text{Diag}(Q_k^x) X' D Y_n Q_n^y b_n. \\
C_{2opt.new} &= \sum_k \mu_k.
\end{aligned}$$

étape 2 : test de convergence

si $|C_{2opt.new} - C_{2opt}| < \epsilon$, (avec $\epsilon = 10^{-10}$ en général[†]), ou si le nombre maximal d'itérations est atteint, on passe à l'étape 3.

Sinon, on pose $C_{2opt.} = C_{2opt.new}$ et on retourne à l'étape 1.

étape 3 : calcul des composantes

$$\begin{aligned}
u_n &= Y_n Q_n^y b_n, \\
t_k &= X_k Q_k^x a_k.
\end{aligned}$$

■

Pour l'initialisation de b , il est aussi possible de prendre le vecteur $\frac{1}{\sqrt{N_m}} \mathbb{1}_{N_m}$.

Croissance de l'algorithme :

Nous allons montrer que cet algorithme est croissant, ce qui assure qu'il s'arrête lorsqu'un maximum local ou local strict est atteint. En changeant le point de départ de l'algorithme, nous avons toujours abouti au même maximum, ce qui laisse croire que le maximum atteint est un maximum global sur la surface définie par les contraintes.

La preuve est faite pour $a = 1$, elle est valable pour les étapes suivantes.

Posons $h_k(a_k) = \mu_k = \sum_{n=1}^N \text{cov}(t_k, u_n)$, où $a_1, \dots, a_{k-1}, a_{k+1}, \dots, a_K, b_1, \dots, b_N$ sont fixés et \bar{a}_k le nouvel a_k dans l'algorithme : $\bar{a}_k = \frac{X_k' D \sum_{n=1}^N Y_n Q_n^y b_n}{\|X_k' D \sum_{n=1}^N Y_n Q_n^y b_n\|_{Q_k^x}}$.

La preuve consiste à montrer que le passage dans la boucle de l'itération (**étape 1**) modifie μ_k de telle manière que $\Delta h_k(a_k) = h_k(\bar{a}_k) - h_k(a_k)$ est non-négatif pour un a_k arbitraire :

$$\begin{aligned}
\Delta h_k(a_k) &= h_k(\bar{a}_k) - h_k(a_k) \\
&= \sum_{n=1}^N u_n' D X_k Q_k^x \frac{X_k' D \sum_{n=1}^N u_n}{\|X_k' D \sum_{n=1}^N u_n\|_{Q_k^x}} - \sum_{n=1}^N u_n' D X_k Q_k^x a_k \\
&= \frac{1}{\|X_k' D \sum_{n=1}^N u_n\|_{Q_k^x}} \left(\|X_k' D \sum_{n=1}^N u_n\|_{Q_k^x}^2 - \|X_k' D \sum_{n=1}^N u_n\|_{Q_k^x} a_k' Q_k^x X_k' D \sum_{n=1}^N u_n \right) \\
&= \|X_k' D \sum_{n=1}^N u_n\|_{Q_k^x} \left(1 - \cos_{Q_k^x}(a_k, X_k' D \sum_{n=1}^N u_n) \right).
\end{aligned}$$

Comme $\cos_{Q_k^x}(a_k, X_k' D \sum_{n=1}^N u_n)$ est le cosinus entre a_k et $X_k' D \sum_{n=1}^N u_n$ au sens de Q_k^x , il est plus petit ou égal à 1. Ce qui implique que h_k croît à chaque itération de l'algorithme.

[†] Nous avons remarqué dans la pratique que 10^{-6} ne suffisait pas pour avoir des objets vraiment stables. Par précaution nous utilisons 10^{-10} .

En notant $h_n(b_n) = \lambda_n$ et $\bar{b}_n = \frac{Y_n' D \sum_{k=1}^K t_k}{\|Y_n' D \sum_{k=1}^K t_k\|_{Q_n^y}}$, on peut montrer de la même manière que $\Delta h_n(b_n)$ est non-négatif.

A chaque fois que a_k ou b_n est mis à jour, le critère croît.

Notons que nous avons choisi de tester la convergence absolue (étape 2 de l'algorithme). En général, la valeur du critère C_{2opt} . est grande (très supérieure à 1) pour les premières valeurs de a , qui sont les étapes qui nous intéressent le plus. Si nous testions la convergence relative, le critère d'arrêt serait $|(C_{2opt.new} - C_{2opt})/C_{2opt.new}| < \epsilon$ et il convergerait moins vite.

Dans la pratique que nous avons de cet algorithme, nous avons toujours eu convergence en moins de 30 itérations et ce quelque soit l'initialisation de b_n . Cela dit, il serait intéressant de voir si l'on ne peut pas prendre un autre critère d'arrêt qui permettrait une convergence plus rapide car à partir de la quatrième itération environ, le critère n'augmente que par centièmes.

3.4.4 Les solutions d'ordre a et propriétés

A partir de maintenant, nous noterons $u_{n,a}$, $t_{k,a}$, $b_{n,a}$, $a_{k,a}$, $\mu_{k,a}$, $\lambda_{n,a}$ et $C_{2opt.a}$ les éléments d'ordre a , pour $a = 1, \dots, A$. Le but est ici d'expliquer la structure commune aux deux ensembles de tableaux et reconstituer chaque tableaux en tenant compte de l'information apportée par les $K + N$ tableaux. Ainsi, nous allons maximiser le critère C_2 en utilisant les résidus de projections D -orthogonales particulières à la place des tableaux X_k et des tableaux Y_n .

On pose $X_k^{(0)} = X_k$, $k = 1, \dots, K$ et $Y_n^{(0)} = Y_n$, $n = 1, \dots, N$.

Nous proposons trois ACIMOG se différenciant par les sous-espaces utilisés.

Définition 3.3 *L'ACIMOG1 d'ordre a , $a = 2, \dots, A$ est l'ACIMOG d'ordre 1 appliquée aux triplets $\{(X_k^{(a-1)}, Q_k^x, D)\}_{k=1, \dots, K}$ et $\{(Y_n^{(a-1)}, Q_n^y, D)\}_{n=1, \dots, N}$ avec $X_k^{(a-1)} = P_{t_{k,a-1}}^\perp X_k^{(a-2)}$, $k = 1, \dots, K$ et $Y_n^{(a-1)} = P_{u_{n,a-1}}^\perp Y_n^{(a-2)}$, $n = 1, \dots, N$, où $P_v^\perp = Id_1 - P_v$ et P_v est le projecteur D -orthogonal sur l'espace engendré par le vecteur v .*

Propriété 3.1 *Avec l'ACIMOG1, les systèmes de composantes $\{t_{k,a}\}_a$, pour $k = 1, \dots, K$ et $\{u_{n,a}\}_a$, pour $n = 1, \dots, N$, sont D -orthogonaux.*

De plus, les systèmes d'axes $\{a_{k,a}\}_a$ sont Q_k^x -orthonormés, pour $k = 1, \dots, K$ et les systèmes d'axes $\{b_{n,a}\}_a$ sont Q_n^y -orthonormés, pour $n = 1, \dots, N$.

Démonstration :

L'orthogonalité des composantes est due à la définition même de l'ACIMOG1 :

$t'_{k,a} D t_{k,a-1} = a'_{k,a} Q_k^x X_k^{(a-1)'} D t_{k,a-1} = a'_{k,a} Q_k^x X_k^{(a-2)'} D P_{t_{k,a-1}}^\perp t_{k,a-1} = 0$ pour tout $a \geq 2$. Les composantes sont ainsi deux à deux orthogonales. On suppose de plus

que $t_{k,a-1}, \dots, t_{k,1}$ sont orthogonales. Pour $l < a$:

$$\begin{aligned} t'_{k,l} D t_{k,a} &= t'_{k,l} D P_{t_{k,a-1}}^\perp P_{t_{k,a-2}}^\perp \dots P_{t_{k,l+1}}^\perp X^{(l)} Q_k^x a_{k,a} \\ &= t'_{k,l} D \sum_{v=l+1}^{a-1} P_{t_{k,v}} X^{(l)} Q_k^x a_{k,a} \\ &= 0, \text{ d'après l'hypothèse d'induction.} \end{aligned}$$

La démonstration est analogue en ce qui concerne les composantes $u_{n,a}$.

En ce qui concerne les axes, nous allons simplement faire les calculs pour les axes $a_k, k = 1, \dots, K$. La démonstration pour les axes $b_n, n = 1, \dots, N$ est identique à l'échange près des éléments.

- Les axes $a_{k,a}, k = 1, \dots, K, a = 1, \dots, A$ sont Q_k^x -normés par définition.
- Montrons l'orthogonalité des axes $\{a_{k,a}\}_a$ pour un k fixé parmi $1, \dots, K$.

D'après l'expression (3.18) et la définition de l'ACIMOG d'ordre $a, a > 1$, (définition 3.3), quelque soit ℓ variant de 1 à $a - 1$:

$$\begin{aligned} \mu_{k,a} \mu_{k,\ell} a'_{k,\ell} Q_k^x a_{k,a} &= a'_{k,\ell} Q_k^x X_k^{(a-1)'} D \sum_{n=1}^N u_{n,a} \\ &= a'_{k,\ell} Q_k^x X_k^{(a-2)'} P_{t_{k,a-1}}^{\perp'} D \sum_{n=1}^N u_{n,a} \\ &= a'_{k,\ell} Q_k^x \prod_{l=\ell}^{a-1} \left(X_k^{(l-1)'} D P_{t_{k,l}}^\perp \right) \sum_{n=1}^N u_{n,a} \\ &= a'_{k,\ell} Q_k^x X_k^{(\ell-1)'} P_{t_{k,\ell}}^{\perp'} \prod_{l=\ell+1}^{a-1} \left(X_k^{(l-1)'} D P_{t_{k,l}}^\perp \right) \sum_{n=1}^N u_{n,a} \\ &= t'_{k,\ell} D P_{t_{k,\ell}}^{\perp'} \prod_{l=\ell+1}^{a-1} \left(X_k^{(l-1)'} D P_{t_{k,l}}^\perp \right) \sum_{n=1}^N u_{n,a} \\ &= 0. \end{aligned}$$

D'où le résultat, car $\mu_{k,a}$ et $\mu_{k,\ell}$ sont non nuls. ■

Définition 3.4 L'ACIMOG2 d'ordre $a, a = 2, \dots, A$ est l'ACIMOG d'ordre 1 appliquée aux triplets $\{(X_k^{(a-1)}, Q_k^x, D)\}_{k=1, \dots, K}$ et $\{(Y_n^{(a-1)}, Q_n^y, D)\}_{n=1, \dots, N}$ avec $X_k^{(a-1)} = P_{t_{a-1}}^\perp X_k^{(a-2)}, k = 1, \dots, K$ et $Y_n^{(a-1)} = P_{u_{a-1}}^\perp Y_n^{(a-2)}, n = 1, \dots, N$, où $t_a = \sum_{k=1}^K t_{k,a}$ et $u_a = \sum_{n=1}^N u_{n,a}$.

Définition 3.5 L'ACIMOG3 d'ordre $a, a = 2, \dots, A$ est l'ACIMOG d'ordre 1 appliquée aux triplets $\{(X_k^{(a-1)}, Q_k^x, D)\}_{k=1, \dots, K}$ et $\{(Y_n^{(a-1)}, Q_n^y, D)\}_{n=1, \dots, N}$ avec $X_k^{(a-1)} = P_{t_{a-1}}^\perp X_k^{(a-2)}, k = 1, \dots, K$ et $Y_n^{(a-1)} = P_{u_{a-1}}^\perp Y_n^{(a-2)}, n = 1, \dots, N$, où $t_a = \sum_{k=1}^K \mu_{k,a} t_{k,a}$ et $u_a = \sum_{n=1}^N \lambda_{n,a} u_{n,a}$.

Les composantes t_a et u_a sont appelées «super-composantes» car elles sont des combinaisons linéaires de toutes les variables de leur ensemble (X ou Y), et aussi respectivement combinaison linéaire des K composantes $t_{k,a}$ et des N composantes $u_{n,a}$.

Dans l'ACIMOG3, la super-composante t_a d'ordre a pondère les composantes partielles par $\mu_{k,a}$. C'est-à-dire que la super-composante t_a tient surtout compte de l'information apportée par les tableaux X_k étant les plus «proches» de l'ensemble des tableaux Y_n . C'est la même chose pour les super-composantes u_a .

Propriété 3.2 *Avec les ACIMOG2 & 3, les systèmes de super-composantes $\{t_a\}_a$ et $\{u_a\}_a$ sont D -orthogonaux.*

Avec ces deux définitions, les systèmes de composantes partielles $\{t_{k,a}\}_a$, et $\{u_{n,a}\}_a$ ne sont plus orthogonaux, ni les systèmes d'axes associés. Par contre, les systèmes de super-composantes $\{t_a\}_a$ et $\{u_a\}_a$ le sont. La démonstration est la même que pour l'orthogonalité des composantes dans l'ACIMOG1. Avec ces méthodes, nous avons «moins d'orthogonalités» ce qui complique énormément l'interprétation. C'est pourquoi nous privilégions la définition 3.3.

3.4.5 Formules de reconstitution

Les orthogonalités présentes dans chacune des ACIMOG permettent d'écrire des modèles, formules de reconstitution, simples. Notons $\mathcal{E}_a^{x,k}$ l'espace sur lequel est projeté le tableau $X_k^{(a-1)}$ lors du passage à l'étape $a + 1$. C'est-à-dire, pour l'ACIMOG1, $\mathcal{E}_a^{x,k} = t_{k,a}$ et pour les ACIMOG 2 et 3 $\mathcal{E}_a^{x,k} = t_a$ pour tout k . De même, notons $\mathcal{E}_a^{y,n}$ l'espace sur lequel est projeté le tableau $Y_n^{(a-1)}$ lors du passage à l'étape $a + 1$: $\mathcal{E}_a^{y,n} = u_{n,a}$ dans le cas de l'ACIMOG1, et $\mathcal{E}_a^{y,n} = u_a$ pour les ACIMOG 2 et 3.

Proposition 3.4 *Les formules de reconstitution des tableaux sont données par :*

$$X_k = \sum_{a=1}^A P_{\mathcal{E}_a^{x,k}} X_k + X_k^{(A)}, \quad (3.22)$$

$$= \sum_{a=1}^A X_k \widehat{\beta}_k^{x,a} + X_k^{(A)}, \text{ avec l'ACIMOG1} \quad (3.23)$$

$$= \sum_{a=1}^A X \widehat{\beta}_k^{x,a} + X_k^{(A)}, \text{ avec les ACIMOG2 et 3.} \quad (3.24)$$

$$Y_n = \sum_{a=1}^A P_{\mathcal{E}_a^{y,n}} Y_n + Y_n^{(A)}, \quad (3.25)$$

$$= \sum_{a=1}^A Y_n \widehat{\beta}_n^{y,a} + Y_n^{(A)}, \text{ avec l'ACIMOG1} \quad (3.26)$$

$$= \sum_{a=1}^A Y \widehat{\beta}_n^{y,a} + Y_n^{(A)}, \text{ avec les ACIMOG2 et 3,} \quad (3.27)$$

avec :

$$\widehat{\beta}_k^{x,a} = \sum_{l=1}^a \delta_l^{x,k} (\mathcal{E}_l^{y,n'} D\mathcal{E}_l^{x,k})^{-1} \mathcal{E}_l^{x,k'} DX_k,$$

$$\widehat{\beta}_n^{y,a} = \sum_{l=1}^a \delta_l^{y,n} (\mathcal{E}_l^{x,n'} D\mathcal{E}_l^{y,n})^{-1} \mathcal{E}_l^{y,n'} DY_n,$$

où :

$$\delta_1^{x,k} = \alpha_1^{x,k},$$

$$\delta_a^{x,k} = \alpha_a^{x,k} - \sum_{l=1}^{a-1} \delta_l^{x,k} (\mathcal{E}_l^{x,k'} D\mathcal{E}_l^{x,k})^{-1} \mathcal{E}_l^{x,k'} DX_k \alpha_a^{x,k} \text{ pour l'ACIMOG1,}$$

$$= \alpha_a^{x,k} - \sum_{l=1}^{a-1} \delta_l^{x,k} (\mathcal{E}_l^{x,k'} D\mathcal{E}_l^{x,k})^{-1} \mathcal{E}_l^{x,k'} DX \alpha_a^{x,k} \text{ pour les ACIMOG2 et 3,}$$

$$\delta_1^{y,n} = \alpha_1^{y,n},$$

$$\delta_a^{y,n} = \alpha_a^{y,n} - \sum_{l=1}^{a-1} \delta_l^{y,n} (\mathcal{E}_l^{y,n'} D\mathcal{E}_l^{y,n})^{-1} \mathcal{E}_l^{y,n'} DY_n \alpha_a^{y,n} \text{ pour l'ACIMOG1,}$$

$$= \alpha_a^{y,n} - \sum_{l=1}^{a-1} \delta_l^{y,n} (\mathcal{E}_l^{y,n'} D\mathcal{E}_l^{y,n})^{-1} \mathcal{E}_l^{y,n'} DY \alpha_a^{y,n} \text{ pour les ACIMOG2 et 3,}$$

et

$$\alpha_a^{x,k} = Q_k^x a_{k,a} \text{ pour l'ACIMOG1,}$$

$$\alpha_a^{x,k'} = \alpha_a^{x'} = [a'_{1,a} \dots a'_{K,a}] \text{Diag}(Q_k^x) \text{ pour l'ACIMOG2,}$$

$$\alpha_a^{x,k'} = \alpha_a^{x'} = [\mu_{1,a} a'_{1,a} \dots \mu_{K,a} a'_{K,a}] \text{Diag}(Q_k^x) \text{ pour l'ACIMOG3,}$$

$$\alpha_a^{y,n} = Q_n^y b_{n,a} \text{ pour l'ACIMOG1,}$$

$$\alpha_a^{y,n'} = \alpha_a^{y'} = [b'_{1,a} \dots b'_{K,a}] \text{Diag}(Q_n^y) \text{ pour l'ACIMOG2,}$$

$$\alpha_a^{y,n'} = \alpha_a^{y'} = [\lambda_{1,a} b'_{1,a} \dots \lambda_{N,a} b'_{N,a}] \text{Diag}(Q_n^y) \text{ pour l'ACIMOG3.}$$

Démonstration :

L'expression des modèles (3.22) et (3.25), sont dus à l'orthogonalité des systèmes $\{\mathcal{E}_a^{x,k}\}_a$ et $\{\mathcal{E}_a^{y,n}\}_a$. Montrons (3.22). Nous avons :

$$\begin{aligned} X_k^{(0)} &= X_k \\ X_k^{(1)} &= X_k - P_{\mathcal{E}_1^{x,k}} X_k \\ X_k^{(2)} &= X_k^{(1)} - P_{\mathcal{E}_2^{x,k}} X_k^{(1)} \\ &\vdots \\ X_k^{(a)} &= X_k^{(a-1)} - P_{\mathcal{E}_a^{x,k}} X_k^{(a-1)}. \end{aligned}$$

En additionnant terme à terme ces équations, nous obtenons $X_k^{(a)} = X_k - \sum_{l=1}^a P_{\mathcal{E}_l^{x,k}} X_k^{(l-1)}$. Grâce aux orthogonalités des composantes données par les propriétés (3.1) et (3.2), cette égalité devient

$$X_k^{(a)} = X_k - \sum_{l=1}^a P_{\mathcal{E}_l^{x,k}} X_k.$$

La démonstration est identique pour la formule de reconstitution de Y_n .

La démonstration de l'expression des coefficients « $\hat{\beta}$ » est similaire à celle de la proposition 5.3 du chapitre 5.6. Les grandes lignes de la démonstration consistent à montrer que $\mathcal{E}_a^{x,k} = X_k^{(a-1)} \alpha_a^{x,k}$ et $\mathcal{E}_a^{y,n} = Y_n^{(a-1)} \alpha_a^{y,n}$ dans l'ACIMOG1 et $\mathcal{E}_a^{x,k} = X_k^{(a-1)} \alpha_a^x$ et $\mathcal{E}_a^{y,n} = Y_n^{(a-1)} \alpha_a^y$ dans les ACIMOG2 et 3. Puis, $\mathcal{E}_a^{x,k} = X_k \delta_a^{x,k}$ et $\mathcal{E}_a^{y,n} = Y_n \delta_a^{y,n}$ dans l'ACIMOG1 et $\mathcal{E}_a^{x,k} = X \delta_a^x$ et $\mathcal{E}_a^{y,n} = Y \delta_a^y$ dans les ACIMOG2 et 3, pour enfin en déduire l'expression des modèles et des $\hat{\beta}_k^{x,a}$ et $\hat{\beta}_n^{y,a}$. ■

Les modèles partiels d'ordre a dans (3.22) et (3.25) sont orthogonaux, puisque les composantes le sont. Cela signifie que la variance de X_k et de Y_n se décompose de manière additive en fonction de a . Chaque étape explique et apporte quelque chose de nouveau par rapport aux précédentes.

3.4.6 Commentaires

Dans l'ACIMOG1, l'orthogonalité des composantes $\{t_{k,a}\}_a$, pour tout k , et des composantes $\{u_{n,a}\}_a$ pour tout n , permet d'étudier les structures internes de chacun des tableaux. Pour chaque tableau, on peut obtenir des cartes des variables grâce aux corrélations (ou covariances) entre les variables du tableau et deux composantes associées. Ces corrélations correspondent aux coordonnées des variables projetées dans le repère orthonormé défini par les deux composantes. Par exemple, les variables de Y_n projetées dans $(u_{n,a}/\|u_{n,a}\|_D, u_{n,\ell}/\|u_{n,\ell}\|_D)$ ont pour coordonnées $r(Y_n, u_{n,a})$ et $r(Y_n, u_{n,\ell})$. Les axes $b_{n,a}$ et $b_{n,\ell}$ ne contiennent pas ces coordonnées. Le schéma de dualité de l'ACP n'est pas valable ici (voir l'annexe C). $b_{n,1}$ contient les coordonnées des variables de $Y_n^{(a-1)}$ projetées sur la super-composante $t_a = \sum_{k=1}^K t_{k,a}$. Inversement, $a_{k,a}$ contient les coordonnées des variables de $X_k^{(a-1)}$ projetées sur la super-composante $u_a = \sum_{n=1}^N u_{n,a}$. Cela ne donne pas les coordonnées des variables projetées dans un repère orthonormé du plan engendré par deux super-composantes, celles-ci n'étant pas orthogonales. Les axes peuvent être utilisés pour donner une carte des variables dans un plan, il faut alors interpréter les résultats avec précaution.

Dans l'ACIMOG2, les coordonnées variables de Y_n projetées dans le repère orthonormé $(t_a/\|t_a\|_D, t_\ell/\|t_\ell\|_D)$ de \mathbb{R}^1 sont données par les corrélations (ou les covariances) $r(t_a, Y_n) = b_{n,a} \lambda_{n,a} / \|t_a\|_D$ et $r(t_\ell, Y_n) = b_{n,\ell} \lambda_{n,\ell} / \|t_\ell\|_D$. Inversement, les coordonnées variables de X_k projetées dans le repère orthonormé $(u_a/\|u_a\|_D, u_\ell/\|u_\ell\|_D)$ de \mathbb{R}^1 sont données par les corrélations (ou les covariances) $r(u_a, X_k) = a_{k,a} \mu_{k,a} / \|u_a\|_D$ et $r(u_\ell, X_k) = a_{k,\ell} \mu_{k,\ell} / \|u_\ell\|_D$. Si les super-composantes t_a et u_a sont suffisamment proches (comme le cherche le critère), ces deux cartes sont superposables.

On peut utiliser les super-composantes t_a (et u_a) pour fournir une carte des individus, comme en régression PLS, bien que leur coordonnées ne correspondent pas à la projection des individus dans un repère orthonormé. (voir annexe C).

Dans l'ACIMOG3, on peut procéder de même, à la différence près que $r(t_a, Y_n) \neq b_{n,a} \lambda_{n,a} / \|t_a\|_D$ et $r(u_a, X_k) \neq a_{k,a} \mu_{k,a} / \|u_a\|_D$.

On remarque que les composantes partielles $t_{k,a}$, pour tout k , sont exprimées en fonction $\sum_{n=1}^N u_{n,a}$. Inversement, les composantes partielles $u_{n,a}$, pour tout n , sont exprimées en fonction de $\sum_{k=1}^K t_{k,a}$. Cela montre comment l'ACIMOG «croise» l'information.

Ceci est la principale différence que l'on peut noter entre l'ACIMOG et l'HPLS (WORLD, KETTANEH & TJESSEM, 1996) d'ordre 1. En effet, bien que la présentation de l'HPLS (§I-5.3) soit relativement floue, nous avons pu noter grâce à l'algorithme que les auteurs proposent (et que nous avons programmé), que les composantes partielles, quelles qu'elles soient, sont exprimées en fonction de la super-composante $t_T = \sum_{k=1}^K t_k$. L'information contenue dans les tableaux Y_n ne semble pas avoir beaucoup d'importance dans le calcul des différents objets.

Si $N = 1$, l'ACIMOG1 ne correspond pas tout à fait à l'ACIMO. Cela dit, les résultats d'ordre 1, sont relativement proches. Les corrélations entre les composantes issues des méthodes sont très proches en pratique. Dans l'ACIMOG1, à l'étape 1, λ_1 est la première valeur propre associée au vecteur propre b_1 de $Y'D \sum_{k=1}^K X_k Q_k^x X_k' D Y Q^y / \mu_k$ avec $\mu_k = \|X_k' D Y n\|_{Q_k^x}$. Alors que le « λ » issu de l'ACIMO, que l'on va noter λ^* , est la plus valeur propre associée au vecteur b^* de $Y'D \sum_{k=1}^K X_k Q_k^x X_k' D Y Q^y$. La matrice à diagonaliser n'est tout à fait la même dans les deux méthodes.

Si $N = 1$ et $K = 1$, l'ACIMOG correspond à l'Analyse Canonique PLS de TENENHAUS (1995).

Et si en plus $X = Y$, nous retrouvons l'ACP.

Si $K = N$ et $X_k = Y_k$, le critère à maximiser devient $var(\sum_{k=1}^K t_k)$. Cela ne correspond pas tout à fait à l'ACP du tableau $X = [X_1 \ \cdots \ X_k]$ à cause de la contrainte de norme unité sur les axes a_k . Les vecteurs a_k ne sont pas solutions d'un problème aux valeurs propres.

S'il n'y a qu'une seule variable y à prédire ($N = 1$ et $M_1 = 1$), l'algorithme n'est plus itératif et les solutions sont exactes.

Chapitre 4

Régression cubique ou multiway regression

Après l'analyse conjointe de tableaux, nous abordons le cas de la régression entre cubes*. La régression «cubique» englobe les méthodes ayant pour but l'explication d'une variable seule (y) ou d'un tableau à deux ou plus entrées par un tableau à plus de deux entrées. Dans ce chapitre, les méthodes sont détaillées pour des tableaux (explicatif et réponse) à trois entrées. Ainsi, ce chapitre regroupe 5 méthodes de régression entre cubes. Elles ont toutes été développées par les chimométriciens qui ont été les premiers à essayer de modéliser une variable réponse par un multi-tableau.

Nous ne proposons ici aucune nouvelle méthode. Les cinq méthodes sont décrites brièvement et nous exposons leur algorithme le plus clairement possible. Dans le paragraphe concernant la PLS Multi-linéaire (NPLS), nous démontrons la validité de l'algorithme grâce à la méthode des multiplicateurs de Lagrange, car cela n'a pas été fait dans la littérature. L'algorithme que nous exposons est valable pour la modélisation d'un cube \underline{Y} par un cube \underline{X} et nous l'avons déduit du programme MATLAB® de son auteur.

4.1 Multiway PLS : MPLS

On peut dire que c'est en 1987 que les premiers travaux portant sur la modélisation de multi-tableaux ont été publiés. WOLD, GELADI, ESBENSEN & ÖHMAN (1987) proposent dans leur article une méthode de décomposition d'un tableau d'ordre supérieur à 2, appelée «Multi-way PCA», en abrégé MPCA, ainsi que l'équivalent pour la modélisation de type PLS d'un tableau d'ordre élevé par un autre, méthode appelée «Multi-way PLS» ou encore MPLS. Leurs méthodes sont inspirées des travaux de LOHMÖLLER & WOLD (1980) présentés lors d'un congrès. Ces derniers ont montré qu'un cube pouvait être décomposé de manière approchée suivant un produit tensoriel entre un vecteur et une matrice, et qu'un cube d'ordre 4 pouvait s'écrire comme le produit tensoriel de deux matrices. Ainsi, ils ont développé des algorithmes pour l'ACP et la régression PLS, et ont montré que les modèles alors obtenus étaient équivalents à ceux obtenus en utilisant l'ACP et la PLS usuelle sur les 2 multi-tableaux dépliés («unfold» dans la littérature anglaise) en matrices.

* Rappel: «cube» est employé par abus de langage pour désigner un tableau de données à trois entrées pouvant être représenté par un parallépipède rectangle.

Un peu plus tard, ESBENSEN & WOLD (1983) ont proposé de résoudre l'ACP et la PLS de multi-tableaux en appliquant ces méthodes aux cubes dépliés en matrices en utilisant un algorithme NIPALS standard. C'est à partir de là que sont nées la MPCA et la MPLS. Les auteurs ont utilisé les résultats de LOHMÖLLER & WOLD (1980) pour construire des algorithmes NIPALS spécifiques afin de décomposer un multi-tableau d'ordre R (i.e à R entrées) en produit tensoriel d'un vecteur et d'un multi-tableau d'ordre $R-1$, appelé «tableau de loading».

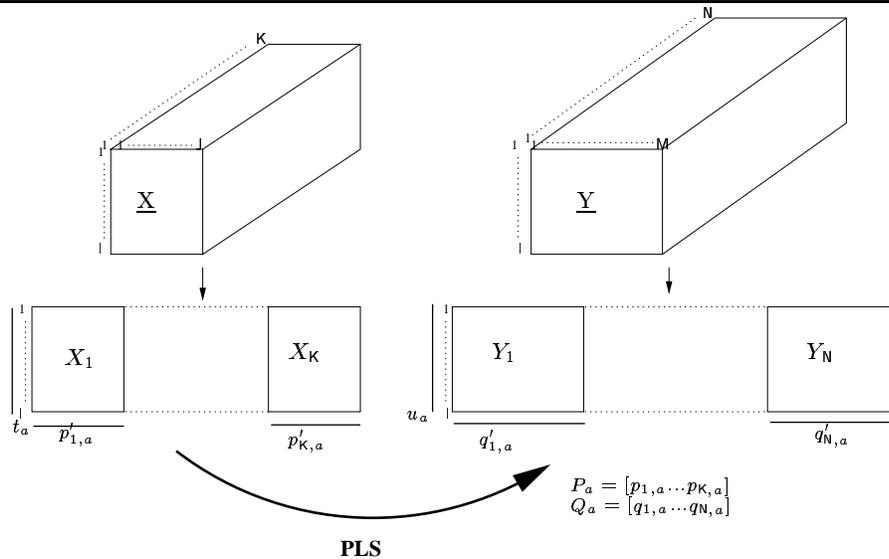
Ainsi, les modèles pour des cubes \underline{X} et \underline{Y} d'ordre R sont les suivants :

$$\begin{aligned} MPCA : \hat{\underline{X}}^A &= \sum_{a=1}^A t_a \otimes \underline{P}_a \\ MPLS : \hat{\underline{X}}^A &= \sum_{a=1}^A t_a \otimes \underline{P}_a \\ \hat{\underline{Y}}^A &= \sum_{a=1}^A u_a \otimes \underline{Q}_a. \end{aligned}$$

Les vecteurs t_a et u_a sont les composantes associées au premier mode (de longueur 1) et le cube \underline{P}_a d'ordre $R-1$ contient ce qui est appelé les «loadings» en régression PLS usuelle et le cube \underline{Q}_a d'ordre $R-1$ contient les poids des variables réponses dans la construction de u_a . Comme dans la régression PLS usuelle, il y a une relation interne entre u_a et t_a : $u_a = b_a t_a$, b_a étant le coefficient de la régression simple de u_a sur t_a . Le modèle pour \underline{Y} peut ainsi s'écrire :

$$\hat{\underline{Y}}^A = \sum_{a=1}^A b_a t_a \otimes \underline{Q}_a.$$

FIG. 4.1 – La MPLS



Lorsque $R = 3$, les cubes $\underline{P}_a = P_a$ et $\underline{Q}_a = Q_a$ sont des matrices de dimensions $J \times K$ et $M \times N$. On remarque facilement l'analogie avec la régression PLS : si les cubes \underline{X} et \underline{Y} sont des matrices, on obtient $\hat{X}^A = \sum_{a=1}^A t_a \otimes p_a = \sum_{a=1}^A t_a p_a'$, p_a étant le vecteur de «loading» de la régression PLS, de longueur J , et $\hat{Y}^A = \sum_{a=1}^A b_a t_a q_a'$, modèle de la régression PLS.

Comme en PLS (et ACP) usuelle, les composantes $\{t_a\}_a$ sont orthogonales entre elles.

Considérons la MPCA d'un cube. Alors \underline{P}_a est une matrice, et si cette matrice est de rang 1, elle pourra s'écrire comme produit tensoriel de deux vecteurs, $\underline{P}_a = v_a \otimes r_a$. Il vient donc que le modèle correspond au modèle PARAFAC.

Ces méthodes sont équivalentes à utiliser l'ACP et la PLS sur les cubes dépliés en matrices ($X = \mathbf{X}_1$ et $Y = \mathbf{Y}_1$), ce qui est illustré par la figure 4.1. Elles sont donc en fait une réécriture des modèles obtenus par une ACP ou une PLS particulière. Elles fournissent des résumés de toutes les variables (t_a), mais pas de résumés des variables par «tranches frontales» des cubes. Les modèles obtenus ne sont pas fonction de chaque tableau frontal (X_k), ni optimisés par bloc. MPLS n'est donc pas réellement une méthode de modélisation cubique. Notons que ses auteurs la présente ainsi tout en citant qu'elle est équivalente à la régression PLS usuelle définie ci-dessus.

L'algorithme

L'algorithme MPLS que nous présentons, dans le cas de deux cubes d'ordre 3, est tiré de l'article de base de MPLS (WOLD, GELADI, ESBENSEN & ÖHMAN, 1987).

Habituellement les données sont centrées et réduites en colonne. Nous noterons X et Y les cubes \underline{X} et \underline{Y} dépliés en matrices, c'est-à-dire, $X = \mathbf{X}_1$ et $Y = \mathbf{Y}_1$, t , la composante associée à \underline{X} et u celle associée au cube réponse \underline{Y} .

étape 0 : initialisation de a

$$a = 0$$

étape 1 : initialisation

$$a = a + 1$$

On pose t_a comme étant la colonne de X ayant la plus grande variance.

étape 2

$w_a = \text{vec}(W_a) = u_a' X / (u_a' u_a)$, w_a est le vecteur des poids, de longueur JK ; W_a est de taille $J \times K$.

Normaliser W : $W_a = W_a / \|W_a\|$ i.e. $w_a = w_a / \|w_a\|$, ou bien normer t_a à l'unité à l'étape 4.

étape 3

Ici, il est possible de décomposer W_a de la forme « sr' » (produit tensoriel de deux vecteurs s et r) avec l'algorithme NIPALS.

étape 4

$$\text{Calculer } t : t_a = X w_a / w_a' w_a$$

étape 5

Calculer le critère de convergence : $d = (t_{new} - t_{old})'(t_{new} - t_{old}) / (t_{new}'t_{new})$

étape 6

Calcul des poids réponses : $q_a = t_a'Y/t_a't_a$, q_a est égal à $vec(Q_a)$ et de longueur MN.

étape 7

Tester la convergence avec d : aller à l'étape 9 si $d < 10^{-10}$, sinon à l'étape 8.

étape 8

Calcul de u_a : $u_a = Yq_a/q_a'q_a$

Retourner à l'étape 2.

étape 9 : après la convergence

Calcul de P , matrice de loading : $vec(P_a) = t_a'X/t_a't_a$

Remarque : on ne décompose pas P_a avec l'aide de NIPALS (ACP), car les composantes t_a ne seraient plus orthogonales.

Si on le désire, on peut ici normer P_a à l'unité ($c = \|P_a\| = \|vec(P_a)\|$, $P_a = P_a/c$), et dans ce cas transformer t_a et W_a pour que les relations entre P_a , t_a et W_a restent toujours vraies : $t_a = t_a c$ et $w_a = w_a c$.

Eventuellement, normer Q_a à l'unité.

Calculer $b_a = t_a'u_a/t_a't_a$. Si q_a n'est pas normé, $b_a = 1$.

étape 10 : calcul des résidus

$\underline{E}_X = \underline{X} - t_a \otimes P_a$, où \otimes désigne le produit tensoriel.

$\underline{E}_Y = \underline{Y} - t_a \otimes Q_a b_a$

étape 11 : passage à l'étape suivante

Pour calculer la composante suivante, poser $\underline{X} = \underline{E}_X$ et $\underline{Y} = \underline{E}_Y$ et aller à l'étape 1.

A l'étape 2, les auteurs proposent de normer soit les vecteurs de poids, soit les composantes. Si les composantes sont normées à l'unité, on retombe dans un problème d'analyse canonique. De plus, à l'étape 3, ils proposent d'utiliser l'algorithme NIPALS pour faire une ACP des vecteurs de poids afin d'expliquer la composante t plus précisément en fonction des deux derniers modes du cube.

4.2 Linear Three-way Decomposition : LTD

STÅHLE (1989) a été le premier à proposer une méthode spécifique aux cubes fournissant des modèles trilinéaires. Cette méthode est basée sur la construction d'un algorithme généralisant celui de PLS (WOLD, 1966b) et permettant de trouver une décomposition trilinéaire d'un cube \underline{X} explicatif qui puisse être utilisé pour prédire un cube \underline{Y} . Cet algorithme va calculer une composante associée au premier mode et un vecteur de poids par mode supplémentaires pour chaque cube. L'algorithme est détaillé ci-dessous. Nous ne donnons pas plus de détails sur cette méthode pour plusieurs raisons : nous n'en avons pas plus et il a été montré l'équivalence de cet algorithme avec celui la PLS multilinéaire décrite au paragraphe 4.3 lorsqu'il n'y a qu'une variable ou matrice à expliquer (DE JONG, 1998).

Pour décrire brièvement cette méthode, il est nécessaire de spécifier quelques notations

supplémentaires. On note t la composante suivant le premier mode (vecteur de taille l), ω^J et ω^K les vecteurs de poids correspondant aux deux autres modes pour les prédicteurs. On note u la composante suivant le premier mode, q^M , q^N les vecteurs des poids pour les deux autres modes pour les réponses du cube \underline{Y} .

L'algorithme

étape 1 : initialisation

On pose u égal à la première colonne de Y . De plus, on fixe les vecteurs de poids ω^J , ω^K , q^M et q^N égaux respectivement aux vecteurs unitaires $\mathbf{1}_J$, $\mathbf{1}_K$, $\mathbf{1}_M$ et $\mathbf{1}_N$.

étape 2 : calcul des poids explicatifs

$\omega^J(j) = u'X_j\omega^K$ où X_j est la $j^{\text{ième}}$ tranche latérale du cube \underline{X} et $\omega^J(j)$ le $j^{\text{ième}}$ élément du vecteur ω^J .

normer ω^J à l'unité

$\omega^K(k) = u'X_k\omega^J$

normer ω^K à l'unité

Remarque : les vecteurs des poids utilisés à droite dans ces 2 égalités sont les poids venant de l'étape a précédente (ou de l'étape d'initialisation lorsque $a = 1$).

étape 3 : calcul de t

$t(i) = \omega^{K'}X_i\omega^J$

étape 4 : calcul des poids réponses

$q^M(m) = t'Y_mq^N$

normer q^M à l'unité

$q^N(n) = t'Y_n^M$

normer q^N à l'unité

Remarque : les poids utilisés à droite dans ces égalités sont les poids venant de l'étape a précédente (ou de l'étape d'initialisation lorsque $a = 1$).

étape 5 : calcul de u

$u(i) = q^{N'}Y_iq^M$

étape 6 : test de la convergence

Si deux itérations successives donnent la même valeur pour tous les vecteurs, la procédure a convergé.

S'il y a convergence, passer à l'étape 7; sinon retourner à l'étape 2.

étape 7

$b = t'u/(t't)$

b est le coefficient de la régression de u sur t .

étape 8 : calcul d'une matrice P afin d'obtenir des composantes t orthogonales

calcul de P , de taille $J \times K$, telle que $vec(P) = X't/(t't)$.

$d = \|P\| = tr(P'P) = vec(P)'vec(P)$, normer P

$t = td$; $\omega^J = \omega^J/\sqrt{d}$; $\omega^K = \omega^K/\sqrt{d}$

Remarque : on peut garder des poids de longueur 1 en ne normant pas P et en ne les transformant pas.

étape 9 : calculs des résidus

$X_i = X_i - t(i)P$

$Y_i = Y_i - bt(i)q^M q^{N'}$

étape 10 : passage à l'étape suivante

Si la dernière composante n'apporte que peu d'information, on arrête la procédure.

Sinon, on calcule les composantes de l'étape en utilisant les résidus calculés à l'étape 9.

4.3 Multilinear PLS : NPLS

Cette méthode est notée en abrégé NPLS, la lettre N faisant référence au nombre de modes du cube explicatif. Elle a été proposée par BRO (1996) comme généralisation directe de la régression PLS usuelle de 2 tableaux.

La NPLS s'applique à des hyper-cubes de tous ordres et le modèle de reconstitution de \underline{X} est multi-linéaire. Dans le cas de cubes de données, les modèles sont tri-linéaires.

Elle présente l'avantage d'avoir la principale propriété de PLS, c'est-à-dire de maximiser la covariance entre les composantes associées aux cubes prédicteurs et réponses.

En comparaison avec des méthodes de type MPLS, c'est-à-dire où les cubes sont dépliés en matrices, l'algorithme de NPLS est plus rapide, car il y a moins de paramètres à estimer, surtout lorsqu'on est en présence de grandes dimensions, (> 100) ce qui est fréquent en chimométrie. Toutefois à cause des contraintes plus fortes imposées pour NPLS, on perd un peu au niveau de l'ajustement.

Comme en LTD, des composantes associées au mode commun des deux cubes et des vecteurs de poids associés à chacun des autres modes sont recherchés. Les notations sont les mêmes que celles utilisées pour décrire la LTD.

Dans la suite nous nous plaçons dans le cas où les cubes sont d'ordre trois, c'est-à-dire des tableaux à trois entrées.

La PLS Multi-linéaire, comme la PLS, est un problème de maximisation d'un critère de covariance :

$$\max_{\omega^J, \omega^K, q^M, q^N} cov(t, u),$$

tel que les quatre vecteurs de poids soient de norme unité, et où les composantes t et u , appartenant à \mathbb{R}^I , sont de la forme, pour $i = 1, \dots, I$: $t(i) = t_i = \sum_{j=1}^J \sum_{k=1}^K x_{ijk} \omega_j^J \omega_k^K$ et

$$u(i) = u_i = \sum_{m=1}^M \sum_{n=1}^N y_{imn} q_m^M q_n^N.$$

Si on note X_i (*resp.* Y_i) la $i^{\text{ème}}$ matrice de taille $J \times K$ du cube \underline{X} (*resp.* la $i^{\text{ème}}$ matrice de

taille $M \times N$ du cube \underline{Y}), alors on peut écrire ces expressions sous forme matricielle :

$$t_i = \omega^J X_i \omega^K$$

et

$$u_i = q^{M'} Y_i q^N, \quad \forall i = 1, \dots, l.$$

Ce qui peut encore s'écrire $t = [X_1 \dots X_K] (\omega^K \otimes \omega^J)$ et $u = [Y_1 \dots Y_N] (q^N \otimes q^M)$, où \otimes est l'opérateur du produit de Kronecker.

La résolution de ce problème se fait grâce à un algorithme itératif, détaillé pour le cas d'une seule variable réponse y ou d'une matrice de variables réponses Y , dans (BRO, 1996, 1998) et discuté dans (SMILDE, 1997) pour le cas d'un seul y .

Justification de l'utilisation de l'algorithme

Nous allons ici montrer que les vecteurs de poids calculés dans l'algorithme ci-après sont bien solutions du problème de maximisation de $cov(t, u)$ sous les contraintes de norme unité pour les vecteurs de poids ω^J , ω^K , q^M et q^N . Cette démonstration nous paraît nécessaire car elle ne figure pas dans les articles de références.

Le critère à maximiser peut être écrit suivant les expressions :

$$\begin{aligned} cov(t, u) &= \sum_{i=1}^l t_i q^{M'} Y_i q^N \\ &= \sum_{i=1}^l t_i q^{N'} Y_i' q^M \text{ (car c'est un réel donc égal à son transposé)} \\ &= \sum_{i=1}^l u_i \omega^{J'} X_i \omega^K \\ &= \sum_{i=1}^l u_i \omega^{K'} X_i' \omega^J \text{ (pour la même raison)} \end{aligned}$$

Ainsi, en utilisant la méthode des multiplicateurs de Lagrange et les notations usuelles :

$$\begin{aligned} \mathcal{L}(\omega^J, \omega^K, q^M, q^N) &= \\ cov(t, u) - \frac{\alpha_J}{2} (\omega^{J'} \omega^J - 1) - \frac{\alpha_K}{2} (\omega^{K'} \omega^K - 1) - \frac{\alpha_M}{2} (q^{M'} q^M - 1) - \frac{\alpha_N}{2} (q^{N'} q^N - 1), \end{aligned}$$

on obtient les équations normales :

$$\begin{aligned}\nabla_{\omega^K} \mathcal{L} &= \sum_{i=1}^I u_i \omega^{J'} X_i - \alpha_K \omega^K = 0 \\ \nabla_{\omega^J} \mathcal{L} &= \sum_{i=1}^I u_i \omega^{K'} X_i' - \alpha_J \omega^J = 0 \\ \nabla_{q^M} \mathcal{L} &= \sum_{i=1}^I t_i q^{N'} Y_i' - \alpha_M q^M = 0 \\ \nabla_{q^N} \mathcal{L} &= \sum_{i=1}^I t_i q^{M'} Y_i - \alpha_N q^N = 0 \\ \omega^{J'} \omega^J - 1 &= 0 \\ \omega^{K'} \omega^K - 1 &= 0 \\ q^{N'} q^N - 1 &= 0 \\ q^{M'} q^M - 1 &= 0.\end{aligned}$$

On en déduit par quelques manipulations simples que les multiplicateurs doivent vérifier $\alpha_J = \alpha_K = \alpha_N = \alpha_M = \alpha = \text{cov}(t, u)$ et les quatre équations aux valeurs propres suivantes, pour que les solutions existent :

$$\begin{aligned}\alpha^2 \omega^J &= \sum_{i_1=1}^I u_{i_1} X_{i_1} \sum_{i_2=1}^I u_{i_2} X_{i_2}' \omega^J \\ \alpha^2 \omega^K &= \sum_{i_1=1}^I u_{i_1} X_{i_1}' \sum_{i_2=1}^I u_{i_2} X_{i_2} \omega^K \\ \alpha^2 q^M &= \sum_{i_1=1}^I t_{i_1} Y_{i_1} \sum_{i_2=1}^I t_{i_2} Y_{i_2}' q^M \\ \alpha^2 q^N &= \sum_{i_1=1}^I t_{i_1} Y_{i_1}' \sum_{i_2=1}^I t_{i_2} Y_{i_2} q^N.\end{aligned}$$

En remarquant que $\sum_{i=1}^I u_i X_i = Z_u$ et $\sum_{i=1}^I t_i Y_i = Z_t$ (voir l'algorithme), on retrouve bien les équations aux valeurs propres cherchées.

S'ils existent, les vecteurs de poids q^N , q^M , ω^K et ω^J sont respectivement les vecteurs propres normés associés à la plus grande valeur propre commune α des matrices $Z_t' Z_t$, $Z_t Z_t'$, $Z_u' Z_u$, $Z_u Z_u'$ car α est aussi le maximum du critère.

L'algorithme pour un cube \underline{X} et un cube \underline{Y}

Rappelons les notations :

Le cube de variables explicatives \underline{X} est de dimension $I \times J \times K$, le cube de variables réponses \underline{Y} est de dimension $I \times M \times N$.

Le nombre total de composantes retenues est noté A et l'indice courant a .

Les vecteurs des poids associés à \underline{X} sont notés ω_a^J et ω_a^K et ses composantes t_a .

Les vecteurs des poids associés à \underline{Y} sont notés q_a^M et q_a^N et ses composantes u_a .

On pose $\omega_a = \omega_a^K \otimes \omega_a^J$, et $q_a = q_a^N \otimes q_a^M$, où \otimes est le produit Kronecker.

On note $X^{(0)} = X = [X_1 \dots X_K]$ et $Y^{(0)} = Y = [Y_1 \dots Y_N]$.

Enfin, $T_a = [t_1 \dots t_a]$, $U_a = [u_1 \dots u_a]$, $Q_a = [q_1 \dots q_a]$, $W_a = [\omega_1 \dots \omega_a]$, $b_a = (T_a' T_a)^{-1} T_a' u_a$, et B_a est la matrice triangulaire supérieure de taille $a \times a$ formée des vecteurs b_1 à b_a .

étape 0 : prétraitement des données

Centrer en colonne les deux cube \underline{X} et \underline{Y} , i.e. $\sum_i x_{ijk} = \sum_i y_{imn} = 0$.

Boucle a

étape 1 : initialisation

u_a est égal à une variable de $\underline{Y}^{(a-1)}$.

étape 2 : calcul de Z_u

Calculer la matrice Z_u telle que $vec(Z_u) = X^{(a-1)'} u_a$.

C'est-à-dire $Z_u = [X_1^{(a-1)'} u_a \dots X_K^{(a-1)'} u_a]$.

Remarque : on a aussi $vec(Z_u) = [vec(X_1^{(a-1)}) \dots vec(X_K^{(a-1)})] u_a$.

étape 3 : calcul des poids

ω_a^J et ω_a^K sont les premiers vecteurs singuliers à droite et à gauche de norme unité de Z_u . C'est-à-dire, ils sont respectivement les premiers vecteurs propres de normes unité de $Z_u Z_u'$ et $Z_u' Z_u$.

étape 4 : calcul de t_a

$t_a = X^{(a-1)} \omega_a$

Remarque : $t_a = \sum_{k=1}^K \omega_{ak}^K X_k^{(a-1)} \omega_a^J$: somme pondérée de «composantes partielles».

étape 5 : calcul de Z_t

$vec(Z_t) = Y^{(a-1)'} t_a$, c'est-à-dire $Z_t = [Y_1^{(a-1)'} t_a \dots Y_N^{(a-1)'} t_a]$.

Remarque : $vec(Z_t) = [vec(Y_1^{(a-1)}) \dots vec(Y_N^{(a-1)})] t_a$.

étape 6 : calcul des poids associés à \underline{Y}

q_a^M et q_a^N sont les premiers vecteurs singuliers à droite et à gauche de norme unité de Z_t . C'est-à-dire, ils sont respectivement les premiers vecteurs propres normés de $Z_t Z_t'$ et $Z_t' Z_t$.

étape 7 : calcul de u_a

$u_a = Y^{(a-1)} q_a$

étape 8 : convergence de u_a

Si, en norme, u_a s'est stabilisé ou si le nombre maximal d'itérations est atteint, continuer.

Sinon, reprendre à l'étape 2.

étape 9 : calcul de B_a

$$b_a = (T_a' T_a)^{-1} T_a' u_a$$

$b_l^{*'} = [b_l' \ 0 \dots 0]$, vecteur de longueur a , pour $l = 1, \dots, a-1$.

$$B_a = [b_1^{*'} \ \dots \ b_{a-1}^{*'} \ b_a].$$

étape 10 : calcul des prédictions

$\hat{Y}^a = T_a B_a Q_a'$. \hat{Y}^a est de dimension $l \times MN$.

Pour obtenir $\hat{\underline{Y}}^a$, il suffit de «replier» correctement \hat{Y}^a .

$\hat{X}^a = \hat{X}^{a-1} + t_a \omega_a' = \sum_{l=1}^a t_a \omega_a' = T_a W_a'$. \hat{X}^a est de dimension $l \times JK$. Il suffit de «replier» correctement cette matrice pour obtenir $\hat{\underline{X}}^a$.

étape 11 : calcul des résidus

$E^{xa} = X - \hat{X}^a$, de taille $l \times JK$,

$E^{ya} = Y - \hat{Y}^a$, de taille $l \times MN$.

Si on veut des composantes supplémentaires, passer à l'étape 12. Sinon, s'arrêter.

étape 12

Poser $X^{(a)} = E^{xa}$ et $Y^{(a)} = E^{ya}$;

$a = a + 1$ et recommencer à partir de l'étape 1.

D'après son auteur, cet algorithme est généralisable pour des cubes explicatifs et réponses d'ordre supérieur. Cependant, l'algorithme correspondant n'est détaillé nulle part.

Nous avons déduit l'algorithme présenté du programme MATLAB[®] de Bro qui est en diffusion libre sur internet[†]. Celui-ci fonctionne pour un cube explicatif d'ordre inférieur ou égal à cinq, et pour un cube réponse d'ordre inférieur ou égal à trois.

La reconstitution \hat{X}_i^A de la matrice X_i , $i = 1, \dots, l$, $i^{\text{ième}}$ tranche horizontale du cube de prédicteurs \underline{X} va alors s'écrire :

$$\hat{X}_i^A = \sum_{a=1}^A t_{ai} \omega_a^J \omega_a^{K'}$$

et la modélisation \hat{Y}_n^A d'une matrice Y_n , $n = 1, \dots, N$, du cube réponse \underline{Y} va s'écrire :

$$Y_n = \sum_{a=1}^A P_{T_a} u_a (q_a^{N'} \otimes q_a^{M'}) + E_n^{ya}.$$

Le système des composantes $\{t_a\}_a$ ne forme pas, *a priori*, un système orthogonal. Ceci, à notre avis, est le gros inconvénient de cette méthode. En effet, chaque étape a n'explique pas une part différente de l'information contenue dans \underline{X} et \underline{Y} . Contrairement à la régression PLS, qu'elle «généralise», les modèles issus de la NPLS ne se décomposent pas en une somme de modèles partiels orthogonaux qui est une de ses principales propriétés.

[†] URL: <http://www.models.kvl.dk/sources/>

Les systèmes des poids $\{\omega_a^J\}_a$ et $\{\omega_a^K\}_a$ ne sont pas orthonormés, les poids sont seulement normés. De même pour le système $\{\omega_a\}_a = \{\omega_a^K \otimes \omega_a^J\}_a$.

Le modèle \widehat{Y}^A peut s'écrire de la forme $X\widehat{\beta}_A$, où $X = [X_1 \ \dots \ X_K]$. En effet, il est facile de montrer qu'une composante t_a s'écrit (SMILDE, 1997) :

$$t_a = X \left[\prod_{l=1}^{a-1} (I - \omega_l \omega_l') \right] \omega_a.$$

Ainsi, en notant \mathcal{W}_A la matrice de dimension $JK \times A$, $[\omega_1 \ \dots \ [\prod_{l=1}^{A-1} (I - \omega_l \omega_l') \omega_A]]$, on a

$$\widehat{Y}^A = X\mathcal{W}_A B_A Q_A' = X\widehat{\beta}_A.$$

Si un nouvel échantillon $X_{i'}$ est mesuré, matrice de taille $J \times K$, on peut ainsi prédire la matrice $\widehat{Y}_{i'}^A$, $M \times N$, correspondante. En effet, X est égale à $[vec(X_1) | \dots | vec(X_K)]'$ et $vec'(\widehat{Y}_{i'}^A) = vec'(X_{i'})\widehat{\beta}_A$.

De manière analogue, il est possible de montrer que :

$$\widehat{X}^A = X \left[\omega_1 \omega_1' + \sum_{a=2}^A \left\{ \prod_{l=1}^{a-1} (I - \omega_l \omega_l') \right\} \omega_a \omega_a' \right].$$

Dans le cas d'un unique y , DE JONG (1998) donne quelques éléments de calculs et montre que l'application de la méthode LTD ou NPLS permet d'obtenir les mêmes modèles d'ajustement pour y . Il arrive à montrer que les composantes t_a issues des deux méthodes engendrent le même espace, celles de LTD formant une base. Il montre aussi que les vecteurs de poids ω^J et ω^K issus de ces deux méthodes sont égaux. Sa démonstration est basée sur le fait que leur calcul ne dépend pas de la déflation de X . C'est-à-dire, il montre que $y^{(a-1)'} X^{(a-1)} = y^{(a-1)'} X$, qui est la seule quantité nécessaire au calcul de ω . Il n'est pas nécessaire de déflater X pour calculer les composantes (car il est possible d'écrire $T_a = X\mathcal{W}_a$). On peut donc aussi dire que le type de déflation appliqué à X importe peu au calcul de ω .

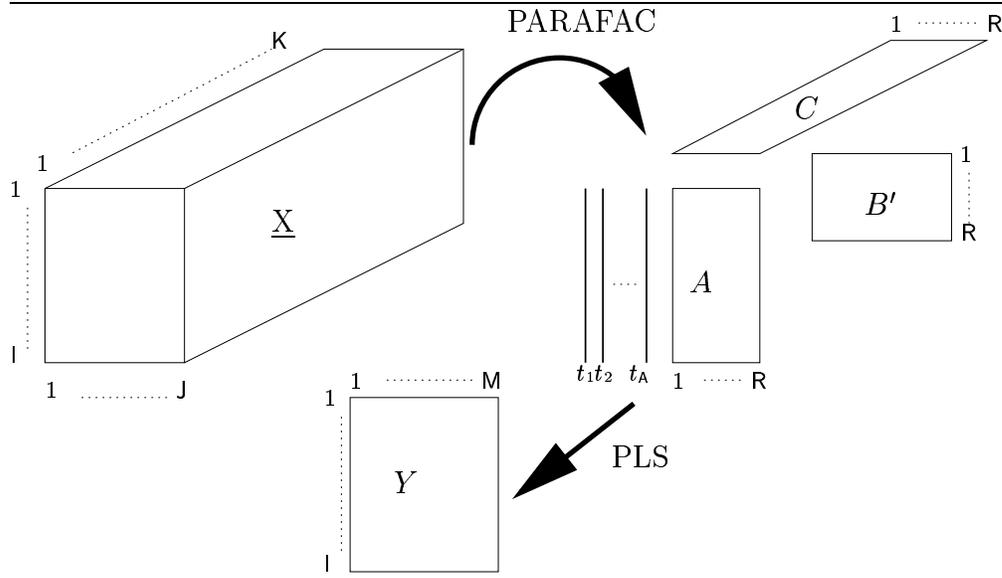
Remarquons simplement que De Jong dit au début de sa publication que tous les résultats qu'il montre dans le cas d'un unique y sont aussi valables pour Y et \underline{Y} , sans toutefois donner des éléments de démonstration. Or, sur une application numérique, nous avons constaté dans le cas de réponses cubiques, que les quantités $X'u_a$ et $X^{(a-1)'} u_a$ ne sont pas égales, sauf si l'on arrondit au nombre entier ! Nous en déduisons donc que ω dépend de la déflation de X . Cependant, il reste à voir si une autre démonstration ne permettrait pas d'affirmer que NPLS et LTD fournissent les mêmes ajustements des réponses et les mêmes ω .

4.4 Régression sur les facteurs de PARAFAC

Nous pensons que cette technique doit être citée bien qu'elle ne présente aucune propriété particulière. Sa description est immédiate.

Proposée par GELADI, XIE, POLISSAR & HOPKE (1998), elle combine une méthode de décomposition d'un cube, PARAFAC avec la contrainte d'avoir des facteurs positifs, (Positive

FIG. 4.2 – Régression sur les paramètres de PARAFAC.



Matrix Decomposition (PAATERO & TAPPER, 1994; PAATERO, 1997)) et la régression PLS de deux matrices. Elle s'applique donc à un cube de prédicteurs et une matrice de réponses. Dans leur papier, ils donnent des exemples où cette technique a été employée sur des données de chimie environnementale.

Le principe est le suivant (Fig.4.2): décomposer de façon trilinéaire les prédicteurs \underline{X} à l'aide de CANDECOMP/PARAFAC (chapitre 1.2) sous les contraintes que tous les facteurs (A , B et C) soient non négatifs, et effectuer une régression PLS du tableau réponse Y sur la matrice des facteurs du premier mode A .

Rappelons que les R facteurs de la matrice A associée au premier mode sont une combinaison linéaire des variables de \underline{X} . Notons $A = XW$ où A est taille $I \times R$, $X = \mathbf{x}_i$ est le tableau déplié le long du troisième mode, et W est de taille $JK \times R$ et dépend de B et C . La régression PLS de Y sur A entraîne l'écriture du modèle de rang A de la forme :

$$\begin{aligned} \hat{Y}^A &= A\hat{\beta}^A \\ &= XW\hat{\beta}^A. \end{aligned}$$

Ainsi, le modèle PLS peut s'écrire en fonction des variables initiales.

4.5 Multiway Covariates Regression Models

SMILDE & KIERS (1999) proposent de généraliser aux cas de tableaux à trois entrées la méthode de régression de DE JONG & KIERS (1992) («Principal Covariates Regression Models»). Nous conserverons, pour simplifier, son nom anglais «Multiway Covariates Regression Models», noté MWcovR.

Les modèles sont décrits pour le cas de deux cubes d'ordre 3, mais peuvent être étendus aux cas de cubes d'ordre supérieur (hyper-cubes), en particulier pour plus de deux cubes d'ordre 3 (SMILDE, WESTERHUIS & BOQUÉ, 2000).

Elle diffère des méthodes MPLS, NPLS, et LTD, vues précédemment dans ce chapitre, en plusieurs points : les \mathbf{A} composantes sont calculées en même temps, au cours d'une seule et même étape, contrairement aux méthodes de type PLS qui les calculent une par une, et aucune structure n'est imposée aux cubes (pas d'orthogonalité).

Le but premier des modèles MWcovR est d'expliquer un cube $\underline{\mathbf{Y}}$ par un cube $\underline{\mathbf{X}}$. Mais avec un choix approprié des paramètres, les modèles de type TUCKER1,2,3 et PARAFAC peuvent être retrouvés pour $\underline{\mathbf{X}}$ et/ou $\underline{\mathbf{Y}}$.

Pour déterminer les modèles, un problème de maximisation est résolu où les cubes $\underline{\mathbf{X}}$ et $\underline{\mathbf{Y}}$ sont préalablement dépliés en matrices le long du premier mode $\mathbf{X} = \mathbf{X}_1$ et $\mathbf{Y} = \mathbf{Y}_1$:

$$\max_W (\alpha R_X^2 + (1 - \alpha) R_Y^2),$$

où α est un réel fixé entre 0 et 1, \mathbf{W} , $\mathbf{KJ} \times \mathbf{A}$, la matrice des poids des composantes ($\mathbf{A} = \mathbf{X}\mathbf{W}$, $\mathbf{l} \times \mathbf{A}$) et les coefficients R_X^2 et R_Y^2 , exprimant le pourcentage de variance expliquée par les modèles, s'écrivent de la manière suivante :

$$\begin{aligned} R_X^2 &= 1 - \|\mathbf{X} - \mathbf{X}\mathbf{W}\mathbf{H}'_X\|^2 / \|\mathbf{X}\|^2 \\ R_Y^2 &= 1 - \|\mathbf{Y} - \mathbf{X}\mathbf{W}\mathbf{H}'_Y\|^2 / \|\mathbf{Y}\|^2. \end{aligned}$$

Les matrices \mathbf{H}'_X et \mathbf{H}'_Y peuvent être choisies suivant différentes expressions pour retrouver l'écriture des modèles PARAFAC, TUCKER1, TUCKER2, ou TUCKER3 (voir KROONENBERG & DE LEEUW, 1980; KROONENBERG, 1983, voir aussi §I-1.1, I-1.2) pour \mathbf{X} et \mathbf{Y} .

Les modèles de reconstitution des réponses et des prédicteurs s'écrivent alors sous la forme :

$$\begin{aligned} \mathbf{X} &= \mathbf{X}\mathbf{W}\mathbf{H}'_X + \mathbf{E}_X \\ \mathbf{Y} &= \mathbf{X}\mathbf{W}\mathbf{H}'_Y + \mathbf{E}_Y, \end{aligned}$$

où \mathbf{E}_X et \mathbf{E}_Y sont les matrices d'erreurs.

SMILDE & KIERS (1999) montrent que ce problème de maximisation est équivalent au problème de minimisation

$$\min_W (\beta \|\mathbf{X} - \mathbf{X}\mathbf{W}\mathbf{H}'_X\|^2 + (1 - \beta) \|\mathbf{Y} - \mathbf{X}\mathbf{W}\mathbf{H}'_Y\|^2) \quad (4.1)$$

qui peut également s'écrire sous la forme

$$\min_W \|\mathbf{Z} - \mathbf{X}\mathbf{W}\mathbf{U}'\|^2,$$

où les matrices \mathbf{Z} et \mathbf{U} sont définies implicitement par (4.1) et $\beta = (\alpha \|\mathbf{Y}\|^2) / (\alpha \|\mathbf{Y}\|^2 + (1 - \alpha) \|\mathbf{X}\|^2)$ est compris entre 0 et 1. Cette dernière expression du critère étant la plus pratique, elle est utilisée pour résoudre le problème d'optimisation initial, à l'aide d'un algorithme de type Moindres Carrés Alternés (voir algorithme ci-après).

Pour choisir la valeur de α la validation croisée est une méthode appropriée. Cependant, cela peut prendre énormément de temps. Dans ce cas, les utilisateurs devront choisir une

valeur de α suivant leur connaissance personnelle des données.

Si $\alpha = 1$, on se ramène à l'étude du cube \underline{X} par les méthodes de TUCKER ou PARAFAC. Si l'on choisit $\alpha = 0$, on se ramène à l'étude du cube \underline{Y} par les méthodes de TUCKER ou PARAFAC avec la contrainte que les composantes $A, I \times A$, appartiennent à l'espace engendré par les colonnes de \underline{X} .

Lorsque les prédicteurs et les réponses sont deux matrices X et Y , MWcovR est exactement la méthode des Principal Covariates Regression Models (DE JONG & KIERS, 1992).

L'algorithme

Cet algorithme est tiré de (SMILDE & KIERS, 1999).

Il s'agit d'un algorithme aux moindres carrés alternés qui va minimiser la fonction de perte

$$f(W) = \|Z - XWU'\|^2$$

où $Z = [\sqrt{\beta}X \quad \sqrt{1-\beta}Y]$ et $U = [\sqrt{\beta}H'_X \quad \sqrt{1-\beta}H'_Y]$.

Choix de P_X et P_Y

Selon le type de modèles (TUCKER ou PARAFAC) que l'on désire pour \underline{X} et \underline{Y} , il faut choisir la manière dont vont être calculées les matrices H'_X et H'_Y .

- Pour un modèle de type TUCKER 3 pour \underline{X} : $H'_X = G_X(C'_X \otimes B'_X)$. $G_X, A \times QR$, est la juxtaposition des tranches frontales de la matrice noyau correspondante de taille $A \times Q \times R$.
- Pour un modèle de type PARAFAC pour \underline{X} : $H'_X = I_A(C'_X \otimes B'_X)$. $I_A, A \times AA$ est la juxtaposition des tranches frontales du cube identité de dimension $A \times A \times A$.
- Pour un modèle de type TUCKER2: $H'_X = G_X(Id \otimes B'_X)$. $G_X, A \times QK$, est la juxtaposition des tranches frontales de la matrice noyau correspondante de taille $A \times Q \times K$.
- Pour un modèle de type TUCKER1: $H'_X = G_X$. $G_X, A \times JK$, est la juxtaposition des tranches frontales de la matrice noyau correspondante de taille $A \times J \times K$.

Les matrices G_X sont les matrices noyaux correspondant à chacun des modèles et $B_X, J \times Q$, et $C_X, K \times R$, sont les matrices des facteurs des modes 2 et 3 correspondants à chacun des modèles (voir chapitre 1.1). Dans le cas de modèles PARAFAC, B_X et B_Y sont respectivement de taille $J \times A$ et $K \times A$.

H_Y s'établit de la même façon que H_X en remplaçant les indices X par Y afin de définir des facteurs (B_Y et C_Y) et une matrice noyau G_Y pour \underline{Y} .

étape 1 : initialisation

Initialiser W, B_X, B_Y, C_X, C_Y .

étape 2

Mettre à jour les matrices noyaux G_X et G_Y

étape 3

Calculer la valeur de la fonction de perte: f_{old} .

étape 4

Calculer W en minimisant $f_{old}(W)$: $W = X^+ ZU'^+$. Le symbole $+$ désigne l'inverse généralisé de Moore-Penrose.

étape 5

Mettre à jour les matrices noyaux G_X et G_Y .

étape 6

Calculer B_X et B_Y avec l'algorithme TUCKALS sur X et sur Y .

étape 7

Calculer C_X et C_Y à l'aide de l'algorithme TUCKALS.

étape 8

Calculer la valeur de la fonction de perte: f_{new} .

étape 9

Comparer f_{old} à f_{new} . S'il y a convergence, s'arrêter. Sinon, retourner à l'étape 4.

Cet algorithme correspond au où cas on veut avoir des modèles de TUCKER3. Dans ce cas, rappelons que la mise à jour des matrices noyaux aux étape 2 et 5 (voir chapitre 1.1) se fait par $G_X = (XW)'X(C_X \otimes B_X)$ et $G_Y = (XW)'Y(C_Y \otimes B_Y)$, comme dans l'algorithme TUCKALS (§1.1). Ainsi, lorsque W est mis à jour à l'étape 4, les matrices noyaux changent, d'où leur actualisation à l'étape suivante.

L'initialisation (étape 1) peut se faire de plusieurs façons. En particulier, en choisissant les B_X , C_X , B_Y et C_Y comme étant les matrices des vecteurs singuliers des tableaux X_J , X_K , Y_M et Y_N , comme cela est fait dans les méthodes de modélisation de type TUCKER. Quant à W , elle peut être choisie comme la matrice des A premiers vecteurs propres de $X_i'X_i$: cela correspond à l'initialisation des facteurs associés au premier mode dans les modèles de TUCKER.

Pour avoir des modèles de type PARAFAC, les étapes 6 et 7 sont changées en étapes PARAFAC/CANDECOMP appropriées; et les étapes 2 et 5 sont omises (voir chapitre 1.2).

D'après ses auteurs, cet algorithme est décroissant monotone et borné par 0.

Chapitre 5

Régression multi-blocs

Sous les termes de «régression multi-blocs», nous regroupons les méthodes mettant en jeu un ensemble de tableaux de données mesurés sur mêmes l individus soit du côté des prédicteurs soit du côté des variables réponses, soit des deux côtés. Ainsi, la modélisation d'une variable y seule par un ensemble de tableaux X_k de variables explicatives, relève de la régression multi-blocs (ou multi-tableaux). De même, pour l'explication d'un ensemble de N tableaux réponses Y_n par un seul tableau de variables explicatives X .

En pratique, en chimométrie, la question est presque toujours de modéliser, dans le but de prédire ensuite, une unique variable réponse y par un ensemble de tableaux de variables explicatives. Les modèles multi-blocs sont utiles dans ce cas pour éviter de sélectionner une par une les variables explicatives très nombreuses si on utilisait un modèle de régression multi-variables tel que celui de la régression PLS usuelle. L'aspect «groupe» des variables permet de simplifier l'interprétation du modèle en étudiant les différents groupes et non les différentes variables et de sélectionner des groupes de variables explicatives importantes (voir chap. 6.3.2) plutôt que de choisir les variables. Il est bien sûr possible que les modèles multi-blocs obtenus soient un peu moins explicatifs que des modèles de régression multi-variables usuels, mais l'interprétation devrait être plus aisée.

Il y a un autre domaine où les méthodes de régression multi-blocs peuvent être utiles : l'analyse sensorielle. En effet, dans ce domaine, il est d'usage de résumer les tableaux de données avant de faire des régressions multi-variables pour expliquer, par exemple, des évaluations sensorielles de produits établies par des consommateurs, par des évaluations sensorielles de ces mêmes produits par des experts (voir l'application III-5). L'évaluation, selon un attribut, des différents produits donnée par plusieurs consommateurs ou experts constitue à chaque fois un tableau. Résumer ces tableaux peut conduire à des pertes d'informations. Les méthodes multi-blocs sont bien adaptées pour éviter cela. Malheureusement, elles sont encore très mal connues en France dans ce domaine de recherche qu'est l'analyse sensorielle.

Ce chapitre est le plus important en volume. Nous avons essayé de donner un maximum de détails sur les méthodes issues de la littérature dont le but est d'expliquer un ou plusieurs tableaux par un ensemble de tableaux prédicteurs.

Ces méthodes, APLS, MBPLS, HPLS, SPLS sont des méthodes algorithmiques et n'optimisent pas initialement de critère. Nous nous sommes attachés à retranscrire pour chaque méthode l'algorithme présenté dans les publications initiales avec cependant parfois quelques petites corrections d'erreurs puis à les commenter.

Les différentes sections ne se déroulent pas toutes de la même manière. Nous avons essayé de clarifier en général certains points de ces algorithmes et avons donné leurs propriétés et l'écriture des modèles obtenus, sans démonstrations, sauf si elles n'ont pas été données (ou établies) dans la littérature.

Ce chapitre contient aussi des méthodes originales REMUB, ACIMO-PLS (VIVIEN & SABATIER, 2001c), ACIMOG-PLS, les deux dernières étant les versions dissymétrisées de l'ACIMO et l'ACIMOG présentées aux chapitres 2.4 et 3.4. Contrairement aux méthodes issues de la littérature, ces nouvelles approches sont toutes basées sur l'optimisation d'un critère de covariance à chaque étape. Ces sous-parties sont beaucoup plus denses et structurées et contiennent les démonstrations de toutes les propriétés et l'écriture détaillée des modèles.

5.1 Approche PLS : APLS

L'approche PLS, à différencier de la régression PLS, a été initialement proposée par WOLD (1975, 1982, 1985) puis étendue par LOHMÖLLER (1989) pour mettre en relation K matrices X_k mesurées sur les mêmes individus. Récemment elle a été reprise en français par TENENHAUS (1999).

L'approche PLS est une méthode algorithmique dans laquelle on suppose que chaque tableau X_k est l'expression d'une variable latente. LOHMÖLLER (1987) en a développé le programme, LVPLS 1.8, fonctionnant sous DOS. Il est disponible gratuitement sur le site internet de l'université de Virginie*.

Dans l'approche PLS, les tableaux X_k , $I \times J_k$, sont supposés centrés et réduits. L'objectif de l'approche PLS est d'estimer les variables latentes ξ_k , centrées et réduites, dont chaque tableau X_k est le reflet.

La première étape est de définir la nature des relations entre les blocs de variables et leur variable latente, et entre les variables latentes elles-mêmes. Ces relations de causalités sont alors résumées par un schéma fléché. La figure 5.1 présente un réseau de causalité entre K matrices explicatives et une matrice Y réponse. Ces relations sont exprimées ainsi :

- si X_k est supposée être une conséquence de ξ_k , X_k est un bloc réflexif. Alors, les variables de X_k sont reliées à leur variable latente ξ_k par la relation

$$X_k = \xi_k \lambda_k' + \epsilon_k, \quad (5.1)$$

où λ_k est un vecteur de taille $J_k \times 1$ à estimer et ϵ_k un terme d'erreur d'espérance nulle et non corrélé avec ξ_k .

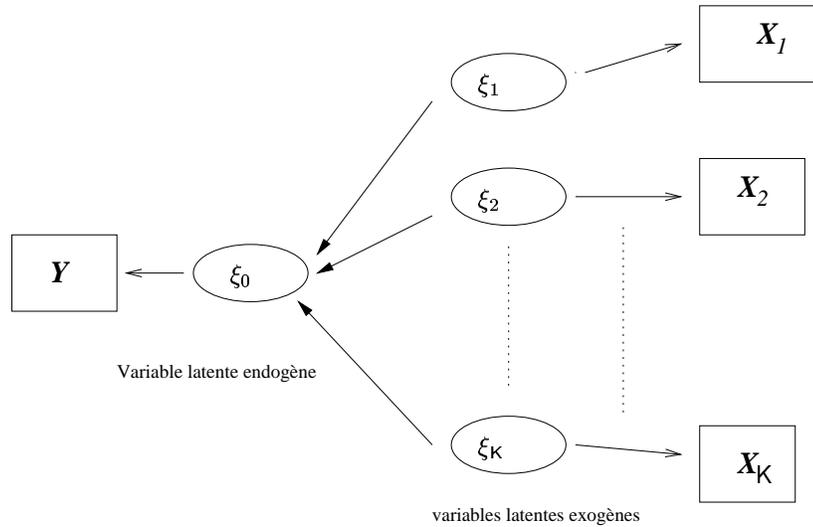
- Réciproquement, si le bloc X_k est supposé être une cause de sa variable latente ξ_k , on dira qu'il est formatif, et il est relié à ξ_k par la relation †

$$\xi_k = X_k \pi_k + e_k \quad (5.2)$$

où π_k est un vecteur de taille $J_k \times 1$ à estimer et e_k un terme d'erreur d'espérance nulle et non corrélé avec X_k .

* <http://kiptron.psyc.virginia.edu/disclaimer.html>

† Remarquons que sous cette hypothèse ξ_k est ce que nous avons l'habitude de nommer «composante».

FIG. 5.1 – Un réseau de causalité avec $K + 1$ blocs réfléchifs

Les signes des coordonnées des vecteurs π_k et λ_k peuvent éventuellement être précisés par l'utilisateur.

Ensuite, il faut décrire les relations entre les variables latentes. On note pour chaque variable latente ξ_k

$$\xi_k = \sum_{\ell=1}^K \beta_{k\ell} \xi_\ell + \zeta_k. \quad (5.3)$$

ζ_k est un terme d'erreur d'espérance nulle et non corrélée avec les variables latentes explicatives, apparaissant à droite de l'équation. Les coefficients $\beta_{k\ell}$ sont à estimer. Ils définissent la liaison entre les variables latentes. Un coefficient $\beta_{k\ell}$ est posé nul si la variable latente ξ_ℓ n'explique pas ξ_k . En particulier, les β_{kk} sont nuls.

Cette relation définit donc les variables latentes exogènes (qui en expliquent d'autres) et les variables latentes endogènes (qui sont expliquées par d'autres variables latentes).

L'utilisateur peut éventuellement préciser le sens des relations en imposant le signe $c_{k\ell}$ de la corrélation entre les variables latentes ξ_k et ξ_ℓ :

si $r(\xi_k, \xi_\ell) > 0$ alors on pose $c_{k\ell} = c_{\ell k} = 1$, si $r(\xi_k, \xi_\ell) < 0$, $c_{k\ell} = c_{\ell k} = -1$, et si $\beta_{k\ell} = 0$ on pose $c_{k\ell} = c_{\ell k} = 0$.

S'ils ne sont pas imposés, ils seront estimés par l'algorithme à partir des données.

Une fois ces hypothèses relationnelles faites, il s'agit d'estimer les variables latentes ξ_k . Pour cela, sont définies les estimations dites externes, $\hat{\xi}_k^e$, et les estimations dites internes, $\hat{\xi}_k^i$. Les $\hat{\xi}_k^e$ sont des combinaisons linéaires des variables :

$$\hat{\xi}_k^e = X_k \omega_k / \|X_k \omega_k\|. \quad (5.4)$$

Les estimations $\hat{\xi}_k^i$ sont liées aux estimations $\hat{\xi}_\ell^e$ des variables latentes liées à ξ_k :

$$\hat{\xi}_k^i = \sum_{\ell | \ell \neq k, c_{\ell k} \neq 0} e_{k\ell} \hat{\xi}_\ell^e / \left\| \sum_{\ell | \ell \neq k, c_{\ell k} \neq 0} e_{k\ell} \hat{\xi}_\ell^e \right\|. \quad (5.5)$$

Les coefficients e_{kl} peuvent être choisis selon trois schémas : le schéma centroïde (proposé par WOLD (1982)), le schéma factoriel et le schéma structurel (proposés par LOHMÖLLER (1989)). Ensuite, les deux estimations, interne et externe, sont mises en relation, suivant l'un des deux modes dits A et B, afin d'obtenir les équations dites de stationnarités, qui permettent de calculer $\hat{\xi}_k^e$ itérativement. Le programme LVPLS 1.8 propose plus de possibilités de relations, en particulier, des relations non-linéaires. TENENHAUS (1999) propose un troisième mode, mode C, qui consiste à effectuer la régression PLS de $\hat{\xi}_k^e$ sur les $\hat{\xi}_\ell^i$ qui lui sont liées. Enfin, on estime λ_k , π_k et β_{ki} par $\hat{\lambda}_k$, $\hat{\pi}_k$ et $\hat{\beta}_{ki}$, à l'aide de régressions ordinaires. Finalement, on aboutit aux modèles suivants pour les blocs endogènes (c'est-à-dire réponses) :

$$\hat{X}_k = \sum_{i=1, \dots, K | c_{ik} \neq 0} (\hat{\beta}_{ki} X_i \omega_i) p_k'$$

Les détails techniques sont donnés dans l'algorithme, ci-après. Des exemples sont traités dans (TENENHAUS, 1999).

Ce dernier montre comment on peut retrouver la régression PLS, et la première composante de l'ACP, de l'Analyse Canonique, de l'Analyse de Redondance (TENENHAUS, 1995), des Analyses Canoniques Généralisées de Horst (HORST, 1961) et Carroll (CARROLL, 1968).

L'algorithme

étape 0 : schéma de causalité

Définir la nature des matrices X_k (réflectifs ou formatifs) et de leurs variables latentes ξ_k (endogène ou exogène).

Définir les liaisons (5.3) entre les variables latentes.

étape 1 : choix du calcul de e_{kl}

- *Les estimations externes*: $\hat{\xi}_k^e = X_k \omega_k / \|X_k \omega_k\|$.
- *Les estimations internes*: $\hat{\xi}_k^e = \sum_{\ell | \ell \neq k, c_{\ell k} \neq 0} e_{k\ell} \hat{\xi}_\ell^e / \|\sum_{\ell | \ell \neq k, c_{\ell k} \neq 0} e_{k\ell} \hat{\xi}_\ell^e\|$.

Avec :

- **schéma centroïde** :

$e_{ki} = c_{ki}$ où les c_{kl} non nuls sont fixés par l'utilisateur égaux à 1 ou -1 , ou bien seront estimés par le signe r_{kl} de la corrélation entre $\hat{\xi}_k^e$ et $\hat{\xi}_\ell^e$.

- **schéma factoriel** :

$e_{kl} = r_{kl} = \text{cor}(\hat{\xi}_k^e, \hat{\xi}_\ell^e)$.

- **schéma structurel** :

$e_{kl} = b_{kl}$ si ξ_ℓ explique ξ_k , où les b_{kl} sont les coefficients de la régression multiple de $\hat{\xi}_k^e$ sur $\hat{\xi}_\ell^i$, estimations internes des ξ_ℓ qui expliquent ξ_k ,

$e_{kl} = r_{kl}$ si ξ_ℓ est expliquée par ξ_k .

étape 2 : choix de la relation entre $\hat{\xi}_k^e$ et $\hat{\xi}_k^i$

- **Mode A** :

$$\hat{\xi}_k^e = X_k X_k' \hat{\xi}_k^i / \|X_k X_k' \hat{\xi}_k^i\|$$

Condition de stationnarité :

$$\hat{\xi}_k^e = X_k X_k' \sum_{\ell | \ell \neq k, c_{k\ell} \neq 0} e_{k\ell} \hat{\xi}_\ell^e / \|X_k X_k' \sum_{\ell | \ell \neq k, c_{k\ell} \neq 0} e_{k\ell} \hat{\xi}_\ell^e\|.$$

$$\text{Ainsi, } \omega_k = X_k' \hat{\xi}_k^i / \|X_k X_k' \sum_{\ell | \ell \neq k, c_{k\ell} \neq 0} e_{k\ell} \hat{\xi}_\ell^e\|.$$

- Mode B :

$$\hat{\xi}_k^e = X_k(X_k'X_k)^{-1}X_k'\hat{\xi}_k^i/\|X_k(X_k'X_k)^{-1}X_k'\hat{\xi}_k^i\|$$

Condition de stationnarité :

$$\hat{\xi}_k^e = X_k(X_k'X_k)^{-1}X_k' \sum_{\ell|\ell \neq k, c_{k\ell} \neq 0} e_{k\ell} \hat{\xi}_\ell^e / \|X_k(X_k'X_k)^{-1}X_k' \sum_{\ell|\ell \neq k, c_{k\ell} \neq 0} e_{k\ell} \hat{\xi}_\ell^e\|.$$

$$\text{Ainsi, } \omega_k = (X_k'X_k)^{-1}X_k'\hat{\xi}_k^i/\|X_k(X_k'X_k)^{-1}X_k' \sum_{\ell|\ell \neq k, c_{k\ell} \neq 0} e_{k\ell} \hat{\xi}_\ell^e\|.$$

étape 3 : estimation de ξ_k

- Initialiser les $\hat{\xi}_k^e$: $\hat{\xi}_k^{e(0)} = X_k^1$, première colonne de X_k .
- Jusqu'à convergence des $\hat{\xi}_k^e$, actualiser $\hat{\xi}_k^{e(a)}$, de l'itération courante, grâce à l'équation de stationnarité choisie à l'étape 2, en utilisant les estimations $\hat{\xi}_\ell^{e(a-1)}$, de l'itération précédente, des variables latentes qui lui sont liés.

étape 4 : estimation des coefficients λ_k , π_k et $\beta_{k\ell}$

Les coefficients estimés $\hat{\lambda}_k$ (*resp.* $\hat{\pi}_k$) sont obtenus par la régression de x_{kh} , variable h de X_k sur $\hat{\xi}_k^e$, (*resp.* de $\hat{\xi}_k^e$ sur X_k).

Les $\hat{\beta}_{ki}$ sont donnés par la régression multiple de $\hat{\xi}_k^e$ sur les $\hat{\xi}_\ell^e$ qui lui sont liées :

$$\hat{\xi}_k^e = \sum_{\ell|\beta_{k\ell} \neq 0} \hat{\beta}_{k\ell} \hat{\xi}_\ell^e.$$

étape 5 : si les $c_{k\ell}$ sont fixés *a priori* :

Vérifier que les $c_{k\ell}$ sont égaux aux $r(\hat{\xi}_k^e, \hat{\xi}_\ell^e)$.

étape 6 : modèles pour les matrices réponses (endogènes) :

- Régression des X_k endogènes sur $\hat{\xi}_k^e$:

$$X_k = \hat{\xi}_k^e p_k'$$

- Expression des modèles :

$$\hat{X}_k = \sum_{\ell|c_{k\ell} \neq 0} (\hat{\beta}_{k\ell} X_\ell \omega_\ell) p_k'.$$

—■—

Si on le désire, le programme LVPLS 1.8 propose de calculer des variables latentes d'ordre supérieur en utilisant les déflations de tous les blocs X_k sur ξ_k ou seulement celles des blocs réponses.

On peut trouver une application récente de l'APLS ainsi que les détails pour l'utilisation du programme LVPLS 1.8 dans (CHATELIN, VINZI & TENENHAUS, 2002).

L'Approche PLS fait des méthodes connues sous le nom de «path modeling with latent variables». Parmi ces méthodes, il y a une autre manière de procéder à l'estimation des variables latentes, en utilisant l'approche du maximum de vraisemblance. Le logiciel LISREL[®] (JÖRESKOG & SÖRBUM, 1984) fonctionne ainsi.

5.2 Multi-Block Regression : MBPLS

WANGEN & KOWALSKI (1988) ont introduit un algorithme PLS pour de multiples blocs basé sur un algorithme tout d'abord présenté par S. Wold, H. Martens et H. Wold, qui porte le nom de «The SW-Harald-HW multiblock algorithm» dans (WOLD, 1984). On notera cette méthode MBPLS.

La MBPLS permet de mettre en relation plusieurs groupes de variables, c'est-à-dire plusieurs matrices. On peut avoir des tableaux seulement prédicteurs, des tableaux seulement réponses et d'autres qui sont à la fois prédicteurs et réponses. Comme dans l'approche PLS, on définit des relations entre eux, mais ici, l'algorithme est de type NIPALS. L'algorithme étant assez lourd, **nous détaillerons ici deux situations : le cas où on possède un ensemble de N blocs réponses Y_n et plusieurs blocs explicatifs X_1, \dots, X_K et son cas particulier lorsque $N = 1$.** Ainsi, l'algorithme que nous donnons est un cas particulier de l'algorithme général auquel le lecteur devra se référer pour une situation plus complexe (WANGEN & KOWALSKI, 1988).

La MBPLS est une méthode «component-wise», des composantes sont calculées étape après étape. Le passage d'une étape a à la suivante, se fait par l'utilisation des déflations des tableaux X_k et Y_n sur des composantes. WANGEN & KOWALSKI (1988) ont proposé des déflations des blocs sur leur propre composante. Dans la littérature, le terme «block score deflation» est employé pour désigner ce type de déflations. Dans le cas où $N = 1$, WESTERHUIS & COENEGRACHT (1997) proposent d'utiliser plutôt la déflation des tableaux sur la super-composante associée aux prédicteurs (combinaison linéaire de toutes les variables explicatives) en montrant qu'ainsi les modèles sont plus explicatifs.

A chaque étape a , deux composantes par bloc et les vecteurs de poids associés sont calculés : une composante qui «explique» (et t_k) et une qui est «expliquée» (u_k^x et u_n). Ces deux types de composantes sont présentes pour décrire les caractères «prédictif» et «réponse» de chaque bloc. De plus, des super-composantes explicatives (combinaison linéaire de tous les prédicteurs de Y_n) et réponse (combinaison linéaire de toutes les variables expliquées par X_k), $t_{T,n}$ et $u_{U,k}$, associées aux matrices $T = [t_1, \dots, t_K, u_1^x, \dots, u_K^x]$ et $U = [u_1, \dots, u_N, t_1^y, \dots, t_N^y]$ sont calculées. Toutes ces composantes sont calculées de manière itérative, comme dans l'algorithme NIPALS.

5.2.1 Cas de K tableaux explicatifs et N tableaux réponses

Nous allons ici reprendre l'algorithme initial pour un ensemble de tableaux prédicteurs X_k et un ensemble de tableaux réponses Y_n .

Nous notons t_k et u_k^x les composantes «explicatives» et «réponses» de X_k . Et t_n^y et u_n celles de Y_n . ω_k sera le facteur associé à X_k et t_k , c_n celui associé à Y_n et u_n . Enfin, nous noterons $t_{T,n}$ et $\omega_{T,n}$ la super-composante explicative et le super-facteur respectivement associés à Y_n et $u_{U,k}$ et $c_{U,k}$ la super-composante réponse et le super-facteur associé à X_k . Cet algorithme est illustré par la figure 5.2.

étape 0 : initialisation

On fixe u_n et t_k égaux à la variable de Y_n et X_k ayant respectivement la plus grande variance.

étape 1 : calcul des composantes «explicatives»

$t_n^y = u_n$ car Y_n est un tableau réponse.

X_k prédisant tous les Y_n , une super-matrice U est utilisée pour calculer les super-composantes.

$$\spadesuit U = [u_1, \dots, u_N, t_1^y, \dots, t_N^y] = [u_1, \dots, u_N, u_1, \dots, u_N].$$

$$c_{U,k} = U^t t_k, \quad c_{U,k} = c_{U,k} / \|c_{U,k}\|.$$

$$u_{U,k} = U c_{U,k}.$$

$$\omega_k = X_k^t u_{U,k}.$$

$$\omega_k = \omega_k / \|\omega_k\|.$$

$$t_k = X_k \omega_k.$$

étape 2 : calcul des composantes «réponses»

$u_k^x = t_k$ (car les X_k ne sont pas prédits).

Y étant prédit par plusieurs tableaux, on utilise des super-composantes $t_{T,n}$ et super-facteurs $\omega_{T,n}$.

$$\spadesuit T = [t_1 | \dots | t_K | u_1^x | \dots | u_K^x] = [t_1 | \dots | t_K | t_1 | \dots | t_K].$$

$$\omega_{T,n} = T^t u_n; \quad \omega_{T,n} = \omega_{T,n} / \|\omega_{T,n}\|,$$

$$t_{T,n} = T \omega_{T,n}.$$

$$c_n = Y^t t_{T,n}; \quad c_n = c_n / \|c_n\|.$$

$$u_n = Y_n c_n (= t_n^y).$$

On teste la convergence de u_n et t_k : s'il y a convergence, on passe à l'étape 3, sinon on recommence à l'étape 1.

étape 3 : calcul des «loadings» prédictes

$$p_k = X_k^t t_k / (t_k^t t_k).$$

étape 4 : calcul des coefficients de régression

$$Y_n \text{ est prédit par } T : b_{T,n} = u_n^t t_{T,n} / (t_{T,n}^t t_{T,n}).$$

étape 5 : calcul des résidus

$$E_k = X_k - t_k p_k^t = X_k - P_{t_k} X_k,$$

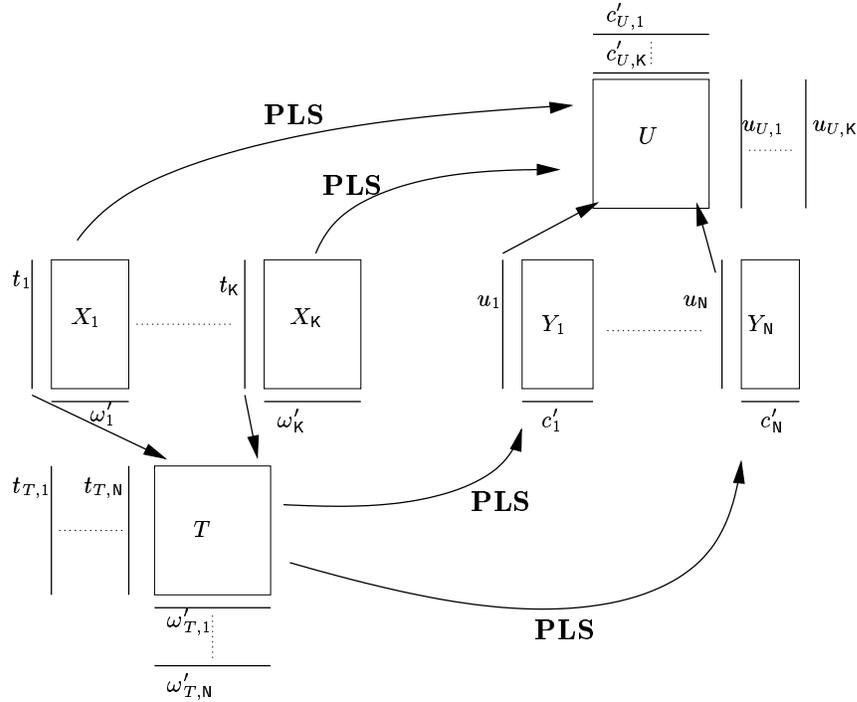
$$E_n = Y - \hat{u}_n c_n^t = Y - P_{t_{T,n}} Y_n, \text{ où } \hat{u}_y \text{ est l'estimation de } u_n : \hat{u}_n = b_{T,n} t_{T,n}. \quad \blacksquare$$

Si l'on désire des composantes supplémentaires, on reprend l'algorithme en utilisant les résidus E_k et E_n à la place de X_k et Y_n .

