

S O M M A I R E

■ **Ojectifs du congrès** p 2

■ **Programme** p 4

■ **Communications orales** p 7

Analyse multivariée p 7

Traitement d'images p 17

Méthodes séparatives multidimensionnelles p 27

Applications industrielles p 33

Méthodes multivoies p 37

Plans d'expériences et plans numériques p 39

Multi-tableaux p 49

■ **Posters** p 65

Les objectifs du Congrès :

Le congrès annuel du Groupe Français de Chimiométrie, **Chimométrie 2010**, vise à rassembler Universitaires et Industriels pour faire le point sur les avancées de la Chimiométrie, allant de la collecte de l'information aussi efficacement et précisément que possible à partir de plans d'expériences, à l'analyse et à la modélisation des données. Ce congrès, qui depuis plusieurs années réunit théoriciens et expérimentateurs de tout domaine, a pour vocation de promouvoir l'utilisation de la Chimiométrie au delà même de l'application des Statistiques dans le domaine de la Chimie. C'est un lieu privilégié de rencontres et d'échanges, qui donne la possibilité aux Universitaires et Industriels, sans restriction disciplinaire, de communiquer en faisant état de leurs expériences, de leurs préoccupations et/ou de leurs besoins actuels et futurs.

En cette année 2010, les objectifs principaux seront :

1- Prendre connaissance des avancées dans différentes thématiques qui seront présentées lors des sessions du congrès. Les thèmes privilégiés cette année seront :

- L'Analyse multivariée,
- Le traitement d'images,
- Les méthodes séparatives multidimensionnelles,
- Les Plans d'expériences et plus particulièrement les plans numériques,
- Les approches multi-tableaux,
- Des applications industrielles,

2- Echanger et communiquer sur la chimométrie appliquée aux différents métiers du monde industriel et académique. L'accent sera mis sur les disciplines et les thèmes suivants :

- Agro-alimentaire,
- Analyse sensorielle,
- Biologie, protéomique, métabolomique,...
- Cosmétiques,
- Développement, optimisation et validation de méthodes analytiques,
- Interprétation des résultats issus de techniques analytiques couplées (LC-MS, GC-MS, GCxGC, etc...),
- Optimisation et suivi des procédés,
- Pharmacie, chimie médicinale,
- Spectroscopies vibrationnelles,
- Techniques séparatives,
- Traitement de l'image.

Il sera aussi mis l'accent sur les méthodes d'apprentissage, de simulation, d'extraction des connaissances, et sur les récents développements concernant l'analyse de données complexes comme les images ou les données issues de la Bio ou Chemo-Informatique.

3- S'informer des développements récents et des nouveautés en matière de logiciels ou d'édition.

Des exposants seront présents lors des deux journées pour présenter les développements récents en matière de logiciels ou d'instrumentation.

Comme chaque année, un Challenge-concours primé est organisé par le comité scientifique pour donner envie aux chimiométriciens de mettre en oeuvre les outils de leur domaine et ainsi de partager avec les concurrents leur pratique face à un jeu de données, mis à disposition sur le site web du congrès.

4 - Offrir aux jeunes chercheurs une opportunité de présenter leurs travaux et leurs compétences.

Les jeunes chimiométriciens de 30 ans ou moins, sont invités et encouragés à présenter leurs travaux. Un prix récompensera la meilleure communication orale.

Publication :

Les communications, orales ou par affiche, présentées lors du congrès pourront être soumises pour publication dans la revue "Chemometrics and Intelligent Laboratory Systems" (Elsevier). Les textes définitifs devront être téléchargés, suivant les instructions indiquées, et au plus tard le 30 Décembre 2010, sur le site : <http://ees.elsevier.com/chemolab> (rubrique "Submit paper")

Les Prix du Congrès :

- Le Prix junior :
- Le Prix du "challenge" :
- Le Prix des meilleurs posters :
Les trois meilleurs posters élus par les congressistes seront primés.
- Des Prix d'honneur attribués par le Centre de Compétence en Chimie et Toxicologie Analytique (Suisse) récompenseront des jeunes chercheurs auteurs de communications orales ou par affiches.