Comme les composantes u_k^x sont égales à t_k et les composantes t_k^y à u_n^y , il revient exactement au même de ne pas les calculer et de remplacer les lignes de l'algorithme précédées de \spadesuit par $T = [t_1 \ \dots \ t_K]$ et $U = [u_1 \ \dots \ u_N]$. C'est à cette dernière définition de l'algorithme que nous ferons référence par la suite.

Cet algorithme ne semble pas optimiser un critère à chaque étape. Cependant, nous avons remarqué que cet algorithme est équivalent à effectuer itérativement les $K+N$ régressions PLS de U sur X_k et de Y_n sur T , jusqu'à convergence des composantes t_k et u_n . En effet, il suffit de montrer que les vecteurs de poids vérifient le problème aux valeurs propres de la régression PLS :

FIG. 5.2 – L'algorithme de la MBPLS pour K matrices explicatives X_k et N tableaux réponses Y_n



à la convergence de l'algorithme, $c_{U,k}$ et ω_k vérifient :

$$\|U' t_k\| \|X'_k u_{U,k}\| \omega_k = X'_k U U' X_k \omega_k, \quad (5.6)$$

$$\|U' t_k\| \|X'_k u_{U,k}\| c_{U,k} = U' X_k X'_k U c_{U,k}. \quad (5.7)$$

Grâce à ces deux équations, $t'_k u_{U,k} = \|U' t_k\| = \|X'_k u_{U,k}\|$. En effet :

$$\begin{aligned} t'_k u_{U,k} &= \omega'_k X'_k U c_{U,k} \\ &= \omega'_k X'_k U U' t_k / \|U' t_k\|. \end{aligned}$$

L'équation (5.6) et le fait que ω_k soit de norme unité mène à :

$$\begin{aligned} t'_k u_{U,k} &= \omega'_k \omega_k \|X'_k u_{U,k}\| \\ &= \|X'_k u_{U,k}\|. \end{aligned}$$

De la même manière en utilisant l'équation (5.7) et $c'_{U,k} c_{U,k} = 1$, nous obtenons :

$$\begin{aligned} u'_{U,k} t_k &= c'_{U,k} U X_k \omega_k \\ &= c'_{U,k} U X_k X'_k u_{U,k} / \|X'_k u_{U,k}\| \\ &= c'_{U,k} c_{U,k} \|U' t_k\| = \|U' t_k\|. \end{aligned}$$

Ceci montre qu'à convergence $(t'_k u_{U,k})^2$ est valeur propre de $X'_k U U' X_k$ et $U' X_k X'_k U$. De manière analogue, il est possible de montrer que $(u'_n t_{T,n})^2 = \|T' u_n\| = \|Y'_n t_{T,n}\|$ est valeur propre de $Y'_n T T' Y_n$ et $T' Y_n Y'_n T$ associée respectivement à c_n et $\omega_{T,n}$. Les problèmes aux valeurs propres à résoudre sont donc ceux de la régression PLS et $(t'_k u_{U,k})^2$ et $(u'_n t_{T,n})^2$ sont maximum. La figure 5.2 illustre l'algorithme.

Nous n'avons pas vu ce résultat dans la littérature.

5.2.2 Cas de K tableaux explicatifs et un tableau réponse (N = 1)

Lorsque $N = 1$, il n'y a plus qu'une seule super-composante $t_T = \sum_{k=1}^K \omega_T(k)t_k$. Ainsi, tous les tableaux réponses Y_n sont déflatés sur la même super-composante t_T .

WESTERHUIS & COENEGRACHT (1997) proposent de «déflater» aussi les tableaux X_k sur t_T au lieu de les «déflater» sur t_k . Ainsi, disent-ils, la reconstitution des tableaux X_k est meilleure en terme de variance expliquée. L'année suivante, WESTERHUIS, KOURTI & MACGREGOR (1998) montrent que la MBPLS avec $N = 1$ ainsi utilisée, fournit des super-composantes t_T et $u = u_U$ identiques aux composantes de la régression PLS de Y sur le tableau conjoint $X = [X_1 \ \cdots \ X_K]$, puis WESTERHUIS & SMILDE (2001) donnent un nouvel algorithme qui consiste à effectuer cette régression PLS, puis à en déduire les composantes par tableaux t_k , et le vecteur des super-poids ω_T . A chaque étape a , ils proposent donc de suivre la démarche suivante :

étape 1 : PLS

Effectuer la régression PLS de Y sur $X = [X_1 \ \cdots \ X_K]$, qui fournit $t_T = X\omega$,
 $u = Yc$.

étape 2 : calcul des composantes par tableaux

$\omega_k = \omega[k]/\|\omega[k]\| \quad k = 1, \dots, K$
 $t_k = X_k\omega_k, \quad k = 1, \dots, K$

étape 3 : calcul des super-poids

$\omega_T = [t_1 \ \cdots \ t_K]' u/u'u$
 $\omega_T = \omega_T/\|\omega_T\|$

L'étape suivante est l'étape du calcul des résidus (la déflation). Celle-ci est la même que l'étape 5 de l'algorithme initial de WANGEN & KOWALSKI (1988) que nous avons exposé page 83, à la différence près que X_k est projeté soit sur t_k , comme dans l'algorithme initial, soit sur t_T , auquel cas $E_k = X_k - P_{t_T}X_k$.

Notons que dans l'étape 3, la division par $u'u$ est inutile, puisque ω_T est réduit juste après.

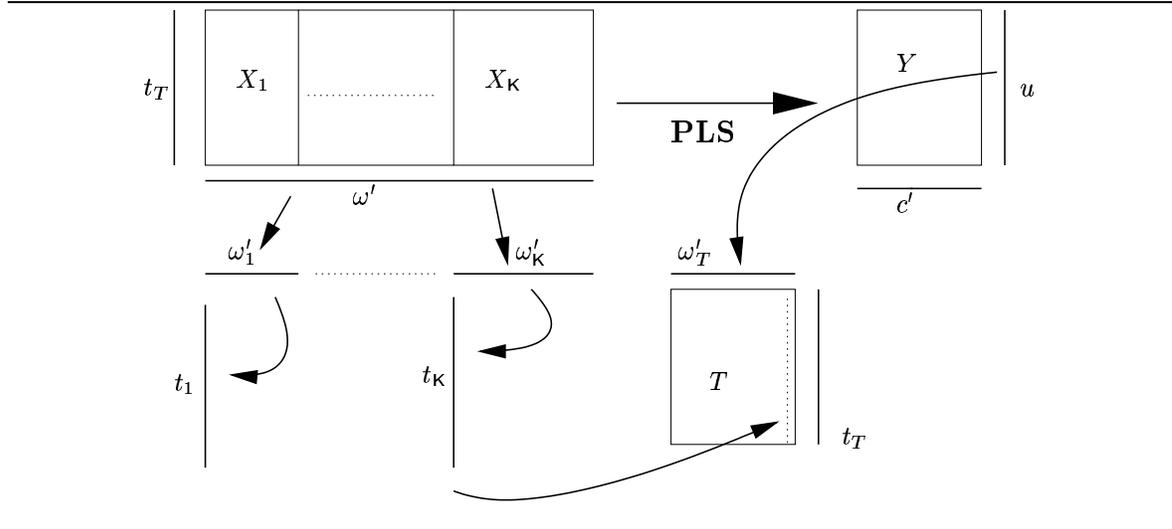
Cet algorithme montre que lorsque $N = 1$, la régression MBPLS n'est pas vraiment «multi-blocs» : les composantes par blocs (t_k) ne sont pas ajustées par tableaux. Elles sont simplement déduites de la super-composante t_T qui, elle, est optimisée pour le tableau X . Ici, les composantes t_k sont uniquement calculées pour donner des éléments d'interprétation supplémentaires dans la régression PLS usuelle de deux tableaux. La figure 5.3 illustre cet algorithme. WESTERHUIS & SMILDE (2001) discutent aussi de la déflation à effectuer sur les tableaux X_k . La déflation sur la super-composante aboutit à une meilleure reconstitution des X_k (WESTERHUIS & COENEGRACHT, 1997), en terme de variance expliquée, mais ils soulignent le problème d'interprétation des résultats dus à la déflation sur cette super-composante t_T : dès la seconde étape, $a = 2$, l'information contenue dans tous les tableaux est mélangée.

Pour résoudre cela, ils proposent une nouvelle modification de l'algorithme où les X_k ne sont pas «déflatés», et où les super-composantes $t_{T,a}$ sont orthogonalisées d'une étape à l'autre. Les composantes par bloc, $t_{k,a} = X_k\omega_{k,a}$ alors obtenues sont différentes des composantes issues de l'algorithme initial (où $t_{k,a} = X_k^{(a-1)}\omega_{k,a}$). Par contre, les super-composantes sont les mêmes. Dans le chapitre 7.1 de cette partie nous montrons ce résultat d'une manière différente de

celle de ses auteurs.

Cette situation ($N = 1$) est la plus fréquente en chimiométrie, domaine où est utilisée et explorée la MBPLS. De plus, cette méthode semble de plus en plus citée comme la méthode de référence parmi les méthodes multi-blocs, la PLS hiérarchique (WOLD et al., 1996) n'ayant pas fait ses preuves (chapitre 5.3). Il nous paraît, à la vue, des résultats précédents que dans ce cas précis, $N = 1$, elle n'est pas vraiment multi-bloc et que de nouvelles approches doivent être étudiées.

FIG. 5.3 – L'algorithme de la MBPLS pour K matrices explicatives et un tableau réponse Y



Lorsque l'on explique N tableaux par K tableaux, les composantes $t_{k,1}, \dots, t_{k,A}$ sont orthogonales, les autres ne le sont pas, A désignant le nombre de déflations effectuées, autrement dit le rang des modèles. Lorsque $N = 1$, selon la déflation utilisée, ce sont soit les composantes $t_{k,1}, \dots, t_{k,A}$ qui sont orthogonales (WANGEN & KOWALSKI, 1988), soit les super-composantes $t_{T,1}, \dots, t_{T,a}$ (WESTERHUIS & COENEGRACHT, 1997).

5.3 PLS hiérarchique : HPLS

Le nom de «PLS Hiérarchique» (HPLS) regroupe en fait plusieurs méthodes du même type, ayant des algorithmes très proches les uns des autres et fournissant des résultats analogues. L'histoire de cette méthode peut se trouver dans (WESTERHUIS, KOURTI & MACGREGOR, 1998). Nous détaillons brièvement ici la dernière en date, proposée par (WOLD, KETTANEH & TJESSEM, 1996), qui semble la plus citée.

La régression PLS Hiérarchique (HPLS) est une méthode proposée par WOLD, KETTANEH & TJESSEM (1996) comme méthode pour résoudre les problèmes d'interprétabilité rencontrés lorsque l'on veut modéliser une matrice de réponses par une matrice de variables explicatives possédant l'une et/ou l'autre un grand nombre de variables, c'est-à-dire plus de 100. Cette situation est fréquemment rencontrée en chimiométrie, avec par exemple des données issues des spectres proches infrarouges (NIRS).

Dans ces situations, on cherche souvent à réduire et/ou à choisir le nombre de variables. Mais

si on en supprime, on risque d'aboutir à des modèles erronés et ne pouvant servir pour faire de la prédiction. Aussi, l'HPLS consiste à utiliser un algorithme de type NIPALS sur des ensembles de variables (X_k et Y_n) préalablement définies au sein de deux principales matrices X et Y . Cet algorithme calcule simultanément, de manière itérative, des résumés partiels (t_k et u_n) des variables de X_k et Y_n et des résumés globaux, c'est-à-dire des résumés de tous les prédicteurs (t_T) et de toutes les réponses (u_U). Ces résumés (ou composantes globales ou encore super-composantes), sont exprimées en fonction des résumés partiels.

L'algorithme est séparé en deux niveaux qui sont appelés le «sous niveau» et le «super-niveau». Lors du sous-niveau, une composante par matrice (X_k et Y_n) est calculée ainsi qu'un vecteur de poids par matrice. Les composantes explicatives constituent alors une matrice T et les composantes réponses, la matrice U . Alors, le super niveau consiste à faire une régression PLS de U sur T , qui fournit la super-composante t_T , résumé de toutes les variables. C'est ainsi que WOLD et al. (1996) présentent cet algorithme.

L'algorithme utilisé consiste à imbriquer plusieurs étapes de type NIPALS. En fait, lors d'une étape a , les composantes partielles et la super-composante t_T sont calculées en même temps. C'est-à-dire, qu'il n'y a pas un algorithme itératif qui résout d'abord le sous-niveau pour ensuite effectuer une PLS, mais un seul et même algorithme itératif qui fait tout en même temps (WOLD et al., 1996).

A l'issue de cette méthode, le résultat est composé de plusieurs matrices de composantes et de vecteurs de poids (une pour chaque matrice X_k et Y_n) issues des sous-niveaux et de super-matrices de composantes et de vecteurs de poids (une pour les réponses et une pour les prédicteurs) issues du dernier super-niveau. On peut donc étudier les relations entre les matrices X et Y (représentant toutes variables explicatives et à expliquer) à l'aide des «super-résultats» et les groupes de variables en particulier (X_k et Y_n) avec les résultats du sous-niveau.

Dans cette la version de l'HPLS décrite ici, les super-composantes explicatives sont orthonormées entre elles : à chaque étape, on procède par déflation des variables sur les super-composantes précédentes.

La principale différence entre la HPLS présentée ici et les plus anciennes consiste justement en l'introduction de cette orthogonalité (WESTERHUIS et al., 1998). Notons que dans la première version de l'HPLS les super-composantes n'étaient pas normées mais les super-poids l'étaient (WOLD, HELLBERG, LUNDSTEDT, SJOSTROM & WOLD, 1987).

L'algorithme décrit dans (WESTERHUIS et al., 1998) est celui de SLAMA (1991) lequel avait alors rajouté l'orthogonalité de super-composantes par rapport à celui de (WOLD et al., 1987).

WOLD et al. (1996) proposent aussi un algorithme similaire pour une ACP multi-blocs Hiérarchique (HPCA).

L'algorithme

Chacun des K blocs explicatifs X_k est de taille $l \times J_k$ et les N blocs réponses Y_n est de taille $l \times M_n$.

Les composantes partielles et poids associés à X_k sont notés t_k et p_k . Ceux pour Y_n sont notés u_n et q_n . On posera $T = [t_1 | \dots | t_K]$ et $U = [u_1 | \dots | u_N]$. La super-composante explicative et

son vecteur de poids sont notés t_T et ω_T . Quant aux super-éléments réponses, ils sont notés u_U et c_U .

Nous ajouterons un a en indice à tous ces éléments lorsqu'on voudra spécifier à quelle étape a ils correspondent.

Les données sont centrées et réduites éventuellement.

étape 0 : initialisation

Poser t_T égale à la variable de X ayant la variance maximale, puis normaliser t_T .

Poser Δ égal 100 $\mathbf{1}_1$, utilisé pour tester la convergence de t_T .

étape 1 : les sous niveaux

$$t_{old} = t_T.$$

$$p'_k = t'_T X_k / t'_T t_T; q'_n = t'_T Y_n / t'_T t_T.$$

$$t_k = X_k p_k / J_k; u_n = Y_n q_n / M_n.$$

étape 2 : tester la convergence de t_T

Si $\Delta' \Delta / t'_T t_T < 10^{-8}$ ou *nombre.itérations* > 200 , aller directement à l'étape 4.

Sinon, continuer.

étape 3 : le Super-Niveau

Calcul de c_U et u_U : $c'_U = t'_T U$; $u_U = U c_U / (c'_U c_U)$.

Normalisation u_U et transformation de c_U pour que $u_U = U c_U$: $d = \|u_U\|$, $u_U = u_U / d$; $c_U = c_U / d^\ddagger$.

$$\omega'_T = u'_U T / u'_U u'_U = u'_U T; t_T = T \omega_T.$$

«Petite étape» pour s'assurer de l'orthogonalité des super-composantes entre elles :

Si $a > 1$

$$t_T = t_T - T_T T'_T t_T \text{ où } T_T = [t_{T1} | \dots | t_{Ta-1}].$$

$$\omega_T = (T'_T T)^{-1} T'_T t_T, \text{ pour que l'on ait toujours } t_T = T \omega_T.$$

Normalisation de t_T : $d = \|t_T\|$, $t_T = t_T / d$, $\omega_T = \omega_T / d^\S$.

Calcul pour que $t'_T u_U = 1$: $d = t'_T u_U$, $u_U = u_U / d$, $c_U = c_U / d^\ddagger$

$$\Delta = t_T - t_{old}$$

nombre.itérations = *nombre.itérations* + 1 et retour à l'étape 1.

étape 4 : calcul des résidus

$$E_k^x = X_k - t_T p'_k = X_k - P_{t_T} X_k.$$

$$E_n^y = Y_n - t_T q'_n = Y_n - P_{t_T} Y_n.$$

étape 5 : passage à l'étape suivante pour une composante supplémentaire

Si l'on désire une composante supplémentaire, reprendre à partir de l'étape 0 en

utilisant les résidus, E_k^x et E_n^y , calculés à l'étape 4 à la place des tableaux X_k et Y_n .

—■—

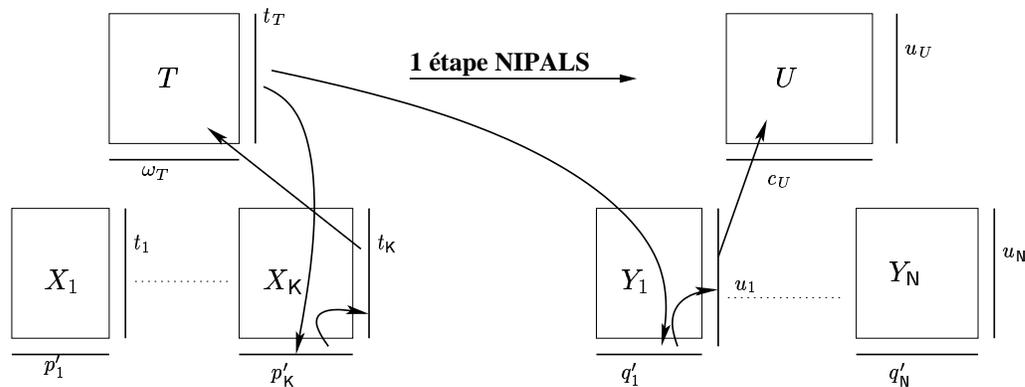
La figure 5.4 illustre cet algorithme.

Notons que la «petite étape» assurant l'orthogonalité des super-composantes explicatives est inutile, puisqu'à chaque étape les tableaux résidus de la projection orthogonale sur la super-composante précédente sont utilisés.

[‡] Dans l'article de base (WOLD et al., 1996), il est noté $c_U = c_U d$, ce qui ne nous paraît pas correct.

[§] Même remarque: dans l'article il est écrit $\omega_T = \omega_T d$.

FIG. 5.4 – L'algorithme de l'HPLS



L'algorithme ayant été détaillé, nous pouvons maintenant donner l'expression des modèles d'ordre A pour les tableaux réponses. L'écriture la plus simple est $\hat{Y}_n^A = \sum_{a=1}^A P_{t_{T,a}} Y_n^{(a-1)}$. On peut aussi écrire $\hat{Y}_n^A = \sum_{a=1}^A P_{t_{T,a}} Y^{(a-1)}$.

Grâce à l'orthogonalité des super-composantes, on a pour les erreurs $E_A^y = Y - \sum_{a=1}^A P_{t_{T,a}} Y$. Il est possible de montrer que chaque composante $t_{T,a}$ est combinaison linéaire des variables initiales (résultat cité mais non exprimé ni démontré par WOLD et al. (1996)). $t_{T,a}$ peut s'écrire de la façon suivante: $t_{T,a} = X\Omega_a$ avec $\Omega_1 = \text{Diag}(p_{k,1})\text{Diag}(1/J_k)\omega_{T,1}$ et $\Omega_a = \text{Diag}(p_{k,a})\text{Diag}(1/J_k)\omega_{T,a} - \sum_{l=1}^{a-1} (\Omega_l \Omega_l' X' X) \text{Diag}(p_{k,a}) \text{Diag}(1/J_k)\omega_{T,a}$ pour $a = 2, \dots, A$, où $\text{Diag}(p_{k,a})$ est la matrice de taille $\sum_k J_k \times K$ diagonale par bloc, le bloc k étant le vecteur $p_{k,a}$.

Ainsi, le modèle \hat{Y}_n^A de rang A peut s'écrire de la forme $X\hat{\beta}_A^n$ avec $\hat{\beta}_A^n = \sum_{a=1}^A \Omega_a \Omega_a' X' Y_n$. En considérant ce modèle, on ne peut pas juger de l'influence d'un groupe de prédicteurs sur le groupe réponse considéré.

Le modèle \hat{Y}_n^A peut aussi s'écrire de deux autres façons, mais plus complexes :

$$\hat{Y}_n^A = \sum_k \left[\sum_a \omega_{T,a}[k] \sum_{k'} \omega_{T,a}[k'] t_k t_{k'}' Y_n \right],$$

$$\hat{Y}_n^A = \sum_k X_k \left[\sum_a \Omega_a[k] \sum_{k'} \Omega_a'[k'] X_{k'}' Y_n \right].$$

Cette méthode ne maximise aucun critère et l'algorithme est très flou : les composantes et les poids sont normés à l'unité puis dénormés puis renormés Seules les super-composantes t_T sont orthogonales. De plus, celles-ci sont normées, ont une covariance égale à 1 avec les super-composantes u_U de même ordre et sont orthogonales aux u_U d'ordre inférieur. Par contre, les super-composantes u_U ne sont plus normées à la sortie de l'algorithme, cela à cause de la fin de l'étape 3. Ni les poids p_k , q_n , w_T , ni c_U ne sont normés ni orthogonaux. L'interprétation paraît alors très compliquée.

De plus, WESTERHUIS et al. (1998) montrent que l'HPLS semble mieux expliquer les tableaux X_k que les tableaux Y_n , alors que son but est justement de modéliser ces derniers. Nous avons

aussi constaté ce phénomène dans la pratique (voir III-4). Nous pensons que cela peut être dû au fait que les composantes u_n et les composantes t_k sont calculées en fonction de la super-composante explicative t_T . L'information n'est pas croisée. La super-composante u_U n'est pas utilisée dans le sous-niveau.

5.4 Serial PLS : SPLS

Cette méthode, notée SPLS, due à BERGLUND & WOLD (1999), est aussi une méthode de type « multi-blocs ». Mais celle-ci ne s'applique qu'au cas d'un seul bloc Y de réponses et plusieurs tableaux (ou blocs) de prédicteurs (formant un cube ou non). Sa grande différence avec les méthodes précédentes, est que les blocs ne sont pas traités de façon parallèle, mais en série, comme le nom de la méthode le laisse entendre. Ceci signifie que les blocs sont utilisés les uns après les autres et pas simultanément comme dans les autres méthodes.

La démarche est la suivante : K régressions PLS sont réalisées successivement en utilisant le bloc X_k et la partie de Y qui n'est pas expliquée par les blocs précédents. Le modèle finalement obtenu peut s'écrire de la forme $\hat{Y} = X_1\beta_1 + X_2\beta_2 + \dots + X_K\beta_K$. Il correspond à la somme des modèles résultant des différentes régressions PLS.

Un algorithme itératif spécifique a été développé pour obtenir le résultat. Il ne consiste pas simplement à effectuer des régressions PLS successives. Les résidus sont mis à jour après chaque itération. Si après les K régressions PLS il n'y a pas convergence du modèle, alors on reprend les PLS successives en utilisant les derniers résidus à la place de Y .

Les auteurs précisent que, bien que ces itérations puissent sembler non nécessaires, ils ont remarqué sur des exemples pratiques que cette procédure aboutissait à de meilleurs modèles en terme de prédiction. Avec les itérations, le premier modèle PLS est lui aussi calculé à partir de résidus, comme les autres. Les auteurs explicitent l'algorithme (voir ci-dessous) pour le cas de blocs prédictifs X_1 et X_2 .

L'algorithme

étape 0 : initialisation

$$F_2 = Y$$

Fixer le nombre de composantes à retenir dans les différentes PLS.

étape 2 : première PLS

Effectuer la régression PLS d'ordre fixé au préalable de F_2 sur X_1 .

Calculer $F_1 = Y - P_{T_1}F_2$, où T_1 est la matrice des composantes PLS retenues.

étape 3 : seconde PLS

Effectuer la régression PLS d'ordre fixé au préalable de F_1 sur X_2 .

Calculer $F_2 = Y - P_{T_2}F_1$, où T_2 est la matrice des composantes PLS retenues.

étape 4 : test de convergence

Tester la convergence de F_2 . S'il n'y a pas convergence, retourner à l'étape 2.

Le nombre de composantes pour chacun des modèles PLS doit être déterminé au début. Pour cela, les auteurs proposent d'utiliser ce qu'ils appellent la « validation croisée totale ». Il faut générer toutes les combinaisons possibles de composantes pour choisir le meilleur modèle SPLS.

Les auteurs expliquent la méthode pour $K = 2$ comme ceci : cela consiste à expliquer par l'un des blocs explicatifs la partie des réponses Y qui n'est pas expliquée par l'autre bloc grâce à deux régressions PLS. Ils stipulent aussi que l'algorithme est « facilement » généralisable au cas de K blocs explicatifs. Mais nous entrevoyons deux solutions pour $K = 3$:

→ faire successivement la PLS du bloc X_k avec la partie de Y qui n'a pas été expliquée par le bloc précédent. Dans ce cas les différents blocs pourraient expliquer la même chose.

→ Ou bien utiliser l'algorithme suivant :

étape 1 : initialisation

$$F_3 = Y$$

fixer le nombre de composantes dans chacune des PLS qui vont suivre.

étape 2 : première PLS

effectuer la régression PLS de F_3 sur X_1 .

$$\text{calculer } F_1 = Y - P_{T_1} F_3$$

étape 3 : seconde PLS

effectuer la régression PLS de F_1 sur X_2 .

$$\text{calculer } F_2 = Y - P_{T_2} F_1 - P_{T_1} F_3$$

étape 4 : troisième PLS

effectuer la régression PLS de F_2 sur X_3

$$\text{calculer } F_3 = Y - P_{T_3} F_2$$

étape 5 : test de la convergence

tester la convergence de F_3 .

S'il n'y a pas convergence, retourner à l'étape 2.

Cet algorithme nous semble correspondre avec l'idée générale de la méthode.

Dans ce type de méthode, c'est le premier bloc de prédicteurs qui expliquera le plus les réponses. Le choix de l'ordre dans lequel on va effectuer les PLS est donc très important. Cette méthode est donc plus appropriée dans les cas où l'ordre des blocs est évident, ou bien lorsqu'on est en présence de peu de blocs. Les auteurs disent eux-mêmes que la méthode doit être très difficile à appliquer lorsqu'on a au moins trois blocs, ne serait ce pour choisir l'ordre, et qu'il faut se limiter à deux blocs explicatifs.

Le choix du nombre de composantes pour chacun des modèles est très compliqué si l'on est en présence de plus de deux blocs, surtout s'il y a beaucoup de variables en jeu.

Les auteurs proposent d'utiliser la SPLS pour étudier la partie linéaire et la partie non linéaire d'un ensemble de données et comparent leur méthode à la régression sur variables latentes avec non-linéarité implicite (INLR) (voir chapitre II-1.5). De plus, ils proposent de l'utiliser pour estimer si des blocs de variables expliquent significativement ou non un groupe de réponses de la manière suivante : commencer par effectuer la régression PLS de Y sur X_1 . Pour savoir si X_2 apporte quelque chose, faire la SPLS avec X_1 et X_2 . Si le modèle obtenu est n'est pas significativement plus explicatif que celui la régression PLS, X_2 n'apporte rien. Cela suppose que l'on privilégie X_1 .

5.5 Analyse de Co-Inertie Multiple Orthogonale PLS : ACIMO-PLS

L'ACIMO-PLS est une extension de l'ACIMO. Elle diffère de l'ACIMO à partir de la deuxième étape, c'est-à-dire à partir des objets d'ordre 2. Grâce à elle, on peut modéliser un ensemble de tableaux Y_1, \dots, Y_N , (c'est-à-dire un multi-tableau d'ordre 3) par un seul tableau X .

Attention : Nous avons inversé les notations par rapport à l'ACIMO qui étudiait K tableaux X_k et un tableau Y . La raison est que lorsqu'on compare un ensemble de tableau à un tableau de référence, on a l'habitude de le noter Y . Lorsqu'on se place dans le cadre de la régression, l'usage est de désigner par « Y » les réponses. D'où les notations choisies ici.

Nous allons chercher des combinaisons linéaires des variables de chacun des tableaux Y_n les plus proches de combinaisons linéaires des variables de X , ces dernières expliquant au mieux X et les Y_n . Il se trouve que l'optimisation par étape du critère proposé, C_1 , avec les déflations appropriées, revient à la résolution de la régression PLS traditionnelle, (WOLD, 1966a; TENENHAUS, GAUCHI & MENARDO, 1995; TENENHAUS, 1998) de $Y = [Y_1 \ \dots \ Y_N]$ sur X , mais fournit des objets supplémentaires. Pour cette raison, nous avons appelé cette technique ACIMO-PLS.

Définition 5.1 *L'ACIMO-PLS d'ordre a est l'ACIMO d'ordre 1 de $\{(Y_n^{(a-1)}, Q_n^y, D)\}_k$ et $(X^{(a-1)}, Q^x, D)$, où $X^{(a-1)} = P_{t_{a-1}}^\perp X^{(a-2)}$ et $Y_n^{(a-1)} = P_{t_{a-1}}^\perp Y_n^{(a-2)}$, pour $2 \leq a \leq A$ et $Y_n^{(0)} = Y_n$ et $X^{(0)} = X$.*

Ainsi, le système de composantes $\{t_a = X^{(a-1)} Q^x a_a\}_a$ est D -orthogonal dans \mathbb{R}^l . De plus, nous sommes en présence de propriétés analogues à celles de la régression PLS.

Propriété 5.1

- $(a_a | a_{a'})_D = \delta_{aa'}$, $a = 1, \dots, A$.
- $(u_{n,a} | t_l)_D = 0$, $l = 1, \dots, a-1$, avec $u_{n,a} = Y_n^{(a-1)} Q_n^y b_{n,a}$.
- $(b_{n,l} | b_{n,a})_{Q_n^y} = (u_{n,l} | t_a)_D / \text{cov}^2(t_a | u_{n,a})$ $l = 1, \dots, a-1$.

Démonstration :

- La démonstration du premier point est identique à la démonstration de la propriété 2.6 de page 37.
- Calculons $(u_{n,a} | t_l)_D$.
 $(u_{n,a} | t_l)_D = t_l' D Y_n^{(a-1)} Q_n^y b_{n,a} = t_l' D P_{t_1, \dots, t_{a-1}}^\perp Y_n Q_n^y b_{n,a} = 0 \quad \forall l = 1, \dots, a-1$.
- On a $(b_{n,l} | b_{n,a})_{Q_n^y} = b_{n,l}' Q_n^y b_{n,a} = \frac{1}{\sqrt{\lambda_{n,a}}} b_{n,l}' Q_n^y Y_n^{(a-1)'} D X^{(a-1)} Q^x a_a$.

Or, $Y_n^{(a-1)'} = Y_n^{(l-1)'} P_{\{t_l, \dots, t_{a-1}\}}^{\perp}$. D'où

$$\begin{aligned} (b_{n,l} | b_{n,a})_{Q_n^y} &= \frac{1}{\sqrt{\lambda_{n,a}}} b'_{n,l} Q_n^y Y_n^{(l-1)'} P_{\{t_l, \dots, t_{a-1}\}}^{\perp} D t_a \\ &= \frac{1}{\sqrt{\lambda_{n,a}}} u'_{n,l} D P_{\{t_l, \dots, t_{a-1}\}}^{\perp} D t_a \\ &= \frac{1}{\sqrt{\lambda_{n,a}}} u'_{n,l} D t_a \text{ car } t_a \text{ est } D\text{-orthogonal aux } t \text{ précédents.} \end{aligned}$$

—■—

5.5.1 Lien avec la régression PLS

Nous allons montrer dans ce paragraphe le lien entre l'ACIMO-PLS et la régression PLS de $(Y, Q_Y = \text{Diag}(Q_n^y), D)$ sur (X, Q^x, D) .

Quelques notations supplémentaires sont nécessaires.

- b_a , de dimension $\sum_{n=1}^N M_n \times 1$, le $a^{\text{ième}}$ axe de Y issu de la régression PLS. b_a est Q_Y -normé.
- $u_a = Y^{(a-1)'} Q_Y b_a$ la $a^{\text{ième}}$ composante PLS de Y .
- $b_a[n]$ le $n^{\text{ième}}$ bloc de longueur M_n de b_a .

Propriété 5.2 *Les axes de co-inertie a_a (resp. les composantes t_a) de l'ACIMO-PLS associés à X sont identiques aux axes d'ordre a (resp. aux composantes d'ordre a) de X issus de la régression PLS de $(Y, Q_Y = \text{Diag}(Q_n^y), D)$ sur (X, Q^x, D) .*

Démonstration :

En régression PLS, l'axe associé au tableau explicatif est le premier vecteur propre normé à l'unité de $X' D Y Q_Y' D X Q^x$. a_1 vérifie aussi cette propriété, donc les axes et composantes d'ordre 1 associées à X sont identiques en ACIMO-PLS et PLS. Les composantes d'ordre a sont identiques car seules les composantes de Y sont utilisées[¶].

—■—

Le critère maximisé (voir annexe A) dans la PLS d'ordre 1 est $\text{cov}(Y Q_Y b, X Q^x a)$ sous les contraintes de norme unité pour les axes a et b $\|a\|_{Q^x}^2 = 1$ et $\|b\|_{Q_Y}^2 = 1$. Les solutions maximisent donc aussi $\text{cov}^2(X Q^x a, Y Q_Y b)$ sous les mêmes contraintes et μ_1 est égal $\text{cov}^2(X Q^x a_1, Y Q_Y b_1)$.

Propriété 5.3 $b_a[n] / \|b_a[n]\|_{Q_n^y} = b_{n,a}$, pour tout b et tout a .

[¶] Rappel : $Y^{(a-1)} = Y - P_{T_{a-1}} Y = [Y_1 - P_{T_{a-1}} Y_1 \dots Y_N - P_{T_{a-1}} Y_N]$ et $X^{(a-1)} = X - P_{T_{a-1}} X$ dans la PLS de Y sur X . T_a désigne ici $[t_1 \dots t_a]$.

Démonstration :

Dans les formules qui suivent, écrire Y (*resp.* Y_n), est équivalent à écrire $Y^{(a-1)}$ (*resp.* $Y_n^{(a-1)}$), d'après les définitions de $Y^{(a-1)}$, $Y_n^{(a-1)}$ et $X^{(a-1)}$.

Comme $b_a = \frac{1}{\sqrt{\mu_a}} Y' D X^{(a-1)} Q^x a_a$, nous en déduisons que $b_a[n] / \|b_a[n]\|_{Q_n^y} = Y_n' D X^{(a-1)} Q^x a_a / \|Y_n' D X^{(a-1)} Q^x a_a\|_{Q_n^y}$. Or, nous avons montré dans le chapitre 2.4 que $\sqrt{\lambda_{n,a}} = \|Y_n' D X^{(a-1)} Q^x a_a\|_{Q_n^y}$. D'où le résultat annoncé.

Rappels : les notations de l'ACIMO sont inversées par rapport à celles utilisées ici, entre « Y » et « X », « b » et « a », k et n , « μ » et « λ ». —■—

Cette propriété implique une relation entre u_a et les $u_{n,a}$. En effet, on a $u_a = Y^{(a-1)} Q_Y b_a = \sum_{n=1}^N Y_n^{(a-1)} Q_n^y b_a[n] = \sum_{n=1}^N \|b_a[n]\|_{Q_n^y} u_{n,a}$, et $\|b_a[n]\|_{Q_n^y} = \sqrt{\lambda_{n,a}} / \sqrt{\mu_a}$.

Remarque : LAFOSSE & HANAFI (1997) montrent que

$$\max_{\|a\|_{Q^x}^2 = \|b\|_{Q^y}^2 = 1} \sum_{n=1}^N \text{cov}^2(X Q^x a, Y_n Q_n^y b_n) = \max_{\|a\|_{Q^x}^2 = \|b\|_{Q^y}^2 = 1} \text{cov}^2(X Q^x a, Y Q_Y b),$$

et une relation entre les solutions qui en découlent : les a sont identiques dans les deux problèmes et $b_n = b[n] / \|b[n]\|_{Q_n^y}$.

5.5.2 Modélisation des tableaux $\{Y_n\}_n$

La modélisation des tableaux Y_n par X est immédiate en appliquant les résultats de la régression PLS, car $\hat{Y}^A = [\hat{Y}_1^A \dots \hat{Y}_N^A]$.

Nous donnons les propriétés qui suivent à titre de rappel. Les démonstrations sont présentées dans (DURAND, ROMAN & VIVIEN, 1998) et sont analogues à celle de la proposition 5.3 de page 99.

Proposition 5.1 *Les composantes $t_a = X^{(a-1)} Q^x a_a$ peuvent s'écrire sous la forme $t_a = X \gamma_a$ pour tout $a = 1, \dots, A$ et γ_a est donné par :*

$$\begin{aligned} \gamma_1 &= Q^x a_1 \\ \gamma_a &= Q^x a_a - \sum_{l=1}^{a-1} \frac{\gamma_l \gamma_l'}{\|t_l\|_D^2} X' D X Q^x a_a \quad a \geq 2. \end{aligned}$$

Proposition 5.2 *Pour chacun des tableaux Y_n , nous avons le modèle linéaire suivant :*

$$\begin{aligned} Y_n &= \sum_{a=1}^A P_{t_a} Y_n + Y_n^{(A)} \doteq \hat{Y}_n^A + Y_n^{(A)} \\ \hat{Y}_n^A &= X \hat{\beta}_n^A \end{aligned}$$

avec $\hat{\beta}_n^A = \sum_{a=1}^A \frac{\gamma_a \gamma'_a}{\|t_a\|_D^2} X' D Y_n$.

5.5.3 Commentaires

Le rang A des modèles ne peut dépasser le rang de X .

L'ACIMO-PLS met en évidence l'optimalité de la régression PLS de $Y = [Y_1 \ \dots \ Y_N]$ sur X .

L'ACIMO-PLS permet de donner des composantes par tableaux. On peut alors représenter sur un même graphique les $l \times (N + 1)$ points dont les coordonnées sont données par $(t_a, u_a), \{(t_a, u_{n,a})\}_n$, ce qui donne une vision globale et partielle (par tableau) de la qualité des régressions partielles de Y et Y_n sur t_a à l'étape a . Si le nuage de points est proche d'une droite, la relation entre les variables de Y et de X est bien linéaire et l'utilisation de la régression linéaire est justifiée. On peut visualiser cette liaison par tableaux Y_n . Si la régression linéaire ne semble pas justifiée, il faut peut-être, à ce moment là, envisager d'utiliser, une extension non linéaire de la PLS, par exemple, la régression PLS Splines ou la régression PLS avec splines additives (DURAND, 2001; DURAND & SABATIER, 1997, voir les chapitres II-1.7 et II-1.6).

Les composantes $\{u_{n,a}\}_a$ ne forment pas un système D -orthogonal de \mathbb{R}^l . Les axes de $\{b_{n,a}\}_a$ ne forment pas un système Q_n^y -orthogonal de \mathbb{R}^{M_n} .

L'orthogonalité des composantes t_a entre elles assurent la décomposition de la variance de chaque tableau. L'orthogonalité des axes a_a permet de définir des repères orthonormés de \mathbb{R}^l . Les composantes étant orthogonales, $t_a/\|t_a\|_D$ et $t_\ell/\|t_\ell\|_D$ définissent un repère orthonormé de \mathbb{R}^l dans lequel on peut projeter les variables explicatives (X) et les variables réponses (Y_n). Si les variables sont réduites, alors, les coordonnées des projections des variables dans ce repère sont données par les corrélations entre les variables et les composantes. Si les variables sont seulement centrées, les coordonnées des points projetés sont données par les covariances entre les variables et les composantes normées à l'unité. De plus, si $\sum_{n=1}^N u_{n,a}$ est très proche de t_a alors la configuration de la carte des variables donnée par la représentation des points de coordonnées (a_a, a_ℓ) , devrait être proche de la configuration des points projetés sur $(t_a/\|t_a\|_D, t_\ell/\|t_\ell\|_D)$. La configuration des points dont les coordonnées sont données par $(b_{n,a}, b_{n,\ell})$ est similaire à la carte des variables réponses projetées dans $(t_a/\|t_a\|_D, t_\ell/\|t_\ell\|_D)$ à un coefficient multiplicatif $(\sqrt{\lambda_{n,a}}/\|t_a\|_D)$ près sur chaque axe.

On peut projeter les l individus de \mathbb{R}^J dans le plan défini par (a_a, a_ℓ) . Mais les coordonnées de ces projections ne sont pas exactement données par les coordonnées des composantes t_a et t_ℓ , excepté pour $a = 1$ (voir annexe C). Comme en régression PLS usuelle on peut cependant utiliser les coordonnées des composantes pour obtenir une carte des individus. Cela montre quels sont les individus qui «construisent» chacune des composantes à chaque étape.

Comme dans l'ACIMO, il est intéressant de comparer la variance d'une composante de co-inertie $u_{n,a}$ (*resp.* t_a) avec la variance de la composante d'inertie de même ordre issue de l'ACP du triplet (Y_n, Q_n^y, D) (*resp.* du triplet (X, Q^x, D)).

Enfin, il semble naturel de comparer la variance des composantes $u_{n,a}$ avec la variance des composantes associées à Y_n de même ordre issues de la régression PLS de Y_n sur X , ainsi qu'à la variance des composantes de même ordre de l'ACPVI de (Y_n, Q_n^y, D) par rapport à X (RAO, 1964). Ces deux dernières analyses visant à expliquer indépendamment chaque tableau Y_n par X , les composantes associées ont toutes les raisons d'être plus explicatives de chaque tableau que les composantes de l'ACIMO-PLS, si bien sûr, X explique un tant soit peu le tableau Y_n . Les composantes de l'ACIMO-PLS expliquent chacun des tableaux en tenant compte des autres.

L'ACIMO-PLS a fait l'objet d'un article (VIVIEN & SABATIER, 2001c), dans lequel un exemple d'application en Analyse Sensorielle est traité. Un second exemple est proposé au chapitre 2 de la troisième partie.

5.6 Analyse de Co-Inertie Multiple Orthogonale Généralisée PLS : ACIMOG-PLS

Dans cette partie, nous présentons l'Analyse de Co-inertie Multiple Orthogonale Généralisée - PLS. Cette partie est plus détaillée que les autres car c'est une méthode originale à laquelle nous avons porté une grande attention. Le but de l'ACIMOG-PLS est d'expliquer l'ensemble des tableaux Y_n par l'ensemble des tableaux X_k , et éventuellement faire de la prédiction. Un certain nombre de méthodes, que nous avons décrites dans ce chapitre n'optimisent pas de critère, mais sont simplement basées sur des modifications de l'algorithme NIPALS de la régression PLS. Comme ces méthodes, l'ACIMOG-PLS est une méthode «component-wise».

5.6.1 Définitions et solutions

Le critère que nous utilisons à chaque étape est le critère $C_2(\{a_k\}, \{b_n\}) = \sum_{k=1}^K \sum_{n=1}^N cov(t_k, u_n)$ utilisé par l'ACIMOG (chapitre 3.4).

Les solutions d'ordre 1 sont les mêmes que celles de l'ACIMOG. La différence des modèles obtenus est due à la déflation utilisée lors du passage d'une étape à l'autre. Rappelons l'expression de ces solutions.

Propriété 5.4 *Les solutions $a_{k,1}$ et $b_{n,1}$ d'ordre 1 de l'ACIMOG(-PLS) maximisent le critère C_2 sous contrainte de norme unité sur a_k et b_n et sont données par :*

$$\begin{aligned} a_{k,1} &= X_k' D \sum_{n=1}^N Y_n Q_n^y b_{n,1} / \mu_{k,1} \\ b_{n,1} &= Y_n' D \sum_{k=1}^K X_k Q_k^x a_{k,1} / \lambda_{n,1}, \end{aligned}$$

où $\mu_{k,1} = cov(t_{k,1}, \sum_{n=1}^N u_{n,1}) = \|X_k' D \sum_{n=1}^N Y_n Q_n^y\|_{Q_k^x}$ et $\lambda_{n,1} = cov(u_{n,1}, \sum_{k=1}^K t_{k,1}) = \|Y_n' D \sum_{k=1}^K X_k Q_k^x a_{k,1}\|_{Q_n^y}$.

Rappelons que $\sum_{n=1}^N \lambda_{n,1} = \sum_{k=1}^K \mu_{k,1}$ est égal au maximum du critère. De plus, les solutions sont obtenues grâce à un algorithme itératif croissant et convergent décrit au chapitre 3.4.

Soit $X_k^{(0)} = X_k, k = 1, \dots, K$ et $Y_n^{(0)} = Y_n, n = 1, \dots, N$.

Le passage à l'étape suivante peut se faire de plusieurs façons, ce qui définira les ACIMOG-PLS 1, 2 et 3.

Définition 5.2 L'ACIMOG-PLS1 d'ordre a est l'ACIMOG d'ordre 1 des triplets

$$\{(X_k^{(a-1)}, Q_k^x, D)\}_{k=1, \dots, K} \text{ et } \{(Y_n^{(a-1)}, Q_n^y, D)\}_{n=1, \dots, N}, \text{ où } X_k^{(a-1)} = P_{T_{a-1}}^\perp X_k^{(a-2)}, \text{ pour } k = 1, \dots, K, \text{ et } Y_n^{(a-1)} = P_{T_{a-1}}^\perp Y_n^{(a-2)}, \text{ pour } n = 1, \dots, N, \text{ avec } T_{a-1} = [t_{1,a-1} \ \dots \ t_{K,a-1}].$$

Définition 5.3 L'ACIMOG-PLS2 d'ordre a est l'ACIMOG d'ordre 1 des triplets

$$\{(X_k^{(a-1)}, Q_k^x, D)\}_{k=1, \dots, K} \text{ et } \{(Y_n^{(a-1)}, Q_n^y, D)\}_{n=1, \dots, N}, \text{ où } X_k^{(a-1)} = P_{t_{a-1}}^\perp X_k^{(a-2)}, \text{ pour } k = 1, \dots, K, \text{ et } Y_n^{(a-1)} = P_{t_{a-1}}^\perp Y_n^{(a-2)}, \text{ pour } n = 1, \dots, N, \text{ avec } t_{a-1} = \sum_{k=1}^K t_{k,a-1}.$$

Définition 5.4 L'ACIMOG-PLS3 d'ordre a est l'ACIMOG d'ordre 1 des triplets

$$\{(X_k^{(a-1)}, Q_k^x, D)\}_{k=1, \dots, K} \text{ et } \{(Y_n^{(a-1)}, Q_n^y, D)\}_{n=1, \dots, N}, \text{ où } X_k^{(a-1)} = P_{t_{a-1}}^\perp X_k^{(a-2)}, \text{ pour } k = 1, \dots, K, \text{ et } Y_n^{(a-1)} = P_{t_{a-1}}^\perp Y_n^{(a-2)}, \text{ pour } n = 1, \dots, N, \text{ avec } t_{a-1} = \sum_{k=1}^K \mu_{k,a-1} t_{k,a-1}.$$

Nous avons les propriétés suivantes :

Propriété 5.5 Les systèmes de composantes $\{t_a\}_a$ sont D -orthogonaux.

Propriété 5.6 Dans l'ACIMOG-PLS1 les espaces $\{T_a\}_a$ sont D -orthogonaux ainsi que les K systèmes de composantes $\{t_{k,a}\}_a$.

Tout d'abord posons :

$$\begin{aligned} \mathcal{E}_a &= t_a \text{ dans le cas de l'ACIMOG-PLS2 et ACIMOG-PLS3,} \\ &= T_a \text{ dans le cas de l'ACIMOG-PLS1.} \end{aligned}$$

Démonstration :

Nous allons montrer ces deux propriétés avec une seule démonstration.

Dans un premier temps, il s'agit de montrer la D -orthogonalité des espaces \mathcal{E}_a .

Remarquons que $\mathcal{E}_a = [X_1^{(a-1)} \ \dots \ X_K^{(a-1)}] \text{Diag}(Q_k^x) a_a$, où :

$$\begin{aligned} a'_a &= [a'_{1,a} \ \dots \ a'_{K,a}] \text{ avec l'ACIMOG-PLS2} \\ &= [\mu_{1,a} a'_{1,a} \ \dots \ \mu_{K,a} a'_{K,a}] \text{ avec l'ACIMOG-PLS3} \\ &= \text{Diag}(a'_{k,a}). \end{aligned}$$

La preuve se fait par récurrence :

$$\mathcal{E}'_1 D \mathcal{E}_2 = \mathcal{E}'_1 D P_{\mathcal{E}_1}^\perp X \text{Diag}(Q_k^x) a_2 = 0.$$

On suppose maintenant que les espaces $\mathcal{E}_1, \dots, \mathcal{E}_{a-1}$ sont D -orthogonaux. Soit $l < a$. Montrons que \mathcal{E}_a est D -orthogonal à \mathcal{E}_l quelque soit $l < a$.

Si $l = a - 1$, $\mathcal{E}'_{a-1} D\mathcal{E}_a = \mathcal{E}'_{a-1} DP_{\mathcal{E}_{a-1}}^\perp X^{(a-2)} \text{Diag}(Q_k^x) a_a = 0$.
Ainsi, pour $l < a - 1$

$$\begin{aligned} \mathcal{E}'_l D\mathcal{E}_a &= \mathcal{E}'_l DP_{\mathcal{E}_{a-1}}^\perp \dots P_{\mathcal{E}_l}^\perp X^{(l-1)} \text{Diag}(Q_k^x) a_a \\ &= \mathcal{E}'_l DP_{\mathcal{E}_{a-2}}^\perp \dots P_{\mathcal{E}_l}^\perp X^{(l-1)} \text{Diag}(Q_k^x) a_a \text{ d'après l'hypothèse} \\ &= \mathcal{E}'_l DP_{\mathcal{E}_l}^\perp X^{(l-1)} \text{Diag}(Q_k^x) a_a \text{ d'après l'hypothèse} \\ &= 0. \end{aligned}$$

Les espaces $\{\mathcal{E}_a\}_a$ sont D -orthogonaux.

La deuxième partie de la propriété 5.6, l'orthogonalité de composantes $\{t_{k,a}\}_a$ avec l'ACIMOG-PLS1, est immédiate car $T'_l DT_a = 0$ pour tout $l \leq a$. —■—

Propriété 5.7 $u'_{n,a} Dt_l$, pour tout $a > 1$ et $l = 1, \dots, a - 1$, avec l'ACIMOG-PLS 2 et 3.

Démonstration :

La propriété 5.5 implique :

$$\begin{aligned} t'_l Du_{n,a} &= t'_l DY_n^{(a-1)} Q_n^y b_{n,a} \\ &= t'_l DP_{t_{a-1}}^\perp \dots P_{t_l}^\perp Y_n^{(l-1)} Q_n^y b_{n,a} \\ &= t'_l DP_{t_l}^\perp Y_n^{(l-1)} Q_n^y b_{n,a} \\ &= 0. \end{aligned}$$
—■—

5.6.2 Les modèles

L'orthogonalité des espaces sur lesquels on projette les variables successivement va permettre d'écrire des modèles dont la variance peut se décomposer additivement comme nous le verrons un peu plus loin.

Nous allons maintenant déterminer l'expression des modèles \widehat{Y}_n^a d'ordre a pour le tableau Y_n . La procédure est la même quelque soit la définition.

On a bien sûr pour tout n ,

$$Y_n^{(A)} = Y_n^{(0)} - \sum_{a=1}^A P_{\mathcal{E}_a} Y_n^{(a-1)},$$

ainsi qu'une expression analogue pour $X_k^{(A)}$.

Les espaces engendrés par \mathcal{E}_a et \mathcal{E}_i pour $i < a$ étant orthogonaux (par construction) et comme $Y^{(a-1)} = P_{\mathcal{E}_{a-1}}^\perp Y_n^{(a-2)}$ quelque soit $a > 1$, on a :

$$Y_n^{(A)} = Y_n^{(0)} - \sum_{a=1}^A P_{\mathcal{E}_a} Y_n^{(0)}.$$

Ainsi, la reconstitution d'ordre A du tableau Y_n s'écrit :

$$Y_n = \widehat{Y}_n^A + E_{n,A}^y,$$

où $\widehat{Y}_n^A = \sum_{a=1}^A P_{\mathcal{E}_a} Y_n^{(0)}$ et $E_{n,A}^y = Y_n^{(A)}$ est l'erreur du modèle non corrélée avec \widehat{Y}_n^A .

Chaque modèle partiel $P_{\mathcal{E}_a} Y_n^{(0)}$ explique une part différente des précédentes de l'information contenue dans Y_n .

Ainsi on peut retenir des composantes étape après étape, de la même manière qu'en régression PLS. Pour cela, on peut s'intéresser à la variance expliquée par le modèle étape après étape :

$$\begin{aligned} \text{var}(\widehat{Y}_n^A) \parallel &= \|\widehat{Y}_n^{(A)}\|_D^2 \\ &= \sum_a \sum_{a'} Y_n^{(0)'} D P_{\mathcal{E}_a} P_{\mathcal{E}_{a'}} Y_n^{(0)} \end{aligned}$$

Comme pour tout $a \neq a'$ les espaces engendrés par $\mathcal{E}_{a'}$ et \mathcal{E}_a sont D -orthogonaux, nous obtenons :

$$\begin{aligned} \text{var}(\widehat{Y}_n^A) &= \sum_a Y_n^{(0)'} D P_{\mathcal{E}_a} P_{\mathcal{E}_a} Y_n^{(0)} \\ &= \sum_a \|P_{\mathcal{E}_a} Y_n^{(0)}\|_D^2. \end{aligned}$$

Cette formule nous dit que l'on a une décomposition de la variance totale de $Y_n^{(0)}$.

Cela fournit une aide pour le choix du nombre A de composantes à retenir.

En procédant de la même manière, la variance expliquée par la reconstitution d'ordre A du tableau X_k s'exprime $\text{var}(\widehat{X}_k^A) = \sum_a \|P_{\mathcal{E}_a} X_k^{(0)}\|_D^2$.

Remarque : Dans l'ACIMOG-PLS2 $\|P_{t_a} Y_n\|_D^2 = \lambda_{n,a}^2 / \|t_a\|_D^2$. En effet, grâce à la propriété

5.5, $\lambda_{n,a} = \|Y_n^{(a-1)'} D t_a\|_{Q^y} = \|Y_n' D t_a\|_{Q^y}$ Or, $\|P_{t_a} Y_n\|_D^2 = \|Y_n' D t_a\|_{Q^y}^2 / \|t_a\|_D^2$. D'où le résultat. Cela signifie, que lorsque les tableaux sont D -centrés, la variance de Y_n expliquée par l'étape a est proportionnelle à $\lambda_{n,a}^2$. A chaque étape, les coefficients $\lambda_{n,a}$ montrent quels sont les tableaux Y_n les mieux et les moins expliqués par l'étape.

5.6.3 Prédiction

Nous allons montrer que l'ACIMOG-PLS permet aussi de prédire \widehat{Y}_n^A .

Commençons par démontrer la propriété qui suit :

Proposition 5.3 *La modélisation \widehat{Y}_n^A d'ordre A du tableau réponse Y_n , $n = 1, \dots, N$, peut s'écrire de la forme $\widehat{Y}_n^A = X \widehat{\beta}_n^A$ où :*

|| Rappelons que la variance est ici égale à la norme car les tableaux sont D -centrés.

$$\begin{aligned} X &= [X_1 \ \dots \ X_K] \\ \widehat{\beta}_n^A &= \sum_{a=1}^A \delta_a (\mathcal{E}'_a D \mathcal{E}_a)^{-1} \mathcal{E}'_a D Y_n^{(0)} \end{aligned}$$

$$\text{où } \begin{cases} \delta_1 = \alpha_1 \\ \delta_a = \alpha_a - \sum_{l=1}^{a-1} \delta_l (\mathcal{E}'_l D \mathcal{E}_l)^{-1} \mathcal{E}'_l D X \alpha_a, \quad \forall a > 1 \end{cases}$$

$$\text{avec } \begin{cases} \alpha_a = \text{Diag}(Q_k^x) \text{Diag}(a_{k,a}) & \text{pour l'ACIMOG - PLS1} \\ = \text{Diag}(Q_k^x) \begin{bmatrix} a_{1,a} \\ \vdots \\ a_{K,a} \end{bmatrix} & \text{pour l'ACIMOG - PLS2,} \\ = \text{Diag}(Q_k^x) \begin{bmatrix} \mu_{1,a} a_{1,a} \\ \vdots \\ \mu_{K,a} a_{K,a} \end{bmatrix} & \text{pour l'ACIMOG - PLS3.} \end{cases}$$

Démonstration :

Nous avons $\mathcal{E}_a = X^{(a-1)} \alpha_a$ $a = 1, \dots, A$.

En effet, dans le cas de l'ACIMOG-PLS2 :

$$\begin{aligned} t_a &= \sum_k t_{k,a} \\ &= \sum_k X_k^{(a-1)} Q_k^x a_{k,a} \\ &= X^{(a-1)} \text{Diag}(Q_k^x) \begin{bmatrix} a_{1,a} \\ \vdots \\ a_{K,a} \end{bmatrix}, \end{aligned}$$

dans le cas de l'ACIMOG-PLS1 :

$$\begin{aligned} T_a &= [t_{1,a} \ \dots \ t_{K,a}] \\ &= [X_1^{(a-1)} Q_1^x a_{1,a} \ \dots \ X_K^{(a-1)} Q_K^x a_{K,a}] \\ &= X^{(a-1)} \text{Diag}(Q_k^x) \text{Diag}(a_{k,a}), \end{aligned}$$

et dans le cas de l'ACIMOG-PLS3 :

$$\begin{aligned} t_a &= \sum_k \mu_{k,a} t_{k,a} \\ &= \sum_k \mu_{k,a} X_k^{(a-1)} Q_k^x a_{k,a} \\ &= \sum_k X_k^{(a-1)} Q_k^x \mu_{k,a} a_{k,a} \\ &= X^{(a-1)} \text{Diag}(Q_k^x) \begin{bmatrix} \mu_{1,a} a_{1,a} \\ \vdots \\ \mu_{K,a} a_{K,a} \end{bmatrix}. \end{aligned}$$

Montrons par induction que $\mathcal{E}_a = X\delta_a$.

$a = 1$: évident.

$a = 2$:

$$\begin{aligned} \mathcal{E}_2 &= X^{(1)}\alpha_2 \\ &= P_{\mathcal{E}_1}^\perp X^{(0)}\alpha_2 \\ &= X\alpha_2 - X\mathcal{E}_1(\mathcal{E}_1'D\mathcal{E}_1)^{-1}\mathcal{E}_1'X\alpha_2 \\ &= X\alpha_2 - X\alpha_1(\mathcal{E}_1'D\mathcal{E}_1)^{-1}\mathcal{E}_1'X\alpha_2 \\ &= X\delta_2. \end{aligned}$$

$a > 2$: On suppose que l'expression est vraie jusqu'à $a - 1$.

$$\begin{aligned} \mathcal{E}_a &= X^{(a-1)}\alpha_a \\ &= P_{\mathcal{E}_{a-1}}^\perp X^{(a-2)}\alpha_a \\ &= P^\perp \mathcal{E}_{a-1} \dots P_{\mathcal{E}_1}^\perp X\alpha_a \\ &= (Id - \sum_{l=1}^{a-1} P_{\mathcal{E}_l})X\alpha_a, \quad \text{car les espaces } \mathcal{E}_l \text{ sont orthogonaux} \\ &= X\alpha_a - \sum_{l=1}^{a-1} \mathcal{E}_l(\mathcal{E}_l'D\mathcal{E}_l)^{-1}\mathcal{E}_l'X\alpha_a \\ &= X\alpha_a - \sum_{l=1}^{a-1} X\delta_l(\mathcal{E}_l'D\mathcal{E}_l)^{-1}\mathcal{E}_l'X\alpha_a \\ &= X\delta_a. \end{aligned}$$

L'expression du modèle \hat{Y}_n^A en découle aisément :

$$\begin{aligned} \hat{Y}_n^A &= \sum_{a=1}^A P_{\mathcal{E}_a} Y_n \\ &= \sum_{a=1}^A X\delta_a(\mathcal{E}_a'D\mathcal{E}_a)^{-1}\mathcal{E}_a'Y_n \\ &= X\hat{\beta}_n^A. \end{aligned}$$

—■—

Le modèle ainsi exprimé permet de faire de la prédiction. Si l'on dispose d'un échantillon supplémentaire $x_{i'}$, de taille $\sum_{k=1}^K J_k \times 1$, la valeur des variables réponses s'estime, pour tout n , par le calcul

$$\hat{y}_n^A(i') = x_{i'}' \hat{\beta}_n^A, \quad (5.8)$$

où $\hat{\beta}_n^A$ est la matrice des coefficients déterminée antérieurement.

Cette propriété permet de réaliser la validation croisée, comme critère de prédiction interne

(voir le chapitre 6).

Comme dans PLS, les coefficients $\widehat{\beta}_n^a$ peuvent s'écrire sous forme matricielle. Commençons par établir de nouvelles notations et démontrer quelques résultats. Soient :

ACIMOG-PLS 2 et 3	ACIMOG-PLS1
$p_a^k = X_k^{(a-1)'} D \frac{t_a}{\ t_a\ _D^2}, J_k \times 1$	$\mathbf{P}_a^k = X_k^{(a-1)'} DT_a(T_a' DT_a)^+, J_k \times K$
$p_a^{k'} = [p_a^{1'} \dots p_a^{K'}], 1 \times \sum_{k=1}^K J_k$	$\mathbf{P}_a = X^{(a-1)'} DT_a(T_a' DT_a)^+, \sum_{k=1}^K J_k \times K$
$p_a = X^{(a-1)'} D \frac{t_a}{\ t_a\ _D^2}, \sum_{k=1}^K J_k \times 1$	$\mathbf{P}_a' = [\mathbf{P}_a^{1'} \dots \mathbf{P}_a^{K'}], K \times \sum_{k=1}^K J_k$
$\mathcal{P}_a^k = [p_1^k, \dots, p_a^k], J_k \times a$	$\mathcal{P}_a^k = [\mathbf{P}_1^k, \dots, \mathbf{P}_a^k], J_k \times Ka$
$\mathcal{P}_a = [p_1, \dots, p_a], \sum_{k=1}^K J_k \times a$	$\mathcal{P}_a = [\mathbf{P}_1, \dots, \mathbf{P}_a], \sum_{k=1}^K J_k \times Ka$
$\mathbf{T}_a = [t_1, \dots, t_a], l \times a$	$\mathbf{T}_a = [T_1, \dots, T_a], l \times Ka$
$\mathbf{W}_a = [\alpha_1, \dots, \alpha_a], \sum_{k=1}^K J_k \times a$	$\mathbf{W}_a = [\alpha_1, \dots, \alpha_a], \sum_{k=1}^K J_k \times Ka$
$\mathbf{W}_a^* = [\delta_1, \dots, \delta_a], \sum_{k=1}^K J_k \times a$	$\mathbf{W}_a^* = [\delta_1, \dots, \delta_a], \sum_{k=1}^K J_k \times Ka.$

Avec ces notations, les modèles de reconstitution des prédicteurs s'écrivent :

ACIMOG-PLS 2 et 3	ACIMOG-PLS1
$\widehat{X}_k^a = \sum_{l=1}^a t_l p_l^{k'}$	$\widehat{X}_k^a = \sum_{l=1}^a T_l \mathbf{P}_l^{k'}$
$\widehat{X}_k^a = \mathbf{T}_l \mathcal{P}_l^{k'}$	$\widehat{X}_k^a = \mathbf{T}_l \mathcal{P}_l^{k'}$
$\widehat{X}^a = \sum_{l=1}^a t_l p_l'$	$\widehat{X}^a = \sum_{l=1}^a T_l \mathbf{P}_l'$
$\widehat{X}^a = \mathbf{T}_l \mathcal{P}_l'$	$\widehat{X}^a = \mathbf{T}_l \mathcal{P}_l'.$

La proposition ci-après établit des résultats d'orthogonalité et d'autres écritures des objets tels que les composantes et les tableaux résidus à chaque étape. Elle permettra ensuite de donner une forme plus simple pour les modèles, à la manière de la régression PLS.

Proposition 5.4

- a) $t_a = \prod_{l=1}^{a-1} (X - t_l p_l') \alpha_l$ dans l'ACIMOG-PLS 2 et 3.
 $T_a = \prod_{l=1}^{a-1} (X - T_l \mathbf{P}_l') \alpha_l$ dans l'ACIMOG-PLS 1.
- b) $X^{(a)} = X \prod_{l=1}^a (Id_{\sum_{k=1}^K J_k} - \alpha_l p_l')$ dans l'ACIMOG-PLS 2 et 3.
 $X^{(a)} = X \prod_{l=1}^a (Id_{\sum_{k=1}^K J_k} - \alpha_l \mathbf{P}_l')$ dans l'ACIMOG-PLS 1.
 $X_k^{(a)} = X_k - X \left[\alpha_1 p_1^{k'} + \sum_{l=1}^{a-1} \prod_{u=1}^l (Id_{\sum_{k=1}^K J_k} - \alpha_u p_u') \alpha_{l+1} p_{l+1}^{k'} \right]$, dans l'ACIMOG-PLS 2 et 3.
- c) $\alpha_a' p_a = 1$ dans l'ACIMOG-PLS 2 et 3.
 $\alpha_a' \mathbf{P}_a = Id_K$, dans l'ACIMOG-PLS 1, si $T_a' DT_a$ est inversible.
 $\alpha_{k,a}' Q_k^x \mathbf{P}_a^k = e_k'$, dans l'ACIMOG-PLS1 si $T_a' DT_a$ est inversible,, où e_k est le vecteur composé de 0 partout sauf en sa $k^{\text{ième}}$ coordonnée qui est égale à 1.
- d) $l \geq a$ $X^{(l)} \alpha_a = 0$ dans l'ACIMOG-PLS 1, 2 et 3.
 $l \geq a$ $X_k^{(l)} Q_k^x a_{k,a} = 0$ dans l'ACIMOG-PLS 1.

- e) $l > a$ $\alpha'_a p_l = 0$ dans l'ACIMOG-PLS 2 et 3.
 $l > a$ $\alpha_a \mathbf{P}_l = 0$ dans l'ACIMOG-PLS 1.
 $l > a$ $a'_{k,a} Q_k^x \mathbf{P}_l^k = 0$ dans l'ACIMOG-PLS 1.
- f) $l > a$ $\alpha'_a \alpha_l = 0$ dans l'ACIMOG-PLS 3 avec $Q_k^x = Id_{J_k}$.
 $l > a$ $\alpha'_a \text{Diag}(Q_k^{x^{-1}}) \alpha_l = 0$ dans l'ACIMOG-PLS 1 et 3.
 $l > a$ $a'_{k,a} Q_k^x a_{k,l} = 0$ dans l'ACIMOG-PLS 1.
- g) $l \geq a$ $t'_a DX^{(l)} = 0$ dans l'ACIMOG-PLS 2 et 3.
 $l \geq a$ $T'_a DX^{(l)} = 0$ dans l'ACIMOG-PLS 1.
 $l \geq a$ $t'_{k,a} DX_k^{(l)} = 0$ dans l'ACIMOG-PLS 1.
- h) $\forall a$ $\delta_a = \prod_{l=1}^{a-1} (Id_{\sum_{k=1}^K J_k} - \alpha_l p'_l) \alpha_a$ dans l'ACIMOG-PLS 2 et 3.
 $\forall a$ $\delta_a = \prod_{l=1}^{a-1} (Id_{\sum_{k=1}^K J_k} - \alpha_l \mathbf{P}'_l) \alpha_a$ dans l'ACIMOG-PLS 1.

Démonstration :

a) ACIMOG-PLS 2 et 3 :

$$\begin{aligned} t_a &= X^{(a-1)} \alpha_a = P_{t_{a-1}}^\perp X^{(a-2)} \alpha_a \\ &= \prod_{l=1}^{a-1} (Id_l - P_{t_l}) X \alpha_a \\ &= \prod_{l=1}^{a-1} (X - t_l p'_l) \alpha_a \text{ car } p_l = X^{(l-1)'} D \frac{t_l}{\|t_l\|_D^2} = X' D \frac{t_l}{\|t_l\|_D^2}. \end{aligned}$$

ACIMOG-PLS 1 : mêmes calculs en remplaçant p_l par \mathbf{P}_l et t_l par T_l .

b) Ici encore, nous montrons la propriété pour l'ACIMOG-PLS 2 et 3, les calculs sont identiques pour l'ACIMOG-PLS 1 en effectuant les remplacements cités ci-dessus.

Pour $a = 1$, la propriété est vérifiée. En effet,

$$X^{(1)} = X - X \alpha_1 p'_1 = X (Id_{\sum_{k=1}^K J_k} - \alpha_1 p'_1).$$

On suppose que la propriété est vraie pour l'étape $a - 1$, et montrons qu'elle est vérifiée pour l'étape a :

$$\begin{aligned} X^{(a)} &= X^{(a-1)} - P_{t_a} X^{(a-1)} = X^{(a-1)} - t_a p'_a \\ X^{(a)} &= X^{(a-1)} - X^{(a-1)} \alpha_a p'_a \\ X^{(a)} &= X \prod_{l=1}^{a-1} (Id_{\sum_{k=1}^K J_k} - \alpha_l p'_l) (I - \alpha_a p'_a) \text{ d'après l'hypothèse de récurrence,} \\ X^{(a)} &= X \prod_{l=1}^a (Id_{\sum_{k=1}^K J_k} - \alpha_l p'_l). \end{aligned}$$

Il reste à montrer la troisième assertion.

Elle est vraie pour $a = 1$ et $a = 2$:

$$\begin{aligned} X_k^{(1)} &= X_k - X\alpha_1 p_2^{k'} \\ X_k^{(2)} &= X_k^{(1)} - t_2 p_2^{k'} = X_k^{(1)} - X^{(1)}\alpha_2 p_2^{k'} \\ &= X_k - X\alpha_1 p_1^{k'} - (X - X\alpha_1 p_1')\alpha_2 p_2^{k'} \\ &= X_k - X \left(\alpha_1 p_1^{k'} + (Id_{\sum_{k=1}^K J_k} - \alpha_1 p_1')\alpha_2 p_2^{k'} \right). \end{aligned}$$

Supposons que la relation soit vraie jusqu'à l'étape $a - 1$, et procédons par induction pour l'établir à l'étape a :

$$\begin{aligned} X_k^{(a)} &= X_k^{(a-1)} - t_a p_a^{k'} = X_k^{(a-1)} - X^{(a-1)}\alpha_a p_a^{k'} \\ &= X_k - X \left[\alpha_1 p_1^{k'} + \sum_{l=1}^{a-2} \prod_{u=1}^l (Id_{\sum_{k=1}^K J_k} - \alpha_u p_u')\alpha_{l+1} p_{l+1}^{k'} \right] - X^{(a-1)}\alpha_a p_a^{k'} \\ \text{or } X^{(a-1)} &= X \prod_{u=1}^{a-1} (Id_{\sum_{k=1}^K J_k} - \alpha_u p_u'), \\ &= X_k - X \left[\alpha_1 p_1^{k'} + \sum_{l=1}^{a-2} \prod_{u=1}^l (Id_{\sum_{k=1}^K J_k} - \alpha_u p_u')\alpha_{l+1} p_{l+1}^{k'} + \prod_{u=1}^{a-1} (Id_{\sum_{k=1}^K J_k} - \alpha_u p_u')\alpha_a p_a^{k'} \right] \\ &= X_k - X \left[\alpha_1 p_1^{k'} + \sum_{l=1}^{a-1} \prod_{u=1}^l (Id_{\sum_{k=1}^K J_k} - \alpha_u p_u')\alpha_{l+1} p_{l+1}^{k'} \right]. \end{aligned}$$

c) ACIMOG-PLS 1:

$$\begin{aligned} \alpha'_a \mathbf{P}_a &= \text{Diag}(a'_{k,a} Q_k^x) \mathbf{P}_a \\ &= \begin{bmatrix} a'_{1,a} Q_k^x & 0 & \dots & 0 \\ 0 & \ddots & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & \dots & 0 & a'_{K,a} Q_K^x \end{bmatrix} T_a (T_a' D T_a)^+ \\ &= T_a' D T_a (T_a' D T_a)^+ = Id_K \text{ si } T_a' D T_a \text{ est inversible.} \end{aligned}$$

Cela implique :

$$a'_{k,a} Q_k^x \mathbf{P}_a^k = t'_{k,a} D T_a (T_a' D T_a)^{-1} = e'_k.$$

ACIMOG-PLS 2 :

$$\begin{aligned} \alpha'_a p_a &= [a'_{1,a} \dots a'_{K,a}]' \text{Diag}(Q_k^x) X^{(a-1)'} D t_a / \|t_a\|_D^2 \\ &= \sum_{k=1}^K a'_{k,a} Q_k^x X_k^{(a-1)'} D t_a / \|t_a\|_D^2 \\ &= \|t_a\|_D^2 / \|t_a\|_D^2 = 1. \end{aligned}$$

ACIMOG-PLS 3 :

$$\begin{aligned}
 \alpha'_a p_a &= [\mu_{1,a} a'_{1,a} \dots \mu_{K,a}] a'_{K,a} \text{Diag}(Q_k^x) X^{(a-1)'} D t_a / \|t_a\|_D^2 \\
 &= \sum_{k=1}^K \mu_{k,a} a'_{k,a} Q_k^x X_k^{(a-1)'} D t_a / \|t_a\|_D^2 \\
 &= \|t_a\|_D^2 / \|t_a\|_D^2 = 1.
 \end{aligned}$$

d) Soit $l \geq a$.

ACIMOG-PLS 2 et 3 :

$$\begin{aligned}
 X^{(l)} \alpha_a &= P_{t_l}^\perp \dots P_{t_a}^\perp X^{(a-1)} \alpha_a \\
 &= P_{t_l}^\perp \dots P_{t_a}^\perp t_a = 0.
 \end{aligned}$$

ACIMOG-PLS 1 :

$$\begin{aligned}
 X^{(l)} \alpha_a &= P_{T_l}^\perp \dots P_{T_a}^\perp X^{(a-1)} \alpha_a \\
 &= P_{T_l}^\perp \dots P_{T_a}^\perp T_a = 0. \\
 X_k^{(l)} Q_k^x a_{k,a} &= P_{T_l}^\perp \dots P_{T_a}^\perp t_{k,a} = 0, \\
 &\text{car } t_{k,a} \in \text{Im}(T_a).
 \end{aligned}$$

e) Soit $l > a$. ACIMOG-PLS 2 et 3 :

$$\begin{aligned}
 \alpha'_a p_l &= \alpha'_a X^{(l-1)'} D P_{t_a}^\perp \dots P_{t_{l-1}}^\perp t_l / \|t_l\|_D^2 \\
 &= t'_a D t_l / \|t_l\|_D^2 = 0.
 \end{aligned}$$

Pour l'ACIMOG-PLS1, de manière analogue, on obtient $\alpha'_a \mathbf{P}_l$ et :

$$\begin{aligned}
 a'_{k,a} Q_k^x \mathbf{P}_l^k &= t'_{k,a} D T_l (T_l' D T_l)^+ = 0, \\
 &\text{car } t_{k,a} \text{ est orthogonale à } t_{k,l} \forall k \text{ et } a < l.
 \end{aligned}$$

f) Soit $l > a$. ACIMOG-PLS 3 avec $Q_k^x = Id_{J_k}$:

$$\begin{aligned}
 \alpha'_a \alpha_l &= \sum_{k=1}^K \mu_{k,a} \mu_{k,l} a'_{k,a} a_{k,l} \\
 &= \sum_{k=1}^K \mu_{k,a} a'_{k,a} X_k^{(a-1)'} D P_{t_a}^\perp \dots P_{t_l}^\perp \sum_{n=1}^N u_{n,l} \\
 &= t'_a D P_{t_a}^\perp \dots P_{t_l}^\perp \sum_{n=1}^N u_{n,l} = 0.
 \end{aligned}$$

De même, $\alpha'_a \text{Diag}(Q_k^{x^{-1}}) \alpha_l = 0$ car $\alpha'_a \text{Diag}(Q_k^{x^{-1}}) \alpha_l = \sum_{k=1}^K \mu_{k,a} \mu_{k,l} \alpha'_{k,a} Q_k^x Q_k^{x^{-1}} Q_k^x a_{k,l}$.
Remarquons, que dans l'ACIMOG-PLS2 on a $\alpha'_a \text{Diag}(\mu_{k,l} / \sum_{k=1}^K \mu_{k,l}) \alpha_l = 0$ avec $Q_k^x = Id_{J_k}$.

Pour l'ACIMOG-PLS 1:

$$\begin{aligned} a'_{k,a} Q_k^x a_{k,l} &= a'_{k,a} Q_k^x X_k^{(l-1)'} D \sum_{n=1}^N u_{n,l} / \mu_{k,l} \\ &= t'_{k,a} D P_{T_a}^\perp \dots P_{T_l}^\perp \sum_{n=1}^N u_{n,l} / \mu_{k,l} = 0. \end{aligned}$$

Ceci implique $\alpha'_a \text{Diag}(Q_k^{x^{-1}}) \alpha_l = 0$ car :

$$\alpha'_a \text{Diag}(Q_k^{x^{-1}}) \alpha_l = \begin{bmatrix} a'_{1,a} Q_1^x a_{1,l} & 0 & \dots & 0 \\ 0 & a'_{2,a} Q_2^x a_{2,l} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & \dots & 0 & a'_{K,a} Q_K^x a_{K,l} \end{bmatrix}.$$

g) Soit $l \geq a$. Pour les ACIMOG-PLS 2 et 3 :

$$\begin{aligned} t'_a D X^{(l)} &= t'_a D (P_{t_a}^\perp X^{(a-1)} - t_{a+1} p'_{a+1} - \dots - t_l p'_l) \\ &= -t'_a D t_{a+1} p'_{a+1} - \dots - t'_a D t_l p'_l = 0. \end{aligned}$$

Pareillement, on montre que $T'_a D X^{(l)} = 0$, $t'_{k,a} D X_k^{(l)} = 0$ et même $t'_{k,a} D X^{(l)} = 0$ dans l'ACIMOG-PLS1.

h) Montrons le résultat pour les ACIMOG-PLS 2 et 3. La démonstration se transcrit facilement pour l'ACIMOG-PLS1.

Nous avons vu que $X^{(a)} = X \prod_{l=1}^a (Id_{\sum_{k=1}^K J_k} - \alpha_l p'_l)$. D'où,

$$t_a = X^{(a-1)} \alpha_a = X \prod_{l=1}^{a-1} (Id - \alpha_l p'_l) \alpha_a.$$

δ_a est le vecteur tel que $t_a = X \delta_a$. En identifiant, on trouve $\delta_a = \prod_{l=1}^{a-1} (Id_{\sum_{k=1}^K J_k} - \alpha_l p'_l) \alpha_a$. ■

Le point **f**) stipule que l'on peut faire une représentation des individus des tableaux explicatifs dans les plans engendrés par (α_l, α_a) avec l'ACIMOG-PLS 3 et $(a_{k,l}, a_{k,a})$ dans l'ACIMOG-PLS 1.

Proposition 5.5 $\forall a \mathbf{W}_a^* = \mathbf{W}_a (P'_a \mathbf{W}_a)^{-1}$ dans l'ACIMOG-PLS 2 et 3.

Démonstration :

Montrer $\mathbf{W}_a^* = \mathbf{W}_a(\mathcal{P}'_a \mathbf{W}_a)^{-1}$ revient à montrer $\mathbf{W}_a = \mathbf{W}_a^*(\mathcal{P}'_a \mathbf{W}_a)$.

$$\begin{aligned} \mathbf{W}_a^*(\mathcal{P}'_a \mathbf{W}_a) &= [\delta_1 \quad \dots \quad \delta_a] \begin{bmatrix} p'_1 \alpha_1 & \dots & p'_1 \alpha_a \\ \vdots & & \vdots \\ p'_a \alpha_1 & \dots & p'_a \alpha_a \end{bmatrix} \\ &= [\delta_1 \quad \dots \quad \delta_a] \begin{bmatrix} 1 & p'_1 \alpha_2 & \dots & p'_1 \alpha_a \\ 0 & 1 & \dots & \\ 0 & 0 & 1 & \\ 0 & \dots & 0 & 1 \end{bmatrix} \\ &\text{d'après la Proposition 5.4 c) et e).} \\ &= [\delta_1 \quad \delta_2 + \delta_1 p'_1 \alpha_2 \quad \dots \quad \delta_a + \sum_{l=1}^{a-1} \delta_l p'_l \alpha_a] \end{aligned}$$

Ainsi, nous devons montrer que $\alpha_a = \delta_a + \sum_{l=1}^{a-1} \delta_l p'_l \alpha_a$. Nous savons que $\alpha_1 = \delta_1$ par définition.

Pour tout $a > 1$:

$$\begin{aligned} \delta_a &= \prod_{l=1}^{a-1} (Id_{\sum_{k=1}^{\kappa} J_k} - \alpha_l p'_l) \alpha_a \\ &= (Id_{\sum_{k=1}^{\kappa} J_k} - \alpha_1 p'_1) \dots (Id_{\sum_{k=1}^{\kappa} J_k} - \alpha_{a-1} p'_{a-1}) \alpha_a \\ &= \prod_{u=1}^{a-2} (Id_{\sum_{k=1}^{\kappa} J_k} - \alpha_u p'_u) \alpha_a - \prod_{u=1}^{a-2} (Id_{\sum_{k=1}^{\kappa} J_k} - \alpha_u p'_u) \alpha_{a-1} p'_{a-1} \alpha_a \\ &= \prod_{u=1}^{a-3} (Id_{\sum_{k=1}^{\kappa} J_k} - \alpha_u p'_u) \alpha_a - \prod_{u=1}^{a-3} (Id_{\sum_{k=1}^{\kappa} J_k} - \alpha_u p'_u) \alpha_{a-2} p'_{a-2} \alpha_a \\ &\quad - \prod_{u=1}^{a-2} (Id_{\sum_{k=1}^{\kappa} J_k} - \alpha_u p'_u) \alpha_{a-1} p'_{a-1} \alpha_a \end{aligned}$$

En procédant par récurrence, nous obtenons :

$$\begin{aligned} \delta_a &= \prod_{u=1}^1 (Id_{\sum_{k=1}^{\kappa} J_k} - \alpha_u p'_u) \alpha_a - \prod_{u=1}^2 (Id_{\sum_{k=1}^{\kappa} J_k} - \alpha_u p'_u) \alpha_2 p'_2 \alpha_a - \dots \\ &\quad - \prod_{u=1}^{a-2} (Id_{\sum_{k=1}^{\kappa} J_k} - \alpha_u p'_u) \alpha_{a-1} p'_{a-1} \alpha_a \\ &= \alpha_a - \alpha_1 p'_1 \alpha_a - \sum_{l=2}^{a-1} \prod_{u=1}^{l-1} (Id_{\sum_{k=1}^{\kappa} J_k} - \alpha_u p'_u) \alpha_l p'_l \alpha_a \\ &= \alpha_a - \delta_1 p'_1 \alpha_a - \sum_{l=2}^{a-1} \delta_l p'_l \alpha_a \\ &= \alpha_a - \sum_{l=1}^{a-1} \delta_l p'_l \alpha_a. \end{aligned}$$

D'où le résultat. La matrice $\mathcal{P}'_a \mathbf{W}_a$ est bien inversible car elle est triangulaire supérieure, avec des 1 sur la diagonale. —■—

Ainsi, $\mathcal{P}'_a \mathbf{W}_a^* = Id_a$: \mathcal{P}'_a est un inverse généralisé à gauche de \mathbf{W}_a^* dans les ACIMOG-PLS 2 et 3.

Nous avons maintenant les outils pour établir une nouvelle écriture des modèles \hat{Y}_n^A des ACIMOG-PLS 2 et 3 :

$$\begin{aligned}\hat{Y}_n^A &= \sum_{a=1}^A P_{t_a} Y_n = \sum_{a=1}^A P_{t_a} Y_n^{(a-1)} \\ &= \sum_{a=1}^A t_a c_a^{n'} = \mathbf{T}_a C_a^{n'},\end{aligned}$$

où $c_a^n = Y_n' D t_a / \|t_a\|_D^2 = Y_n^{(a-1)'} D t_a / \|t_a\|_D^2$ et $C_a^n = [c_1^n \dots c_a^n]$. Ce sont les équivalents de p_a^k et \mathbf{P}_a^k pour les prédicteurs. Nous obtenons ainsi, $\hat{Y}_n^A = X \mathbf{W}_A (\mathcal{P}'_A \mathbf{W}_A)^{-1} C_A^{n'} = X \hat{\beta}_n^A$ et $\hat{\beta}_n^A = \mathbf{W}_A (\mathcal{P}'_A \mathbf{W}_A)^{-1} C_A^{n'}$. Ceci montre que les résultats et propriétés de la régression PLS se généralisent au cas de l'ACIMOG-PLS.

Proposition 5.6 *Si $A = rg(X)$ alors $\hat{\beta}_n^A = (X' D X)^+ X' D Y_n$, c'est-à-dire, les modèles des ACIMOG-PLS 2 et 3 sont égaux aux modèles de la régression multi-réponses ordinaire de Y_n sur X .*

Dès que $rg(\mathbf{T}_a) = rg(X)$ alors $\hat{\beta}_n^a = (X' D X)^+ X' D Y_n$ dans l'ACIMOG-PLS1.

Démonstration :

Pour l'ACIMOG-PLS 2 et 3 :

Soit $A = rg(X)$. On a $\mathbf{T}_A = [t_1 \ \dots \ t_a] = X \mathbf{W}_A^*$. \mathbf{T}_A engendre un sous-espace de $Im(X)$. Or, les composantes t_a forment une base orthogonale de l'espace engendré par \mathbf{T}_A , d'où cet espace est de dimension A . Ainsi, \mathbf{T}_A et X engendrent le même espace et $P_X = P_{\mathbf{T}_A}$. Les modèles s'écrivent donc $\hat{Y}_n^A = P_{\mathbf{T}_A} Y_n = P_X Y_n = X (X' D X)^+ X Y_n$ qui est le modèle de la régression multi-réponses ordinaire de Y_n sur X .

Pour l'ACIMOG-PLS1 :

Soit $u = rg(X) = rg(\mathbf{T}_a)$. Comme $\mathbf{T}_a = [t_1 \ \dots \ t_a] = X \mathbf{W}_a^*$, l'espace engendré par \mathbf{T}_a , $Im(\mathbf{T}_a)$ est inclus dans l'espace engendré par X , $Im(X)$. Puisque les deux espaces ont la même dimension u , ils engendrent le même espace et $P_X = P_{\mathbf{T}_a}$. D'où le résultat. —■—

5.6.4 Interprétation des modèles

La propriété 5.3 nous montre que les variables réponses de chaque tableau Y_n sont modélisées comme combinaison linéaire de chacune des variables explicatives. Un tel modèle permet

de juger de l'influence d'une variable explicative dans le modèle. En règle générale, ce n'est pas cela qui importe lorsqu'on modélise des données d'une telle forme. Nous aimerions connaître l'influence d'un tableau X_k sur chaque tableau réponse Y_n . La section 6.3.2 du chapitre 6 discute de cette question.

5.6.5 Commentaires

Dans l'ACIMOG-PLS1, l'orthogonalité des axes $a_{k,a}$ et $a_{k,\ell}$ ($\ell \neq a$) dans \mathbb{R}^{J^k} leur permet de définir un repère orthonormé de \mathbb{R}^{J^k} sur lequel on peut projeter les l individus. Cependant, les coordonnées des points projetés ne sont pas données par les coordonnées des composantes $t_{k,a}$ et $t_{k,\ell}$ (voir annexe C). On utilise toutefois ces composantes pour effectuer une carte des individus «vus» par chacun des tableaux explicatifs, comme cela est pratiqué en régression PLS et dans toutes les méthodes multi-tableaux. L'orthogonalité des composantes $t_{k,a}$ et $t_{k,\ell}$ ($\ell \neq a$) dans \mathbb{R}^1 leur permet de définir un repère orthonormé (en les normant à l'unité) dans lequel on peut projeter les variables explicatives du tableau X_k . Si les tableaux sont réduits, les coordonnées des variables projetées dans ce repère sont données par les corrélations entre les variables de X_k et les composantes $t_{k,a}$ et $t_{k,\ell}$. Ces cartes des variables permettent d'interpréter les composantes de chaque tableau explicatif. Si les données ne sont pas réduites, on utilise les covariances entre les variables et les composantes normées à l'unité.

Dans l'ACIMOG-PLS1, les super-composantes $t_a = \sum_{k=1}^K t_{k,a}$ sont aussi orthogonales entre elles dans \mathbb{R}^1 . t_a et t_ℓ permettent de définir un repère orthonormé dans lequel on peut projeter les variables réponses. Les coordonnées de ces points projetés sont données par les corrélations des variables réponses avec les super-composantes t_a et t_ℓ (si les tableaux sont réduits, sinon on utilise les covariances). Ces coordonnées sont proportionnelles aux coordonnées des vecteurs $b_{n,a} = Y_n' D t_a / \|Y_n' D t_a\|_{Q_n^y}$.

Les coordonnées du vecteur $a_{k,a}$, elles, ne sont pas proportionnelles aux corrélations entre t_a et les variables de X_k . Mais si t_a et $u_a = \sum_{n=1}^N u_{n,a}$ sont très proches, cela devrait être presque le cas. S'il n'y a qu'une seule variable réponse y à expliquer, si t_a explique bien y , les positions relatives des points dont les coordonnées sont celles de $a_{k,a}$, qui sont proportionnelles à $X_k' D y$, doivent être similaires à celles données par les corrélations entre X_k et t_a .

Dans l'ACIMOG-PLS3, nous avons vu avec la proposition 5.4 que les vecteurs α_a de $\mathbb{R}^{\sum_{k=1}^K J^k}$ sont orthogonaux pour la métrique identité lorsque $Q_k^x = Id_{J^k}$. Rappelons que ces vecteurs sont tels que $t_a = X^{(a-1)} \alpha_a$. Cette orthogonalité permet à deux vecteurs α_a et α_ℓ de définir un repère orthonormé dans le quel on peut projeter les individus «globaux» de $\mathbb{R}^{\sum_{k=1}^K J^k}$. Les coordonnées sur l'axe a des points projetés ne sont malheureusement pas égales aux coordonnées du vecteur $t_a / \|\alpha_a\| = t_a / \sum_{k=1}^K \mu_{k,a}^2$, excepté lorsque $a = 1$. Les points dont les coordonnées sont données par (t_a, t_ℓ) permettent (à un coefficient près) de visualiser les points individus dans $\mathbb{R}^{\sum_{k=1}^K J^k}$. Cependant, les proximités entre les points sont à interpréter avec précautions puisqu'ils sont représentés dans un repère orthonormé alors qu'ils ne sont pas les projections des individus dans un repère orthonormé.

Inversement, l'orthogonalité des super-composantes t_a entre elles permet de définir des repères orthonormés dans \mathbb{R}^1 dont lesquels on peut projeter les variables. Les coordonnées des points-variables projetés sont obtenues par les corrélations (ou covariances) des variables avec les super-composantes.

Dans l'ACIMOG-PLS2 les problèmes subsistent pour faire des cartes des individus. Nous n'avons pas d'orthogonalité ni dans \mathbb{R}^{J_k} ni dans $\mathbb{R}^{\sum_{k=1}^K J_k}$. Cependant, nous pouvons faire une carte des individus «globaux» grâce aux points de coordonnées (t_a, t_ℓ) dans un repère orthonormé. Mais en faisant attention à l'interprétation de la proximité entre les points car ces coordonnées sont normalement valables dans un repère non orthogonal.

Les super-composantes t_a de l'ACIMOG-PLS3 mettent en valeur dans leur expression, les composantes partielles $t_{k,a}$ qui sont le plus covariées avec l'ensemble des tableaux réponses. Tandis que dans l'ACIMOG-PLS2, toutes les composantes $t_{k,a}$, $k = 1, \dots, K$ ont la même importance dans le calcul de t_a .

Dans l'ACIMOG-PLS2,3, les super-composantes t_a forment une base de ImX . X étant au plus de rang $\min(l-1, \sum_k J_k)$, A ne peut dépasser cette valeur. Dans l'ACIMOG-PLS1, le rang de T_a est au plus $\min(l-1, Ka)$. De plus, il au plus égal au rang de X . Aussi, A ne peut dépasser la valeur pour laquelle $K(A-1) \leq rg(X) \leq KA$.

Si $I < K$, et que les tableaux sont D -centrés, $Im(X)$ est de dimension $l-1$ (en supposant que les individus sont linéairement indépendants). Alors, l'ACIMOG-PLS1 ne présente pas grand intérêt car l'espace engendré par $T_1 = [t_{1,1} \dots t_{K,1}]$ sera sûrement de dimension $l-1$ et dès la deuxième étape il sera impossible de trouver un des composantes $t_{k,2}$ engendrant un espace orthogonal à T_1 .

Lorsque $N = 1$ et $M_1 = 1$, c'est-à-dire lorsque les réponses sont composées d'une seule et unique variable y , la résolution de l'algorithme de l'ACIMOG-PLS n'est ni itératif, ni une diagonalisation. La solution est directe, et revient à faire K régressions PLS séparées de y sur les différents X_k à l'étape 1, uniquement. Aux étapes suivantes, l'ACIMOG-PLS1 permet de trouver l'information contenue dans les X_k qui soit la plus proche de y et qui n'a pas été détectée par toutes les composantes $t_{k,a}$ précédentes. Ainsi, en très peu d'étapes, nous obtenons un modèle pour y très explicatif. Bien sûr, à l'étape a , la part de variation de y expliquée par la composante $t_{k,a}$ est au mieux aussi grande que la part de variation expliquée par la composante de même ordre associée à X_k dans la régression PLS de y sur X_k . Remarquons de plus, que les composantes $t_{k,a}$ pour un a fixé peuvent expliquer en partie la même information.

Toujours lorsque $N = 1$ et $M_1 = 1$, l'ACIMOG-PLS3 donne les mêmes modèles que MPLS (WOLD et al., 1987). En effet, la pondération des $t_{k,a}$ par $\mu_{k,a}$ dans l'expression de t_a se simplifie car $t_{k,a} = \frac{1}{\mu_{k,a}} X_k Q_k^x X_k' D y$. Alors, $t_a = \sum_{k=1}^K X_k Q_k^x X_k' D y$ et la composante explicative de même ordre issue de la MPLS sont proportionnelles et engendrent le même espace. Ce qui engendre des modèles identiques.

Dans le cas où tous les J_k sont égaux à 1, chaque \hat{Y}_n^1 dans l'ACIMOG-PLS1 correspond au modèle de la régression linéaire multi-réponses : $P_{[x_1 \dots x_K]} Y_n$.

5.6.6 A propos de la déflation et variantes

Comme l'ont souligné WESTERHUIS & SMILDE (2001), les méthodes de régression multi-tableaux où la déflation se fait sur les super-composantes engendrent des problèmes au niveau de l'interprétation dans le bloc X_k .

En effet, considérons les deux premières étapes. Les tableaux utilisés à l'étape 2 sont $X_k^{(1)}$ et $Y_n^{(1)}$. Cela signifie que

$t_{k,2} = \sum_{k=1}^K (X_k - P_{t_1}) Q_k^x a_{k,2} = \sum_{k=1}^K (X_k - \sum_{k'=1}^K X_{k'} a_{k',1} t_1' D X_{k'} / \|t_1\|_D^2) Q_k^x a_{k,2}$: la composante $t_{k,2}$ associée au tableau k dépend des variables de tous les autres tableaux. L'information est mélangée. Une solution serait de déflater les tableaux X_k sur leur propre composante, $t_{k,a}$, et les Y_n sur la super-composante t_a .

C'est ce que font WANGEN & KOWALSKI (1988) dans leur version de la MBPLS avec $N = 1$. Dans ce cas, à chaque X_k on enlève toute l'information apportée par la composante t_k à l'étape précédente, et à Y_n , que l'on veut expliquer, on n'enlève seulement qu'une partie de l'information apportée par chacun des X_k puisque t est une combinaison linéaire des t_k . On perd donc de l'information à chaque étape, et WESTERHUIS & COENEGRACHT (1997) ont montré que ces modèles de la MBPLS étaient moins explicatifs qu'en projetant tous les tableaux sur la super-composante t_a . Rajoutons que les modèles ne sont pas construits par les mêmes composantes, ce qui peut compliquer l'interprétation.

Si l'on n'utilise pas les ACIMOG-PLS2 et 3 mais l'ACIMOG-PLS1, aucun de ces problèmes n'existe. En effet, tous les tableaux sont déflatés sur le même espace à chaque étape, il n'y a donc pas de perte d'information, et de plus, les composantes $t_{k,a}$ du bloc X_k sont orthogonales aux composantes de ce même bloc obtenues dans les étapes précédentes et à celles des autres blocs des étapes inférieures. Il n'y a pas de mélange de l'information et on est assuré que chacune des composantes $t_{k,a}$ explique quelque chose de nouveau, qui n'a pas été expliqué précédemment.

Parlons alors de l'ACIMOG et de ses déflations. Les déflations se réalisent soit par bloc (chaque tableau déflaté sur sa propre composante) soit sur des super-composantes (les prédicteurs sur le même t et les réponses sur le même u). La première méthode ne possède pas de problème de mélange d'information. Par contre, les ACIMOG 2 et 3 ont ce problème.

Pour la MBPLS, avec $N = 1$, WESTERHUIS & SMILDE (2001) proposent de ne pas déflater les prédicteurs, mais d'orthogonaliser les super-composantes t_T d'une étape à l'autre. Une variante de l'ACIMOG-PLS 2 et 3 est de faire de même. Notons que cela n'est pas applicable à l'ACIMOG 2 et 3 car à chaque étape le critère à optimiser serait le même. Il faudrait optimiser le critère avec des contraintes d'orthogonalité sur les super-composantes, qui est un problème moins simple à résoudre.

Définition 5.5 Une variante de l'ACIMOG-PLS 2 et 3 d'ordre $a > 1$ est définie par la maximisation du critère $\sum_{k=1}^K \sum_{n=1}^N cov(X_k Q_k^x a_k, Y_n^{(a-1)} Q_n^y b_n)$ sous les contraintes $\|a_k\|_{Q_k^x}^2 = 1$ et $\|b_n\|_{Q_n^y}^2 = 1$ où $Y_n^{(a-1)} = P_{t_{a-1}^*}^\perp Y_n^{(a-2)}$, t_a^* étant la super-composante d'ordre a définie

comme suit :

$$\begin{aligned} t_1^* &= t_1 \\ t_2^* &= t_2 - t_1^* t_1^{*'} D t_2 / \|t_1^*\|_D^2 = P_{t_1^*}^\perp t_2 \\ &\vdots \\ t_a^* &= P_{t_1^*, \dots, t_{a-1}^*}^\perp t_a \end{aligned}$$

avec $t_a = \sum_{k=1}^K X_k Q_k^x a_{k,a}$ pour l'ACIMOG-PLS2 et $t_a = \sum_{k=1}^K \text{cov}(X_k Q_k^x a_{k,a}, \sum_{n=1}^N Y_n^{(a-1)} Q_n^y b_{n,a}) X_k Q_k^x a_{k,a}$ dans le cas de l'ACIMOG-PLS3.

Les super-composantes t_a^* sont donc définies grâce à la méthode d'orthogonalisation de Gram-Schmidt.

Les modèles alors obtenus s'écrivent $Y_n = \sum_{a=1}^A P_{t_a^*} Y_n + Y_n^{(A)}$, somme de modèles partiels orthogonaux.

Proposition 5.7 *Les super-composantes issues des ACIMOG-PLS 2 et 3 sont égales à celles issues de leur variante, et les modèles \hat{Y}_n^a sont les mêmes.*

Démonstration :

Nous notons $t_a, a_{k,a}, b_{n,a}, \mu_{k,a}$ et $\lambda_{n,a}$ les objets d'ordre a issus d'une ACIMOG-PLS 2 ou 3, $t_a^*, a_{k,a}^*, b_{n,a}^*, \mu_{k,a}^*$ et $\lambda_{n,a}^*$ les objets d'ordre a issus de leur variante. Pour la première étape, les deux techniques sont identiques $t_1^* = t_1$.

A l'étape a les solutions $a_{k,a}$ et $b_{n,a}$ de la méthode originelle vérifient le système d'équations :

$$\begin{aligned} a_{k,a} &= X_k^{(a-1)'} D \sum_{n=1}^N Y_n^{(a-1)} Q_n^y b_{n,a} / \mu_{k,a} \\ &= X_k' D P_{t_1^*, \dots, t_{a-1}^*}^\perp \sum_{n=1}^N P_{t_1^*, \dots, t_{a-1}^*}^\perp Y_n Q_n^y b_{n,a} / \mu_{k,a} \\ &= X_k' D \sum_{n=1}^N Y_n^{(a-1)} Q_n^y b_{n,a} / \mu_{k,a} \\ b_{n,a} &= Y_n^{(a-1)'} D \sum_{k=1}^K X_k^{(a-1)} Q_k^x a_{k,a} / \lambda_{n,a} \\ &= Y_n' D P_{t_1^*, \dots, t_{a-1}^*}^\perp \sum_{k=1}^K P_{t_1^*, \dots, t_{a-1}^*}^\perp X_k Q_k^x a_{k,a} / \lambda_{n,a} \\ &= Y_n^{(a-1)'} D \sum_{k=1}^K X_k Q_k^x a_{k,a} / \lambda_{n,a}, \end{aligned}$$

car les composantes t_{a-1} sont D -orthogonales aux composantes d'ordre inférieur. Ceci implique aussi que $\mu_{k,a} = \text{cov}(X_k Q_k^x a_{k,a}, \sum_{n=1}^N Y_n^{(a-1)} Q_n^y b_{n,a})$ et $\lambda_{n,a} =$

$cov(Y_n^{(a-1)}, \sum_{k=1}^K X_k Q_k^x a_{k,a})$, qui sont aussi les expressions de $\mu_{k,a}^*$ et $\lambda_{n,a}^*$. Si $Y_n^{(a-1)}$ est le même dans les deux méthodes, alors $\mu_{k,a} = \mu_{k,a}^*$ et $\lambda_{n,a} = \lambda_{n,a}^*$. Les solutions $a_{k,a}^*$ et $b_{n,a}^*$ de la variante vérifient le système d'équations :

$$\begin{aligned} a_{k,a}^* &= X_k' D \sum_{n=1}^N Y_n^{(a-1)} Q_n^y b_{n,1} / cov(X_k Q_k^x a_{k,a}, \sum_{n=1}^N Y_n^{(a-1)} Q_n^y b_{n,a}) \\ b_{n,a}^* &= Y_n^{(a-1)'} D \sum_{k=1}^K X_k Q_k^x a_{k,1} / cov(Y_n^{(a-1)}, \sum_{k=1}^K X_k Q_k^x a_{k,a}). \end{aligned}$$

Ainsi, $a_{k,a} = a_{k,a}^*$ et $b_{n,a}^* = b_{n,a}$ si $Y_n^{(a-1)}$ est le même dans les deux méthodes.

Pour $a = 2$, $Y_n^{(1)} = P_{t_1}^\perp Y_n = P_{t_1^*}^\perp Y_n$. Donc, $a_{k,2} = a_{k,2}^*$, $b_{n,2}^* = b_{n,2}$, $\mu_{k,2} = \mu_{k,2}^*$ et $\lambda_{n,2} = \lambda_{n,2}^*$. La super-composante issue de la méthode originale est $t_2 = \sum_{k=1}^K X_k^{(1)} Q_k^x a_{k,2} = \sum_{k=1}^K P_{t_1}^\perp X_k Q_k^x a_{k,2}$, qui est exactement la valeur de t_2^* .

Par induction, on montre que $t_a = t_a^*$, $\mu_{k,a} = \mu_{k,a}^*$, $\lambda_{n,a} = \lambda_{n,a}^*$, $a_{k,a}^* = a_{k,a}$ et $b_{n,a}^* = b_{n,a}$.

On suppose que ces égalités sont vérifiées jusqu'à l'étape $a - 1$. Montrons qu'elles sont exactes à l'étape a . Dans l'ACIMOG-PLS, $Y_n^{(a-1)} = P_{t_1, \dots, t_{a-1}}^\perp Y_n$. Ce qui est égal, d'après l'hypothèse d'induction, à $Y_n^{(a-1)} = P_{t_1^*, \dots, t_{a-1}^*}^\perp Y_n$. Ce qui implique les égalités sur les axes et les coefficients μ_k et λ_n .

Ainsi, dans la variante de l'ACIMOG-PLS 2,

$$\begin{aligned} t_a^* &= P_{t_1^*, \dots, t_{a-1}^*}^\perp \sum_{k=1}^K X_k Q_k^x a_{k,a}^* \\ &= P_{t_1^*, \dots, t_{a-1}^*}^\perp \sum_{k=1}^K X_k Q_k^x a_{k,a} \\ &= P_{t_1, \dots, t_{a-1}}^\perp \sum_{k=1}^K X_k Q_k^x a_{k,a} \\ &= \sum_{k=1}^K X_k^{(a-1)} Q_k^x a_{k,a} \\ &= t_a. \end{aligned}$$

Les super-composantes d'ordre a de l'ACIMOG-PLS2 et de sa variantes sont donc identiques. Pour l'ACIMOG-PLS 3, le raisonnement est le même. Il suffit d'utiliser $t_a^* = P_{t_1^*, \dots, t_{a-1}^*}^\perp \sum_{k=1}^K \mu_{k,a}^* X_k Q_k^x a_{k,a}^*$ et le fait que $\mu_{k,a} = \mu_{k,a}^*$.

Ainsi, les modèles \hat{Y}_n^A obtenus sont les mêmes que l'on utilise l'ACIMOG-PLS ou sa variante. —■—

Par contre, les composantes par blocs ne sont pas identiques. Dans la variante, elles sont égales à $X_k Q_k^x a_{k,a}$ alors que dans la méthode originale elles sont égales à $X_k^{(a-1)} Q_k^x a_{k,a}$. Ainsi, les

composantes par bloc de la variante ne dépendent que des variables de leur propre groupe. L'information n'est plus mélangée d'une étape à l'autre.

Cependant, puisque les modèles des deux approches sont égaux, il ne nous paraît pas nécessaire de procéder ainsi, car les composantes par blocs ne sont orthogonales entre elles dans aucune des approches. De même, cela ne permet pas d'obtenir des axes orthogonaux. Cela ne permet pas non plus de résoudre les problèmes liés aux représentations graphiques des individus décrits dans les commentaires (section précédent) et dans l'annexe C. De plus, le pourcentage de variance expliquée du tableau X_k par une étape est égale à $\text{var}(P_{t_a}^* X_k) / \text{var}(X_k)$. Cela signifie qu'il est calculé à partir des super-composantes. Il ne semble donc pas logique d'utiliser $X_k Q_k^x a_{k,a}$ pour interpréter ce qu'il se fait dans X_k plutôt que $X_k^{(a-1)} Q_k^x a_{k,a}$, qui sert à la construction de la super-composante qui «explique» X_k .

Une telle variante dans l'ACIMOG-PLS1 n'est d'aucune utilité car les composantes par blocs sont déjà orthogonales. Une telle variante ne ferait que supprimer cette propriété d'orthogonalité.

5.7 REgression Multi Blocs : REMUB

Nous proposons ici une autre nouvelle méthode de régression Multi-Bloc. Pour la construire, nous avons changé de critère. En effet, l'étude de l'ACIMOG-PLS2 et 3 et des méthodes telles que HPLS (WOLD et al., 1996) nous ont amené à penser qu'ils faudrait trouver les meilleures super-composantes, combinaisons linéaires des composantes partielles, les composantes partielles étant les meilleurs résumés des tableaux.

5.7.1 La méthode

Définition 5.6 *L'étape a de la REgression Multi-Bloc est définie par la maximisation du critère*

$$C_3(\{a_k\}, \{b_n\}, \{\alpha_k\}, \{\nu_n\}) = \sum_{k=1}^K \alpha_k \underbrace{a_k' Q_k^x X_k^{(a-1)'}}_{t_k'} D \sum_{n=1}^N \nu_n \underbrace{Y_n^{(a-1)} Q_n^y b_n}_{u_n} = \text{cov}\left(\sum_{k=1}^K \alpha_k t_k, \sum_{n=1}^N \nu_n u_n\right)$$

sous les contraintes de normes :

$$\sum_{k=1}^K \alpha_k^2 = 1, \quad (5.9)$$

$$\sum_{n=1}^N \nu_n^2 = 1, \quad (5.10)$$

$$a_k' Q_k^x a_k = 1, \quad k = 1, \dots, K \quad (5.11)$$

$$b_n' Q_n^y b_n = 1, \quad n = 1, \dots, N \quad (5.12)$$

$$\text{avec } X^{(0)} = X_k; \quad X_k^{(a-1)} = X_k^{(a-2)} - P_{\sum_{k=1}^K \alpha_{k,a} t_{k,a}} X_k^{(a-2)} \quad \text{et } Y_n^{(0)} = Y_n;$$

$$Y_n^{(a-1)} = Y_n^{(a-2)} - P_{\sum_{k=1}^K \alpha_{k,a} t_{k,a}} Y_n^{(a-2)}.$$

Nous n'allons donner que les solutions d'ordre 1, les autres en découleront.
Le lagrangien est :

$$\begin{aligned} \mathcal{L}(\{a_k\}, \{b_n\}, \alpha, \nu) &= \sum_{k=1}^K \alpha_k a'_k Q_k^x X'_k D \sum_{n=1}^N \nu_n Y_n Q_n^y b_n - \sum_{k=1}^K \frac{\mu_k}{2} (a'_k Q_k^x a_k - 1) \\ &\quad - \sum_{n=1}^N \frac{\lambda_n}{2} (b'_n Q_n^y b_n - 1) - \frac{\beta}{2} (\alpha' \alpha - 1) - \frac{\gamma}{2} (\nu' \nu - 1). \end{aligned} \quad (5.13)$$

Il mène aux équations normales suivantes :

$$\nabla_{a_k} \mathcal{L} = \sum_{n=1}^N \nu_n \alpha_k Q_k^x X'_k D Y_n Q_n^y b_n - \mu_k Q_k^x a_k = 0 \quad (5.14)$$

$$\nabla_{b_n} \mathcal{L} = \sum_{k=1}^K \alpha_k \nu_n Q_n^y Y'_n D X_k Q_k^x a_k - \lambda_n Q_n^y b_n = 0 \quad (5.15)$$

$$\nabla_{\alpha_k} \mathcal{L} = \sum_{n=1}^N \nu_n a'_k Q_k^x X'_k D Y_n Q_n^y b_n - \beta \alpha_k = 0 \quad (5.16)$$

$$\nabla_{\nu_n} \mathcal{L} = \sum_{k=1}^K \alpha_k b'_n Q_n^y Y'_n D X_k Q_k^x a_k - \gamma \nu_n = 0 \quad (5.17)$$

$$a'_k Q_k^x a_k - 1 = 0 \quad (5.18)$$

$$b'_n Q_n^y b_n - 1 = 0 \quad (5.19)$$

$$\sum_{k=1}^K \nu_k^2 - 1 = 0 \quad (5.20)$$

$$\sum_{n=1}^N \beta_n^2 - 1 = 0. \quad (5.21)$$

Les points critiques vérifient les relations :

$$a_k = \frac{1}{\mu_k} \alpha_k X'_k D \sum_{n=1}^N \nu_n Y_n Q_n^y b_n \quad (5.22)$$

$$b_n = \frac{1}{\lambda_n} \nu_n Y'_n D \sum_{k=1}^K \alpha_k X_k Q_k^x a_k \quad (5.23)$$

$$\alpha_k = \frac{1}{\beta} t_k' D \sum_{n=1}^N \nu_n u_n \quad (5.24)$$

$$\nu_n = \frac{1}{\gamma} u_n' D \sum_{k=1}^K \alpha_k t_k \quad (5.25)$$

$$\mu_k = \alpha_k t_k' D \sum_{n=1}^N \nu_n u_n \quad (5.26)$$

$$\lambda_n = \nu_n u_n' D \sum_{k=1}^K \alpha_k t_k \quad (5.27)$$

$$\beta = \gamma = \sum_{k=1}^K \alpha_k a_k' Q_k^x X_k' D \sum_{n=1}^N \nu_n Y_n Q_n^y b_n. \quad (5.28)$$

Ainsi, les équations (5.24) et (5.25) mènent à l'écriture matricielle :

$$\alpha = \frac{1}{\beta} [t_1 \ \cdots \ t_K]' D [u_1 \ \cdots \ u_N] \nu \quad (5.29)$$

$$\nu = \frac{1}{\gamma} [u_1 \ \cdots \ u_N]' D [t_1 \ \cdots \ t_K] \alpha. \quad (5.30)$$

Ces résultats s'obtiennent aisément, aussi nous ne détaillerons pas les calculs ici.

Remarque : L'optimum du critère est égal à $\beta = \gamma = \sum_{k=1}^K \mu_k = \sum_{n=1}^N \lambda_n$. Celui-ci est bien un maximum car nous choisissons β comme la plus grande valeur propre de $[t_1 \ \cdots \ t_K]' D [u_1 \ \cdots \ u_N] [u_1 \ \cdots \ u_N]' D [t_1 \ \cdots \ t_K]$ associée à α .

Pour approcher les solutions, un algorithme croissant et convergent (afin de maximiser la fonction objectif assurément) si possible doit être mis en place. On peut noter que les vecteurs α et ν sont solutions de la régression PLS de $[u_1 \ \cdots \ u_N]$ sur $[t_1 \ \cdots \ t_K]$ et les vecteurs a_k et b_n sont solutions de la première étape de l'ACIMOG appliquée aux tableaux $\alpha_k X_k$ et $\nu_n Y_n$ munis des métriques D , Q_k^x et Q_n^y , d'où l'algorithme qui suit.

5.7.2 L'algorithme

La figure 5.5 illustre cet algorithme.

étape 0 : initialisation

Les tableaux X_k , Y_n sont supposés centrés et éventuellement réduits en colonne pour D .

$$\alpha = \left(\frac{1}{\sqrt{K}} \ \cdots \ \frac{1}{\sqrt{K}} \right)'$$

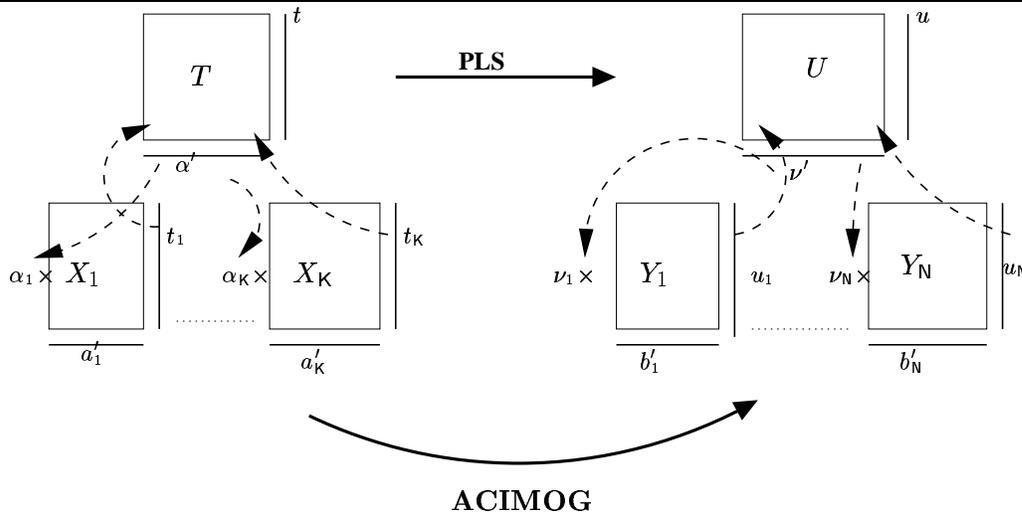
$$\nu = \left(\frac{1}{\sqrt{N}} \ \cdots \ \frac{1}{\sqrt{N}} \right)'$$

$$\mathcal{X}_k = \alpha_k X_k; \mathcal{Y}_n = \nu_n Y_n$$

$$b_n = \left(\frac{1}{\sqrt{M_n}} \ \cdots \ \frac{1}{\sqrt{M_n}} \right)'$$

$$a_k = \mathcal{X}_k' D \sum_{n=1}^N \mathcal{Y}_n b_n; a_k = a_k / \|a_k\|_{Q_k^x}$$

FIG. 5.5 – L'algorithme REMUB

**étape 1: étape 1 de l'ACIMOG**

$\mathcal{X}_k = \alpha_k X_k$; $\mathcal{Y}_n = \nu_n Y_n$
 $crit0 = \sum_{k=1}^K a'_k Q_k^x \mathcal{X}'_k D \sum_{n=1}^N \mathcal{Y}_n Q_n^y b_n$
tant que $|crit1 - crit0| \geq crit1 \epsilon$
faire :
 $b_n = \mathcal{Y}'_n D \sum_{k=1}^K \mathcal{X}_k Q_k^x a_k$
 $\lambda_n = \|b_n\|_{Q_n^y}$; $b_n = b_n / \lambda_n$
 $a_k = \mathcal{X}'_k D \sum_{n=1}^N \mathcal{Y}_n Q_n^y b_n$
 $\mu_k = \|a_k\|_{Q_k^x}$; $a_k = a_k / \mu_k$
 $t_k = \mathcal{X}_k Q_k^x a_k$; $u_n = \mathcal{Y}_n Q_n^y b_n$
 $crit1 = \sum_{k=1}^K t'_k D \sum_{n=1}^N u_n$
 $crit0 = crit1$

étape 2: actualisation des composantes

$t_k = t_k / \alpha_k$
 $u_n = u_n / \nu_n$

étape 3: étape NIPALS

tant que $|crit2 - crit1| \geq crit2 \epsilon$
faire :
 $\alpha = [t_1 \ \dots \ t_K]' D [u_1 \ \dots \ u_N] \nu$
 $\alpha = \alpha / \|\alpha\|$
 $\nu = [u_1 \ \dots \ u_N]' D [t_1 \ \dots \ t_K] \alpha$
 $\nu = \nu / \|\nu\|$
 $crit2 = \nu' [u_1 \ \dots \ u_N]' D [t_1 \ \dots \ t_K] \alpha$
 $crit1 = crit2$

étape 4 : test de la convergence

Si $|crit1 - crit| \leq crit1\epsilon$, aller à l'étape 4.

Sinon, $crit = crit1$ et retourner à l'étape 1.

étape 5 : Passage à l'ordre supérieur

$X_k = X_k - P_{\sum_{k=1}^K \alpha_k t_k} X_k$; $Y_n = Y_n - P_{\sum_{k=1}^K \alpha_k t_k} Y_n$
retourner à l'étape 0

—■—

Monotonicit  et convergence :

Propri t  5.8 *Cet algorithme fait cro tre le crit re C_3 et converger vers un maximum local.*

D monstration :

L'actualisation de a_k et b_n fait cro tre le crit re car c'est une  tape ACIMOG (chap. 3.4). De plus, l'actualisation de ν et α le fait aussi augmenter car c'est une  tape NIPALS. En effet, notons $\bar{\alpha} = \frac{T'DU\nu}{\|T'DU\nu\|}$, $T = [t_1 \ \cdots \ t_K]$ et $U = [u_1 \ \cdots \ u_N]$. Au moment de l'actualisation de α , l'accroissement du crit re s' crit :

$$h(\bar{\alpha}) - h(\alpha) = \bar{\alpha}'T'DU\nu - \alpha'T'DU\nu$$

La fonction h est  gale au crit re lorsque toutes les variables sauf α sont fix es.

$$\begin{aligned} h(\bar{\alpha}) - h(\alpha) &= \frac{1}{\|T'DU\nu\|} \nu'U'DTT'DU\nu - \alpha'T'DU\nu \\ &= \|T'DU\nu\| (1 - \cos(\alpha, T'DU\nu)) \\ &\geq 0. \end{aligned}$$

L'accroissement du crit re,   chaque actualisation de α est positif. De plus, $|h(\alpha)|$ est born e par $\|T'DU\nu\|$. Un r sultat similaire s'obtient facilement pour l'accroissement du crit re lors de la mise   jour de ν . L' tape NIPALS est bien croissante et convergente.

Il reste   montrer que l'algorithme converge. Or, $h(\alpha)$ est  gal   la fonction objectif $C_3(\{a_k\}, \{b_n\}, \{\alpha_k\}, \{\nu_n\})$. D'o , le crit re  tant positif,

$$C_3 \leq \|T'DU\nu\| \tag{5.31}$$

$$\leq \|Diag(a'_k Q_k^{x^{1/2}})Diag(Q_k^{x^{1/2}})X'DYDiag(Q_n^{y^{1/2}})Diag(Q_n^{y^{1/2}} b_n)\nu\| \tag{5.32}$$

$$\leq \|Diag(Q_k^{x^{1/2}} a_k)\| \|Diag(Q_k^{x^{1/2}})X'DYDiag(Q_n^{y^{1/2}})\| \|Diag(Q_n^{y^{1/2}} b_n)\| \|\nu\| \tag{5.33}$$

d'apr s l'in galit  de Cauchy-Schwarz,

$$\leq KN \|Diag(Q_k^{x^{1/2}})X'DYDiag(Q_n^{y^{1/2}})\|. \tag{5.34}$$

La fonction objectif est croissante et born e, donc converge vers un maximum local.

—■—

5.7.3 Propriétés et modèles

Des propriétés similaires à celles de l'ACIMOG-PLS sont valables pour REMUB. Aussi, nous les énoncerons simplement sans les démontrer.

A chaque étape, l'algorithme construit les super-composantes $t_a = T_a \alpha_a = \sum_{k=1}^K \alpha_{k,a} t_{k,a}$ et $u_a = U_a \nu_a = \sum_{n=1}^N \nu_{n,a} u_{n,a}$. Elles sont optimales car le critère C_3 maximisent leur covariances $cov(t_a, u_a)$ et les composantes $t_{k,a}$ et $u_{n,a}$ le sont aussi car C_3 est optimisé par blocs, et de plus elles sont solutions d'une ACIMOG particulière à chaque étape a .

Propriété 5.9 *Les super-composantes $\{t_a\}$ sont D -orthogonales.*

Propriété 5.10 *Les coefficients $\alpha_{k,a}$ sont positifs et proportionnels aux coefficients $\mu_{k,a}$. Les coefficients $\nu_{n,a}$ sont positifs et proportionnels aux coefficients $\lambda_{n,a}$.*

Démonstration :

Nous allons démontrer le résultat pour $\alpha_{k,1}$, le résultat se transposant facilement pour les étapes supérieures et pour $\nu_{n,a}$.

$$\begin{aligned} \alpha_k &= t'_k D u / \beta, \\ \text{et} \\ \mu_k &= \alpha_k t'_k D u. \end{aligned}$$

D'où, $\alpha_k^2 = \mu_k / \beta$.

De plus, μ_k est positif car égal à $\|\alpha_k X'_k D u\|_{Q_k^x}$. D'où $\alpha_k = \|X'_k D u\|_{Q_k^x} / \beta$ qui est positif car $\beta = \sum_{k=1}^K \mu_k$.

■

Les modèles s'écrivent de la même manière qu'en ACIMOG-PLS, comme somme de modèles partiels de rang D -orthogonaux et comme combinaison linéaire des variables initiales :

$$\widehat{Y}_n^A = \sum_{a=1}^A P_{t_a} Y_n + E_n^A \quad (5.35)$$

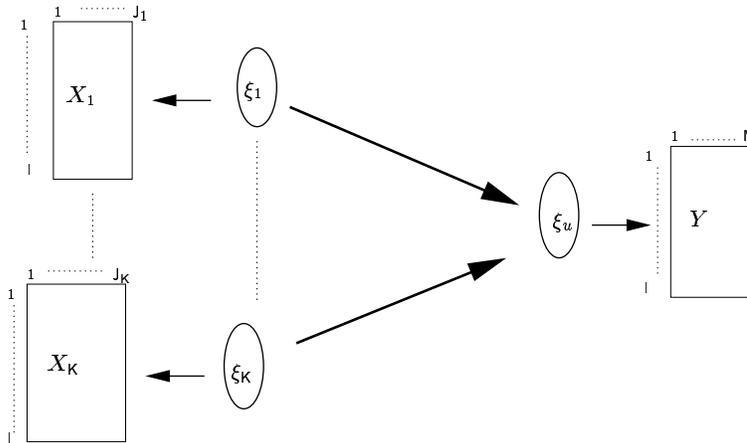
$$= X \widehat{\beta}_n^A. \quad (5.36)$$

Nous avons remarqué dans l'ACIMOG-PLS2 que la variance expliquée de Y_n par l'étape a était proportionnel à λ_n . Dans REMUB, nous avons une relation du même type : $\|P_{t_a} Y_n\|_D^2 = \lambda_{n,a} \beta / \|t_a\|_D^2$ ou encore $\|P_{t_a} Y_n\|_D^2 = \beta^2 \nu_{n,a}^2 / \|t_a\|_D^2$. Ces relations sont obtenues grâce aux équations (5.19) et (5.23) et à la relation entre $\nu_{n,a}$ et $\lambda_{n,a}$.

Ce résultat donne un rôle à $\lambda_{n,a}$: ils permettent de visualiser quels sont les tableaux Y_n les mieux reconstitués par l'étape a .

Dans la pratique, nous avons remarqué, tout comme en ACIMOG-PLS, que les coefficients $\mu_{k,a}$ (ou $\alpha_{k,a}$) semblent indiquer quels sont les tableaux X_k qui sont les mieux reconstitués par l'étape a , lorsque celle-ci a un bon pouvoir explicatif, c'est-à-dire lorsque t_a et u_a sont assez « proches ».

FIG. 5.6 – Schéma relationnel: $K + 1$ tableaux réfléchifs, K variables latentes exogènes ξ_k et une variable latente endogène ξ_u



TAB. 5.1 – Corrélations entre les variables latentes de l'APLS et les composantes de REMUB pour $a = 1$

	t_1	t_2	u	ξ_1^e	ξ_2^e	ξ_u^e
t_1	1	0.314	0.433	-1	-0.313	-0.432
t_2		1	0.769	0.314	-1	-0.763
u_1			1	0.430	-0.768	-1
ξ_1^e				1	-0.313	-0.431
ξ_2^e					1	0.762
ξ_u^e						1

Les super-composantes t_a permettent de donner une représentation des individus «compromis». Pour expliquer leurs positions, on utilise $cov(t_a/\|t_a\|_D, X_k)$. Si les données sont réduites, alors il s'agit de corrélations. Cette covariance est liée à la variance expliquée de X_k par la composante, de la même manière qu'en PLS ou ACIMOG-PLS: $\|P_{t_a} X_k\|_D^2 = \sum_{j=1}^{J_k} cov^2(x_k^j, t_a/\|t_a\|_D)$, x_k^j désignant la $j^{\text{ième}}$ variable de X_k .

Les super-composantes t_a donnent de l'importance aux composantes $t_{k,a}$ les plus covariées avec la super-composante u_a : $t_a = \sum_{k=1}^K \frac{cov(t_{k,a}, u_a)}{cov(t_a, u_a)} t_{k,a}$. Cela signifie aussi que les tableaux X_k dont la composante $t_{k,a}$ est fortement covariée avec u_a seront sans doute mieux expliqués par la composante t_a .

5.7.4 Cas particuliers

Considérons le schéma relationnel de la figure 5.6. Une solution pour le résoudre est d'utiliser l'approche PLS (chapitre 5.1) en suivant l'approche centroïde et le mode A. Les résultats obtenus par l'APLS donnent les mêmes résultats que REMUB à l'étape 1. Les étapes suivantes sont différentes à cause des déflations utilisées. En notant t_k les composantes issues de REMUB associées à X_k et u celle associée à Y , ξ_k^e les estimations externes des variables latentes associée à X_k suivant le schéma centroïde et le mode A, et ξ_u^e celle associée à Y , nous avons $r(t_k, t_\ell) = r(\xi_k^e, \xi_\ell^e)$ quelques soient $k \neq \ell$ et $r(t_k, u) = r(\xi_k^e, \xi_u^e)$.

Nous proposons un exemple des résultats obtenus par les 2 approches sur les données de

Russett que l'on peut trouver dans (TENENHAUS, 1998, p172). Dans cet exemple, $K = 2$, $J_1 = 3$, $J_2 = 4$ et $M = 6$. Le tableau 5.1 contient les corrélations entre les variables latentes et les composantes issues de REMUB à la première étape.

Chapitre 6

Autour de la pratique

Dans ce chapitre, nous nous attachons aux questions qui se posent lors de la manipulation pratique des méthodes que nous avons proposées en particulier ACIMOG(-PLS), REMUB. Nous exposons ici beaucoup de méthodes relatives à la régression PLS car nos méthodes en sont en quelques sortes des généralisations. Le chapitre se décompose en trois parties.

La première concerne le prétraitement des données. Nous rappellerons les prétraitements classiques en régression PLS et les méthodes dérivées de la régression PLS utilisées pour prétraiter des données de spectres infra-rouges en chimiométrie. Nous exposerons aussi ce qui peut être fait en régression multi-tableaux.

La seconde partie concerne la validation des modèles. Nous faisons référence au choix du nombre de composantes A à retenir dans les modèles. Nous rappellerons les critères classiques utilisés en régression PLS et exposerons la manière de les utiliser en régression multi-tableaux et parlerons rapidement de critères utilisés en analyse cubique.

Enfin, la troisième partie est consacrée à la sélection de variables en régression PLS et, par extension, à la sélection de blocs en régression PLS multi-tableaux.

6.1 Prétraitement des données

Avant toute analyse, se pose la question du prétraitement des données. Habituellement, les tableaux sont centrés en colonne par rapport à la moyenne. Ainsi, les produits scalaires sont interprétés en terme de covariance. Lorsque les données ne sont pas mesurées sur une même échelle, il est d'usage de les réduire en colonne, *i.e.* diviser chaque colonne par son écart-type. En analyse cubique, les types de centrages et de réductions ne sont pas aussi évidents. Ils peuvent être effectués par colonne ou par tranche.

En analyse multi-tableau, le prétraitement se fait en colonne. De plus, si les tableaux ont un nombre de variables très différent les uns des autres, il est préconisé de les pondérer pour limiter une influence disproportionnée de celui ayant le plus grand nombre de variables. On peut, par exemple, diviser chaque tableau par la racine carrée de son inertie.

En chimiométrie, la spectrographie proche infra-rouge est très utilisée, car facile à mettre en oeuvre et de faible coût, pour expliquer des mesures chimiques ou de l'activité biologique, Y . Chaque échantillon est passé au spectromètre de réflexion. Celui-ci illumine chaque

échantillon par une source de radiations monochromatiques couvrant la région comprise entre 400nm et 2500nm, le plus souvent. Une partie de ces rayons est absorbée par l'échantillon, une autre partie est transmise à travers lui, et une dernière partie est réfléchiée. La lumière réfléchiée est «mesurée» en unité de réflectance tous les 2nm, en général. Pour chaque échantillon un spectre de 1050 points (longueurs d'ondes) est obtenu. Ces spectres composent un tableau X ayant 1050 colonnes qui correspondent à la description des 1050 longueurs d'ondes du spectre proche infra-rouge (NIRS) compris entre 400 nanomètres et 2500 nanomètres, mesuré tous les 2 nanomètres. Une ligne de ce tableau représente le spectre d'une observation. En général le tableau n'est pas réduit en colonne car cela supprimerait les variations que l'on cherche justement à identifier, mais uniquement centré. Cependant, ce tableau contient énormément de bruit qu'il faut supprimer pour espérer avoir des modèles avec un bon pouvoir de prédiction. Il existe des méthodes, dites de filtrage, dont le but est d'enlever une partie de cette variation systématique.

Récemment, les chimiométriciens ont commencé à développer des méthodes de filtrage du NIRS basée sur la régression PLS. Le but de ces méthodes étant de supprimer le bruit, l'information comprise dans le NIRS qui n'explique par les variables réponses Y . La méthode OSC, «Orthogonal Signal Correction», (WOLD, ANTTI, LINDGREN & ÖHMAN, 1998) a été la première à paraître. Nous détaillons ici son algorithme :

- 1 Eventuellement centrer et réduire les données.
- 2 t est la première composante principale de X
- 3 $t^* = P_Y^\perp t$, t est orthogonalisée à Y
- 4 $t^{**} = XX^{-1}t^*$ ou $t = TT^{-1}t^*$ où T contient les premières composantes issues de la régression PLS de X sur t^* .
- 5 Si $\|t^{**} - t\|/\|t^{**}\| < 10^{-6}$ on continue, sinon on recommence en **3**
- 6 $X_{osc} = X - P_t^\perp X$

Les composantes OSC suivantes s'obtiennent en utilisant l'algorithme à partir de la matrice filtrée X_{osc} .

Les auteurs soulignent qu'il y a plusieurs solutions pour la composante OSC lorsque $l < J$. SVENSSON, KOURTI & MACGREGOR (2002) le montrent numériquement, en prenant divers point de départ pour t à la première étape. En général, la composante finale t n'est pas orthogonale à Y . Et lorsque $l - 1 > J$, (WESTERHUIS, DE JONG & SMILDE, 2001), t n'appartient pas non plus à l'espace engendré par X . Ce qui signifie que de l'information qui peut être importante pour prédire Y peut être ôtée de X .

Dans le même numéro de revue, SJÖBLOM, SVENSSON, JOSEFSON, KULLBERG & WOLD (1998) annoncent qu'ils appliquent la méthode OSC de (WOLD et al., 1998) mais l'algorithme qu'ils présentent n'est pas le même. Une étape supplémentaire a été rajoutée avant l'étape **4** : **3bis** $\omega = X't^*/\|X't^*\|$, $t = X\omega$. Et les étapes **2** – **3** sont répétées jusqu'à convergence. Ils utilisent 15 étapes PLS pour l'étape **4**, et l'étape **5** est supprimée. Au départ, la composante t explique la plus grande part de variance possible de X , puis devient orthogonale à Y , pour à la fin ne plus l'être.

L'année suivante, ANDERSSON (1999) propose la méthode de filtrage DO, «Direct Orthogonalization». Il dit de cet algorithme qu'il assure que l'information de X qui est corrélée parfaitement avec Y n'est pas supprimée pendant le prétraitement. Au lieu d'orthogonaliser une composante à Y comme c'est le cas dans l'OSC, il orthogonalise la matrice X elle-même. La (les) composante(s) résultant de cet algorithme ne décrivent pas l'information de X et ne

sont pas orthogonales aux réponses Y . Cependant, elles servent à filtrer le spectre.

FEARN (2000) définit clairement le problème cherché en OSC. Il maximise $\omega'X'X\omega$ sous la contrainte d'orthogonalité $\omega'X'Y = 0$ et de norme unité $\omega'\omega = 1$. Cette méthode présente l'avantage de ne pas être itérative, et réduit donc le temps de calcul des composantes OSC.

WESTERHUIS et al. (2001) font une revue des algorithmes, agrémentés de commentaires techniques, des méthodes parues avant 2001, avant de proposer à leur tour un algorithme nommé DOSC, «Direct Orthogonal Signal Correction».

La dernière méthode en date est l'O-PLS, «orthogonal projection to latent structure», de TRYGG & WOLD (2002), amélioré par la suite en l'O2-PLS (TRYGG, 2002). O-PLS peut être considérée comme une méthode de filtrage ou comme une méthode d'analyse à part entière. Elle est basée sur le rajout de plusieurs étapes dans l'algorithme PLS. Le filtrage du tableau X se fait à l'intérieur d'une étape de la régression PLS. L'avantage est que le calcul des composantes OSC (à chaque étape PLS) n'est pas itératif comme c'est le cas en OSC. Les composantes OSC sont orthogonales à Y et expliquent aussi une part de la variation de X . O-PLS permet aussi une interprétation de la variation de X qui est orthogonale à Y , en faisant une ACP sur l'information de X que l'on vient de lui enlever. La méthode O2-PLS permet de chercher l'information de X orthogonale à Y , l'information de Y orthogonale à X , l'information conjointe et donne des modèles prédictifs dans les deux sens.

SVENSSON et al. (2002) donnent une rapide récapitulation des méthodes de WOLD et al. (1998), SJÖBLOM et al. (1998), FEARN (2000), ANDERSSON (1999) et TRYGG & WOLD (2002), puis font une comparaison numérique de ces méthodes. Il apparaît alors clairement que les méthodes DO et OSC de SJÖBLOM et al. (1998) ne fournissent pas des composantes orthogonales aux réponses. Donc, il se peut que l'on enlève de X de l'information nécessaire à la prédiction de Y . De plus, il apparaît sur l'exemple traité, que la méthode DO procède à un filtrage très «violent». Après une composante seulement, plus de 90% de la variance de X lui a été otée. Les méthodes O-PLS, DO et celle de Fearn, qui enlèvent l'information orthogonale à chaque étape PLS, permettent de réduire le nombre de composantes dans le modèle PLS à chaque fois qu'une composante OSC supplémentaire est utilisée. De plus, le minimum de la racine carrée de l'erreur de prédiction moyenne n'est pas changé pour le modèle PLS retenu quelque soit le nombre de composantes OSC utilisées (SVENSSON et al., 2002). TRYGG & WOLD (2002) disent ainsi, que l'on peut toujours obtenir un modèle PLS construit à partir d'une seule composante sans dégrader l'erreur de prédiction, si tant est que l'on utilise le nombre suffisant de composantes OSC.

De plus, VERRON (2002) a montré que le modèle final pour Y peut s'écrire comme une combinaison linéaires des variables initiales (avant correction), ce qui est très utile pour faire de la prédiction à partir d'un nouvel échantillon.

En régression multi-tableau mettant en jeu une matrice NIRS parmi les tableaux de prédicteurs, on peut se demander comment effectuer le filtrage. La méthode la plus directe est d'effectuer un des prétraitements décrit ci-dessus en utilisant le tableau X_k contenant les données NIRS et le tableau conjoint des réponses.

Si l'on découpe un tableau NIRS en plusieurs blocs X_k , afin d'y appliquer l'ACIMOG-PLS (chap. 5.6), VERRON (2002) a développé la méthode O-ACIMOG, par analogie à O-PLS. Il s'agit d'une méthode de prétraitement des X_k de type multi-bloc basé sur O-PLS, ayant pour but de simplifier les modèles obtenus par l'application de l'ACIMOG-PLS. L'idée sous-jacente est d'enlever à chaque bloc X_k la variation qui est orthogonale aux Y_n , c'est-à-dire le «bruit»

qui peut gêner l'obtention de modèles stables. Pour cela, l'étape a de l'algorithme ACIMOG-PLS est modifiée de façon à calculer des composantes «OSC», $t_{k,a}^{osc} = X_k^{(a-1)} a_{k,a}^{osc}$, où $X_k^{(a-1)}$ représente la déflation du tableau X_k sur les super-composantes «OSC» précédentes.

A l'étape $a = 1^*$, O-ACIMOG procède ainsi :

après le calcul d'une composante t_k par l'ACIMOG-PLS, des vecteurs de poids a_k^{osc} sont recherchés de façon à maximiser $(cov(t_k, X_k Q_k^x a_k^{osc}))^2$ sous les contraintes $\|a_k^{osc}\|_{Q_k^x}^2 = 1$ et $y_n' D X_k Q_k^x a_k^{osc} = 0$ pour tout n . Dans cette même étape $a = 1$, on peut calculer d'autres composantes $t_k^{osc} = X_k Q_k^x a_k^{osc}$ en maximisant ce même critère en prenant $X_k = X_k - P_{t_{osc}} X_k$. Lorsqu'on passe à l'étape suivante $a = 2$, on utilise alors l'ACIMOG-PLS sur les tableaux $X_k = X_k - P_{T_{osc}} X_k - P_t X_k$ où t est la super-composante de l'ACIMOG-PLS et T_{osc} est la matrice constituée des composantes «OSC» que l'on vient de calculer. Notons, qu'ici nous avons décrit l'O-ACIMOG en utilisant des super-composantes, c'est-à-dire qu'il s'agit de l'O-ACIMOG pour l'ACIMOG-PLS2 ou 3. Pour la version O-ACIMOG ayant pour but l'application de l'ACIMOG-PLS1, il faut remplacer dans ce qui précède t_{osc} par $[t_1^{osc} \dots t_K^{osc}]$. Cependant, dans une telle situation, il reste encore à résoudre le problème du choix de découpage des spectres NIRS X en bloc X_k . Pour l'instant, cela doit être fait par des connaissances *a priori* des spectres.

On peut penser que le développement de méthodes multi-tableaux pour le filtrage du NIRS va être de plus en plus dense dans les années qui viennent. La spectrométrie permettant de mesurer quantité de variables à faible coût, il est nécessaire de développer des méthodes permettant l'obtention de modèles stables et exploitables. L'utilisation de méthodes de régression multi-tableaux en combinaison avec un filtrage adapté devrait permettre cela.

6.2 Validation des modèles et choix du nombre de composantes

Dans les méthodes «component-wise», il faut à un moment ou un autre s'arrêter de générer des composantes. Pour cela, il faut être capable de savoir si le modèle auquel on ajoute une étape a , modélise une part importante de l'information contenue dans les données et s'il a un bon pouvoir de prédiction. Selon la situation et le désir de l'expérimentateur, l'une ou l'autre des propriétés du modèle sera favorisée.

Les démarches utilisées en régression PLS usuelle peuvent aisément être généralisées au cas de multi-tableaux.

6.2.1 Critère d'ajustement

Si l'on cherche à avoir des modèles bien ajustés aux données mais non nécessairement prédictifs, le coefficient de corrélation multiple (variance expliquée) est un outil pour choisir le nombre A de composantes à retenir, que ce soit en régression PLS ou régression multi-tableau. En ACIMOG-PLS et REMUB, tout comme en régression PLS, la variance du modèle

* Nous prenons l'étape $a = 1$ pour simplifier. Ainsi, l'indice a de objets est omis. Les résultats restent valables pour les étapes a d'ordres supérieurs en utilisant pour X_k et Y_n les tableaux déflatés de façon appropriée.

se décompose de façon additive car les modèles partiels sont orthogonaux (chap. 5.6) :

$$\begin{aligned}
 R_{Y_n}^2(a) &= \frac{\|\hat{Y}_n^a\|_D^2}{\|Y_n\|_D^2} \\
 &= \sum_{\ell=1}^a \frac{\|P_{\mathcal{E}_\ell} Y_n\|_D^2}{\|Y_n\|_D^2} \\
 &= R_{Y_n}^2(a-1) + \frac{\|P_{\mathcal{E}_a}\|_D^2}{\|Y_n\|_D^2}.
 \end{aligned}$$

Nous rappelons que \mathcal{E}_a est égal à la super-composante t_a dans le cas de REMUB, ACIMOG-PLS2 et 3, et à $[t_{1,a} \dots t_{\kappa,a}]$ dans l'ACIMOG-PLS1.

Comme les tableaux sont D -centrés, les D -normes au carré ci-dessus sont des D -variances. Le coefficient de corrélation multiple croît avec a . Lorsqu'on n'a qu'un seul tableau réponse (comme en régression PLS), on peut choisir le nombre d'étapes A tel que $R_{Y_1}^2(A+1) - R_{Y_1}^2(A)$ ne soit plus très grand. Le choix est finalement assez subjectif. Lorsqu'on est en présence de plusieurs tableaux réponses, on cherche généralement à avoir le même nombre de composantes dans les N modèles. On préférera choisir une valeur de A telle que la «majorité» des modèles soient bien expliqués.

LAZRAQ & CLÉROUX (2001) ont établi un test pour tester la signification des composantes successives. Il faut ici faire l'hypothèse que $(Y \ X)$ suit une loi multinormale. Le test est basé sur le coefficient de redondance, *i.e* le pourcentage de variance expliquée. A l'étape a , une observation de la statistique de test est donnée par $V_{obs} = RI(Y, t_a)/(1 - RI(Y, t_a))$ où $RI(Y, t_a) = \frac{tr(Y' t_a (t_a' t_a)^{-1} t_a' Y)}{tr(Y' Y)}$. Les hypothèses du test sont $H_0 : RI(Y, t_a) = 0$ contre $H_1 : RI(Y, t_a) > 0$. L'hypothèse nulle H_0 est rejetée au niveau α si $V_{obs} > c_\alpha$. Le seuil c_α est déterminé algorithmiquement en utilisant la loi sous H_0 de $RI(Y, t_a) - c_\alpha(1 - RI(Y, t_a))$, qui est une combinaison linéaire de lois du χ^2 . Ce test permet ainsi de savoir si la variance de Y expliquée par la composante t_a est significativement non nulle. Le nombre A d'étapes est tel que la variance expliquée par la composante t_{A+1} n'est pas significativement non nulle.

6.2.2 Validation Croisée

La validation croisée est une méthode de validation interne de modèle. C'est un critère permettant de d'évaluer la consistance interne des données. Elle a été initialement proposée par (STONE, 1974) pour l'ACP. L'idée générale est de constituer successivement des groupes d'individus tests à partir des données et de tester la capacité du modèle à les prédire. Elle a ensuite été grandement adaptée au cas de la régression PLS usuelle. Il y a plusieurs manière de procéder.

Dans ce qui suit, nos notations correspondent au cas où les groupes d'individus tests ne sont composés que d'un seul individu i choisi aléatoirement («leave-one-out cross-validation»). Il est facilement possible de les adapter au cas de groupes aléatoires d'individus.

Dans la littérature anglophone, on trouve les appellations «full cross-validation» et «step-wise cross-validation» (WAKELING & MORRIS, 1993) en ce qui concerne la validation croisée en régression PLS usuelle. La différence ne nous paraît pas très claire quant aux critères utilisés. Cependant nous avons retenu que le premier type de validation croisée se réalise une fois un ensemble de modèles PLS déterminés. Alors, des *PRESS* sont calculés (Prédiction Residual Sum et Squares) et permettent le choix du A optimal. Le second type de validation

croisée se fait à l'intérieur de la régression PLS, à chaque étape, et un critère permet de dire si la composante de l'étape courante est significative ou non. Si elle ne l'est pas l'algorithme s'arrête. C'est de cette manière que procède le logiciel SIMCA-P®.

En régression PLS usuelle, sont utilisés les PRESS par variable et le PRESS total pour le tableau Y des réponses. En régression multi-tableaux, nous avons les PRESS par variable, par tableau et le PRESS total. Dans ce qui suit, nous nous plaçons dans la situation de régression multi-tableaux. Pour le calcul des *PRESS* dans la régression PLS usuelle, il suffit de ne considérer qu'un seul tableau de réponses.

Pour une étape a , posons $E_{n,a}^y(i) = Y_n(i) - X(i)\widehat{\beta}_n^a(-i)$ où $\widehat{\beta}_n^a(-i)$ est la matrice des coefficients du modèle obtenus par une des ACIMOG-PLS ou REMUB avec les données auxquelles l'observation i a été préalablement ôtée. $E_{n,a}^y$ est une matrice de taille $1 \times M_n$. $X(i)$ et $Y_n(i)$ sont respectivement les $i^{\text{ièmes}}$ lignes de X et Y_n . Le *PRESS* pour la variable m du groupe n est défini par

$$PRESS_n^m(a) = \|E_{n,a}^y(\cdot, m)\|_D^2,$$

où $E_{n,a}^y(\cdot, m)$ est la colonne m de $E_{n,a}^y$. Si D est la métrique uniforme, comme c'est généralement le cas, $PRESS_n^m(a) = \frac{1}{1} \sum_{i=1}^1 E_{n,a}^y(i, m)^2$. Le *PRESS* pour le groupe n est défini par

$$PRESS_n(a) = \sum_{m=1}^{M_n} PRESS_n^m(a) = \|E_{n,a}^y\|_D^2$$

qui est aussi égal à $\sum_{i=1}^1 \frac{1}{1} \|E_{n,a}^y(i)\|^2$ si D est la métrique uniforme. Le *PRESS* total, ou global, est donné par

$$PRESS(a) = \sum_{n=1}^N PRESS_n(a) = \left\| \begin{bmatrix} E_{1,a}^y & \dots & E_{N,a}^y \end{bmatrix} \right\|_D^2.$$

Plaçons nous maintenant dans le cas des méthodes ACIMOG (chap. 3.4). Les modèles reconstituant les tableaux X_k et Y_n , selon la méthode, s'écrivent de la forme $X\widehat{\beta}_k^a$ (ou $X_k\widehat{\beta}_k^a$) et $Y\widehat{\beta}_n^a$ (ou $Y_n\widehat{\beta}_n^a$). Nous définissons comme précédemment $E_{k,a}^x$ et $E_{n,a}^y$ et les *PRESS* par variable

$$PRESS_k^j(a) \text{ et } PRESS_n^m(a),$$

par tableau :

$$PRESS_k(a) = \sum_{j=1}^{J_k} PRESS_k^j(a) \text{ et } PRESS_n(a) = \sum_{m=1}^{M_n} PRESS_n^m(a)$$

et par ensemble (X ou Y):

$$PRESS_X(a) = \sum_{k=1}^K PRESS_k(a) \text{ et } PRESS_Y(a) = \sum_{n=1}^N PRESS_n(a).$$

Le *PRESS* global est alors donné par la somme des *PRESS* : $PRESS(a) = PRESS_X(a) + PRESS_Y(a)$.

Il y a plusieurs manières de choisir le nombre optimal A d'étapes. On peut choisir A tel que $PRESS(A)$ soit minimal globalement ou soit égal au premier minimum. Plusieurs auteurs ont proposés des critères de choix basés sur la pente présente entre deux valeurs de $PRESS$ successives. WAKELING & MORRIS (1993) en citent quelques-uns tirés de la littérature. Citons celui proposé par EASTMENT & KRZANOWSKI (1982) et repris par OSTEN (1988). Ces derniers construisent une statistique de test F_a . Si F_a dépasse le quantile d'ordre 0.95 d'une loi de Fisher à (d_1, d_2) degrés de liberté, alors la composante d'ordre a est considérée comme significative :

$$F_a = \left(\frac{(PRESS_{a-1} - PRESS_a)}{d_1} \right) / \frac{PRESS_a}{d_2},$$

avec $d_1 = J$ le nombre de variables explicatives,
 $d_2 = |J - aJ|$.

Dans les applications de la Partie III, le nombre A que nous avons retenu est celui qui minimise globalement le « $PRESS$ total».

Une méthode de validation très répandue en régression PLS est l'utilisation du coefficient de corrélation multiple «cross-validé» (Cross-validated R^2), noté Q^2 . Cette méthode est utilisée dans SIMCA-P®. Ce coefficient est aisément généralisable au cas de la régression multi-tableaux. On peut ainsi définir le Q^2 par variable, $Q_{n,m}^2$, le Q^2 par tableau, Q_n^2 , le Q^2 total, de la même manière que les $PRESS$:

$$\begin{aligned} Q_{n,m}^2(a) &= 1 - PRESS_n(a)^m / RSS_n^m(a-1), \\ Q_n^2(a) &= 1 - PRESS_n(a) / RSS_n(a-1), \\ Q^2(a) &= 1 - PRESS(a) / RSS(a-1), \end{aligned}$$

où RSS désigne la somme des carrés des résidus du modèle :

$$\begin{aligned} RSS_n^m(a) &= \frac{1}{I} \sum_{i=1}^I (Y_n(i,m) - \hat{Y}_n^a(i,m))^2 \\ &= \|Y_n(,m) - \hat{Y}_n^a(,m)\|_D^2 \\ RSS_n(a) &= \sum_{m=1}^{M_n} RSS_n^m(a) \\ &= \|Y_n - \hat{Y}_n^a\|_D^2 \\ RSS(a) &= \sum_{n=1}^N RSS_n(a) \\ &= \|Y - \hat{Y}^a\|_D^2. \end{aligned}$$

En régression PLS, une composante t_a est significative si $Q^2(a) \geq \alpha$. α est une valeur arbitraire. La dernière valeur proposée, sur l'expérience personnelle de S. Wold, est 0.0975. On peut dire que t_a est significative si au moins un des $Q_{1,m}^2$ est supérieur à $\alpha = 0.0975$.

En régression multi-tableaux, on peut appliquer les mêmes règles. Le Q^2 cumulé est parfois utilisé. Il permet de mesurer l'apport global de l'ensemble des a premières composantes PLS : $Q_{cum,n}^2(a) = 1 - \prod_{i=1}^a (1 - Q_n^2(a))$. En régression PLS, si $Q_{cum,1}^2(a) \leq 0.5$, le tableau Y ($N = 1$) est bien modélisé par les a premières composantes. En régression multi-tableaux,

nous pouvons définir le Q^2 cumulé total $Q_{cum}^2(a) = 1 - \prod_{l=1}^a (1 - Q^2(a))$ et appliquer la règle précédente pour savoir si l'ensemble des tableaux est correctement modélisé.

En analyse cubique, les trois modes sont pris en compte dans la modélisation. Aussi, faire la validation croisée par rapport aux individus n'est pas plus justifiée que de la faire par rapport à un autre mode. LOUWERSE, SMILDE & KIERS (1999) proposent deux manières de réaliser la validation croisée afin d'obtenir le meilleur nombre de facteurs (ou composantes) pour les modèles TUCKER3. Leur première méthode est appelée validation croisée EM-Tucker3, pour «expectation maximization approach». Cette technique consiste à enlever tour à tour, des ensembles de valeurs du cube \underline{X} initial, jusqu'à ce tous les éléments aient été enlevés une fois, et à estimer chaque groupe enlevé, en les considérant comme des valeurs manquantes dans \underline{X} grâce à un algorithme aux moindres carrés alternés prenant en compte les valeurs manquantes. Cela permet de calculer une valeur du PRESS dépendant du nombre de facteurs retenus pour chaque mode.

La seconde méthode est une généralisation de la validation croisée de (EASTMENT & KRZANOWSKI, 1982) pour l'ACP, où ils enlèvent une ligne et une colonne au lieu de simplement enlever une ligne. Ainsi, LOUWERSE et al. (1999) proposent d'utiliser comme échantillon test à la fois une tranche frontale, une tranche horizontale et une tranche verticale du cube \underline{X} , et de les prédire à partir des éléments restants. En répétant cela plusieurs fois jusqu'à ce que toutes les tranches aient été prédites une fois, une valeur du PRESS est obtenue, dépendant du nombre de facteurs dans le modèle. Cette méthode est appelée validation croisée LBO-Tucker3. LBO est l'abréviation de «Leave-bar-out». Dans cette méthode, la valeur de la prédiction est un peu plus compliquée à obtenir. Le lecteur se référera à la publication (LOUWERSE et al., 1999). Une statistique W est calculée, à la manière de la statistique F_a décrite précédemment, pour choisir le nombre de composantes. Si W est significativement plus grand que 1, alors une composante supplémentaire est incluse dans le modèle. On commence par une composante par mode, puis on ajoute tour à tour une composante dans chacun des modes. On peut aussi retenir le nombre de composantes pour lequel le PRESS est minimum.

Un commentaire s'impose sur le prétraitement des données. Centrer et éventuellement réduire les données après avoir enlevé l'échantillon test, ne donnera pas le même résultat que si cela n'est pas fait. D'après LOUWERSE et al. (1999), il est préférable de prétraiter les données restantes après avoir sélectionné un groupe test pour une question d'indépendance entre l'échantillon test et l'échantillon qui prédit. Cette méthode est sans doute la plus courante et c'est celle que nous avons appliquée dans nos programmations.

6.3 Sélection

6.3.1 Sélection de variables en régression PLS

En régression multiple, se pose souvent la question de sélectionner les variables importantes. En régression PLS, les méthodes peuvent se classer dans deux grandes catégories :

- où les variables sont choisies une fois le modèle établi
- où les variables sont choisies à chaque étape, en modifiant les vecteurs de poids.

Toutes les méthodes que nous énonçons ici ont été développées par leurs auteurs pour le cas d'une seule variable réponse y et une matrice de prédicteurs $X = [x_1 \dots x_j]$.

GAUCHI & CHAGNON (2001) donnent une revue des méthodes de sélection de variables pré-

sentes dans le logiciel SAS[®]. Ces méthodes consistent à établir un modèle PLS avec le nombre optimal de composantes selon la validation croisée, sur l'ensemble des prédicteurs ou sur un sous-ensemble présélectionné. Ensuite, des critères basés sur les coefficients du modèle ou sur des corrélations sont utilisés pour supprimer des variables et le modèle PLS est recalculé. Les auteurs préconisent l'utilisation de la méthode BQ, *backward* – Q_{cum}^2 . Le modèle PLS avec tous les prédicteurs est premièrement calculé, où le nombre de composantes est estimé par validation croisée. Le coefficient de corrélation «cross-validé» cumulé est déterminé. Puis, la variable présentant le plus petit coefficient en valeur absolue dans le modèle est éliminée, et un nouveau modèle et un nouveau Q_{cum}^2 sont estimés. Cette opération est répétée jusqu'à ce que toutes les variables aient été ôtées. Les variables retenues, et donc le modèle, correspondent au plus petit ensemble de variables dont le Q_{cum}^2 est maximum.

FRANK (1987) a été le premier à proposer une méthode de sélection de variables pour la régression PLS agissant à chaque étape a . Cette méthode, appelée Intermediate Least Squares Regression (ILS), permet de retrouver la régression PLS, lorsqu'aucune des variables n'est supprimée, et la «régression par étape» (SR) lorsqu'une seule variable est gardée par étape, d'où son appellation. Elle consiste à poser certains poids du vecteur ω_a à zéro de la manière suivante : à chaque étape a , les valeurs absolues $\omega_a[j]$ sont classées par ordre croissant. Puis, les α plus petites valeurs sont annulées. α est choisi avant de commencer la régression et il peut prendre toutes les valeurs de 0 à $J - 1$. Le vecteur des poids ω_a est ensuite renormé. α se choisit par validation croisée. Cela signifie qu'il faut essayer toutes les valeurs possibles de α avec les valeurs possibles de A . Cette procédure est assez lourde en calculs. La conclusion de la publication est que dans certains cas, ILS donne de meilleurs modèles en terme de pouvoir prédictif que la PLS.

Plusieurs techniques de sélection basées sur l'annulation de certains poids ont par la suite été développées. LINDGREN, GELADI, RÄNNAR & WOLD (1994) modifient la procédure de choix de α , qui devient visuelle. Ils proposent d'enlever les variables dont les poids $\omega_a[j]$ sont les plus petits, (ou les plus grands) en testant quel est le seuil par validation-croisée avec un pas de 0.01 unité. Puis, à l'aide d'un graphique on choisit la valeur de α et par conséquent le modèle.

TOMINAGA & FUJIWARA (1997) et FORINA, CASOLINO & PIZZARO MILLAN (1999) présentent des algorithmes modifiés, respectivement appelés Prediction-Weighted PLS (PWPLS) et Iterative Predictor Weighting PLS (IPW PLS), dans lesquels les poids ω_a sont modifiés à chaque étape pour tenir compte de l'importance des variables et construire des modèles plus prédictifs. Cela revient en fait à sélectionner les prédicteurs les plus importants à chaque étape, et à les pondérer plus fortement. Dans PWPLS, l'importance des variables est jugée par des corrélations tandis que dans IPWPLS elle est jugée par la valeur des coefficients du modèle à l'étape a (rapport en valeur absolue du coefficient de la variable x_j sur la somme des valeurs absolues de tous les coefficients).

La dernière méthode de sélection dont nous allons parler est un critère permettant de juger l'importance d'un prédicteur dans la construction du modèle de la régression PLS avec plus d'une réponse. Ce critère, Variable Importance in the Prediction (VIP) est utilisé par le logiciel SIMCA-P[®]. On peut trouver sa description dans l'ouvrage de (TENENHAUS, 1998).

Le *VIP* de la variable j du tableau X à l'étape a est donné par :

$$VIP_j(a) = \sqrt{J \frac{In(Y)}{In(P_{t_1, \dots, t_a} Y)} \sum_{l=1}^a \frac{In(P_{t_l} Y)}{In(Y)} \omega_a^2[j]}.$$

$\frac{In(P_{t_1, \dots, t_a} Y)}{In(Y)}$ est la variance de Y expliquée par les a premières composantes de la régression PLS, encore appelée redondance. On a $\sum_{j=1}^J VIP_j^2(a) = J$: à l'étape a , les variables ayant un VIP^2 supérieur à 1 sont les plus importantes dans la construction de Y .

6.3.2 Sélection de blocs en ACIMOG-PLS et REMUB

Dans le cas de la régression multi-blocs, se pose naturellement la question de développer des critères de sélection de blocs influents. En effet, si un groupe de variables n'est pas important dans la modélisation, il n'est plus nécessaire pour l'expérimentateur de les mesurer, ce qui peut ainsi permettre des économies de temps et d'argent.

Ainsi, une méthode peut être d'essayer de modéliser un ou plusieurs groupes de variables réponses Y_n par plusieurs groupes de variables explicatives X_k . Si certains de ces groupes apparaissent comme peu ou pas influents, alors on peut refaire l'analyse multi-blocs en les supprimant.

6.3.2.1 Critère basé sur les coefficients du modèle

Le premier critère de sélection de blocs auquel nous avons pensé est basé sur la valeur absolue des coefficients du modèle. En régression PLS, nous savons que les coefficients sont plus ou moins bien interprétables pour les modèles de rang faible, contrairement à ceux de la régression multiple ordinaire. C'est de là que nous est venue l'idée.

Notre modèle ne peut pas s'exprimer comme une combinaison linéaire des tableaux X_k . Cela dit, on peut tout de même chercher un moyen raisonnable de quantifier l'importance de chaque X_k .

Le modèle à l'étape A, \hat{Y}_n^A , peut s'écrire de la forme (*c.f.* chapitres 3.4, 5.7) :

$$\hat{Y}_n^A = \sum_{k=1}^K X_k \hat{\beta}_n^A[k], \quad (6.1)$$

où $\hat{\beta}_n^A[k]$ est le k -ième bloc de taille $J_k \times M_n$.

La norme au carré de la matrice $\hat{\beta}_n^A[k]$ peut être utilisée pour quantifier l'importance du tableau X_k dans ce modèle. En effet, cette norme au carré se décompose de la manière suivante :

$$\|\hat{\beta}_n^A\|^2 = tr \left(\hat{\beta}_n^{A'} \hat{\beta}_n^A \right) \quad (6.2)$$

$$= tr \left(\sum_{k=1}^K \hat{\beta}_n^{A'}[k] \hat{\beta}_n^A[k] \right) \quad (6.3)$$

$$= \sum_{k=1}^K tr \left(\hat{\beta}_n^{A'}[k] \hat{\beta}_n^A[k] \right) \quad (6.4)$$

$$= \sum_{k=1}^K \|\hat{\beta}_n^A[k]\|^2. \quad (6.5)$$

A chaque étape a , on peut donc calculer ce que représente la norme au carré de la matrice des coefficients correspondant au tableau X_k par rapport à la norme au carré de la matrice $\widehat{\beta}_n^a$. C'est cette quantité, en pourcentage, que représente le coefficient $B_n^a[k]$ défini ci-dessous :

$$B_n^a[k] = \frac{\|\widehat{\beta}_n^{a'}[k]\|^2}{\|\widehat{\beta}_n^a\|^2} \times 100. \quad (6.6)$$

La représentation graphique, pour un tableau Y_n , de $B_n^a[k]$ pour tout k en fonction de a permet de comparer «l'influence» des différents tableaux explicatifs X_k .

Le graphique de $\|\widehat{\beta}_n^a\|^2$ en fonction de a peut éventuellement être utilisé pour le choix du nombre A , en le choisissant tel que l'évolution de ces coefficients semble stabilisée. Lorsque le rang de X est atteint, les coefficients sont ceux de la régression multiple ordinaire : $\|\widehat{\beta}_n^a\|^2$ a tendance à être croissante en fonction de a vers la norme des coefficients de la régression multiple, bien que l'on puisse observer une décroissance à une étape.

Malheureusement, ce critère ne semble pas être efficace de manière systématique. En effet, si les variables des différents blocs prédicteurs sont de natures différentes et non réduites, étudier la valeur absolue des coefficients n'a pas de sens. De plus, à la première étape de l'ACIMOG-PLS2, ces coefficients ne dépendent ni de n ni de k si les métriques Q_k^x sont les métriques identités, qui est la situation la plus courante. En effet,

$$\begin{aligned} \|\widehat{\beta}_n^1[k]\|^2 &= \text{tr}(Y_n' D t_1 \delta_1' [k] \delta_1 [k] t_1' D Y_n) / \|t_1\|_D^4, \\ &= \text{tr}(Y_n' D t_1 a'_{k,1} Q_k^x Q_k^x a_{k,1} t_1' D Y_n) / \|t_1\|_D^4, \\ &= a'_{k,1} Q_k^x Q_k^x a_{k,1} \text{tr}(Y_n' D t_1 t_1' D Y_n) / \|t_1\|_D^4, \\ &= a'_{k,1} Q_k^{x^2} a_{k,1} \|P_{t_1} Y_n\|_D^2 / \|t_1\|_D^2. \end{aligned}$$

Si les métriques Q_k sont les métriques identités, $\|\widehat{\beta}_n^1[k]\|^2 = \|P_{t_1} Y_n\|_D^2 / \|t_1\|_D^2$, impliquant que $\|\widehat{\beta}_n^a\|^2 = K \|P_{t_1} Y_n\|_D^2 / \|t_1\|_D^2$ quelque soit k et $B_n^1[k] = 100/K$, quelque soient n et k . Autrement dit, tous les tableaux X_k modélisent avec la même importance tous les tableaux réponses Y_n . Notons cependant, que ce résultat est à corroborer avec celui de la construction des super-composantes t_a dans lesquelles les composantes t_k sont prises entièrement et non pondérées. Si Q_k^x n'est pas la métrique identité, $B_n^1[k]$ ne dépend pas de n . Cela signifie qu'un tableau X_k influe de la même manière sur tous les Y_n . Pour les étapes suivantes, cela n'est plus vrai. Des calculs similaires montrent que $\|\widehat{\beta}_n^1[k]\|^2 = \frac{\mu_{k,1}^2}{\|t_1\|_D^2} a'_{k,1} Q_k^{x^2} a_{k,1} \|P_{t_1} Y_n\|_D^2$ dans l'ACIMOG-PLS3. Si Q_k^x est la métrique identité, $B_n^1[k] = 100\mu_{k,1}^2 / \sum_{k=1}^K \mu_{k,1}^2$, et ne dépend pas de n .

Malgré ceci, dans le cas d'une unique réponse y , et avec l'utilisation de l'ACIMOG-PLS1, nous avons constaté une similitude dans le classement des tableaux influents donné par ce coefficient, et celui donné par les corrélations partielles $r(y, t_k | t_{k'} k' \neq k)$.

Avec l'ACIMOG-PLS3, il existe une écriture complémentaire intéressante des modèles \widehat{Y}_n^A .

En effet, en partant de l'écriture du modèle :

$$\begin{aligned}\widehat{Y}_n^A &= \sum_{a=1}^A P_{t_a} Y_n \\ &= \sum_{a=1}^A t_a (t'_a D t_a)^{-1} t'_a D Y_n \\ &= \sum_{a=1}^A \sum_{k=1}^K \mu_{k,a} t_{k,a} (t'_a D t_a)^{-1} t'_a D Y_n.\end{aligned}\quad (6.7)$$

$$= \sum_{k=1}^K [t_{k,1} | \dots | t_{k,A}] \Theta_k^n, \quad (6.8)$$

où

$$\Theta_k^n = \begin{bmatrix} \mu_{k,1} (t'_1 D t_1)^{-1} t'_1 D Y_n \\ \vdots \\ \mu_{k,A} (t'_A D t_A)^{-1} t'_A D Y_n \end{bmatrix}.$$

L'expression (6.7) permet, pour chaque modèle partiel de rang 1, de classer par ordre d'influence les tableaux X_k , en comparant les normes des vecteurs $\mu_{k,a} t_{k,a} (t'_a D t_a)^{-1} t'_a D Y_n$, pour a fixé.

En identifiant un tableau X_k au tableau $[t_{k,1} | \dots | t_{k,A}]$, (6.8), on peut utiliser la norme au carré de la matrice Θ_k^n comme indicateur d'influence du tableau X_k sur le modèle. Cette norme, $\|\Theta_k^n\|^2$ se décompose en la somme des normes au carré de ses lignes :

$$\begin{aligned}\|\Theta_k^n\|^2 &= tr \left(\Theta_k^{n'} \Theta_k^n \right) \\ &= \sum_{a=1}^A tr \left[(\mu_k Y_n' D t_a (t'_a D t_a)^{-1} t'_{k,a}) (\mu_{k,a} t_{k,a} (t'_a D t_a)^{-1} t'_a D Y_n) \right] \\ &= \sum_{a=1}^A tr \left[(\mu_{k,a} t_{k,a} (t'_a D t_a)^{-1} t'_a D Y_n) (\mu_k Y_n' D t_a (t'_a D t_a)^{-1} t'_{k,a}) \right] \\ &= \sum_{a=1}^A \|\mu_{k,a} t_{k,a} (t'_a D t_a)^{-1} t'_a D Y_n\|^2.\end{aligned}$$

Lorsque $N = 1$ et $M_n = 1$, le modèle \widehat{y}^a de l'ACIMOG-PLS1 peut s'écrire comme une somme pondérée des composantes partielles de tous ordres :

$$\widehat{y}^a = \sum_{a=1}^A \sum_{k=1}^K \theta_k^a t_{k,a}. \quad (6.9)$$

Comme $\widehat{y}^a = \sum_{a=1}^A T_a (T'_a D T_a)^{-1} T'_a D y$, le résultat est immédiat en notant $\theta_{k,a}$ la $k^{\text{ième}}$ ligne de $(T'_a D T_a)^{-1} T'_a D y$. $\theta_{k,a}$ peut servir d'indicateur pour mesurer l'importance d'un bloc X_k à l'étape a . Nous avons remarqué par la pratique qu'à la première étape le classement donné par $\theta_{k,a}$ était similaire à celui donné par $B_y^1[k]$ (voir chap. 3 de la Partie III pour une application).

6.3.2.2 Critère : Bloc Importance in Prediction

Le critère précédant ne semblant pas suffisant et peu fiable dans certains cas, nous avons développé pour l'ACIMOG-PLS1, l'ACIMO-PLS3 et REMUB (§5.6 et 5.7), un critère nommé BIP pour Block Importance in the Prediction, dont le nom fait référence au critère VIP (§6.3.1) utilisée en PLS, ainsi qu'un critère BIP basé sur le coefficient RV (ROBERT & ESCOUFIER, 1976), que nous noterons BIPrv, pour l'ACIMOG-PLS1.

Définition 6.1 Les coefficients BIP et BIPrv sont définis de la manière suivante :

$$BIP_k^n(A) = \sqrt{\mathbb{K} \sum_{a=1}^A \frac{\mu_{k,a} \text{In}(P_{t_a} Y_n)}{\sum_{k'} \mu_{k',a} \text{In}(P_{t_1, \dots, t_A} Y_n)}}, \quad \text{pour l'ACIMOG-PLS3}, \quad (6.10)$$

$$BIP_k^n(A) = \sqrt{\mathbb{K} \sum_{a=1}^A \frac{\|t_{k,a}\|_D^2 \text{In}(P_{T_a} Y_n)}{\|T_a\|_D^2 \text{In}(P_{T_1, \dots, T_A} Y_n)}}, \quad \text{pour l'ACIMOG-PLS1}, \quad (6.11)$$

$$BIP_k^n(A) = \sqrt{\mathbb{K} \sum_{a=1}^A \alpha_{k,a}^2 \frac{\text{In}(P_{t_a} Y_n)}{\text{In}(P_{t_1, \dots, t_A} Y_n)}}, \quad \text{pour REMUB}, \quad (6.12)$$

$$BIPrv_k^n(A) = \mathbb{K} \sum_{a=1}^A \frac{\text{var}(t_{k,a}) \text{RV}(\widehat{Y}_n^A, t_{k,a})}{\sqrt{\sum_a \text{var} V(T_a)} \text{RV}(\widehat{Y}_n^A, T_1, \dots, T_A)}, \quad \text{pour l'ACIMOG-PLS1}. \quad (6.13)$$

Lorsque les métriques Q_n^y sont les métriques identité, alors les inerties In apparaissant dans les formules sont égales à la D -norme au carré de l'objet correspondant, car nous supposons, depuis de début ce travail, que les tableaux de données sont centrés pour D .

Quelques commentaires s'imposent. Chaque formule est constituée d'une partie mesurant la part de variation de Y_n expliquée par l'étape a par rapport au modèle total, pondérée par une quantité mesurant la contribution du tableau k dans la construction de la super-composante (ou du tableau T_a pour l'ACIMOG-PLS1) à l'étape a .

Prenons le cas de l'ACIMOG-PLS3 avec la formule (6.10). $\text{In}(P_{t_a} Y_n) / \text{In}(P_{t_1, \dots, t_A} Y_n)$ représente, en proportion, ce qui, dans la variation de Y_n , est expliqué par le modèle partiel d'ordre a dans le modèle d'ordre A . $\mu_{k,a} / \sum_k \mu_{k,a}$ mesure la contribution de la composante $t_{k,a}$ à la construction de la super-composante $t_a = \sum_{k=1}^K \mu_{k,a} t_{k,a}$ (qui sert à expliquer Y_n à l'étape a). La somme indexée par a permet de mesurer l'influence du tableau k en «cumulé». Nous notons «cumulé» entre guillemets car nous n'avons pas $BIP_k^n(A) =$

$$BIP_k^n(A-1) + \sqrt{\mathbb{K} \frac{\mu_{k,A} \text{In}(P_{t_A} Y_n)}{\sum_k \mu_{k,A} \text{In}(P_{t_1, \dots, t_A} Y_n)}}.$$

Ainsi, dans (6.11) la contribution de la composante $t_{k,a}$ à la construction de T_a (qui est l'objet qui modélise à chaque étape) est mesuré par $\|t_{k,a}\|_D^2 / \|T_a\|_D^2$ car $\|T_a\|_D^2 = \sum_{k=1}^K \|t_{k,a}\|_D^2$. Dans (6.12) c'est $\alpha_{k,a}^2$ qui est utilisé car $t_a = \sum_{k=1}^K \alpha_{k,a} t_{k,a}$ et $\sum_{k=1}^K \alpha_{k,a}^2 = 1$.

Propriété 6.1 $\sum_{k=1}^K (BIP_k^n(a))^2 = \mathbb{K}$ et $\sum_{k=1}^K BIPrv_k^n(a) = \mathbb{K}$.

Alors, si $BIP_k^n(a)$ et/ou $BIPrv_k^n(a)$ est très supérieur à 1, le tableau X_k est bien plus important que les autres dans la modélisation de Y_n .

Démonstration :

Pour l'ACIMOG-PLS 3:

$$\sum_{k=1}^K (BIP_k^n(A))^2 = K \sum_k \sum_{a=1}^A \frac{\mu_{k,a} In(P_{t_a} Y_n)}{\sum_{k'=1}^K \mu_{k',a} In(P_{t_1, \dots, t_A} Y_n)} \quad (6.14)$$

$$= K \sum_{a=1}^A \frac{In(Y_n)}{In(P_{t_1, \dots, t_A} Y_n)} \quad (6.15)$$

= K, car les super-composantes sont D -orthogonales.

La preuve pour le BIP de l'ACIMOG-PLS1 est similaire. Comme les matrices T_a sont D -orthogonales, $\sum_a In(P_{T_a} Y_n) = In(P_{T_1, \dots, T_A} Y_n)$.

$$\sum_k (BIPrv_k^n(A)) = K \sum_{a=1}^A \frac{\sum_k \overbrace{\sqrt{\text{var}V(t_{k,a})}}^{\text{var}(t_{k,a})} RV(Y_n^A, t_{k,a})}{\sqrt{\sum_a \text{var}V(T_a)} RV(Y_n^A, [T_1 \dots T_a])} \quad (6.16)$$

$$= K \frac{\sum_a \sum_k \text{cov}V(Y_n^A, t_{k,a})}{\sqrt{\sum_a \text{var}V(T_a)} RV(Y_n^A, [T_1 \dots T_a]) \sqrt{\text{var}V(Y_n^A)}} \quad (6.17)$$

$$= K \frac{\text{cov}V(Y_n^A, [T_1 \dots T_a])}{\underbrace{\sqrt{\sum_a \text{var}V(T_a)}}_{\sqrt{\text{var}V([T_1 \dots T_a])}} RV(Y_n^A, [T_1 \dots T_a]) \sqrt{\text{var}V(Y_n^A)}} \quad (6.18)$$

$$= K \frac{RV(Y_n^A, [T_1 \dots T_a])}{RV(Y_n^A, [T_1 \dots T_a])} = K. \quad (6.19)$$

■

Remarquons que le critère $BIP_k^n(1)$ ne dépend pas de n .

6.3.2.3 Utilisation du coefficient RV

Etant donné qu'il nous a été impossible jusqu'alors d'établir un critère pertinent pour l'ACIMOG-PLS2, nous avons pensé à utiliser les coefficients RV (ROBERT & ESCOUFIER, 1976) de plusieurs façons. En effet, ce coefficient permet de déterminer la proximité entre deux tableaux mesurés sur les mêmes individus, en évaluant la proximité des configurations des individus vus par les deux tableaux. Il peut s'utiliser de trois manières ici :

$$RVx_k^a(k) = RV(\hat{Y}_n^a, X_k) \quad (6.20)$$

$$RVres_n^a(k) = RV(\hat{Y}_n^a, X_k^{(a-1)}) \quad (6.21)$$

$$RV\beta_n^a(k) = RV(\hat{Y}_n^a, X_k \hat{\beta}_n^a[k]). \quad (6.22)$$

Le premier, $RVxk_n^a(k)$, mesure à l'étape a , la proximité entre le modèle d'ordre a de Y_n et les X_k initiaux. Le second, $RVres_n^a(k)$ mesure la proximité entre le modèle d'ordre a de Y_n et les $X_k^{(a-1)}$, c'est-à-dire les tableaux prédicteurs utilisés à l'étape a , autrement dit ceux qui servent à la construction des modèles partiels d'ordre a . Ce coefficient donne l'importance du tableau k à l'étape a uniquement. Enfin, le dernier, $RV\beta_n^a(k)$, mesure la proximité entre le modèle d'ordre a de Y_n et la partie du modèle relative au tableau prédicteur X_k , soit $X_k\widehat{\beta}_n^a[k]$. Il est plus élevé que $RVxk_n^a$ en général car les coefficients du modèles sont pris en compte. On peut raisonnablement penser que si X_k est important pour la modélisation de Y_n alors $X_k\widehat{\beta}_n^a[k]$ devrait être proche de \widehat{Y}_n^a en terme de configuration des individus, ce qui est mesuré par le coefficient RV.

Avec les ACIMOG-PLS2 et 3, tous ces critères sont indépendant de n à l'étape 1. Cela vient de la déflation qui est faite sur une super-composante. En effet :

$$\begin{aligned} covV(\widehat{Y}_n^1, X_k) &= tr(X_k Q_k^x X_k' D P_{t_1} Y_n Q_n^y Y_n' D P_{t_1}) \\ &= tr(P_{t_1} X_k Q_k^x X_k' D P_{t_1} Y_n Q_n^y Y_n' D P_{t_1}) \\ &= \frac{1}{(t_1' D t_1)^3} tr(t_1 t_1' D X_k Q_k^x X_k' D t_1 t_1' D Y_n Q_n^y Y_n' D t_1 t_1' D) \\ &= \frac{1}{(t_1' D t_1)^3} (t_1' D X_k Q_k^x X_k' D t_1 t_1' D Y_n Q_n^y Y_n' D t_1 t_1' D t_1) \\ &= \frac{1}{(t_1' D t_1)^2} (t_1' D X_k Q_k^x X_k' D t_1) (t_1' D Y_n Q_n^y Y_n' D t_1) \\ &= \frac{1}{(t_1' D t_1)^2} \|X_k' D t_1\|_{Q_k^x}^2 \|Y_n' D t_1\|_{Q_n^y}^2. \end{aligned}$$

Avec un calcul analogue, nous obtenons d'autre part :

$$varV(\widehat{Y}_n^1) = \frac{1}{(t_1' D t_1)^4} \|Y_n' D t_1\|_{Q_n^y}^4.$$

D'où :

$$\begin{aligned} RVxk_n^1(k) &= (t_1' D t_1)^2 \|X_k' D t_1\|_{Q_k^x}^2 \|Y_n' D t_1\|_{Q_n^y}^2 / \left((t_1' D t_1)^2 \|Y_n' D t_1\|_{Q_n^y}^2 \sqrt{varV(X_k)} \right) \\ &= \|X_k' D t_1\|_{Q_k^x}^2 / \sqrt{varV(X_k)} \\ &= RVres_n^1(k), \end{aligned}$$

ce qui est indépendant de n .

De la même manière, en utilisant l'expression de $\widehat{\beta}_n^1[k] = Q_k^x a_{k,1} t_1' D Y_n / (t_1' D t_1)$, pour l'ACIMOG-PLS2 et $\widehat{\beta}_n^1[k] = \mu_{k,1} Q_k^x a_{k,1} t_1' D Y_n / (t_1' D t_1)$ pour l'ACIMOG-PLS3 (voir §5.6.3, page 99), nous obtenons dans les deux cas $RV\beta_n^1(k) = (t_{k,1}' D t_1)^2 / (t_{k,1}' D t_{k,1})$.

Il se pose ensuite la question du seuil à partir duquel un de ces coefficient RV est significativement important pour la modélisation de l'ensemble des Y_n . En effet, lorsqu'on fait une sélection de blocs, il est nécessaire de refaire l'analyse avec les blocs sélectionnés. Or, lorsqu'il y a un grand nombre de tableaux réponses, il n'est pas rare qu'un tableau X_k ne soit important que pour certains groupes de réponses, ou inversement, qu'un tableau X_k soit important pour la plupart des Y_n mais pas pour tous. Ainsi, il faut sélectionner des X_k qui

soient globalement importants.

Une première méthode, subjective, est de dénombrer, pour chaque étape a et chaque tableau X_k , les coefficients RV supérieur à 0.5, et ceux supérieurs à 0.7. Ainsi, X_k peut être considéré comme très important si le nombre des coefficients RV supérieurs à 0.7 associés à X_k est supérieur à $N/2$, et moyennement important si le nombre de coefficients RV supérieurs à 0.5 associés à X_k est supérieur à $N/2$. Plus les nombres K et N de tableaux sont grands, plus la sélection sera difficile. En pratique, nous avons choisi $N/2$ pour ne pas être trop sélectifs. Les analyses répétées permettent d'affiner la sélection.

Une méthode plus rigoureuse est de tester la significativité des coefficients RV, en utilisant les tests de permutations, que nous décrivons dans le paragraphe suivant.

6.3.2.4 Test de permutations

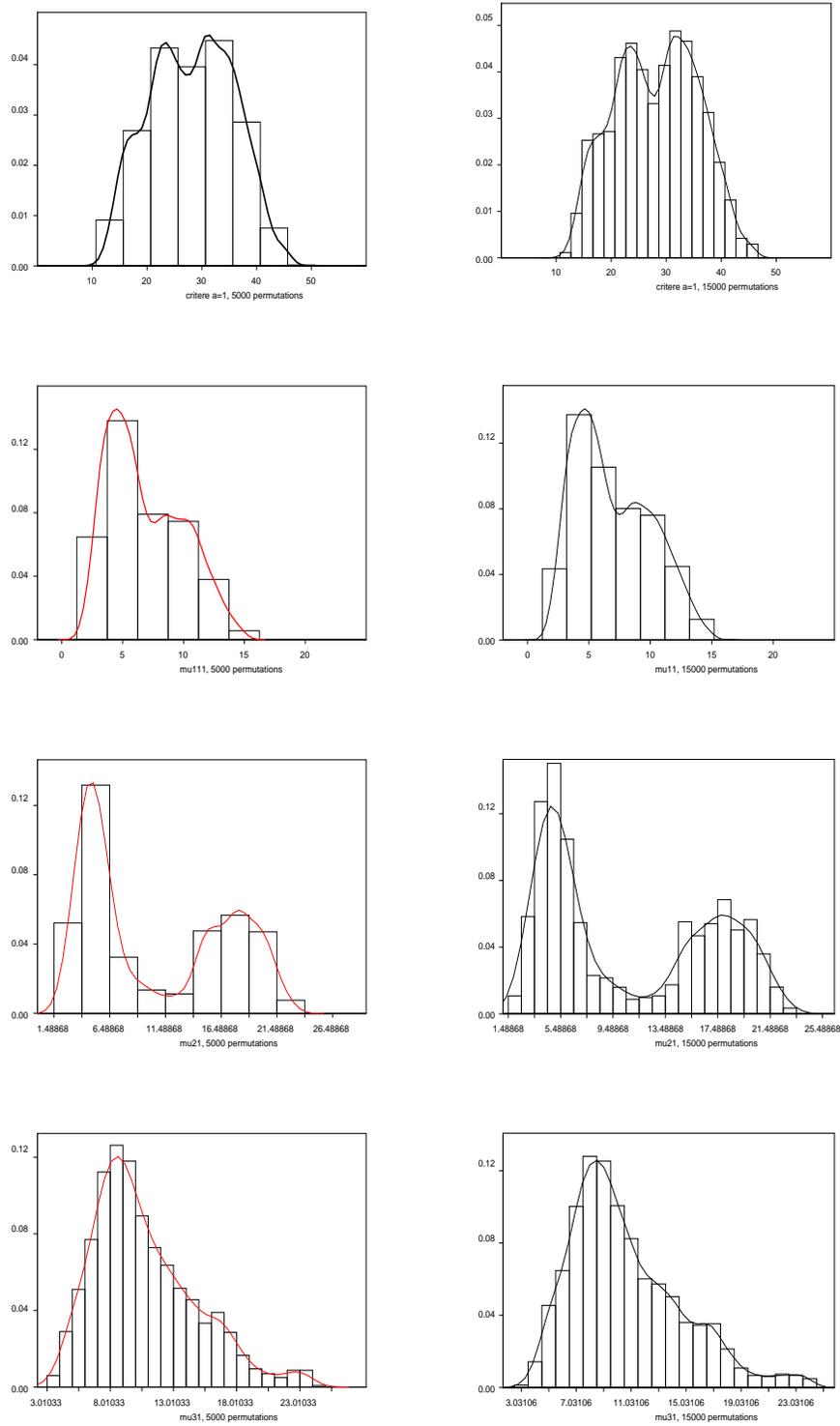
KAZI-AOUAL, HITIER, SABATIER & LEBRETON (1995) ont mis en place des tests de permutations pour tester l'indépendance entre deux tableaux X et Y . Une des statistiques de test proposée est le coefficient RV entre 2 matrices. L'hypothèse est la suivante : si ces tableaux sont indépendants, effectuer des permutations des individus sur un des tableaux ne devrait pas affecter le coefficient RV ainsi que les résultats d'une analyse de co-inertie. KAZI-AOUAL et al. (1995) ont calculé les trois premiers moments (moyenne, écart-type et asymétrie) des statistiques de test s'écrivant sous la forme $tr(AW)$ où A est une matrice centrée et symétrique, et $W = YY'D$, sous hypothèse de permutation des individus de la matrice X . Grâce à ces trois moments, la loi de la statistique de test peut être approchée par une loi de Pearson type III, dont la densité est donnée par $f(x) = \left(\frac{x-\alpha}{\beta}\right)^{\rho-1} e^{-\frac{x-\alpha}{\beta}} / (\beta\Gamma(\rho))$, avec $\alpha \in \mathbb{R}$, $\beta > 0$ et $\rho > 0$ tels que $E(X) = \alpha + \rho\beta$, $V(X) = \rho\beta^2$ et l'asymétrie $\gamma_1 = 2/\sqrt{\rho}$.

Nous avons essayé de mettre un place un test «global» afin de mettre en évidence la pertinence d'un ou plusieurs blocs de prédicteurs. Si, après permutation des individus des matrices X_k , le critère de l'ACIMOG-PLS ou REMUB ne bouge pas, cela signifie que les X_k sont indépendants des Y_n et qu'il n'y a pas de sens à établir un tel modèle. Malheureusement, le critère ne peut s'écrire de la forme $tr(AW)$. Il faut effectuer les permutations et estimer les moments de la loi sous permutation à l'aide des résultats.