Comité scientifique :

Président :

Michelle SERGENT - *Université Paul Cézanne, Marseille*

Membres :

Philippe BASTIEN - *l'Oréal Recherche, Aulnay-sous-Bois*

Dominique BERTRAND - *INRA, Nantes*

Delphine BOUVERESSE - *INRA, Paris*

Jacques R. CHRETIEN - *BIOCHEMICS, Orléans*

Pierre DARDENNE - *CRA, Gembloux*

Ludovic DUPONCHEL - *LASIR- Université de Lille I*

Max FEINBERG - *INRA, Paris*

Philippe GIAMARCHI - *Université de Bretagne Occidentale Brest*

Pierre LANTERI - *Université Lyon1/CPE Lyon*

Georges OPPENHEIM - *Université Paris Sud - Orsay*

Jean-Michel ROGER - *Unite Mixte de Recherche Cemagref - SupAgro, Montpellier*

Serge RUDAZ - *Université de Genève*

Douglas N. RUTLEDGE - *INRA-AgroParisTech, Paris*

Claude-Alain SABY - *TOTAL France, Solaize*

Gilbert SAPORTA - *CNAM, Paris*

Magida ZEAITER - *GlaxoSmithKline, UK*

Comité d'organisation :

Président :

Jérôme VIAL - *ESPCI ParisTech, Paris*

Membres :

Servane BIANCIARDI - *SfDS, Paris*

Delphine BOUVERESSE - *INRA/AgroParisTech, Paris*

Christophe CORDELLA - *INRA/AgroParisTech, Paris*

Jacques GOUPY - *ReConFor, Paris*

Pierre LANTERI - *Université Lyon1/CPE Lyon*

Douglas N. RUTLEDGE - *INRA/AgroParisTech, Paris*

Michelle SERGENT - *Université Paul Cézanne, Marseille*

Exposition

Les congressistes pourront s'informer des développements récents en instrumentations et des nouveautés en logiciel ou en édition. Une exposition permettra un contact direct avec les commerciaux des sociétés d'instrumentation et éditrices de logiciels :



PROGRAMME

Jeudi 2 Décembre 2010

08h30 **Accueil des participants**

09h00 **Ouverture du Congrès et présentation de l'ESPCI par le Pr. A. Jardy**

Analyse multivariée

(Chairmen: Douglas N. Rutledge, Philippe Bastien)

09h30 **Conférence plénière : Regularized generalized canonical correlation analysis**

Michel Tenenhaus

10h15 **SmartICA, une application Web pour l'Analyse en Composantes Indépendantes**

Luc Jouneau - INRA France

10h40 **Pause café / Posters / Exposition**

11h25 **Etude des propriétés écotoxicologiques des solvants dans le cadre de la réglementation REACH**

Amélie Levet - Laboratoire des Sciences Analytiques, Université Lyon 1, France

11h50 **VODKA-PLSR, une famille de modèles PLSR dérivés de l'algorithme NIPALS**

Jean-Claude Boulet - INRA, France

12h15 **Déjeuner / Posters / Exposition**

Traitement d'images

(Chairmen: Claude-Alain Saby, Jean-Michel Roger)

14h00 **Discrimination des variétés d'olives par imagerie numérique des noyaux et analyse chimiométrique**

Pierre Vanloot - CNRS UMR 6263, ISM2, équipe AD2EM groupe Systèmes Chimiques Complexes, case 451, Université Paul Cézanne, Avenue Escadrille Normandie Niémen, 13397 Marseille cedex 20, France

14h25 **Relating hyperspectral images having different spatial resolutions**

Fatma Allouche - INRA - Nantes France

14h50 **Increasing the spatial resolution in vibrational microscopic imaging using post-processing algorithms**

Ludovic Duponchel - Laboratoire de Spectrochimie Infrarouge et Raman (LASIR CNRS UMR 8516)

15h15 **Pause café / Posters / Exposition**

Méthodes séparatives multidimensionnelles

(Chairmen: Serge Rudaz, Max Feinberg)

16h00 **Comparaison de différentes approches statistiques pour évaluer l'orthogonalité de séparations chromatographiques: application aux systèmes en phase inverse.**

Ramia Al Bakain - Laboratoire des Sciences Analytiques, Bioanalytiques et Miniaturisation UMR- CNRS -UPMC- PECSA, Ecole Supérieure de Physique et de Chimie Industrielles de la Ville de Paris, ParisTech, 10 Rue Vauquelin, 75005 Paris, France