Les premiers résultats obtenus sous S-Plus[®], avec 5000 et 15000 permutations, $K = 3$ tableaux X_k de taille 8×10 et $N = 2$ tableaux réponses Y_n de taille 8×10 , semblent indiquer que les lois des «critères permutés» et des coefficients μ_k «permutés» ne peuvent être ajustées par une loi de Pearson Type III. Par l'utilisation de la fonction native au logiciel S-Plus permettant d'estimer une densité par la méthode des histogrammes à noyau gaussien, indique qu'il serait plus juste de les estimer par un mélange de lois normales.

Des investigations futures plus poussées devraient nous permettre de trouver un moyen de calculer la probabilité de non dépassement d'un quantile d'une telle loi.

FIG. 6.1 – Histogramme et estimation de la densité de la loi suivie par le critère et les $\mu_{k,1}$ avec 5000 et 15000 permutations



Chapitre 7

Résumés et comparaisons

Pour clore cette première partie, nous proposons de montrer l'équivalence entre REMUB, PLS, MBPLS et l'ACIMO pour $N = 1$, à la première étape seulement pour certaines, pour toutes les étapes pour d'autres.

Puis, nous résumons en tableaux les principales propriétés des méthodes présentées dans cette partie.

7.1 Relation entre MBPLS-REMUB-PLS-ACIMO

Nous allons montrer que la MBPLS (chapitre 5.2) est équivalente à plusieurs méthodes dont la PLS usuelle, REMUB (chapitre 5.7), ACIMO (Chapitre 2.4) lorsque $N = 1$, autrement dit lorsqu'il n'y a qu'un seul tableau réponse à expliquer Y .

7.1.1 REMUB-MBPLS

Rappelons que REMUB consiste à maximiser, sous les contraintes de norme unité pour α , ν , a_k et b_n , le critère $C_3 = \sum_{k=1}^K \sum_{n=1}^N \alpha_k \nu_n \text{cov}(X_k Q_k^x a_k, Y_n Q_n^y b_n)$. Sans perte de généralité, on peut omettre les métriques Q_k^x et Q_n^y . Il suffit de les considérer comme partie intégrante des tableaux X_k et Y_n . Dans la situation où $N = 1$, ν est un scalaire égal à 1 et la super-composante u est égale à u_1 . Nous noterons donc u cette composante. Les solutions a_k et $b_1 = b$ après convergence de l'algorithme vérifient les équations de l'ACIMOG appliqué aux tableaux $\alpha_k X_k$ et νY :

$$a_k = \frac{1}{\text{cov}(\alpha_k t_k, u_1 \nu)} X_k' D Y b \nu = X_k' D Y b \nu / \|X_k' D Y b \nu\| \quad (7.1)$$

$$b = \frac{1}{\text{cov}(\sum_{k=1}^K \alpha_k t_k, \nu u_1)} Y' D \sum_{k=1}^K \alpha_k X_k a_k = \frac{1}{\text{cov}(t, u)} Y' D T \alpha \quad (7.2)$$

t étant la super-composante explicative. $cov(t,u)$ est le maximum du critère. De plus, les solutions α et ν vérifient les équations de la régression PLS usuelle de u_1 sur $T = [t_1 \ \cdots \ t_K]$:

$$\alpha = \frac{1}{cov(t,u)} T' D u_1 \nu \quad (7.3)$$

$$\nu = \frac{1}{cov(t,u)} u_1' D T \alpha = \frac{1}{cov(t, \nu u_1)} cov(u_1, t) = 1/\nu \quad (7.4)$$

Cette dernière équation (7.4) montre que $\nu = 1$ (car ν est positif). Ainsi, α est solution du problème aux valeurs propres :

$$\alpha = \frac{1}{cov^2(t,u)} T' D Y Y' D T \alpha. \quad (7.5)$$

$cov^2(t,u)$ est la plus grande valeur propre (car c'est l'optimum du critère) et α est le premier vecteur propre de $T' D Y Y' D T$. De plus, en utilisant les équations (7.2) et (7.3), on obtient :

$$b = \frac{1}{cov^2(t,u)} Y' D T T D Y b. \quad (7.6)$$

Ainsi, b et α sont les solutions de la régression PLS de Y sur T .

Passons maintenant à la MBPLS. Nous avons montré au chapitre 5.2 que celle-ci revenait à effectuer jusqu'à convergence les régressions PLS usuelles de $U = [u_1 \ \cdots \ u_N]$ sur X_k et Y_n sur T . Dans le cas où $N = 1$ cela revient donc à effectuer les régression PLS de $u = Yc$ sur X_k pour obtenir les poids ω_k de norme 1 et les composantes t_k et de Y sur T pour obtenir la super-composante $t_T = T\omega_T$ et le vecteur de poids c de norme unité de la composante $u = Yc$. Cela se traduit par les équations suivantes à la convergence :

$$\omega_k = \frac{1}{cov(t_k, u)} X_k' u = X_k' u / \|X_k' u\| \quad (7.7)$$

$$c = \frac{1}{cov(t_T, u)} Y' T \omega_T \quad (7.8)$$

$$\omega_T = \frac{1}{cov(t_T, u)} T' Y c, \quad (7.9)$$

où $cov^2(t_T, u)$ est la plus grande valeur propre de $Y' T T' Y$ et $T' Y Y' T$ associée respectivement aux vecteurs propres c et ω_T .

Ainsi, les solutions de REMUB et de MBPLS vérifient les mêmes problèmes aux valeurs propres et donc $\alpha = \omega_T$, $a_k = \omega_k$ et $b = c$. Les solutions d'ordre supérieurs de REMUB sont égales aux solutions d'ordre supérieur de MBPLS dans laquelle la déflation est faite sur la super-composante t_T (WESTERHUIS & COENEGRACHT, 1997).

7.1.2 MBPLS-ACIMO

Les solutions ω_k et c d'ordre 1 de MBPLS sont égales aux solutions a_k et b d'ordre 1 de l'ACIMO (Chapitre 2.4). Rappelons que l'ACIMO consiste en la maximisation de $C_1 =$

$\sum_{k=1}^K cov^2(X_k Q_k^x a_k, Y Q b)$ sous contrainte de norme unité des axes a_k et b . Les solutions d'ordre 1 de l'ACIMO sont données par :

$$a_k = X_k' D Y b / \|X_k' D u\| = X_k' D u / (t_k' D u) \quad (7.10)$$

$$b = Y' D \sum_{k=1}^K t_k' D u t_k / \|T' D u\|^2 = Y' D \sum_{k=1}^K X_k X_k' D Y b / \|T' D u\|^2. \quad (7.11)$$

$\|T' D u\|^2$ est égal à l'optimum du critère et b est le vecteur propre de $Y' D \sum_{k=1}^K X_k X_k' D Y$ associé à la plus grande valeur propre $\|T' D u\|^2$. Dans MBPLS, c est relié à ω_k par :

$$c(t_T' u) = Y' T \omega_T \quad (7.12)$$

$$= Y' \sum_{k=1}^K \omega_T[k] t_k, \quad (7.13)$$

or, $\omega_T = T' u / t_T' u$, d'où :

$$c(t_T' u)^2 = Y' \sum_{k=1}^K t_k' u t_k \quad (7.14)$$

$$= Y' \sum_{k=1}^K t_k' u X_k \omega_k \quad (7.15)$$

$$= Y' \sum_{k=1}^K t_k' u X_k X_k' u / (t_k' u) \text{ d'après (7.7)} \quad (7.16)$$

$$= Y' \sum_{k=1}^K X_k X_k' Y c. \quad (7.17)$$

c est le premier vecteur propre de $Y' D \sum_{k=1}^K X_k X_k' D Y$ associé à $(t_T' u)^2$, comme b dans l'ACIMO, donc $b = c$ et $u_{mbpls} = u_{acimo}$. Les équation (7.7) et (7.10) étant identiques, $a_k = \omega_k$, à l'étape 1. Les déflations n'étant pas les mêmes dans l'ACIMO et MBPLS, les solutions d'ordres supérieurs ne sont pas identiques.

Cependant, ceci montre que lorsque $N = 1$, MBPLS revient, à la première étape, à maximiser le critère $C_1 = \sum_{k=1}^K cov^2(X_k \omega_k, Y c)$ sous les contraintes de normes unité pour les vecteurs ω_k et c . Les solutions d'ordres supérieurs sont alors obtenues en remplaçant X_k et Y par les résidus de la projection orthogonale des variables sur la super-composante $t_{T,a} = \sum_{k=1}^K t_{k,a} t_{k,a}' u_a / (t_{T,a}' u)^2$, $X_k^{(a-1)} = X_k^{(a-2)} - P_{t_{T,a-1}} X_k^{(a-2)}$ et $Y^{(a-1)} = Y^{(a-2)} - P_{t_{T,a-1}} Y^{(a-2)}$.

7.1.3 MBPLS-PLS

L'équation (7.11) est aussi l'équation du problème aux valeurs propres de la régression PLS de Y sur le tableau conjoint X permettant le calcul de b . Ainsi, la composante u_1 d'ordre 1 associée aux réponses de REMUB, de l'ACIMO et de la MBPLS est la composante de la régression PLS usuelle de Y sur X . Il reste à montrer maintenant que la super-composante d'ordre 1 est égale à la composante t associée à X de la régression PLS.

La super-composante de MBPLS t_T vérifie :

$$t_T = \sum_{k=1}^K t_k \omega_T[k] \quad (7.18)$$

$$= \sum_{k=1}^K t_k t'_k u / (t'_T u) \quad (7.19)$$

$$= \sum_{k=1}^K X_k \omega_k t'_k u / (t'_T u) = \sum_{k=1}^K X_k X'_k Y Y' t_T / (t'_T u)^2. \quad (7.20)$$

Comme $(t'_T u)^2$ est la plus grande valeur propre de $Y' X X' Y$, c'est aussi la plus grande valeur propre de $X X' Y Y'$. Ainsi, t_T vérifie le problème aux valeurs propres de la régression PLS de Y sur X .

Il y a une relation entre les poids ω_k et ω_T de MBPLS ($t_T = T \omega_T$) et ceux de la régression PLS ω_{pls} ($t_T = X \omega_{pls}$) :

$$\begin{aligned} \omega_k &= X'_k u / (t'_T u) \\ &= \frac{t'_T u}{t'_k u} \omega_{pls}[k] = \omega_{pls}[k] / \|\omega_{pls}[k]\| \end{aligned} \quad (7.21)$$

$$\begin{aligned} \omega_T &= T' Y c / (t'_T u) \\ &= \text{diag}(\omega'_k) X' Y Y' t_T / (t'_T u) \\ &= \text{diag}(\omega'_k) X' Y Y' X \omega_{pls} / (t'_T u) \text{ car } t_T = t_{pls} \\ &= \text{diag}(\omega'_k) \omega_{pls}, \end{aligned} \quad (7.22)$$

d'où,

$$\omega_T[k] = u' X_k \omega_{pls}[k] / (t'_k u) = \omega_{pls}[k] u' X_k / \|u' X_k\|. \quad (7.23)$$

Il est possible de montrer que $\omega_T[k] = t'_k u / t'_T u$ et que ω_T est bien de norme unité en utilisant le fait $t'_k u = \|X'_k u\|$ et $t'_T u = \|T' u\|$. En effet $t'_k u = u' X_k X'_k u / t'_k u$ et $t'_T u = u' T T' u / t'_T u$.

Ainsi, lorsque $N = 1$, la régression PLS de Y sur le tableau conjoint X fournit les mêmes prédictions que la MBPLS. Ce résultat et les relations (7.21) et (7.23) justifient l'algorithme MBPLS proposé par WESTERHUIS & SMILDE (2001). La démonstration de l'équivalence des méthodes fournie par WESTERHUIS et al. (1998) n'est pas exacte. Ils annoncent un résultat sans le démontrer pour dire que les matrices $Y Y' X X'$ et $Y Y' X \text{Diag}(\omega_k) \text{Diag}(\omega'_k) X'$ ont les mêmes valeurs propres et vecteurs propres. Le résultat, comme nous l'avons vu, est exact, mais leur justification ne l'est pas. Quant à la super-composante t_T la démonstration ne figure pas, le résultat est simplement cité.

Alors que nous avons déjà fait la démonstration, QIN, VALLE & PIOVOSO (2001) l'ont publiée, ayant eux aussi remarqué que la démonstration de WESTERHUIS et al. (1998) était fautive. Toutefois, leur démonstration reste incomplète car ils ne montrent pas que t_T et t_{pls} sont associées à la même valeur propre. Dans PLS, on sait que la valeur propre associée à t_{pls} est la plus grande car les solutions de la régression PLS sont issues d'un problème de maximisation sous contraintes, dont le maximum est justement cette valeur propre. Mais MBPLS n'est pas un problème de maximisation par sa définition. C'est justement ce que nous voulons montrer ici avec l'équivalence de MBPLS et PLS.

TAB. 7.1 – Les critères

Numéro	Critère
0	Pas de critère
1	$\min_{A,B,C} \ X_1 - AG(B' \otimes C')\ ^2$
2	$\min_{A,B,C} \ X_1 - A(C \odot B)'\ ^2$
3	$\min_{A,B,D_k^A,H,D_k^B} \sum_{k=1}^K \ X_k - AD_k^A H D_k^B B'\ ^2$
4	$\max_{a_k,v} \sum_{k=1}^K \pi_k \text{cov}^2(t_k,v)$
5	$\max_{a_k,b} \sum_{k=1}^K \text{cov}^2(t_k,u)$
6	$\max_{a_k,b_n} \sum_{k=1}^K \sum_{n=1}^N \text{cov}(t_k,u_n)$
8	$\max_{\omega,c} \text{cov}(t,u)$
9	$\max_{a_k,b_n,\alpha,\nu} \sum_{k=1}^K \sum_{n=1}^N \alpha_k \nu_n \text{cov}(t_k,u_n)$
10	$\max_{\alpha,\beta} \sum_{k=1}^K \sum_{n=1}^N \alpha_k \beta_n \text{tr}(W_k^x D W_n^y D)$
11	$\min_{\alpha,\beta} \ \sum_{k=1}^K \alpha_k W_k^x D - \sum_{n=1}^N \beta_n W_n^y D\ ^2$
12	$\min_W \ Z - X W U'\ ^2$
13	$\min_{W_1,\dots,W_L,W} \sum_{l=1}^L \alpha_l \ X_l - X_l W_l Z_l'\ ^2 + \alpha_T \ T - T W Z_T'\ ^2$
14	$\max_{\beta} (\beta' C \Pi C \beta) / (\beta' \Pi^{-1} \beta)$

7.2 Résumé en tableaux

Nous donnons ici un résumé sous forme de tableaux des méthodes basées sur l'optimisation d'un critère et algorithmiques citées dans cette partie.

Les tableaux (7.1) et (7.2) contiennent respectivement les différents critères et contraintes utilisés. Les tableaux (7.3) et (7.4) résument les méthodes d'analyse conjointe et de régression. Ils contiennent les auteurs, le type de la méthode (multi-tableaux ou cubique) le critère, les contraintes, les orthogonalités éventuelles, la déflation éventuelle, quelques notations et certaines propriétés.

Nous distinguons ω_T de ω : le premier est de dimension $K \times 1$, il est calculé à partir de la matrice $T = [t_1, \dots, t_k]$, le second est de dimension $\sum_k J_k \times 1$, il est calculé à partir de $X = [X_1, \dots, X_K]$. Dans tous les cas $X_k^{(0)} = X_k$ et $Y_n^{(0)} = Y_n$.

La méthode SPLS (§5.4) (BERGLUND & WOLD, 1997) ne figure pas dans le tableau 7.4 car celle-ci peut difficilement se résumer ainsi.

7.2.1 Commentaires

D'une manière générale, nous pouvons dire que les méthodes cubiques diffèrent des méthodes multi-tableaux de part leur mise en oeuvre. En effet, en analyse cubique, le critère utilisé est basé sur la minimisation de l'erreur du modèle (excepté pour NPLS), et on recherche des facteurs associés à chacun des modes. En analyse multi-tableaux, lorsqu'un critère est optimisé, on préfère utiliser un critère de covariance (excepté STATIS) qui cherche des composantes (résumés des tableaux) qui soient le plus proche possible. La «longueur» du multi-tableau n'est pas prise en compte dans les calculs.

TAB. 7.2 – Les contraintes

Numéro	Contraintes
0	Pas de contraintes
1	$A'A = Id_P, B'B = Id_Q, C'C = Id_R$
2	D_k^A diagonale $P \times P, D_k^B$ diagonale $Q \times Q$
3	$\omega'\omega = 1$
4	$c'c = 1$
5	$\omega'_T\omega_T = 1$
6	$\ a_k\ _{Q_k^x}^2 = 1$ (ou $\ \omega_k\ ^2 = 1$)
7	$\ b_n\ _{Q_n^y}^2 = 1$ (ou $\ c_n\ ^2 = 1$)
8	$\ v\ _D^2 = 1$
9	$\ b\ _{Q^y}^2 = 1$
10	$\sum_{k=1}^K \alpha_k^2 = \sum_{n=1}^N \nu_n^2 = 1$
11	$\ w^J\ = \ w^K\ = 1$
12	$\ q^N\ = \ q^M\ = 1$
13	$\ \omega_{T,n}\ = 1; \ c_{U,k}\ = 1$
13	$\alpha'\alpha = 1, \beta'\beta = 1$

De plus, en analyse cubique, en général, toutes les solutions de toutes les dimensions sont obtenues en même temps. En analyse multi-tableaux, on procède étape par étape, en opérant par déflations des tableaux. Ces déflations font que les solutions sont emboîtées. Contrairement aux méthodes cubiques (sauf NPLS) où les trois premiers facteurs d'un modèle à quatre facteurs par mode ne sont pas les mêmes que les facteurs d'un modèle à trois facteurs par mode. En ce sens, NPLS est une méthode cubique fonctionnant comme les méthodes multi-tableaux.

En analyse multi-tableaux, d'une manière générale, la déflation d'un bloc (prédicteur ou non) sur sa propre composante à chaque étape, implique l'orthogonalité des axes qui lui sont associés dans \mathbb{R}^{J_k} ou \mathbb{R}^{M_n} . Lorsque la déflation est effectuée sur les super-composantes, seules celles-ci sont orthogonales dans \mathbb{R}^1 .

L'orthogonalité des composantes (partielles ou non) implique la décomposition de la variance des modèles des tableaux qu'elles reconstituent, en fonction des étapes a . De plus, cela permet de construire un repère orthonormé de \mathbb{R}^1 dans lequel on peut projeter les variables, et dont les coordonnées sont données par des corrélations. Par exemple, dans l'ACIMOG1, les variables de Y_n projetées dans le repère $(u_{n,a}/\|u_{n,a}\|_D, u_{n,\ell}/\|u_{n,\ell}\|_D)$ ont pour coordonnées les corrélations (ou covariances) entre Y_n et les composantes $(u_{n,a}/\|u_{n,a}\|_D$ et $u_{n,\ell}/\|u_{n,\ell}\|_D)$. De même, dans l'ACIMOG-PLS2, les variables des tableaux X_k projetées dans le repère $(t_a/\|t_a\|_D, t_\ell/\|t_\ell\|_D)$ ont pour coordonnées les corrélations (ou covariances) entre les variables de $X = [X_1 \cdots X_K]$ et les super-composantes $t_a/\|t_a\|_D$ et $t_\ell/\|t_\ell\|_D$ (se référer à l'annexe C pour plus d'explications).

Les méthodes telles que l'ACOM ou l'ACO utilisent la déflation des individus des tableaux sur les axes associés. Ceci permet la décomposition de l'inertie des tableaux par rapport aux axes d'ordre a . De plus, cela implique que les individus du tableau X_k projetés dans le repère orthonormé $(a_{k,a}, a_{k,\ell})$ ont pour coordonnées les composantes $t_{k,a}$ et $t_{k,\ell}$. Le but de ces méthodes est donc d'étudier les nuages des individus de \mathbb{R}^{J_k} et non les variables, comme cela est fait le plus souvent.

Tableau 7.3 – Les méthodes d'analyse conjointe. «MB» désigne une méthode multi-tableaux et «Cb» une méthode cubique. Une expression entre parenthèses signifie qu'elle peut être choisie. S'il s'agit d'un critère, cela signifie que les auteurs ne posent pas le problème ainsi. «NE» désigne des solutions non emboîtées, et «E», emboîtées. «ALS» précise que l'algorithme est de type ALS, «ITER.» qu'il est itératif et «Diago.» qu'il s'agit d'une diagonalisation.

Nom	Type	Auteurs	Crit.	Cont.	Orthogonalités	Notations	Déflation	Remarques
TUCKER3	Cb	TUCKER (1966)	1	1	$\{a_p\}_p, \{b_q\}_q, \{c_r\}_r$	$GP = A'X(C \otimes B)$	aucune	solutions N . ALS
PARAFAC/ GANDECOMP	Cb	HARSHMAN CARROLL & CHANG (1970)	2	0	$\{a_p\}_p, \{b_q\}_q, \{c_r\}_r$		aucune	solutions N . ALS.
PARATUCK2	Cb	HARSHMAN & LUNDY (1996)	3	2	$\{a_p\}_p, \{b_r\}_r$		aucune	solutions N . ALS.
ACOM	MB	CHESSEL & HANAFI (1996)	4	6, 8	$\{a_{k,\alpha}\}_\alpha, \{v_\alpha\}_\alpha$	$t_{k,\alpha} = X^{(\alpha-1)} Q_k^\alpha a_{k,\alpha}$	$X_k^{(\alpha)} = X_k^{(\alpha-1)} P_{a_{k,\alpha}}^\perp$	solutions E. Diago.
ACO	MB	LAFOSSE & HANAFI (1997)	5	6, 9	$\{a_{k,\alpha}\}_\alpha, \{b_\alpha\}_\alpha$	$u_\alpha = Y Q^y b_\alpha$ $t_{k,\alpha} = X^{(\alpha-1)} Q_k^\alpha a_{k,\alpha}$	$X_k^{(\alpha)} = X_k^{(\alpha-1)} P_{a_{k,\alpha}}^\perp$	solutions E. Diago.
ACIMO	MB	§2.4	5	6, 9	$\{t_{k,\alpha}\}_\alpha, \{u_\alpha\}_\alpha, \{a_{k,\alpha}\}_\alpha$	$t_{k,\alpha} = X_k^{(\alpha-1)} Q_k^\alpha a_{k,\alpha}$ $u_\alpha = Y^{(\alpha-1)} Q^y b_\alpha$	$X_k^{(\alpha)} = P_{t_{k,\alpha}}^\perp X_k^{(\alpha-1)}$ $Y^{(\alpha)} = P_{u_\alpha}^\perp Y^{(\alpha-1)}$	solutions E. Diago.
ACIMOG1	MB	§3.4	6	6, 7	$\{t_{k,\alpha}\}_\alpha, \{u_{n,\alpha}\}_\alpha, \{a_{k,\alpha}\}_\alpha$	$t_{k,\alpha} = X_k^{(\alpha-1)} Q_k^\alpha a_{k,\alpha}$ $u_{n,\alpha} = Y_n^{(\alpha-1)} Q_n^y b_{n,\alpha}$	$X_k^{(\alpha)} = P_{t_{k,\alpha}}^\perp X_k^{(\alpha-1)}$ $Y^{(\alpha)} = P_{u_{n,\alpha}}^\perp Y^{(\alpha-1)}$	solutions E. ITER.
ACIMOG2	MB	§3.4	6	6, 7	$\{t_{k,\alpha}\}_\alpha$	$t_{k,\alpha} = X_k^{(\alpha-1)} Q_k^\alpha a_{k,\alpha}$ $u_{n,\alpha} = Y_n^{(\alpha-1)} Q_n^y b_{n,\alpha}$ $t_\alpha = \sum_{k=1}^K t_{k,\alpha}$ $u_\alpha = \sum_{n=1}^N u_{n,\alpha}$	$X_k^{(\alpha)} = P_{t_\alpha}^\perp X_k^{(\alpha-1)}$ $Y^{(\alpha)} = P_{u_\alpha}^\perp Y^{(\alpha-1)}$	solutions E. ITER.
ACIMOG3	MB	§3.4	6	6, 7	$\{t_\alpha\}_\alpha$	$t_{k,\alpha} = X_k^{(\alpha-1)} Q_k^\alpha a_{k,\alpha}$ $u_{n,\alpha} = Y_n^{(\alpha-1)} Q_n^y b_{n,\alpha}$ $t_\alpha = \sum_{k=1}^K \mu_{k,\alpha} t_{k,\alpha}$ $u_\alpha = \sum_{n=1}^N \lambda_{n,\alpha} u_{n,\alpha}$ $\mu_{k,\alpha} = \text{cov}(t_{k,\alpha}, \sum_{n=1}^N u_{n,\alpha})$ $\lambda_{n,\alpha} = \text{cov}(u_{n,\alpha}, \sum_{k=1}^K t_{k,\alpha})$	$X_k^{(\alpha)} = P_{t_\alpha}^\perp X_k^{(\alpha-1)}$ $Y^{(\alpha)} = P_{u_\alpha}^\perp Y^{(\alpha-1)}$	solutions E. ITER.
DOACT	MB	§3.3	10, 11	13	$\{\alpha_\alpha\}_\alpha, \{\beta_\alpha\}_\alpha$		aucune	α_α et β_α : Diago.
MWMBcovC*	Cb	SMILDE et al. (2000)	13	0	$\{t_{T,\alpha}\}_\alpha, \{b_{i,q}\}_q, \{c_{i,r}\}_r, \{t_{i,\alpha}\}_\alpha$	$T_I = X_I W_I, T_T = T W,$ $\alpha_T + \sum_{I=1}^L \alpha_I = 1$	aucune	solutions N . ALS
STATIS : étape compromis	MB	LAVIT (1988)	14	0		$\beta = \alpha p_1$ $W = \sum_{k=1}^k \alpha \alpha_k \pi_k W_k^x$ $p_1 = (\alpha_1, \dots, \alpha_K)'$ premier vecteur propre normé de CII	aucune	

* Z_I est pris de façon à construire un modèle de type TUCKER1, 2, 3 ou PARAFAC (voir §3.1). X_I est de taille $l \times J_I K_I$, dépliage du cube \underline{X}_I .

Tab. 7.4 – Les méthodes de régression. «MB» désigne une méthode multi-tableaux et «Cb» une méthode cubique. Une expression entre parenthèse signifie qu'elle peut être choisie. S'il s'agit d'un critère, cela signifie que les auteurs ne posent pas le problème ainsi. «NE» désigne des solutions non emboîtées, et «E» emboîtées. «ALS» désigne que l'algorithme est de type ALS, «ITER.» qu'il est itératif et «Diago», qu'il s'agit d'une diagonalisation.

Nom	Type	Auteurs	Crit.	Cont.	Orthogonalités	Notations	Définition	Remarques
NPLS	Cb	Bro (1996)	8	11 et 12	aucune	$t_a = X^{(\alpha-1)} \omega_k^k \otimes \omega_j^j$ $u_a = Y^{(\alpha-1)} q_a^k \otimes q_a^j$ $\omega_a = \omega_a^k \otimes \omega_a^j$	$X_k^{(\alpha)} = X_k^{(0)} - T_a \omega_a$ $Y_n^{(\alpha)} = Y_n^{(0)} - \sum_{a=1}^k P_{T_a} u_a q_a^k \otimes q_a^{j'}$	solutions NE. ITER.
LTD	Cb	Stahle (1989)	0 (8)	11 et 12	$\{t_a\}_a$	$t_a(\hat{\theta}) = \omega_a^k X^{(\alpha-1)} \omega_a^j$ $u_a(\hat{\theta}) = q_a^k Y^{(\alpha-1)} q_a^j$ $P = X^{(\alpha-1)'} \dots X^{(\alpha-1)'} [t_a / \ t_a\ ^2]$	$X_k^{(\alpha)} = X_k^{(\alpha-1)} - t_a(\hat{\theta}) P$ $Y_n^{(\alpha)} = Y_n^{(\alpha-1)} - u_a' t_a(\hat{\theta}) q_a^k q_a^{j'}$	solutions E; égale à NPLS
MPLS	Cb	Wold et al. (1987)	(8)	6, 7	$\{t_a\}_a$	$t_a = X_k^{(\alpha-1)'} \omega_a$	$X_k^{(\alpha)} = P_{t_a}^\perp X_k^{(\alpha-1)}$	égale à PLS conjointe
MBPLS	MB	WANGEN & KO-WALSKI (1988)	0	6, 7, 14	$\{t_{k,a}\}_a$	$t_{k,a} = X_k^{(\alpha-1)} \omega_{k,a}$ $t_{T_n,a} = T_{a-1} \omega_{T_n,a}$	$X_k^{(\alpha)} = P_{t_{k,a}}^\perp X_k^{(\alpha-1)}$ $Y_n^{(\alpha)} = P_{t_{T_n,a}}^\perp Y_n^{(\alpha-1)}$	solutions E. ITER.
MBPLS N=1	MB	WANGEN & KO-WALSKI (1988)	0	6, 7, 5, 4	$\{t_{k,a}\}_a, \{\omega_{k,a}\}_a$	$t_{k,a} = X_k^{(\alpha-1)} \omega_{k,a}$ $u_{a} = Y^{(\alpha-1)} c_a$ $t_{T,a} = \sum_k \omega_{T,a} [k] t_{k,a}$	$X_k^{(\alpha)} = P_{t_{k,a}}^\perp X_k^{(\alpha-1)}$ $Y^{(\alpha)} = P_{t_{T,a}}^\perp Y^{(\alpha-1)}$	solutions E. optimise aussi le critère 5.
MBPLS N=1	MB	WESTERHUIS & SMULDE (2001)	8	3, 4	$\{t_{T,a}\}_a, \{\omega_a\}_a$	$t_{T,a} = X^{(\alpha-1)} \omega_a$ $\omega_{k,a} = \omega_a [k] / \ \omega_a [k]\ $ $t_{k,a} = X_k^{(\alpha-1)} \omega_{k,a}$	$X_k^{(\alpha)} = P_{t_{T,a}}^\perp X_k^{(\alpha-1)}$ $Y^{(\alpha)} = P_{t_{T,a}}^\perp Y^{(\alpha-1)}$	
HPLS	MB	Wold et al. (1996)	0	4, 5, 6, 7	$\{T_a\}_a$	$T_{T,a} = \sum_k \omega_a [k] t_{k,a}$ $t_{k,a} = X_k^{(\alpha-1)} \omega_{k,a}$	$X_k^{(\alpha)} = P_{T_a}^\perp X_k^{(\alpha-1)}$ $Y_n^{(\alpha)} = P_{T_a}^\perp Y_n^{(\alpha-1)}$	algo compliqué; information mal prise en compte
ACIMO-PLS	MB	VIVIEN & SABATIER (2001c), §5.5	5, (8)	4 et 7	$\{t_a\}_a, \{a_a\}_a$	$t_a = X^{(\alpha-1)} Q_a^x a_a$ $u_{n,a} = Y_n^{(\alpha-1)} Q_a^y b_{n,a}$	$X^{(\alpha)} = P_{t_a}^\perp X^{(\alpha-1)}$ $Y_n^{(\alpha)} = P_{t_a}^\perp Y_n^{(\alpha-1)}$	solutions E. Diago.
ACIMOG-PLS1	MB	\$5.6	6	6 et 7	$\{T_a\}_a, \{t_{k,a}\}_a, \{a_{k,a}\}_a$	$T_a = \sum_k [t_{1,a}, \dots, t_{k,a}]$ $t_{k,a} = X_k^{(\alpha-1)} Q_k^x a_{k,a}$ $u_{n,a} = Y_n^{(\alpha-1)} Q_n^y b_{n,a}$	$X_k^{(\alpha)} = P_{T_a}^\perp X_k^{(\alpha-1)}$ $Y_n^{(\alpha)} = P_{T_a}^\perp Y_n^{(\alpha-1)}$	solutions E. ITER.
ACIMOG-PLS2	MB	\$5.6	6	6 et 7	$\{t_a\}_a$	$t_a = \sum_{k=1}^K t_{k,a}$ $t_{k,a} = X_k^{(\alpha-1)} Q_k^x a_{k,a}$ $u_{n,a} = Y_n^{(\alpha-1)} Q_n^y b_{n,a}$	$X_k^{(\alpha)} = P_{t_a}^\perp X_k^{(\alpha-1)}$ $Y_n^{(\alpha)} = P_{t_a}^\perp Y_n^{(\alpha-1)}$	solutions E. ITER.
ACIMOG-PLS3	MB	\$5.6	6	6 et 7	$\{t_a\}_a$	$t_a = \sum_{k=1}^K \mu_{k,a} t_{k,a}$ $\mu_{k,a} = \cos(t_{k,a}, \sum_{r=1}^k t_{r,a})$ $u_{n,a} = Y_n^{(\alpha-1)} Q_n^y b_{n,a}$	$X_k^{(\alpha)} = P_{t_a}^\perp X_k^{(\alpha-1)}$ $Y_n^{(\alpha)} = P_{t_a}^\perp Y_n^{(\alpha-1)}$	solutions E. ITER.
REMUB	MB	\$5.7	9	6, 7, 10	$\{t_a\}_a$	$t_a = \sum_{k=1}^K \alpha_{k,a} t_{k,a}$ $t_{k,a} = X_k^{(\alpha-1)} Q_k^x a_{k,a}$ $u_{n,a} = Y_n^{(\alpha-1)} Q_n^y b_{n,a}$	$X_k^{(\alpha)} = P_{t_a}^\perp X_k^{(\alpha-1)}$ $Y_n^{(\alpha)} = P_{t_a}^\perp Y_n^{(\alpha-1)}$	solutions E. ITER.

7.2.2 Quelle méthode choisir ?

Devant le grand nombre de méthodes existantes, l'utilisateur se demande forcément comment choisir la meilleure méthode pour son problème. Malheureusement, il n'existe pas de réponse universelle. L'utilisateur doit donc choisir entre la facilité d'interprétation, l'optimisation de critère, la simplicité de mise en oeuvre et la qualité de la modélisation.

- **Analyse conjointe** : si les données du problème forment un tableau à trois entrées, il est plus intéressant de privilégier une méthode cubique si l'on veut tenir compte de l'effet des trois modes. La méthode la plus facile à mettre en oeuvre est sans doute PARAFAC. Les méthodes TUCKER ont «l'inconvénient» de posséder une matrice noyau difficilement interprétable. Cependant, les modèles cubiques sont difficiles à interpréter à cause de la perte de la dualité (FRANC, 1992). On peut donc se reporter sur l'utilisation d'une méthode multi-tableaux, en choisissant bien le sens dans lequel on découpe le cube de données.

Les méthodes STATIS et ACOM permettent l'étude d'un ensemble de tableaux en procédant de manière complètement différente. L'ACOM est une méthode «component-wise» alors que STATIS est basée sur les opérateurs «WD» ou «VQ» et la recherche d'un opérateur compromis. Les méthodes ACO et ACIMO permettent la comparaison d'un ensemble de tableaux avec un tableau de référence: l'ACO cherche à étudier les individus dans \mathbb{R}^k tandis que l'ACIMO cherche à étudier les variables dans \mathbb{R}^1 .

Si deux ensembles de tableaux sont à comparer, alors on peut utiliser DO-ACT, qui fonctionne comme la méthode STATIS, ou une des méthodes ACIMOG. Le peu d'expérience que nous en avons ne nous permet pas de dire si l'une des approches est préférable à l'autre. Si l'on veut comparer deux tableaux à trois entrées en tenant compte de tous les modes, la seule méthode, difficile à mettre en oeuvre, est MWMBcovC.

- **Régression** : si les variables explicatives forment un tableau à trois entrées, la méthode NPLS est bien adaptée. Elle optimise un critère et généralise la régression PLS. Cependant, on peut lui reprocher de ne pas fournir de composantes orthogonales ce qui ne permet pas de décomposer la variance du modèle et n'assure pas que chaque composante modélise une part différente de l'information contenue dans les données. La méthode MPLS n'est pas réellement cubique car elle est équivalente à la régression PLS usuelle. Elle n'est donc d'aucune utilité. Lorsque les données ne sont pas cubiques, l'expérimentateur a le choix parmi un grand nombre de méthodes. La méthode HPLS n'est pas conseillée étant donné les mauvais résultats qu'elle peut donner (voir exemple III-4). MBPLS donne de bons résultats, bien que n'optimisant pas de critère, (voir III-4), mais engendre un grand nombre de composantes, ce qui n'en simplifie pas l'interprétation. Parmi les méthodes que nous avons proposées, le choix n'est pas immédiat. L'ACIMO-PLS convient s'il n'y a qu'un seul tableau de variables explicatives et plusieurs tableaux réponses. La méthode ACIMOG-PLS1, optimise un critère et en très peu d'étapes engendre des modèles ayant un important pourcentage de variance expliquée. Si le nombre d'individus est plus petit que le nombre de tableaux prédicteurs, elle n'est pas applicable. Les méthodes ACIMOG-PLS2, 3 et REMUB optimisent toutes les trois un critère. Elles sont basées sur l'obtention de super-composantes dont le calcul diffère selon la méthode. Basées sur la déflation des tableaux sur les super-composantes explicatives, elles permettent d'utiliser les mêmes composantes pour construire tous les modèles. Ce qui n'est pas le cas lorsque la déflation est faite par bloc, comme avec MBPLS. Cependant, l'interprétation des super-composantes n'est pas toujours très simple.

Deuxième partie

... Aux modèles non linéaires

Introduction

Cette partie est la plus courte des trois parties que contient cette thèse, mais n'est pas pourtant d'un intérêt moindre. Nous nous intéressons ici la modélisation non-linéaire de multi-tableaux. A ce jour, il n'existe à notre connaissance aucune méthode de régression spécifique pour cette situation. Dans la littérature, sont citées des techniques possibles, mais rarement leur application, basées sur l'utilisation des prédicteurs au carré, au cube et les termes d'interaction. Ces termes sont placés dans différents blocs X_k et il est suggéré d'utiliser des méthodes multi-blocs telles que MBPLS ou SPLS. De telles approches sont limitées. Elles ne cherchent que la non-linéarité polynômiale. De plus, ce ne sont pas des méthodes non-linéaires multi-tableaux proprement dites. Pour cela, nous proposons une méthode spécifique (SARMUB), optimisant un critère sous contraintes à chaque étape, généralisant l'ACIMOG-PLS et l'AS-PLS. Elle est présentée en détail dans le chapitre 2.2. Dans le premier chapitre, nous donnons une revue de quelques méthodes non-linéaires basées sur la régression PLS de deux tableaux X et Y , la régression PLS étant le support des méthodes de régression multi-blocs. Le second chapitre concerne la modélisation non-linéaire de plusieurs tableaux. Nous verrons que les premières propositions de méthodes non-linéaires multi-tableaux concernent l'analyse conjointe de tableaux. Après avoir présenté SARMUB en détail, nous proposons trois approches plus simples que SARMUB, basées sur l'utilisation du codage des prédicteurs dans une base de B-splines.

SARMUB n'étant pas entièrement aboutie pour l'instant, comme nous le verrons au cours du chapitre 2, nous sommes dans l'incapacité de conseiller une approche plutôt qu'une autre.

Chapitre 1

Modèles non linéaires PLS en régression : cas de deux tableaux

La régression PLS linéaire étant efficace, dans les années 1980, les chimiométriciens ont commencé à développer des méthodes de modélisation non-linéaire basées sur des extensions de la régression PLS. Il existe bien sûr beaucoup d'autres méthodes paramétriques ou non-paramétriques pour la régression non-linéaire, lesquelles nécessitent en général un nombre d'individus grand par rapport au nombre de variables explicatives. Nous nous focaliserons ici sur celles généralisant la régression PLS qui supporte le cas, de plus en plus courant, où le nombre d'observations est plus petit que le nombre de variables explicatives et qui est la base de notre travail.

La régression PLS usuelle contient plusieurs expressions de la «linéarité» (voir annexe A) :

- les modèles $\widehat{Y}^{(a)} = X\widehat{\beta}^a$ sont linéaires en les variables et en les coefficients.
- La composante PLS $t_a = X\omega_a^*$ est linéaire en les variables : c'est la «relation externe».
- La composante PLS $u_a = b_a t_a + \text{erreur}_a$ est linéaire en t_a : c'est la «relation interne». Selon la définition du vecteur de poids associé à u_a , le réel b_a peut être ou non égal à 1.

Les méthodes non-linéaires basées sur la régression PLS rendent non-linéaire la relation entre $\widehat{Y}^{(a)}$ et X , parfois aussi celle entre $\widehat{Y}^{(a)}$ et les coefficients du modèles. Mais surtout, elles sont basées sur la délinéarisation des relations «externe» et «interne».

Ainsi, un certain nombre de méthodes consiste à rendre non-linéaire la relation entre la composante PLS explicative t et la composante associée aux réponses u (QPLS2, NLPLS, SPL-PLS, CNLPLS). D'autres sont basées sur une transformation *a priori* des variables explicatives avant l'utilisation de la régression PLS (INLR, GIFL-PLS, PLSS) ou encore utilisent un nouveau critère à optimiser dont la résolution s'apparente à celle de la régression PLS (ASPLS). Ces dernières délinéarisent la relation «externe» en optimisant un critère.

Nous nous contentons d'exposer ici le principe des méthodes qui sont essentiellement algorithmiques. Elles ne doivent être utilisées que s'il est prouvé que la liaison entre les variables explicatives et à expliquer n'est pas linéaire. De plus, il est impossible de dire si une méthode est plus efficace qu'une autre. Cela dépend des jeux de données. Cependant, certaines d'entre elles, basées sur une transformations des prédicteurs avant l'application de la régression PLS usuelle, sont inapplicables en pratique lorsque le nombre de variables explicatives est élevé, car elles engendrent une trop grosse augmentation de dimension.

Les méthodes basées sur les fonctions splines, plutôt que les fonctions quadratiques, sont cependant recommandées étant donné leur flexibilité (DE BOOR, 1978).

Les notations utilisées dans ce chapitre sont les suivantes :

- X est la matrice centrée des prédicteurs de taille $I \times J$,
- Y est la matrice centrée des réponses de taille $I \times M$,
- a est l'indice des étapes, A le nombre final d'étapes,
- ω_a est le vecteur des poids associé à la composante t_a ,
- $t_a = X\omega_a$ (ou $t_a = X^{(a-1)}\omega_a$ si X est «déflaté») est la composante PLS,
- q_a est le vecteur de poids associé à la composante réponse u_a ,
- $u_a = Yq_a$ (ou $u_a = Y^{(a-1)}q_a$ si Y est «déflaté») est la composante associée aux réponses.

1.1 PLS Quadratique

La première méthode de régression PLS non-linéaire a été proposée par WOLD, KETTANEH-WOLD & SKAGERBERG (1989) : Quadratic PLS (QPLS). En régression PLS, il y a une relation linéaire entre les composantes t_a et u_a , c'est la «relation interne». L'idée des auteurs est de délinéariser la relation interne par un modèle polynômial d'ordre 2 à l'aide de nouvelles étapes dans l'algorithme de la régression PLS linéaire. Ils sont les initiateurs de ce type de méthodes. L'algorithme (WOLD et al., 1989) est le suivant :

étape 0 : prétraitement

centrer et éventuellement réduire X et Y .

Poser $a = 0$

étape 1

$a = a + 1$ et $iter = 0$

a) u est égal à la colonne de Y de plus grande variance.

b) $\omega' = u'X/u'u$, $\omega = \omega'/\|\omega\|$

c) $t = X\omega/(\omega'\omega)$

étape 2

$iter = iter + 1$

Estimer c par les moindres carrés, $\hat{u} = r = c_0 + c_1t + c_2t^2$ où $c = (c_0 \ c_1 \ c_2)'$.

$u_{old} = u = \hat{u}$

étape 3

a) $q = Y'r/(r'r)$, $q = q/\|q\|$

b) $u = Yq/q'q$

étape 4

Estimer c par les moindres carrés : $\hat{u} = c_0 + c_1t + c_2t^2$

$u_{new} = u = \hat{u}$

étape 5 : mise à jour de ω

Calculer la correction $d\omega$ (à l'aide de la linéarisation de Newton-Raphson) par :

$v' = Z'u/u'u$; $v = v/\|v\|$

$s = Zv/v'v$; $b = s'u/s's$

$\Delta\omega = J$ premiers éléments de bw
 Voir détail de cette étape après l'algorithme
 Actualiser ω : $\omega = \omega + \Delta\omega$; $\omega = \omega / \|\omega\|$

étape 6

$$t = X\omega / (\omega' \omega)$$

étape 7 : test de la convergence

Si $\|u_{new} - u_{old}\| / \|u_{old}\| < 10^{-10}$, on passe à l'étape suivante,
 sinon, si $iter > 50$ ou 100 , on passe à l'étape suivante,
 sinon, on repart de l'étape 2.

étape 8

$$r = \hat{u}, q = Y'r / (r'r), q = q / \|q\|, u = Yq / q'q$$

$$c = \left([\mathbf{1} \ t \ t^2]' [\mathbf{1} \ t \ t^2] \right)^{-1} [\mathbf{1} \ t \ t^2]' u$$

étape 9

$$p = X't / (t't)$$

étape 10 : calcul des résidus

$$E^x = X - tp', E^y = Y - rq' = Y - P_r Y$$

étape 11 : passage à l'étape suivante

On pose $X = E^x$ et $Y = E^y$ et on recommence à partir de l'étape 1.

—■—

L'étape 5 de cet algorithme nécessite quelques explications. Lorsque que la relation interne est approximée par une fonction non-linéaire, implicitement on affecte la relation interne, car ω est mis à jour par $X'u$. Ce qui justifie la nécessité de rajouter une étape d'actualisation de ω après l'approximation de la fonction quadratique. WOLD et al. (1989) font référence à une communication de FRANK (1989), dans laquelle cette actualisation n'est pas faite. Cette méthode est référencée par LQ-PLS, qui signifie Linear Quadratic PLS. LQ-PLS serait moins stable. Détaillons l'étape 5. A l'entrée de cette étape, $u = \hat{u} + e$ avec $\hat{u} = f(t) = c_0 + c_1 t + c_2 t^2$. f va être considérée comme une fonction de ω et c , $f(\omega, c)$. Ici, f est quadratique, mais elle peut être remplacée par n'importe quelle autre fonction différentiable en ω . Une linéarisation de Newton-Raphson est utilisée autour du point $\hat{u} = f(t) = f(\omega_{00}, c_{00})$. Ce point étant connu, u est approximé par : $u = \hat{u} + \frac{\partial f}{\partial \omega} |_{00} \Delta\omega + \frac{\partial f}{\partial c} |_{00} \Delta c$, où l'indice 00 signifie que les dérivées partielles sont calculées en (ω, c) tels que $\hat{u} = f(\omega_{00}, c_{00})$. Cette décomposition devient, en notant $c = c_{00} + \Delta c$:

$$u = [\mathbf{1} \ t_{00} \ t_{00}^2] c + [c_{100} X + 2c_{200} t_{00} * X] \Delta\omega \quad (1.1)$$

$$= Z \begin{bmatrix} \Delta\omega \\ c_{00} + \Delta c \end{bmatrix}. \quad (1.2)$$

Pour évaluer $\Delta\omega$, WOLD et al. (1989) effectuent une étape de régression PLS de u sur $Z = [c_{100} X + 2c_{300} t_{00} * X \ \mathbf{1} \ t \ t^2]$, de dimension $l \times J + 3$. « $t * X$ » désigne la multiplication terme à terme de t avec chacune des colonnes de X . $\Delta\omega$ est alors égal aux J premières lignes de bw car cette régression PLS implique $u = bs = Zbv$: s est la composante PLS associée à Z et b est le coefficient de la régression simple de u sur t (relation interne) (voir étape 5).

Une remarque s'impose ici. Cette méthode est tirée de la publication (WOLD et al., 1989). BAFFI, MARTIN & MORRIS (1999a) reprennent cet algorithme et ont mal retranscrit cette étape. En effet, en considérant que (1.1) s'écrit

$$u = [\hat{u} \quad \mathbf{1} \quad t_{00} \quad t_{00}^2 \quad c_{100}X + c_{200} * t_{00}X] \begin{bmatrix} 1 \\ \Delta c \\ \Delta \omega \end{bmatrix},$$

ils écrivent, à tort, que pour calculer $\Delta\omega$, WOLD et al. (1989) effectuent la régression de u sur $Z^* = [\hat{u} \quad \mathbf{1} \quad t_{00} \quad t_{00}^2 \quad c_{100}X + c_{200}t_{00} * X]$, qui ne donne pas exactement les mêmes résultats (le $\Delta\omega$ obtenu ainsi est proportionnel à celui de WOLD et al. (1989)).

BAFFI et al. (1999a) proposent trois autres manières de définir cette matrice Z , définissant ainsi trois variantes de QPLS. Leurs motivations sont liées principalement à deux questions : pourquoi ne pas faire la régression aux moindres carrés ordinaires de u sur Z pour évaluer $\Delta\omega$, et pourquoi utiliser Δc dans les calculs si c n'est pas actualisé ?

Ainsi, leur première variante consiste à mettre à jour $\Delta\omega$ par la partie correspondante de $(Z^{*'}Z^*)^{-1}Z^{*'}u$. Dans la deuxième variante, les termes en Δc sont négligés. $\Delta\omega$ est calculé comme précédemment avec $Z^* = [\hat{u} \quad c_{100}X + c_{200}t_{00} * X]$. Dans leur troisième et dernière variante, dont ils comparent les performances numériquement avec QPLS, ils utilisent $Z^* = e = u - \hat{u} = c_{100}X + c_{200}t_{00} * X$. Alors, $\Delta\omega = (Z^{*'}Z^*)^{-1}Z^{*'}(u - \hat{u}) = Z^{*'}Z^*)^{-1}Z^{*'}(Yq - \hat{u})$. Cette dernière version, basée sur l'erreur e d'ajustement de u , semble améliorer les RMSEP (residual mean square error in prediction) des modèles dans la pratique.

1.2 Non Linear PLS : NLPLS

En 1990, FRANK (1990) propose la méthode de régression non-linéaire dérivée de PLS, appelée Non-Linear PLS (NLPLS).

Il propose de délinéariser la relation interne de PLS à l'aide de lisseurs, à la manière de SMART, (Smooth Multiple Additive Regression Technique) (FRIEDMAN, 1984), pour approximer la relation interne.

La modélisation \hat{y}^m de la variable m des réponses Y par la méthode SMART est de la forme

$$\sum_{a=1}^A q_a(m) f_a^m(X\omega_a^m)$$

où le coefficient $q_a(m)$, le vecteur ω_a^m et la fonction f_a^m sont obtenus simultanément par la minimisation d'un critère aux moindres carrés. Avec SMART, chaque variable réponse est ici modélisée par une fonction non linéaire (en général différente) d'une combinaison linéaire (différente) des prédicteurs à valeur dans \mathbb{R}^l . Pour chaque étape a , trois étapes sont enchaînées : étant donné les coefficients ω_a et les fonctions non-linéaires f_a , les variables Y^m sont régressées séparément sur les fonctions f_a , ce qui permet de déterminer une valeur de q_a . Puis f_a est évaluée par une procédure de lissage en fonction de ω_a et q_a . Enfin, les vecteurs de poids ω_a sont recalculés par une la méthode de Gauss-Newton. L'auteur compare ces trois étapes aux étapes de la régression PLS (FRANK, 1990) : la première étape correspond à la manière dont est calculée les poids q_a associés à la composante u_a , avec la particularité que $f_a(t_a) = t_a$. La

deuxième étape correspond au calcul de la relation interne. Et la dernière étape correspond au calcul de la composante t_a .

C'est cela qui a inspiré la méthode NLPLS. Ci-dessous nous donnons l'algorithme, retranscrit à partir des publications (FRANK, 1990, 1995).

étape 0 : prétraitement

centrer et réduire X et Y .

Poser $a = 1$

étape 1 : étape NIPALS

poser u_a égal à la première colonne de Y .

Répéter jusqu'à convergence de u_a :

- $\omega_a = X'u_a$; $\omega_a = \omega_a / \|\omega_a\|$
- $t_a = X\omega_a$
- $q_a = Y't_a$; $q_a = q_a / \|q_a\|$
- $u_a = Yq_a$

étape 2 : procédure de lissage

Estimer la fonction f_a par une procédure de lissage :

$\hat{u}_a = f_a(t_a)$; $f_a(t_a) = E(u_a|t_a)$

étape 3 : passage à l'étape suivante

$y^m = y^m - q_a(m)f_a(t_a)$

Si on fait une régression PLS linéaire, en plus, on prend les résidus de X :

$X = X - P_{t_a}X$.

$a = a + 1$, puis on retourne à l'étape 1.

—■—

Le modèle d'ordre A s'écrit alors $\hat{Y}^A = \sum_{a=1}^A u_a q'_a = \sum_{a=1}^A f_a(t_a) q'_a$, comme le modèle PLS usuel.

Les composantes t_a ne sont pas orthogonales, car X n'est pas déflaté. La déflation de tableau X sur t_a rendrait les composantes orthogonales, mais le modèle \hat{Y}^A ne serait toujours pas une somme de modèles partiels orthogonaux. Cela compliquerait le modèle un peu plus.

1.3 SPLine PLS : SPLPLS

WOLD (1992), qui s'intéresse toujours à la délinéarisation de la relation interne dans PLS, propose cette fois d'utiliser les fonctions splines (WOLD, 1974; WEGMAN & WRIGHT, 1983) linéaires, quadratiques, ou cubiques, plutôt qu'une fonction quadratique comme il l'avait précédemment fait (WOLD et al., 1989). Cette méthode est désignée par SPL-PLS.

Pour modéliser la relation interne entre u et t , l'auteur propose d'utiliser des fonctions splines de degré 1, 2 ou 3, à L noeuds fixes ξ_ℓ , codées dans la base des puissances tronquées (voir annexe B). Ainsi, à l'étape a , la composante u_a sera liée à t_a par

$$u_a = b_0 + b_1 t_a + b_2 t_a^2 + b_3 t_a^3 + \sum_{\ell=1}^L b_{\ell+3} (t_a - \xi_\ell)_+^3 + \epsilon_a \quad (1.3)$$

Les noeuds étant fixés, ce modèle est linéaire en ses $L + 3 + 1$ coefficients b_0, \dots, b_{L+3} , lesquels peuvent donc être estimés par les moindres carrés.

Pour utiliser cette modélisation, l'auteur préconise peu de noeuds (inférieur à 15) et si possible correspondants à des observations des données ou le plus proche possible. Lorsqu'un grand nombre de noeuds sont utilisés il conseille d'utiliser le codage dans une base de B-splines (voir annexe B).

WOLD (1974) recommande de prendre au moins 5 à 10 observations entre chaque noeuds avec un nombre maximum de noeuds autour de $L = l/7$ ou $L = \sqrt{l}$ lorsque $l > 100$. Cependant, WOLD (1992) recommande ici le minimum de noeuds, autour de $l/25$, ou même mieux de choisir L , ainsi que le degré des polynômes (0, 1, 2, ou 3) par validation croisée, en commençant par des polynômes de degré 0 avec aucun noeud.

L'algorithme

étape 0 : prétraitement et initialisation

centrer et éventuellement réduire X et Y
poser $a = 0$

étape 1

$a = a + 1$

Si $a > 1$, déflation de Y :

- $Q'_{a-1} = (SS')^{-1}S'Y$ où $S, l \times (a-1)$ est la matrice des splines de l'étape précédente
- $F_{a-1} = Y - SQ'_{a-1}$.

étape 2 : une itération NIPALS

$u_a = y^1$, première colonne de Y

$\omega_a = X'u_a/u'_a u_a$; $\omega_a = \omega_a/\|\omega_a\|$

$t_a = X\omega_a/\omega'_a \omega_a$

étape 3 : estimation des coefficients splines par les moindres carrés

$t_{a,old} = t_a$

$\hat{u}_a = s(t_a) = S_a$ avec

$S_a = b_0 + b_1 t_a + b_2 t_a^2 + b_3 t_a^3 + \sum_{\ell=1}^L b_{\ell+3} (t_a - \xi_\ell)_+^3$

$(b_0, \dots, b_{L+3})' = (\Omega'\Omega)^{-1}\Omega'u_a$ où

$\Omega = [\mathbf{1} \quad t_a \quad t_a^2 \quad t_a^3 \quad (t_a - \xi_1)_+^3 \quad \dots \quad (t_a - \xi_L)_+^3]$, de dimension $l \times (L + 4)$

étape 4 : actualisation de q_a et u_a

$q_a = S'_a F_{a-1} / S'_a S_a$; $q_a = q_a / \|q_a\|$

$u_a = F_{a-1} q_a / q'_a q_a$

étape 5 : actualisation de ω_a

$\nu_j = \|t_a\|/\|x^j\|$, x^j désignant la $j^{\text{ième}}$ colonne de X

$r_{j+} = u's(x^j \nu_j) / (\|u_a\| \|s(x^j \nu_j)\|)$

$r_{j-} = -u's(-x^j \nu_j) / (\|u_a\| \|s(x^j \nu_j)\|)$

$\omega_a(j) = \|x^j\| \max(r_{j+}, r_{j-})$; $\omega_a = \omega_a / \|\omega_a\|$

étape 6

$t_a = X\omega_a/\omega'_a \omega_a$

étape 7 : test de la convergence

si $\|t_a - t_{a,old}\|/\|t_a\| < 10^{-10}$ continuer à l'étape 8
 sinon retourner à l'étape 3.

étape 8 : déflation de X , passage à l'étape suivante

$$p_a = X't_a/t_a't_a$$

$$X = X - t_a p_a'$$

Actualiser $S = [SS_a]$

Retourner à l'étape 1.

—■—

Dans l'étape 5, Wold veut être cohérent avec le principe de Heisenberg («H-principe») de PLS d'HÖSKULDSSON (1988, 1992a) qui est la maximisation « $\Delta(\text{fit})/\Delta(\text{var})$ », $\Delta(\text{fit})$ représentant ce qui améliore l'ajustement du modèle et $\Delta(\text{var})$ la variance des paramètres du modèle. Pour satisfaire le principe H, on calcule la covariance entre u_a et les x^j transformés par la spline déterminée à l'étape 4. Dans PLS, ω est calculé par la covariance entre les x^j et u , c'est-à-dire que l'élément j de ω est calculé grâce à $\|x^j\|r(x^j, u)$. C'est ce qui va être utilisé ici. La pondération de x^j par ν_j entraîne que $x^j\nu_j$ à la même variance que t_a et ainsi $s(x^j\nu_j)$ est comparable à $s(t_a)$. C'est un calcul nécessaire, sinon, les noeuds peuvent ne plus être «au bon endroit». De plus, comme $s(-t) \neq -s(t)$, les deux corrélations r_{j+} et r_{j-} sont calculées, la plus forte en valeur absolue étant retenue pour le calcul de ω_a .

1.4 Continuum Non Linear PLS : CNLPLS

TAAVITSAINEN & KORHONEN (1992) et HAARIO & TAAVITSAINEN (1994) proposent d'utiliser des critères permettant d'obtenir un continuum de modèles de régression, pour exprimer une liaison non-linéaire entre les composantes PLS u_a et t_a . Leurs critères se rapprochent d'une certaine façon de celui défini par STONE & BROOKS (1990), permettant d'obtenir, entre autres, la régression PLS et/ou la régression sur composantes principales.

Ainsi, TAAVITSAINEN & KORHONEN (1992) proposent de minimiser

$$C_1(\omega, q, b) = \left[\frac{\|u - f(t, b)\|}{\|u\|} \right]^{2\alpha} \|t\|^{2(\alpha-1)},$$

sous les contraintes de norme unité pour les vecteurs de poids ω et q . Ici, on a $t = X\omega$, $u = Yq$. b sont les paramètres, à estimer, de la fonction de lien f entre u et t : $u = f(t, b) + e$, e étant une erreur. Le paramètre α , $0 \leq \alpha \leq 1$, définit un ensemble de modèles et donne un poids entre la maximisation de l'ajustement de la relation interne et la variance de la composante t .

Les auteurs nomment cette méthode NonLinear PLS (NLPLS), mais comme cette abréviation est déjà utilisée pour désigner la méthode de FRANK (1990), nous la désignerons par CNLPLS1, pour Continuum Non Linear PLS, première définition.

HAARIO & TAAVITSAINEN (1994) proposent des exemples d'application de CNLPLS mais utilisent un autre critère, qu'ils avaient cité dans la première publication (TAAVITSAINEN &

KORHONEN, 1992). Nous désignerons cette seconde méthode par CNLPLS2 (les auteurs la désignant par CNLPLS). Il s'agit de minimiser le critère :

$$C_2(\omega, q, b) = (r(u, f(t, b)))^{2\alpha} (||t|| ||u||)^{2(1-\alpha)},$$

sous les contraintes de norme unité $||q|| = ||\omega|| = 1$.

Avec ce critère, lorsque $\alpha = 0.5$, on retrouve la régression PLS.

L'algorithme ci-dessous est valable pour les deux méthodes.

L'algorithme

étape 0 : prétraitement

Centrer X et Y et éventuellement les réduire.

$$a = 0$$

étape 1 : initialisation de l'étape a

$$a = a + 1$$

Initialiser ω_a et q_a par les solutions de régression PLS linéaire de Y sur X .

Effectuer la régression non-linéaire de u_a sur t_a pour estimer b : $\hat{u}_a = f(t_a, b)$.

étape 2

$$t_a = X\omega_a$$

$$u_a = Yq_a$$

$$\hat{u}_a = f(t_a, b)$$

étape 3 : passage en coordonnées polaires

Coder ω_a en coordonnées polaires θ_{ω_a} :

$$\theta_{\omega_a}(1) \in [-\pi, \pi] \text{ et } \theta_{\omega_a}(j) \in [0, \pi] \text{ pour } j = 2, \dots, J - 1$$

Si $M > 1$, coder q_a en coordonnées polaires θ_{q_a} :

$$\theta_{q_a}(1) \in [-\pi, \pi] \text{ et } \theta_{q_a}(m) \in [0, \pi] \text{ pour } m = 2, \dots, M - 1$$

Sinon, $q_a = \mathbf{1}$.

étape 4 : optimisation du critère

Utiliser une procédure d'optimisation en θ_{ω_a} , θ_{q_a} et b pour :

- CNPLS1 : minimiser sans contraintes, $\left[\frac{||u_a - \hat{u}_a||}{||u_a||} \right]^{2\alpha} ||t_a||^{2(\alpha-1)}$.

- CNLPLS2 : maximiser sans contraintes, $(r(u_a, \hat{u}_a))^{2\alpha} (||t_a|| ||u_a||)^{2(1-\alpha)}$.

Ici, après convergence, la fonction de lien est réduite par la minimisation de $||u_a - c_0 - c_1 f(t_a, b)||^2$ par rapport à c_0 et c_1 , pour conserver la propriété d'invariance de la corrélation par rapport à la norme.

étape 5 : retour aux coordonnées dans la base initiale

Revenir, par la procédure inverse, aux coordonnées de ω_a et q_a dans la base initiale.

étape 6

$$t_a = X\omega_a, u_a = Yq_a, \hat{u}_a = f(t_a, b)$$

$$p_a = X't_a/t_a't_a, \hat{q}_a = Y'\hat{u}_a/\hat{u}_a'\hat{u}_a$$

étape 7 : passage à l'étape suivante

$$Y = Y - \hat{u}_a \hat{q}'_a$$

- CNLPLS1 : $X = X - t_a p'_a$

Retourner à l'étape 1 sur l'on veut une étape supplémentaire.

—■—

Les performances de CNLPLS1 sont étudiées dans (TAAVITSAINEN & KORHONEN, 1992) en utilisant des fonctions de liens quadratiques et logistiques. Les fonctions de liens peuvent être de différentes natures selon l'étape a . Ils comparent les résultats de CNLPLS1 avec ceux de la régression PLS linéaire. Il apparaît qu'elle est efficace lorsque les non-linéarités sont fortes et qu'en une ou deux étapes les modèles CNLPLS1 sont aussi ou mieux explicatifs et stables que ceux de la régression PLS.

De plus, les auteurs soulignent des problèmes d'optimisation lors de l'utilisation de fonctions logistiques. Ils conseillent, par expérience, d'utiliser comme valeurs initiales dans l'algorithme, pour un α quelconque, les solutions obtenues par des modèles (liens logistiques) avec $\alpha = 0.9$ ou 1, qui posent moins de problèmes de convergence vers un minimum global.

Les performances de CNLPLS1, CNLPLS2 et PLS sont comparées dans (HAARIO & TAAVITSAINEN, 1994) en utilisant des fonctions de liens quadratiques, logistiques, exponentielles et gaussiennes.

Les auteurs expliquent que X n'est pas déflaté ici, car l'expérience montre que la partie expliquée de X peut être significative pour l'étape suivante si elle représente une forme différente de non-linéarité.

Notons que CNLPLS1 fait référence à la régression PLS linéaire seulement par son calcul de q_a , \hat{q}_a et l'initialisation de son algorithme. Le critère de la régression PLS ne peut pas être obtenu par une valeur particulière de α , comme c'est le cas avec CNLPLS2.

Ce type de méthode présente l'inconvénient de dépendre du choix de α . Soit, l'expérimentateur sait quel type de régression il veut effectuer, soit α doit être choisi par validation croisée, en plus du nombre A d'étapes. Cela peut devenir très long en temps de calcul, surtout si les données ont une taille importante.

1.5 Implicit Non Linear latent variable Regression : INLR

BERGLUND & WOLD (1997) proposent une méthode simple et rapide pour modéliser non linéairement un tableau Y par un tableau prédictor X . Il s'agit d'étendre le tableau de prédictors avec les termes au carré, et éventuellement les termes d'ordre supérieurs, mais en omettant les termes croisés. Cette méthode est appelée Implicit Non Linear latent variable Regression (INLR).

Les motivations de cette méthode sont multiples. Les méthodes de modélisation PLS non-linéaire utilisant des transformations non-linéaires, (par exemple SPL-PLS) sont souvent si flexibles qu'elles engendrent un sur-ajustement des données. De plus, les différents paramètres doivent être estimés par validation croisée, ainsi que le nombre a de composantes. Cela peut être long et fastidieux dans certains cas.

Une manière d'obtenir des modèles non-linéaires est d'étendre la matrice des prédicteurs avec leur carrés et leur termes croisés, pour ensuite utiliser la régression PLS linéaire. Cependant, les auteurs soulignent que ce type de méthode ne peut pas être appliqué lorsque le nombre de prédicteurs est important, comme c'est généralement le cas en chimométrie. En effet, alors, le nombre de variables correspondant aux termes croisés $(J(J-1)/2)$ devient vite très important devant le nombre des variables restantes $(2J)$, qui sont les variables initiales et leurs carrés. Ceci implique que les termes croisés ont tendance à dominer les autres prédicteurs, ajoute du bruit au modèle et engendre de mauvais modèles, en terme d'ajustement et de prédiction.

En régression PLS linéaire, rappelons qu'il y a une relation interne linéaire entre u_a et t_a . Comme $\hat{Y}^A = \sum_{a=1}^A u_a q'_a$, cela implique que le modèle \hat{Y} est linéaire en les composantes t_a . Une solution pour modéliser une liaison interne non-linéaire est d'utiliser un modèle quadratique. Ce qui est le cas dans QPLS (§1.1). BERGLUND & WOLD (1997) présentent ici une autre manière de le faire. Ils soulignent que si l'on désire un modèle \hat{Y} polynômial et quadratique en les variables latentes (composantes) t_a , il suffit d'introduire les carrés des prédicteurs dans la matrice X : si X est de rang A alors $\hat{X}^A = X = \sum_{a=1}^A t_a p'_a$ et l'élément i de la variable $(x^j)^2$ vaut $(x^j(i))^2 = \sum_{a=1}^A \sum_{b=1}^A t_a(i)t_b(i)p_a(k)p_b(k) + \text{erreur}$.

Pour que INLR soit efficace deux hypothèses doivent être satisfaites :

- X , $l \times J$, est de rang A bien plus petit que J ,
- X a une structure latente : $X = \sum_{a=1}^A t_a p'_a$.

Ainsi, en effectuant la régression PLS des réponses Y sur le tableau $[X \quad X^2]$ de taille $l \times 2J$, les composantes u_a^{inlr} (qui sont une fonction linéaire des composantes t_a^{inlr} , elles-mêmes combinaisons linéaires des variables x^j et $(x^j)^2$) seront une fonction quadratique des variables latentes t_a . Ainsi, implicitement, une relation non linéaire entre les composantes u_a^{inlr} et la structure latente des prédicteurs X , est faite.

Ainsi, la méthode INLR doit être efficace lorsqu'il existe une structure latente de faible dimension qui décrit très bien les prédicteurs X . Alors, si l'utilisation de la régression PLS linéaire ne donne pas de bons résultats, et que la liaison entre u_a et t_a ne semble pas bien modélisée (par une régression linéaire) alors INLR peut être appliquée, c'est-à-dire, la régression PLS des réponses sur la matrice étendue de ses termes au carré (ou au cube) peut être envisagée.

Les points forts de cette méthode sont la simplicité de son utilisation et sa stabilité par rapport aux valeurs aberrantes en comparaison aux méthodes utilisant les fonctions plus flexibles que sont les splines.

Les auteurs présentent trois exemples d'application dans leur publication.

1.6 Splines Additives en régression PLS

La méthode ASPLS, Additive Splines for Partial Least Squares (DURAND & SABATIER, 1997) fournit une technique généralisant celle de la régression PLS linéaire tout en ne procédant comme aucune des méthodes présentées précédemment. Le critère de covariance de PLS

est optimisé, la relation externe de PLS est délinéarisée par l'utilisation de transformations splines optimisées et les composantes t_a obtenues sont orthogonales, préservant la décomposition de la variance totale de Y .

Ici, l'espace \mathbb{R}^l des variables est muni de la métrique D , en général égale à la métrique uniforme Id_1/l .

Pour chaque prédicteurs x^j ($j^{\text{ième}}$ colonne de X), une base, de B -splines est calculée (le degré d_j des splines, le nombre L_j et la position des noeuds ξ_ℓ^j sont choisis par l'expérimentateur). Le vecteur colonne $S^j = B_j a^j$, B_j est la matrice de codage des x^j dans la base de B -splines et représente une transformation spline de x^j définie par le vecteur de coefficients splines a^j . On note $S = [S^1, \dots, S^J]$, où les matrices B_j sont centrées pour que S^j le soit aussi. Ces matrices B_j seront utilisées à chaque étape a de l'ASPLS. L'ASPLS va consister à chercher les vecteurs a_a^j optimaux selon le critère ci-dessous, c'est-à-dire des fonctions splines optimales, ainsi que les vecteurs de poids ω_a et q_a , tout en s'assurant que t_a soit orthogonale aux composantes t des étapes précédentes.

L'étape a de l'ASPLS consiste en la maximisation du critère

$$C(a^1, \dots, a^J, \omega, q) = \text{cov}(t, u) = w' S' D Y^{(a-1)} q$$

sous les contraintes $\|\omega\|^2 = \|q\|^2 = 1$, $\|B_j a^j\|^2 = \|x^j\|^2$ et $t' D t_\ell = 0$ pour $\ell = 1, \dots, a-1$. $Y^{(a-1)}$ désigne la déflation de Y sur les composantes t_1, \dots, t_{a-1} . La contrainte $\|B_j a^j\|^2 = \|x^j\|^2$ permet de définir un compact, assurant l'existence de la fonction spline $S^j = B_j a^j$. Ce problème d'optimisation se résout algorithmiquement. Pour les démonstrations de la nature et de l'existence des solutions et de la convergence de l'algorithme se référer à (DURAND & SABATIER, 1997).

L'algorithme

étape 0 : prétraitement

Centrer X et Y .

Choisir le degré des splines d_j , le nombre L_j et les positions des L_j noeuds ξ_ℓ^j .

Déterminer les bases de B -splines associées aux prédicteurs x^j , donnant le codage B_j .

étape 1 : étape $a = 1$

Poser $S_1 = X$ et $u_1 = y^1$, première colonne de Y .

• Jusqu'à la convergence du critère faire :

• • une étape PLS (1 composante) de Y sur S_1

• • $S_1^j = \text{sgn}(\omega_1[j]) \frac{\|x^j\|_D}{\|P_{B_j} u_1\|} P_{B_j} u_1$

(• calcul non indispensable de $a_1^j = \text{sgn}(\omega_1[j]) (B_j' D B_j)^+ B_j' D u_1 \|x^j\| / \|P_{B_j} u_1\|_D$)

$t_1 = S_1 \omega_1$

étape 2 : étape $a > 1$

étape 2a : initialisation

$a = a + 1$ utiliser l'étape 1 pour initialiser ω_a , S_a^j et q_a

$u_a = Y q_a$, $t_a = S_a \omega_a$

étape 2b

Jusqu'à convergence du critère effectuer :

- Pour $\ell = 1, \dots, a-1$ $\beta_\ell = -\frac{1}{\|t_\ell\|_D^2} \sum_{j=1}^J \text{sgn}(\omega_a[j]) \omega_\ell[j] S_\ell^{j'} D S^j \|P_{B_j} v_a\|_D / \|x^j\|_D$

où $v_a = u_a$ lors du premier passage dans cette boucle, et $v_a = u_a - \sum_{\ell=1}^{a-1} \beta_\ell t_\ell$ sinon.

- $\omega_a = S'_a D (u_a - \sum_{\ell=1}^{a-1} \beta_\ell t_\ell)$
- $\omega_a = \omega_a - P_{S'_a D T_{a-1}} \omega_a$ où $T_{a-1} = [t_1, \dots, t_{a-1}]$; $\omega_a = \omega_a / \|\omega_a\|$
- $q_a = Y' D S_a \omega_a$; $q_a = q_a / \|q_a\|$
- $u_a = Y q_a$
- $t_a = S_a \omega_a$

étape 2c

tester la convergence du critère. Si non convergence aller en 2d. Sinon en 3.

étape 2d

Pour $\ell = 1, \dots, a-1$ $\beta_\ell = -\frac{1}{\|t_\ell\|_D^2} \sum_{j=1}^J \text{sgn}(\omega_a[j]) \omega_\ell[j] S_\ell^{j'} D S^j \|P_{B_j} v_a\|_D / \|x^j\|_D$

$$v_a = u_a - \sum_{\ell=1}^{a-1} \beta_\ell t_\ell$$

$$S_a^j = \text{sgn}(\omega_a[j]) P_{B_j} v_a \|x^j\|_D / \|P_{B_j} v_a\|_D$$

$$t_a = S_a \omega_a$$

$$S_a^j = S_a^j + \omega_a[j] B_j B_j' D \sum_{\ell=1}^{a-1} \gamma_\ell^j t_\ell \text{ où le vecteur } \gamma^j = (\gamma_1^j, \dots, \gamma_\ell^j)' \text{ est la}$$

$$\text{solution du système } T_{a-1}' D B_j B_j' D T_{a-1} \gamma^j = -\frac{1}{\omega_a[j]^2} T_{a-1}' D t_a$$

Retourner à l'étape 2.

étape 3 : passage à l'étape suivante

$Y = Y - P_{t_a} Y$ puis retourner en 2.

■

Dans l'étape 2b, le troisième calcul sert à orthogonaliser ω_a à l'espace engendré par les colonnes de $S'_a D T_{a-1}$. Ceci est dû à la contrainte d'orthogonalité des composantes t_a avec les précédentes (DURAND & SABATIER, 1997).

Les calculs des coefficients splines a_a^j sont omis dans l'algorithme car ils ne sont pas indispensables.

Le calcul de S_a^j (étape 1 et première mise à jour dans 2d) tiennent compte de la contrainte $\|S_a^j a_a^j\| = \|x^j\|$.

Les composantes $\{t_a\}_a$ sont D -orthogonales. De plus, $t_a = \sum_{j=1}^J S_a^j \omega_a[j] = \sum_{j=1}^J B_j a_a^j \omega_a[j]$. En notant $s_a^j(x^j) = B_j a_a^j$ la fonction spline optimale qui transforme le prédicteur x^j à l'étape a , on a $t_a = \sum_{j=1}^J \omega_a[j] s_a^j(x^j)$. t_a est une somme de fonctions coordonnées qui vont servir à mesurer l'influence d'un prédicteur sur la composante t_a , au travers du graphique $(x^j, \omega_a[j] s_a^j(x^j))$. La modélisation d'ordre A de la variable m de Y , y^m est donnée par $\hat{y}^{m,A} = P_{t_a} y^{m,(A-1)} = \sum_{j=1}^J \varphi_j^m(x^j)$ où $\varphi_j^m(x^j) = B^j (\sum_{a=1}^A \omega_a[j] \frac{t_a^j}{\|t_a\|_D^2} D Y^{m,(a-1)} a_a^j)$ est une fonction spline coordonnée. Celles-ci permettent donc de déterminer quelles sont les variables x^j les plus influentes non-linéairement sur la réponse y^m .

1.7 Régression PLS Spline

La régression PLSS (PLS Spline) (DURAND, 1997, 1999, 2001) est une extension de la régression PLS linéaire au cas non-linéaire, basée sur le recodage des prédicteurs. Ici, ce n'est pas la liaison interne entre les composantes PLS t_a et u_a qui est délinéarisée, comme dans la plupart des méthodes de régression PLS non-linéaires.

Ici, chaque variable x^j de la matrice X centrée et réduite sont codées dans la base, dans l'espace des fonctions splines de dimension $d_j + 1 + L_j$, de B -splines de degré d_j , avec L_j noeuds ξ_j intérieurs à l'étendue de x^j . Ce codage est désigné par une matrice B_j de dimension $l \times (d_j + 1 + L_j)$ (voir annexe B).

$B = [B_1 \dots B_J]$, matrice de taille $l \times \sum_{j=1}^J (d_j + 1 + L_j)$, contient les codages de chaque prédicteur x^j dans la base des B -splines.

Alors, la régression PLSS consiste à effectuer la régression PLS de Y sur B centrée pour D . Ainsi, le modèle d'ordre A pour la $m^{\text{ième}}$ variable y^m de Y s'écrit de la forme $\hat{y}^{m^A} = \sum_{j=1}^J B_j \hat{\beta}^{m^A}[j] = \sum_{j=1}^J s_m^A(x_j)$. $s_m^A(x_j)$ est la fonction spline coordonnée associée à x^j . Ainsi, chaque variable réponse est modélisée par une somme de transformations splines des prédicteurs.

Dans les autres généralisations de PLS au cas non-linéaire, la composante t_a est toujours une combinaison linéaire des variables initiales $X\omega_a^*$. Ici, t_a est une somme de transformations splines des variables initiales : $t_a = B\omega_a^* = \sum_{j=1}^J B^j \omega_a^*[j] = \sum_{j=1}^J \phi_a^j(x^j)$.

Le graphique $(x^j, \phi_a^j(x^j))$ est alors utilisé pour interpréter l'influence du prédicteur x^j sur la composante t_a . De même, le graphique $(x^j, s_m^A(x_j))$ est utilisé pour interpréter l'influence non-linéaire et additive du prédicteur x^j sur la réponse y^m .

Il n'existe pas encore de technique précise permettant un choix optimal de la position des noeuds, du nombre de noeuds et du degré des splines, qui doivent être choisis avant d'effectuer la régression PLS. En général, comme dans toutes les méthodes mettant en jeu des splines, on commence par un degré faible et un, deux ou trois noeuds. La position des noeuds dans le programme de l'auteur est telle qu'ils soient équi-espacés.

Toutefois, citons les travaux de MOLINARI (2000) dans lesquels il propose un algorithme appelé MBOK (Multivariate Bounded Optimal Knots) permettant de choisir la position optimale des noeuds, ainsi que le degré des splines, pour le cas de la régression multiple splines ordinaire. L'auteur signale que la méthode est surtout adaptée pour les petits jeux de données. Son algorithme pourrait peut être s'adapter ici.

Malgré la forte augmentation de dimension des «prédicteurs», la régression PLS utilisée sur le codage spline engendre des modèles nécessitant beaucoup moins de composantes pour être stables que dans la régression PLS usuelle.

Les performances de l'ASPLS (§1.6) et de PLSS sont comparées sur un exemple dans (DURAND, 1997, 1999). L'auteur conseille l'utilisation de PLSS en général et l'utilisation d'ASPLS seulement si PLSS ne donne pas d'assez bons résultats, car ASPLS étant une mé-

thode itérative, elle peut être assez longue à exécuter. D'autres applications sont présentes dans (DURAND, 2000, 2001).

1.8 GIFI PLS

GIFI-PLS (ERIKSSON, JOHANSSON, LINDGREN & WOLD, 2000; BERGLUND, KETTANEH-WOLD, UPPGÅRD, WOLD, BENDWELL & CAMERON, 2001) consiste à adopter l'approche de GIFI (GIFI, 1990; MICHAILIDIS & DE LEEUW, 1998) pour étudier la non-linéarité, présente dans les prédicteurs X , qui explique Y .

L'approche GIFI est basée sur le codage «flou» des données (initialement qualitatives (MICHAILIDIS & DE LEEUW, 1998)). Comme le notent Michailidis et De Leeuw, toute variable continue peut être vue comme une variable qualitative ayant de nombreuses classes. Aussi, cette approche a été beaucoup utilisée, notamment en ACP pour étudier les non-linéarités présentes dans les données.

Tout naturellement, ERIKSSON et al. (2000) ont défini la méthode GIFI-PLS qui consiste à effectuer la régression PLS linéaire usuelle des réponses Y sur le tableau des variables explicatives transformées dans un codage disjonctif complet.

Le nombre L_j de classes par variables ne doit pas être trop grand car il doit y avoir un nombre raisonnable d'observations par classe. ERIKSSON et al. (2000) ont choisi de prendre L_j tel que $L_j \leq l/10$. Ainsi, la régression PLS usuelle est effectuée sur le tableau de codage disjonctif complet de X , de taille $l \times \sum_{j=1}^J L_j$. Les auteurs proposent 4 exemples relevant du domaine de l'analyse QSAR (quantitative structure-activity relationship) en utilisant le même nombre de classes par variables.

Le type de non-linéarité présente dans x^j influençant chaque variable réponse est étudié par la représentation en bâton des coefficients des modèles GIFI-PLS. Pour chaque variable x^j , il y a un bâton (coefficient) par classe. Si les coefficients sont tous «presque nuls», x^j n'est pas influente. Si les extrémités des bâtons forment une droite, l'influence de x^j est linéaire. D'une manière générale, les extrémités des bâtons définissent le type de non-linéarité et la hauteur des bâtons l'importance de l'influence de la variable (ou classe) dans la modélisation.

BERGLUND et al. (2001) testent quatre types de codages sur des données simulées et des données réelles et comparent les performances de la régression PLS linéaire, INLR (§1.5) et de GIFI-PLS au travers des quatre codages. Le codage GIFI1 est le codage disjonctif complet utilisé par ERIKSSON et al. (2000). Le codage GIFI2 est du même type que GIFI1 mais en utilisant la valeur de l'observation au lieu des «1». C'est-à-dire «GIFI2= X *GIFI1». Le codage GIFI3 d'une variable x^j engendre une matrice à $L_j - 1$ colonnes. Elle est constituée de 1 et de 0. Les colonnes représentent les bornes supérieures des classes, et la valeur 1 est prise lorsque $x^j(i)$ est supérieur à la borne de la classe. Ainsi, la matrice de codage est constituée de «1» dans sa partie triangulaire inférieure gauche. Enfin, le codage GIFI4 est donné par « $[x^j, x^j * GIFI3]$ ». La matrice correspondante présente L_j colonnes et contient les valeurs des observations et des 0. La dernière ligne ne contient que la valeur $x^j(l)$ et la première ligne contient $x^j(1)$ dans la première colonne et des 0 ailleurs.

GIFI-PLS semble donner de meilleurs résultats que INLR. Les modèles sont mieux expliqués

avec moins de composantes.

Les différents auteurs s'accordent pour dire que GIFI-PLS donne de bons résultats lorsque les prédicteurs sont corrélés et qu'il y a une non-linéarité marquée mais pas trop forte entre Y et X . S'ils y a beaucoup de variables et/ou d'individus il vaut mieux utiliser une autre méthode ainsi que lorsque les non-linéarités sont importantes. De plus, GIFI-PLS permet de modéliser des non-linéarités non continues, ce que ne permettent pas les autres méthodes.

1.9 Discussion

Nous avons présenté ici un ensemble de méthodes de régression non-linéaire multivariées basées de la régression PLS linéaire. La régression PLS qui possède de bonnes propriétés, puisqu'elle arrive à donner de bons modèles même lorsque les prédicteurs sont très corrélés, en présence de bruit, et lorsqu'il y a peu d'observations, a ainsi inspiré bon nombre de chercheurs, en particuliers chimiométriciens, pour créer des modèles non-linéaires ayant les propriétés de PLS.

Nous pouvons classer ces méthodes dans trois grands groupes. Le premier est celui représentant les méthodes estimant la relation interne $\hat{u} = f(t)$. La majorité des méthodes sont construites ainsi. WOLD et al. (1989) en ont été les initiateurs avec la PLS quadratique, qui consiste à supposer que f est polynômiale quadratique. BAFFI et al. (1999a) l'ont ensuite reprise en modifiant l'algorithme au niveau de la procédure d'actualisation du vecteur de poids ω . Pour des raisons de flexibilité insuffisante de QPLS, WOLD (1992) propose d'estimer f par une fonction spline (SPL-PLS). Dans ces deux méthodes des hypothèses sont faites sur la nature de f , ce qui peut ne pas convenir. Ainsi, FRANK (1990, 1995) utilise la procédure de lissage SMART pour estimer f (NLPLS). TAAVITSAINEN & KORHONEN (1992); HAARIO & TAAVITSAINEN (1994) utilisent le principe d'un continuum de modèles de régression pour utiliser deux nouveaux critères permettant d'optimiser f (CNLPLS). Cependant, ici, les suppositions sont nombreuses. Il faut émettre une hypothèse quant à la nature de f (les auteurs utilisent des fonctions exponentielle, quadratique et gaussienne) et choisir en plus du nombre d'étapes A , le coefficient α du continuum.

Il existe d'autres techniques que nous n'avons pas détaillées dans ce chapitre, qui trouvent leur place dans ce groupe. En effet, récemment, LI, MEI & CONG (1999) ont utilisé un algorithme génétique pour la résolution du lien non-linéaire entre t et u (NPLSNGA). Les réseaux de neurones ont aussi été beaucoup utilisés pour la modélisation de cette relation interne f . QIN & MACAVOY (1992) en sont les initiateurs (NNPLS). Cette approche a été reprise ensuite par MALTHOUSE, MAH & TAMHANE (1994); MALTHOUSE, TAMHANE & MAH (1997); BAFFI, MARTIN & MORRIS (1999b, 2000). Enfin, HASSEL, MARTIN & MORRIS (2002) proposent l'utilisation d'un nouvel algorithme, basé sur le critère de la variance réciproque, pour modéliser f .

Dans un deuxième groupe, nous pouvons classer les méthodes basées sur une «expansion» de la matrice des prédicteurs. Cela revient en fait à délinéariser la relation externe de PLS. C'est-à-dire que $t = G(X)\omega$ où G est une fonction matricielle. La première méthode en date, dont nous n'avons pas parlé, est connue dans la littérature sous le nom de CARSO (CLEMENTI, CRUCIANI, CURTI & SKAGERBERG, 1989). Elle consiste à introduire les prédicteurs

au carré ainsi que les termes d'interaction, au côté des variables initiales dans X , pour ensuite effectuer une régression PLS linéaire. Avec la même idée, BERGLUND & WOLD (1997) proposent de n'utiliser que les termes au carré et/ou au cube au côté des X . Leur justification est, qu'ainsi, les modèles des réponses sont tout de même fonction des termes croisés entre les variables latentes sous-jacentes aux prédicteurs. Le principal avantage est que la matrice étendue est beaucoup moins grosse qu'avec l'utilisation de tous les termes croisés entre prédicteurs. D'autres méthodes sont basées sur le codage (flou) des prédicteurs. (ERIKSSON et al., 2000; BERGLUND et al., 2001) utilisent le codage GIFI (GIFI-PLS). Cette méthode semble particulièrement utile lorsque la non-linéarité présente est modérée et/ou discontinue. DURAND (2001), avec PLSS, utilise le codage des prédicteurs dans la base de B-splines. Les modèles obtenus ne sont sensibles aux valeurs extrêmes que localement. Ces méthodes sont assez efficaces dans beaucoup de situations mais sont difficilement applicables si le nombre de prédicteurs est grand. Les modèles deviennent très compliqués à interpréter.

Enfin, notre troisième groupe correspond aux techniques utilisant un nouveau critère. DURAND & SABATIER (1997) maximisent le critère de la régression PLS tout en optimisant une liaison externe ($G(X)$) non-linéaire à l'aide de B-splines. Cette méthode possède les bonnes propriétés de PLS (optimalité des solutions, orthogonalité des t , décomposition de la variance de Y etc) sans étendre la matrice X . Pour ces raisons, nous avons essayé le généraliser l'AS-PLS au cas de multi-tableaux (voir chapitre suivant).

Une méthode dont nous n'avons pas discuté, mais beaucoup citée dans la littérature est la régression PLS quadratique de HÖSKULDSSON (1992b). Il utilise la procédure de modélisation de Heisenberg, connue sous le nom de «H-principe». La procédure de Heisenberg établit une démarche générale à suivre pour extraire un modèle des données. Cela consiste à maximiser « $\Delta(\textit{fit})/\Delta(\textit{variance})$ ». $\Delta(\textit{fit})$ représente ce qui améliore l'ajustement du modèle, et $\Delta(\textit{variance})$ représente la variance des coefficients du modèle. Il a aussi utilisé ce principe pour construire des modèles polynômiaux cubiques (HÖSKULDSSON, 1998).

Les généralisations non-linéaires sont nombreuses et la relation interne de PLS est très sollicitée pour cela. Celles-ci semblent être utilisées pour modéliser des non-linéarités fortes. D'une manière générale les méthodes utilisant les transformations splines sont plus flexibles et ainsi arrivent à mieux cerner la non-linéarité. Cependant, cela peut aussi être un inconvénient, avec le risque de sur-ajuster les données. Les modèles quadratiques sont moins flexibles mais non adaptés à tous types de non-linéarités. Les méthodes étendant la matrice des prédicteurs sont aussi moins flexibles, mais non applicables si le nombre de prédicteurs est important.

Finalement, aucune méthode n'est meilleure qu'une autre. L'efficacité d'une méthode dépend du type de non-linéarité, de son intensité, du nombre de variables et du bruit.

Chapitre 2

Modélisation non linéaire de plusieurs tableaux

Le chapitre précédent a donné un vue générale sur les méthodes PLS non linéaires entre deux tableaux. Nous nous intéressons maintenant au cas où l'on a plus de deux tableaux. Les méthodes sont peu nombreuses et nous n'en n'avons pas trouvé concernant la régression de multi-tableaux. Par contre, nous pouvons en citer trois concernant l'étude d'un ensemble de K tableaux ou d'un cube.

VAN DER BURG (1988); VAN DER BURG & DIJKSTERHUIS (1989); VAN DER BURG, DE LEEUW & DIJKSTERHUIS (1994) proposent un critère de type GIFI pour définir une analyse canonique généralisée, OVERALS, voir aussi MICHAÏLIDIS & DE LEEUW (1998). Cette méthode consiste en la recherche de composantes par tableau, combinaisons linéaires d'un codage des variables initiales, qui soient les plus proches d'une variable auxiliaire de norme unité. Comme les méthodes GIFI (MICHAÏLIDIS & DE LEEUW, 1998), OVERALS est dite non-linéaire, car les composantes par tableaux ne sont pas linéaires en les variables initiales, mais construites à partir d'un codage des variables. Il existent plusieurs méthodes de ce type qui sont des cas particuliers de OVERALS, dont on peut trouver les références dans (VAN DER BURG et al., 1994).

Récemment, deux publications sont apparues concernant les méthodes TUCKER3 et PARAFAC utilisant des contraintes non-linéaires. TIMMERMAN & KIERS (2002) proposent de délinéariser les modèles TUCKER3 (§I-1.1) et PARAFAC (§I-1.2) à l'aide de B-splines. Pour cela il proposent de contraindre les composantes associées au troisième mode, C , d'être la transformation par des fonctions splines des colonnes de \mathbf{X}_K , $K \times IJ$. C'est-à-dire que $C = \mathbf{B}W$, où \mathbf{B} est le codage de \mathbf{X}_K dans une base de B-splines, et W la matrice des coefficients splines optimale au sens d'un modèle de TUCKER3 ou PARAFAC. Les solutions sont obtenues par un algorithme de type moindres carrés alternés. Ils montrent que les modèles PARAFAC et TUCKER3 obtenus avec de telles contraintes sont identiques en commençant par ajuster \mathbf{X}_K par des transformations splines codées dans la même base, $S = \mathbf{B}U$, $K \times IJ$ avec $U = (\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\mathbf{X}_K$, et rechercher le modèle TUCKER3 ou PARAFAC sans contraintes de S . Ce résultat permet une mise en oeuvre rapide de cette méthode.

REIS & FERREIRA (2002) proposent un modèle PARAFAC avec des contraintes de lissage sur les composantes associées au premier mode du cube, A , utilisant les B-splines. Des splines de lissage (DE BOOR, 1978) codées dans une base de B-splines sont recherchées pour ajuster A .

A notre connaissance il n'existe pas de méthodes de régression non-linéaire multi-blocs. Dans ce chapitre, nous proposons une méthode Splines Additives en Régression Multi-blocs, SARMUB, généralisant à la fois l'ACIMOG-PLS (§I-5.6) au cas non-linéaire et l'ASPLS (§II-1.6) au cas de plus de deux tableaux. SARMUB, comme l'ACIMOG-PLS, possède plusieurs variantes. Une de ces variantes, SARMUB2 présente encore quelques problèmes non résolus que nous décrirons.

La dernière partie de ce chapitre expose trois approches de modélisation non-linéaire multi-tableaux utilisant les méthodes ACIMOG-PLS et ACIMO-PLS (§I-5.5).

2.1 Splines Additives en Régression Multi-Bloc : SARMUB

Ce chapitre regroupe deux méthodes de régression non-linéaire multi-blocs utilisant les B-splines. Nous les avons nommées Splines Additives en Régression Multi-block, noté en abrégé SARMUB. Ces méthodes sont les versions non-linéaires des méthodes de régression linéaire multi-blocs précédemment décrites ACIMOG-PLS1 et ACIMOG-PLS2. Nous verrons cela un peu plus loin. De la même façon que dans ACIMOG, la première étape ($a = 1$) est commune à SARMUB1 et 2. La différence se fait dans le choix de la déflation des variables réponses lors du passage à une étape d'ordre supérieur.

Les méthodes de régression non-linéaire spécialement conçues pour les multi-tableaux sont inexistantes. Le principe consiste actuellement à utiliser les carrés, les termes croisés, *etc*, d'une matrice de prédicteurs X et d'appliquer une méthode de modélisation multi-tableaux linéaire. Cela dit, de telles pratiques sont surtout énoncées dans la littérature mais peu appliquées. Seuls BERGLUND & WOLD (1997) le font avec la SPLS (§I-5.4).

Nous avons voulu ici développer une méthode spécifique à la régression non-linéaire multi-tableaux. Il est bien sûr possible de «délinéariser» la relation interne entre les super-composantes u_a et t_a issues de méthodes de régression linéaire multi-blocs (telles que celles décrites dans I-5), mais nous avons préféré développer une méthode basée sur l'optimisation d'un critère, à chaque étape, fournissant des modèles non-linéaires et optimisant une relation externe non-linéaire. Pour cette raison, nous avons choisi d'adapter la technique ASPLS (§1.6) à l'ACIMOG-PLS. Ceci revient à généraliser la méthode ASPLS au cas de plusieurs tableaux, c'est-à-dire, utiliser le codage par des transformations splines des prédicteurs. Nous utilisons le même critère que celui l'ACIMOG où les matrices de prédicteurs sont remplacées par les matrices S_k , matrices de codage splines. On note $S_k = \begin{bmatrix} B_k^1 a_k^1 & \cdots & B_k^{J_k} a_k^{J_k} \end{bmatrix}$, où B_k^j , pour $j = 1, \dots, J_k$ est la matrice de codage de la variable (colonne) j du tableau de prédicteur X_k dans une base de B-splines normalisées, et a_k^j le vecteur des coefficients splines.

Pour chaque prédicteurs x_k^j , sont utilisés L_k^j noeuds intérieurs ξ_k^j , et des splines de degré d_k^j pour construire la base de B-splines. Ainsi la matrice B_k^j est de dimension $l \times (L_k^j + d_k^j + 1)$.

2.1.1 La première étape, $a = 1$

Définition 2.1 L'étape 1 de SARMUB est définie par la maximisation de la fonction objectif :

$$\begin{aligned} \mathcal{C}_4(\{w_k\}, \{b_n\}, \{a_k^j\}) &= \sum_{k=1}^K \sum_{n=1}^N b'_n Y'_n D S_k w_k = \sum_{k=1}^K \sum_{n=1}^N \text{cov}(u_n, t_k) \\ &= \sum_{k=1}^K \sum_{n=1}^N \underbrace{b'_n Y'_n}_{u'_n} D \underbrace{\sum_{j=1}^{J_k} B_k^j a_k^j}_{S_k^j} w_k[j], \end{aligned}$$

sous les contraintes de normes :

$$b'_n b_n = 1, \quad (2.1)$$

$$w'_k w_k = 1, \quad (2.2)$$

$$S_k^{j'} D S_k^j = v_k^j, \quad (2.3)$$

où v_k^j désigne le carré de la norme de la j ème variable de X_k .

La contrainte (2.3) est nécessaire pour assurer l'existence de la solution $S_k^j = B_k^j a_k^j$. En effet, l'ensemble $\{S_k^j | S_k^{j'} D S_k^j = v_k^j\}$ définit un compact.

La méthode des multiplicateurs de Lagrange nous permet d'écrire le Lagrangien :

$$\begin{aligned} \mathcal{L}(\{w_k\}, \{b_n\}, \{a_k^j\}) &= \sum_{k=1}^K \sum_{n=1}^N b'_n Y'_n D \sum_{j=1}^{J_k} B_k^j a_k^j w_k[j] \\ &\quad - \sum_{k=1}^K \frac{\mu_k}{2} (w'_k w_k - 1) - \sum_{n=1}^N \frac{\lambda_n}{2} (b'_n b_n - 1) - \sum_{k=1}^K \sum_{j=1}^{J_k} \frac{\beta_k^j}{2} (a_k^{j'} B_k^{j'} D B_k^j a_k^j - v_k^j), \end{aligned}$$

où μ_k , $k = 1, \dots, K$, λ_n , $n = 1, \dots, N$ et β_k^j , $k = 1, \dots, K$, $j = 1, \dots, J_k$, sont les multiplicateurs associés aux contraintes (2.2), (2.1) et (2.3). Les points critiques sont solutions des équations normales :

$$\nabla_{w_k} \mathcal{L} = \sum_{n=1}^N S_k' D Y_n b_n - \mu_k w_k = 0, \quad k = 1, \dots, K \quad (2.4)$$

$$\nabla_{b_n} \mathcal{L} = \sum_{k=1}^K Y_n' D S_k w_k - \lambda_n b_n = 0, \quad n = 1, \dots, N \quad (2.5)$$

$$\nabla_{a_k^j} \mathcal{L} = w_k[j] \sum_{n=1}^N B_k^{j'} D Y_n b_n - \beta_k^j B_k^{j'} D B_k^j a_k^j = 0, \quad j = 1, \dots, J, k = 1, \dots, K \quad (2.6)$$

$$w'_k w_k - 1 = 0, \quad k = 1, \dots, K$$

$$b'_n b_n - 1 = 0, \quad n = 1, \dots, N$$

$$a_k^{j'} B_k^{j'} D B_k^j a_k^j - v_k^j = 0, \quad j = 1, \dots, J, \quad k = 1, \dots, K.$$

w_k et b_n se déterminent immédiatement en utilisant respectivement les équations (2.4) et (2.5) :

$$w_k = \frac{1}{\mu_k} \sum_{n=1}^N S'_k D Y_n b_n, \quad k = 1, \dots, K \quad (2.7)$$

$$b_n = \frac{1}{\lambda_n} \sum_{k=1}^K Y'_n D S_k w_k, \quad n = 1, \dots, N. \quad (2.8)$$

D'après la contrainte (2.2) et l'équation (2.7), $\mu_k^2 = \|\sum_{n=1}^N S'_k D Y_n b_n\|^2$, ce qui nous permet d'imposer à μ_k d'être positif et égal à $\|\sum_{n=1}^N S'_k D Y_n b_n\|$. De même, on impose à λ_n d'être positif et égal à $\|\sum_{k=1}^K Y'_n D S_k w_k\|$. De plus, en pré-multipliant respectivement les équations (2.4) et (2.5) par w'_k et b'_n et en utilisant les contraintes (2.2) et (2.1), nous obtenons :

$$\begin{aligned} \mu_k &= \sum_{n=1}^N b'_n Y'_n D S_k w_k = t'_k D \sum_{n=1}^N u_n = \text{cov}(t_k, \sum_{n=1}^N u_n), \quad k = 1, \dots, K \\ \lambda_n &= \sum_{k=1}^K b'_n Y'_n D S_k w_k = u'_n D \sum_{k=1}^K t_k = \text{cov}(u_n, \sum_{k=1}^K t_k), \quad n = 1, \dots, N. \end{aligned}$$

Déterminons maintenant les multiplicateurs β_k^j . Pré-multiplions (2.6) par $a_k^{j'}$:

$$a_k^{j'} \nabla_{a_k^j} \mathcal{L} = w_k[j] a_k^{j'} B_k^{j'} D \sum_{n=1}^N Y_n b_n - \beta_k^j a_k^{j'} B_k^{j'} D B_k^j a_k^j = 0 \quad (2.9)$$

$$= w_k[j] \mu_k w_k[j] - \beta_k^j v_k^j = 0, \quad \text{d'après (2.7) et (2.3)}. \quad (2.10)$$

D'où l'expression pour le multiplicateur :

$$\beta_k^j = w_k^2[j] \mu_k / v_k^j, \quad k = 1, \dots, K, \quad j = 1, \dots, J. \quad (2.11)$$

Comme μ_k est positif, β_k^j l'est aussi. Une autre expression pour μ_k se déduit de (2.10). En effet, celle-ci nous dit que $v_k^j \beta_k^j = \sum_{n=1}^N u'_n D B_k^j a_k^j w_k[j]$, d'où $\sum_{j=1}^J v_k^j \beta_k^j = \sum_{n=1}^N u'_n D t_k = \mu_k$.

Enfin, il reste à déterminer a_k^j . Pour cela, nous allons utiliser (2.6). Comme $B_k^{j'} D B_k^j$ n'est forcément inversible, nous ne pouvons pas utiliser son inverse, mais plutôt l'inverse de Moore-Penrose. Une solution pour a_k^j est donnée à partir de (2.6), d'après la définition de l'inverse généralisé de Moore-Penrose (voir par exemple MAGNUS & NEUDECKER, 1988), par :

$$a_k^j = \frac{1}{\beta_k^j} w_k[j] (B_k^{j'} D B_k^j)^+ B_k^j D \sum_{n=1}^N u_n, \quad j = 1, \dots, J, \quad k = 1, \dots, K. \quad (2.12)$$

Ce résultat, utilisé avec la contrainte (2.3), implique :

$$\begin{aligned}
\beta_k^{j2} &= \frac{w_k^2[j]}{v_k^j} \sum_{n=1}^N u'_n D B_k^j \underbrace{(B_k^{j'} D B_k^j)^+ B_k^{j'} D B_k^j (B_k^{j'} D B_k^j)^+}_{(B_k^{j'} D B_k^j)^+, \text{ par définition de}} B_k^{j'} D \sum_{n=1}^N u_n \\
&\quad \text{l'inverse de Moore-Penrose} \\
&= \frac{w_k^2[j]}{v_k^j} \sum_{n=1}^N u'_n D B_k^j (B_k^{j'} D B_k^j)^+ B_k^{j'} \sum_{n=1}^N u_n \\
&= \frac{w_k^2[j]}{v_k^j} \|P_{B_k^j} \sum_{n=1}^N u_n\|_D^2,
\end{aligned}$$

d'où,

$$\beta_k^j = \frac{|w_k[j]|}{\sqrt{v_k^j}} \|P_{B_k^j} \sum_{n=1}^N u_n\|_D, \quad j = 1, \dots, J, \quad k = 1, \dots, K, \quad (2.13)$$

car nous avons imposé précédemment à β_k^j d'être positif. Ainsi, S_k^j peut être déterminé sans calculer les vecteurs des coefficients splines a_k^j ni les multiplicateurs β_k^j :

$$\begin{aligned}
S_k^j &= B_k^j a_k^j = P_{B_k^j} \sum_{n=1}^N u_n w_k[j] / \beta_k^j \text{ d'après (2.12),} \\
&= \text{sgn}(w_k[j]) v_k^j \frac{P_{B_k^j} \sum_{n=1}^N u_n}{\|P_{B_k^j} \sum_{n=1}^N u_n\|_D}, \quad j = 1, \dots, J, k = 1, \dots, K. \quad (2.14)
\end{aligned}$$

Il reste à montrer que les points critiques théoriques sont bien des maxima locaux. Ceci se fera par la démonstration de la croissance stricte et de la convergence d'un algorithme itératif permettant de les approximer.

Notons que $\sum_{k=1}^K \mu_k = \sum_{n=1}^N \lambda_n$ est égal à l'optimum, comme dans l'ACIMOG. Comme tous les multiplicateurs μ_k et λ_n sont positifs, il est raisonnable de penser que les points critiques ainsi obtenus seront au maximum de la fonction objectif.

2.1.1.1 L'algorithme

Tout comme dans l'ASPLS, les solutions que nous cherchons ne sont pas celles d'un problème «simple», comme par exemple une recherche de vecteurs propres. Aussi, nous proposons un algorithme itératif pour les déterminer.

Celui-ci se décompose en deux parties. La première, consiste à calculer de manière itérative les objets b_n , w_k , μ_k et λ_n , avec les S_k , $k = 1, \dots, K$ fixés, c'est-à-dire à a_k^j , $j = 1, \dots, J$, $k = 1, \dots, K$ fixés. La deuxième partie consiste à actualiser les a_k^j grâce aux objets issus de la première partie. En effet, la fonction objectif peut s'écrire plus simplement $C_4(w, b, a_1^1, \dots, a_K^J) = b' Y D S(a_1^1, \dots, a_K^J) w$. Si b et w sont fixés à l'optimum, maximiser \mathcal{L} sur a_k^j seulement est identique à maximiser \mathcal{L} selon toutes les variables. De même, si a_k^j est fixé à l'optimum, maximiser C_4 dans les directions de b_n et w_k donne les mêmes solutions qu'en maximisant \mathcal{L} par rapport à toutes les variables.

La première partie, revient donc à effectuer la première étape de l'ACIMOG sur Y_1, \dots, Y_N et S_1, \dots, S_K , ce qui semble assez raisonnable et intuitif.

étape 0 : initialisation

Les tableaux X_k et Y_n sont D -centrés en colonnes et éventuellement réduits.

$$S_k = X_k; b_n = (1/\sqrt{M_n}, \dots, 1/\sqrt{M_n})$$

Première partie

étape 1 : étape 1 de l'ACIMOG

$$Z = [S_1 | \dots | S_K]$$

tant que $opt1 = \sum_{k=1}^K \mu_k = \sum_{k=1}^K w'_k Z'_k D \sum_{n=1}^N Y_n b_n$ n'est pas stable

faire :

$$\begin{aligned} w_k &= Z'_k D \sum_{n=1}^N Y_n b_n \\ \mu_k &= \|w_k\|; w_k = w_k / \mu_k \\ b_n &= Y'_n D \sum_{k=1}^K Z_k w_k \\ \lambda_n &= \|b_n\|; b_n = b_n / \lambda_n \end{aligned}$$

Seconde partie

étape 2 : calcul des β_k^j

$$\beta_k^j = w_k^2[j] \mu_k / v_k^j \text{ (équation (2.11) ou bien par (2.13)).}$$

étape 3 : calcul des vecteurs de coefficients splines

$$\begin{aligned} a_k^j &= \frac{1}{\beta_k^j} w_k[j] (B_k^{j'} D B_k^j)^+ B_k^j D \sum_{n=1}^N Y_n b_n, \quad j = 1, \dots, J, \quad k = 1, \dots, K \\ S_k^j &= B_k^j a_k^j \end{aligned}$$

étape 4 : actualisation des composantes et critère

$$\begin{aligned} t_k &= S_k w_k; u_n = Y_n b_n \\ opt2 &= \sum_{k=1}^K t'_k D \sum_{n=1}^N u_n \end{aligned}$$

étape 5 : test de la convergence relative

Si $|opt2 - opt1| \leq |opt2| \epsilon$, il y a convergence relative du critère et on passe à l'étape $a = 2$ de SARMUB.

S'il n'y a pas convergence, on retourne à l'étape 1.



L'initialisation de b_n (étape 0) proposée en est une parmi d'autres.

2.1.1.2 Monotonie et convergence

Propriété 2.1 *L'algorithme précédent est croissant et borné, donc convergent.*

Démonstration :

Lors de la première étape, le critère croît d'après la propriété de croissance et de convergence de l'algorithme de l'ACIMOG.

Montrons que l'actualisation de a_k^j , avec $a_{k'}^{j'}$ fixés pour $k' \neq k$ et $j' \neq j$. No-

tons $\bar{a}_k^j = \frac{w_k[j] \sqrt{v_k^j}}{|w_k[j]| \|P_{B_k^j} \sum_{n=1}^N u_n\|_D} (B_k^{j'} D B_k^j)^+ B_k^j D \sum_{n=1}^N u_n$ l'actualisation de a_k^j . Soit

$h_k^j(a) = \sum_{n=1}^N u_n' DB_k^j a$. On a la relation $C_4 = \sum_{k=1}^K \sum_{j=1}^{J_k} h_k^j(a_k^j)$. Si h_k^j croît quelque soient k et j alors le critère croît. Calculons l'accroissement de h_k^j .

$$\begin{aligned}
\Delta(h_k^j) &= h_k^j(\bar{a}_k^j) - h_k^j(a_k^j) \\
&= \sum_{n=1}^N u_n DB_k^j w_k[j] \left(\bar{a}_k^j - a_k^j \right) \\
&= \sum_{n=1}^N u_n DB_k^j w_k[j] \left(\operatorname{sgn}(w_k[j]) \frac{\sqrt{v_k^j}}{\|P_{B_k^j} \sum_{n=1}^N u_n\|_D} (B_k^{j'} DB_k^j)^+ B_k^{j'} D \sum_{n=1}^N u_n - a_k^j \right) \\
&= \sum_{n=1}^N u_n D |w_k[j]| \sqrt{v_k^j} \left(\frac{P_{B_k^j} \sum_{n=1}^N u_n}{\|P_{B_k^j} \sum_{n=1}^N u_n\|_D} - \operatorname{sgn}(w_k[j]) \sum_{n=1}^N u_n' D S_k^j / \sqrt{v_k^j} \right) \\
&= \left\| \sum_{n=1}^N u_n \right\|_D |w_k[j]| \sqrt{v_k^j} \left(R \left(\sum_{n=1}^N u_n, B_k^j \right) - r \left(\sum_{n=1}^N u_n, \operatorname{sgn}(w_k[j]) S_k^j \right) \right)
\end{aligned}$$

où $R \left(\sum_{n=1}^N u_n, B_k^j \right)$ désigne le coefficient de corrélation multiple entre $\sum_{n=1}^N u_n$ et l'espace engendré par B_k^j . Par définition, celui-ci est supérieur à la corrélation entre $\sum_{n=1}^N u_n$ et tout vecteur appartenant à l'image de B_k^j , et, en particulier, il est supérieur à $r \left(\sum_{n=1}^N u_n, \operatorname{sgn}(w_k[j]) S_k^j \right)$. L'accroissement $\Delta(h_k^j)$ est donc positif et h_k^j croît.

De plus, $h_k^j(a_k^j) = \left\| \sum_{n=1}^N u_n \right\|_D |w_k[j]| \sqrt{v_k^j} r \left(\sum_{n=1}^N u_n, \operatorname{sgn}(w_k[j]) S_k^j \right)$ implique

$$-\left\| \sum_{n=1}^N u_n \right\|_D |w_k[j]| \sqrt{v_k^j} \leq h_k^j(a_k^j) \leq \left\| \sum_{n=1}^N u_n \right\|_D |w_k[j]| \sqrt{v_k^j}.$$

$h_k^j(a_k^j)$ est bornée, et donc la fonction objectif aussi. ■

Ainsi, le critère C_4 , que nous cherchons à maximiser est borné et croît à chaque passage dans une itération de l'algorithme. Nous sommes donc sûrs d'atteindre un maximum local si ce n'est global.

2.1.1.3 Passage à l'étape suivante

Nous voulons maintenant chercher d'autres super-composantes t_a expliquant une partie de l'inertie des Y_n non encore expliquée par les composantes des étapes précédentes. Comme dans les méthodes PLS, l'idée est de projeter les variables réponses Y_n sur l'espace engendré par les composantes explicatives précédentes. Nous ne faisons pas de déflations des B_k^j . Nous voulons conserver ce codage des prédicteurs dans une base de B-splines. Si la déflation était faite, les S_k^j ne représenteraient plus une transformation de la variable x_k^j par une spline. Il faut ici rajouter une contrainte d'orthogonalité entre la composante t_a cherchée et les précédentes.

SARMUB étant définie comme une généralisation de l'ACIMOG-PLS, nous avons tenté de résoudre le problème dans les trois situations proposées dans L'ACIMOG-PLS, à savoir :

1. projection des variables sur l'espace engendré par les composantes partielles explicatives $T = [t_1 | \dots | t_K]$.

2. projection des variables sur la super-composante explicative $t = \sum_{k=1}^K t_k$.

3. projection des variables sur la super-composante explicative $t = \sum_{k=1}^K \text{cov}(t_k, \sum_{n=1}^N u_n) t_k$.

Les deux premiers cas ont conduit à la définition de SARMUB1 et SARMUB2. Tandis que nous n'avons pas réussi à résoudre le dernier. L'expression de la super-composante $t_a = \sum_{k=1}^K \text{cov}(t_k, \sum_{n=1}^N u_n) t_k$ complique énormément les calculs au niveau de la contrainte d'orthogonalité 2.31 (voir la description de SARMUB2). Nous n'avons pas réussi à résoudre les équations normales.

2.1.2 SARMUB1, étapes $a > 1$

Définition 2.2 L'étape a , $1 < a \leq A$, de SARMUB1 est définie par la maximisation de la fonction objectif $C_5(w_k, b_n, a_k^j) = \sum_{n=1}^N \underbrace{b'_n Y_n^{(a-1)'}}_{u_n} D \sum_{k=1}^K \underbrace{\sum_{j=1}^{J_k} B_k^j a_k^j w_k[j]}_{t_k}$, avec $Y_n^{(a-1)} =$

$P_{T_{a-1}}^\perp Y_n^{(a-2)}$ sous les contraintes :

$$w'_k w_k = 1, \quad k = 1, \dots, K \quad (2.2)$$

$$b'_n b_n = 1, \quad n = 1, \dots, N \quad (2.1)$$

$$S_k^{j'} S_k^j = v_k^j, \quad k = 1, \dots, K, \quad j = 1, \dots, J \quad (2.3)$$

$$T' D T_\ell = 0, \quad \ell = 1, \dots, a-1, \quad (2.15)$$

avec $T = [t_1 | \dots | t_K]$.

Rappelons que la contrainte (2.15) impose à chaque composante partielle t_k d'expliquer une partie de l'information non encore expliquée par toutes les composantes partielles des étapes précédentes. Il en découle que les K composantes $t_{k,a}$ vont être orthogonales à l'ensemble des composantes partielles précédentes. Nous sommes obligés d'imposer cette orthogonalité pour avoir la décomposition de la variance totale de Y_n , car nous ne projetons pas la matrice B_k^j sur le sous espace de \mathbb{R}^1 engendré par les colonnes de T .

2.1.2.1 Les solutions

Le Lagrangien associé à ce problème s'écrit :

$$\begin{aligned} \mathcal{L}(\{w_k\}, \{b_n\}, \{a_k^j\}) &= \sum_{n=1}^N \sum_{k=1}^K b'_n Y_n' D \sum_{j=1}^{J_k} B_k^j a_k^j w_k[j] - \sum_{k=1}^K \frac{\mu_k}{2} (w'_k w_k - 1) \\ &- \sum_{n=1}^N \frac{\lambda_n}{2} (b'_n b_n - 1) - \sum_{\ell=1}^{a-1} \underbrace{\text{tr}(\Gamma'_\ell T'_\ell D T)}_{(*)} - \sum_{k=1}^K \sum_{j=1}^{J_k} \frac{\beta_k^j}{2} (a_k^{j'} B_k^{j'} D B_k^j a_k^j - v_k^j) \end{aligned}$$

Les coefficients μ_k , λ_n , β_k^j , et les matrices Γ_l sont les multiplicateurs de Lagrange associés aux contraintes. Γ_ℓ est une matrice de taille $\mathbf{K} \times \mathbf{K}$. Notons $\Gamma'_\ell = [\gamma_{1,\ell} \dots \gamma_{\mathbf{K},\ell}]$. Avant d'établir les équations normales, nous allons réécrire (★) comme une somme multiple, afin que la dérivation soit plus facile à réaliser :

$$\Gamma'_\ell T'_\ell DT = [\Gamma'_\ell T'_\ell Dt_1 | \dots | \Gamma'_\ell T'_\ell Dt_{\mathbf{K}}] \quad (2.16)$$

$$= \left[\sum_{r=1}^{\mathbf{K}} \gamma_{r,\ell} t'_{r,\ell} Dt_1 | \dots | \sum_{r=1}^{\mathbf{K}} \gamma_{r,\ell} t'_{r,\ell} Dt_{\mathbf{K}} \right] \quad (2.17)$$

$$= \left[\sum_{r=1}^{\mathbf{K}} \gamma_{r,\ell} [k] t'_{r,\ell} Dt_{k'} \right]_{k,k'} \quad (2.18)$$

pour $k = 1, \dots, \mathbf{K}$ et $k' = 1, \dots, \mathbf{K}$. D'où la trace :

$$tr(\Gamma'_\ell T'_\ell DT) = \sum_{k=1}^{\mathbf{K}} \sum_{r=1}^{\mathbf{K}} \gamma_{r,\ell} [k] t'_{r,\ell} Dt_k.$$

Ainsi, nous obtenons les équations normales :

$$\nabla_{w_k} \mathcal{L} = S'_k D \sum_{n=1}^{\mathbf{N}} u_n - \mu_k w_k - \sum_{\ell=1}^{a-1} \sum_{r=1}^{\mathbf{K}} S'_k D \gamma_{r,\ell} [k] t_{r,\ell} = 0 \quad (2.19)$$

$$\nabla_{b_n} \mathcal{L} = \sum_{k=1}^{\mathbf{K}} Y'_n D S_k w_k - \lambda_n b_n = 0 \quad (2.20)$$

$$\begin{aligned} \nabla_{a_k^j} \mathcal{L} = & w_k [j] B_k^{j'} D \sum_{n=1}^{\mathbf{N}} u_n - \beta_k^j B_k^j D B_k^j a_k^j \\ & - \sum_{\ell=1}^{a-1} \sum_{r=1}^{\mathbf{K}} w_k [j] B_k^{j'} D \gamma_{r,\ell} [k] t_{r,\ell} = 0 \end{aligned} \quad (2.21)$$

$$T' D T_\ell = 0, \quad \ell = 1, \dots, a-1,$$

$$w'_k w_k - 1 = 0, \quad k = 1, \dots, \mathbf{K}$$

$$b'_n b_n - 1 = 0, \quad n = 1, \dots, \mathbf{N}$$

$$a_k^{j'} B_k^{j'} D B_k^j a_k^j - v_k^j = 0, \quad j = 1, \dots, \mathbf{J}, \quad k = 1, \dots, \mathbf{K}.$$

w_k et b_n s'obtiennent immédiatement à partir des équations (2.19) et (2.20) :

$$w_k = \frac{1}{\mu_k} S'_k D \underbrace{\left[\sum_{n=1}^{\mathbf{N}} u_n - \sum_{\ell=1}^{a-1} \sum_{r=1}^{\mathbf{K}} \gamma_{r,\ell} [k] t_{r,\ell} \right]}_{g_k} \quad (2.22)$$

$$b_n = \frac{1}{\lambda_n} \sum_{k=1}^{\mathbf{K}} Y'_n D S_k w_k. \quad (2.23)$$

Ces deux équations utilisées avec les contraintes (2.2) et (2.1) impliquent :

$$\mu_k^2 = \|S'_k Dg_k\|^2 \quad (2.24)$$

$$\lambda_n^2 = \left\| \sum_{n=1}^N Y'_n D t_k \right\|^2, \quad (2.25)$$

ce qui nous amènera à imposer à μ_k et λ_n d'être positifs et respectivement égaux à $\|S'_k Dg_k\|$ et $\left\| \sum_{n=1}^N Y'_n D t_k \right\|$.

En pré-multipliant les équations (2.19) et (2.20) respectivement par w'_k et b'_n et en utilisant les contraintes (2.2) et (2.1), on obtient les expressions de μ_k et λ_n :

$$\begin{aligned} \mu_k &= \left[\sum_{n=1}^N u'_n - \sum_{l=1}^{a-1} \sum_{r=1}^K \gamma_{r,l}[k] t'_{r,l} \right] D S_k w_k \\ &= g'_k D t_k \end{aligned} \quad (2.26)$$

$$\lambda_n = u'_n D \sum_{k=1}^K t_k. \quad (2.27)$$

On retrouve toujours que $\sum_{n=1}^N \lambda_n = \sum_{k=1}^K \mu_k$ est égal à l'optimum du critère. Pour ce qui est de $\sum_{n=1}^N \lambda_n$, le résultat est trivial. Quant à $\sum_{k=1}^K \mu_k$, le résultat est obtenu aisément en remarquant que $\sum_{k=1}^K \sum_{l=1}^{a-1} \sum_{r=1}^K \gamma_{r,l}[k] t'_{r,l} D t_k = \text{tr}(\Gamma_l T'_l D T)$ qui est nul d'après la contrainte d'orthogonalité (2.15).

La procédure pour calculer a_k^j et β_k^j est la même que celle utilisée pour $a_{k,1}^j$ et $\beta_{k,1}^j$. On trouve :

$$\begin{aligned} \beta_k^j &= g'_k D S_k^j w_k[j] / v_k^j, \\ &= \mu_k w_k^2[j] / v_k^j, \\ &= \sqrt{\frac{w_k^2[j] \|P_{B_k^j} g_k\|_D^2}{v_k^j}} \end{aligned} \quad (2.28)$$

$$\begin{aligned} a_k^j &= \frac{1}{\beta_k^j} (B_k^{j'} D B_k^j)^+ B_k^{j'} D g_k w_k[j], \\ &= \frac{v_k^j}{\mu_k w_k[j]} (B_k^{j'} D B_k^j)^+ B_k^{j'} D g_k. \end{aligned} \quad (2.29)$$

(2.28) nous amène à choisir β_k^j positif et égal à $\frac{|w_k[j]| \|P_{B_k^j} g_k\|_D}{\sqrt{v_k^j}}$.

Passons maintenant au calcul de Γ_ℓ qui est le plus compliqué.

Soit s appartenant à $\{1, 2, \dots, a-1\}$. Après avoir divisé (2.21) par $w_k[j]$, puis multiplié à

gauche par $a_{k,s}j'w_{k,s}[j]$ et enfin sommé sur l'indice j , nous obtenons :

$$\begin{aligned} \sum_{\ell=1}^{a-1} \sum_{r=1}^K \sum_{j=1}^{J_k} w_{k,s}[j] a_{k,s}^{j'} B_k^{j'} D \gamma_{r,\ell}[k] t_{r,\ell} &= \\ \sum_{j=1}^{J_k} w_{k,s}[j] a_{k,s}^{j'} B_k^{j'} D \sum_{n=1}^N u_n - \sum_{j=1}^{J_k} \frac{\beta_k^j}{w_k[j]} w_{k,s}[j] a_{k,s}^{j'} B_k^{j'} D B_k^j a_k^j, \\ \sum_{\ell=1}^{a-1} \sum_{r=1}^K \gamma_{r,\ell}[k] t'_{k,s} D t_{r,\ell} &= t'_{k,s} D \sum_{n=1}^N u_n - \sum_{j=1}^{J_k} \frac{\beta_k^j}{w_k[j]} w_{k,s}[j] S_{k,s}^{j'} D S_k^j. \end{aligned}$$

Or, d'après la contrainte (2.15), $t'_{k,s} D t_{r,\ell}$ est nul pour tout $\ell \neq s$. De plus, $t'_{k,s} D \sum_{n=1}^N u_n = 0$. En effet, $u_n = P_{T_{a-1}}^\perp \dots P_{T_1}^\perp Y_n b_n$. D'où, $T'_s D u_n = T'_s D P_{T_s}^\perp \dots Y_n b_n = 0$, car T_s est orthogonal à toutes les matrices T_ℓ , pour $\ell \neq s$.

D'où :

$$\begin{aligned} \sum_{r=1}^K \gamma_{r,s}[k] t'_{k,s} D t_{r,s} &= - \sum_{j=1}^{J_k} \frac{\beta_k^j}{w_k[j]} w_{k,s}[j] S_{k,s}^{j'} D S_k^j, \\ t'_{k,s} D T_\ell \gamma_{k,s}^* &= - \sum_{j=1}^{J_k} \frac{\beta_k^j}{w_k[j]} w_{k,s}[j] S_{k,s}^{j'} D S_k^j, \end{aligned}$$

où $\gamma_{k,s}^*$ désigne la $k^{\text{ième}}$ colonne de Γ_s . Multiplions à gauche la dernière équation par $T'_s D t_{k,s}$:

$$T'_s D t_{k,s} t'_{k,s} D T_s \gamma_{k,s}^* = - \sum_{j=1}^{J_k} \frac{\beta_k^j}{w_k[j]} w_{k,s}[j] T'_s D t_{k,s} S_{k,s}^{j'} D S_k^j.$$

Nous avons un système de la forme $U = Vx$ à résoudre. La matrice $V = T'_s D t_{k,s} t'_{k,s} D T_s$ est de rang 1 donc non inversible. D'après la définition de l'inverse de Moore-Penrose, une solution pour $x = \gamma_{k,s}^*$ est donnée par $V^+ U$, soit :

$$\gamma_{k,s}^* = -(T'_s D t_{k,s} t'_{k,s} D T_s)^+ T'_s D t_{k,s} \sum_{j=1}^{J_k} \frac{\beta_k^j}{w_k[j]} w_{k,s}[j] S_{k,s}^{j'} D S_k^j, \quad s = 1, \dots, a-1.$$

Nous voyons donc que les points critiques de la fonction objectif ne sont pas solutions d'un système simple à résoudre. Il est nécessaire de construire un algorithme itératif pour les approximer.

2.1.2.2 L'algorithme

L'algorithme va se décomposer en deux parties, comme lors de la première étape. La première partie consiste à calculer les axes w_k et b_n et les multiplicateurs Γ_ℓ en fixant les vecteurs de coefficients splines a_k^j . La deuxième partie va consister à calculer ces coefficients avec w_k , b_n et Γ_ℓ fixés.

étape 0 : initialisation

L'étape 1 de SARMUB appliquée aux X_k et aux $Y_n^{(a-1)}$, nous donne une initialisation de w_k , b_n , μ_k , λ_n et S_k .

$$g_k = \sum_{n=1}^N u_n; \beta_k^j = g'_k D S_k^j w_k[j] / v_k^j.$$

Première partie**étape 1**

tant que $opt1 = \sum_{k=1}^K t'_k D \sum_{n=1}^N u_n$ ne converge pas,

faire :

$$\gamma_{k,s}^* = -(T'_s D t_{k,s} t'_{k,s} D T_s)^+ T'_s D t_{k,s} \sum_{j=1}^{J_k} \frac{\beta_k^j}{w_k[j]} w_{k,s}[j] S_{k,s}^{j'} D S_k^j, \quad s = 1, \dots, a-1$$

$$g_k = \sum_{n=1}^N u_n - \sum_{s=1}^{a-1} \sum_{r=1}^K \gamma_{k,s}^*[r] t_{r,s}$$

$$w_k = S'_k D g_k$$

$$\spadesuit w_k = w_k - P_{S'_k D} [T_1 \quad \dots \quad T_{a-1}] w_k$$

$$\mu_k = \|w_k\|$$

$$w_k = w_k / \mu_k$$

$$b_n = Y'_n D \sum_{k=1}^K S_k w_k; \lambda_n = \|b_n\|$$

$$b_n = b_n / \lambda_n$$

$$opt1 = \sum_{n=1}^N b'_n Y'_n D \sum_{k=1}^K S_k w_k$$

étape 2 : calcul des β_k^j

$$\beta_k^j = g'_k D S_k^j w_k[j] / v_k^j$$

étape 3 : calcul de Γ_s

$$\gamma_{k,s}^* = -(T'_s D t_{k,s} t'_{k,s} D T_s)^+ T'_s D t_{k,s} \sum_{j=1}^{J_k} \frac{\beta_k^j}{w_k[j]} w_{k,s}[j] S_{k,s}^{j'} D S_k^j, \quad s = 1, \dots, a-1$$

étape 4 : actualisation

$$g_k = \sum_{n=1}^N u_n - \sum_{s=1}^{a-1} \sum_{r=1}^K \gamma_{k,s}^*[r] t_{r,s}$$

$$\beta_k^j = g'_k D S_k^j w_k[j] / v_k^j$$

Seconde partie**étape 5 : calcul des coefficients splines**

$$a_k^j = \frac{1}{\beta_k^j} (B_k^{j'} D B_k^j)^+ B_k^{j'} D g_k w_k[j]$$

étape 6 : actualisation

$$S_k^j = B_k^j a_k^j; t_k = S_k w_k$$

$$\spadesuit \nu_k^j = W_k^{j+} l_k^j$$

$$\spadesuit a_k^j = a_k^j + w_k[j] B_k^{j'} D [T_1 \quad \dots \quad T_{a-1}] \nu_k^j$$

étape 7 : test de la convergence

$$opt2 = \sum_{k=1}^K t'_k D \sum_{n=1}^N u_n$$

Si $|opt2 - opt1| \leq opt2\epsilon$ on arrête.

Sinon, on retourne à l'étape 1.

Les calculs précédés de ♠ sont présentes pour assurer l'orthogonalité des matrices T_s (2.15). En effet, à chaque fois que les vecteurs de poids w_k et les coefficients splines sont actualisés, cela influe sur les composantes t_1, \dots, t_K recherchées et donc sur les orthogonalités imposées. Il faut ainsi tenir compte de la contrainte (2.15) dans le calcul de w_k et a_k^j .

Celle-ci peut s'écrire :

$$\begin{aligned} T'D [T_1 \ \cdots \ T_{a-1}] &= [0 \ \cdots \ 0] \\ \text{Diag}(w'_k) [S_1 \ \cdots \ S_K]' D [T_1 \ \cdots \ T_{a-1}] &= [0 \ \cdots \ 0]. \end{aligned}$$

Chaque ligne de la matrice $T'D [T_1 \ \cdots \ T_{a-1}]$ est un vecteur ligne nul de longueur $a - 1$. C'est-à-dire, pour tout k ,

$$\begin{aligned} [0 \ \cdots \ 0 \ w'_k \ 0 \ \cdots] [S_1 \ \cdots \ S_K]' D [T_1 \ \cdots \ T_{a-1}] &= 0 \\ [0 \ \cdots \ 0 \ w'_k S'_k D [T_1 \ \cdots \ T_{a-1}] \ 0 \ \cdots] &= 0 \end{aligned}$$

ce qui est équivalent à

$$w'_k S'_k D [T_1 \ \cdots \ T_{a-1}] = 0, \quad k = 1, \dots, K.$$

Cette équation nous dit que le vecteur w_k est orthogonal à l'espace engendré par toutes les composantes explicatives précédentes, soit par $[T_1 \ \cdots \ T_{a-1}]$. D'où l'actualisation de w_k par $P_{S'_k D [T_1 \ \cdots \ T_{a-1}]}^\perp w_k$ dans l'étape de l'algorithme.

Détaillons maintenant l'actualisation de a_k^j dans l'étape 5. En développant l'expression de la contrainte (2.15),

$$\begin{aligned} T'_\ell D \sum_{j=1}^{J_k} B_k^j a_k^j w_k[j] &= 0, \quad k = 1, \dots, K, \quad \ell = 1, \dots, a-1 \\ T'_\ell D B_k^j a_k^j w_k[j] &= - \sum_{j'=1; j' \neq j}^{J_k} T'_\ell D B_k^{j'} a_k^{j'} w_k[j'] \quad k = 1, \dots, K, \quad \ell = 1, \dots, a-1. \end{aligned}$$

Considérons que $j' \neq j$ sont fixés, et que nous cherchons à actualiser a_k^j , pour un certain k . Alors, en notant le vecteur ligne de constantes de longueur K (qui dépend aussi de j et de l'étape courante a),

$$c_{k,\ell} = - \sum_{j'=1; j' \neq j}^{J_k} a_k^{j'} B_k^{j'} D T_\ell w_k[j']$$

et la matrice de dimension $(L_k^j + 1 + d_k^j) \times K(a-1)$ (qui dépend aussi de l'étape courante a et j) $V_k = [w_k[j] B_k^j D T_1 \ \cdots \ w_k[j] B_k^j D T_{a-1}]$, nous devons résoudre l'équation matricielle suivante :

$$a_k^j V_k = [c_{k,1} \ \cdots \ c_{k,a-1}] = \mathbf{c}_k \quad (2.30)$$

Pour vérifier l'orthogonalité de la matrice T_a avec les précédentes, T_ℓ , $a_{k,a}^j$ doit vérifier les $K(a-1)$ équations

$$a_k^j v_k^s = \mathbf{c}_k(s),$$

v_k^s désignant la $s^{\text{ième}}$ colonne de V_k et $\mathbf{c}_k(s)$ le $s^{\text{ième}}$ élément du vecteur $\mathbf{c}_k = [c_{k,1} \ \dots \ c_{k,a-1}]$, s variant de 1 à $K(a-1)$. Ainsi, il doit appartenir à une intersection de $K(a-1)$ hyperplans \mathcal{H}_s de $\mathbb{R}^{L_k^j+1+d_k^j}$, chacun des hyperplans \mathcal{H}_s étant orthogonal à la colonne $v_{k,s}$. En effet, les vecteurs x de \mathbb{R}^n solutions de l'équation $x'u = cste$ définissent un hyperplan de \mathbb{R}^n orthogonal à u . La projection orthogonale d'un vecteur quelconque y de $\mathbb{R}^{L_k^j+1+d_k^j}$ sur l'intersection des hyperplans \mathcal{H}_s , est donnée par l'addition de y avec une combinaison linéaire des vecteurs colonnes de V_k . Ainsi, pour actualiser a_k^j , nous allons chercher le vecteur ν_k tel que $a_k^j + w_k[j]B_k^{j'} D [T_1 \ \dots \ T_{a-1}] \nu_k = a_k^j + V_k \nu_k$ appartienne à l'intersection des \mathcal{H}_s , c'est-à-dire soit solution de l'équation (2.30) :

$$\begin{aligned} a_k^{j'} V_k + w_k[j] \nu_k' [T_1 \ \dots \ T_{a-1}]' D B_k^j &= - \sum_{j'=1, j' \neq j}^{J_k} a_k^{j'} B_k^{j'} D [T_1 \ \dots \ T_{a-1}] w_k[j'], \\ w_k[j] a_k^{j'} B_k^{j'} D [T_1 \ \dots \ T_{a-1}] + w_k^2[j] \nu_k' [T_1 \ \dots \ T_{a-1}]' D B_k^j B_k^{j'} D [T_1 \ \dots \ T_{a-1}] &= \\ &- \sum_{j'=1, j' \neq j}^{J_k} a_k^{j'} B_k^{j'} D [T_1 \ \dots \ T_{a-1}] w_k[j'], \\ w_k^2[j] \nu_k' [T_1 \ \dots \ T_{a-1}]' D B_k^j B_k^{j'} D [T_1 \ \dots \ T_{a-1}] &= \\ &- \sum_{j=1}^{J_k} a_k^{j'} B_k^{j'} D [T_1 \ \dots \ T_{a-1}] w_k[j], \\ \nu_k &= - \frac{1}{w_k^2[j]} \left([T_1 \ \dots \ T_{a-1}]' D B_k^j B_k^{j'} D [T_1 \ \dots \ T_{a-1}] \right)^+ [T_1 \ \dots \ T_{a-1}]' D t_k \\ \nu_k &= W_k^{j+} l_k^j. \end{aligned}$$

Lorsque w_k et t_k ont été actualisés à l'issue de la première partie de l'algorithme, seuls les vecteurs $l_k^j = -\frac{1}{w_k^2[j]} [T_1 \ \dots \ T_{a-1}]' D t_k$ changent. Lorsque la composante t_k devient orthogonale à toutes les composantes précédentes $[T_1 \ \dots \ T_{a-1}]$, ν_k devient nul, et par conséquent a_k^j devient fixe. Au fur et à mesure des itérations de l'algorithme, le vecteur ν_k évolue dans $Im(W_k^{j+})$ jusqu'à devenir le vecteur nul. Alors, les solutions optimales sont atteintes.

2.1.3 SARMUB2, étapes $a > 1$

Définition 2.3 *L'étape a de SARMUB2 est définie par la maximisation de la fonction objectif $C_5(w_k, b_n, a_k^j) = \sum_{n=1}^N b'_n Y_n^{(a-1)'} D \sum_{k=1}^K \sum_{j=1}^{J_k} B_k^j a_k^j w_k[j]$, avec $Y_n^{(a-1)} = P_{t_{a-1}}^\perp Y_n^{(a-2)}$ sous les contraintes :*

$$w'_k w_k = 1, \quad k = 1, \dots, K \quad (2.2)$$

$$b'_n b_n = 1, \quad n = 1, \dots, N \quad (2.1)$$

$$S_k^{j'} S_k^j = v_k^j, \quad k = 1, \dots, K, \quad (2.3)$$

$$t' D t_l = 0, \quad l = 1, \dots, a-1, \quad (2.31)$$

avec $t = \sum_{k=1}^K t_k$.

La démarche pour déterminer les points critiques est exactement la même que pour SARMUB1.

2.1.3.1 Les solutions

Le Lagrangien s'écrit :

$$\begin{aligned} \mathcal{L}(\{w_k\}, \{b_n\}, \{a_k^j\}) &= \sum_{n=1}^N \sum_{k=1}^K b'_n Y'_n D \sum_{j=1}^{J_k} B_k^j a_k^j w_k[j] - \sum_k \frac{\mu_k}{2} (w'_k w_k - 1) \\ &- \sum_{n=1}^N \frac{\lambda_n}{2} (b'_n b_n - 1) - \sum_{l=1}^{a-1} \alpha_l t'_l D t - \sum_{k=1}^K \sum_{j=1}^{J_k} \frac{\beta_k^j}{2} (a_k^{j'} B_k^{j'} D B_k^j a_k^j - v_k^j) \end{aligned} \quad (2.32)$$

avec $t_l = \sum_{k=1}^K t_{k,a}$ pour $l = 1, \dots, a-1$.

Les équations normales permettant de calculer les points critiques sont :

$$\nabla_{w_k} \mathcal{L} = S'_k D \sum_{n=1}^N u_n - \mu_k w_k - \sum_{l=1}^{a-1} \alpha_l S'_k D t_l = 0 \quad (2.33)$$

$$\nabla_{b_n} \mathcal{L} = \sum_{k=1}^K Y'_n D S_k w_k - \lambda_n b_n = 0 \quad (2.34)$$

$$\nabla_{a_k^j} \mathcal{L} = w_k[j] B_k^{j'} D \sum_{n=1}^N u_n - \beta_k^j B_k^{j'} D B_k^j a_k^j - \sum_{\ell=1}^{a-1} \alpha_\ell w_k[j] B_k^{j'} D t_\ell = 0 \quad (2.35)$$

$$t'_l D t_\ell = 0, \quad \ell = 1, \dots, a-1,$$

$$w'_k w_k - 1 = 0, \quad k = 1, \dots, K$$

$$b'_n b_n - 1 = 0, \quad n = 1, \dots, N$$

$$a_k^{j'} B_k^{j'} D B_k^j a_k^j - v_k^j = 0, \quad j = 1, \dots, J, \quad k = 1, \dots, K.$$

w_k et b_n s'obtiennent directement des équations (2.33) et (2.34) :

$$\begin{aligned} w_k &= \frac{1}{\mu_k} S'_k D \left[\underbrace{\sum_{n=1}^N u_n - \sum_{l=1}^{a-1} \alpha_l t_l}_g \right] \\ b_n &= \frac{1}{\lambda_n} \sum_{k=1}^K Y'_n D S_k w_k. \end{aligned}$$

En procédant de la même manière que pour SARMUB1, nous trouvons :

$$\begin{aligned}
\mu_k &= t'_k Dg \\
&= \|S'_k Dg\| \\
\lambda_n &= b'_n Y'_n D \sum_{k=1}^K t_k \\
&= \|Y'_n Dt\| \\
\beta_k^j &= \mu_k w_k^2[j] / v_k^j \\
&= \|P_{B_k^j} g\|_D |w_k[j]| / \sqrt{v_k^j} \geq 0 \\
a_k^j &= \frac{1}{\beta_k^j} (B_k^{j'} DB_k^j)^+ B_k^{j'} Dg \\
&= \frac{\text{sgn}(w_k[j])}{\|P_{B_k^j} g\|_D} \sqrt{v_k^j} (B_k^{j'} DB_k^j)^+ B_k^{j'} Dg \\
\alpha_l &= -\frac{1}{\|t_l\|_D^2} \sum_{k=1}^K \sum_{j=1}^{J_k} \beta_k^j \frac{w_{k,l}[j]}{w_k[j]} S_{k,l}^{j'} DS_k^j.
\end{aligned}$$

Nous pouvons noter deux différences entre SARMUB1 et SARMUB2 : la première est le passage d'un vecteur g_k à un vecteur g commun pour tous les w_k . Ensuite, les multiplicateurs de Lagrange associés aux contraintes d'orthogonalité qui étaient des matrices deviennent des scalaires, ce qui simplifie les calculs.

2.1.3.2 Problèmes rencontrés

Pour approximer les solutions théoriques, l'idée est de mettre en place un algorithme similaire à celui de SARMUB1. Ainsi se pose la question de l'actualisation des w_k tenant compte de la contrainte d'orthogonalité (2.31), et de l'actualisation des a_k^j en tenant compte de cette contrainte. Pour ce dernier, une procédure strictement similaire nous conduit à actualiser a_k^j par

$$a_k^j + w_k[j] B_k^{j'} D \sum_{\ell=1}^{a-1} \nu_\ell t_\ell, \quad \nu_\ell \in \mathbb{R},$$

où les ν_ℓ sont solutions du système, à $(a-1)$ équations à $(a-1)$ inconnues :

$$\underbrace{\begin{bmatrix} t'_1 DB_k^j B_k^{j'} Dt_1 & \cdots & \cdots & t'_1 DB_k^j B_k^{j'} Dt_{a-1} \\ \vdots & t'_\ell DB_k^j B_k^{j'} Dt_\ell & \cdots & t'_\ell DB_k^j B_k^{j'} Dt_{a-1} \\ \vdots & \ddots & & \vdots \\ \vdots & & & t'_{a-1} DB_k^j B_k^{j'} Dt_{a-1} \end{bmatrix}}_{W_k^j} \begin{bmatrix} \nu_1 \\ \vdots \\ \nu_{a-1} \end{bmatrix} = \underbrace{\begin{bmatrix} -t'_1 Dt / w_k^2[j] \\ \vdots \\ -t'_{a-1} Dt / w_k^2[j] \end{bmatrix}}_{i_k^j} \quad (2.36)$$

En effet, en utilisant un calcul analogue au développement de l'équation (2.15), la contrainte (2.31) mène à :

$$\begin{aligned} & [t_1 \ \cdots \ t_{a-1}]' D B_k^j a_k^j w_k[j] = \\ & - [t_1 \ \cdots \ t_{a-1}]' D \sum_{k'=1; k' \neq k}^K \sum_{j=1}^{J_{k'}} B_{k'}^j a_{k'}^j w_{k'}[j] \\ & - [t_1 \ \cdots \ t_{a-1}]' D \sum_{j'=1; j' \neq j}^{J_k} B_k^{j'} a_k^{j'} w_k[j']. \end{aligned} \quad (2.37)$$

En supposant fixés $a_k^{j'}$ pour $j' \neq j$ et $j' \in \{1, \dots, J_k\}$, ainsi que $a_{k'}^j$ pour tout $k' \neq k$ et tout $j \in \{1, \dots, J_{k'}\}$, nous cherchons à actualiser a_k^j . Celui-ci vérifie l'équation $a_k^j V_k = c_k^j$ où $V_k = w_k[j] B_k^j D [t_1 \ \cdots \ t_{a-1}]$ et la constante c_k^j est égale à :

$$- [t_1 \ \cdots \ t_{a-1}]' D \sum_{k'=1; k' \neq k}^K \sum_{j=1}^{J_{k'}} B_{k'}^j a_{k'}^j w_{k'}[j] - [t_1 \ \cdots \ t_{a-1}]' D \sum_{j'=1; j' \neq j}^{J_k} B_k^{j'} a_k^{j'} w_k[j'].$$

D'où, a_k^j doit appartenir à l'intersection de $a - 1$ hyperplans, orthogonale à l'espace engendré par les colonnes de V_k . Ainsi, $\nu_k^j = [\nu_1^j \ \cdots \ \nu_{a-1}^j]'$ est le vecteur tel que $a_k^j + w_k[j] B_k^j D \sum_{\ell=1}^{a-1} \nu_\ell t_\ell$ vérifie (2.37). D'où le résultat $\nu_k^j = W_k^{j+} l_k^j$.

Pour ce qui est l'actualisation des w_k , un problème se pose. En effet, (2.31) nous permet d'écrire :

$$[w'_1 \ \cdots \ w'_K] \begin{bmatrix} S'_1 \\ \vdots \\ S'_K \end{bmatrix} D [T_1 \ \cdots \ T_{a-1}] = 0$$

Cela signifie que le vecteur $w = [w'_1 \ \cdots \ w'_K]'$, qui rappelons-le, est de norme unité par bloc, est orthogonal à l'espace engendré par $S' D T_{a-1}$, où S désigne le tableau conjoint des S_k et T_{a-1} celui des super-composantes t_1, \dots, t_{a-1} . Si, pour actualiser w , nous utilisons le résidu $w - P_{S' D T_{a-1}} w$, alors le nouveau w n'est plus normé par bloc, et si on le norme ainsi, il ne sera plus dans l'espace voulu.

Ce que nous pouvons dire, c'est que, à l'optimum, w est orthogonal au sens de D_w à $S' D T_{a-1}$, avec :

$$D_w = \begin{bmatrix} \frac{Id_{J_1}}{\|w_1\|} & 0 & \cdots & 0 \\ 0 & \ddots & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & \frac{Id_{J_K}}{\|w_K\|} \end{bmatrix} \quad (2.38)$$

Si les vecteurs w_k sont correctement initialisés, sous entendu «initialisés proche de l'optimum», alors, on peut espérer qu'un algorithme similaire à celui de SARMUB1, dans lequel l'actualisation de w se ferait par $w - P_{S' D T_{a-1}} w$, avec

$$P_{S' D T_{a-1}} = S' D T_{a-1} D_w (T_{a-1}' D S D_w S' D T_{a-1})^+ T_{a-1}' D S D_w,$$

converge. C'est cet algorithme que nous détaillons ci-après. Pour l'instant, nous avons pu constater par la pratique qu'il convergeait selon les données. Nous pensons chercher une meilleure initialisation pour w_k afin de résoudre ce problème.

2.1.3.3 Proposition d'un algorithme

étape 0 : initialisation de w_k , b_n , μ_k , λ_n et S_k

Effectuer la première étape de SARMUB sur X_k et $Y_n^{(a-1)}$.
 $g = \sum_{n=1}^N u_n$; $\beta_k^j = g' D S_k^j w_k[j] / v_k^j$

étape 1

tant que $opt1 = \sum_{k=1}^K t'_k D \sum_{n=1}^N u_n$ ne converge pas,

Faire :

$$\begin{aligned} \alpha_\ell &= -\frac{1}{\|t_\ell\|_D^2} \sum_{k=1}^K \sum_{j=1}^{J_k} \beta_k^j \frac{w_{k,\ell}[j]}{w_k[j]} S_{k,\ell}^{j'} D S_k^j, \quad \ell = 1, \dots, a-1 \\ g &= \sum_{n=1}^N u_n - \sum_{\ell=1}^{a-1} \alpha_\ell t_\ell \\ w_k &= S_k' D g \\ \mu_k &= \|w_k\| \\ \spadesuit w &= w - P_{S'D} [t_1 \ \dots \ t_{a-1}] w \\ t_k &= S_k w_k / \mu_k; \quad w_k = w_k / \mu_k \\ b_n &= Y_n' D \sum_{k=1}^K S_k w_k; \quad \lambda_n = \|b_n\| \\ b_n &= b_n / \lambda_n \\ opt1 &= \sum_{n=1}^N b_n' Y_n' D \sum_{k=1}^K S_k w_k \end{aligned}$$

étape 2 : calcul des β_k^j

$$\beta_k^j = g' D S_k^j w_k[j] / v_k^j$$

étape 3 : calcul de α_ℓ

$$\alpha_\ell = -\frac{1}{\|t_\ell\|_D^2} \sum_{k=1}^K \sum_{j=1}^{J_k} \beta_k^j \frac{w_{k,\ell}[j]}{w_k[j]} S_{k,\ell}^{j'} D S_k^j, \quad \ell = 1, \dots, a-1$$

étape 4 : actualisation

$$\begin{aligned} g &= \sum_{n=1}^N u_n - \sum_{\ell=1}^{a-1} \alpha_\ell t_\ell \\ \beta_k^j &= g' D S_k^j w_k[j] / v_k^j \end{aligned}$$

étape 5 : calcul des coefficients splines

$$a_k^j = \frac{1}{\beta_k^j} (B_k^{j'} D B_k^j) + B_k^{j'} D g w_k[j]$$

étape 6 : actualisation

$$\begin{aligned} S_k^j &= B_k^j a_k^j; \quad t_k = S_k w_k \\ \spadesuit \nu_k^j &= W_k^{j+l_k^j} \\ \spadesuit a_k^j &= a_k^j + w_k[j] B_k^{j'} D [t_1 \ \dots \ t_{a-1}] \nu_k^j \end{aligned}$$

étape 7 : test de la convergence

$$opt2 = \sum_{k=1}^K t'_k D \sum_{n=1}^N u_n$$

Si $|opt2 - opt1| \leq opt2\epsilon$ on arrête.

Sinon, on retourne à l'étape 1.

2.1.4 Nature des solutions d'ordre a : la matrice Hessienne

Pour les étapes $a > 1$, il ne nous a pas été possible de démontrer que l'algorithme est croissant. Donc, s'il converge, nous ne sommes pas sûrs que les solutions obtenues soient au maximum du critère. C'est pourquoi, nous nous sommes posé la question « *s'ils existent, ces points critiques sont-ils au maximum de notre fonction objectif?* ». Pour répondre à cela, nous avons calculé la matrice hessienne en les points critiques (solutions des équations normales). Une condition suffisante pour qu'ils soient bien au maximum, est que la matrice Hessienne \mathcal{H} soit semi-définie négative sur l'ensemble des points x tels que $\mathcal{G}x = 0$, où \mathcal{G} est la matrice jacobienne associée aux contraintes. Si $x'\mathcal{H}x \leq 0$ pour tout point x vérifiant les contraintes, alors la condition suffisante est vérifiée.

Dans l'écriture de \mathcal{H} qui suit, $w_k, \lambda_n, b_n, \mu_k, \alpha_k^j, \beta_k^j$ et α_l font référence aux points critiques vérifiant les équations normales.

$$\mathcal{H} = \begin{bmatrix} -D_\mu & S'DY & D_{gB} \\ Y'DS & -D_\lambda & Y'DB_w \\ D'_{gB} & B'_wDY & -V_{\beta B} \end{bmatrix} \quad (2.39)$$

avec :

$$\begin{aligned} S &= [S_1 \quad \cdots \quad S_K] \\ Y &= [Y_1 \quad \cdots \quad Y_N] \end{aligned}$$

$$D_\mu = \begin{bmatrix} Id_{J_1} \mu_1 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & Id_{J_k} \mu_k & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & Id_{J_K} \mu_K \end{bmatrix}$$

$$D_\lambda = \begin{bmatrix} Id_{M_1} \lambda_1 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & Id_{M_n} \lambda_n & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & Id_{M_N} \lambda_N \end{bmatrix}$$

$$B_w = \begin{bmatrix} B_1^1 w_1[1] & B_1^2 w_1[2] & \cdots & B_k^j w_k[j] & \cdots & B_K^{J_K} w_K[J_K] \end{bmatrix}$$

$$D_{gB} = \begin{bmatrix} g'DB_1^1 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & g'DB_1^2 & \ddots & & & \vdots \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & g'DB_k^j & \ddots & \vdots \\ \vdots & & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & g'DB_K^{J_K} \end{bmatrix} \text{ pour SARMUB 2}$$

et

$$D_{gB} = \begin{bmatrix} g'_1DB_1^1 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & g'_1DB_1^2 & \ddots & & & \vdots \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & g'_kDB_k^j & \ddots & \vdots \\ \vdots & & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & g'_KDB_K^{J_K} \end{bmatrix} \text{ pour SARMUB 1}$$

$$V_{\beta B} = \begin{bmatrix} \beta_1^1 B_1^1{}' DB_1^1 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \beta_1^2 B_1^2{}' DB_1^2 & \ddots & & & \vdots \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \beta_k^j B_k^j{}' DB_k^j & \ddots & \vdots \\ \vdots & & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & \beta_K^{J_K} B_K^{J_K}{}' DB_K^{J_K} \end{bmatrix}$$

Tous les termes diagonaux de \mathcal{H} sont négatifs, car μ_k , λ_n et β_k^j sont positifs. Nous pouvons ainsi affirmer que la matrice n'est pas semi définie positive, et que les points critiques ne se trouvent pas sur un minimum. Si nous pouvions montrer que la matrice symétrique

$$\mathcal{H}_1 = \begin{bmatrix} 0 & S'DY & D_{gB} \\ Y'DS & 0 & Y'DB_w \\ D'_{gB} & B'_wDY & 0 \end{bmatrix} \quad (2.40)$$

est semi-définie négative, alors, comme la somme de matrices semi-définies négatives est semi-définie négative, le résultat cherché serait atteint.

Le fait qu'il nous paraît impossible de montrer que \mathcal{H} est semi-définie négative nous amène à conclure qu'elle est soit effectivement semi-définie négative, et nous n'arrivons pas à le montrer, et les points critiques que nous cherchons à approximer algorithmiquement sont bien au maximum, et alors si notre algorithme ne les trouve pas c'est qu'il n'est pas bon (dans SARMUB 1 et 2); soit \mathcal{H} n'est pas toujours semi-définie négative, et si l'algorithme converge il se peut qu'il ait trouvé un point selle.

Si l'algorithme ne converge pas, cela peut aussi être parce qu'il n'y pas de solutions. Malheureusement nous n'avons nulle part montré l'existence des solutions d'ordre a . Par contre, les solutions d'ordre 1 existent et sont atteintes car l'algorithme est croissant et convergent.

Notre faible expérience de l'algorithme SARMUB ne permet pas l'heure actuelle de dire s'il est convergent dans la plupart des cas ou non.

2.1.5 Propriétés et modèles

Grâce à la contrainte imposée, les super-composantes t_a de SARMUB2 sont orthogonales, et les composantes $t_{k,a}$ sont orthogonales aux composantes $t_{k',\ell}$, pour $\forall k'$ et $\ell = 1, \dots, a - 1$ dans SARMUB1. Ainsi, dans SARMUB, nous pouvons définir les super-composantes $t_a = \sum_{k=1}^K t_{k,a}$ qui sont aussi orthogonales entre elles. Notons que nous pouvons également les définir comme une somme pondérée (par $\mu_{k,a}$ par exemple), elles seront aussi orthogonales. Ainsi, nous pouvons faire une représentation «moyenne» des individus vus par les prédicteurs grâce aux super-composantes : (t_a, t_ℓ) . Dans SARMUB1, une étude plus précise des individus peut être effectuée pour chaque tableau X_k grâce à la représentation de $(t_{k,a}, t_{k,\ell})$.

Comme dans ASPLS (DURAND & SABATIER, 1997) (§1.6), les composantes $t_{k,a}$ s'expriment additivement en fonction de transformations splines des prédicteurs x_k^j :

$$\begin{aligned} t_{k,a} &= S_{k,a} w_{k,a} \\ &= \sum_{j=1}^{J_k} B_k^j a_{k,a}^j w_{k,a}(j) \\ &= \sum_{j=1}^{J_k} w_{k,a}(j) s_{k,a}^j(x_k^j) \\ &= \sum_{j=1}^{J_k} \varphi_{k,a}^j(x_k^j). \end{aligned}$$

Chaque composante $t_{k,a}$ est une somme pondérée par $w_{k,a}(j)$ des transformations splines optimales des prédicteurs du tableau X_k , $s_{k,a}^j(x_k^j)$. Ces transformations splines ont été optimisées de façon à modéliser le mieux possible les variables réponses Y_n . $\varphi_{k,a}^j(x_k^j) = w_{k,a}(j) s_{k,a}^j(x_k^j)$ est une fonction coordonnée des prédicteurs, dans la base de B-splines définie par B_k^j , car $\varphi_{k,a}^j = B_k^j (a_{k,a}^j w_{k,a}(j))$. $w_{k,a}(j)$ mesure l'importance «non-linéaire» de la variable x_k^j dans la construction de $t_{k,a}$. De plus, $s_{k,a}^j(x_k^j)$ représente la forme de la non-linéarité du prédicteur qui est utilisée à l'étape a pour la construction des composantes. Ainsi, le graphique $(x_k^j, \varphi_{k,a}^j(x_k^j))$ permet d'évaluer l'importance, en terme de non-linéarité, du prédicteur x_k^j dans la construction de la composante $t_{k,a}$.

Les super-composantes $t_a = \sum_{k=1}^K t_{k,a}$ s'écrivent donc :

$$t_a = \sum_{k=1}^K \sum_{j=1}^{J_k} \varphi_{k,a}^j(x_k^j).$$

De façon analogue, les modèles de chaque réponse se décomposent additivement en fonction de transformations splines des prédicteurs. De plus, grâce à l'orthogonalité des espaces sur lesquels les tableaux Y_n sont «déflatés», le modèle est une somme de sous modèles partiels orthogonaux : à chaque étape une partie de la variation de Y_n non encore expliquée par les étapes précédentes est modélisée.

Comme dans l'ACIMOG-PLS, en notant $\mathcal{E}_a = t_a$ dans SARMUB2 et $\mathcal{E}_a = T_a$ dans SARMUB1, les modèles s'écrivent :

$$\begin{aligned} Y_n &= \widehat{Y}_n^A + E_{n,A} \\ &= \sum_{a=1}^A P_{\mathcal{E}_a} Y_n^{(a-1)} + Y_n^{(A)} \\ &= \sum_{a=1}^A \mathcal{E}_a (\mathcal{E}_a' D \mathcal{E}_a)^+ \mathcal{E}_a' D Y_n^{(a-1)} + Y_n^{(A)}. \end{aligned}$$

Dans SARMUB2, nous avons :

$$\begin{aligned} \widehat{Y}_n^A &= \sum_{a=1}^A t_a t_a' D Y_n^{(a-1)} / \|t_a\|_D^2 \\ &= \sum_{a=1}^A \sum_{k=1}^K S_{k,a} w_{k,a} t_a' D Y_n^{(a-1)} / \|t_a\|_D^2 \\ &= \sum_{a=1}^A \sum_{k=1}^K \left(\sum_{j=1}^{J_k} B_k^j a_{k,a}^j w_{k,a}[j] \right) t_a' D Y_n^{(a-1)} / \|t_a\|_D^2 \end{aligned} \quad (2.41)$$

$$= \sum_{k=1}^K \sum_{j=1}^{J_k} B_k^j \left(\sum_{a=1}^A a_{k,a}^j w_{k,a}[j] t_a' D Y_n^{(a-1)} / \|t_a\|_D^2 \right). \quad (2.42)$$

Notons que le passage de (2.41) à (2.42) est dû à l'utilisation de la même base de B-splines à chaque étape a . Si les prédicteurs avaient été «déflatés» à chaque étape, B_k^j dépendrait de a et les modèles seraient beaucoup plus difficiles à interpréter.

En notant $\tilde{a}_{k,A}^{j,m} = \sum_{a=1}^A a_{k,a}^j w_{k,a}[j] t_a' D y_n^{m(a-1)} / \|t_a\|_D^2$, la variable m du tableau Y_n est modélisée à l'étape A par :

$$\begin{aligned} \widehat{y}_n^{m,A} &= \sum_{k=1}^K \sum_{j=1}^{J_k} B_k^j \tilde{a}_{k,A}^{j,m} \\ &= \sum_{k=1}^K \sum_{j=1}^{J_k} f_{k,A}^{j,m}(x_k^j). \end{aligned}$$

$f_{k,A}^{j,m}(x_k^j)$ est une fonction spline coordonnée de x_k^j . Elle est fonction des transformations splines optimales $s_{k,a}^j(x_k^j)$:

$$f_{k,A}^{j,m}(x_k^j) = \sum_{a=1}^A \frac{w_{k,a}(j)}{\|t_a\|_D^2} t_a' D y_n^{m(a-1)} s_{k,a}^j(x_k^j).$$

$f_{k,A}^{j,m}(x_k^j)$ est une somme des transformations splines optimales du prédicteur x_k^j pondérées par l'importance de x_k^j dans la composante $t_{k,a}$, donc dans t_a , ($w_{k,a}[j]$) multiplié par la proximité de t_a avec la réponse y_n^m à l'étape a ($t_a' D y_n^{m(a-1)} / \|t_a\|_D^2$).

Le graphique $(x_k^j, f_{k,A}^{j,m}(x_k^j))$ permet de visualiser l'importance non-linéaire de x_k^j sur y_n^m .

Dans SARMUB1, nous avons :

$$\begin{aligned} \widehat{Y}_n^A &= \sum_{a=1}^A T_a (T_a' D T_a)^+ T_a' D Y_n^{(a-1)} \\ &= \sum_{a=1}^A \left[\sum_{j=1}^{J_1} B_1^j a_{1,a}^j w_{1,a}[j] \quad \dots \quad \sum_{j=1}^{J_K} B_K^j a_{K,a}^j w_{K,a}[j] \right] (T_a' D T_a)^+ T_a' D Y_n^{(a-1)}. \end{aligned}$$

En notant $M_{n,a} = (T_a' D T_a)^+ T_a' D Y_n^{(a-1)}$, et $M_{n,a}^m(k)$ la $k^{\text{ième}}$ ligne de la $m^{\text{ième}}$ colonne de $M_{n,a}$, nous obtenons :

$$\begin{aligned} \widehat{y}_n^{mA} &= \sum_{k=1}^K \sum_{a=1}^A \sum_{j=1}^{J_1} B_k^j a_{k,a}^j w_{k,a}[j] M_{n,a}^m(k) \\ &= \sum_{k=1}^K \sum_{j=1}^{J_k} B_k^j \left(\sum_{a=1}^A a_{k,a}^j w_{k,a}[j] M_{n,a}^m(k) \right) \\ &= \sum_{k=1}^K \sum_{j=1}^{J_k} B_k^j a_{k,A}^{j,m} \\ &= \sum_{k=1}^K \sum_{j=1}^{J_k} g_{k,A}^{j,m}(x_k^j). \end{aligned}$$

Les modèles ainsi obtenus sont du même type que dans SARMUB2. Notons que la fonction coordonnée $g_{k,A}^{j,m}(x_k^j)$ tient ici compte des proximités entre $y_n^{m(a-1)}$ et chaque composante $t_{k,a}$. Dans SARMUB2, $f_{k,A}^{j,m}(x_k^j)$ tenait compte des proximités entre $y_n^{m(a-1)}$ et t_a . De la même façon que dans SARMUB2, $(x_k^j, g_{k,A}^{j,m}(x_k^j))$ permet de visualiser l'importance non-linéaire de x_k^j sur y_n^m .

2.1.6 Commentaires

La principale différence entre SARMUB et ASPLS appliquée aux tableaux conjoints, est que l'ajustement des paramètres se fait par tableaux. S'il n'y a qu'une seule variable réponse y à expliquer, les deux approches sont équivalentes à la première étape. Mais elles ne le sont plus dès la seconde étape à cause de la contrainte d'orthogonalité des (super-)composantes.

Le choix du nombre d'étapes A peut être fait par validation croisée ainsi qu'en s'aidant des pourcentages de variance des Y_n expliquée à chaque étape. De plus, l'utilisateur doit choisir le nombre de noeuds L_k^j à utiliser pour le codage de chacun des prédicteurs ainsi que

leurs positions et le degré d_k^j des splines. Il n'existe pas encore de méthode permettant de choisir le meilleur nombre de noeuds ni la meilleure place pour la situation dans laquelle on se trouve. Il est envisageable d'étendre la méthode MBOK (Multivariate Bounded Optimal Knots) applicable à la régression spline multivariée ordinaire (MOLINARI, 2000), pour les cas de multi-réponses et de multi-tableaux. Cependant, cela risque de prendre énormément de temps de calcul. Aussi, il est recommandé à l'utilisateur de commencer par essayer des noeuds équi-espacés, et en faible nombre. Le degré des splines doit d'abord être pris à 0 pour tester la performance des modèles linéaires, puis si l'utilisation des modèles non-linéaires est justifiée, essayer des degrés supérieurs. L'utilisation de la validation croisée pour choisir L_k^j et d_k^j est possible en théorie, mais s'avère difficile en pratique car l'algorithme itératif nécessite beaucoup de temps de calcul, surtout si le nombre de variables par tableau est assez grand et les tableaux nombreux.

Comme avec l'ACIMOG-PLS, il faudrait essayer de trouver un outil pour mesurer l'influence d'un tableau k sur les modèles \widehat{Y}_n^A . Le seul outil disponible pour l'instant est l'étude des fonctions coordonnées qui permettent de mesurer l'influence d'un prédicteur sur les modèles.

Nous ne pouvons malheureusement pas présenter à l'heure actuelle un exemple d'application pratique pertinent. SARMUB est toujours en cours de développement.....

Une autre généralisation de ASPLS pourrait se faire en utilisant le critère de l'ACIMO-PLS (§I-5.5), $\sum_{n=1}^N cov^2(t_k, u)$, dans le cas où il n'y a qu'un seul tableau de prédicteurs. Nous ne présenterons pas la résolution d'un tel problème ici, bien que nous l'ayons fait. Les calculs s'établissent aussi facilement qu'avec l'ASPLS, à la différence près que le u dans ASPLS est remplacé par $\sum_{n=1}^N Y_n Y_n' D t$ dans l'ACIMO-ASPLS. Alors, pour l'étape $a = 1$, la composante t obtenue par l'ASPLS sur X et le tableau conjoint Y est identique à la composante issue de l'ACIMO-ASPLS. Les composantes u_n de l'ACIMO-ASPLS sont proportionnelles au $n^{\text{ième}}$ bloc de la composante u issue de ASPLS. Pour les étapes suivantes, les composantes t doivent être assez proches sans être égales, car $u = \sum_{n=1}^N Y_n b_n = Y Y' D t$ dans ACIMO-ASPLS, et dans l'ASPLS $u = Y Y' D t / \|Y' D t\|$. Comme le facteur qui les différencie ne peut être factorisé dans le calcul de S^j , les composantes t devraient être légèrement différentes, et cette différence devrait s'accroître avec a .

2.2 Autres approches

Nous proposons dans cette partie plusieurs manières d'utiliser le codage des prédicteurs dans une base de B-splines en combinaison avec les méthodes de régression linéaire que nous avons proposé dans la première partie de cette thèse : ACIMO-PLS et ACIMOG-PLS.

Les notations sont les mêmes que précédemment : chaque prédicteur centré et éventuellement réduit x_k^j est codé dans une base de B-splines normalisée de degré d_k^j et à l'aide de L_k^j noeuds intérieurs ξ_k^j . Ce codage définit la matrice B_k^j de dimension $L_k^j + 1 + d_k^j$. Les noeuds ne sont pas considérés comme des paramètres inconnus, ils doivent être fixés par l'utilisateur.

Les tableaux Y_n sont centrés et éventuellement réduits.

La première technique que nous proposons peut être référencée par ACIMO-PLSS. Elle généralise l'ACIMO-PLS en procédant comme PLSS (DURAND, 2001) (§1.7). Ici, nous supposons que $K = 1$ c'est-à-dire que nous sommes en présence d'un seul tableau de prédicteur et de N tableaux réponses. L'ACIMO-PLSS est alors définie par l'ACIMO-PLS appliquée aux tableaux (non centrés et non réduits) $B = [B^1 \dots B^J]$ et Y_n (centrés et/ou réduits) munis des métriques $D_n Q^y Q^x = Id$. Les composantes t_a obtenues s'écrivent $t_a = B^{(a-1)} a_a$ où $B^{(a-1)} = P_{t_1, \dots, t_{a-1}}^\perp B$. t_a peut s'écrire aussi $t_a = B a_a^* = \sum_{j=1}^J B^j a_a^*[j]$. $B^j a_a^*[j] = s_a^j(x^j)$ est la fonction spline coordonnée qui a été optimisée selon le critère de l'ACIMO-PLS, dans le but d'expliquer les Y_n . Elle permet d'étudier l'influence du prédicteur x^j sur la composante t_a . De même, les modèles $\hat{Y}_n^A = \sum_{a=1}^A P_{t_a} Y_n$ peuvent s'écrire comme une somme de fonctions splines coordonnées. Ecrivons le modèle pour la variable m de Y_n :

$$\begin{aligned} \hat{y}_n^{m,A} &= \sum_{a=1}^A \sum_{j=1}^J B^j a_a^*[j] t_a' D y_n^m / \|t_a\|_D^2 \\ &= \sum_{j=1}^J B^j \left(\sum_{a=1}^A a_a^*[j] t_a' D y_n^m / \|t_a\|_D^2 \right) \\ &= \sum_{j=1}^J B^j \tilde{a}_A^{j,m} = \sum_{j=1}^J f_A^{j,m}(x^j). \end{aligned}$$

La fonction spline coordonnée $f_A^{j,m}(x^j) = B^j \tilde{a}_A^{j,m}$ est utile pour évaluer l'influence non-linéaire de la variable j sur la variable réponse y_n^m .

Cette approche est possible lorsque le nombre de prédicteurs n'est pas trop élevé.

Une autre approche est d'utiliser ces mêmes codages des prédicteurs B^j en combinaison avec l'ACIMOG-PLS, en appliquant cette dernière aux tableaux B^1, \dots, B^J et Y_1, \dots, Y_N munis des métriques D , Q_n^y et $Q_n^x = Id$. Alors les composantes explicatives $t_{j,a}$ du bloc B^j s'écrivent : $t_{j,a} = B^{j(a-1)} a_{j,a} = B^j a_{j,a}^*$. $t_{j,a}$ est donc une fonction spline coordonnée de x^j et permet de voir quelle est la non-linéarité capturée par $t_{j,a}$.

De plus les modèles se décomposent de la façon suivante, en notant $\mathcal{E}_a = t_a = \sum_{j=1}^J t_{j,a}$ pour l'application de l'ACIMOG-PLS2, $\mathcal{E}_a = \sum_{j=1}^J \mu_{j,a} t_{j,a}$ pour l'application de l'ACIMOG-PLS3 et $\mathcal{E}_a = [t_{1,a} \dots t_{J,a}]$ pour l'ACIMOG-PLS1 :

$$\begin{aligned} \hat{Y}_n^A &= \sum_{a=1}^A \mathcal{E}_a (\mathcal{E}_a' D \mathcal{E}_a)^+ \mathcal{E}_a' D Y_n \\ &= \sum_{a=1}^A [B^1 a_{1,a}^* \dots B^J a_{J,a}^*] (\mathcal{E}_a' D \mathcal{E}_a)^+ \mathcal{E}_a' D Y_n \text{ pour l'ACIMOG-PLS1} \end{aligned}$$

$$\begin{aligned}\widehat{Y}_n^A &= \sum_{a=1}^A \sum_{j=1}^J B^j a_{j,a}^* (\mathcal{E}'_a D \mathcal{E}_a)^+ \mathcal{E}'_a D Y_n \text{ pour l'ACIMOG-PLS2} \\ \widehat{Y}_n^A &= \sum_{a=1}^A \sum_{j=1}^J \mu_{j,a} B^j a_{j,a}^* (\mathcal{E}'_a D \mathcal{E}_a)^+ \mathcal{E}'_a D Y_n \text{ pour l'ACIMOG-PLS3.}\end{aligned}$$

Notons $M_{n,a} = (\mathcal{E}'_a D \mathcal{E}_a)^+ \mathcal{E}'_a D Y_n$, et $M_{n,a}^m[j]$ le $j^{\text{ième}}$ bloc ligne de la $m^{\text{ième}}$ colonne. Alors, le modèle pour la $m^{\text{ième}}$ variable de Y_n à l'étape A devient :

$$\begin{aligned}\widehat{y}_n^{m,A} &= \sum_{j=1}^J B^j \sum_{a=1}^A (a_{j,a}^* M_{n,a}^m[j]) \text{ pour l'ACIMOG-PLS1} \\ \widehat{y}_n^{m,A} &= \sum_{j=1}^J B^j \left(\sum_{a=1}^A a_{j,a}^* M_{n,a}^m \right) \text{ pour l'ACIMOG-PLS2} \\ \widehat{y}_n^{m,A} &= \sum_{j=1}^J B^j \left(\sum_{a=1}^A \mu_{j,a} a_{j,a}^* M_{n,a}^m \right) \text{ pour l'ACIMOG-PLS3} \\ &= \sum_{j=1}^J B^j \check{a}_A^{j,m} = \sum_{j=1}^J g_A^{j,m}(x^j).\end{aligned}$$

Comme précédemment la fonction coordonnée $g_A^{j,m}(x^j)$ permet d'évaluer l'influence non-linéaire de la variable x^j sur la variable y_n^m . La différence entre $g_A^{j,m}(x^j)$ et $f_A^{j,m}(x^j)$ de l'ACIMO-PLSS réside dans le calcul des vecteurs des coefficients splines $\check{a}_A^{j,m}$ et $\tilde{a}_A^{j,m}$, qui diffère selon les deux approches.

La dernière approche s'applique lorsque l'on est en présence de K tableaux de prédicteurs. Elle est envisageable uniquement lorsque le nombre de prédicteurs par tableau et le nombre de tableaux prédicteurs ne sont pas trop élevés. Il s'agit d'appliquer l'ACIMOG-PLS sur les K tableaux $B_k = \begin{bmatrix} B_k^1 & \dots & B_k^k \end{bmatrix}$ et les Y_n munis des métriques D , $Q_k^x = Id$ et Q_n^y . On peut désigner cette approche par ACIMOG-PLSS.

Les modèles obtenus s'écrivent comme une somme de fonctions splines coordonnées des prédicteurs x_k^j . Nous ne détaillons pas leur écriture. Elle s'obtient en rajoutant l'indice k et une somme sur k dans l'écriture des modèles obtenus par l'approche précédente.

Cette dernière technique est en théorie applicable, mais très vite la méthode peut nécessiter beaucoup de temps en calculs si le nombre de variable est élevé car l'ACIMOG-PLS est une méthode itérative.

Troisième partie

Applications

Introduction

Cette partie contient des applications de méthodes dont nous avons parlé dans les parties précédentes, et plus particulièrement de applications de nos méthodes, à des données réelles. Les problèmes posés par ces données relèvent le l'écologie, l'analyse sensorielle, la chimiométrie ou encore la biométrie. Les chapitres correspondent chacun à une application indépendante, et peuvent donc être lus dans n'importe quel ordre. Certaines de ces applications ont été présentées lors de congrès et certaines ont fait l'objet de collaborations extérieures.

Nous avons programmé toutes les méthodes utilisées en langage S-Plus[®] 5.1 et 6.0 pour fonctionner sur un système Linux. Mais, ces programmes devraient fonctionner sous S-Plus[®] 2000 pour Windows[™] ou encore sous R* après quelques modifications mineures. Ces programmes, décrits succinctement dans le tableau 1, seront disponibles prochainement sur internet à l'adresse <http://perso.wanadoo.fr/myrtouille> et peuvent m'être demandés par e-mail (myrtouille@wanadoo.fr ou mvivien@pharma.univ-montp1.fr).

* Clone de S-Plus[®] disponible gratuitement sur internet à l'adresse <http://cran.r-project.org>.

TAB. 1 – *Les méthodes que nous avons programmées en S-Plus®*

Programme	Description
acimo	résultats numériques de l'ACIMO
acimo.plot	effectue les graphiques pour l'ACIMO
acimopls	résultats numériques de l'ACIMO-PLS
acimopls.plot	effectue les graphiques de l'ACIMO-PLS
acimopls cv	effectue la validation-croisée de l'ACIMO-PLS
acom	résultats numériques de l'ACOM
concord	résultats numériques de l'ACO
acimogen	résultats numériques de l'ACIMOG-PLS1, 2 et 3
acimogen.graph	effectue les graphiques de l'ACIMOG-PLS. A compléter.
acimogen cv	effectue la validation croisée de l'ACIMOG-PLS
acimogen0	résultats numériques de l'ACIMOG 1, 2 et 3
acimogen0.graph	effectue les graphiques de l'ACIMOG
remub	résultats numériques de REMUB
doact	résultats numériques et graphiques de DO-ACT
npls	résultats numériques de NPLS
npls cv	validation-croisée de NPLS
hpls	résultats numériques de HPLS
hpls cv	validation croisée de HPLS
mbpls1	résultats numériques de MBPLS avec un tableau réponse Y et déflation sur la super-composante t_T . A compléter
mbpls	résultats numériques de MBPLS avec N tableaux réponses et K tableaux explicatifs. A compléter
statis	résultats numériques et graphiques de STATIS défini par LAVIT (1988)
statis dual	résultats numériques et graphiques de STATIS dual défini par LAVIT (1988)
acimopls s	résultats numériques de l'ACIMO-PLSS
acimopls s cv	validation-croisée pour l'ACIMO-PLSS
acimopls s.plot	effectue les graphiques de l'ACIMO-PLSS
sarmub2	résultats numériques de SARMUB2. Algorithme à améliorer
sarmub1	résultats numériques de SARMUB1. A compléter
OSC	prétraitement osc de WOLD et al. (1998)
BIP	calcul des coefficients BIP
CrossValR2	calcul des Q^2 à partir des résultats de la validation croisée

Chapitre 1

Biométrie : ACIMO et STATIS

Ce chapitre présente un exemple de l'application de l'Analyse de Co-Inertie Multiple Orthogonale à l'intrastructure de STATIS. La méthode a été présentée dans le chapitre 2.5 de la première partie.

Les données sont des données psychométriques qui ont été étudiées par M.O. Pernin dans sa thèse (PERNIN, 1986). Elles sont extraites d'une grande base de données construite à partir d'une étude sur la croissance de mêmes enfants suivis de leur naissance à la fin de leur adolescence, entre 1953 et 1975. Cette enquête a été conduite par Sempé jusqu'en 1975, et tous les détails de cette enquête peuvent être trouvés dans (SEMPÉ et al., 1979). Tous les enfants sont nés à Paris et pesaient à peu près le même poids à leur naissance.

Le jeu de données que nous étudions ici est constitué de la description morphologique de $l = 30$ fillettes par $J = 8$ variables (tableau 1.1) mesurées tous les ans pendant $K = 12$ ans : de leur quatrième à leur quinzième anniversaire. Le tableau de données prend donc la forme d'un cube (Fig. 1.1)

Ces données ont été étudiées par différentes analyses cubiques et multi-bloc dans (SEMPÉ, KROONENBERG, LAVIT, PERNIN, LEWI, CALOMNE, VAN HOOF, PONTIER & MINEO, 1989), incluant la méthode STATIS.

Le but de l'étude est de savoir s'il est possible de distinguer différents âges correspondants à des étapes de croissance, et de mettre en évidence des variables morphologiques typiques pour ces âges.

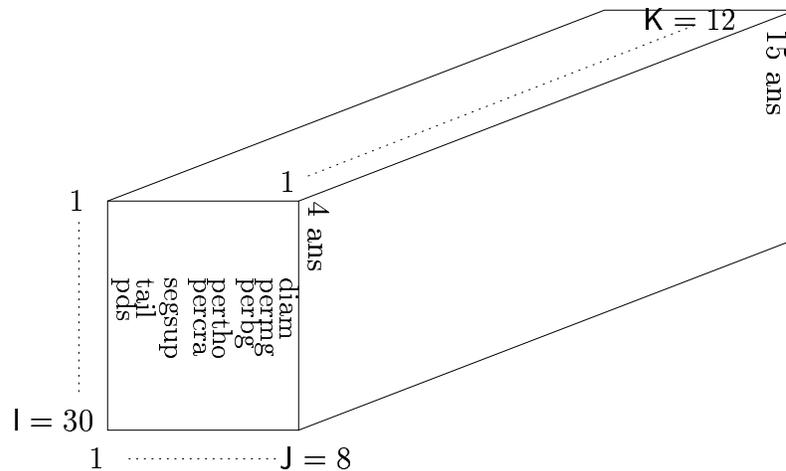
Les $K = 12$ tableaux ont été centrés et réduits pour la métrique uniforme $D = Id_{30}/30$.

Nous allons d'abord appliquer la méthode STATIS, sans toutefois trop en détailler les ré-

TAB. 1.1 – Codage et description des $J = 8$ variables

1	pds	poids (dcg)
2	tail	taille, mesurée en position couchée (mm)
3	segsup	longueur du segment supérieur, mesurée en position assise (mm)
4	percra	périmètre crânien (mm)
5	pertho	périmètre thoracique (mm)
6	perbg	périmètre du bras gauche (mm)
7	permg	périmètre du mollet gauche (mm)
8	diam	diamètre maximum du bassin (mm)

FIG. 1.1 – Les données de croissance



sultats que l'on peut trouver dans (LAVIT, 1988), puis appliquerons la méthode ACIMO à l'étape de l'intrastructure de STATIS, comme nous l'avons détaillé dans le chapitre 2.5. Nous appliquerons la méthode STATIS dual et y appliquerons ensuite l'ACIMO pour «corriger» les trajectoires de STATIS dual.

1.1 STATIS et l'ACIMO

La méthode STATIS a été appliquée de la manière suivante :

la matrice de pondération des 12 tableaux est choisie uniforme, $\Pi = Id_{12}/12$, et nous utilisons les opérateurs $W_k D$ non réduits.

Les résultats de STATIS ne seront pas étudiés en détails. Le lecteur pourra trouver l'interprétation complète dans (LAVIT, 1988). Nous détaillerons plus l'application de l'ACIMO pour modifier les trajectoires de STATIS.

Le tableau 1.2 donne les variances des premières composantes principales issues de l'ACP de chaque tableau X_k , ($var.acp_k$), les variances des composantes t_k issues de l'ACIMO appliquée à STATIS comme décrit dans la première partie de cette thèse, au chapitre 2.5, ($var.t_k$), et les coefficients μ_k qui mesurent la proximité entre les composantes t_k et la composante associée au compromis u issue de l'ACIMO. Le tableau montre que les composantes t_k ont une variance très proche de celle de la première composante principale de X_k , c'est-à-dire que t_k explique une grosse partie de l'information contenue dans X_k . A l'étape 2, ce résultat est encore vrai. Les coefficients μ_k nous indiquent à la première étape que les tableaux X_1 , X_{10} et X_{11} correspondant aux âges 4, 14 et 15 ans, sont les plus éloignés de la composante compromis u_1 du compromis Y_c . A l'étape 2, u_2 est plus proche des tableaux X_{10} et X_{11} , c'est-à-dire que l'étape 2 explique principalement les variations de ces deux tableaux. u_1 et $u_{c,1}$ ont une corrélation de 1 (arrondi à la sixième décimale), et u_2 et $u_{c,2}$ une corrélation de 0.998. Alors, les composantes de l'ACIMO et les composantes principales de Y_c sont colinéaires et expliquent la même structure commune aux X_k . Mais, l'ACIMO nous dit aussi que la seconde variation commune est surtout liée aux deux derniers âges, soit 14 et 15 ans.

La figure 1.2 représente les 96 variables-compromis vues par l'ACIMO, dont les coordonnées

TAB. 1.2 – Pour chaque X_k , pour les étapes 1 et 2, variance de la composante principale ($var.acp_k$), variance de t_k ($var.t_k$), covariance entre t_k et u (μ_k), et pourcentage de variance de X_k expliquée par la composante t_k ($iner.exp$).

étape 1, k	1	2	3	4	5	6	7	8	9	10	11	12
$var.acp_k$	4.90	5.22	4.90	5.50	5.35	5.04	5.09	5.45	5.35	5.42	4.97	4.45
$var.t_k$	4.87	5.19	4.89	5.50	5.34	5.01	5.06	5.43	5.31	5.38	4.93	4.39
μ_k	18.86	23.01	22.05	26.44	25.31	24.21	24.03	25.43	23.78	23.77	21.48	18.45
$iner.exp$	61.16	65.18	61.15	68.79	66.87	62.86	63.57	68.14	66.80	67.67	62.00	55.37
étape 2, k	1	2	3	4	5	6	7	8	9	10	11	12
$var.acp_k$	1.17	1.09	1.33	0.84	1.14	1.42	1.36	1.14	1.28	1.33	1.71	1.98
$var.t_k$	1.07	0.98	1.30	0.82	1.14	1.41	1.35	1.14	1.27	1.33	1.71	1.99
μ_k	0.60	0.71	1.13	0.77	1.14	1.54	1.44	1.22	1.31	1.40	1.66	2.01
$iner.exp$	14.08	13.07	16.49	10.39	14.31	17.83	16.98	14.33	16.02	16.72	21.46	24.95

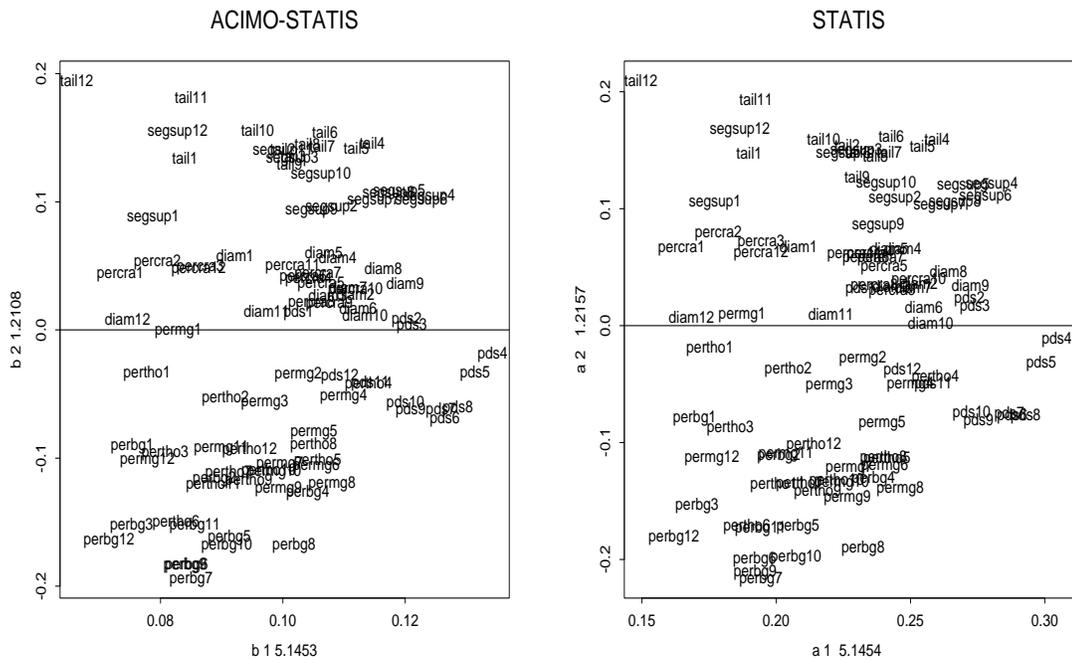
TAB. 1.3 – Variance intra-groupe pour les individus 9, 11, 27 et 18, calculée par rapport à la position compromis, pour les axes 1 et 2 pour les deux méthodes.

individu	axe 1				axe 2			
	9	11	18	27	9	11	18	27
STATIS	0.64	0.81	0.38	0.64	0.70	0.84	2.05	0.96
OMCIA-STATIS	0.05	0.05	0.085	0.15	0.16	0.08	0.6	0.24

sont données par les axes (b_1, b_2) , et vues par STATIS en représentant les deux premiers axes principaux de Y_c (ACP). A la vue de ces dessins, nous voyons que les variables se regroupent par type de variables, soit en 8 groupes. Ainsi, nous pouvons utiliser cette représentation pour interpréter les positions des observations en fonctions des variables morphologiques. Les deux dessins sont équivalents. Ceci montre encore une fois la proximité des résultats associés au compromis dus à l'ACIMO avec ceux dus à l'analyse du compromis dans STATIS. Ainsi, le premier axe est représenté par le poids, et le second axe oppose les variables de taille (*tail* et *segsup*) aux variables de corpulences (*perbg*, *permg*, *pertho*).