16h25 **Stratégie chimiométrique pour l'analyse de chromatogrammes bidimensionnels.**

Isabelle Rivals- Équipe de Statistique Appliquée, Ecole Supérieure de Physique et de Chimie Industrielles de la Ville de Paris, ParisTech, 10 Rue Vauquelin, 75005 Paris

Applications industrielles

(Chairman: Ludovic Duponchel)

16h50 **Utilising technology to simplify process optimisation, improvement and control**

Malcolm Moore - SAS JMP

17h15 **Challenge**

Pierre Dardenne - Walloon Agricultural Research Centre, Gembloux, Belgium

18h15 **Cocktail sponsorisé par JMP**

Vendredi 3 Décembre 2010

Méthodes multivoies

(Chairmen: Delphine Bouveresse, Philippe Giamarchi)

08h30 **Conférence plénière : Advanced trends of multivariate curve resolution in process modelling and hyperspectral image resolution**

Anna de Juan, Chemometrics group, Dept. Analytical Chemistry, Universitat de Barcelona, Espagne

Plans d'expériences et plans numériques

(Chairmen: Michelle Sergent, Pierre Lantéri)

09h15 **Construction de plans uniformes en grande dimension**

Jenny Santiago - Laboratoire de la Méthodologie et Recherche Expérimentale France

09h40 **High dimensional space filling design analysis**

Adrian Azarian - ONERA France

10h05 **Pause café / Posters / Exposition**

10h50 **Potentiel des plans numériques de type Space Filling Design pour l'analyse de sensibilité et la construction de métamodèles dans le domaine de l'optique interférentielle**

Olivier Vasseur - ONERA, France

11h15 **DOE: from empiricism to tailor-made, the story of the optimisation of red blood cells culture**

Laurence Guyonneau-Harmand – Université Pierre et Marie Curie, France

Multi-tableaux

(Chairmen: Christophe Cordella, Dominique Bertrand)

11h40 **On the study of the influence of geographical factors on skin surface lipids (SSLs): Multivariate analysis of Raman and GC-MS data**

Rime Michael-Jubeli – Université de Paris Sud, Châtenay-Malabry, France

12h05 **Analyse de plans d'expériences en situation multivariée : une contribution des méthodes multi-blocks.**

Gérard Mazerolles - INRA, France

12h30 **Déjeuner / Posters / Exposition**

14h00 **Discrimination PLS en présence de données à trois entrées**

Samia-Samar Ouertani - ONIRIS Nantes, France et Université de Genève, Suisse

14h25 **Combinaison de several mass spectrometry ionisation modes:**

A multiblock analysis for a rapid characterisation of the red wine polyphenolic composition

Sébastien Preys - Ondalys, France

14h50 **Applications de modèles de mélanges gaussiens pour la classification supervisée de données spectroscopiques (de grande dimension)**

Cyril Ruckebusch - Lasir CNRS univ Lille1, France

15h15 **Analyse Canonique Généralisée à noyau**

Arthur Tenenhaus - Supelec, France

15h40 **Pause café / Posters / Exposition**

16h25 **Remise des Prix - Clôture**

17h00 **Fin**

Contacts

www.chimietrie.fr

www.chimietrie.fr/chemom2010

ANALYSE MULTIVARIÉE

CHAIRMEN:
DOUGLAS N. RUTLEDGE
PHILIPPE BASTIEN

Regularized generalized canonical correlation analysis

Michel Tenenhaus

41 rue des Francs Bourgeois, 75004 Paris, tenenhaus@hec.fr

Regularized generalized canonical correlation analysis (RGCCA) is a generalization of regularized canonical correlation analysis to three or more sets of variables. It constitutes a general framework for many multi-block data analysis methods. It combines the power of multi-block data analysis methods (maximization of well identified criteria) and the flexibility of PLS path modeling (the researcher decides which blocks are connected and which are not). Searching for a fixed point of the stationary equations related to RGCCA, a new monotone convergent algorithm, very similar to the PLS algorithm proposed by Herman Wold, is obtained. Finally, a practical example is discussed.

Keywords: Generalized canonical correlation analysis, Multi-block data analysis, PLS path modeling, Regularized canonical correlation analysis.

Références

[1] A. Tenenhaus and M. Tenenhaus. Regularized generalized canonical correlation analysis. *Psychometrika* (in press).