La figure 1.3 représente la position compromis des individus, les 30 filles, au centre des «étoiles», pour les deux méthodes. Une «étoile» représente une fillette. Aux extrémités des branches des «étoiles» se situent la position de la fillette en chacune des années. Nous pouvons remarquer que les positions-compromis sont identiques dans les deux méthodes et que la variance des composantes-compromis aussi. Par contre, la position des individus selon les années est différente. Sur le premier axe, les positions relatives des fillettes autour de leur compromis est à peu près identique pour les deux méthodes. Mais elles sont différentes à partir du second axe. La variance intra-groupe (*i.e.* intra-fillette) par rapport à la position compromis semble être plus petite dans la représentation par l'ACIMO. En effet, cela est confirmé par le tableau 1.3 qui donne la variance intra-groupe par rapport au compromis. Le compromis est plus représentatif des trajectoires de l'ACIMO que de celles de STATIS. Dans tous les exemples que nous avons étudiés, la majorité des variances intra de l'ACIMO étaient plus petites que celles de STATIS. Mais, malheureusement, ce résultat n'est pas systématique. Dans STATIS, les tableaux 11 et 12 sont les plus éloignés du compromis. Dans l'ACIMO, comme le second axe les explique, ils sont proches de la position compromis. Les trajectoires peuvent être déduites de ces dessins. Figure 1.4 donne celles des individus 9, 11, 18 et 27.

FIG. 1.2 – Représentation des variables du compromis par l'ACIMO, (b_1, b_2) , et par STATIS, les deux premiers axes principaux de Y_c . Le nombre écrit à côté de l'axe est sa variance.



Considérons la fillette 11. Sa trajectoire est dans le cadran inférieur gauche du dessin. Sa position, est surtout due à sa petite taille et à son faible poids par rapport aux autres fillettes. La trajectoire de STATIS nous indique que la fillette 11 commence à grandir à 7 ans et être plus grande que la moyenne. Mais d'après la trajectoire de l'ACIMO, ce n'est pas vrai. En retournant aux données initiales, nous avons pu voir que cette fille est toujours plus petite que la moyenne et que ses caractéristiques en terme de *perbg* et *segsup* sont proches de la moyenne. Cela explique pourquoi la trajectoire de l'ACIMO est en forme de spirale autour de la position compromis. De la même manière, nous avons remarqué que la fillette 9 est plus stable dans sa croissance que ce qu'il y paraît avec la trajectoire de STATIS. D'un autre côté, la trajectoire de la fillette 18 est quasiment identique à celle de l'ACIMO. Un élément de réponse est que la plupart des différences entre les deux méthodes apparaissent à partir du second axe. Et comme la fillette 18 n'est pas bien représentée sur le premier axe, seules les informations concernant les variables *tail*, *segsup*, *perbg* sont expliquées dans les trajectoires, comme dans l'ACIMO. Pour les fillettes qui sont bien expliquées par le premier axe, la variation due au poids n'existe plus pour la seconde étape dans l'ACIMO, mais pas pour STATIS : nous pouvons dire que dans STATIS l'information est un peu mélangée.

FIG. 1.3 — *Position des individus compromis et positions vues par chaque tableaux dans l'ACIMO, $(u_1/\|u_1\|_D, u_2/\|u_2\|_D)$ et $(t_{k,1}/\|t_{k,1}\|_D, t_{k,2}/\|t_{k,2}\|_D)$, et dans STATIS, $(u_{c,1}, u_{c,2})$ et $(W_k Dc_1/\sqrt{\lambda_{c,1}}, W_k Dc_2/\sqrt{\lambda_{c,2}})$.*

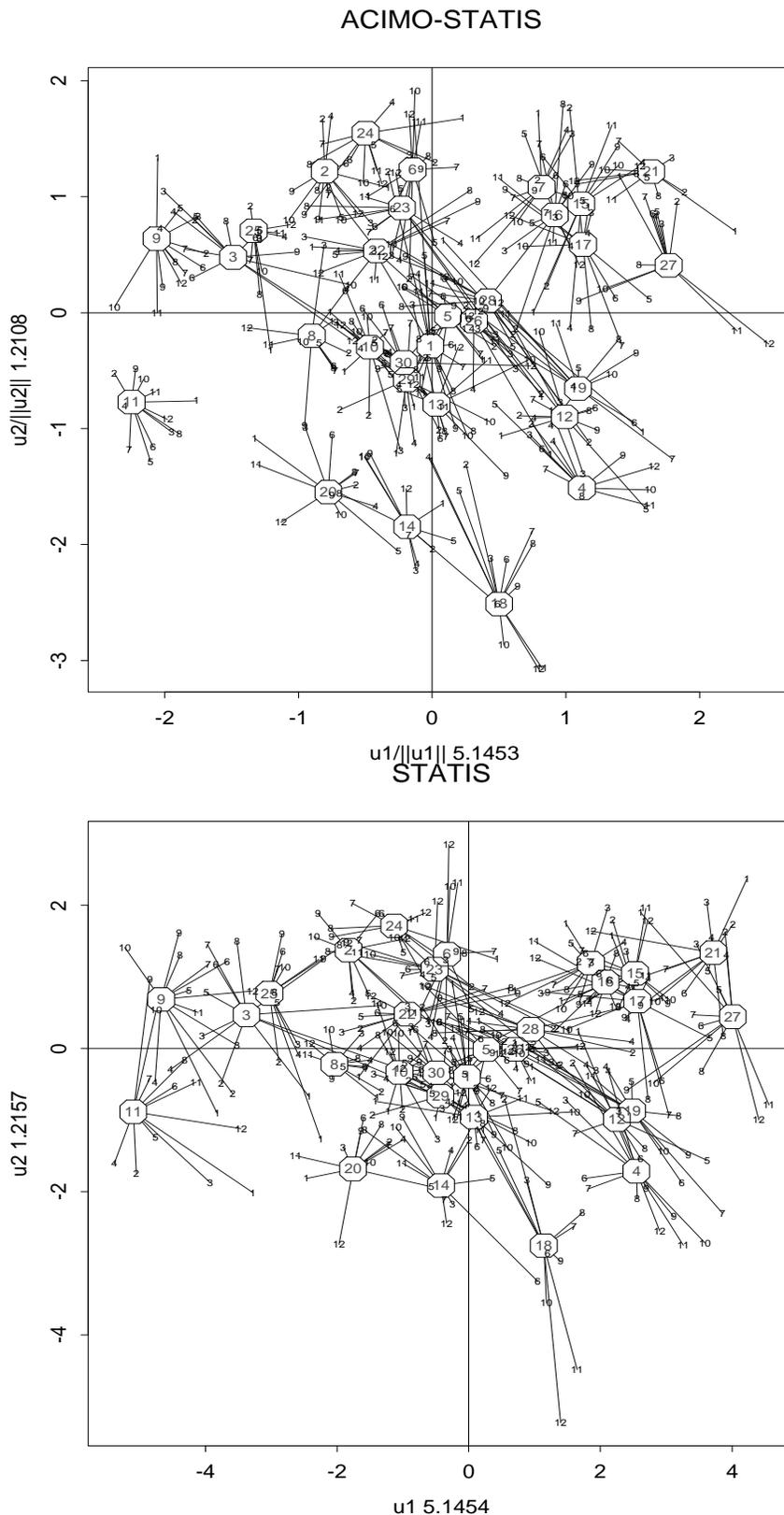
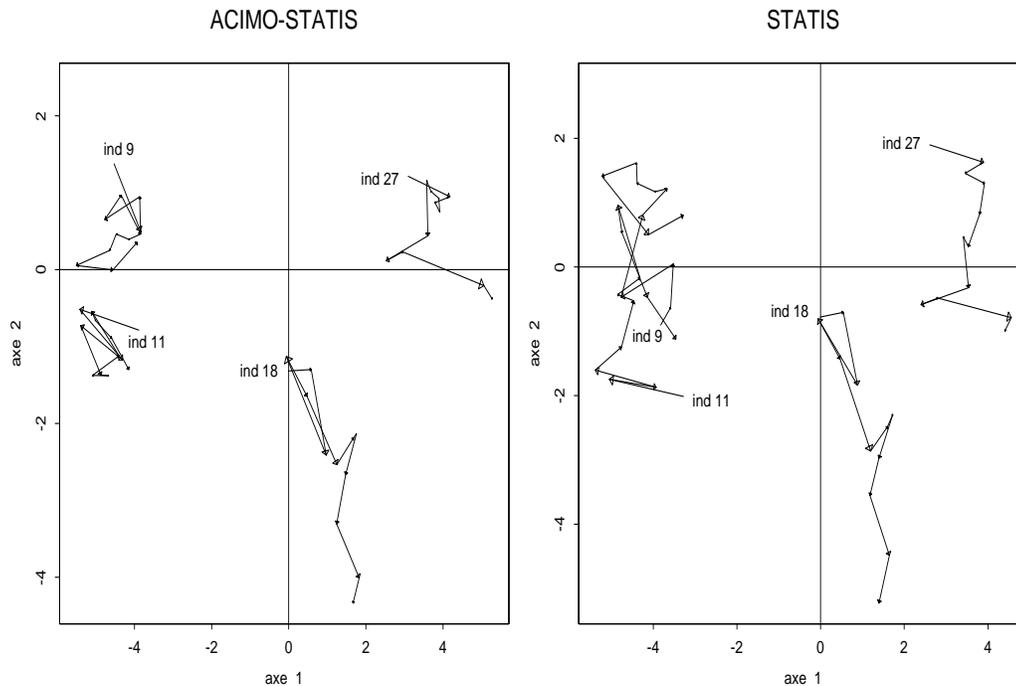


FIG. 1.4 – Trajectoires de STATIS et l'ACIMO des individus 9, 11, 18 et 27.



1.2 STATIS dual et ACIMO

Avec STATIS, les individus étaient globalement étudiés et nous cherchions à expliquer leur différences de croissances les uns par rapport aux autres en fonction des années et des 8 variables morphologiques «générales». Avec STATIS dual, c'est l'évolution de ces 8 variables que nous voulons expliquer en fonction du temps et ce pour les 30 fillettes.

Les conclusions obtenues sont les mêmes qu'avec STATIS, à savoir, que la seconde dimension (deuxième axe) explique ce qu'il y a en commun entre les variables à 14 et 15 ans. Le premier axe explique la structure commune à tous les âges.

Le tableau 1.4 donne les variances des premières composantes principales (var_{acp_k}) issues des ACP non centrées des triplets $(X'_k, Id_{30}/30, Id_8)$ où X'_k est centré en colonne et réduit. Nous rappelons que l'ACIMO est dans ce cas appliquée aux tableaux X'_k centrés et réduits en ligne pour la métrique $D_k = D$. Ainsi, les composantes t_k et u obtenues ne sont pas centrées mais les axes (ou vecteurs de poids) a_k et b le sont. La variance de t_k est donnée par $var.t_k$. Comme pour STATIS, ces variances sont presque égales à celles des composantes principales des ACP. Cela signifie que les composantes de l'ACIMO sont très explicatives de leur groupe (tableau). μ_k mesure, dans un certain sens, la proximité entre X'_k et $Y'c$ mais n'est plus une covariance car les tableaux X'_k ne sont pas centrés en colonnes. Le classement par ordre d'importance des X'_k donné par les μ_k correspond encore à celui donné par $var.t_k$. $iner.exp$ est le pourcentage de variance de X'_k expliquée par la composante $t_{k,a}$, $a = 1, 2$. Dans cet exemple, presque toute la variation commune aux 12 tableaux est expliquée par le premier axe. A la première étape, X_1 et X_{12} sont les plus éloignés du compromis. A la seconde étape, nous remarquons que les deux

TAB. 1.4 – Pour chaque X'_k , pour les étapes 1 et 2, variance des composantes principales ($var.acp_k$), variance des composantes t_k ($var.t_k$), μ_k qui n'est pas une covariance, et pourcentage de variance de X'_k expliquée par la composante t_k ($iner.exp$).

étape 1, k	1	2	3	4	5	6	7	8	9	10	11	12
$var.acp_k$	0.01	0.00	0.01	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.03
$var.t_k$	0.01	0.00	0.01	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.02
μ_k	25.11	26.78	25.03	28.18	27.45	25.79	26.11	27.95	27.45	27.79	25.38	22.39
$iner.exp$	98.70	98.90	98.73	99.08	98.98	98.79	98.82	99.04	98.97	99.02	98.80	98.63
étape 2, k	1	2	3	4	5	6	7	8	9	10	11	12
$var.acp_k$	0.15	0.14	0.17	0.11	0.14	0.18	0.17	0.14	0.16	0.17	0.21	0.24
$var.t_k$	0.14	0.13	0.16	0.10	0.14	0.18	0.17	0.14	0.16	0.17	0.21	0.24
μ_k	1.32	1.31	1.60	1.02	1.43	1.77	1.68	1.43	1.61	1.66	2.12	2.41
$iner.exp$	0.47	0.42	0.60	0.32	0.43	0.58	0.55	0.45	0.49	0.55	0.85	1.33

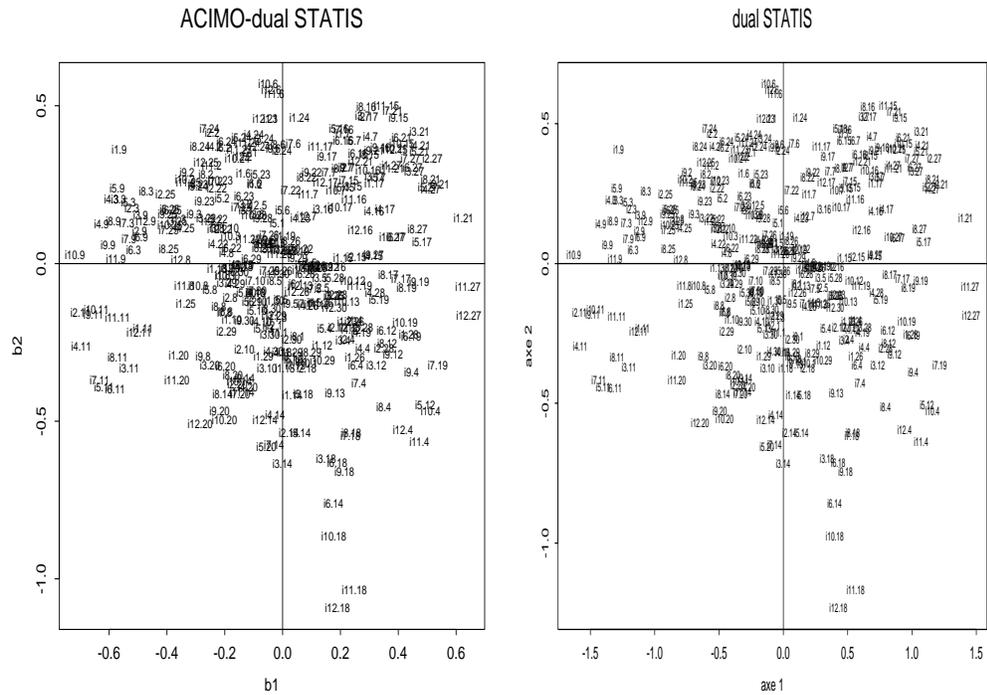
derniers tableaux (X_{11} et X_{12}) sont les plus proches du compromis. La très faible variation expliquée par cette étape est contenue dans ces deux tableaux. La figure 1.5 montre que pour les deux méthodes, tous les individus (filles) sont bien représentés par le premier axe et que la fillette 18 est la seule qui représente le second axe. Avec la figure 1.6, nous pouvons remarquer que le premier axe explique la variable *pds* et que le second axe explique les variables *perbg* (la sixième variable) qui est opposée à la variable *tail*. Contrairement à STATIS dual, l'ACIMO nous dit que les tableaux 11 et 12 ne sont pas si éloignés de la position compromis sur le second axe. Cela correspond à ce que nous avons remarqué au dessus. Selon le premier axe, ces deux tableaux sont éloignés de la position compromis. De plus, cette figure montre que la variance intra-groupe par rapport à la position compromis (un groupe correspondant à une variable) est plus petite en utilisant l'ACIMO qu'en utilisant STATIS dual usuel : les «étoiles» sont moins étalées.

Les trajectoires pour les variables *pds*, *tail*, *perbg* et *diam* sont données par la figure 1.7. Nous pouvons noter que la trajectoire de *perbg*, selon le deuxième axe, est constante tout au long des années avec l'ACIMO. Nous pouvons faire la même constatation en ce qui concerne le poids, *pds* le long du premier axe.

1.3 Conclusion

Les trajectoires qui sont habituellement tracées dans les méthodes STATIS et STATIS dual (LAVIT, 1988) ne sont pas justifiées, dans le sens où ces trajectoires n'optimisent pas de critère : elles sont données par les images des composantes principales du compromis par les opérateurs $W_k D$ sur le plan principal du compromis. L'utilisation de l'ACIMO permet de les «corriger». L'ACIMO donne des positions compromis proches des positions compromis de STATIS (dual) et les composantes $t_{k,a}$ peuvent être considérées comme «corrigeant correctement» les trajectoires. «Corrigeant correctement» est employé dans le sens que les composantes $\{t_{k,a}\}_a$ expliquent différentes parties de X_k et satisfont une optimisation étape par étape. Les deux types de trajectoires sont très proches sur le premier axe, où l'information la plus importante, souvent évidente, est détectée. Mais des différences apparaissent pour certains individus ou variables (selon que l'on soit dans le cas STATIS ou STATIS dual) à partir du second axe, et la variance intra-groupe d'individus (ou variables) est améliorée pour la ma-

FIG. 1.5 — Représentation des individus compromis (fillettes) donnée par l'ACIMO, (b_1, b_2) , et par STATIS dual, les deux premières composantes principales de Y_c .



porité d'entre eux, c'est-à-dire qu'elle est réduite avec l'utilisation de l'ACIMO. De plus, ces trajectoires décrivent les données d'une meilleure façon, plus juste. Cependant, quelquefois, les trajectoires usuelles de STATIS sont quasiment égales à celles obtenues par l'utilisation de l'ACIMO. Cela signifie, dans ce cas, que les trajectoires de STATIS sont «bonnes» bien qu'elles aient été tracées sans aucune optimalité. Pour d'autres figures, les trajectoires obtenues par les deux techniques sont très différentes à partir du second axe. Alors, l'utilisation de l'ACIMO donne celles qui sont le mieux ajustées aux données.

FIG. 1.6 — Positions des variables compromises et des variables par année, selon l'ACIMO, $(u_1/\|u_1\|_D, u_2/\|u_2\|_D)$ et $(t_{k,1}/\|t_{k,1}\|_D, t_{k,2}/\|t_{k,2}\|_D)$, et selon STATIS dual, $(u_{c,1}^d, u_{c,2}^d)$ et $(V_k C_1^d/\sqrt{\lambda_{c,1}^d}, V_k C_2^d/\sqrt{\lambda_{c,2}^d})$.

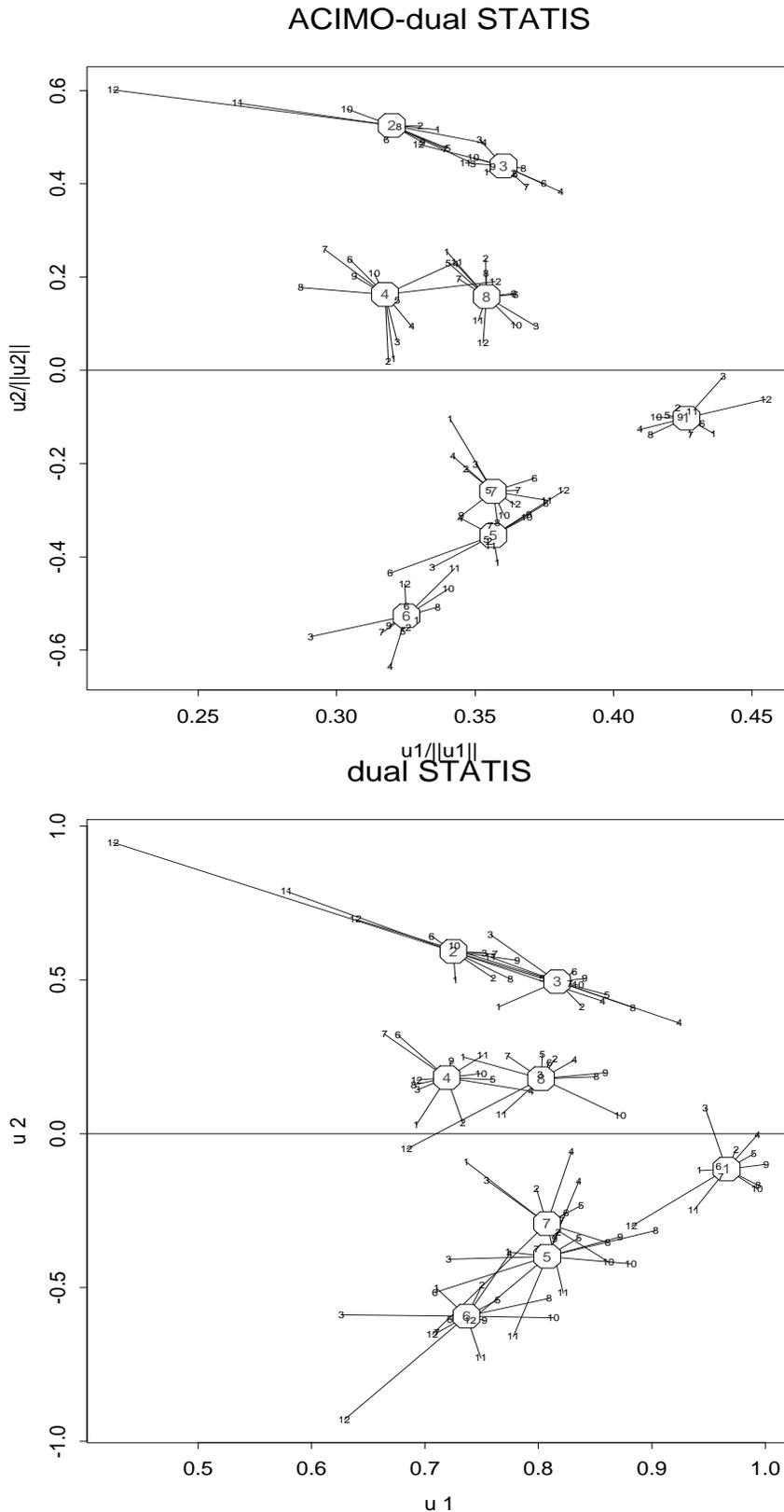
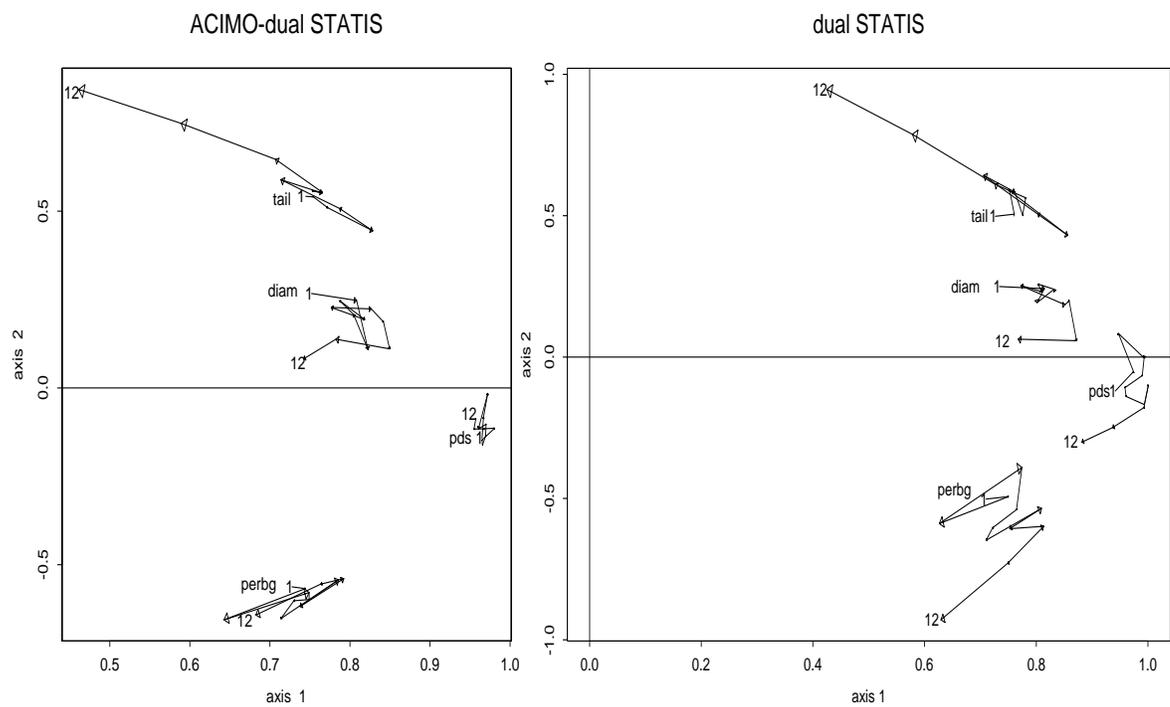


FIG. 1.7 – Trajectoires pour le poids (*pds*), la taille (*tail*) et le diamètre du bras gauche (*perbg*), et le diamètre du bassin (*diam*).



Chapitre 2

Environnement : ACIMO-PLS et ACIMO-PLSS

Les données que nous utilisons ici sont des données environnementales évolutives. Elles nous ont été fournies par M. Simier et D. Nandris, de l'IRD de Montpellier. SIMIER, BLANC, PELLEGRIN & NANDRIS (1999) les ont analysées à l'aide de STATICO (I-3.2).

2.1 Présentation des données

Depuis 1992, une enquête épidémiologique est menée par l'IRD sur des caféiers (*Coffea Arabica*) en Nouvelle Calédonie répartis selon $l = 15$ sites. Les données que nous avons concernent l'année 1994. Chaque année, à intervalles de huit temps réguliers ($N = 8$), $M = 6$ variables pathologiques sont mesurées (Tab. 2.1). Deux variables climatologiques, mesurées en continu, correspondent à la hauteur des pluies cumulées et à la température minimale. De plus, $K = 6$ variables environnementales ont été relevées sur les 15 sites au début de l'étude, décrivant le milieu (Tab. 2.1).

Des études effectuées par les propriétaires des données n'ont pas montré d'influence majeure des variables climatologiques (dépendant du temps k), aussi, nous ne les utiliserons pas. Nous voulons donc étudier l'influence des variables de milieu sur la présence de maladies. Les variables de milieu forment le tableau X de prédicteurs et les variables pathologiques forment $N = 8$ tableaux (8 dates) de taille 15×6 décrivant l'évolution des maladies. Pour faire cela, nous allons utiliser l'ACIMO-PLS (§I-5.5) puis l'ACIMO-PLSS (§II-2.3).

2.2 ACIMO-PLS

Les données sont centrées et réduites pour la métrique $D = Id_{15}/15$. Les métriques Q^x et Q_n^y sont les métriques identités. L'inertie des tableaux Y_n , avant réduction, a tendance à croître, puis décroître à partir de $n = 6$. Pour cette raison, nous avons choisi de les réduire. Les inerties initiales sont respectivement égales à 231.9, 578.2, 740.9, 900.7, 1110.4, 1176.5, 1076.6 et 885.6. Cela signifie que le nombre de caféiers malades croît très vite pour certaines parcelles durant les 6 premières observations de la maladie, puis à la 7^{ième} mesure la tendance diminue, c'est-à-dire que des caféiers d'autres parcelles deviennent malades eux aussi. Après

TAB. 2.1 – Description des variables

Variables Pathologiques (Y)		
Code	Variable	Description
NSSROU	rouille	Maladie la plus répandue et dommageable
NSSANT	anthracnose	Maladie fréquente et peu dommageable
NSScerco	cercosporiose	Maladie rare mais sévère
%CFneoMAL	maladie feuilles	% cum. des feuilles devenues malades
%CFTMal	chute feuilles malades	% cum. feuilles malades tombées
%CFTSain	chute feuilles saines	% cum. feuilles saines tombées
Variables de milieu (X)		
Code	Variable	
fert	fertilité du sol	
struct	structure du sol	
cap	capacité rétention d'eau	
pH	pH moyen du sol	
alt	altitude	
ombr	degré d'ombrage	

réduction, tous les tableaux ont, bien sûr, une inertie égale à 6.

A chaque étape a de l'ACIMO-PLS, nous pouvons comparer la variance de la composante $u_{n,a}$, associée à Y_n avec la variance de la composante associée à Y_n issue de la $a^{\text{ième}}$ étape de la régression PLS de Y_n sur X , ainsi qu'avec la variance de la composante principale d'ordre a de Y_n . Plus $u_{n,a}$ a une variance proche de celle de la composante principale, plus est explicative de Y_n . Ces résultats apparaissent dans le tableau 2.2. Ces résultats nous permettent d'étudier la qualité des composantes $u_{n,a}$.

Il apparaît au travers de ces résultats que les deux premiers tableaux ($n = 1$ et $n = 2$) se démarquent des tableaux Y_4 à Y_8 , ces derniers semblent avoir une structure commune. Y_3 se situe en ces deux groupes. L'axe 1 explique la structure commune aux tableaux Y_4 à Y_8 : les cov^2 à l'étape 1 sont les plus élevées et sont assez proches les unes des autres pour ces 5 tableaux. Alors que les tableaux Y_1 et Y_2 ont les plus petites cov^2 , c'est-à-dire que leurs composantes $u_{n,1}$ sont les moins covariées avec la composante explicative t_1 . Sur l'axe 2, la séparation est moins nette, tandis que l'axe 3 explique ce qu'il y a de commun entre Y_1 et Y_2 et éventuellement Y_3 .

Les composantes $u_{n,1}$ pour $n = 4$ à $n = 8$ ont une variance pas très éloignée de celle de la composante principale du tableau qu'elles représentent. Cela signifie qu'elles sont bien explicatives de leur groupe, en plus d'être proche de t . Et cette variance est encore plus proche de la variance des composantes PLS.

Ces premiers résultats correspondent avec l'évolution des inerties des tableaux avant réduction. En effet, les deux premiers tableaux avaient une faible inertie (231.9 et 578.2), tandis que celle des tableaux 4 à 8 étaient comprises entre 885 et 1111.

Les pourcentages de variance expliquée par les modèles en fonction de a sont donnés dans le tableau 2.3. Nous retrouvons les résultats précédents, à savoir, que les 3 premières dates sont nettement moins bien modélisées par les variables de milieu que les 5 dates suivantes

TAB. 2.2 – Résultats numériques. μ_a donne l'optimum du critère à l'étape a . cov^2 donne les covariances carrées partielles $cov^2(u_{n,a}, t_a)$. *VarianceACP* donne la variance des composantes principales (ACP) d'ordre a de Y_n . *Variance* donne la variance de $u_{n,a}$, composante de l'ACIMO-PLS. *VariancePLS* donne la variance des composantes PLS issues des 8 régressions PLS de Y_n sur X .

n	1	2	3	4	5	6	7	8
	étape 1							$\mu_1 = 23.105$
cov^2	1.923	1.548	2.218	3.086	3.511	3.704	3.639	3.477
<i>VarianceACP</i>	2.592	2.456	2.212	2.530	2.705	2.797	2.891	2.881
<i>VariancePLS</i>	1.658	1.623	1.878	2.232	2.461	2.529	2.603	2.508
<i>Variance</i>	1.671	1.434	1.834	2.224	2.444	2.494	2.539	2.426
	étape 2							$\mu_2 = 6.619$
cov^2	0.958	0.765	0.848	0.855	0.876	0.766	0.778	0.772
<i>VarianceACP</i>	1.433	1.470	1.720	1.576	1.599	1.663	1.644	1.671
<i>VariancePLS</i>	2.248	1.785	1.563	1.599	1.584	1.648	1.770	1.881
<i>Variance</i>	1.612	1.359	1.518	1.598	1.632	1.643	1.749	1.827
	étape 3							$\mu_3 = 4.606$
cov^2	0.818	0.954	0.628	0.432	0.454	0.501	0.453	0.366
<i>VarianceACP</i>	1.229	0.962	1.375	1.213	1.294	1.078	1.065	1.130
<i>VariancePLS</i>	1.028	0.855	0.742	1.012	1.210	0.934	0.925	1.030
<i>Variance</i>	1.537	1.729	0.961	0.982	1.201	1.093	1.165	1.349

TAB. 2.3 – Pourcentage de variance expliquée par les modèles en fonction de a

	a = 1	a = 2	a = 3	a = 4	a = 5
X	34.14	49.38	71.41	87.48	95.37
Y₁	16.16	34.16	44.97	50.28	51.96
Y₂	13.00	27.38	39.98	41.78	45.54
Y₃	18.63	34.58	42.88	46.68	50.63
Y₄	25.93	42.00	47.71	50.64	52.99
Y₅	29.50	45.95	51.97	53.95	55.50
Y₆	31.13	45.52	52.14	54.20	56.46
Y₇	30.58	45.21	51.19	54.00	56.73
Y₈	30.58	45.21	51.19	52.18	56.43

(Fig 2.1).

Au regard de ces variances expliquées, nous aurions tendance à croire que les modèles à retenir sont ceux d'ordre 3, car on observe une stabilisation du pourcentage cumulé des variances expliquées (Fig 2.1). L'utilisation de la validation croisée nous dit que ces modèles sont effectivement les moins instables en terme de prédiction (Fig. 2.2). La valeur du PRESS à l'étape 1 est proche de celui à l'étape 3. Cependant, nous n'avons pas retenu $A = 1$ car les pourcentages de variance expliquée sont assez faibles.

Ces résultats ne sont pas très bons, car le rang des modèles est au maximum 6. Cela peut venir de plusieurs causes. La modélisation linéaire ne convient peut être pas, et pour cela nous avons essayé l'ACIMO-PLSS (voir section suivante), soit les variables de milieu n'ont pas un très grand pouvoir de prédiction sur les variables pathologiques.

La figure 2.4 nous permet d'étudier la qualité des modèles partiels d'ordre 1. Le même

FIG. 2.1 – Evolution des pourcentages de variance expliquée par les modèles \hat{Y}_n^a en fonction de a

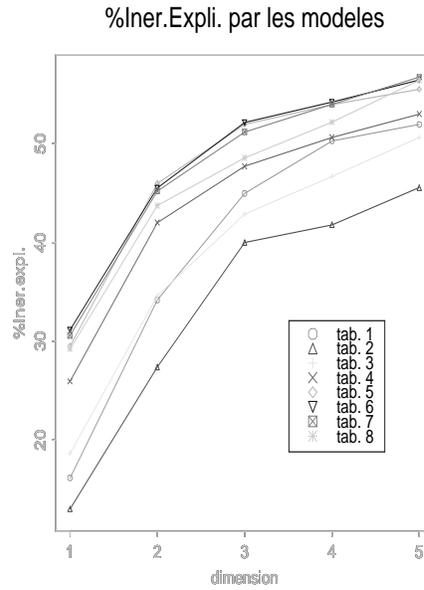


FIG. 2.2 – PRESS total et PRESS par variable en fonction de a

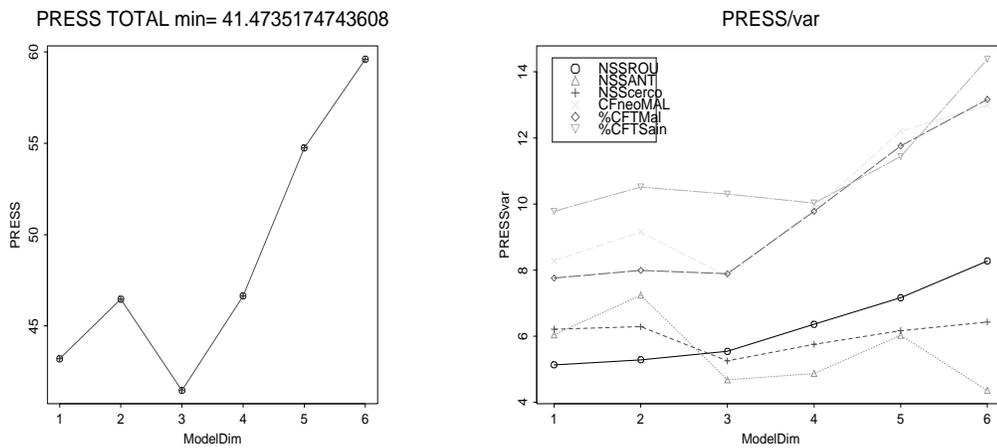
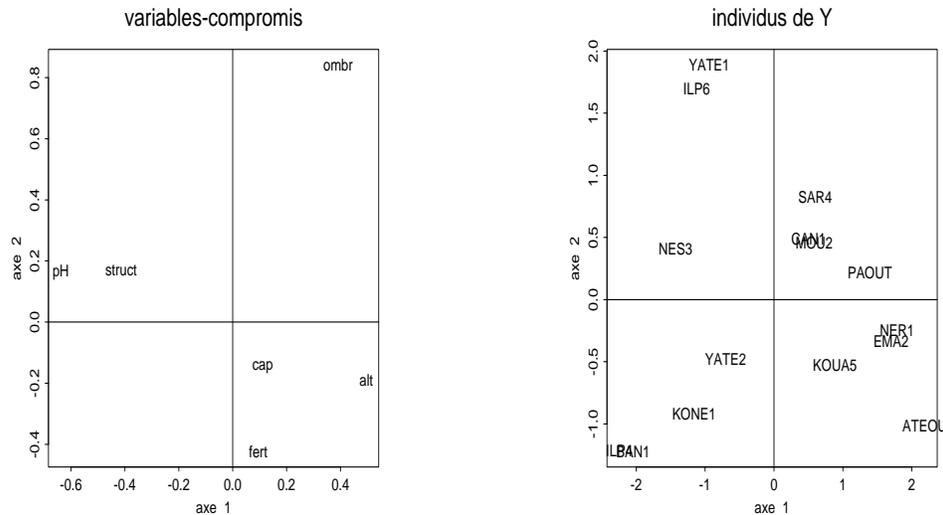


FIG. 2.3 – Représentation des variables explicatives et des individus prédicteurs dans le plan 1,2.



type de graphique est faisable pour chaque dimension. Le premier graphique représente la régression simple de la composante PLS u_1 sur t_1 . Les suivants représentent la régression simple des composantes partielles ACIMO-PLS $u_{n,1}$ sur t_1 . Rappelons (voir I-5.5) que t_1 est égale à la composante de la régression PLS usuelle du tableau conjoint Y sur X , et que $u_1 = \sum_{n=1}^N \|b_1[n]\| u_{n,1} = \sum_{n=1}^N \sqrt{\lambda_{n,1}} / \sqrt{\mu_1} u_{n,1}$, où b_1 désigne le vecteur de poids associé à Y ($u_1 = Yb_1$) de la régression PLS. La valeur *coeff* au dessus des graphiques correspond à $\|b_1[n]\|$. L'ACIMO-PLS met en évidence que la composante PLS u_1 tient plus compte des tableaux Y_n qui sont le mieux expliqués par t_1 , au détriment des autres.

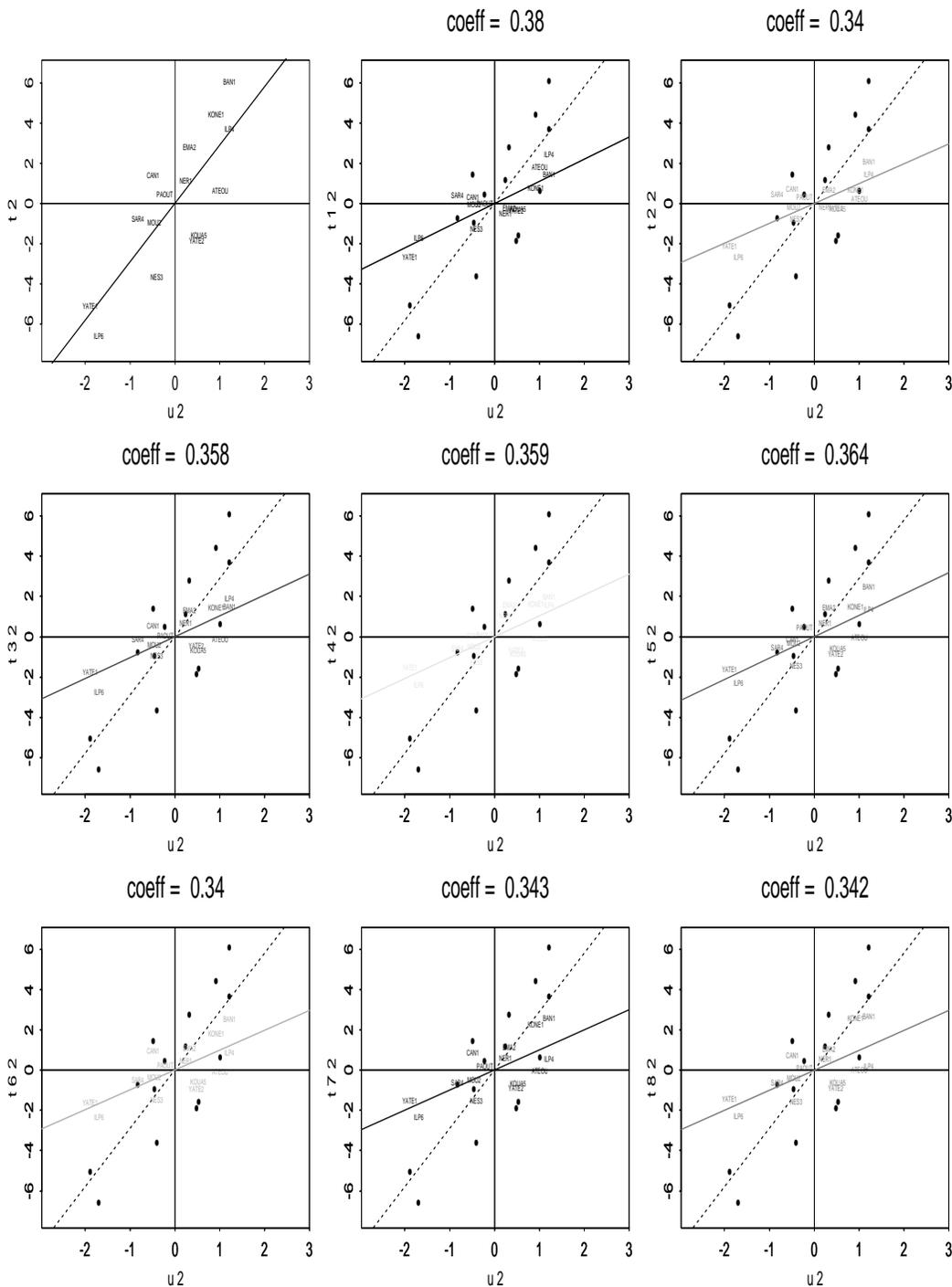
Les tableaux Y_1 et Y_2 , qui ont les plus petits *coeff*, présentent des modèles partiels de rang 1 de moins bonne qualité que celui du tableau Y_6 , par exemple. La qualité de la régression de u_1 sur t_1 pourrait être meilleure : on remarque quelques observations au niveau du quadrant inférieur gauche qui s'éloignent de la droite de régression. Peut-être qu'il y a une légère non-linéarité dans les données.

Il apparaît (Fig. 2.5) que les modèles partiels d'ordre 2 ne sont pas de très bonne qualité. De plus, ici c'est Y_1 qui est le mieux modélisé.

La figure 2.3 permet l'interprétation des axes 1 et 2. La représentation des variables explicatives est donnée par les vecteurs de poids (a_1, a_2) et la représentation des sites est faite par (t_1, t_2) , comme en régression PLS. L'axe 1 oppose de fortes valeurs de pH et de structure aux sites ayant une haute altitude. L'axe 2 oppose les sites à fort degré d'ombrage à ceux ayant un sol très fertile. Si nous faisons la représentation du plan (a_2, a_3) nous verrions que l'axe 3 représente les sites fertiles et ayant une grande capacité de rétention d'eau.

Les données étant évolutives dans le temps, il est intéressant de voir la trajectoire des variables pathologiques. La figure 2.6 représente la corrélation des variables réponses des 8 dates avec les composantes explicatives 1 et 2. D'une manière générale, nous voyons que les deux premières dates se démarquent des autres pour toutes les variables sauf pour *NSScerco*. La date 3 est en général entre les deux groupes de dates. La maladie *rouille* est très liée au

FIG. 2.5 – Qualité des modèles partiels d'ordre 2. Régressions simples de $u_2, u_{1,2}, \dots, u_{8,2}$ sur t_2 . $\text{coeff} = \|b_2[n]\|$. La droite en pointillés et les points dans les 8 derniers graphiques représentent la régression de u_2 sur t_2 (premier graphique).



premier axe, surtout à partir de la date 4. Le nombre de feuilles saines tombées, $\%CFTSain$, est lié au deuxième axe pour les dates 3 à 8 et est complètement orthogonal au nombre de feuilles malades tombées, $\%CFTMal$, qui est lié à l'axe 1. La *cercosporiose*, *cercospora*, est stable dans le temps et ne caractérise pas plus l'axe 1 que l'axe 2, s'opposant tout de même à l'*anthracnose*.

En conclusion, nous pouvons dire que l'ACIMO-PLS a décelé une forte structure commune aux dates 4 à 8, qui est due à la maladie *rouille* et la maladie *anthracnose*, croissantes. Les feuilles malades (sans doute de la rouille) tombent de plus en plus avec le temps, à partir de la date 4. Cependant, nous pouvons noter une petite accalmie de la maladie à partir de la date 7. Les parcelles malades de la rouille sont celles à haute altitude, plutôt ombragées, ayant un sol à faible pH et mal structuré; en opposition aux parcelles malades de l'*anthracnose*. La maladie *cercosporiose* est présente à toutes les dates de manière stable, dans les parcelles très ensoleillées ayant un sol au pH plutôt élevé. Les caféiers perdant leurs feuilles saines se situent dans des régions plutôt fertiles et très ensoleillées. Cette caractéristique ne paraît pas liée à une maladie particulière.

L'ACIMO-PLS, facile à obtenir à partir d'une régression PLS usuelle, fournit des outils permettant l'étude des modèles ajustés par tableaux réponses. La PLS seule ne permet pas.

Les résultats obtenus ici sont cohérents avec ceux de SIMIER et al. (1999). Cependant les modèles ne sont pas très bons pour faire de la prédiction. La qualité des modèles partiels de rang 1 n'est pas très bonne. Aussi, nous avons appliqué l'ACIMO-PLSS.

2.3 L'ACIMO-PLSS

Nous avons essayé l'ACIMO-PLSS avec des splines de degrés 0, 1 et 2 en combinaison avec 1, 2 ou 3 noeuds équi-espacés intérieurs. De plus, nous avons essayé de ne coder que les trois variables *fert*, *pH* et *alt* qui sont continues alors que les trois autres ne prennent que 3 ou 4 valeurs. Les résultats, en terme de variance expliquée et de PRESS, ne sont réellement meilleurs (Tab. 2.4). Aussi, nous ne donnerons pas une analyse détaillée.

On remarque (Tab. 2.4), cependant, que les codages des prédicteurs dans une base de B-splines de degré 1 avec deux noeuds intérieurs, donnent les meilleurs résultats en terme de variance expliquée des réponses. Les dates 1 et 2 sont moins bien modélisées qu'avec l'ACIMO-PLS, leur structure étant différentes des autres dates. A la première étape, aucun tableau n'est significativement mieux expliqué par un tel modèle non linéaire. Mais à partir de la seconde étape, les dates 3 à 8 sont mieux expliquées ainsi qu'avec un modèle linéaire. Cela correspond avec ce que nous avons constaté dans la section précédente. A la première étape, les modèles partiels d'ordre 1 de l'ACIMO-PLS étaient de qualité moyenne mais à partir de $a = 2$ leur qualité devenait médiocre.

Dans le tableau 2.4, la situation 8 correspond au cas où seules les variables *pH*, *fert* et *alt* ont été codées à l'aide de splines de degré 1 et 1 noeud, et où la position de ce noeud a été choisie comment étant celle qui donnait en moyenne les meilleures reconstitutions des tableaux réponses Y_n à l'étape $a = 1$. Dans la situation 9, les noeuds pour ces variables sont choisis de la même façon, et un noeud intérieur placé au milieu est choisi pour les variables *ombr*, *cap* et *struct*.

FIG. 2.6 – *Evolution des variables pathologiques en fonction du temps. Corrélations entre les variables et t_1 et t_2 .*

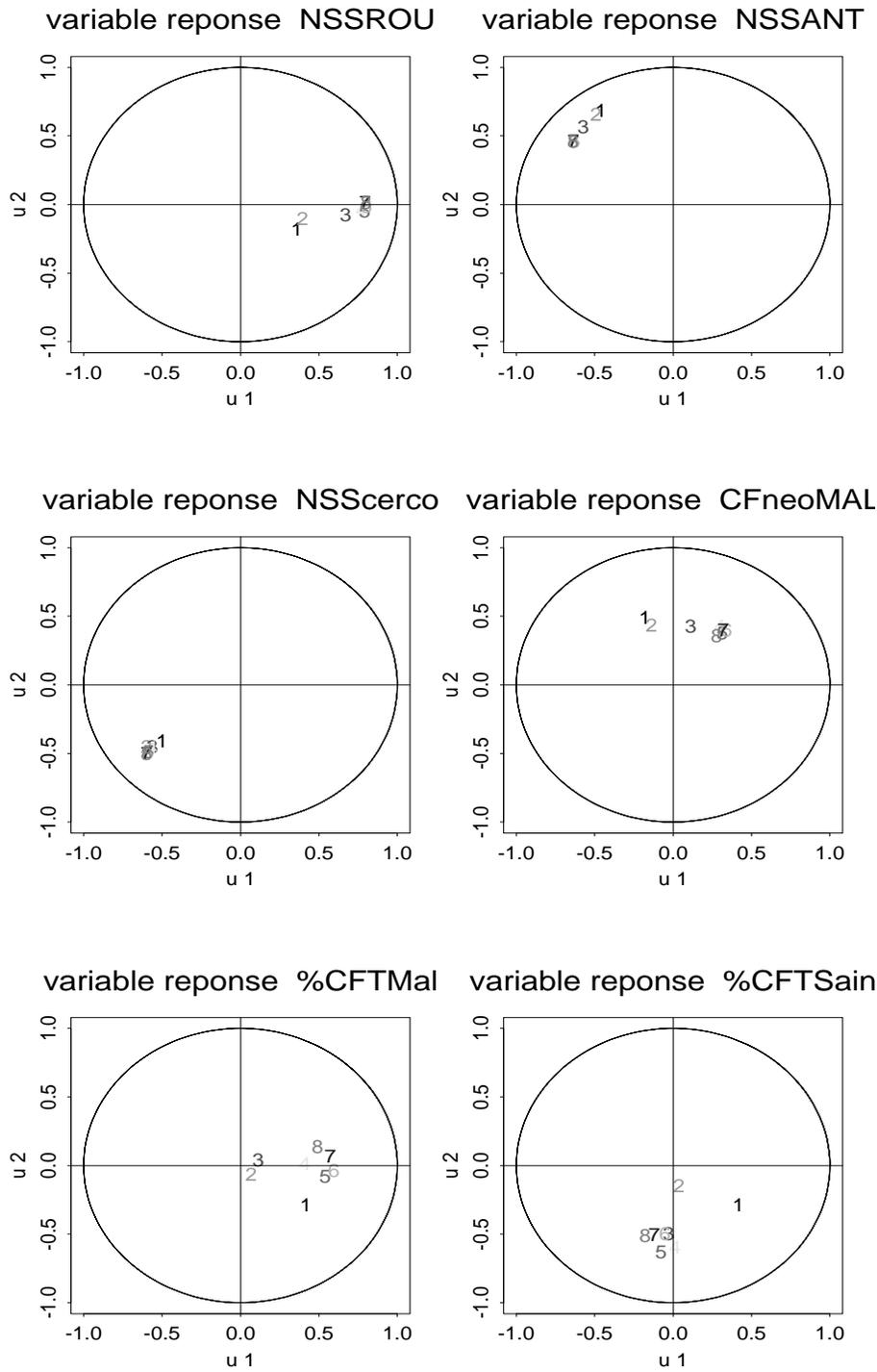
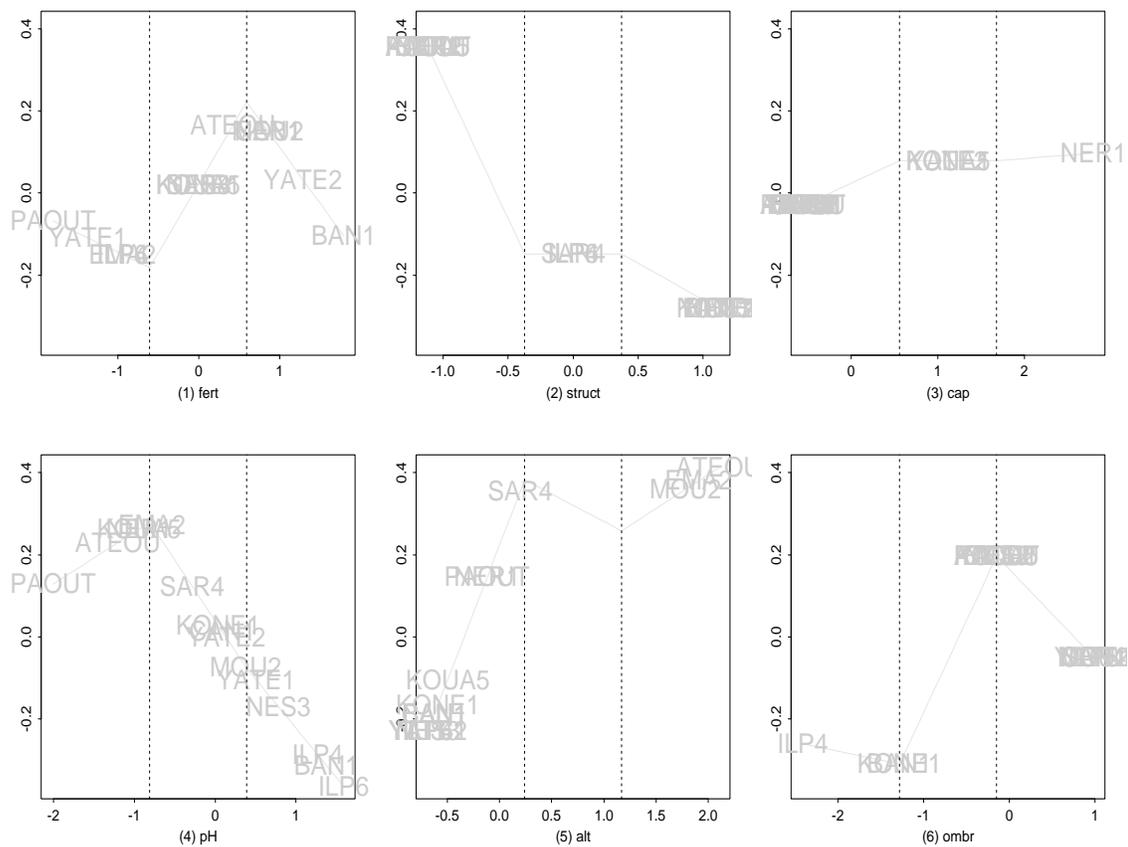


FIG. 2.7 – Influence des prédicteurs sur t_1 , dans l'ACIMO-PLSS, 2 noeuds et ordre 2.



Ces données donnent un exemple de situation où soit le modèle non-linéaire n'apporte pas plus que le modèle linéaire, même si celui-ci n'est pas très bon, soit la forme de non-linéarité est mal choisie.

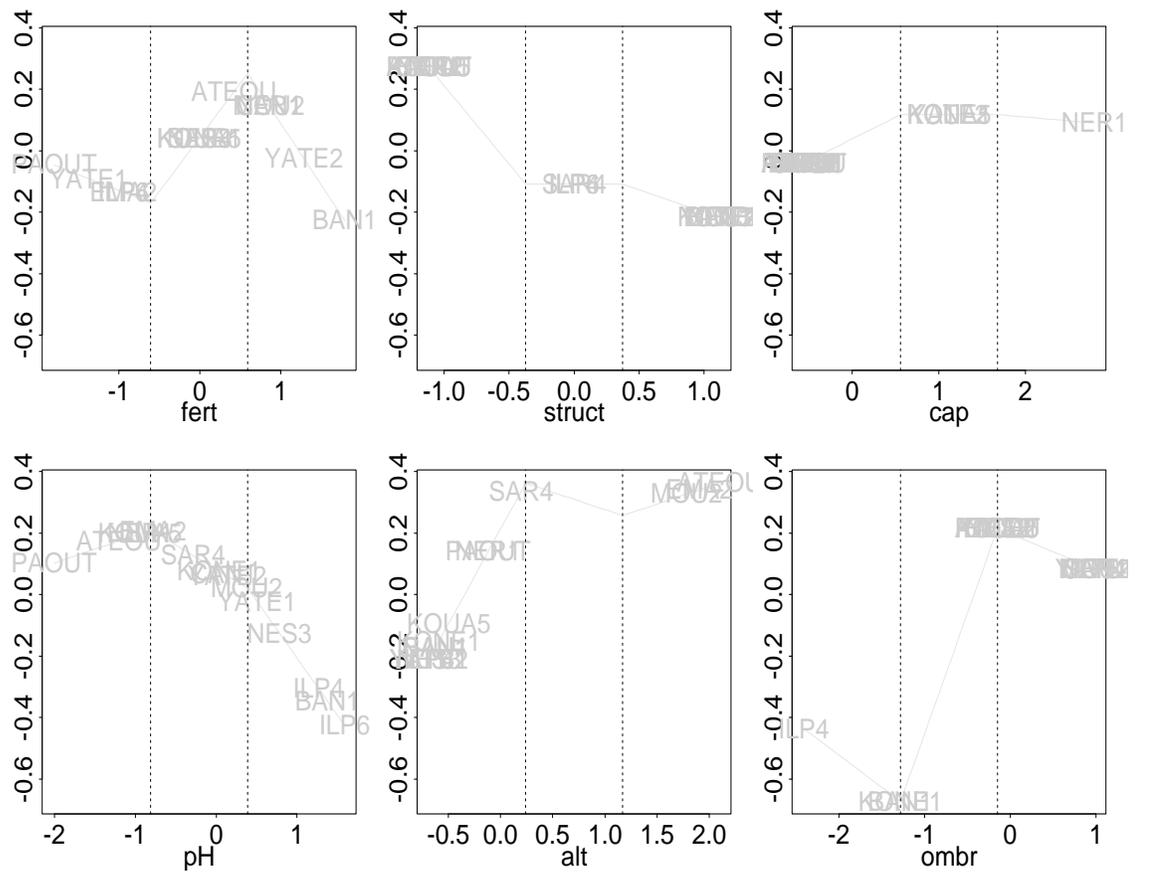
Les figures 2.7 et 2.8 représentent respectivement les fonctions coordonnées des prédicteurs influant sur t_1 et sur la variable $NSSROU$ à la date 6 pour un modèle avec trois composantes. L'expression de ces fonctions coordonnées est détaillée dans le chapitre II-2.2. L'amplitude d'une fonction coordonnée indique son influence. Ainsi, nous pouvons classer par ordre d'importance décroissante l'influence des prédicteurs sur t_1 : *struct*, *alt*, *pH*, *ombr*, *fert* et *cap*. La non-linéarité est forte pour *alt* et *fert*. La fonction coordonnée de *pH* est presque linéaire. Sa non-linéarité est due à un seul individu, le site *PAOUT*.

Avec trois composantes, à la date 6, la variable $NSSROU$, la rouille, est surtout due aux prédicteurs *ombr*, *pH*, *alt* et *fert*. La liaison entre $NSSROU$ et le *pH*, est quasi-linéaire, de même avec *struct*.

TAB. 2.4 – *Pourcentage de variance expliquée par les modèles pour les trois premières dimensions. «oi,nj» signifie ACIMO-PLSS avec codage des variables dans une base de B-splines utilisant j noeuds intérieurs équi-espacés et des splines d'ordre i. Les situations 8 et 9 sont particulières et sont expliquées dans le texte.*

	1) ACIMO-PLS			2) ACIMO-PLSS : o2,n1			3) ACIMO-PLSS : o2,n2		
	a = 1	a = 2	a = 3	a = 1	a = 2	a = 3	a = 1	a = 2	a = 3
X	34.14	49.38	71.41	24.36	47.38	58.61	26.15	42.48	52.42
Y₁	16.16	34.16	44.97	13.46	19.89	41.57	11.22	18.38	43.53
Y₂	13.00	27.38	39.98	17.28	24.40	38.14	14.54	22.24	39.16
Y₃	18.63	34.58	42.88	21.23	27.31	39.99	18.69	30.16	45.20
Y₄	25.93	42.00	47.71	26.39	33.11	47.59	25.62	38.20	55.61
Y₅	29.50	45.95	51.97	28.52	36.62	53.49	29.53	43.21	62.62
Y₆	31.13	45.52	52.14	29.65	40.41	54.44	31.25	47.14	64.09
Y₇	30.58	45.21	51.19	28.97	41.36	55.15	30.71	48.33	65.59
Y₈	30.58	45.21	51.19	27.57	41.46	54.73	29.27	48.47	65.58
	4) ACIMO-PLSS : o2,n3			5) ACIMO-PLSS : o3,n1			6) ACIMO-PLSS : o3,n2		
	a = 1	a = 2	a = 3	a = 1	a = 2	a = 3	a = 1	a = 2	a = 3
X	22.38	41.66	51.77	26.24	45.86	59.93	24.94	43.68	57.32
Y₁	14.18	19.20	39.31	12.39	20.77	39.94	13.47	21.12	38.32
Y₂	17.50	22.98	34.94	15.29	24.63	36.76	16.20	24.51	34.95
Y₃	22.02	29.54	41.32	19.59	28.12	39.20	20.60	28.89	38.21
Y₄	27.30	36.20	51.42	25.64	33.96	46.77	26.25	34.85	46.10
Y₅	29.70	40.45	58.22	28.72	37.55	52.76	29.14	38.77	52.51
Y₆	30.60	43.82	58.66	30.24	41.22	53.46	30.38	42.31	53.08
Y₇	29.99	44.41	59.99	29.53	41.79	54.19	29.62	42.81	54.01
Y₈	28.70	44.66	60.18	28.09	41.58	53.73	28.21	42.76	53.93
	7) ACIMO-PLSS : o3,n3			8) o2,n1, NdsMax+Lin.			9) o2,n1, NdsMax+Mil.		
	a = 1	a = 2	a = 3	a = 1	a = 2	a = 3	a = 1	a = 2	a = 3
X	25.70	42.78	54.58	33.24	55.69	80.52	25.34	48.46	60.36
Y₁	12.92	30.57	40.82	10.04	29.26	37.27	14.04	22.02	43.75
Y₂	15.37	24.09	36.95	11.54	24.50	34.05	16.61	25.42	38.89
Y₃	19.92	26.56	41.75	16.54	29.46	36.16	22.10	28.89	40.81
Y₄	25.95	35.87	50.76	22.37	35.42	40.76	27.85	34.66	47.78
Y₅	29.19	40.14	57.55	25.07	39.58	44.29	30.33	37.97	53.38
Y₆	30.51	43.48	58.25	26.49	39.68	46.16	31.42	41.20	54.09
Y₇	29.79	43.94	59.74	28.03	39.14	47.30	30.56	41.76	54.65
Y₈	28.40	43.98	60.01	28.07	37.25	46.29	29.03	41.60	54.20

FIG. 2.8 – Influence des prédicteurs sur la rouille, *NSSROU*, à la date 6, avec trois composantes.



Chapitre 3

Ecologie : ACIMOG-PLS1

Les données réelles que nous allons utiliser ici sont des données environnementales que nous avons traité en collaboration avec Nabila Mazouni (CEPRALMAR) et Jean-Claude Gaertner (COM) dans le cadre du Programme National - Environnement Côtier, Activité de Recherche Thématique n°4 de l'année 2000, au sein de l'équipe Mazouni, Sabatier, Gaertner et Durbec. Ces résultats sont contenus dans la bilan d'activité 2000 de l'ART4 (MAZOUNI, VIVIEN, SABATIER, GAERTNER & DURBEC, 2001) et ont été présentés aux XXXIIIèmes Journées de Statistiques (VIVIEN & SABATIER, 2001b). Une publication est en cours de préparation.

3.1 Les données

Il s'agit de modéliser l'abondance de la population de palourdes *Ruditapes decussatus* à partir des facteurs de l'habitat (granulométrie, profondeur, richesse nutritive) et de la structure des communautés benthiques dans l'étang de Thau.

Les données disponibles ont été recueillies par Nabila Mazouni dans le cadre du Programme Palourde qui participe au Chantier Lagunes-Méditerranéennes du PNEC. Un échantillonnage de plusieurs compartiments de l'écosystème a été réalisé au cours des années 1998 et 1999. L'ensemble des prélèvements a été réalisé selon une grille d'échantillonnage unique avec un pas de 0.5 min pour la latitude et 0.66 min pour la longitude. Les stations, qui représentent les individus ($l = 64$), ont toutes été positionnées de façon précise à l'aide d'un GPS (Global Position System) équipé d'un différentiel. Le tableau 3.1 récapitule la composition des données. Les variables ont été réparties en groupes par N. Mazouni après des études statistiques préliminaires et des connaissances spécifiques des données. Les 4 dernières variables mesurées

TAB. 3.1 – *Composition des données et unités de valeur*

Compartiment	Nombre de variables	Type de variables	unités
MACROFAUNE	37	Taxons	Individus.m ²
SEDIMENTS	4	Tailles de particules	Pourcentages (%)
COLONNE D'EAU	5	Profondeur	Mètres (m)
		Oxygène dissous	mg.l ⁻¹
		Chlorophylles	µg.l ⁻¹
		Phéopigments	µg.l ⁻¹
		Carbone organique particulaire	µg.l ⁻¹

TAB. 3.2 – *Groupes et codage des variables*

groupe	type	variables
y	<i>palourdes</i>	Mpal
X_1	<i>bivalves</i>	Mcom Mlu Mcl Mdt Mca Mmy
X_2	<i>polychètes errantes</i>	PEap PEph PEsy PEne PEgl PEeu
X_3	<i>espèces liées aux substrat sableux</i>	PSow PSte Ccrab2 Eophi SIPON PHORO
X_4	<i>polychètes sédentaires</i>	PSar PSsp PSci PSop PSca PSma PSsa PSam PSse
X_5	<i>lien avec macrophytes</i>	Mga Mpl Cis2 Ccrev2 Eechin
X_6	<i>granulométrie et profondeur</i>	Z %coqs %sables %silts %argiles
X_7	<i>le reste</i>	Cbe Plath CNID ASCID

sur la colonne d'eau n'ont pas été utilisées. La variable y à expliquer est la variable *palourde* qui mesure l'abondance de cette espèce dans l'étang Thau en ses $l = 64$ sites. Le tableau 3.2 donne ces différents groupes X_k ($K = 7$) et le codage des variables dont on peut trouver la correspondance dans le tableau 3.3.

Avant de passer aux résultats, précisons que nous avons fait plusieurs études avant de retenir cette configuration des données. Nous avons fait des analyses avec transformation de la variable *palourde* par $\ln(1 + Mpal)$ uniquement d'une part, et appliqué cette transformation à toutes les variables d'abondance, d'autre part. Les résultats n'étaient pas meilleurs en terme de variance expliquée. Nous avons aussi rajouté un bloc X_8 contenant les coordonnées géographiques (x, y) des stations dans l'étang de Thau. Ceci a à peine amélioré la variance expliquée de la variable réponse. Enfin, nous avons rajouté un bloc X_9 contenant les coordonnées au carré et croisées (x^2, xy, y^2) des coordonnées géographiques des stations : cela n'a pas amélioré la variance expliquée de la réponse, celle-ci à même diminuée. Pour toutes ces raisons, nous avons gardé les résultats de l'analyse avec les 7 groupes décrits plus haut.

3.2 Application de l'ACIMOG-PLS1

Les données ont été centrées et réduites par la pondération uniforme. Nous avons appliqué l'ACIMOG-PLS 1 (Première Partie, chapitre 5.6). Dans la situation présente, $N = 1$, le modèle \hat{y} de l'ACIMOG-PLS3 est équivalent à celui de la régression PLS usuelle de y sur tous les prédicteurs. Nous donnerons à titre indicatif les pourcentages de reconstitution des données obtenues par l'ACIMOG-PLS3 (tableau 3.5) mais ne détaillerons pas les résultats. Pour une application de l'ACIMOG-PLS2, se reporter à l'application du chapitre 5 de cette partie.

Le tableau 3.4 donne les pourcentages de variance des tableaux expliquée par le modèle en fonction de l'étape a pour l'ACIMOG-PLS1. Dès la première étape, ces résultats montrent que la variable *palourde* est expliquée à 54.87%.

L'utilisation de la validation croisée (tableau 3.6) nous a amené à retenir une seule composante par tableau.

La figure 3.1 nous montre qu'à cette étape, d'après le critère $B^1[k]$, basé sur la valeur absolue

TAB. 3.3 – Correspondance codage-variable

Mollusques		Polychètes Errantes		Polychètes sédentaires	
Mpal	Palourde	PEap	aphriditidés	PSar	aricidés
Mcom	corbula	PEph	phyllodidés	PSsp	spionnidés
Mlu	lucinidés	PEsy	syllidés	PSci	cirratulidés
Mcl	clovisses	PEne	néreidés	PSop	ophélidés
Mdt	tellines	PEgl	glycérédés	PSca	capitellidés
Mca	cardidés	PEeu	eunicidés	PSma	maldanidés
Mmy	mytilidés			PSow	owenidés
Mga	gastéropodes			PSsa	sabellidés
Mpl	placophores			PSam	amphipectinidés
				PSst	terrébellidés
				PSse	serpulidés
Crustacés		Echinodermes		Autres espèces	
Cbe	bernard l'hermite	Eophi	ophirures	Plath	plathelminthes
Ccrab2	crabes	Eechin	echinidés	CNID	cnidaires
Cis2	isopodes			ASCID	ascidies
Ccrev2	amphipodes			SIPON	siponculidés
				PHORO	phoronidés
Granulométrie		Colonne d'eau			
%coqs.<1.6	coquilles	C	carbone or		
%sables	sables	Chl a	chlorophylle a		
%silts	silts	Pheo a	phéopigments		

TAB. 3.4 – Pourcentage d'inertie des tableaux expliquée par le modèle en fonction de a ($\%R^2$) par l'ACIMOG-PLS1

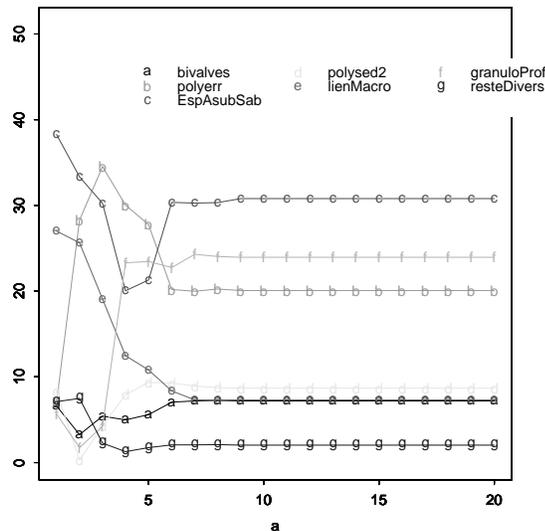
a	1	2	3	4	5	6	7	8
%y	54.87	74.17	83.30	90.40	91.57	92.69	92.89	92.90
%X ₁	27.59	61.54	78.58	90.33	96.31	100	100	100
%X ₂	52.76	69.89	82.83	91.53	97.09	100	100	100
%X ₃	35.62	75.31	81.95	92.27	98.86	100	100	100
%X ₄	28.34	50.05	65.32	76.72	85.14	90.68	93.92	97.60
%X ₅	33.68	52.19	71.07	89.24	100	100	100	100
%X ₆	63.64	88.44	96.63	100	100	100	100	100
%X ₇	35.79	61.59	91.01	100	100	100	100	100

TAB. 3.5 – Pourcentages d'inertie des tableaux expliquée en fonction de a par l'ACIMOG-PLS3

a	1	2	3	4	5	6	7	8
%y	50.42	66.39	76.14	81.36	84.56	86.95	88.59	89.71
%X ₁	4.56	9.54	15.55	26.44	38.51	42.07	44.63	47.73
%X ₂	34.87	39.56	41.89	46.49	47.69	49.93	54.45	58.51
%X ₃	9.37	16.77	18.54	19.92	37.53	51.91	55.61	61.96
%X ₄	12.22	16.87	19.69	23.61	26.62	19.04	34.79	36.00
%X ₅	4.07	5.91	13.27	20.61	21.52	24.75	27.06	33.90
%X ₆	4.01	17.60	36.43	59.15	67.25	70.20	70.95	73.08
%X ₇	5.68	12.62	13.89	16.03	19.73	21.66	26.66	32.06

TAB. 3.6 – PRESS totaux en fonction de a

a	1	2	3	4	5	6	7	8	9	10
PRESS	1.02	1.10	1.17	1.2	1.32	1.35	1.43	1.44	1.45	1.45

FIG. 3.1 – Evolution du coefficient $B^a[k]$ en fonction de a 

des coefficients du modèle, décrit dans le chapitre 6.3.2 de la première partie, ce sont les espèces liées aux substrat sableux puis les espèces liées aux macrophytes qui sont les plus influentes dans les modèles. Nous pouvons constater que les corrélations partielles $r(t_{k,1,y}|t_{l,1}l \neq k)$ mettent en avant les deux mêmes tableaux avec $r(t_{3,1,y}|t_{l,1}l \neq 3) = 0.33$ et $r(t_{5,1,y}|t_{l,1}l \neq 5) = 0.31$, les autres corrélations partielles étant comprises entre 0.12 et 0.2. Nous avons pu constater sur d'autres exemples que les tableaux significativement plus importants que les autres par le critère $B_n^a[k]$ étaient les mêmes (et dans le même ordre d'importance) que ceux donnés par les corrélations partielles.

En regardant de plus près les coefficients du modèle d'ordre 1, tableau 3.7, on voit que les variables *Ccrab2* (qui fait partie des espèces liées au substrat sableux) et *Mpl* (lien avec macrophytes) ont les plus gros coefficient en valeur absolue. Cela explique les résultats obtenus avec le coefficient $B^a[k]$.

Nous avons aussi mentionné dans la première partie (chapitre 6.3.2) que le modèle (d'ordre 1) pouvait être exprimé comme combinaison linéaire des composantes $t_{k,1}$. Ces coefficients θ_k sont respectivement égaux à 0.11, 0.12, 0.27, 0.13, 0.23, 0.10, 0.12. Les deux plus importants correspondent une fois encore aux tableaux X_3 et X_5 .

Intéressons-nous maintenant aux critères *BIP*, *BIPrv* et aux coefficients *RV*, tableau 3.8. A l'étape 1, les critères *BIP* et *BIPrv* mettent particulièrement en avant le tableau X_2 , *polychètes errantes*. Ce tableau est aussi celui qui a un coefficient *RV* élevé avec \hat{y} . X_3 et X_5 font partis des *BIP* et *BIPrv* les plus élevés mais ne sont pas les plus élevés. Ainsi, X_3 et X_5 ont les plus gros coefficients mais ne sont pas les plus influents dans la construction du modèle d'ordre 1 d'après le *BIP* et *BIPrv*. Il s'agirait plutôt de X_2 et X_4 . Nous ne les avons pas

TAB. 3.7 – Coefficients du modèle d'ordre 1

Mcom	Mlu	Mcl	Mdt	Mca	Mmy			
-0.032	-0.041	0.086	-0.002	0.004	0.05			
PEap	PEph	PEsy	PEne	PEgl	PEeu			
0.019	0.029	0.048	-0.009	0.053	0.083			
PSow	PSte	Ccrab2	Eophi	SIPON	PHORO			
0.017	0.037	0.262	-0.02	0.041	0.01			
PSar	PSsp	PSci	PSop	PSca	PSma	PSsa	PSam	PSse
-0.001	0.047	0.035	0.023	-0.004	0.068	-0.026	0.079	-0.016
Mga	Mpl	Cis2	Ccrev2	Eechin				
0.005	0.113	-0.064	-0.092	0.161				
Z	%coqs.>1.6	%sables	%silts	%argiles				
0.002	-0.015	0.081	-0.056	-0.03				
Cbe	Plath	CNID	ASCID					
-0.028	0.066	0.09	0.014					

TAB. 3.8 – Valeur des BIP, BIP_{pv}, RV_{xk}, RV_{res}, RV_β pour les étapes 1 et 2

a = 1	BIP	BIP _{pv}	RV _{xk}	RV _{res}	RV _β
X ₁	0.74	0.44	0.09	0.09	0.31
X ₂	1.30	2.71	0.50	0.50	0.61
X ₃	0.88	1.18	0.24	0.24	0.56
X ₄	1.06	1.43	0.25	0.25	0.48
X ₅	0.882	0.64	0.17	0.17	0.31
X ₆	1.21	0.24	0.05	0.05	0.06
X ₇	0.79	0.36	0.10	0.10	0.21
a = 2					
X ₁	0.83	0.46	0.06	0.01	0.22
X ₂	1.19	2.58	0.35	0.07	0.36
X ₃	0.96	1.17	0.15	0.04	0.40
X ₄	1.07	1.39	0.18	0.03	0.07
X ₅	0.88	0.65	0.11	0.02	0.13
X ₆	1.17	0.27	0.03	0.03	0.05
X ₇	0.83	0.37	0.65	0.02	0.08

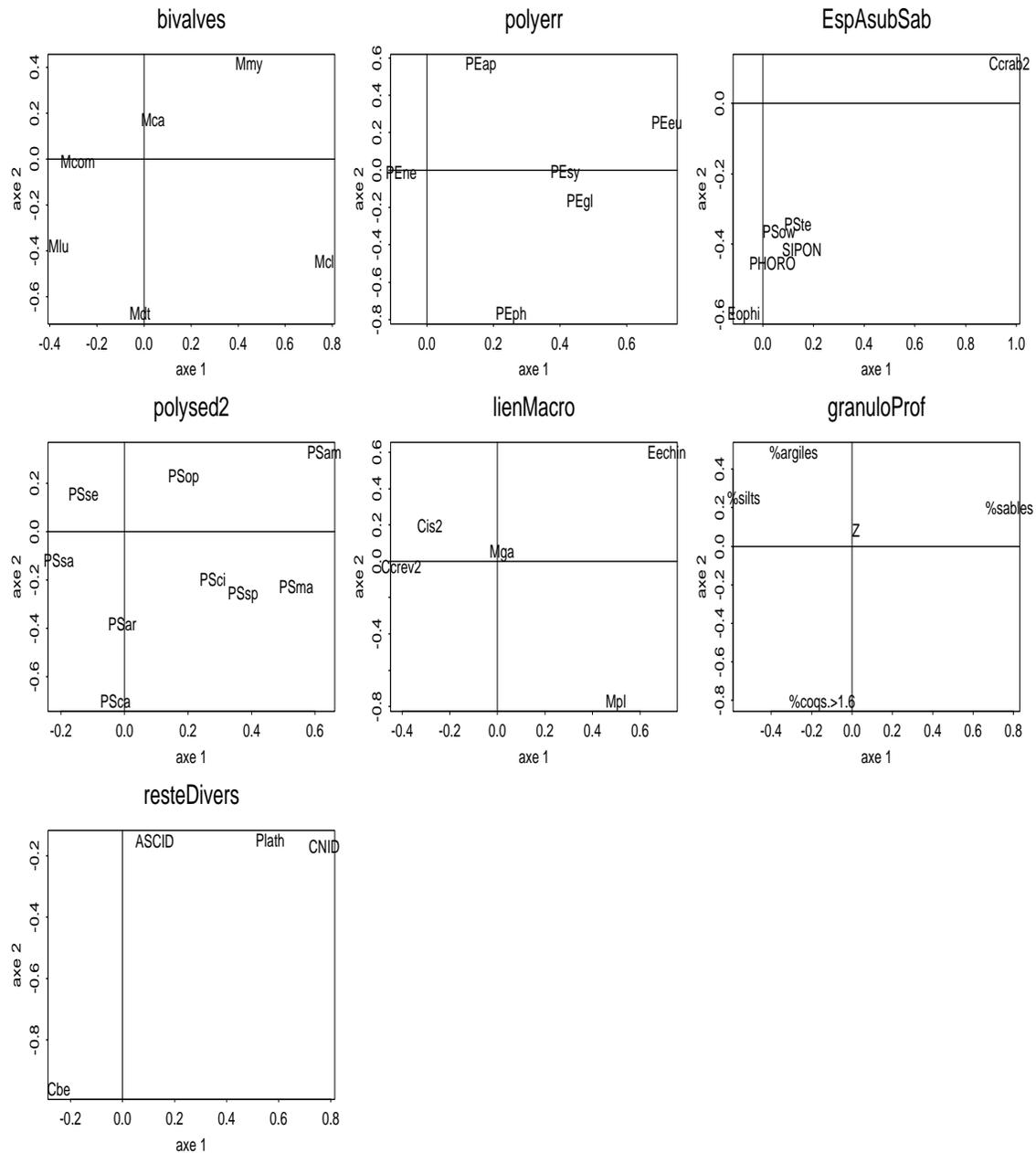
représentées ici, mais le graphique des corrélations des prédicteurs avec les super-composantes $t_a = \sum_{k=1}^K t_{k,a}$ montrent que les variables de X_2 font partie des plus corrélées avec t_1 .

Enfin, la figure 3.3 représente la cartographie des composantes $t_{k,1}$. Cette figure a été réalisée grâce au logiciel ADE-4* en diffusion libre sur internet. Les stations représentées par les croix sont des stations qui n'ont pas pu être échantillonnées. Ce graphique nous montre quelles sont les stations qui possèdent les plus faibles et les plus grosses valeurs dans la composante $t_{k,1}$. En associant ce graphique avec le graphique des variables, figure 3.2, nous pouvons expliquer cette cartographie.

Les fortes valeurs positives (les «gros» ronds noirs) dans $t_{k,1}$ sont expliquées par les variables du tableau k se situant à l'extrême droite du graphique des variables pour le tableau k . Inversement, les fortes valeurs négatives (les «gros» carrés) sont expliqués par les variables situées à l'extrême gauche du graphique des variables pour le tableau k . Le centre des graphiques de la figure 3.2, (0,0), ne représente pas la moyenne. L'important est juste de savoir qu'elles sont les variables qui se démarquent des autres et celles qui s'opposent. Nous pouvons constater que cette représentation des variables donnée par $(a_{k,1}, a_{k,2})$ respecte les positions des variables sur les axes données par les corrélations des variables X_k avec ses composantes $t_{k,1}$ et $t_{k,2}$ (Tab. 3.9). Ceci est d'autant plus respecté que le tableau X_k est bien reconstitué par l'étape. Ainsi, sur le premier axe, nous notons quelques différences de signe (pour *Mca* et *Mdt*), et donc de «position» entre la représentation graphique (Fig. 3.2) et les corrélations (Tab. 3.9), pour le tableau X_1 .

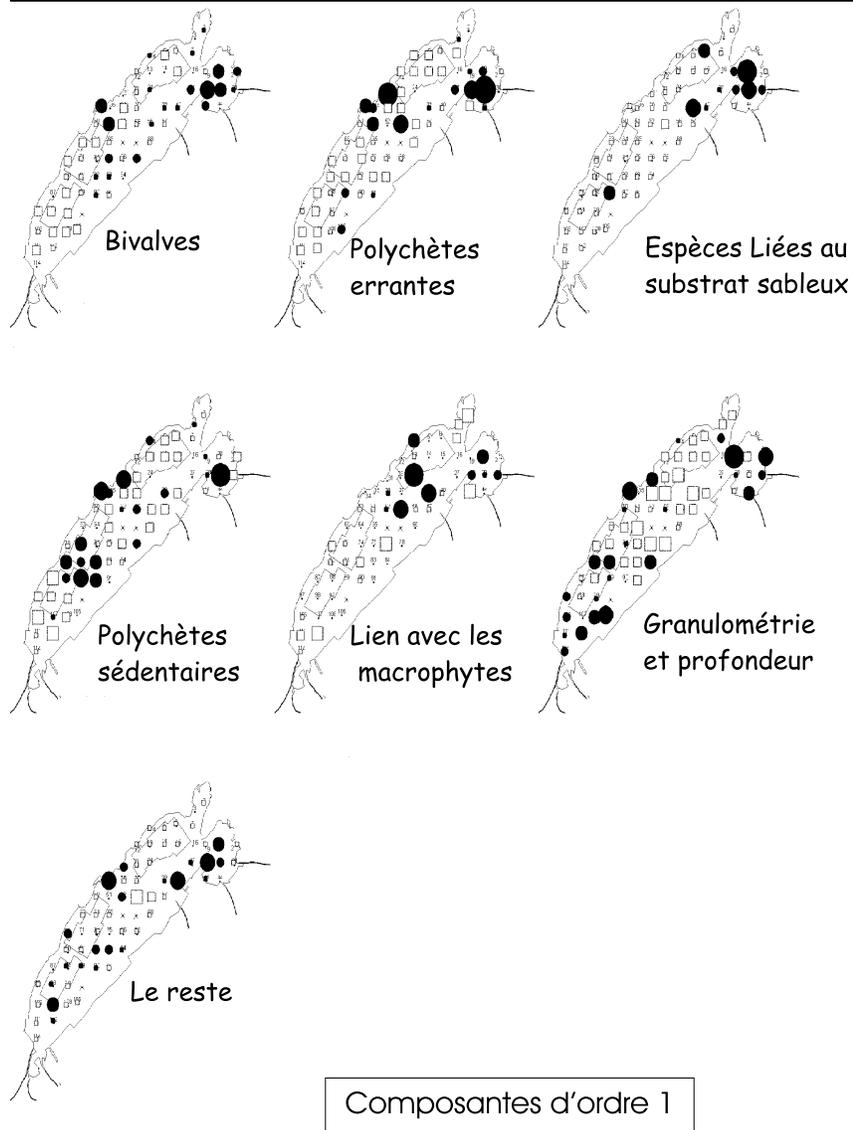
* <http://pbil.univ-lyon1.fr/ADE-4/ADE-4F.html>

FIG. 3.2 – Représentation des variables prédictrices du tableau X_k par les coordonnées des vecteurs $(a_{k,1}, a_{k,2})$.



TAB. 3.9 – Corrélations entre X_k et les composantes $t_{k,1}$ et $t_{k,2}$.

Bivalves		Polychètes Errantes				Esp. liées substrat sab.		
	$r(X_1, t_{1,1})$	$r(X_1, t_{1,2})$		$r(X_2, t_{2,1})$	$r(X_2, t_{2,2})$		$r(X_3, t_{3,1})$	$r(X_3, t_{3,2})$
Mcom	0.005	-0.35	PEap	0.09	0.64	PSow	0.16	-0.71
Mlu	-0.143	-0.6	PEph	0.628	-0.51	PSte	0.34	-0.62
Mcl	0.769	-0.55	PEsy	0.76	0.06	Ccrab2	0.96	0.18
Mdt	0.126	-0.73	PEne	-0.13	-0.01	Eophi	-0.07	-0.38
Mca	0.188	-0.28	PEgl	0.86	-0.10	SIPON	0.29	-0.31
Mmy	0.52	0.24	PEeu	0.91	0.05	PHORO	0.14	-0.65
Polychètes séden.			Lien macrophytes			Granulo. et Prof.		
	$r(X_4, t_{4,1})$	$r(X_4, t_{4,2})$		$r(X_5, t_{5,1})$	$r(X_5, t_{5,2})$		$r(X_6, t_{6,1})$	$r(X_6, t_{6,2})$
PSAR	-0.25	-0.44	Mga	-0.16	0.12	Z	-0.52	0.30
PSsp	0.61	-0.39	Mpl	0.52	-0.59	%coqs	0.34	-0.86
Psci	0.32	-0.29	Cis2	-0.35	0.28	%sables	0.98	0.19
PSop	0.36	0.31	Ccrev2	-0.46	0.03	%silts	-0.91	0.33
PSca	0.30	-0.75	Eechin	0.72	0.54	%argiles	-0.87	0.36
PSma	0.64	-0.10						
PSsa	-0.31	-0.26						
PSam	0.70	0.25						
PSse	-0.25	0.09						
Reste divers								
	$r(X_7, t_{7,1})$	$r(X_7, t_{7,2})$						
Cbe	-0.17	-0.83						
Plath	0.54	-0.17						
CNID	0.68	-0.08						
ASCID	0.33	-0.35						

FIG. 3.3 – Cartographie des composantes d'ordre 1

Chapitre 4

Chimiométrie : ACIMOG-PLS, HPLS, MBPLS

Les données présentées appartiennent au groupe ALTADIS. Nous présentons ici les premiers résultats obtenus à la suite de notre collaboration avec B. Vidal et J.-P. Biesse, laquelle a pour but de donner suite au travail de stage de DESS de R. Laroche (LAROCHÉ, 2001) au sein du centre de recherche d'ALTADIS à Orléans. Nous avons l'autorisation de présenter ici les résultats mais les données sont confidentielles. Aussi nous cacherons la nature précise de certaines variables, dont les variables réponses.

Le but est de modéliser des variables en fonction de variables chimiques, de variables de thermolyse et de spectres proche infrarouge (NIRS), mesurés sur une variété de tabac. Les variables réponses sont au nombre de 6, réparties en deux groupes Y_1 ($M_1 = 2$) et Y_2 ($M_2 = 4$). Les variables chimiques, bloc X_1 , sont au nombre de $J_1 = 50$, le bloc X_2 , contient $J_2 = 29$ variables issue de thermolyse, et le dernier bloc X_3 contient les spectres infrarouge à $J_3 = 1050$ longueurs d'ondes, de 400nm à 2498nm avec une résolution de 2nm. Tous les blocs sont mesurés sur $l = 24$ tabacs de même variété issus de 9 origines géographiques différentes.

Nous avons tout d'abord considéré les blocs de variables tels que nous venons de les décrire et appliqué différentes méthodes multi-blocs. Les premières analyses, effectuées par R. Laroche, consistaient en des régressions PLS d'une des réponses sur les différents blocs prédicteurs, chercher à sélectionner les variables importantes et à affiner les modèles grâce au VIP inclus dans le logiciel SIMCA-P[®].

En chimiométrie, il est d'usage de ne pas réduire en colonnes les données NIRS, mais d'y appliquer un prétraitement particulier. La réduction des variables NIRS entraîne une perte de l'information contenue dans les spectres. Les prétraitements de type Orthogonal Signal Correction (§I-6.1) sont de plus en plus utilisés. Nous avons ainsi appliqué différentes méthodes multi-tableaux, MBPLS, HPLS et ACIMOG-PLS 1, 2, 3 sur les données centrées et divisées par la racine carrée de leur inertie (afin d'essayer de réduire l'importance d'un bloc qui pourrait être fonction de son nombre de variables) d'une part, et d'autre part en ayant enlevé une partie du bruit présent dans le NIRS par la méthode OSC de S. Wold d'origine, que nous avons programmée. Nous savons que cette OSC n'est pas la meilleure (WOLD et al., 1998;

SVENSSON et al., 2002; VERRON, 2002) et possède quelques inconvénients. Nous avons essayé un autre prétraitement qui n'a pas amélioré les variances expliquées par les modèles. Nous n'avons pas encore essayé la méthode O-ACIMOG de VERRON (2002), qui devrait donner de meilleurs résultats en combinaison avec l'ACIMOG-PLS.

4.1 Résultats numériques

Le tableau 4.1 donne les pourcentages de variance expliquée par les modèles en fonction de a pour les méthodes ACIMOG-PLS1, 2, 3, HPLS et PLS. Les étapes soulignées sont celles retenues par la validation croisée. La validation croisée pour MBPLS n'a pas été faite, car notre programme, non complet, ne le permet pas encore.

Nous remarquons que l'HPLS donne des résultats très mauvais, et que la validation croisée retient plus de 8 composantes. Comme nous l'avons souligné dans le chapitre qui lui est consacré, I-5.3, l'HPLS explique surtout les tableaux des prédicteurs et beaucoup moins les réponses. De plus, elle semble passer à côté de la structure commune à modéliser aux deux premières étapes.

La MBPLS donne les meilleurs résultats à la première étape (mis à part ACIMOG-PLS1) mais devient légèrement moins explicative que les ACIMOG-PLS 2 et 3 dès $a = 2$. Cela vient sûrement de la déflation utilisée dans MBPLS. Mais, ces modèles sont difficilement interprétables car les tableaux sont chacun expliqués par une composante différente et seules les composantes des X_k sont orthogonales. Cela signifie que les modèles partiels pour Y_n n'expliquent pas obligatoirement une part différente de la variance de Y_n . Remarquons de plus, que les prédicteurs sont très vite reconstitués. Les modèles réponses paraissent très proches de ceux obtenus par la régression PLS des tableaux conjoints.

A la première étape, les modèles PLS réponses sont plus explicatifs que ceux des autres méthodes (mis à part MBPLS et ACIMOG-PLS1), ensuite l'ACIMOG-PLS3 est légèrement plus performante. Les résultats de la PLS montrent ici que cette méthode n'est pas forcément plus performante, en terme de variance expliquée, qu'une méthode multi-bloc. D'où l'intérêt de méthodes de régression multi-tableaux adaptées.

L'ACIMOG-PLS1 donne évidemment les meilleurs résultats, car à chaque étape les modèles partiels sont expliqués par un espace de plus grande dimension (1, 2 ou 3) que dans les autres méthodes (1). Les modèles de l'ACIMOG-PLS3 sont assez proches de ceux de l'ACIMOG-PLS2 excepté à la première étape où X_1 est bien mieux expliqué (+19%).

Dans le tableau 4.2, nous voyons qu'aux deux premières étapes, les super-composantes t_a et u_a ne sont pas corrélées. Cela explique les résultats du tableau précédent. Rappelons que l'HPLS ne tient pas compte de u_a dans le calcul des $t_{k,a}$ ce qui peut, comme ici, entraîner de pauvres performances prédictives de la part de t_a . Les corrélations entre u_a et t_a pour l'ACIMOG-PLS1 n'ont que peu de sens, car ce ne sont pas les super-composantes qui servent à modéliser les tableaux. On remarque que les premières composantes de PLS sont les plus corrélées mais qu'à partir de $a = 2$ ce n'est plus le cas. Cela semble dire que les déflations correctement choisies (construction des super-composantes) permettent d'aboutir à des super-composantes plus explicatives des réponses en comparaison avec une PLS ordinaire.

TAB. 4.1 – Pourcentage de variance expliquée par les modèles ACIMOG-PLS1, 2, 3, HPLS, MBPLS et PLS en fonction de a . Les modèles retenus par validation-croisée sont soulignés.

a	1	2	3	4	5	6	7	8
ACIMOG-PLS1								
%Y ₁	<u>90.45</u>	92.52	94.30	97.02	97.37	97.59	99.96	100
%Y ₂	<u>41.92</u>	62.11	70.22	83.36	96.01	97.19	99.60	100
%X ₁	74.91	91.13	95.46	98.22	99.33	99.83	99.95	100
%X ₂	75.88	88.33	97.17	98.79	99.57	99.76	99.91	100
%X ₃	83.93	94.11	98.90	99.70	99.84	99.93	99.98	100
ACIMOG-PLS2								
%Y ₁	62.66	<u>90.16</u>	90.30	93.04	93.50	94.42	94.98	96.04
%Y ₂	28.55	<u>36.42</u>	44.55	61.19	65.45	68.86	72.30	77.89
%X ₁	33.04	64.88	80.59	86.46	89.35	92.79	94.17	95.45
%X ₂	22.29	52.00	74.17	81.58	90.92	93.05	94.78	96.22
%X ₃	43.73	79.67	85.61	91.61	92.95	96.57	97.45	97.75
ACIMOG-PLS3								
%Y ₁	81.20	<u>91.08</u>	92.08	92.91	93.41	94.36	95.01	95.33
%Y ₂	28.94	<u>37.65</u>	57.19	60.28	64.28	68.30	71.13	78.61
%X ₁	27.31	62.48	68.55	83.95	87.17	91.59	93.78	95.32
%X ₂	10.25	54.88	65.63	82.71	92.40	94.55	96.03	96.91
%X ₃	55.12	79.13	86.13	93.49	94.01	96.10	97.02	97.36
HPLS								
%Y ₁	0.40	9.76	67.34	67.76	75.36	76.07	87.97	91.08
%Y ₂	14.05	16.74	23.94	27.59	36.25	43.30	57.98	58.72
%X ₁	13.45	66.42	88.81	89.60	89.98	94.40	96.78	97.92
%X ₂	70.47	74.62	77.13	89.73	96.88	97.03	97.23	98.32
%X ₃	5.40	20.72	65.40	67.72	71.24	73.50	84.70	85.12
MBPLS								
%Y ₁	86.55	89.83	91.03	92.74	93.83	94.02	95.06	96.13
%Y ₂	30.72	37.46	57.72	62.36	65.40	74.67	79.82	84.73
%X ₁	45.69	87.05	92.03	95.00	97.70	99.14	99.47	99.87
%X ₂	69.04	79.36	87.08	96.94	97.71	98.34	99.08	99.23
%X ₃	63.84	90.26	98.37	98.85	99.56	99.73	99.79	99.82
PLS								
%Y ₁	86.05	<u>90.65</u>	91.64	92.77	92.90	94.15	94.22	94.87
%Y ₂	29.95	<u>37.30</u>	54.61	61.48	66.65	68.86	75.93	82.95
%X ₁	25.10	61.31	72.49	84.72	88.23	91.82	93.41	95.23
%X ₂	6.83	56.77	69.00	82.25	90.52	94.45	95.46	96.68
%X ₃	58.34	78.14	86.16	93.06	93.89	96.31	97.09	97.48

TAB. 4.2 – Corrélations entre les super-composantes t_a et u_a

u_a	HPLS			ACIMOG-PLS1			ACIMOG-PLS2			ACIMOG-PLS3			PLS		
	t_1	t_2	t_3	t_1	t_2	t_3	t_1	t_2	t_3	t_1	t_2	t_3	t_1	t_2	t_3
	0.50	0.31	0.81	0.78	0.58	0.45	0.78	0.83	0.52	0.87	0.86	0.61	0.89	0.775	0.575

4.2 Résultats graphiques de l'ACIMOG-PLS

Voyons de plus près les résultats des ACIMOG-PLS.

L'ACIMOG-PLS1 présente l'avantage de permettre une interprétation plus approfondie des modèles par la visualisation de l'information présente dans les tableaux prédicteurs à chaque étape. Elle permet de nombreuses représentations graphiques grâce à l'orthogonalité de ses composantes partielles et de ses super-composantes. Les corrélations entre les variables de chaque bloc k et les super-composantes d'une part, et les corrélations entre les variables d'un bloc et ses composantes partielles (t_k) permettent l'interprétation des t_a et $t_{k,a}$ (fig. 4.1).

Les ACIMOG-PLS2 et 3 ne permettent que l'interprétation des super-composantes, c'est-à-dire de l'information «moyenne» des X_k servant à prédire les réponses.

Ces graphiques sont dans les figures 4.2 et 4.3.

Nous nous devons de rappeler que les données ne sont pas réduites en colonnes. Les coordonnées des projections des variables dans le plan défini par deux composantes orthogonales sont données par les covariances entre les variables les composantes. Cependant, comme certaines variables chimiques et de thermolyse ont une très forte variance par rapport aux variables dont, en particulier, le NIRS, nous avons représenté les corrélations pour plus de lisibilité. Les variables les plus importantes en terme de covariances font aussi partie de celles ayant les plus fortes corrélations. Ces graphiques doivent s'interpréter en terme de directions seulement, mise à part pour les variables que nous savons être importantes (mais que nous ne donnerons pas ici). Nous n'allons pas faire une interprétation détaillée car les données sont confidentielles. Mais nous présentons quelques grandes lignes aidant à l'interprétation.

D'une manière générale, les super-composantes des trois méthodes montrent que l'on peut faire quatre groupes de tabacs. Deux groupes s'opposant sur l'axe 2 (Bey1, Bey16, Bey19 et Bey6, Bey10, Bey8), et un groupe ayant seulement une variation sur l'axe 2 (Bey2, Bey11, Bey5, Bey24, Bey13, Bey12). Le quatrième groupe de tabac se situe au centre, à la moyenne. De plus, le tabac 1 a des caractéristiques particulières, car il se démarque des autres.

Les graphiques des ACIMOG-PLS 2 et 3 ne sont pas significativement différents. Seul l'axe 2 est inversé. La première super-composante est expliquée par une partie des longueurs d'ondes des spectres et quelques variables chimiques. Le second par quelques variables chimiques et une variable de thermolyse. Globalement la thermolyse ne semble pas intervenir. Notons que c'est le tableau (X_2) le moins bien reconstitués aux deux premières étapes par ces ACIMOG-PLS. Ici aussi, les variables NIRS ayant de très petites variances et petites valeurs, sur le graphique des covariances, elles n'étaient pas lisibles. Mais elles étaient plus importantes que les autres variables qui ne «ressortaient» pas.

L'ACIMOG-PLS1 permet d'identifier quels sont les tabacs qui sont liés aux longueurs d'ondes et aux variables chimiques expliquant l'étape 1. Ici, la deuxième étape est expliquée par quelques variables de thermolyse. Les deux premiers graphes de la figure 4.1 donnent une sorte de vision moyenne des graphes situés en dessous. Notons une stabilité des spectres à la première étape entre longueurs d'ondes 600nm et 1100nm environ.

FIG. 4.1 – ACIMOG-PLS1. Représentation des corrélations entre les variables et les super-composantes, et les différentes composantes t_k aux étapes 1 et 2 et représentation des individus grâce aux super-composantes et composantes t_k .

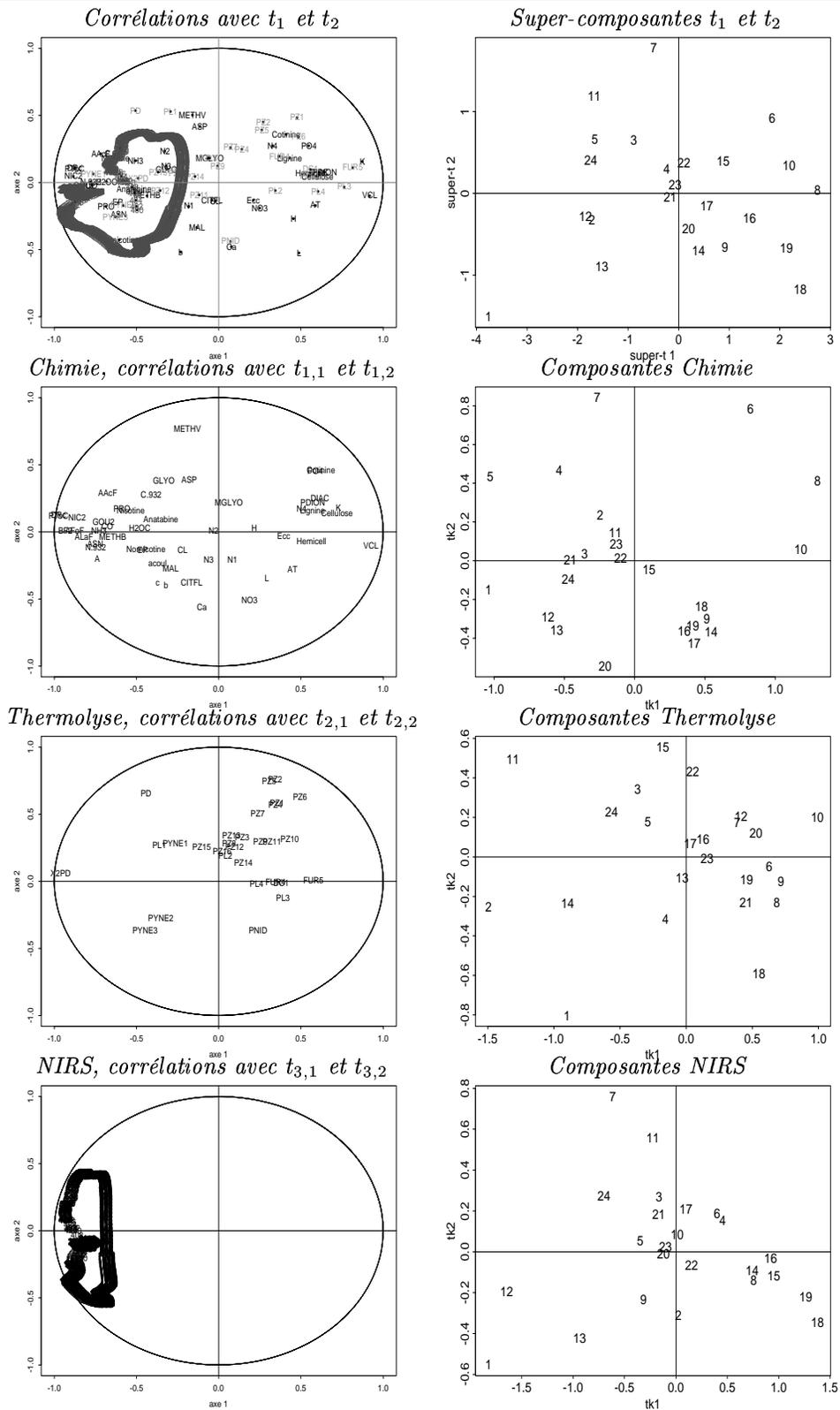


FIG. 4.2 — ACIMOG-PLS2. Représentation des corrélations entre les variables et les super-composante et représentation des individus grâce aux super-composantes.

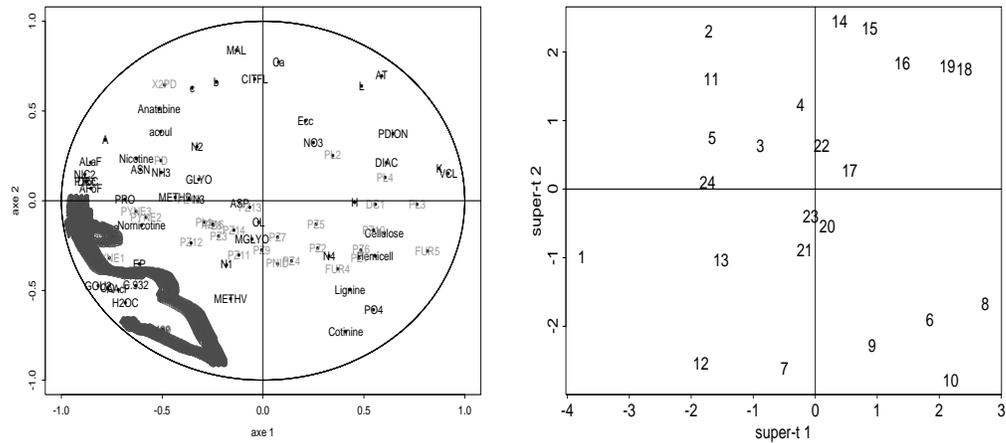
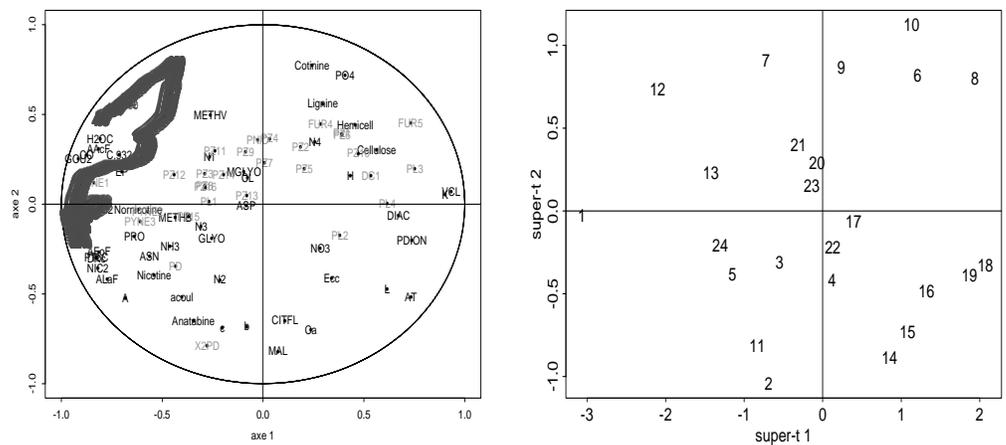


FIG. 4.3 — ACIMOG-PLS3. Représentation des corrélations entre les variables et les super-composantes et représentation des individus grâce aux super-composantes.



4.3 Importance des tableaux

Comme nous en avons discuté dans le chapitre I-6.3.2, il est intéressant de savoir si un groupe de variables est plus important que les autres dans les modèles. Pour les chimométriciens, il serait intéressant que le NIRS, qui sont des données peu coûteuses, faciles et rapides à obtenir, suffise à la construction de modèles prédictifs.

Le tableau 4.3 donne les résultats numériques obtenus par les différents critères BIP, BIP-RV, coefficients RV et $B_n^a[k]$. Dans quasiment tous les cas le NIRS (X_3) est le plus important, avec une importance plus marquée avec les ACIMOG-PLS1 et 3 aux deux premières étapes.

Nous remarquons dans le tableau les «propriétés gênantes» des différents critères pour $a = 1$: le BIP est indépendant de n , et avec les ACIMOG-PLS2 et 3 les critères sont indépendants de n voire de n et de k . Ces résultats nous ont tout de même amenés à penser, que peut-être, le tableau NIRS (X_3) était important à cause de sa taille, bien que nous ayons ramené son inertie à 1. Aussi, nous avons refait les analyses sans l'utiliser.

4.4 Sans le NIRS

Le tableau 4.4 contient les pourcentages de variance expliquée par les modèles. La première surprenante constatation est que les résultats de l'HPLS sont égaux avec ou sans le NIRS (à comparer avec le tableau 4.1). Ceci montre à quel point les résultats de cette méthode sont douteux.

Ensuite, nous constatons que les Y_n sont nettement moins bien expliqués sans le NIRS et que les modèles sont moins stables.

Nous pouvons donc conclure que le NIRS contient bien de l'information essentielle à l'explication des réponses étudiées.

4.5 Pour conclure

Ces résultats sont les premiers d'une plus ample étude menée en collaboration avec le groupe ALTADIS. Ils sont assez prometteurs quant aux capacités de modélisation des ACIMOG-PLS. L'ACIMOG-PLS1 permet d'obtenir rapidement des modèles très explicatifs et stables. Les spectres NIRS apparaissent comme contenant beaucoup d'information utile.

Il serait intéressant de conduire des analyses plus poussées en effectuant des groupes de variables plus petits, que ce soit pour les variables chimiques, thermolyses ou NIRS. Une analyse n'utilisant que le NIRS découpé selon des intervalles de longueurs d'ondes égaux, ou d'une autre manière plus liées aux connaissances des spectres, devrait être menée. Elle pourrait sans doute permettre l'identification des zones contenant l'information utile. Ceci sera fait dans l'avenir.

De plus, notre collaboration doit nous mener vers l'analyse et la comparaison des résultats en fonction de diverses variétés de tabacs.

TAB. 4.3 – Etude de l'influence des tableaux X_k dans les modèles ACIMOG-PLS à l'aide des critères définis §I-6.3.2

a		1			2			3		
ACIMOG-PLS1	X₁	X₂	X₃	X₁	X₂	X₃	X₁	X₂	X₃	
B_n^a Y₁	1.86	8.53	89.61	3.40	24.83	71.22	9.35	46.01	44.64	
Y₂	9.18	17.44	73.38	23.52	7.24	69.23	41.99	7.86	50.44	
BIP Y₁	0.89	0.94	1.15	0.90	0.94	1.15	0.90	0.95	1.14	
Y₂	0.89	0.94	1.15	0.97	0.95	1.07	0.96	1.01	1.02	
BIP_{rv} Y₁	0.67	0.07	2.26	0.67	0.08	2.25	0.67	0.08	2.25	
Y₂	0.51	0.75	1.73	0.57	0.75	1.67	0.59	0.75	1.66	
RV_{xk} Y₁	0.29	0.05	0.90	0.28	0.04	0.88	0.28	0.05	0.87	
Y₂	0.28	0.41	0.74	0.23	0.30	0.52	0.20	0.26	0.45	
RV_{res} Y₁	0.29	0.05	0.90	0.00	0.01	0.01	0.00	0.02	0.00	
Y₂	0.29	0.41	0.74	0.17	0.13	0.74	0.10	0.01	0.00	
RVβ Y₁	0.49	0.05	0.99	0.46	0.00	0.97	0.64	0.06	0.97	
Y₂	0.39	0.51	0.78	0.20	0.33	0.73	0.31	0.24	0.60	
ACIMOG-PLS2	X₁	X₂	X₃	X₁	X₂	X₃	X₁	X₂	X₃	
B_n^a Y₁	33.33	33.33	33.33	29.74	11.57	58.69	29.86	11.77	58.36	
Y₂	33.33	33.33	33.33	31.40	21.62	46.97	14.43	31.11	54.45	
RV_{xk} Y₁	0.48	0.31	0.62	0.29	0.05	0.89	0.30	0.05	0.88	
Y₂	0.48	0.31	0.62	0.56	0.37	0.77	0.29	0.50	0.62	
RV_{res} Y₁	0.48	0.31	0.62	0.20	0.16	0.26	0.00	0.00	0.00	
Y₂	0.48	0.31	0.62	0.17	0.14	0.22	0.15	0.14	0.22	
RVβ Y₁	0.78	0.46	0.66	0.86	0.27	0.95	0.87	0.36	0.95	
Y₂	0.78	0.46	0.66	0.76	0.46	0.73	0.33	0.53	0.63	
ACIMOG-PLS3	X₁	X₂	X₃	X₁	X₂	X₃	X₁	X₂	X₃	
B_n^a Y₁	21.65	7.70	70.65	17.74	10.45	71.81	12.39	11.26	76.35	
Y₂	21.65	7.70	70.65	21.42	7.21	71.37	12.02	28.28	59.70	
BIP Y₁	0.94	0.72	1.26	0.94	0.78	1.23	0.94	0.78	1.22	
Y₂	0.94	0.72	1.26	0.94	0.83	1.19	0.95	0.90	1.13	
RV_{xk} Y₁	0.40	0.14	0.79	0.30	0.05	0.89	0.27	0.05	0.90	
Y₂	0.40	0.14	0.79	0.57	0.37	0.76	0.24	0.34	0.55	
RV_{res} Y₁	0.40	0.14	0.79	0.07	0.07	0.08	0.00	0.00	0.01	
Y₂	0.40	0.14	0.79	0.18	0.19	0.20	0.10	0.16	0.21	
RVβ Y₁	0.71	0.23	0.87	0.78	0.30	0.97	0.78	0.13	0.98	
Y₂	0.71	0.23	0.87	0.67	0.38	0.77	0.14	0.40	0.73	

FIG. 4.4 – Pourcentage de variance expliquée par les modèles sans le NIRS. Les modèles soulignés sont ceux retenus par le PRESS global de la validation croisée

a	1	2	3	4	5	6	7	8
ACIMOG-PLS1								
%Y ₁	42.47	80.40	<u>86.65</u>	90.51	93.30	94.58	97.13	97.65
%Y ₂	22.41	40.36	<u>61.72</u>	67.02	81.79	89.95	92.30	95.54
%X ₁	56.33	88.74	93.16	97.04	98.66	99.53	99.65	99.91
%X ₂	75.59	84.48	96.04	97.84	98.92	99.30	99.59	99.73
ACIMOG-PLS2								
%Y ₁	21.42	74.59	80.51	84.38	87.55	90.74	92.52	92.91
%Y ₂	19.92	26.84	37.02	53.24	62.10	68.06	70.20	79.83
%X ₁	42.88	60.80	83.73	89.69	92.58	93.13	95.61	96.94
%X ₂	42.90	65.50	76.09	82.93	88.68	95.84	97.34	97.92
ACIMOG-PLS3								
%Y ₁	26.83	73.44	79.79	85.15	<u>88.22</u>	89.92	92.45	92.52
%Y ₂	19.15	25.80	36.31	54.54	<u>63.61</u>	68.63	71.62	82.75
%X ₁	47.46	66.06	85.65	90.04	92.66	93.07	95.82	97.22
%X ₂	34.55	61.03	75.62	82.32	86.73	96.81	97.25	97.47
HPLS								
%Y ₁	<u>0.40</u>	9.75	66.72	67.14	74.78	75.48	86.29	89.66
%Y ₂	<u>14.05</u>	16.73	23.75	27.37	36.00	43.01	56.56	57.30
%X ₁	13.44	66.42	88.81	89.60	89.98	94.40	96.79	97.93
%X ₂	70.47	74.62	77.14	89.74	96.89	97.03	97.24	98.31

Chapitre 5

Analyse Sensorielle : ACIMOG-PLS2 et 3, REMUB, ACIMOG, DO-ACT

L'exemple présenté ici rentre dans le cadre de l'analyse sensorielle.

Ces données appartiennent au Laboratoire Evaluation Sensorielle de NESTLE France et nous ont été gracieusement fournies par Pascale Lacroix et Anne Jauneau-Marvier. Elles sont en partie confidentielles, aussi, les labels des variables et des produits n'apparaîtront pas tous explicitement.

Ces données ont été recueillies et préalablement étudiées par Francis Sune dans sa thèse (SUNE, 2001) à l'aide d'analyses statistiques simples (SUNE, LACROIX & HUON DE KERMADEC, 2002) usuellement utilisées en Analyse Sensorielle (URDAPILLETA, TON NU, SAINT DENIS & HUON DE KERMADEC, 2001; SUNE et al., 2002).

Une partie des résultats présentés ici ont fait l'objet d'une communication aux 7^{ièmes} Journées Agro-Industrie et Méthodes Statistiques (VIVIEN & SABATIER, 2002).

5.1 Présentation des données

Le jeu de données est constitué de deux panels de consommateurs enfants ($M = 51$ et $M = 55$), l'un «naïf», l'autre «préparé», évaluant $l = 8$ produits (chocolats), p_1, \dots, p_8 , selon $N = 23$ descripteurs. Les 14 premiers sont notés sur une échelle variant de 0 à 10 et les 9 suivants sont des variables qualitatives avec 2 à 5 modalités. Nous avons aussi étudié les données en supprimant les tableaux réponses des descripteurs qualitatifs, et les résultats étaient très proches numériquement de ceux obtenus ici.

Le détail des méthodes utilisées pour le choix des attributs est donné dans (SUNE et al., 2002).

On peut retenir que ces attributs ont été choisis par les enfants. De plus, nous remarquerons que le nombre d'enfants participants à l'opération était de 139 initiés et 122 non initiés. Mais les données n'étant pas complètes pour tous les enfants, et notre programme ne fonctionnant pas encore avec des valeurs manquantes, nous n'avons pas utilisé les enfants qui dont les données recueillies présentaient au moins une réponse manquante. L'échantillon a été réduit de plus de moitié, mais nous verrons que les résultats obtenus concordent tout de même avec ceux présentés dans (SUNE, 2001; SUNE et al., 2002).

Nous possédons de plus un ensemble de notes, comprises entre 1 et 7, évaluant $K = 27$ propriétés, de ces mêmes produits, données par $J = 10$ experts.

TAB. 5.1 – Correspondance entre les labels utilisés et les attributs sensoriels (tableau)

Experts		Consommateurs	
Code	Attribut	Code	Attribut
a	brillant	Q1	hédonique
b	intensité couleur	Q21	surface blanche/foncée (intensité couleur)
c	homogénéité couverture	Q22	surface lisse/bosselée (homogénéité couverture)
d	quantité d'éclats	Q23	surface céréales peu/beaucoup (quantité d'éclats)
e	grosseur éclats	Q24	surface céréales petites/grosses (grosseur des éclats)
f	homogénéité couleur	Q3	odeur faible/forte (intensité odeur)
g	quantité fourrage	Q41	pas dur /très dur (dureté)
h	hauteur du chocolat	Q42	pas croustillant/ très croustillant (croustillant)
i	intensité couleur fourrage	Q23	pas fondant/ très fondant (fondant)
j	croquant	Q44	pas moelleux/ très moelleux (moelleux)
k	croustillant	Q51	goût faible/fort (intensité goût)
l	fondant	Q52	goût lait faible/fort (goût lait)
m	lisse	Q53	goût café faible/fort (goût café)
n	collant	Q54	pas sucré/ très sucré (sucre)
o	résistance fourrage	Q6	type de chocolat (5 modalités)
p	résistances grains	Q7	riche en lait ou non
q	gras	Q81	bon pour grandir ou non
r	lait concentré sucré	Q82	donne des forces ou non
s	noisette	Q91	plaisir journée
t	praliné	Q92	plaisir au goûter
u	caramel	Q93	emmener à l'école
v	vanille	Q10	convient à qui (3 modalités)
w	intensité goût chocolat	Q11	préférence (4 modalités)
x	malte		
y	céréales		
z	sucre		
A	gaufrette		

Les correspondances entre les attributs sensoriels et les labels utilisés sont données dans le tableau 5.1. Les attributs *Q6* à *Q11* sont liés à l'avis personnel de l'enfant quant au plaisir, à l'utilité et autres critères.

Avec ce type de données, il se pose deux types de questions :

- peut-on expliquer les évaluations des enfants par les évaluations des experts? (Situation 1, Fig.5.1),
- est-ce que les consommateurs naïfs et les consommateurs préparés évaluent les produits de la même manière? (Situation 2, Fig.5.2).

Pour répondre à celles-ci, il est possible d'agencer les données de plusieurs manières. Pour les consommateurs, nous pouvons considérer *M* tableaux réponses de taille $l \times N$ ou 23 tableaux de taille $l \times M$. Quant aux données sensorielles, nous pouvons considérer un tableau par juge ou bien un tableau par propriété.

La disposition des données dépend de(s) la question(s) que l'on se pose.

Concernant la première question (Situation 1), il y a quatre possibilités pour chacun des deux panels :

1. *Quelle est la réaction de chaque consommateur (quant à sa description générale des pro-*

FIG. 5.1 – Situation 1 : explication des évaluations consommateurs par les évaluations des experts

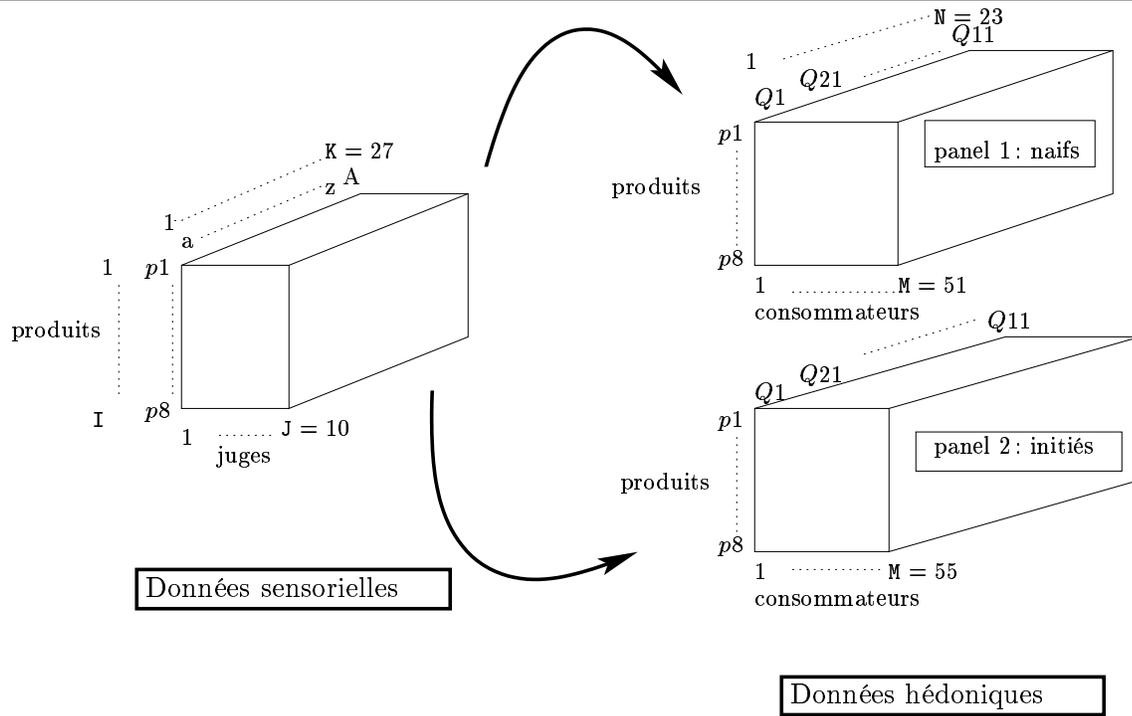
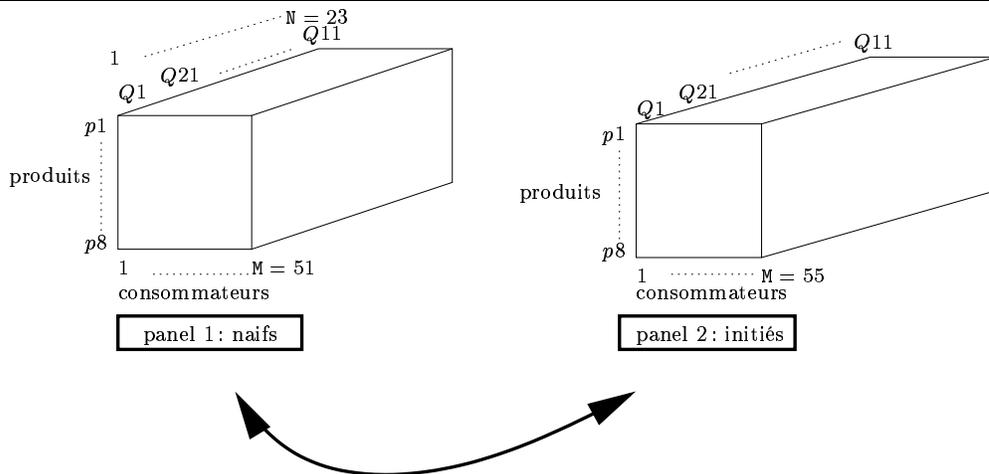


FIG. 5.2 – Situation 2 : comparaison des panels de consommateurs naïfs et initiés



duits) en fonction de chacune des propriétés des produits évaluées par les juges? Ici, les données hédoniques sont constituées d'un ensemble de $M = 55$ (ou $M = 51$) tableaux à $N = 23$ colonnes et les données sensorielles, d'un ensemble de $K = 27$ tableaux de propriétés à $J = 10$ colonnes.

2. Quelle est la réaction de chaque consommateur (quant à sa description générale des produits) en fonction de l'évaluation «générale» des différents juges? Dans ce cas on considèrera plutôt les données sensorielles comme un ensemble de 10 tableaux (les J juges) à $K = 27$ colonnes chacun.

3. Comment l'ensemble des consommateurs réagit face à chaque descripteur en fonction de l'évaluation «générale» des différents juges? On gardera la disposition précédente pour les données sensorielles et on considèrera les données hédoniques comme un ensemble de $N = 23$ (descripteurs) tableaux ayant chacun M colonnes.

4. Comment l'ensemble des consommateurs réagit face à chaque descripteur en fonction de chacune des propriétés des produits évaluées par les juges? Ici, on considère les données hédoniques selon la disposition précédente et les données sensorielles comme dans la première question 1.

C'est à cette dernière question que nous voulons répondre. Celle-ci nous semblant, ici, la plus pertinente.

En analyse multi-blocs, la disposition des données est importante et en général les tableaux de prédicteurs et/ou de réponses n'ont pas tous les mêmes variables. L'effet «juge» (dans notre cas) n'est pas pris en compte. Comme nous sommes dans le cas de données cubiques, nous pourrions appliquer les méthodes de régression «multiway» (Chapitre 4), et ainsi nous aurions un modèle tenant à la fois compte de l'effet «juge» et de l'effet «propriété». Cependant, nous allons appliquer ici l'ACIMOG-PLS 2 et 3 et REMUB. L'ACIMOG-PLS1 n'est pas applicable ici car $l (= 8)$ est inférieur à $K (= 27)$.

Concernant la situation 2, nous avons choisi d'arranger les données de façon à ce que l'on ait un tableau par descripteur. Ainsi, grâce aux méthodes ACIMOG (Chapitre 3.4) et DO-ACT (Chapitre 3.3) nous allons essayer de voir si les deux panels d'enfants ont des réponses similaires et pour quels descripteurs en particulier.

Il est aussi possible de prendre un tableau par consommateurs mais le sens nous paraît moins évident.

5.2 Situation 1: explication des réponses des consommateurs par les évaluations des experts

Dans cette partie, les tableaux X_k désignent les tableaux des évaluations sensorielles des juges, et les tableaux Y_n ceux des enfants, (selon l'analyse, ce seront les enfants initiés ou non initiés).

En pratique en analyse sensorielle, il est d'usage d'utiliser des tableaux résumés pour faire l'analyse (URDAPILLET et al., 2001). Ainsi, (SUNE et al., 2002) ont moyenné par attribut les évaluations des juges et des enfants, afin d'obtenir deux tableaux X de taille $l \times K$ et

Y de taille $l \times N^*$. Ensuite, ils ont effectué une régression PLS entre X et Y . Afin d'affiner les modèles, des régressions PLS par type d'attribut (aspect, texture, flaveur) ont été réalisées.

Nous proposons d'appliquer l'ACIMOG-PLS (I.5.6) pour expliquer les évaluations des consommateurs par les évaluations des experts. Nous allons faire deux analyses, une pour les enfants initiés et l'autre pour les non initiés. Ainsi, nous verrons que la réalisation d'une seule analyse, l'ACIMOG-PLS, permet d'aboutir aux mêmes résultats qu'avec l'utilisation de plusieurs méthodes comme l'ont fait les auteurs : ce qui présente un gain de temps.

5.2.1 Prétraitement

Rappelons les notations. Les données sensorielles forment un ensemble de $K = 27$ tableaux «produits \times juges» de taille $l \times J = 8 \times 10$. Le panel 1 de $M = 51$ consommateurs naïfs et le panel 2, constitué de $M = 55$ consommateurs préparés, forment chacun un ensemble de $N = 23$ tableaux descripteurs «produits \times consommateurs» de taille $l \times M$. Les 8 produits sont notés p_1, \dots, p_8 , les 27 propriétés sensorielles sont désignées par les 26 lettres de l'alphabet en minuscule et A : les propriétés de a à i décrivent l'aspect, j à q la texture, et r à A la flaveur. Les 10 juges du tableau représentant la propriété a sont identifiés par les label a_1, \dots, a_{10} . De même pour les autres propriétés. Enfin, les descripteurs hédoniques sont désignés par la lettre Q suivie de deux chiffres, le premier désignant le groupe auquel il est rattaché. Q_1 est la variable hédonique, Q_{21} à Q_{3} sont des descripteurs de l'aspect, Q_{41} à Q_{44} décrivent la texture, Q_{51} à Q_{54} la flaveur et Q_6 à Q_{11} sont des variables concernant «l'avis» des enfants à propos des chocolats.

Tous les tableaux sont centrés pour la métrique $D = Id_8/8$. Ce centrage implique qu'un maximum de 7 super-composantes non nulles peut être calculées. Ils ne sont pas réduits car certaines variables seraient alors de variance nulle.

Le centrage impliquant que chaque tableaux X_k et Y_n est au plus de rang 7, qui est inférieur au nombre de tableaux (K et N), la méthode ACIMOG-PLS1 ne peut être appliquée car dès la première étape la variance totale des tableaux serait expliquée, ce qui ne présente pas grand intérêt.

5.2.2 Application de l'ACIMOG-PLS 2 et 3

Le tableau 5.2 donne le pourcentage de variance de chacun des tableaux descripteurs Y_n des deux panels expliquée par le modèle d'ordre a de l'ACIMOG-PLS2.

Nous remarquons que dès la deuxième étape certains descripteurs sont assez bien reconstitués Q_{21} , Q_{23} , Q_{52} , Q_7 (intensité couleur, quantité éclats, goût lait, riche en lait) pour le panel 2 et Q_{21} , Q_{23} , Q_{24} , Q_{42} , Q_{52} , Q_6 pour le panel 1 (intensité couleur, quantité éclats, grosseur éclats, croustillant, goût lait, type de chocolat). A partir de la troisième étape, ils sont tous assez bien expliqués. Toutefois, on note une légère différence selon les deux panels. Le descripteur Q_{21} est reconstitué à plus de la moitié (54%) par le premier axe pour le panel 1, et pour le panel 2, ce descripteur est très peu reconstitué (7%) par le premier axe et de 55% par le deuxième axe.

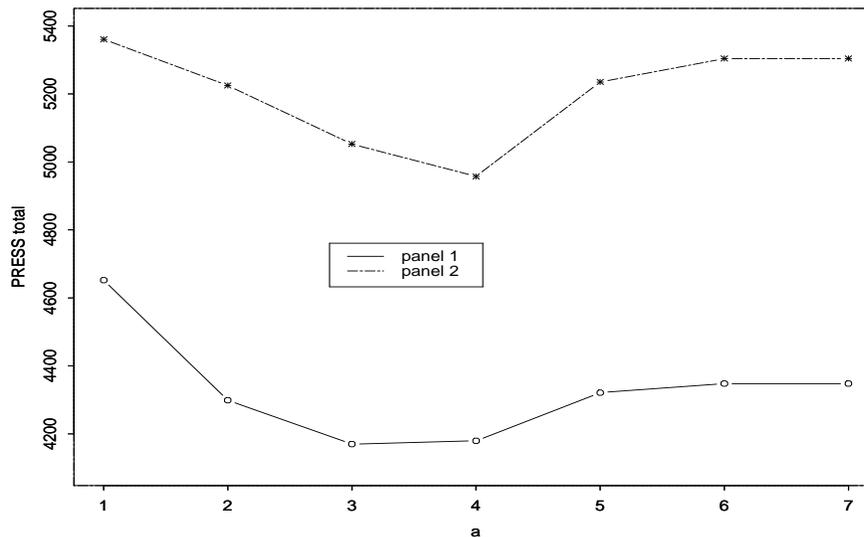
La validation croisée nous incite à retenir 3 étapes pour le panel 1, et 4 étapes pour le panel 2

* (SUNE et al., 2002) n'utilisent que les attributs «réellement» sensoriels pour les enfants, c'est-à-dire les 13 attributs Q_2 à Q_{54} . Quant aux juges, les $K = 27$ sont utilisés.

TAB. 5.2 – Inertie de Y_n expliquée par les modèles en fonction de a par l'ACIMOG-PLS2

a	Panel 1 : naïfs							Panel 2 : initiés						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Q1	23.08	52.08	57.16	69.81	82.27	91.91	100	15.98	39.63	57.91	70.99	81.26	90.20	100
Q21	54.37	76.45	81.98	91.26	96.87	98.00	100	6.97	61.98	79.20	90.10	93.68	96.09	100
Q22	11.95	30.68	56.38	66.69	78.72	93.29	100	31.23	42.40	60.45	67.92	77.05	90.12	100
Q23	28.73	73.26	82.45	86.37	91.08	97.49	100	49.52	59.57	79.10	86.80	91.50	97.86	100
Q24	20.53	60.23	75.73	82.47	91.65	96.38	100	35.54	47.43	70.83	82.10	91.66	97.28	100
Q3	16.36	30.14	47.65	61.19	72.59	82.61	100	18.77	29.31	42.76	57.00	70.00	86.47	100
Q41	14.36	42.48	57.21	68.57	80.88	87.53	100	23.69	41.79	56.38	65.80	75.03	89.56	100
Q42	42.03	54.26	66.12	79.50	88.87	94.29	100	34.57	44.66	51.13	85.30	91.54	97.70	100
Q43	13.21	41.16	52.04	63.61	70.10	86.67	100	21.90	41.94	51.66	65.95	80.44	90.01	100
Q44	14.70	35.98	48.42	61.37	74.72	84.81	100	20.27	36.03	49.23	63.55	72.25	85.90	100
Q51	15.44	27.34	38.55	52.16	64.70	84.86	100	16.32	36.34	48.62	57.40	72.41	84.39	100
Q52	26.24	38.79	58.20	72.18	82.66	92.78	100	37.89	56.01	64.99	74.78	84.14	93.49	100
Q53	18.74	27.82	42.11	57.36	76.86	88.03	100	8.40	34.44	60.88	70.46	84.13	94.63	100
Q54	22.40	31.30	44.36	60.21	73.38	87.01	100	18.78	33.26	48.43	62.38	74.82	88.31	100
Q6	17.55	51.17	75.07	80.17	85.88	92.25	100	14.59	47.26	70.41	80.54	88.00	91.88	100
Q7	19.65	31.21	44.29	68.49	83.57	94.74	100	40.11	55.89	68.35	78.74	87.20	95.74	100
Q81	18.47	35.00	47.39	60.92	75.80	86.97	100	15.10	29.05	43.51	58.78	70.57	87.66	100
Q82	19.58	38.78	48.14	64.18	75.81	86.09	100	16.96	34.30	46.21	63.40	74.70	85.60	100
Q91	12.92	31.59	39.57	52.74	70.26	86.99	100	11.29	35.18	49.35	59.86	72.22	86.27	100
Q92	17.18	32.14	51.41	62.82	72.34	87.82	100	23.01	35.20	43.07	65.29	77.74	89.57	100
Q93	18.10	36.86	48.74	61.42	72.12	81.94	100	13.12	27.49	41.87	53.21	71.73	87.64	100
Q10	11.57	30.29	37.67	53.62	69.73	89.31	100	17.88	31.04	42.84	54.59	72.93	90.69	100
Q11	8.73	24.14	41.35	53.64	68.78	81.48	100	18.95	30.20	42.98	60.85	74.34	86.38	100

FIG. 5.3 – Evolution du PRESS total dans l'ACIMOG-PLS2



(Fig. 5.3). Ces résultats ne sont que très moyens, étant donné que le rang maximal possible des modèles est 7. Cela signifie qu'ils n'ont pas un bon pouvoir prédictif bien qu'ils reconstituent les données.

Il est aussi possible d'étudier le PRESS par tableau réponse, et de retenir un modèle d'ordre différent pour chacun d'eux.

Le critère Q^2 (I-6.2), dit que seulement la première composante est significative ($Q_n^2(1) \geq 0.0975$) et uniquement pour les tableaux $Q22$, $Q23$, $Q24$, $Q52$ et $Q7$ pour le panel 2, et pour le panel 1 : ce sont les tableaux étant expliqués à plus de 30%. Cela confirme que nos modèles ne sont pas très «bons».

La figure 5.4 représente pour le panel 1, les individus donnés par (t_1, t_2) et les covariances entre tous les attributs sensoriels (variables explicatives) et les super-composantes t_1 et t_2 normées à l'unité : on peut ainsi visualiser la position relative des 8 produits (vus par les deux premières super-composantes) et interpréter ces positions grâce à la représentation des variables explicatives. La figure 5.5 représente la même chose pour le panel 2. Les 10 notes pour un même attribut sensoriel sont groupées. Ainsi, nous pouvons dire que la première composante du panel 1 oppose le produit 5, caractérisé par l'attribut r (lait concentré sucré), aux autres produits. La deuxième composante sépare les produits 2 et 4, caractérisés par l'attribut g (quantité de fourrage), au produit 1 caractérisé par l'attribut d (quantité d'éclats). Quant au panel 2, la super composante t_1 oppose les produits 4 (et 5) aux produits 3 et 7, tandis que l'axe 2 oppose les produits 1 et 2. Le produit 5 est encore caractérisé par l'attribut r et le produit 1 par l'attribut d . Le produit 4, n'est pas caractérisé par les mêmes attributs selon les 2 panels : g et d (quantité de fourrage et quantité d'éclats) pour le premier panel, et d , k , e (quantité d'éclats, croustillant, grosseur des éclats) et éventuellement g (quantité fourrage) pour le second panel.

Considérons le descripteur $Q21$, qui est bien reconstitué dès la deuxième dimension dans les deux panels. La figure 5.6 donne pour le panel 1, l'évolution en fonction de a de $\|\hat{\beta}_n^a\|^2$ et $B_n^a[k]$. Ainsi, nous pouvons lire qu'à l'étape 3, la valeur du descripteur hédonique $Q21$ (couleur de la surface du chocolat) est surtout due aux propriétés sensorielles b (intensité de la couleur) et g (quantité fourrage) car ils ont les plus gros coefficients.

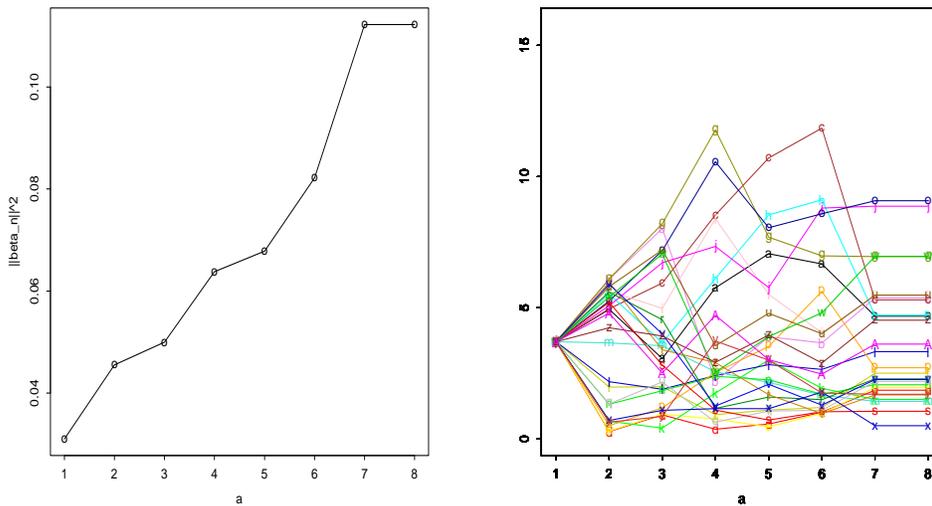
En ce qui concerne l'ACIMOG-PLS3, le tableau 5.3 donne les pourcentages de variance des Y_n expliquées. Nous pouvons constater que les résultats sont très proches de ceux issus de l'ACIMOG-PLS2. Pour le Panel 1, à la deuxième étape ($a = 2$) ceux sont les mêmes descripteurs qui sont expliqués à plus de 50% avec des $\%R^2$ similaires (entre les deux ACIMOG-PLS) à 1 ou 2 % près. Pour le panel 2, la tendance est la même mais les écarts entre les deux ACIMOG-PLS sont plus marqués. Les corrélations entre les super-composantes (Tab. 5.4) le confirment. De plus, les corrélations entre les composantes $t_{k,a}$ d'un même tableau à chaque étape, sont supérieures à 0.9.

Nous ne présentons pas les graphiques pour l'ACIMOG-PLS3, ceux-ci étant similaires à ceux de l'ACIMOG-PLS2.

Remarque : Dans l'ACIMOG-PLS2, les coefficients $\mu_{k,a}$, qui sont égaux à $\sum_{n=1}^N cov(u_{n,a}, t_{k,a})$, permettent de savoir quels sont les tableaux X_k , ici les attributs, qui expliquent les axes, en valeur absolue (Fig.5.7) : les plus gros coefficients $\mu_{k,a}$ correspondent aux tableaux X_k les plus covariés avec les composantes, c'est-à-dire à ceux qui sont le mieux reconstitués par la composante. Les coefficients $\lambda_{n,a}$ permettent de savoir quels sont les tableaux Y_n qui sont le mieux expliqués par la composante t_a (voir I-5.6).

L'ACIMOG-PLS3 étant finalement assez proche de l'ACIMOG-PLS2, ces résultats restent

FIG. 5.6 – Panel 1: représentation de l'évolution en fonction de a de $\|\hat{\beta}_n^a\|^2$ et $B_n^a[k]$ pour le descripteur Q21



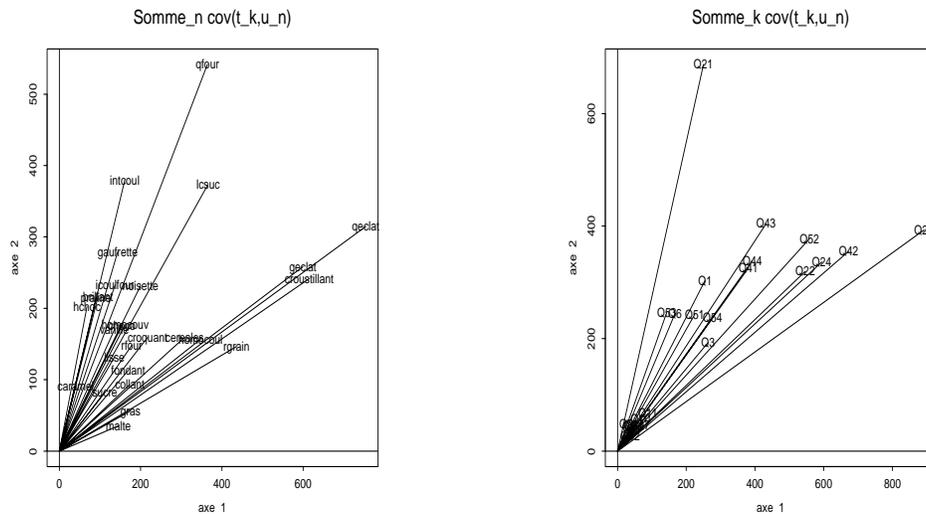
TAB. 5.3 – Pourcentage de variance expliquée des Y_i par l'ACIMOG-PLS3

	Panel 1 : non initiés							Panel 2 : initiés						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Q1	18.82	50.25	54.67	69.50	82.30	91.76	100	16.00	40.88	57.45	69.91	81.04	90.05	100
Q21	43.90	76.80	80.16	91.68	96.81	97.98	100	3.96	64.20	75.59	87.38	93.76	96.08	100
Q22	12.87	32.18	57.13	65.98	78.61	93.24	100	27.70	38.81	58.60	67.65	77.12	90.04	100
Q23	41.03	72.15	81.64	86.46	91.14	97.51	100	62.29	66.36	78.71	87.20	91.42	97.91	100
Q24	29.76	58.11	74.67	82.78	91.95	96.43	100	48.96	53.57	69.69	81.88	91.49	97.27	100
Q3	14.77	30.2	46.96	60.50	72.23	82.28	100	17.61	29.08	41.53	56.53	70.00	86.50	100
Q41	12.53	42.80	57.22	67.78	80.41	87.52	100	25.05	41.87	59.42	66.81	75.29	89.40	100
Q42	44.86	51.90	64.30	80.93	89.30	94.31	100	34.96	53.01	61.05	86.53	91.50	97.77	100
Q43	14.23	40.12	50.93	63.32	70.00	86.59	100	23.34	44.51	54.50	65.96	80.49	89.93	100
Q44	13.27	36.16	48.49	60.95	74.28	84.75	100	20.22	36.59	52.12	63.83	72.11	85.94	100
Q51	14.25	26.37	37.99	51.79	64.31	84.74	100	15.41	35.86	48.15	57.06	72.15	84.26	100
Q52	25.56	39.29	59.82	72.49	82.51	92.74	100	31.92	51.23	64.19	73.97	84.29	93.52	100
Q53	16.78	28.27	42.74	56.94	76.74	87.97	100	7.07	31.76	64.34	71.24	83.48	94.68	100
Q54	22.74	31.78	45.30	60.79	73.75	86.89	100	18.45	33.61	48.05	62.77	74.60	88.22	100
Q6	11.73	53.04	74.94	79.92	85.52	92.11	100	15.47	49.99	71.92	80.99	88.12	91.71	100
Q7	20.96	32.51	47.14	68.65	83.37	94.53	100	31.09	45.16	64.09	78.22	87.10	95.78	100
Q81	18.77	35.57	48.19	60.56	75.22	86.95	100	15.20	29.18	43.28	58.31	70.71	87.65	100
Q82	19.86	39.93	49.31	64.24	76.11	86.18	100	16.69	34.60	46.80	62.58	74.67	85.52	100
Q91	12.64	30.79	39.30	53.68	71.04	86.90	100	10.40	36.76	47.90	59.03	72.01	86.17	100
Q92	15.90	30.21	50.84	63.04	72.16	87.67	100	20.63	32.90	41.87	66.28	77.21	89.59	100
Q93	17.60	37.73	49.22	60.75	71.92	81.92	100	12.61	29.49	44.71	54.89	71.63	87.74	100
Q10	11.32	29.74	37.43	53.27	69.33	89.18	100	15.63	29.00	41.82	54.42	72.67	90.59	100
Q11	9.50	24.74	41.27	53.92	69.20	81.43	100	19.84	30.90	41.61	61.50	74.39	86.27	100

TAB. 5.4 – Corrélations entre les super-composantes t des ACIMOG-PLS2 et ACIMOG-PLS3.

a	1	2	3	4	5	6	7
Panel 1	0.98	-0.98	-1	0.99	1	1	1
Panel 2	0.97	0.97	0.95	0.95	0.98	1	1

FIG. 5.7 – Coefficients $(\mu_{k,1}, \mu_{k,2})$, et $(\lambda_{n,1}, \lambda_{n,2})$ de l'ACIMOG-PLS2 pour le panel 2



TAB. 5.5 – Attributs sensoriels explicatifs importants dans l'ACIMOG-PLS2 selon les critères de sélection de blocs

	Naïfs		Initiés	
	importants	peu importants	importants	peu importants
$RV x_k^1(k)$	b q r v	“les autres”	d e f l n q	“les autres”
$RV res_n^2(k)$	d e i l	“les autres”	b i r v	“les autres”
$RV res_n^3(k)$		“tous”		“tous”
$RV x_k^3(k)$	d g l m n o	a c h k p s t u w x y z A	$RV res_n^4(k)$	“tous”
$RV \beta_n^3(k)$	b d l m n o	a c h k s t u w x y z A	$RV x_k^4(k)$	f g l m n q r a c h k o p s t u w x y A
			$RV \beta_n^4(k)$	l m n v a c e h k o p s t u w x y A

aussi valables pour l'ACIMOG-PLS3.

Pour affiner les modèles, nous pouvons procéder à la sélection de tableaux X_k , c'est-à-dire de descripteurs sensoriels. Pour cela, nous allons utiliser les critères définis au chapitre I-6.3.2. Le critère BIP n'est pas utilisable avec l'ACIMOG-PLS2, mais l'est pour l'ACIMOG-PLS3. Les critères basés sur l'utilisation des coefficients RV peuvent être utilisés quelle que soit la méthode. Ainsi, leur application nous a permis de retenir les attributs sensoriels explicatifs contenus dans le tableau 5.5. L'ACIMOG-PLS 2 a été faite à nouveau en utilisant uniquement les 14 attributs b, d, e, f, g, i, j, l, m, n, o, q, r, v. Les variances expliquées par les modèles sont en général plus élevées (tableau 5.6). De plus, la validation croisée indique que 2 étapes suffisent pour obtenir des modèles stables. Ainsi, la sélection de blocs, a permis de stabiliser de façon adéquate les modèles.

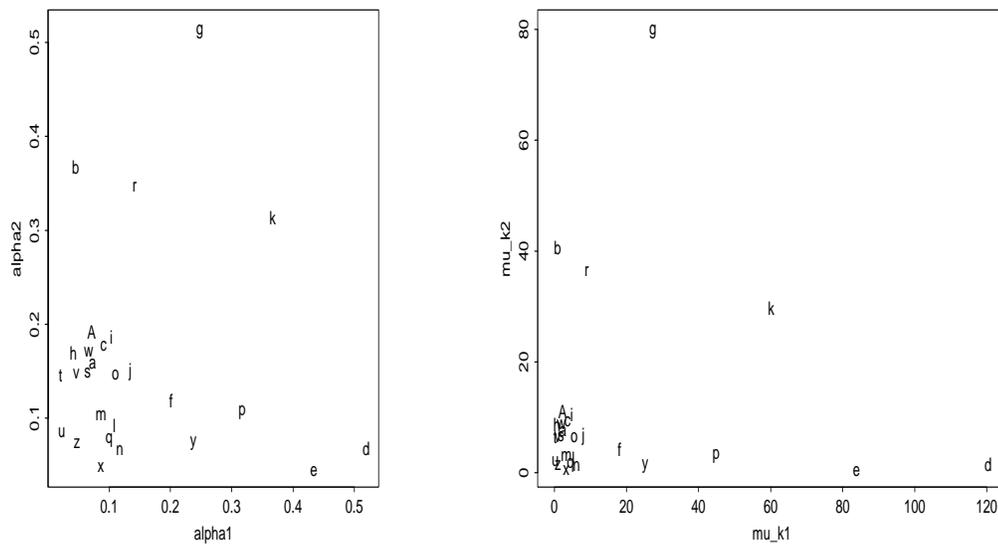
TAB. 5.6 – Variance des Y_n expliquée par l'ACIMOG-PLS2, avec les 14 attributs sélectionnés

a	Panel 1 : naïfs							Panel 2 : initiés						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Q1	32.69	51.68	59.73	74.46	83.58	92.79	100	24.23	48.70	60.00	72.69	85.18	93.88	100
Q21	36.67	80.69	86.67	93.85	96.70	99.10	100	18.52	71.15	82.12	89.34	91.53	96.36	100
Q22	24.24	36.57	56.71	66.49	84.23	91.31	100	32.29	41.70	60.32	67.79	79.23	90.31	100
Q23	28.12	69.17	76.86	82.42	87.22	95.60	100	38.74	67.28	77.38	82.20	88.43	96.43	100
Q24	25.57	58.28	78.56	82.84	86.98	94.88	100	30.59	55.52	73.10	80.87	87.61	96.72	100
Q3	15.67	30.93	44.06	61.60	83.27	88.87	100	12.40	30.58	43.68	57.47	73.24	85.56	100
Q41	29.17	41.30	53.93	68.13	76.77	88.19	100	38.52	46.57	58.67	68.06	82.59	91.65	100
Q42	7.31	35.82	44.78	60.18	67.06	92.82	100	10.59	31.62	40.66	65.68	73.46	92.98	100
Q43	25.54	38.13	47.90	62.50	79.55	89.28	100	20.50	38.68	48.75	64.36	76.94	89.96	100
Q44	22.63	38.29	48.45	61.46	74.89	87.01	100	21.34	33.11	45.17	59.96	72.90	82.77	100
Q51	11.92	28.13	44.32	54.53	71.72	87.28	100	15.44	32.90	48.24	58.56	74.64	85.32	100
Q52	14.15	35.44	62.81	73.27	80.93	90.77	100	22.77	40.27	63.99	76.00	83.74	90.77	100
Q53	9.89	26.03	43.07	65.71	76.09	89.62	100	7.26	23.79	57.05	74.92	86.01	95.59	100
Q54	8.67	30.38	46.14	57.88	71.12	87.72	100	13.10	38.24	51.28	64.69	79.34	90.76	100
Q6	46.34	60.01	73.56	80.41	89.04	93.88	100	42.48	65.67	73.96	83.01	88.04	95.90	100
Q7	10.54	27.38	56.39	70.79	78.74	94.06	100	23.80	32.95	67.91	77.88	85.90	93.88	100
Q81	17.21	34.28	50.10	64.49	74.46	86.73	100	15.06	28.33	43.95	59.58	78.22	89.83	100
Q82	20.08	39.45	51.54	65.19	74.67	85.19	100	14.32	31.35	46.02	62.37	77.93	87.60	100
Q91	19.77	30.96	45.07	56.44	72.69	87.38	100	13.69	34.56	49.02	60.48	76.19	84.29	100
Q92	14.40	27.21	53.00	61.29	76.00	85.76	100	16.67	34.16	47.10	63.74	77.01	88.21	100
Q93	19.91	37.75	47.75	62.47	74.98	83.05	100	16.33	32.16	44.03	57.36	71.80	85.64	100
Q10	18.52	29.38	46.07	57.68	75.06	89.68	100	16.29	25.67	42.76	57.10	74.00	88.88	100
Q11	17.60	30.03	42.90	56.39	71.09	87.24	100	16.52	30.46	45.55	57.11	71.33	86.34	100

TAB. 5.7 – Pourcentage des variances des Y_n expliquées par REMUB

a	Panel 1 : non initiés							Panel 2 : initiés						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Q1	14.53	49.94	54.20	68.89	81.80	91.49	100	16.25	43.28	56.20	69.45	80.77	89.94	100
Q21	0.86	78.01	79.90	91.79	96.85	97.94	100	2.27	66.76	70.95	86.32	93.83	96.11	100
Q22	20.42	32.26	57.89	66.07	79.73	93.05	100	25.95	37.36	58.87	67.95	77.54	89.94	100
Q23	66.62	72.62	81.53	86.46	91.22	97.63	100	66.78	71.07	80.16	87.23	91.33	97.92	100
Q24	52.74	58.64	75.40	82.67	92.16	96.54	100	54.76	58.97	72.22	81.58	91.37	97.23	100
Q3	9.81	30.55	46.57	60.14	72.71	81.59	100	15.96	28.70	41.06	56.47	69.83	86.53	100
Q41	20.82	42.67	56.85	67.46	80.25	87.42	100	27.18	43.12	60.23	66.91	75.37	89.28	100
Q42	28.70	50.74	63.21	81.84	89.60	94.25	100	29.63	53.09	65.93	86.90	91.47	97.82	100
Q43	25.85	39.92	50.49	63.03	69.98	86.14	100	23.26	44.24	55.21	65.96	80.43	89.87	100
Q44	17.21	36.03	48.00	61.077	74.30	84.57	100	20.03	35.82	52.63	63.93	72.20	86.02	100
Q51	10.39	26.01	38.47	51.67	63.61	84.24	100	14.86	34.83	47.76	57.08	71.89	84.26	100
Q52	15.57	38.70	60.95	72.54	82.32	92.77	100	26.82	45.05	63.66	73.52	84.29	93.53	100
Q53	7.76	27.87	42.40	56.14	76.09	88.07	100	7.53	27.66	64.60	71.14	83.81	94.60	100
Q54	13.70	31.79	45.79	60.94	73.88	87.04	100	17.51	34.07	46.45	62.65	74.65	88.10	100
Q6	13.94	54.37	74.80	80.05	85.67	91.88	100	18.17	54.63	71.16	80.95	88.28	91.64	100
Q7	18.15	31.83	48.89	68.37	82.79	94.35	100	25.65	36.51	62.70	77.80	86.88	95.81	100
Q81	18.62	35.53	48.42	60.35	74.51	86.90	100	15.29	29.32	43.54	57.98	70.26	87.60	100
Q82	21.17	40.04	49.48	63.98	76.20	86.17	100	15.78	33.83	46.87	62.25	74.66	85.45	100
Q91	16.58	30.47	39.57	53.64	71.30	86.69	100	9.92	37.60	48.04	58.64	72.12	86.13	100
Q92	12.21	29.55	51.74	63.10	72.19	87.75	100	18.90	30.64	40.60	66.30	77.16	89.67	100
Q93	18.44	37.75	49.02	60.28	71.90	82.14	100	12.34	30.39	45.18	54.98	71.83	87.78	100
Q10	15.27	29.63	38.06	52.93	68.64	88.52	100	14.76	27.26	41.94	54.33	72.30	90.54	100
Q11	14.88	25.35	41.34	53.96	67.00	80.85	100	19.72	31.74	42.78	61.86	74.29	86.26	100

FIG. 5.8 – Représentation des coefficients $(\mu_{k,1}, \mu_{k,2})$ et (α_1, α_2) de REMUB pour le panel 2



5.2.3 Application de REMUB

REMUB donne des variances expliquées comparables quel que soit le panel (tab.5.7). Remarquons que le descripteur $Q21$ (intensité couleur) du panel 1 n'est quasiment pas expliqué à la première étape alors qu'il l'était beaucoup avec les ACIMOG-PLS.

Les coefficients $\alpha_{k,a}$ et $\mu_{k,a}$ sont proportionnels (Fig.5.8). Nous constatons que les plus gros coefficients correspondent aux attributs les plus covariés avec les super-composantes (Fig.5.9). La représentation des produits dans le premier plan (Fig.5.9) est très proche de celle fournie par les ACIMOG-PLS.

5.3 Situation 2 : analyse de la proximité des panels de consommateurs initiés et non initiés : ACIMOG, DO-ACT

Dans la section précédente, nous avons vu que les modèles pour les enfants initiés étaient très proches de ceux des enfants non initiés. Pour être sûr qu'il n'y a pas de grandes différences entre les deux panels de consommateurs, nous allons utiliser les méthodes ACIMOG (I-3.4) et DO-ACT (I-3.3) pour chercher s'il y a une structure commune aux deux panels de consommateurs et si oui, en chercher la cause.

5.3.1 ACIMOG

Nous allons ici appliquer les ACIMOG 1 et 2 aux évaluations des enfants. Le panel 1 sera désigné par les tableaux X_k et le panel 2 par les tableaux Y_k . Tout comme avec l'utilisation de l'ACIMOG-PLS, les tableaux sont uniquement centrés pour la métrique uniforme.

Nous n'allons pas détailler les résultats mais nous verrons que les ACIMOG1, laquelle utilise des déflations sur les composantes de chacun des blocs, et ACIMOG2 sont complémentaires :

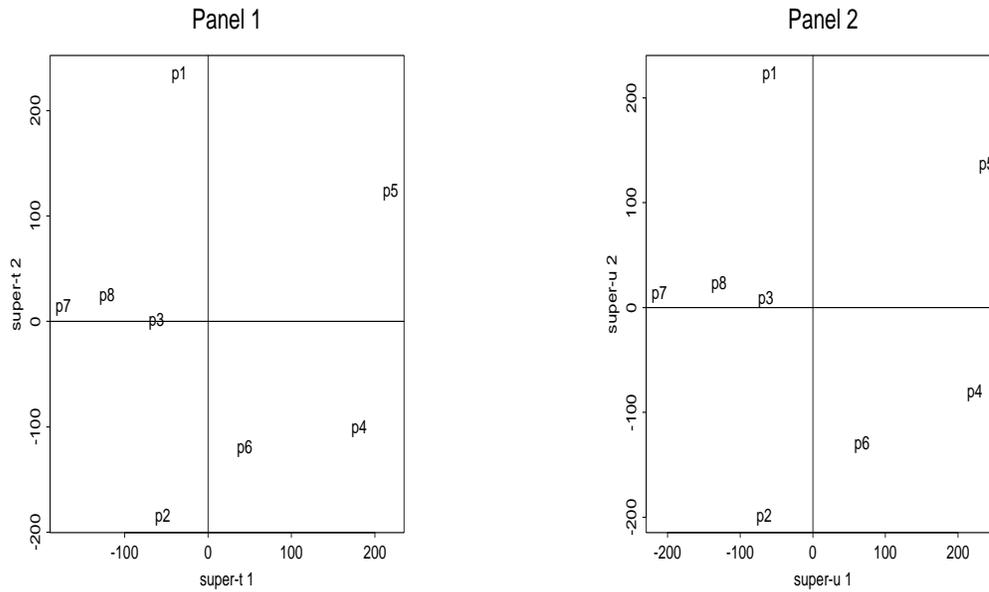
TAB. 5.8 – Variances expliquées des enfants non initiés par les ACIMOG 1 et 2

a	ACIMOG1							a	ACIMOG2						
	1	2	3	4	5	6	7		1	2	3	4	5	6	7
Q1	18.90	67.70	76.44	86.99	92.48	96.93	100	Q1	13.54	52.53	56.70	68.77	78.59	91.83	100
Q21	91.73	93.65	95.43	96.32	97.61	99.05	100	Q21	18.09	79.47	89.41	95.80	97.19	98.75	100
Q22	48.42	61.81	70.10	79.09	85.58	93.40	100	Q22	19.97	35.38	58.71	68.42	86.20	92.71	100
Q23	74.29	79.20	83.84	90.50	94.30	97.05	100	Q23	54.96	71.30	81.93	85.71	91.83	96.96	100
Q24	67.38	75.84	84.86	89.47	94.10	98.31	100	Q24	41.15	60.47	76.91	84.29	91.16	95.80	100
Q3	15.01	30.38	49.33	64.36	79.52	93.54	100	Q3	10.73	30.02	49.65	63.43	80.80	86.00	100
Q41	22.51	52.63	66.34	76.41	84.19	93.24	100	Q41	13.35	37.87	55.37	68.99	79.85	88.58	100
Q42	65.95	72.50	81.93	86.82	91.49	97.08	100	Q42	49.81	56.41	68.25	80.64	90.99	94.94	100
Q43	27.45	46.00	57.92	70.02	84.12	90.64	100	Q43	18.19	40.17	51.30	63.57	75.30	87.92	100
Q44	18.54	45.19	61.45	72.92	81.09	92.03	100	Q44	13.69	35.88	50.55	63.43	75.15	85.80	100
Q51	17.02	31.078	49.24	65.37	78.85	89.10	100	Q51	11.59	27.25	39.51	52.59	62.91	86.17	100
Q52	39.34	54.17	67.54	76.96	85.58	92.65	100	Q52	23.32	42.45	60.13	74.50	82.32	92.42	100
Q53	23.90	51.60	66.56	76.81	86.48	91.55	100	Q53	13.74	27.63	43.09	61.22	73.85	88.37	100
Q54	26.89	41.51	57.89	71.87	79.73	91.02	100	Q54	22.69	33.22	45.01	60.86	74.36	86.97	100
Q6	24.26	67.88	79.25	84.97	90.54	95.77	100	Q6	5.29	49.77	77.12	82.92	88.68	94.02	100
Q7	44.29	59.04	71.68	81.51	89.69	97.50	100	Q7	22.31	34.07	46.39	72.46	81.91	94.96	100
Q81	21.59	38.82	58.79	68.64	82.01	90.41	100	Q81	18.58	34.95	48.54	63.86	73.14	87.34	100
Q82	21.42	46.42	62.13	72.61	83.92	94.10	100	Q82	19.92	35.89	46.51	64.42	76.14	86.12	100
Q91	22.51	44.39	67.80	81.97	94.13	96.55	100	Q91	13.58	33.61	40.81	54.64	71.52	87.83	100
Q92	22.35	41.60	65.53	78.36	86.49	95.25	100	Q92	13.58	33.91	51.09	62.36	75.69	88.48	100
Q93	18.59	42.87	51.00	65.77	77.45	88.19	100	Q93	16.76	32.79	44.52	58.74	71.87	82.18	100
Q10	13.37	37.71	64.82	77.98	83.55	96.13	100	Q10	12.01	29.56	39.80	57.18	72.67	89.97	100
Q11	18.27	37.71	51.02	61.30	81.21	89.54	100	Q11	10.79	24.61	40.72	52.70	63.19	83.15	100

TAB. 5.9 – Variances expliquées des enfants initiés par les ACIMOG 1 et 2

a	ACIMOG1							a	ACIMOG2						
	1	2	3	4	5	6	7		1	2	3	4	5	6	7
Q1	22.03	54.30	67.27	77.41	86.94	90.46	100	Q1	14.37	41.09	57.08	70.81	78.95	91.78	100
Q21	84.19	86.63	90.95	93.59	95.37	97.66	100	Q21	15.04	72.22	83.44	91.55	94.95	96.51	100
Q22	44.35	58.56	69.10	77.85	85.81	94.37	100	Q22	23.53	39.06	59.18	70.73	84.60	90.64	100
Q23	73.84	79.73	85.77	90.29	94.07	98.97	100	Q23	54.57	69.42	79.49	85.96	90.97	97.23	100
Q24	69.29	73.49	80.90	86.28	94.57	99.20	100	Q24	45.57	61.78	74.76	83.11	88.67	97.10	100
Q3	20.59	31.83	44.37	58.58	72.17	85.42	100	Q3	19.37	29.67	41.92	56.75	68.18	86.17	100
Q41	41.86	58.58	72.52	80.66	89.30	95.44	100	Q41	17.43	43.24	64.28	72.76	81.59	90.41	100
Q42	72.79	78.26	84.38	91.71	94.67	98.39	100	Q42	51.32	56.71	68.05	84.89	91.45	96.71	100
Q43	27.03	48.29	60.96	71.43	82.89	90.56	100	Q43	25.19	44.55	55.51	65.63	77.36	90.56	100
Q44	3.54	43.36	56.76	72.68	83.52	92.37	100	Q44	19.66	36.05	52.86	64.50	74.81	85.02	100
Q51	20.75	38.76	55.40	65.43	82.49	88.59	100	Q51	16.58	35.94	49.05	58.57	66.89	84.17	100
Q52	43.16	59.85	71.92	80.83	86.60	91.83	100	Q52	31.23	47.32	60.23	74.34	83.05	93.35	100
Q53	32.53	57.88	82.53	90.63	94.51	98.19	100	Q53	13.14	30.27	60.33	68.48	87.17	94.71	100
Q54	28.86	42.51	55.72	68.61	79.51	89.28	100	Q54	19.97	34.03	48.10	64.15	79.05	89.92	100
Q6	12.25	70.65	79.12	84.04	88.66	96.26	100	Q6	9.96	51.66	75.03	83.48	90.51	94.00	100
Q7	53.77	68.57	79.90	85.10	94.56	99.26	100	Q7	24.79	37.71	56.80	76.88	83.33	94.47	100
Q81	4.42	34.16	54.23	67.32	79.34	86.54	100	Q81	13.92	27.22	42.03	59.06	68.47	88.23	100
Q82	21.11	39.61	52.41	66.74	84.48	92.56	100	Q82	18.92	35.35	47.03	61.95	73.19	86.63	100
Q91	25.07	39.56	55.58	70.20	84.12	92.23	100	Q91	14.59	37.30	47.51	60.85	73.65	85.95	100
Q92	23.82	44.55	60.47	78.74	86.41	96.65	100	Q92	18.78	29.68	39.50	62.56	74.44	88.54	100
Q93	15.95	45.99	58.45	72.90	82.82	91.23	100	Q93	12.43	26.09	40.62	51.72	72.93	87.15	100
Q10	23.05	40.92	51.60	67.36	78.39	91.22	100	Q10	14.17	26.42	40.15	56.22	69.49	90.28	100
Q11	21.94	43.14	62.71	74.69	86.07	93.24	100	Q11	20.34	32.88	46.07	65.02	74.06	86.92	100

FIG. 5.10 – Représentation des super-composantes de l'ACIMOG2 pour le panel 1 (non-initiés) et le panel 2 (initiés)



enfants. Les attributs les mieux expliqués par une étape sont ceux dont les cartes des produits vues par les deux panels sont très proches. Ainsi, on note la similitude des représentations sur l'axe 1 pour $Q21$, $Q23$, $Q24$ et $Q42$. Les enfants, initiés ou non, sont du même avis pour ces descripteurs, qui décrivent la surface du chocolat. Le quatrième descripteur de surface $Q22$, (homogénéité de la couverture), est aussi assez bien reconstitué par l'étape 1 de l'ACIMOG1, et les représentations des produits sont proches.

Nous voyons que les enfants ont du mal à évaluer le caractère moelleux, $Q44$. Ce tableau est mal expliqué par les premières étapes de l'ACIMOG1, et la représentation des produits présente quelques différences entre les deux panels.

Dans l'ACIMOG1, chaque composante explique quelque chose de nouveau à chaque étape. Une représentation étoilée telle que la figure 5.13 permet de «comparer» la position de chaque produit (pour chaque descripteur) avec la position moyenne du produit (vue par tous les descripteurs). Le numéro au centre d'une «étoile» représente le produit vu par la super-composante normée à l'unité, et les extrémités des branches sont les positions des individus vus par les composantes partielles de norme unité, c'est-à-dire pour chaque descripteur.

5.3.2 DO-ACT

Rappelons que, selon les notations du chapitre I-3.3, $K = N = 23$.

Tout comme dans l'ACIMOG (paragraphe précédent), le premier ensemble de tableaux X_k correspond au panel 1 de consommateurs, soit les enfants naïfs, et les tableaux Y_k correspondent au panel 2, soit les enfants initiés.

Nous allons utiliser ici la méthode DO-ACT (voir I-3.3) pour étudier la proximité des deux panels. Rappelons qu'elle est composée de trois étapes, comme STATIS : interstructure, compromis, intrastructure.

FIG. 5.11 – ACIMOG1. Représentations des produits pour chacun des attributs décrits par le Panel 1, données par $(t_{k,1}, t_{k,2})$.

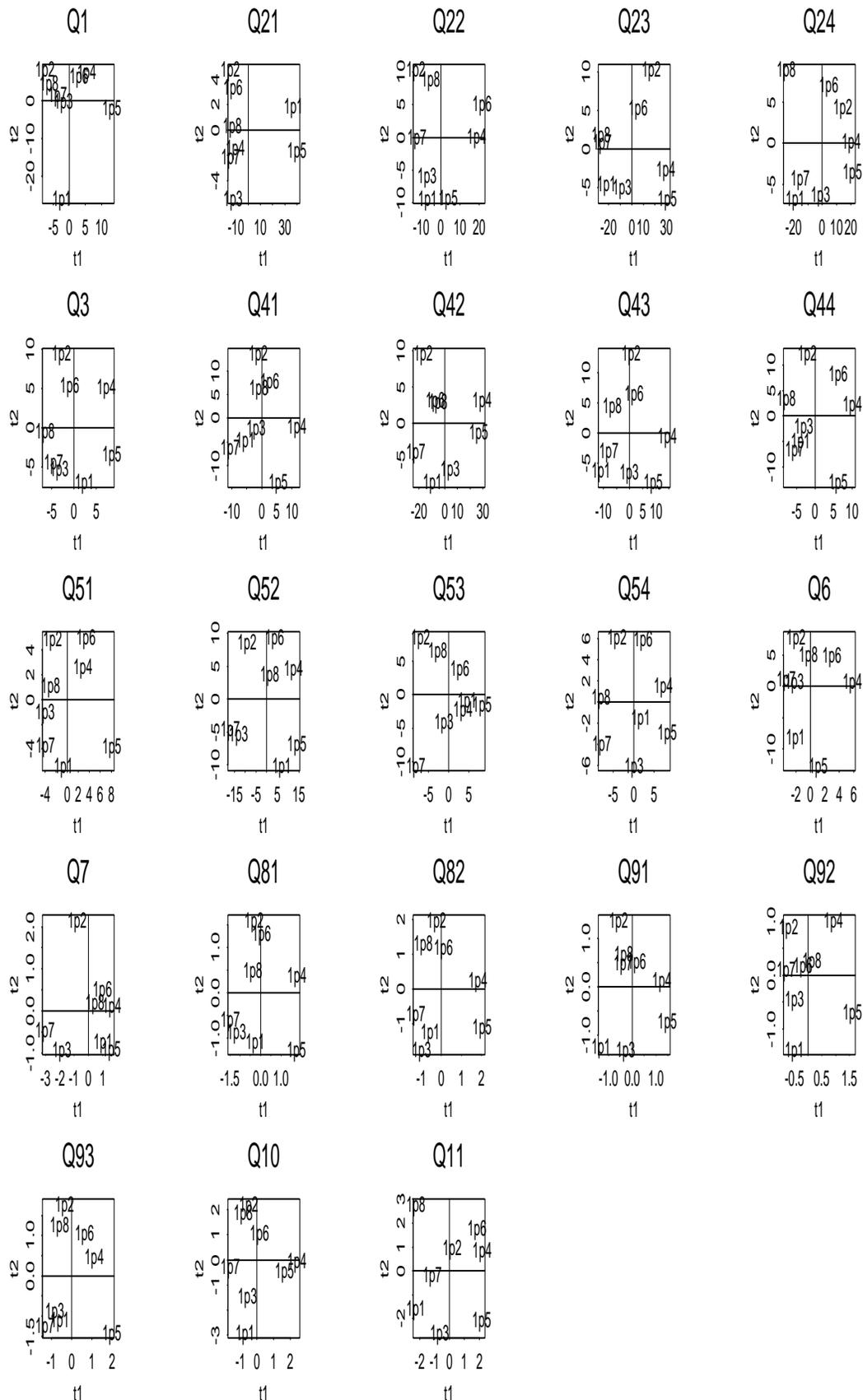


FIG. 5.12 – ACIMOG1. Représentations des produits pour chacun des attributs décrits par le panel 2, données par $(u_{n,1}, u_{n,2})$.

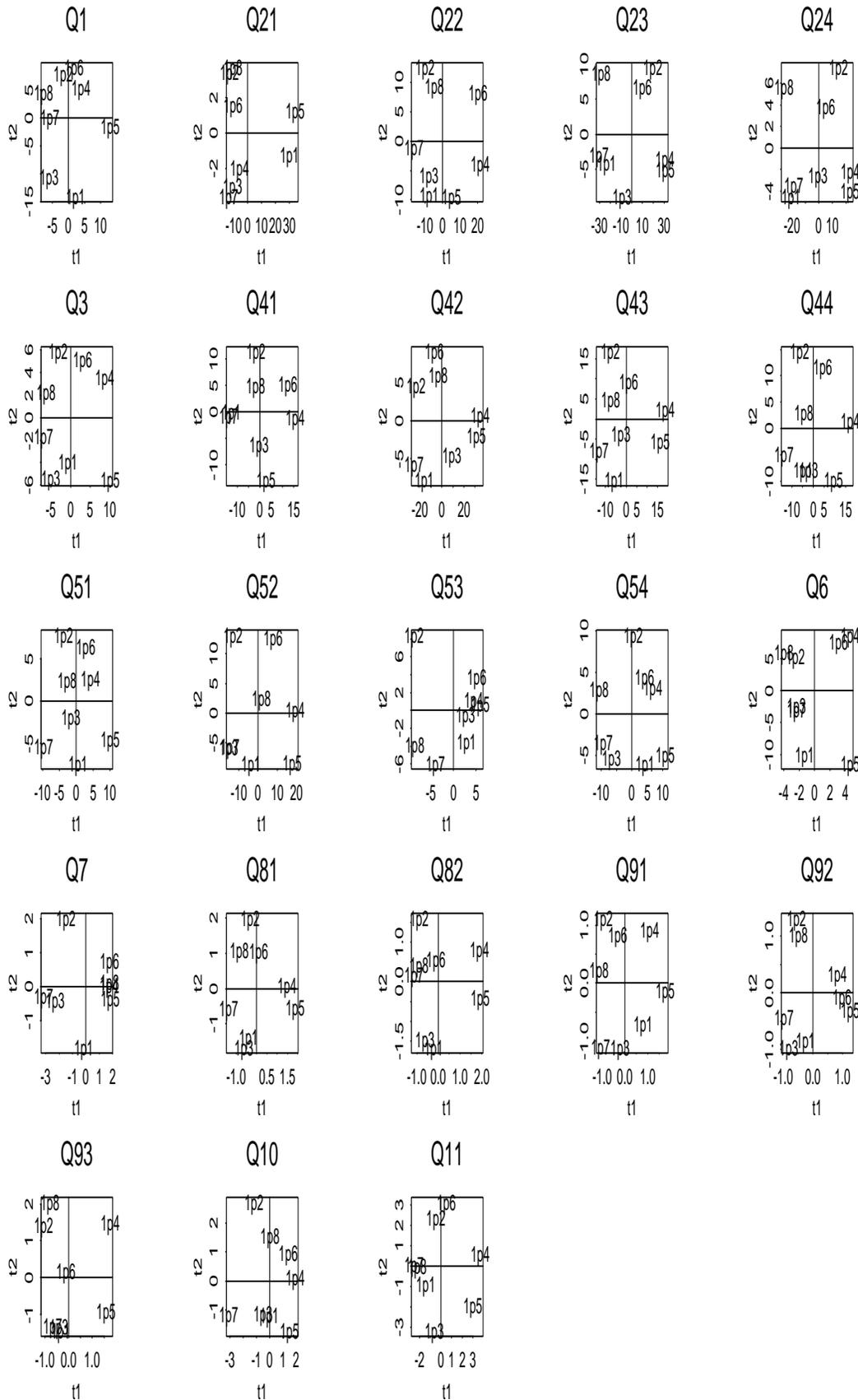
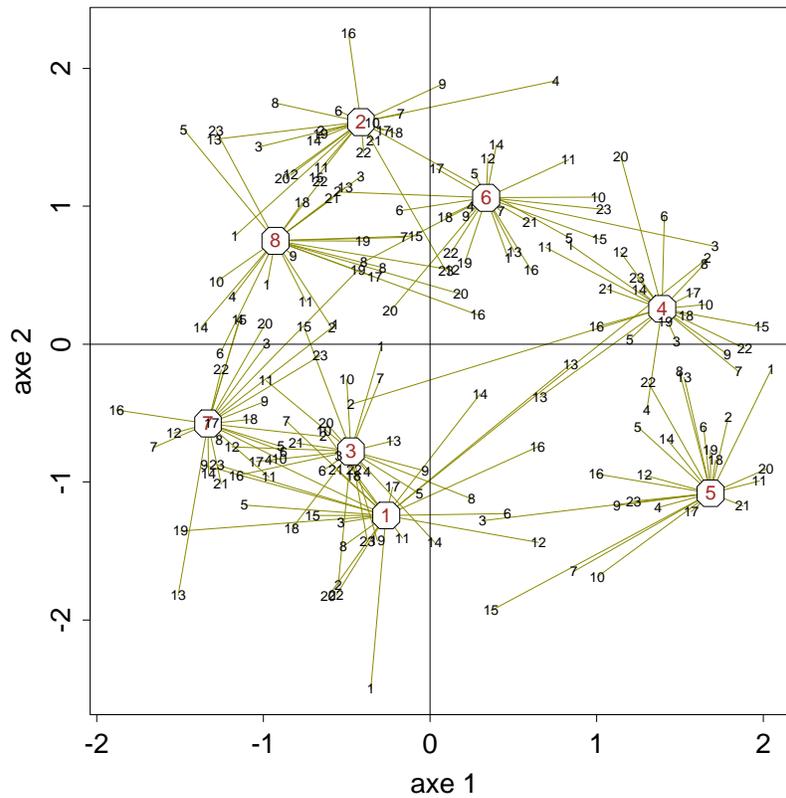


FIG. 5.13 – Représentation étoilée des produits vus par les enfants non-initiés dans l'ACIMOG1, donnée par $(t_1/\|t_1\|_D, t_2/\|t_2\|_D)$ (au centre des étoiles), et $(t_{k,2}/\|t_{k,2}\|_D, t_{k,2}/\|t_{k,2}\|_D)$, où $t_a = \sum_{k=1}^K t_{k,a}$.



TAB. 5.10 – Normes des opérateurs associés aux tableaux X_k et Y_k avant réduction

X_k	Q1	Q21	Q22	Q23	Q24	Q3	Q41	Q42	Q43	Q44	Q51	Q52
	139.60	533.05	226.66	520.13	315.88	93.39	108.95	337.85	114.92	99.80	51.38	167.57
	Q53	Q54	Q6	Q7	Q81	Q82	Q91	Q92	Q93	Q10	Q11	
	64.30	56.45	66.177	3.53	2.55	2.88	1.82	1.60	2.54	5.76	8.70	
	Q1	Q21	Q22	Q23	Q24	Q3	Q41	Q42	Q43	Q44	Q51	Q52
Y_k	98.56	394.95	262.43	621.86	364.34	75.68	168.27	500.07	186.71	163.29	65.94	216.16
	Q53	Q54	Q6	Q7	Q81	Q82	Q91	Q92	Q93	Q10	Q11	
	69.71	89.10	61.57	4.28	3.20	2.66	2.17	1.41	3.00	5.51	9.16	

DO-ACT a été appliquée sur les tableaux centrés en colonnes pour la métrique uniforme (D), avec des métriques de pondérations des variables et des tableaux égales à l'identité. De plus, les calculs sont faits avec les coefficients RV, c'est-à-dire que les opérateurs sont normés.

Les opérateurs $W_k^x D = X_k' X_k D$ d'une part, et les opérateurs $W_k^y D$ d'autre part, ont des normes très disparates (tableau 5.10). Les tableaux Q6 à Q11 (X ou Y) sont notés sur 2 à 5 modalités, ce qui entraîne forcément une norme plus faible que celle des opérateurs des tableaux Q1 à Q54 qui sont notés de 0 à 9. Ceci justifie l'emploi des opérateurs normés dans les calculs.

Les coefficients RV entre les tableaux X_k et Y_k varient de 0.10 à 1: ainsi certains couples de tableaux sont très proches, alors que d'autres sont très éloignés.

La première étape de DO-ACT est l'étude de l'interstructure. La figure 5.14 représente le premier plan de l'interstructure pour les deux panels. Ce premier plan, comme dans STATIS, explique la plus grande part de la variabilité des opérateurs (ici 96.95% pour l'axe 1 et 1.5% pour le second). Les attributs les moins bien représentés par le premier axe mais plutôt par le second sont les mêmes chez les deux panels (Q1, Q21, Q23, Q24, Q42 et Q6). Une très grosse majorité des attributs sont représentés par le premier axe, mais ils ne sont pas tous placés de la même manière dans le plan, selon qu'ils soient du panel 1 ou 2 (par exemple Q3, Q10, Q91 et Q11). Toutefois, ce graphique nous amène à penser que les deux panels d'enfants répondent globalement de la même manière aux questions qui leur sont posées.

Remarquons, que pour le panel 2, l'attribut Q3, qui a une forte coordonnée sur l'axe 1, possède 19 coefficients RV (avec les attributs du panel 1) sur 23 supérieurs à 0.6, la plupart d'entre eux étant supérieurs à 0.7, alors que, inversement, l'attribut Q21 n'en a que 3.

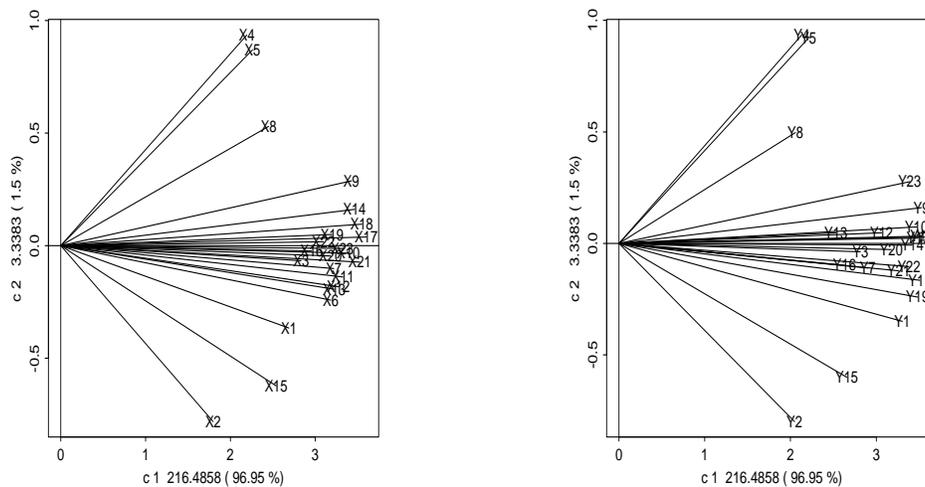
Passons à la seconde étape: le compromis.

L'inertie du compromis $W_X D$, pour le panel 1, est de 9.87 et celle de $W_Y D$, panel 2, est de 9.96, et le coefficient RV entre eux deux vaut 0.99: ces deux compromis sont très proches. Le coefficient RV entre les deux compromis issus de deux STATIS (sur le panel 1 et sur le panel 2) est de 0.98.

Les coefficients pondérant les opérateurs $W_k^x D$ et $W_k^y D$ dans leur compromis sont donnés dans le tableau 5.11. Les tableaux X_k les plus importants dans la construction du compromis $W_X D$ sont Q81 et Q82 avec un coefficient $\alpha_k = 0.24$. Tandis que les tableaux Y_k les plus importants sont Q3, Q43, Q44, Q81, Q82 et Q91 avec un coefficient $\beta_k = 0.24$. Ainsi, ces couples d'attributs sont les plus importants dans la construction de $W_X D W_Y D$ avec un coefficient $\alpha_k \beta_{k'} = 0.24 \times 0.24 = 0.06$.

De plus, nous remarquons que dans les deux compromis, les coefficients sont compris entre 0.12 et 0.24, et on ne note pas de différences majeures entre les deux ensembles de tableaux. Si tous les tableaux avaient la même importance, tous les coefficients seraient égaux à $1/\sqrt{23} = 0.21$:

FIG. 5.14 – Interstructure de DO-ACT. Le panel 1 est représenté par les X_k , le panel 2 par les Y_k



les coefficients n'en sont pas si loin. Certains tableaux sont donc plus importants que les autres mais ce n'est pas flagrant.

Enfin, la troisième étape, l'intrastructure, va nous permettre d'étudier les positions des produits les uns par rapport aux autres et selon chaque tableau.

Les positions relatives de produits compromis sont similaires dans les deux panels (Fig.5.15). Les chocolats 5 et 1 s'opposent aux autres sur l'axe 2 et plus particulièrement au chocolat 2. L'axe 1 oppose les chocolats 1, 3 et 7 aux 4, 6 et 5. Cette configuration est la même (au sens de l'axe 2 près) que celle de REMUB (Fig.5.9). A nouveau ici, nous constatons la proximité des deux panels.

La figure 5.16 donne la position des produits dans chaque panel et permet une comparaison entre tableaux.

5.4 Conclusion

Grâce aux méthodes ACIMOG-PLS, ACIMOG et DO-ACT, nous avons abouti aux mêmes conclusions que SUNE et al. (2002). Les deux panels de consommateurs fournissent les mêmes évaluations des chocolats, qu'ils soient initiés ou non. De plus, les enfants ont plus de facilités avec les attributs visuels ($Q21$, $Q22$, $Q23$, $Q24$) d'après l'ACIMOG1.

Les méthodes de régression ont montré un lien entre le descripteur sensoriel *lait concentré sucré*, et le descripteur noté par les enfants *goût de lait*. De plus, il est ressorti que les descripteurs $Q42$, *croustillant*, et les descripteurs visuels $Q21$, $Q23$, $Q24$ (liés à la couleur et aux céréales) sont surtout liés aux descripteurs sensoriels des juges concernant la couleur (b) et les céréales.

Ainsi, nous montrons qu'il n'est pas nécessaire de résumer les données sensorielles (ce qui peut engendrer des pertes d'informations), de pratiquer plusieurs analyses, et de construire

TAB. 5.11 – Coefficients α_k et β_k des compromis de DO-ACT

k	X_k	α_k	Y_k	β_k
1	Q1	0.18	Q1	0.22
2	Q21	0.12	Q21	0.14
3	Q22	0.19	Q22	0.19
4	Q23	0.15	Q23	0.14
5	Q24	0.15	Q24	0.15
6	Q3	0.22	Q3	0.24
7	Q41	0.22	Q41	0.20
8	Q42	0.17	Q42	0.14
9	Q43	0.23	Q43	0.24
10	Q44	0.23	Q44	0.24
11	Q51	0.23	Q51	0.24
12	Q52	0.22	Q52	0.21
13	Q53	0.22	Q53	0.17
14	Q54	0.23	Q54	0.23
15	Q6	0.17	Q6	0.18
16	Q7	0.20	Q7	0.18
17	Q81	0.24	Q81	0.24
18	Q82	0.24	Q82	0.24
19	Q91	0.22	Q91	0.24
20	Q92	0.22	Q92	0.22
21	Q93	0.24	Q93	0.22
22	Q10	0.21	Q10	0.23
23	Q11	0.22	Q11	0.23

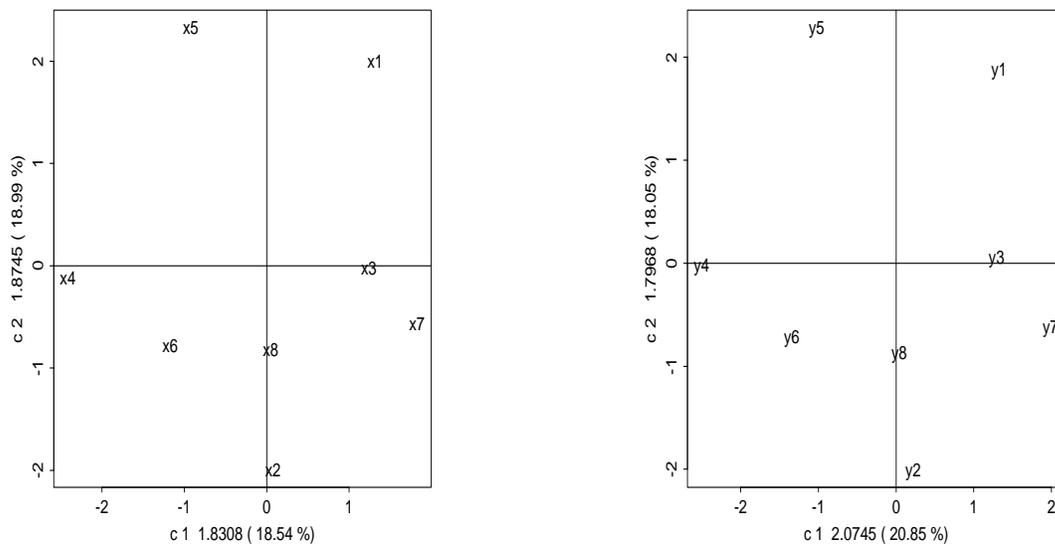
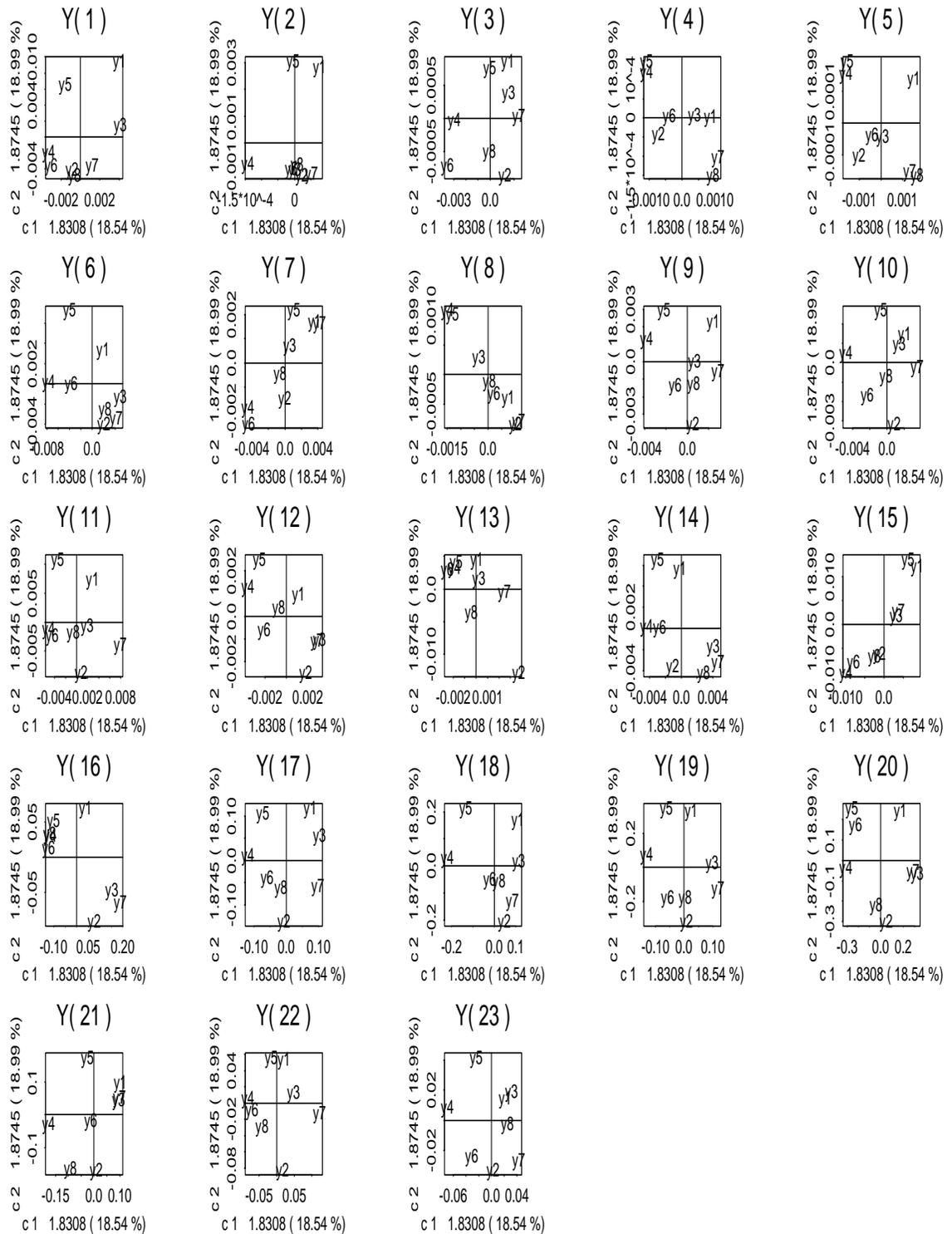
FIG. 5.15 – Positions des produits selon les compromis $W_X D$ et $W_Y D$ de DO-ACT

FIG. 5.16 – Positions des individus pour chaque Y_k (panel 2)



plusieurs modèles, pour aboutir aux résultats : il suffit d'appliquer l'ACIMOG-PLS. Ainsi, grâce à une seule et unique méthode, nous pouvons obtenir de bon modèles en terme d'explication donnant une vision d'ensemble sur les liens entre les tableaux. La sélection successive de blocs peut permettre d'affiner les modèles et de les stabiliser.

Les méthodes d'analyse conjointe, quant à elles, permettent de voir s'il est intéressant de passer du temps à initier des enfants à évaluer un produit, ou si cela n'est pas nécessaire.

Cependant, les méthodes que nous avons appliquées nécessitent encore quelques investigations pour faciliter un peu plus leur interprétation. Les premiers résultats sont prometteurs, puisqu'ils concordent avec ce qui avait déjà été fait.

Discussion et conclusion

A theory has only the alternative of being right or wrong. A model has a third possibility: it may be right but irrelevant.

Manfred Eigen

Dans ce travail, nous avons présenté un grand nombre de méthodes factorielles d'analyse de multi-tableaux pour la régression ou l'analyse de co-inertie, certaines étant issues de la littérature et douze nouvelles, en comptant les variantes.

Nous avons surtout insisté sur le domaine de la régression multi-tableaux en proposant cinq méthodes de type PLS construisant un modèle linéaire en les variables initiales ainsi que des critères aidant à la sélection de blocs de prédicteurs, et deux méthodes pour la modélisation non-linéaire.

Nous avons vu que les méthodes «cubiques» (chapitres I-1, I-4, I-3) sont généralement basées sur la recherche simultanée d'ensembles de composantes associés à chacun des modes des cubes à l'aide de la minimisation de l'erreur d'ajustement du modèle choisi. Ceci est réalisé à l'aide d'un algorithme aux moindres carrés alternés (ALS). La plupart du temps les composantes ne sont pas orthogonales, à moins qu'on ne l'impose, et chaque composante n'explique pas forcément quelque chose de nouveau. Alors que les méthodes multi-tableaux (chapitres I-2, I-3, I-5) sont de type «component-wise»: A composantes par tableaux et A super-composantes sont recherchées les unes après les autres, à l'aide d'algorithmes de type NIPALS (WOLD, 1966a).

Dans le chapitre I-2 nous avons proposé l'Analyse de Co-Inertie Multiple Orthogonale (ACIMO) comme alternative à l'Analyse de CO-inertie Multiple (ACOM) (CHESSEL & HANAFI, 1996) et à l'Analyse de COncordance (ACO) (LAFOSSE & HANAFI, 1997) qui cherchent à étudier les nuages des individus et ne fournissent donc pas des composantes orthogonales. L'ACIMO optimise le même critère de covariance que l'ACO mais l'utilisation de déflations dans l'espace des variables (\mathbb{R}^1) mène à des solutions différentes à partir de l'étape $a = 2$. De plus, nous avons montré comment utiliser l'ACIMO pour l'intrastructure de STATIS (§I-2.5). Les composantes u_a issues de l'ACIMO sont très proches, en pratique, des composantes prin-

principales de même ordre du compromis de STATIS. Ce phénomène résulte de la propriété 2.7 de la page 39 que nous avons établie. La pratique nous montre, comme dans l'exemple III-1 de la troisième partie, que les trajectoires obtenues par l'ACIMO sont souvent plus proches du point compromis et que les différentes trajectoires sont un peu plus distinctes les unes des autres.

Pour l'étude de la proximité de deux multi-tableaux (éventuellement deux cubes), chapitre 3, nous avons proposé deux méthodes aux approches différentes. La DOuble - Analyse Conjointe de Tableaux (DO-ACT, §I-3.3) travaille sur les opérateurs (comme STATIS). Elle cherche un opérateur compromis par multi-tableaux, chacun étant une combinaison linéaire des opérateurs des tableaux qu'il doit résumer, qui soient les plus proches possible en norme. La résolution de la minimisation mène à un problème aux valeurs propres, ce qui fait de DO-ACT une méthode rapide à mettre en oeuvre. Les possibilités graphiques sont assez nombreuses (§III-5). L'Analyse de Co-Inertie Multiple Orthogonale Généralisée, ACIMOG, (§I-3.4), procède comme les méthodes PLS, par la recherche de composantes par tableaux qui soient le plus proche possible en terme de covariance. Trois types de déflations engendrent trois ACIMOG différentes. La première engendre des ensembles de composantes orthogonales par tableaux. La seconde et la troisième engendrent des super-composantes orthogonales par multi-tableaux. Ces deux dernières donnent une représentation compromis de chaque multi-tableaux. En les comparant, il est possible de voir où se situent les différences et les proximités de ces deux ensembles de tableaux.

Ces deux nouvelles méthodes ne sont pas vraiment en concurrence avec celles issues de la littérature, STATICO (§I-3.2) et les modèles MultiWay Multi-Block covariates Components (§I-3.1). En effet, STATICO étudie les configurations des variables par couple de tableaux et MWMBcovC ne s'applique qu'à un ensemble de cubes et cherche des modèles de type TUCKER ou PARAFAC pour chacun d'eux ainsi qu'un ensemble de super-composantes. Soulignons que cette méthode a énormément de paramètres et semble difficile à mettre en oeuvre. Alors que DO-ACT s'intéresse aux individus (au travers des opérateurs «WD») et l'ACIMOG cherche la proximité entre les variables des tableaux de deux ensembles n'ayant pas forcément la même longueur.

Le plus gros de cette thèse concerne la régression multi-tableaux. Dans le chapitre 5 de la première partie, nous avons proposé cinq méthodes (Analyse de Co-Inertie Multiple Orthogonale - PLS, Analyse de Co-Inertie Multiple Orthogonale Généralisée - PLS1, 2, 3, et REgression MUltri-Blocs) d'analyse factorielle permettant d'obtenir des modèles linéaires en les prédicteurs et dans la seconde partie, nous avons proposé deux méthodes (Splines Additives en Régression MUltri-Blocs 1 et 2) pour obtenir des modèles non-linéaires grâce à l'utilisation de transformations splines. Les ACIMOG-PLS font l'objet d'un article soumis pour publication dans une revue internationale et ont fait l'objet de présentations orales à des congrès (VIVIEN & SABATIER, 2001a,b, 2002).

Lorsque l'on veut expliquer une ensemble de tableaux réponses par un tableau prédicteur, l'ACIMO-PLS (VIVIEN & SABATIER, 2000, 2001c) est facile d'accès. En effet, celle-ci optimise un critère de covariances au carré mesurant la proximité entre des composantes de chaque tableau réponse et de la composante explicative. Nous avons montré au chapitre 5.5 que les modélisations obtenues étaient les mêmes que celles issues d'une régression PLS usuelle du tableau conjoint des réponses expliqué par les prédicteurs. Cela montre une certaine optimalité de la régression PLS. L'ACIMO-PLS fournit des objets supplémentaires permettant

l'étude plus précise et la comparaison de chaque tableau réponse. Dans le chapitre I-7.1 nous avons montré que lorsque $N = 1$, c'est-à-dire que l'on n'a qu'un seul tableau de réponses, les solutions d'ordre 1 sont égales aux solutions d'ordre 1 de la MBPLS. Lorsque le nombre N de tableaux à expliquer est supérieur à un, nous avons proposé les ACIMOG-PLS1, 2, 3 et REMUB comme alternative à la PLS Hiérarchique (HPLS, §I-5.3), la PLS en série (SPLS, I-5.4) et la MBPLS (§I-5.2). En effet, ces méthodes présentent toutes un ou plusieurs inconvénients. Le principal défaut que nous leur attribuons est leur manque d'optimalité. En effet, aucune n'a été définie par l'optimisation d'un critère à chaque étape. Il se trouve que dans le cas particulier où $N = 1$, MBPLS avec déflation sur la super-composante (WESTERHUIS & COENEGRACHT, 1997), est équivalente à une régression PLS usuelle (WESTERHUIS et al., 1998; WESTERHUIS & SMILDE, 2001; QIN et al., 2001, voir aussi §I-5.2 et I-7.1). Ce qui lui confère une certaine optimalité. Mais le critère de la régression PLS n'est en aucun cas ajusté par blocs de prédicteurs. Cela veut dire que les composantes t_k associées aux blocs de prédicteurs dans MBPLS ne sont pas recherchées en tenant compte de l'aspect «bloc» alors que le but des méthodes multi-blocs est principalement de considérer les prédicteurs par groupes. Dans le cas où $N > 1$, ce n'est plus pareil. Le principal inconvénient est que les tableaux de prédicteurs ne sont pas déflatés sur les mêmes composantes que les tableaux de réponses. Il est plus difficile d'interpréter les modèles, d'autant plus que les modèles partiels des variables réponses ne sont pas orthogonaux. Cependant, comme nous l'avons montré dans le chapitre I-5.2, MBPLS possède l'atout de consister en plusieurs régressions PLS «simultanées». Pour ce qui est de la méthode HPLS ((WOLD et al., 1996), §I-5.3), il a été prouvé par WESTERHUIS et al. (1998) et par l'application III-4, qu'elle n'était en général pas efficace dès les premières étapes. Les méthodes ACIMOG-PLS1, 2, 3 et REMUB ont le mérite d'optimiser un critère à chaque étape, de fournir soit des super-composantes explicatives orthogonales soit des composantes par tableaux explicatifs orthogonales et de construire des modèles se décomposant en somme de modèles partiels (associés à une étape a) orthogonaux. Ainsi, chaque nouvelle (super-)composante apporte une information nouvelle dans le modèle. De plus, l'orthogonalité des composantes permet d'obtenir des représentations graphiques correspondant à la projection des variables sur le plan engendré par deux composantes. La différence entre les quatre méthodes est en fait la manière de calculer les super-composantes.

L'ACIMOG-PLS1 va chercher, à chaque étape, des composantes explicatives et des composantes par tableau réponse qui soient le plus proche possible en terme de covariance. Les composantes explicatives d'ordre a sont orthogonales à toutes les composantes explicatives d'ordres inférieurs. Ceci implique une très forte reconstitution de la variance des tableaux réponses dès les premières étapes. La limite de cette méthode se situe au niveau de la taille des matrices de données : si le nombre d'individus est plus petit que le nombre de tableaux de prédicteurs, cette méthode ne sert à rien car les tableaux seront entièrement reconstitués à la première étape. De plus, il vaut mieux que le nombre d'observations soit plus grand que le nombre de tableaux de prédicteurs pour éviter les problèmes habituels de la régression multi-réponses ordinaire.

L'ACIMOG-PLS2 va chercher les composantes partielles telles que leur super-composante associée $t = \sum_{k=1}^K t_k$ et $u = \sum_{n=1}^N u_n$ soient le plus proche possible en terme de covariance. Les super-composantes explicatives sont orthogonales entre elles. La définition de la super-composante est telle que chaque tableau est considéré avec une importance égale. Aucune information apportée par un tableau n'est privilégiée ou négligée. Cependant, avec cette méthode nous avons rencontré des problèmes d'interprétation, en particulier lors de la recherche

de tableaux influents, que nous discutons un peu plus loin.

L'ACIMOG-PLS3 possède les mêmes solutions d'ordre 1 que les ACIMO-PLS1 et 2. Les super-composantes calculées ici tiennent compte de la contribution d'un tableau au critère optimisé. C'est-à-dire que la super-composante t tient plus compte des composantes partielles t_k qui sont les plus covariées avec les composantes u_n associées aux réponses. Ces pondérations étant toutes positives et n'étant pas orthogonales d'une étape à l'autre, elles ne peuvent pas servir à représenter une carte des tableaux prédicteurs. Mais, à chaque étape, elles sont un indicateur des tableaux les plus explicatifs en valeur absolue.

REMUB utilise un autre critère. Elle cherche à la fois les composantes par tableaux et les pondérations servant à construire les super-composantes à partir des composantes partielles, tels que la super-composante explicative et la super-composante réponse soient le plus proche en terme de covariance. Cela revient à chercher les composantes par blocs qui donnent le meilleur résumé de son multi-tableau dans le but d'expliquer l'un par l'autre. Les super-composantes explicatives sont orthogonales entre elles et sont utilisées pour construire les modèles. Les pondérations obtenues sont toutes positives. Nous sommes donc dans le même cas que l'ACIMOG-PLS3 où ces pondérations ne peuvent pas servir pour une carte des tableaux. Mais ils donnent une indication sur les tableaux les plus sollicités lors d'une étape. Nous avons montré que REMUB revient, à l'optimum, à effectuer simultanément la première étape de l'ACIMOG-PLS sur des tableaux particuliers et une régression PLS entre composantes. Cependant, les modèles obtenus sont généralement moins explicatifs que ceux des ACIMOG-PLS2 et 3 du fait du nombre de contraintes plus important.

Ces quatre méthodes sont obtenues par un algorithme itératif convergent.

Par rapport, aux modèles de MBPLS, rien ne permet de dire si une méthode est meilleure qu'une autre dans l'absolu. Il apparaît dans l'application en chimométrie (§III-4) que MBPLS a de bonnes performances. L'utilisateur doit choisir s'il veut privilégier l'interprétation, l'optimalité ou la performance du modèles. Selon les données, il est évident que certains modèles peuvent être plus adaptés que d'autres.

Pour aider à l'utilisation des méthodes multi-blocs, nous nous sommes intéressés à la sélection de tableaux prédicteurs, par analogie à la sélection de variables en régression multiple. Ce problème n'a jamais été abordé dans la littérature. Il nous paraît pertinent de s'y pencher. Contrairement à la régression multiple où le rajout d'une variable dans le modèle augmente la variance expliquée, le rajout d'un tableau dans l'étude multi-bloc peut faire décroître le taux de variance expliquée d'un tableau réponse. Il est donc nécessaire de choisir les tableaux les plus importants pour stabiliser les modèles avant d'envisager de faire de la prédiction. De plus, s'il s'avère que certains groupes de variables ne sont pas nécessaires, l'expérimentateur n'aura plus à les mesurer, ce qui peut être un gain de temps considérable. Nous avons donc proposé les critères BIP (Block Importance in the Prediction) pour les ACIMOG-PLS1 et 3 et REMUB, un BIP basé sur le coefficient RV pour l'ACIMOG-PLS1, ainsi que diverses utilisations du coefficient RV. Ces critères présentent quelques inconvénients (§I-6.3.2), et ont été utilisés dans les applications III-4, III-5 et III-3.

Au travers des exemples, nous avons constaté qu'il arrivait que les conclusions soient différentes selon les critères. De plus amples investigations dans ce domaine doivent être faites. De plus, il serait intéressant de mettre en place des tests pour tester la signification des différents coefficients RV.

D'autre part, il faudrait essayer de mettre en place un test de permutation pour vérifier la

validité du modèle ACIMOG-PLS (et/ou REMUB). Les premiers résultats obtenus (§I-6.2) semblent indiquer que la loi du «critère permuté» peut être estimé par un mélange de lois normales.

De plus, il reste à savoir si les ACIMOG-PLS et REMUB sont sensibles à des éventuelles corrélations entre tableaux de prédicteurs et à l'intérieur des tableaux. Des simulations devraient permettre cela en générant des échantillons issus de lois multi-normales. Par ailleurs, il faut aussi étudier l'«influence» du nombre de tableaux. Il paraît assez naturel de penser qu'il est plus difficile de bien modéliser (en terme de variance expliquée) un ensemble de tableaux réponses dont le nombre est élevé, qu'un tableau seul ou une unique variable réponse.

Outre la proposition de nouvelles méthodes, dans la première partie nous avons apporté de nouvelles démonstrations concernant l'ACOM, l'ACO, HPLS et NPLS.

Le travail effectué en régression multi-bloc linéaire, nous a permis d'entamer la recherche de méthodes de régression non-linéaire multi-blocs. Après une revue des méthodes non-linéaires PLS, nous avons choisi de généraliser à la fois l'ASPLS (DURAND & SABATIER, 1997, voir II-1.6) et les ACIMOG-PLS 1 et 2. Nous avons appelé ces méthodes Splines Additives en Régression Multi-Bloc 1 et 2 (SARMUB1, 2). La méthode consiste à rechercher à la fois des composantes par tableaux et des transformations splines optimales des prédicteurs, codés dans une base de B-spline. L'avantage des splines est qu'elles sont à support local. L'influence des valeurs extrêmes est donc moins importante sur le modèle. De plus, les B-splines, elles, sont faciles à calculer. SARMUB1 et 2 nécessitent un algorithme itératif, il est vrai, assez coûteux en temps de calcul. Nous n'avons encore pu résoudre un problème au niveau de l'algorithme de SARMUB2, qui cherche des super-composantes orthogonales. L'algorithme que nous avons pour l'instant ne converge pas dans tous les cas. Nous n'avons pas à l'heure actuelle pu tester SARMUB1 et 2 sur des données multi-tableaux pertinentes.

Dans SARMUB, nous avons considéré le nombre de noeuds et l'ordre des splines à utiliser fixés. Cependant, il est bien connu que la position des noeuds joue un rôle important dans la qualité de la modélisation. Il est envisageable d'adapter au cas multi-tableaux la méthode MBOK (MOLINARI, 2000) qui permet de chercher la meilleure position des noeuds en régression spline multivariée. Cependant, comme l'annonce son auteur, cette méthode est surtout intéressante lorsque la taille du tableau de données n'est pas trop importante, à cause du temps de calcul que nécessite cet algorithme. Nous pourrions tout de même chercher dans cette direction une méthode ou un algorithme pour choisir la meilleure position des noeuds pour chaque prédicteur.

Dans la troisième partie, nous avons appliqué nos méthodes dans divers domaines de recherche. Nous avons essayé de présenter les possibilités graphiques de ces méthodes et de les comparer numériquement avec d'autres. Les résultats obtenus par les applications que nous avons effectuées en collaboration avec des organismes de recherches ou industriels les ont intéressés, bien qu'en général les résultats ne soient que moyens (§III-2, III-3, III-4). Cela nous conforte dans l'idée que ces méthodes sont à développer. Notre travail avec le groupe ALTADIS continue et les résultats obtenus par les ACIMOG-PLS sont prometteurs. Leur présentation devraient être effectuée par un article dans une revue appliquée spécialisée dans l'étude des tabacs.

Le travail avec les données d'analyse sensorielle (§III-5) doit continuer pour essayer de peau-

finer les modèles au moyen de sélection de blocs explicatifs.

Il reste à tester les méthodes SARMUB sur des données multi-tableaux présentant réellement de la non-linéarité. A l'heure actuelle, nous n'avons pu trouver des données adaptées.

Les performances des méthodes d'analyse des données ne peuvent être jugées que par leur application. Nous n'avons finalement pas eu beaucoup d'expérience de nos méthodes et ne demandons qu'à parfaire cela. Nous pensons que les méthodes multi-blocs ont un grand intérêt pour les praticiens. Pour qu'elles leur soient accessibles, nous devons développer leurs côté graphique et pratique afin de faciliter leur interprétation. Pour cela, nous devons travailler ensemble, statisticiens et praticiens, afin de mieux comprendre leurs attentes et tenter d'y répondre.

Annexe A

La régression PLS

Nous donnons ici l'algorithme de la régression PLS et ses principales propriétés. Tous ces résultats peuvent se trouver dans (TENENHAUS, 1998).

La régression PLS est une méthode de régression linéaire d'une matrice Y sur une matrice X . Les tableaux de dimension $I \times M$ et $I \times J$, sont supposés centrés en colonne. Dans cette situation, elle connue sous le nom de PLS2. Lorsqu'il n'y a qu'une variable réponse y , elle est désignée PLS1.

Il s'agit de rechercher des A composantes $t_a = X^{(a-1)}\omega_a$ et $u_a = Y^{(a-1)}q_a$ résumant les tableaux X et Y dans le but de faire de la prédiction. Les tableaux $X^{(a-1)}$ et $Y^{(a-1)}$ désignent les déflations des tableaux initiaux $X = X^{(0)}$ et $Y = Y^{(0)}$.

La méthode originale a été présentée sous forme algorithmique, ensuite il a été montré qu'elle était un problème d'optimisation sous contraintes.

A.1 PLS : méthode algorithmique

La régression PLS1 a été présentée par WOLD (1966b) sous forme algorithmique. Celui-ci est de type NIPALS (WOLD, 1966a). La version multivariée (PLS2) est attribuée à WOLD, MARTENS & WOLD (1983).

L'algorithme :

étape 0 : initialisation

Centrer et réduire X et Y

étape 1

$$a = a + 1$$

$u_a = y_1$, première colonne de Y

étape 2

$$\omega_a = X'u_a/u'au_a$$

$$\omega_a = \omega/\|\omega_a\|$$

$$t_a = X\omega_a$$

$$q_a = Y't_a/t'_at_a$$

$$\spadesuit q_a = q_a/q'_aq_a$$

$$u_a = Yq_a/q'_aq_a$$

étape 3 : test de la convergence

si u_a à convergé en norme, on continue. Sinon on retourne en 2.

étape 4

$$p_a = X't_a/t_a't_a$$

$$\spadesuit b_a = t_a'u_a/t_a't_a$$

étape 5 : passage à l'étape suivante

$$X^{(a)} = X - t_ap_a'$$

$$Y^{(a)} = Y - b_at_aq_a'$$

et retourner en 1, avec $X = X^{(a)}$ et $Y = Y^{(a)}$.

La ligne précédée de \spadesuit n'est pas toujours effectuée. Cela dépend des auteurs. Elle permet de normer le vecteur de pour q_a . Si q_a est normé à l'unité, $u_a = Yq_a$.

Les divisions par $u_a'u_a$ et $t_a't_a$ dans les mises à jour de ω_a et c_a sont conçues pour le cas où il y a des valeurs manquantes. Elles ne sont pas utiles en général puisqu'on norme les vecteurs de poids à l'étape suivant cette mise à jour.

On a $t_ap_a' = P_{t_a}X^{(a-1)}$ et $b_at_aq_a' = P_{t_a}Y^{(a-1)}$.

Le coefficient b_a est le coefficient de la régression simple de u_a sur t_a . Si le vecteur q_a n'est pas normé à l'unité dans l'algorithme, $b_a = 1$. Ainsi, l'algorithme PLS sous-entend une régression simple de u_a sur t_a : $\hat{u}_a = b_at_a$. Cette relation est appelée la «relation interne». Si la qualité de cette régression est bonne, cela signifie qu'il y a bien une relation linéaire entre les variables.

A.2 PLS : une optimisation

La régression PLS peut être définie comme le problème d'optimisation sous contraintes à chaque étape a suivant :

$$\max_{\omega, q} \text{cov}(t, u) = \text{cov}(X\omega, Yq)$$

avec $\|\omega\|^2 = 1 = \|q\|^2$.

Pour l'étape a , il suffit de remplacer X et Y par $X^{(a-1)}$ et $Y^{(a-1)}$.

Les solutions d'ordre a vérifient les équations* :

$$\lambda_a^2 \omega_a = X^{(a-1)'} Y^{(a-1)} Y^{(a-1)'} X^{(a-1)} \omega_a \quad (\text{A.1})$$

$$\lambda_a^2 q_a = Y^{(a-1)'} X^{(a-1)} X^{(a-1)'} Y^{(a-1)} q_a \quad (\text{A.2})$$

où $\lambda_a = \text{cov}(t_a, u_a)$. Ainsi, les vecteurs de poids ω_a et q_a sont les vecteurs propres de norme unité associée à la plus grande valeur propre commune λ_a^2 des matrices définies dans les équations ci-dessus.

De plus, $\lambda_a = \|X^{(a-1)'} u_a\| = \|Y^{(a-1)'} t_a\|$. Il est alors aisé de vérifier l'équivalence des solutions obtenues par ces problèmes aux valeurs propres et l'algorithme défini plus haut, dans le cas où il n'y a pas de valeurs manquantes dans les données.

Il est possible d'utiliser des métriques Q^x et Q^y dans les espaces des individus \mathbb{R}^J et \mathbb{R}^M , et une métrique D dans l'espace des variables \mathbb{R}^I . Alors tous les produits scalaires dans \mathbb{R}^I s'effectuent au sens de D et $t_a = X^{(a-1)} Q^x \omega_a$, $u_a = Y^{(a-1)} Q^y q_a$ et les vecteurs sont de Q^x - et Q^y -norme unité.

* Elles peuvent s'obtenir grâce à la méthode des multiplicateurs de Lagrange où $\lambda_a/2$ est le multiplicateur associé à une des contraintes de l'étape a , $\|\omega_a\|^2 - 1 = 0$, par exemple.

A.3 Commentaires

La déflation du tableau Y peut être omise sans affecter les composantes t_a , et donc les modèles. De même DAYAL & MACGREGOR (1997) ont montré que la déflation de X peut ne pas être effectuée si Y est déflaté. Les différents objets ω_a , u_a et q_a sont identiques à ceux de la régression PLS utilisant les deux déflations et les composantes t_a sont obtenues par orthogonalisation à chaque étape.

En régression PLS1 et 2, il est possible de montrer les propriétés suivantes :

$$\begin{aligned} t'_\ell t_a &= 0, & \ell < a \\ \omega'_a p_a &= 1 \\ \omega'_\ell X^{(a)'} &= 0, & \ell \leq a \\ \omega'_\ell p_a &= 0, & \ell < a \\ \omega'_a \omega_\ell &= 0 \\ t'_\ell X^{(a)} &= 0, & \ell \leq a \end{aligned}$$

La composante $t_a = X^{(a-1)'} \omega_a$ peut s'écrire comme combinaison linéaire des variables initiales : $t_a = X \omega_a^*$, où $\omega_a^* = \prod_{\ell=1}^{a-1} (Id_J - \omega_\ell p'_\ell) \omega_a$, et $\omega_1^* = \omega_1$.

On peut aussi écrire la formule de récurrence (voir par exemple DURAND et al., 1998) $\omega_a^* = \left(Id_J - \sum_{\ell=1}^{a-1} \frac{\omega_\ell^* \omega_\ell^{*'}}{\|t_\ell\|^2} X' X \right) \omega_a$.

Le modèle d'ordre A pour Y s'écrit $Y = \sum_{a=1}^A P_{t_a} Y^{(a-1)} + Y^{(A)}$. Les modèles partiels de rang 1 $P_{t_a} Y^{(a-1)} = P_{t_a} Y$ sont orthogonaux et expliquent une part différente de la variance de Y à chaque étape. L'erreur d'ajustement $Y^{(A)}$ est orthogonale à $\hat{Y}^A = \sum_{a=1}^A P_{t_a} Y$. \hat{Y}^A peut aussi s'écrire de la forme $X \hat{\beta}^A$, où les coefficients $\hat{\beta}^A = \sum_{a=1}^A \frac{\omega_a^* \omega_a^{*'}}{\|t_a\|^2} X' Y$.

Lorsque $A = rg(X)$, le modèle PLS est égal au modèle de régression multi-réponses ordinaire (DE JONG, 1995; TENENHAUS, 1998).

En notant $W_a^* = [\omega_1^* \dots \omega_a^*]$, $W_a = [\omega_1 \dots \omega_a]$, $P_a = [p_1 \dots p_a]$, $Q_a = [q_1 \dots q_a]$, $T_a = [t_1 \dots t_a]$, il est possible de démontrer que :

$$\begin{aligned} T_a &= X W_a (P'_a W_a)^{-1} \\ \hat{\beta}^a &= W_a (P'_a W_a)^{-1} Q'_a \\ &= W_a^* Q'_a. \end{aligned}$$

Dans le cas de PLS1, il suffit de remplacer u_a par $y^{(a-1)}$ pour obtenir les solutions. La régression PLS1 possède des propriétés particulières discutées dans (TENENHAUS, 1998).

Il y a plusieurs expressions de la linéarité en régression PLS2. La relation interne, $\hat{u}_a = b_a t_a$, la relation externe, $t_a = X \omega_a^*$ et le modèle linéaire $\hat{Y}^A = X \hat{\beta}^A$.

Annexe B

Rappels mathématiques

B.1 La méthode des multiplicateurs de Lagrange

La définition que nous proposons est issue de (MAGNUS & NEUDECKER, 1988, page 131). La méthode des multiplicateurs de Lagrange est due au théorème de Lagrange qui donne une condition nécessaire pour l'existence d'un minimum c d'une fonction ϕ sous les contraintes $g(c) = 0$:

Theorème B.1 Soit $g : S \rightarrow \mathbb{R}^m$ un fonction vectorielle définie sur un ensemble S de \mathbb{R}^n , avec $n > m$, et soit c un point intérieur à S . Supposons :

- (i) $g(c) = 0$,
 - (ii) g est différentiable dans une n -boule $B(c)$,
 - (iii) la matrice jacobienne $\mathbf{D}g$, $m \times n$, de g , est continue en c , et
 - (iv) $\mathbf{D}g(c)$ est de plein rang ligne m .
- De plus, soit $\phi : S \rightarrow \mathbb{R}$, une fonction réelle définie sur S et supposons que
- (v) ϕ est différentiable en c ,
 - (vi) $\phi(x) \geq \phi(c)$ pour tout $x \in B(c)$ tel que $g(x) = 0$ (c est un minimum).

Alors, il existe un unique vecteur λ de \mathbb{R}^m satisfaisant les n équations : $\mathbf{D}\phi(c) - \lambda' \mathbf{D}g(c) = 0$.

La condition (vi) peut être remplacée par $\phi(x) \leq \phi(c)$ pour tout $x \in B(c)$ tel que $g(x) = 0$, c est alors un maximum.

λ est le vecteur des multiplicateurs de Lagrange associés aux contraintes définies par g . En pratique, on commence par définir le Lagrangien $\mathcal{L}(x) = \mathbf{D}\phi(x) - \lambda' \mathbf{D}g(x)$, où λ est constant. Puis on dérive \mathcal{L} par rapport à x et on pose les résultats de ces dérivations à 0. Ces équations définissent, avec les m contraintes un système de $m + n$ équations. Ce sont les équations normales. λ et x sont déterminés par leur résolution.

Le vecteur $x = c$ obtenu définit les points critiques de la surface définie par les contraintes. Il faut ensuite déterminer si ce sont des points selles, maxima ou minima, par exemple en étudiant la positivité de la matrice Hessienne de ϕ en c .

Lorsque ϕ est une fonction matricielle à valeur dans \mathbb{R} ($\phi : S \in \mathbb{R}^{n \times q} \rightarrow \mathbb{R}$) et $G : S \rightarrow \mathbb{R}^{m \times q}$ une fonction matricielle définissant les contraintes, la méthode est la même mais fait intervenir les dérivations de matrices. Les multiplicateurs de Lagrange définissent une matrice Γ et le Lagrangien s'écrit $\mathcal{L}(X) = \phi(X) - \text{tr}(\Gamma' G(X))$.

B.2 Les splines de régression

Cette rapide description des fonctions splines a été réalisée en majorité à partir de (DE BOOR, 1978) et (DURAND, 1997).

Une fonction spline est une fonction polynômiale par morceaux où ces «morceaux» sont reliés en des points appelés noeuds. Il existe plusieurs de fonctions splines dont les plus connues sont les splines de régression et les splines de lissage. Les splines de lissages sont ce que l'on appelle des «lisseurs» («smoothers») utilisé en estimation fonctionnelle non-paramétrique. Le degré de lissité de la spline est contrôlé par un réel positif, alors que les splines de régression sont contrôlées par le nombre et la positions des noeuds. Nous ne parlerons ici que des splines de régression. Pour une description des splines de lissage, se référer à (DE BOOR, 1978) et (BESSE & THOMAS-AGNAN, 1989) pour une revue bibliographique en français.

Une fonction spline $s(\cdot)$, utilisée pour transformer un vecteur x , polynôme par morceaux, est définie par leur degré d et L noeuds distincts, ξ_ℓ , en lesquels les polynômes sont raccordés avec un certain degré de lissité, contrôlé par l'ordre de multiplicité des noeuds. Nous ne considérerons ici que des noeuds simples (multiplicité 1), qui rend la fonction spline $s(\cdot)$ continue et deux fois dérivable. Ainsi, nous sommes en présence de L noeuds distincts de multiplicité 1 et $s(x)$ appartient à l'espace des fonctions splines polynômiales de degré d associées aux noeuds ξ_ℓ , noté $\mathcal{S}(\xi_1, \dots, \xi_\ell)$, de dimension $d + 1 + L$.

$s(x)$ peut être exprimée dans une base $\{B_r(x)\}_{r=1, \dots, 1+d+L}$ et s'écrit alors

$$s(x) = \sum_{r=1}^{1+d+L} \alpha_r B_r(x).$$

Le vecteur $\tilde{a} = (\alpha'_1, \dots, \alpha'_{1+d+L})'$ est le vecteur des coefficients splines. Ils déterminent $s(x)$. On appelle matrice de codage de $x \in \mathbb{R}^1$ la matrice de dimension $1 \times (1 + d + L)$, $B = [B_1(x) \ \dots \ B_{1+d+L}(x)]$. B dépend implicitement des noeuds ξ_ℓ . Ainsi, $S = s(x) = B\tilde{a}$. Si l'on veut ajuster $y \in \mathbb{R}^1$ par $s(x)$ et que les noeuds sont fixés, alors \tilde{a} est déterminé par les moindres carrés : $\tilde{a} = (B'B)^+ B'y$.

Une première base de $\mathcal{S}(\xi_1, \dots, \xi_\ell)$ est la base des *puissances tronquées* :

$$\{B_r(x)\}_{r=1, \dots, 1+d+L} = \{1, x, \dots, x^d, (x - \xi_1)_+^d, \dots, (x - \xi_L)_+^d\},$$

où $(x - \xi_\ell)_+ = x - \xi_\ell$ si $x \geq \xi_\ell$, 0 sinon. Cette base n'est pas la plus utilisée en pratique car elle pose des problèmes d'inversibilité de matrices.

Une seconde base est la base des *B-splines*. Pour la définir, nous devons définir la partition étendue des noeuds. Posons $\xi_0 = a$, la plus petite coordonnée de x , et $\xi_{L+1} = b$ la plus grosse coordonnée de x . La partition étendue contient $2(d+1)+L$ noeuds τ_k , avec $\tau_1 = \dots = \tau_{d+1} = a$, $\tau_{L+(d+1)+1} = \dots = \tau_{L+2(d+1)} = b$ et $\tau_{d+1+\ell} = \xi_\ell$. L'ensemble des fonctions, appelées *B-splines*, $\{B_r^{d+1}\}_{r=1, \dots, 1+d+L}$ définies par :

$$B_r^{d+1}(x) = (-1)^{d+1} (\tau_{r+d+1} - \tau_r) [\tau_r, \dots, \tau_{r+d+1}] (t - \tau)_+^d$$

forme une base. La fonction réelle $(z_1, \dots, z_p) \rightarrow [z_1, \dots, z_p]f$ est la fonction «différences divisées» :

$$[z_1, \dots, z_p]f = \frac{[z_2, \dots, z_p]f - [z_1, \dots, z_{p-1}]f}{z_p - z_1} = \sum_{i=1}^p \frac{f(z_i)}{\prod_{j=1, j \neq i}^p (z_i - z_j)}.$$

Outre le fait qu'elles soient conditionnées et faciles à calculer (par des formules de récurrence (DE BOOR, 1978)), les B-splines sont attractives pour leurs propriétés :

- support local : $B_r^{d+1}(x) = 0$ si x est en dehors de l'intervalle $[\tau_r, \tau_{r+d+1}]$. Une mesure n'influe que sur $d + 1$ fonctions splines de base.
- $\sum_{r=1}^{1+d+L} B_r^{d+1}(x) = 1$, pour $x \in [a, b]$ et $B_r^{d+1}(x) \geq 0$ pour $x \in [\tau_r, \tau_{r+d}]$. Elles forment une partition de l'unité.
- Pour $d > 0$, il existe une spline identité définie par $x = \sum_{r=1}^{1+d+L} \beta_r B_r^{d+1}(x)$ avec $\beta_r = \frac{1}{d} \sum_{k=1}^d \tau_{r+k}$. Les coefficients β_r sont les *valeurs nodales*.

En analyse multivariée, si l'on est amené à coder une matrice X , de dimension $l \times J$, on choisit une séquence de noeuds et leur position pour chaque variable x_j , colonne j de X , puis on les code dans une base de splines. A chaque variable x_j est associée une matrice de codage B_j de dimension $l \times (1 + d_j + L_j)$, et une transformation spline de x_j est notée $S_j = B_j a_j$. La matrice $B = [B_1 \ \dots \ B_J]$ est une matrice de codage de X . Les coefficients de la régression spline multi-variables aux moindres carrés de Y sur X sont donnés par $\hat{\beta} = (B'B)^+ B'Y$.

Annexe C

A propos des représentations graphiques ou la perte de la dualité

En Analyse en Composantes Principales, les représentations graphiques sont dues au schéma de dualité bien connu (CAILLIEZ & PAGÈS, 1976). Notons a_m l'axe principal d'ordre m de Q -norme unité associé à la matrice X centrée pour D ($m^{\text{ième}}$ vecteur propre de $X'DXQ$ associé à la $m^{\text{ième}}$ valeur propre $\lambda_m = \text{var}(t_m)$) et $t_m = XQa_m$ la composante principale associée.

En ACP, les axes a_m sont Q -orthonormés dans \mathbb{R}^J . Ainsi, deux axes a_m et a_ℓ définissent un repère orthonormé de \mathbb{R}^J dans lequel on peut projeter les I individus de X . Alors, les I points projetés ont pour coordonnées, dans ce repère, les coordonnées des composantes principales t_a et t_ℓ . En effet, en notant $X(i)$ la $i^{\text{ième}}$ ligne de X ,

$$\begin{aligned} P_{a_m, a_\ell} X'(i) &= P_{a_m} X'(i) + P_{a_\ell} X'(i) \\ &= a_m a_m' Q X'(i) + a_\ell a_\ell' Q X'(i) = t_m(i) a_m + t_\ell(i) a_\ell. \end{aligned}$$

Ceci justifie la représentation graphique des individus effectuée en ACP.

Les composantes t_a sont D -orthogonales dans \mathbb{R}^I . Ainsi, les deux composantes de norme unité $t_m/\|t_m\|_D$ et $t_\ell/\|t_\ell\|_D$ forment un repère orthonormé de \mathbb{R}^I dans lequel on peut projeter les J variables de X . Les coordonnées des J points projetés sont alors données par les corrélations (si les données sont réduites en colonnes) ou par les covariances (si les données ne sont pas réduites) entre les variables x^j de X et les composantes $t_m/\|t_m\|_D$ et $t_\ell/\|t_\ell\|_D$. En effet, en supposant X réduite en colonne,

$$\begin{aligned} P_{t_m, t_\ell} x^j &= t_m t_m' D x^j / \|t_m\|_D^2 + t_\ell t_\ell' D x^j / \|t_\ell\|_D^2 \\ &= \frac{t_m}{\|t_m\|_D} r(x^j, t_m) + \frac{t_\ell}{\|t_\ell\|_D} r(x^j, t_\ell) \\ &= \frac{t_m}{\|t_m\|_D} \sqrt{\lambda_m} a_m(j) + \frac{t_\ell}{\|t_\ell\|_D} \sqrt{\lambda_\ell} a_\ell(j). \end{aligned}$$

Ainsi, l'orthogonalité des axes principaux d'une part, et l'orthogonalité des composantes principales d'autre part, permettent la projection des individus et variables initiales dans des repères orthonormés et les bonnes propriétés de l'ACP (une diagonalisation pour obtenir toutes les composantes) font que les coordonnées de ces projections sont fournies par les composantes principales et les axes principaux.

Mais, ceci ne marche plus en régression PLS ni en analyse multi-tableaux.

C.1 En régression PLS

Que se passe-t-il en régression PLS? Notons ω_a l'axe de norme unité d'ordre a de la régression PLS associé à X et $t_a = X^{(a-1)}Q^x\omega_a$ la composante PLS explicative. Les axes ω_a sont Q -orthonormés dans \mathbb{R}^J (voir l'annexe A) et les composantes t_a sont D -orthogonales dans \mathbb{R}^I . En procédant comme en ACP, les coordonnées des J variables projetées dans le repère orthonormé $t_a/\|t_a\|_D$ et $t_\ell/\|t_\ell\|_D$ sont données par les corrélations (ou covariance) entre les variables déflatées (ou initiales) et les composantes de norme unité :

$$\begin{aligned}
 P_{t_a, t_\ell} x^j &= t_a t'_a D x^j / \|t_a\|_D^2 + t_\ell t'_\ell D x^j / \|t_\ell\|_D^2 \\
 &= \frac{t_a}{\|t_a\|_D} r(x^j, t_a) + \frac{t_\ell}{\|t_\ell\|_D} r(x^j, t_\ell) \\
 &= \frac{t_a}{\|t_a\|_D} t'_a D(x^j - P_{t_1, \dots, t_{a-1}} x^j) / \|t_a\|_D + \frac{t_\ell}{\|t_\ell\|_D} t'_\ell D(x^j - P_{t_1, \dots, t_{\ell-1}} x^j) / \|t_\ell\|_D \\
 &= \frac{t_a}{\|t_a\|_D} t'_a D x^{j(a-1)} / \|t_a\|_D + \frac{t_\ell}{\|t_\ell\|_D} t'_\ell D x^{j(\ell-1)} / \|t_\ell\|_D \\
 &= \frac{t_a}{\|t_a\|_D} r(x^{j(a-1)}, t_a) + \frac{t_\ell}{\|t_\ell\|_D} r(x^{j(\ell-1)}, t_\ell).
 \end{aligned}$$

Comme $\omega_a = X^{(a-1)'} D u_a / \|X' D u_a\|$, les coordonnées des vecteurs ω_a (et celles des ω_a^* tels que $t_a = X \omega_a^*$) ne correspondent pas aux coordonnées des variables projetées dans le repère orthonormé formé par les composantes, contrairement à l'ACP. Dès qu'un tableau est ajouté dans le schéma de l'analyse, la dualité est affectée. Par contre, les variables réponses projetées dans ce repère ($t_a/\|t_a\|_D, t_\ell/\|t_\ell\|_D$) ont des coordonnées proportionnelles aux vecteurs de poids q_a (où $u_a = Y Q^y q_a$) et q_ℓ associés aux réponses. En effet, les coordonnées des M variables réponses projetées sur la composante $t_a/\|t_a\|_D$ sont données par $Y' D t_a / \|t_a\|_D = \|t'_a D Y\|_{Q^y} q_a / \|t_a\|_D$. Ainsi, les coordonnées des vecteurs q_a et q_ℓ fournissent une représentation dilatée des projections des variables réponses.

De plus, à cause de la déflation des variables explicatives X dans l'algorithme PLS, les coordonnées des individus projetés dans le repère Q^x -orthonormé (ω_a, ω_ℓ) de \mathbb{R}^J ne sont pas données par les coordonnées des composantes associées. En effet, pour un individu i , on a $P_{\omega_a, \omega_\ell} X'(i) = \omega_a \omega'_a Q^x X'(i) + \omega_\ell \omega'_\ell Q^x X'(i)$. Pour la première dimension $a = 1$, on a bien $t_1(i) = X(i) Q \omega_1$, mais pas pour les dimensions suivantes, à cause de la déflation de X sur les composantes. On a $\omega'_\ell Q^x X'(i) = \omega'_\ell Q^x (X^{(\ell-1)'})'(i) + X' P'_{t_1, \dots, t_{\ell-1}}(i) = t'_\ell(i) + \omega'_\ell Q^x X' P'_{t_1, \dots, t_{\ell-1}}(i)$. Ainsi, la carte des individus effectuée usuellement en régression PLS est «biaisée», et d'autant plus qu'on utilise des composantes d'ordre élevé.

Il a été montré que l'on n'est pas obligé de déflater le tableau X (DAYAL & MACGREGOR, 1997) si l'on déflate le tableau Y . Alors les axes ω_a obtenus sont les mêmes et les modèles obtenus aussi. On obtient les composantes t_a de PLS en les orthogonalisant à chaque étape* et un ensemble de composantes $t_a^* = X Q^x \omega_a$ qui ne sont pas D -orthogonales mais qui engendrent le même espace que les t_a . Ainsi, les coordonnées des I points-individus projetés dans le repère Q^x -orthonormé (ω_a, ω_ℓ) correspondent aux coordonnées des composantes t_a^* et t_ℓ^* .

La carte des individus en PLS est effectuée par analogie à l'ACP, mais en réalité elle est «biaisée». Il est impossible d'avoir à la fois une carte des individus et une carte des variables «non biaisée» obtenues de manière simple. On peut se demander s'il ne faut pas utiliser à la

* C'est sur ces composantes que le tableau Y est déflaté.

fois t_a et t_a^* , les deux ensembles de composantes engendrant le même espace. La représentation des individus de X en tant que points projetés doit être réalisée grâce aux coordonnées des vecteurs $t_a + P_{t_1, \dots, t_{a-1}} X Q^x \omega_a$.

C.2 En analyse multi-tableaux

En analyse multi-tableaux, il y a les mêmes problèmes qu'en régression PLS, car nous effectuons des déflations et parce qu'il y a plusieurs tableaux à analyser. Les problèmes sont encore plus importants lorsque les déflations sont effectuées sur des super-composantes et non pas sur la composante de chaque tableau.

Considérons les méthodes ACIMOG, ACIMOG-PLS, ACIMO, ACIMO-PLS et REMUB. Dans ces méthodes, à chaque tableau sont associés un axe ($a_{k,a}$ pour X_k , $b_{n,a}$ pour Y_n) et une composante ($t_{k,a} = X_k^{(a-1)} Q_k^x a_{k,a}$ et $u_{n,a} = Y_n^{(a-1)} Q_n^y b_{n,a}$). Dans le cas de l'ACIMO-PLS, où il n'y a qu'un seul tableau explicatif X il suffit de supprimer l'indice k pour retrouver les notations du chapitre I-5.5.

De plus, on calcule des super-composantes, t_a et u_a , qui, d'une manière générale, s'écrivent $t_a = X^{(a-1)} \text{Diag}(Q_k^x) a_a$ et $u_a = Y^{(a-1)} Q_n^y b_a$. La définition des vecteurs a_a et b_a de $\mathbb{R}^{\sum_{k=1}^K J_k}$ et $\mathbb{R}^{\sum_{n=1}^N M_n}$ dépend de la méthode.

Dans chaque méthode, on engendre soit des composantes par blocs D -orthogonales soit des super-composantes D -orthogonales, grâce à la déflation choisie. Lorsque la déflation est faite par bloc, les axes associés sont orthogonaux par bloc. Lorsque la déflation est faite sur une super-composante, en général, ni les axes a_a , ni b_a ne sont orthogonaux, sauf cas particulier. Ainsi, dans l'ACIMOG1 et l'ACIMO les composantes sont orthogonales par blocs ainsi que les axes. Dans l'ACIMOG-PLS1 et l'ACIMO-PLS les composantes explicatives sont orthogonales par bloc, ainsi que les axes. Dans l'ACIMOG2 et 3, les super-composantes sont orthogonales, et enfin dans l'ACIMOG-PLS 2 et 3 et REMUB, les super-composantes explicatives sont orthogonales.

- Si la déflation est faite par bloc, alors, par exemple, les composantes $\{t_{k,a}\}_a$ forment un système D -orthogonal de \mathbb{R}^l .

Alors, le système d'axe $\{a_{k,a}\}_a$ est Q_k^x -orthogonal dans \mathbb{R}^{J_k} . En raisonnant comme dans PLS, les coordonnées des points-individus projetés dans le repère Q_k^x -orthogonal de \mathbb{R}^{J_k} ne sont pas données par les coordonnées des composantes $t_{k,a}$ sauf pour $a = 1$. Les représentations graphiques sont «biaisées». Cependant, c'est la seule manière simple possible de procéder dans bien des cas.

Par contre, les coordonnées des projections des variables de X_k dans le repère orthonormé ($t_{k,a}/\|t_{k,a}\|_D, t_{k,\ell}/\|t_{k,\ell}\|_D$) de \mathbb{R}^l sont données par les corrélations (ou les covariances) entre les variables de X_k et les composantes normées à l'unité. Ce ne sont pas *a priori* les coordonnées des axes $a_{k,a}$ car $a_{k,a}$ est proportionnel à $X_k^{(a-1)'} D \sum_{n=1}^N u_{n,a}$ [†].

Dans l'ACIMOG-PLS1, avec une seule variable réponse y à expliquer, si les composantes sont bien explicatives de y , on peut s'attendre à ce que les coordonnées des vecteurs $a_{k,a}$ donnent une configuration des variables similaire à celle donnée par les corrélations.

- Si la déflation est faite sur les super-composantes, alors celles-ci, par exemple, t_a forment un système D -orthogonal de \mathbb{R}^l . Les coordonnées des points-variables projetés dans le re-

[†] Dans REMUB, il y a des pondérations à rajouter.

père $(t_a/\|t_a\|_D, t_\ell/\|t_\ell\|_D)$ sont données par les corrélations des variables de X avec les super-composantes de norme unité. Ce ne sont pas les coordonnées des axes a_a , ni de $a_{k,a}$. Dans le cas de la régression, les réponses sont aussi déflatées sur les t_a , et les coordonnées des points-variables projetés sont obtenues de la même manière et elles sont aussi proportionnelles aux coordonnées des axes $b_{n,a}$ dans l'ACIMOG-PLS2. Le graphique donné par les coordonnées des vecteurs $b_{n,a}$ donne une image dilatée des variables réponses projetées dans le repère engendré par deux super-composantes explicatives normées.

Dans l'ACIMOG-PLS3, on a en plus l'orthogonalité des axes a_a de $\mathbb{R}^{\sum_{k=1}^K J_k}$ si $Q_k^x = Id_{J_k}$. Mais on retrouve le même problème que dans la régression PLS, à cause de la déflation des X_k : les coordonnées des points individus dans le repère orthonormé de $\mathbb{R}^{\sum_{k=1}^K J_k}$ (a_a, a_ℓ) ne sont pas les coordonnées des super-composantes.

Dans les autres méthodes, les axes a_a n'étant pas orthogonaux, nous ne connaissons pas de repère orthonormé de $\mathbb{R}^{\sum_{k=1}^K J_k}$ dans lequel nous pourrions projeter simplement les individus.

Dans les méthodes d'analyse conjointe de tableaux telles que l'ACIMO ou l'ACIMOG, les mêmes raisonnements peuvent être faits pour les variables Y_n en utilisant soit les composantes $u_{n,a}$, soit les super-composantes u_a et les axes associés.

C.3 Pour conclure

Pour conclure, dans le cas de la régression multi-variables (telle que PLS) ou multi-blocs, le schéma de dualité de l'ACP ne tient plus. L'orthogonalité des composantes nous permet d'étudier les variables grâce aux corrélations. Mais les axes associés ne sont plus *a priori* les coordonnées des projections des variables dans le repère défini par un couple de composantes. De plus, toutes les cartes des individus effectuées grâce aux composantes sont «biaisées». Il faut donc interpréter la proximité entre les points avec précaution, en particulier pour les dimensions supérieures à 1.

Dans le cas de la régression, pour faire des graphiques simplement, on peut envisager de ne pas déflater les prédicteurs, puis orthogonaliser les composantes explicatives afin d'obtenir une base. Cependant, il semble étrange d'utiliser les composantes orthogonalisées pour construire les modèles et interpréter les composantes non orthogonales...

Il faut sinon corriger les graphiques afin d'avoir les «vraies» coordonnées des individus projetés.

D'autre part, on peut constater que si les déflations sont faites par blocs (et non sur une super-composante) on obtient des axes associés orthogonaux.

Dans les méthodes de régression multi-blocs de la littérature, le problème est analogue pour la construction d'une carte des individus par tableau ou «globaux». Tout le monde, par analogie à l'ACP, utilise les coordonnées des (super-)composantes pour les effectuer. Nous faisons de même tout en admettant qu'il y a un problème de «biais» dans ce type de représentations, qui sont les plus simples. Par la suite, nous corrigerons les représentations des individus.

Bibliographie

- ANDERSSON, C.A. (1999) – Direct orthogonalization – *Chemometrics and Intelligent Laboratory Systems*, 47, 51–63.
- ANDERSSON, C.A. & BRO, R. (1998) – Improving the speed of multiway algorithms: Part I. Tucker3 – *Chemometrics and Intelligent Laboratory Systems*, 42, 93–103.
- BAFFI, G., MARTIN, E.B., & MORRIS, A.J. (1999a) – Non-linear projection to latent structure revisited: the quadratic PLS algorithm – *Computers & Chemical Engineering*, 23, 395–411.
- BAFFI, G., MARTIN, E.B., & MORRIS, A.J. (1999b) – Non-linear projection to latent structure revisited (the neural networks PLS algorithm) – *Computers & Chemical Engineering*, 23, 1293–1307.
- BAFFI, G., MARTIN, E.B., & MORRIS, A.J. (2000) – Non-linear dynamic projection to latent structures modelling – *Chemometrics and Intelligent Laboratory Systems*, 52, 5–22.
- BERGLUND, A., KETTANEH-WOLD, N., UPPGÅRD, L.-L., WOLD, S., BENDWELL, N., & CAMERON, D.R. (2001) – The GIFI approach to non-linear PLS modeling – *Journal of Chemometrics*, 15, 321–336.
- BERGLUND, A. & WOLD, S. (1997) – INLR, implicit non-linear latent variable regression – *Journal of Chemometrics*, 11, 141–156.
- BERGLUND, A. & WOLD, S. (1999) – A serial extension of multiblock PLS – *Journal of Chemometrics*, 13, 461–471.
- BESSE, Ph. & THOMAS-AGNAN, C. (1989) – Le lissage par fonctions splines en statistique. Revue bibliographique – *Statistique et Analyse des Données*, 14, 55–84.
- BRO, R. (1996) – Multiway calibration. Multilinear PLS – *Journal of Chemometrics*, 10, 47–61.
- BRO, R. (1997) – PARAFAC. Tutorial and applications – *Chemometrics and Intelligent Laboratory Systems*, 38, 149–171.
- BRO, R. (1998) – *Multi-way Analysis in the Food Industry. Models, algorithms and applications* – Thèse de Doctorat, University of Amsterdam, the Netherlands. 300p.
- CAILLIEZ, F. & PAGÈS, J.-P. (1976) – *Introduction à l'analyse des données* – SMASH, Paris – 616p.
- CARROLL, J.-D. (1968) – A generalization of canonical correlation analysis to three or more sets or variables. *Proceedings of the 76th convention of the American Psychological Association*, number 3, (pp. 227–228).
- CARROLL, J.D. & CHANG, J.J. (1970) – Analysis of individual differences in multidimensional scaling via an n-way generalization of Eckart-Young decomposition – *Psychometrika*, 35, 283–319.
- CASIN, Ph. (1996) – L'analyse en composantes principales généralisée – *Revue de Statistique Appliquée*, XLIV(3), 63–81.
- CASIN, Ph. (2001) – A generalization of principal component analysis to K sets of variables – *Computational Statistics & Data Analysis*, 35, 417–428.

- CATTELL, R.B. (1944) – Parallell proportional profiles and other principles for determining the choice of factors by rotation – *Psychometrika*, 9, 267–283.
- CHATELIN, Y.-M., VINZI, V.E., & TENENHAUS, M. (2002) – *State-of-art on PLS Path Modeling through the available software* – Rapport technique n°CR764, HEC, Paris.
- CHESEL, D. & HANAFI, M. (1996) – Analyses de la co-inertie de K nuages de points – *Revue de Statistique Appliquée*, XLIV(2), 35–60.
- CHESEL, D. & MERCIER, P. (1993) – Couplage de triplets statistiques et liaison espèces - environnement. Lebreton, J.-D. & Asselain, B. (édité par), *Biométrie et Environnement*, (pp. 15–44)., Paris. Masson.
- CLEMENTI, S., CRUCIANI, G., CURTI, G., & SKAGERBERG, B. (1989) – PLS response surface optimization: the Carso procedure – *Journal of Chemometrics*, 3, 499–509.
- DAYAL, B.S. & MACGREGOR, J.F. (1997) – Improved PLS algorithm – *Journal of Chemometrics*, 11(73-85).
- DE BOOR, C. (1978) – *A practical guide to splines* – Springer-Verlag.
- DE JONG, S. (1995) – PLS shrinks – *Journal of Chemometrics*, 9, 323–326.
- DE JONG, S. (1998) – Short communication regression coefficients in multilinear pls – *Journal of Chemometrics*, 12, 77–81.
- DE JONG, S. & KIERS, H.A.L. (1992) – Principal covariates regression. part I. theory – *Chemometrics and Intelligent Laboratory Systems*, 14, 155–.
- DURAND, J.-F. (1997) – *Additive modeling of multivariate data by splines functions*. Rapport de Recherche n°97.04, ENSAM-INRA Montpellier.
- DURAND, J.-F. (1999) – PLS and multivariate additive spline modeling. Tenenhaus, M. & Morineau, A. (édité par), *Les méthodes PLS, Symposium International PLS'99*, (pp. 1–20). CISIA-CERESTA.
- DURAND, J.-F. (2000) – *La régression Partial Least Squares Spline. Guide d'utilisation sous S-Plus* – Rapport de Recherche n°00.06. ENSAM-INRA-UMII.
- DURAND, J.-F. (2001) – Local polynomial additive regression through PLS and splines : PLSS – *Chemometrics and Intelligent Laboratory Systems*, 58, 235–246.
- DURAND, J.-F., ROMAN, S., & VIVIEN, M. (1998) – *Guide d'utilisation de la régression PLS sous Splus* – Rapport technique n°98-06, ENSAM-INRA-UMII, Montpellier.
- DURAND, J.-F. & SABATIER, R. (1997) – Additive splines for partial least squares regression – *Journal of the American Statistical Association*, 92(440), 1546–1554.
- EASTMENT, H.T. & KRZANOWSKI, W.J. (1982) – Cross-validatory choice of the number of components from principal component analysis – *Technometrics*, 24, 73–77.
- ERIKSSON, L., JOHANSSON, E., LINDGREN, F., & WOLD, S. (2000) – GIF-PLS: modelling non-linearities and discontinuities in QSAR – *Quantitative Structure-Activity Relationship*, 19, 345–355.
- ESBENSEN, K. & WOLD, S. (1983) – SIMCA, MACUP, SEPLS, GDAM, & UNFOLD: the way towards regionalised principal components analysis and subconstrained N-way decomposition. *Proceedings of Symposium on Applied Statistics*, (pp. 11–36)., Stavanger, Norway. June 12-14.
- ESCOFIER, B. & PAGÈS, J. (1984) – L'analyse factorielle multiple: une méthode de comparaison de groupes de variables. Diday, E. (édité par), *Data Analysis and Informatics III*, (pp. 41–55)., North-Holland. Elsevier.

- ESCOFIER, B. & PAGÈS, J. (1986) – *Analyses factorielles simples et multiples: objectifs, méthodes et interprétation* – Dunod – 241p.
- FEARN, T. (2000) – On orthogonal signal correction – *Chemometrics and Intelligent Laboratory Systems*, 50, 47–52.
- FORINA, M., CASOLINO, C., & PIZZARO MILLAN, C. (1999) – Iterative predictor weighting (IPW) PLS: a technique for the elimination of useless predictors in regression problems – *Journal of Chemometrics*, 13, 165–184.
- FRANC, A. (1992) – *Etude algébrique des Multitableaux: apport de l'algèbre tensorielle* – Thèse de Doctorat, Université Montpellier II - 362p.
- FRANK, I.E. (1987) – Intermediate least squares regression method – *Chemometrics and Intelligent Laboratory Systems*, 1, 233–242.
- FRANK, I.E. (1989) – *Nonlinear PLS model* – Talk given at the IX ICCCRE. Lago del Garda, Italie.
- FRANK, I.E. (1990) – A nonlinear PLS model – *Chemometrics and Intelligent Laboratory Systems*, 8, 109–119.
- FRANK, I.E. (1995) – Modern nonlinear regression methods – *Chemometrics and Intelligent Laboratory Systems*, 27, 1–9.
- FRIEDMAN, J.H. (1984) – SMART – Technical Report n°1, Department of Statistics, Stanford University.
- GAUCHI, J.-P. & CHAGNON, P. (2001) – Comparison of selection methods of explanatory variables in PLS regression with application to manufacturing process data – *Chemometrics and Intelligent Laboratory Systems*, 58, 171–193.
- GELADI, P., XIE, Y.-L., POLISSAR, A., & HOPKE, Ph. (1998) – Regression on parameters from three-way decomposition – *Journal of Chemometrics*, 12, 337–354.
- GIFI, A. (1990) – *Nonlinear multivariate analysis* – Wiley, New York.
- GLAÇON, F. (1981) – *Analyse conjointe de plusieurs matrices de données: Comparaison de différentes méthodes* – Thèse de Doctorat, Université Scientifique et Médicale de Grenoble. 73p.
- GOWER, J.C. (1975) – Generalised procrustes analysis – *Psychometrika*, 40, 33–51.
- HAARIO, H. & TAAVITSAINEN, V.-M. (1994) – Nonlinear data analysis. II. Examples on new link functions and optimization aspects – *Chemometrics and Intelligent Laboratory Systems*, 23, 51–64.
- HARSHMAN, R.A. (1970) – Foundations of the PARAFAC procedure: models and conditions for an explanatory multi-modal factor analysis – *UCLA Working Papers in Phonetics*, 16, 1–84.
- HARSHMAN, R.A. (1972a) – Determination and proof of minimum uniqueness condition for PARAFAC1 – *UCLA Working Papers in Phonetics*, 22, 111–117.
- HARSHMAN, R.A. (1972b) – PARAFAC2: Mathematical and technical notes – *UCLA Working Papers in Phonetics*, 22, 30–44.
- HARSHMAN, R.A. (1978) – *Models for analysis of asymmetrical relationships among N objects or stimuli* – Paper presented at the first joint meeting of the Psychometric Society and the Society for Mathematical Psychology, Hamilton, Ontario.
- HARSHMAN, R.A., GREEN, P.E., WIND, Y., & LUNDY, M.E. (1982) – A model for the analysis of asymmetric data in marketing research – *Marketing Science*, 1, 205–242.

- HARSHMAN, R.A. & HONG, S. (2002) – "stretch" vs "slice" methods for representing three-way structure via matrix notation – *Journal of Chemometrics*, 16, 198–205.
- HARSHMAN, R.A. & LUNDY, M.E. (1984) – The PARAFAC model for three-way factor analysis and multidimensional scaling. Law, H.G., Snyder, Jr., C.W., Hattie, J.A., & McDonald, R.P. (édité par), *Research methods for multimode data analysis*, (pp. 122–215). New York: Praeger.
- HARSHMAN, R.A. & LUNDY, M.E. (1994) – PARAFAC: Parallel factor analysis – *Computational Statistics & Data Analysis*, 18, 39–72.
- HARSHMAN, R.A. & LUNDY, M.E. (1996) – Uniqueness proof for a family of models sharing features of tucker's three-mode factor analysis and PARAFAC/CANDECOMP – *Psychometrika*, 61(1), 133–154.
- HASSEL, P.A., MARTIN, E.B., & MORRIS, J. (2002) – Non-linear partial least squares. Estimation of the weight vector – *Journal of Chemometrics*, 16, 419–426.
- HENRION, R. (1994) – Tutorial. N-way principal component analysis. Theory, algorithms and application – *Chemometrics and Intelligent Laboratory Systems*, 25, 1–23.
- HENRION, R. (1998) – Simultaneous simplification of loadings and core matrices in N-way PCA: application to chemometric data arrays – *Fresenius Journal of Analytical Chemistry*, 361, 15–22.
- HORST, P. (1961) – Relation among m sets of variables – *Psychometrika*, 38, 433–451.
- HÖSKULDSSON, A. (1988) – PLS regression methods – *Journal of Chemometrics*, 2, 211–228.
- HÖSKULDSSON, A. (1992a) – The H-principle in modelling with application to chemometrics – *Chemometrics and Intelligent Laboratory Systems*, 14, 139–154.
- HÖSKULDSSON, A. (1992b) – Quadratic PLS regression – *Journal of Chemometrics*, 6, 307–334.
- HÖSKULDSSON, A. (1998) – The Heisenberg modelling procedure and application to nonlinear modelling – *Chemometrics and Intelligent Laboratory Systems*, 44, 15–30.
- ISRAELSSON, A. (1969) – Three-way (or second order) component analysis. Wold, H. & Lyttkens, E. (édité par), *Non-linear iterative partial Least-squares (NIPALS) estimation procedures*, tome 43, (pp. 29–51). Bulletin of the International Statistical Institute.
- JÖRESKOG, K.G. & SÖRBUM, D. (1984) – *LISREL VI: Analysis of linear structural relationships by maximum likelihood, instrumental variables and least squares methods* – Scientific Software, Mooresville.
- KAZI-AOUAL, F., HITIER, S., SABATIER, R., & LEBRETON, J.-D. (1995) – Refined approximations to permutation tests for multivariate inference – *Computational Statistics & Data Analysis*, 20, 643–656.
- KIERS, H.A.L. (1988) – Comparison of "anglo-saxon" and "french" three-mode methods – *Statistique et Analyse de Données*, 13(3), 14–32.
- KIERS, H.A.L. (1991) – Hierarchical relations among three-way methods – *Psychometrika*, 56(3), 449–470.
- KIERS, H.A.L. (1993) – An alternating least squares algorithm for PARAFAC2 and three-way DEDICOM – *Computational Statistics & Data Analysis*, 62, 579–598.
- KIERS, H.A.L. (2000a) – Some procedures for displaying results from three-way methods – *Journal of Chemometrics*, 14, 151–170.

- KIERS, H.A.L. (2000b) – Towards a standardized notation and terminology in multiway analysis – *Journal of Chemometrics*, 14, 105–122.
- KIERS, H.A.L. & KRIJNEN, W.P. (1991) – An efficient algorithm for PARAFAC of three-way data with large numbers of observation units – *Psychometrika*, 56, 147–152.
- KIERS, H.A.L., KROONENBERG, P.M., & TEN BERGE, J.M.F. (1992) – An efficient algorithm for TUCKALS3 on data with large numbers of observation units – *Psychometrika*, 57, 415–422.
- KIERS, H.A.L., TEN BERGE, J.M.F., & BRO, R. (1999) – PARAFAC2. Part I. A direct fitting algorithm for the PARAFAC2 model – *Journal of Chemometrics*, 13, 275–294.
- KROONENBERG, P.M. (1983) – *Three-mode Principal Component Analysis - Theory and applications* – DSWO Press, Leiden – 398p.
- KROONENBERG, P.M. (1984) – Three-mode principal component analysis: Illustrated with an example from attachment theory. Law, H.G., C.W Snyder, J., Hattie, J.A., & McDonald, R.P. (édité par), *Research methods for multimode data analysis*, (pp. 64–103)., New York. Praeger.
- KROONENBERG, P.M. (1992) – Three-mode component models. A survey of the litterature – *Statistica Applicata*, 4(4), 619–633.
- KROONENBERG, P.M. (1994) – The TUCKALS line. A suite of programs for three-way data analysis – *Computational Statistics & Data Analysis*, 18, 73–96.
- KROONENBERG, P.M. & DE LEEUW, J. (1980) – Principal component analysis of three-mode data by means of alternating least squares algorithms – *Psychometrika*, 45, 69–97.
- KROONENBERG, P.M., TEN BERGE, J.M.F., BROUWER, P., & KIERS, H.A.L. (1989) – Gram-Schmidt versus Bauer-Rutishauser in alternating least-squares algorithms for three-mode principal component analysis – *Computational Statistics Quarterly*, 5, 81–87.
- KRUSKAL, J.B. (1977) – Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics – *Linear Algebra and Its Applications*, 18, 95–138.
- KRUSKAL, J.B. (1989) – Rank, decomposition and uniqueness for three-way and N-way arrays. Coppi, R. & Balasco, S. (édité par), *Multway data analysis*, (pp. 107–118). Elsevier, North-Holland.
- LAFOSSE, R. & HANAFI, M. (1997) – Concordance d'un tableau avec K tableaux : Définition de K+1 uples synthétiques – *Revue de Statistique Appliquée*, XLV(4), 111–126.
- LAROCHE, R. (2001) – *Contribution à l'élaboration des modèles prédictifs de l'activité biologique spécifique des tabacs en fonction des données chimiques, thermolyse et spectrales* – Rapport de stage de DESS, Université Montpellier II. 65p.
- LAVIT, Ch. (1988) – *Analyse conjointe de tableaux quantitatifs* – Masson, Paris – 252p.
- LAVIT, Ch., ESCOUFIER, Y., SABATIER, R., & TRAISSAC, P. (1994) – The ACT (STATIS method) – *Computational Statistics & Data Analysis*, 18, 97–119.
- LAW, H.G., SNYDER, C.W., HATTIE, J.A., & McDONALDS, R.P. (édité par) (1984) – *Research methods for multimode data analysis* – Praeger, New York.
- LAZRAQ, A. & CLÉROUX, R. (2001) – The PLS multivariate regression model: testing the significance of successive PLS components – *Journal of Chemometrics*, 15, 523–536.
- LEURGANS, S.E., ROSS, R.T., & ABEL, R.B. (1993) – A decomposition for 3-way arrays – *SIAM Journal on Matrix Analysis and Applications*, 14, 1064–1083.

- L'HERMIER DES PLANTES, H. (1976) – *Structuration des tableaux à trois indices de la statistique: Théorie et application d'une méthode d'analyse conjointe* – Thèse de Doctorat, Université des Sciences et Techniques du Languedoc, Montpellier. 98p.
- LI, T., MEI, H., & CONG, P. (1999) – Combining nonlinear PLS with the numeric genetic algorithm for QSAR – *Chemometrics and Intelligent Laboratory Systems*, 45, 177–184.
- LINDGREN, F., GELADI, P., RÄNNAR, S., & WOLD, S. (1994) – Interactive variable selection (IVS) for PLS. part 1: theory and algorithms – *Journal of Chemometrics*, 8, 349–363.
- LOHMÖLLER, J.B. (1987) – LVPLS Program manual, version 1.8 – Zentralarchiv für Empirische Sozialforschung, Köln.
- LOHMÖLLER, J.B. (1989) – *Latent variables Path modelling with Partial Least Squares* – Physica-Verlag, Heidelberg.
- LOHMÖLLER, J.B. & WOLD, H. (1980) – Three-mode path models with latent variables and partial least squares (PLS) parameter estimation.(revised 1982) – Presented at the European Meeting of the Psychometrics Society, Groningen, Netherlands. also *Research Report n°80.03*, München: University of the Federal Armed Forces. Department of Education.
- LOUWERSE, D.J., SMILDE, A.K., & KIERS, H.A.L. (1999) – Cross validation of multiway component models – *Journal of Chemometrics*, 13, 491–510.
- MAGNUS, J.R. & NEUDECKER, H. (1988) – *Matrix differential calculus with applications in statistics and econometrics* – Wiley – Revised in 1991, Reprinted in 1995 - 391p.
- MALTHOUSE, E.C., MAH, R.S.H., & TAMHANE, A.C. (1994) – Nonlinear partial least squares using neural networks – *InCINC'94, the first International Chemometrics Internet Conference*. www.emsl.pnl.gov:2080/docs/incinc/alt_nns.html.
- MALTHOUSE, E.C., TAMHANE, A.C., & MAH, R.S.H. (1997) – Nonlinear partial least squares – *Computers & Chemical Engineering*, 21, 875–890.
- MAZOUNI, N., VIVIEN, M., SABATIER, R., GAERTNER, J.-C., & DURBEC, J.-P. (2001) – Organisation spatiale de la macrofaune benthique de l'étang de thau. *Méthodes et outils pour l'étude de l'influence des facteurs hydroclimatiques et anthropiques sur la variabilité spatio-temporelle des populations et écosystèmes marins, Bilan 2000 et Projets 2001 du PNEC/ART4*, (pp. 57–62). Coordinateurs: B Beliaeff et J.-P Durbec.
- MICHAILIDIS, G. & DE LEEUW, J. (1998) – The GIFI system of descriptive multivariate analysis – *Statistical Science*, 13(4), 307–336.
- MOLINARI, N. (2000) – *Optimisation des noeuds en régression spline: applications aux biostatistiques* – Thèse de Doctorat, Université Montpellier I. 147p.
- OSTEN, D.W. (1988) – Selection of optimal regression models via cross-validation – *Journal of Chemometrics*, 2, 39–48.
- PAATERO, P. (1997) – A weighted non-negative least squares algorithm for three-way PARAFAC factor analysis – *Chemometrics and Intelligent Laboratory Systems*, 38, 223–242.
- PAATERO, P. & TAPPER, U. (1994) – Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values – *Environmetrics*, 5, 111–126.
- PERNIN, M.-O. (1986) – *Contribution à la méthodologie d'analyse de données longitudinales. Exemple de croissance chez l'être humain* – Thèse de Doctorat, Université de Lyon I. 237p.
- QANNARI, E.M., COURCOUX, Ph., LEJEUNE, M., & MAYSTRE, O. (1997) – Comparaison de trois stratégies de détermination d'un compromis en évaluation sensorielle – *Revue de Statistique Appliquée*, XLV(1), 61–74.

- QIN, S.J. & MACAVOY, T.J. (1992) – Non-linear PLS modelling using neural networks – *Computers & Chemical Engineering*, 16, 379–391.
- QIN, S.J., VALLE, S., & PIOVOSO, M.J. (2001) – On unifying multiblock analysis with application to decentralized process monitoring – *Journal of Chemometrics*, 15, 715–742.
- RAO, C.R. (1964) – The use and interpretation of principal component analysis in applied research – *Sankhya*, A(26), 329–359.
- RAO, C.R. & MITRA, S. (1971) – *Generalized inverse of matrices and its applications* – Wiley, New York.
- REIS, M.M. & FERREIRA, M.C. (2002) – PARAFAC with splines: a case study – *Journal of Chemometrics*, 16, 444–450.
- ROBERT, P. & ESCOUFIER, Y. (1976) – A unifying tool for linear multivariate statistics methods: the RV coefficient – *Applied Statistics*, 25(3), 257–265.
- RUTISHAUSER, H. (1969) – Computational aspects of L.R. Bauer’s simultaneous iteration method – *Numerish Mathematik*, 13, 4–13.
- SABATIER, R. (1993) – Critères et contraintes pour l’ordination simultanée de k tableaux. Lebreton, J.-D. & Asselain, B. (édité par), *Biométrie et environnement*, (pp. 100–121)., Paris. Masson.
- SAPORTA, G. (1990) – *Probalibilités, analyse de données et statistique* – Technip, Paris – 493p.
- SCHWARTZ, H.R., RUTISHAUSER, H., & STIEFEL, E. (1968) – *Numerik symmetrischer Matrizen* – FRG: Teubner, Stuttgart.
- SEMPÉ, M., KROONENBERG, P.M., LAVIT, Ch., PERNIN, M.-O., LEWI, P.J., CALOMNE, G., VAN HOOF, J., PONTIER, J., & MINEO, A. (1989) – Multivariate and longitudinal data on growing children. Janssen, Marcotorchino, & Proth (édité par), *Data analysis: the ins and outs of solving real problems*, (pp. 7–112).
- SEMPÉ, M., PEDRON, G., & ROY-PERNOT, M.P. (1979) – Auxologie, méthode et séquences – Laboratoire THERAPLIX, Paris.
- SIMIER, M., BLANC, L., PELLEGRIN, F., & NANDRIS, D. (1999) – Approche simultanée de K couples de tableaux : Application à l’étude des relations pathologie végétale- environnement – *Revue de Statistique Appliquée*, XLVIII(1), 31–46.
- SJÖBLOM, J., SVENSSON, O., JOSEFSON, M., KULLBERG, H., & WOLD, S. (1998) – An evaluation of orthogonal signal correction applied to calibration transfer of near infrared spectra – *Chemometrics and Intelligent Laboratory Systems*, 44, 229–244.
- SLAMA, C.F. (1991) – *Multivariate statistical analysis of data from an industrial fluidized catalytic process using PCA and PLS* – Master’s Thesis, McMaster University, Hamilton, Ontario.
- SMILDE, A.K. (1997) – Comments on multilinear PLS – *Journal of Chemometrics*, 11, 367–377.
- SMILDE, A.K. & KIERS, H.A.L. (1999) – Multiway Covariates regression models – *Journal of Chemometrics*, 13, 31–48.
- SMILDE, A.K., WESTERHUIS, J.A., & BOQUÉ, R. (2000) – Multiway multiblock component and covariates regression models – *Journal of Chemometrics*, 14, 301–331.
- STÄHLE, L. (1989) – Aspects of the analysis of three-way data – *Chemometrics and Intelligent Laboratory Systems*, 7, 95–100.

- STONE, M. (1974) – Cross-Validatory choice and assessment of statistical predictions – *Journal of the Royal Statistical Society - serie B*, 36, 111–147.
- STONE, M. & BROOKS, R.J. (1990) – Continuum regression: cross-validated sequentially constructed prediction embracing ordinary least squares, partial least squares and principal components regression – *Journal of the Royal Statistical Society - serie B*, 52(2), 237–269.
- SUNE, F. (2001) – *L'attribut multidimensionnel: comparaison entre les attributs sensoriels donnés par les enfants pour décrire un produit alimentaire appartenant à leur univers, à ceux utilisés par les experts* – Thèse de Doctorat, Université Nice-Sofia Antipolis. 430p.
- SUNE, F., LACROIX, P., & HUON DE KERMADEC, F. (2002) – A comparison of sensory attribute use by children and experts to evaluate chocolate – *Food Quality and Preference*, 13, 545–553.
- SVENSSON, O., KOURTI, T., & MACGREGOR, J.F.. (2002) – An investigation of orthogonal signal correction algorithms and their characteristics – *Journal of Chemometrics*, 16, 176–188.
- TAAVITSAINEN, V.-M. & KORHONEN, P. (1992) – Nonlinear data analysis with latent variable – *Chemometrics and Intelligent Laboratory Systems*, 14, 185–194.
- TEN BERGE, J.M.F., DE LEEUW, J., & KROONENBERG, P.M. (1987) – Some additional results on principal components analysis of three-mode data by means of alternating least squares algorithms – *Psychometrika*, 52, 183–191.
- TEN BERGE, J.M.F. & KIERS, H.A.L. (1991b) – Some clarifications of the CANDECOMP algorithm applied to INDSCAL – *Psychometrika*, 56, 317–326.
- TENENHAUS, M. (1995) – *A partial least squares approach to multiple regression, redundancy analysis and canonical analysis* – Les Cahiers de la Recherche HEC. CR 500/1995.
- TENENHAUS, M. (1998) – *La régression PLS, théorie et pratique* – Technip, Paris – 254p.
- TENENHAUS, M. (1999) – L'approche PLS – *Revue de Statistique Appliquée*, XLVII(2), 5–40.
- TENENHAUS, M., GAUCHI, J.-P., & MENARDO, C. (1995) – Régression PLS et applications – *Revue de Statistique Appliquée*, XLIII, 7–63.
- THIOULOUSE, J., CHESSEL, D., DOLÉDEC, S., & OLIVIER, J.-M. (1997) – ADE-4: a multivariate analysis and graphical display software – *Statistics and Computing*, 7, 75–83.
- TIMMERMAN, M.E. & KIERS, H.A.L. (2002) – Three-way component analysis with smoothness constraints – *Computational Statistics & Data Analysis*, 40, 447–470.
- TOMINAGA, Y. & FUJIWARA, I. (1997) – Prediction-weighted partial least squares regression method (PWPLS) – *Chemometrics and Intelligent Laboratory Systems*, 38, 139–144.
- TRYGG, J. (2002) – O2-PLS for qualitative and quantitative analysis in multivariate calibration – *Journal of Chemometrics*, 16, 283–293.
- TRYGG, J. & WOLD, S. (2002) – Orthogonal projections to latent structure – *Journal of Chemometrics*, 16, 119–128.
- TUCKER, L.R. (1958a) – Determination of parameters of functional relation by factor analysis – *Psychometrika*, 23, 19–23.
- TUCKER, L.R. (1958b) – An inter-battery method of factor analysis – *Psychometrika*, 23, 111–136.
- TUCKER, L.R. (1963) – Implications of factor analysis of three-way matrices for measurement of change. Harris, C.W. (édité par), *Problems in measuring change*, (pp. 122–137)., Madison. University of Wisconsin Press.

- TUCKER, L.R. (1964) – The extension of analysis to three-dimensional matrices. *Contributions to mathematical psychology*, (pp. 110–127)., New York. Holt, Rinehart, Winston.
- TUCKER, L.R. (1966) – Some mathematical notes on three-mode factor analysis – *Psychometrika*, 31, 279–311.
- TUCKER, L.R. & MESSICK, S. (1963) – An individual difference model for multidimensional scaling – *Psychometrika*, 26, 333–367.
- URDAPILLETA, I., TON NU, C., SAINT DENIS, Ch., & HUON DE KERMADEC, F. (édité par) (2001) – *Traité d'évaluation sensorielle. Aspects cognitifs et métrologiques des perceptions* – Dunod, Paris – 556p.
- VAN DER BURG, E. (1988) – *Non linear canonical correlation and some related methods* – DSWO Press, Leiden.
- VAN DER BURG, E., DE LEEUW, J., & DIJKSTERHUIS, G. (1994) – OVERALS. Nonlinear canonical correlation with k sets of variables – *Computational Statistics & Data Analysis*, 18, 141–163.
- VAN DER BURG, E. & DIJKSTERHUIS, G.B. (1989) – Nonlinear canonical correlation of multiway data. Coppi, R. & Balasco, S. (édité par), *Multiway data analysis*, (pp. 245–255). elsevier.
- VERRON, T. (2002) – *Nouvelles approches PLS pour le traitement et l'analyse multi-tableaux de spectres NIRS* – Rapport de stage de DEA de Biostatistique, Université Montpellier II. 65p.
- VIVIEN, M. (1999) – *Nouvelles approches en analyse multi-tableaux* – Rapport de stage de DEA de Biostatistique, Université Montpellier II. 96p.
- VIVIEN, M. & SABATIER, R. (2000) – Nouvelle extension de la régression PLS. *Actes des 6èmes Journées Européennes Agro-Industrie et Méthodes Statistiques*, (pp. 5.0–5.9)., Pau. 19-21 Janvier.
- VIVIEN, M. & SABATIER, R. (2001a) – Modélisation de multi-tableaux: l'ACIMOG-PLS. *Actes de Chimométrie 2001*, CNAM; Paris. 4-5 Déc.
- VIVIEN, M. & SABATIER, R. (2001b) – Nouvelle méthode de régression pour les multi-tableaux: l'ACIMOG-PLS. *Actes des XXXIIIèmes Journées de Statistique, ASU*, (pp. 707–710)., Nantes. 14-18 Mai.
- VIVIEN, M. & SABATIER, R. (2001c) – Une extension multi-tableaux de la régression PLS – *Revue de Statistique Appliquée*, XLIX(1), 31–54.
- VIVIEN, M. & SABATIER, R. (2002) – Analyse de Co-Inertie Multiple Orthogonale Généralisée 2: mise en évidence d'une relation entre données sensorielles et données hédoniques. *Actes des 7èmes Journées Européennes Agro-Industrie et Méthodes Statistiques*, Lille. 16-18 Janv.
- WAKELING, I.N. & MORRIS, J.J. (1993) – A test of significance for partial least squares regression – *Journal of Chemometrics*, 7, 291–304.
- WANGEN, L.E. & KOWALSKI, B.R. (1988) – A multiblock partial least squares algorithm for investigating complex chemical systems – *Journal of Chemometrics*, 3, 3–20.
- WEGMAN, E.J. & WRIGHT, I.W. (1983) – Splines in statistics – *Journal of the American Statistical Association*, 78(382), 351–365.
- WESTERHUIS, J.A. & COENEGRACHT, P.M.J. (1997) – Multivariate modelling of the pharmaceutical two-step process of wet granulation and tableting with multiblock partial least squares – *Journal of Chemometrics*, 11(5), 379–392.

- WESTERHUIS, J.A., DE JONG, S., & SMILDE, A.K. (2001) – Direct orthogonal signal correction – *Chemometrics and Intelligent Laboratory Systems*, 56, 13–25.
- WESTERHUIS, J.A., KOURTI, T., & MACGREGOR, J.F. (1998) – Analysis of multiblock and hierarchical PCA and PLS models – *Journal of Chemometrics*, 12, 301–321.
- WESTERHUIS, J.A. & SMILDE, A.K. (2001) – Short Communication. Deflation in multiblock PLS – *Journal of Chemometrics*, 15, 485–493.
- WOLD, H. (1966a) – Estimation of principal component and related models by iterative least squares. Krishnaiah, P.-R. (édité par), *Multivariate Analysis*, (pp. 391–420)., New-York. Academic Press.
- WOLD, H. (1966b) – Nonlinear estimation by iterative least squares procedures. David, F. (édité par), *Research papers in statistics, festschrift for Jerzy Neyman*, (pp. 411–444)., New York. Wiley.
- WOLD, H. (édité par) (1975) – *Modelling in Complex Situations with Soft Information*, Toronto, Canada. Third World Congress of Econometric Society.
- WOLD, H. (1982) – Soft modeling: the basic design and some extensions. K.G Jöreskog & Wold, H. (édité par), *System under indirect observation, vol. 2*, (pp. 1–54)., North-Holland, Amsterdam.
- WOLD, H. (1985) – Partial Least Squares. Kotz, S. & Johnson, N.L. (édité par), *Encyclopedia of Statistical Sciences, Vol. 6*, (pp. 581–591)., New York. John Wiley & Sons.
- WOLD, S. (1974) – Spline functions in data analysis – *Technometrics*, 16(1), 1–11.
- WOLD, S. (1984) – Three PLS algorithms according to SW. Wold, S. (édité par), *MULDAST Proceedings*, (pp. 26–30). *Tech. Rep.* Research Group for Chemometrics, Umeå University.
- WOLD, S. (1992) – Nonlinear partial least squares modelling II. Spline inner relation – *Chemometrics and Intelligent Laboratory Systems*, 14, 71–84.
- WOLD, S., ANTTI, H., LINDGREN, F., & ÖHMAN, J. (1998) – Orthogonal signal correction of near-infrared spectra – *Chemometrics and Intelligent Laboratory Systems*, 44, 175–185.
- WOLD, S., GELADI, P., ESBENSEN, K., & ÖHMAN, J. (1987) – Multiway Principal Components- and PLS- Analysis – *Journal of Chemometrics*, 1, 41–56.
- WOLD, S., HELLBERG, S., LUNDSTEDT, T., SJOSTROM, M., & WOLD, H. (1987) – PLS modeling with latent variable in two or more dimensions. *Proc. Symp. on PLS Model Building: theory and Application*, Frankfurt am Main. also *Tech. Rep.*, Department of Organic Chemistry, Umeå University (1987).
- WOLD, S., KETTANEH, N., & TJESSEM, K. (1996) – Hierarchical multiblock PLS and PC models for easier model interpretation and as an alternative to variable selection – *Journal of Chemometrics*, 10, 463–482.
- WOLD, S., KETTANEH-WOLD, N., & SKAGERBERG, B. (1989) – Nonlinear PLS modeling – *Chemometrics and Intelligent Laboratory Systems*, 7, 53–65.
- WOLD, S., MARTENS, H., & WOLD, H. (1983) – The multivariate calibration problem in chemistry solved by the PLS method. Ruhe, A. & Kågstrøm, B. (édité par), *Proceedings of the Conference on Matrix Pencils*, (pp. 286–293)., Heidelberg. Lectures Notes in Mathematics, Springer Verlag.