SmartICA, une application web pour l'Analyse en Composantes Indépendantes

M. Luc Jouneau¹ - M^{lle}. Faten Ammari² - M. Douglas N. Rutledge²

¹INRA, UR0892 ; Unité de Virologie et Immunologie Moléculaires, Jouy-en-Josas, 78352, France

²AgroParisTech/INRA UMR1145 Génial, Laboratoire de Chimie Analytique, 16 rue Claude Bernard, Paris, 75005, France

Keywords: Analyse en Composantes Indépendantes.

1- Introduction

L'Analyse en Composantes Indépendantes (ACI) est utilisée pour dégager d'une matrice de donnée, les signaux purs qui y sont mélangés de manière linéaire^[1].

Le premier objectif de l'application SmartICA est de permettre aux scientifiques de réaliser une analyse ACI sur leurs données sans connaissance des langages utilisés sous les logiciels spécialisés (R, MatLab). Cela doit permettre une meilleure diffusion de cette analyse dans les domaines scientifiques où elle est encore sous utilisée, comme par exemple, en biologie.

Un deuxième objectif de l'application est de proposer à l'utilisateur un certain nombre d'outils qui facilite l'interprétation des résultats rendus par l'ACI.

2- Fonctionnalités proposées par SmartICA

Lors de la soumission, l'utilisateur peut demander à associer une couleur aux variables quantitatives ou qualitatives (afin de mettre en évidence un résultat rendu par une autre analyse, par exemple une analyse de variance).

L'utilisateur peut associer aux variables quantitatives différentes informations (nom, description, ...) qui vont permettre de mieux identifier celles qui contribuent le plus à la construction des différents composants indépendants rendus par l'ACI.

Pour chacun des composants indépendants, l'utilisateur peut demander à visualiser la projection des variables qualitatives ou quantitatives. Le logiciel trie les variables par leur valeur rendue par l'ACI pour un composant donné et permet ainsi de sélectionner celles qui participent le plus à la construction du composant.

L'utilisateur peut alors obtenir une liste des variables les plus extrêmes, soit les valeurs les plus basses, soit les valeurs plus hautes, soit un mixte des deux :



Figure 1 - Projection des variables quantitatives sur un des composants indépendants rendus par l'ACI. Les variables apparaissent triées selon leur valeur attribuée par l'ACI pour le composant indépendant sélectionné.

La zone grise représente la sélection des variables les plus contributrices choisie par l'utilisateur.

Une fois cette sélection effectuée, l'utilisateur peut demander à obtenir la liste des variables sélectionnées avec les informations qui leur sont rattachées (si l'utilisateur a fourni ces informations lors de la soumission).

Il peut demander à voir les intensités de signal pour ces variables. SmartICA présente un tableau dans lequel pour chaque variable quantitative, est calculée la valeur minimale et la valeur maximale de signal, soumis à l'ACI. La valeur minimale sera affichée avec un fond de cellule de couleur bleu, tandis que la valeur maximale sera figurée avec un fond rouge. Les autres valeurs auront une valeur de fond intermédiaire. Cette mise en valeur des données permet de rapidement identifier quelles sont les variables qualitatives qui sont mises en opposition dans le composant indépendant sélectionné :

	F1	F2	M1	M2
SC_XIST_BOVIN.3.44	18.08	18.37	0.01	0.0
SC_XIST_BOVIN.1.44	19.09	19.43	2.56	5.56
SC_NEFH_BOVIN.1.1	16.24	15.98	1.8	3.56
SC_XIST_BOVIN.17.44	12.84	12.71	0.01	0.0
SC_STMN2_BOVIN.1.2	9.95	9.99	0.01	3.1
SC_XIST_BOVIN.10.44	7.0	8.91	0.01	0.0
SC_XIST_BOVIN.2.44	9.95	10.17	2.56	1.79
SC_FMO2_BOVIN.1.25	16.79	15.88	7.69	8.23
SC_STMN2_BOVIN.2.2	11.4	10.97	4.61	4.29

Figure 2 - Sur cet exemple, on met clairement en évidence une opposition entre les variables qualitatives {F1 ; F2} et {M1 ; M2}.

Au sein de ces groupes, l'utilisateur peut demander au logiciel de dégager des variables quantitatives qui sont fortement corrélées. Une matrice de corrélation entre les variables quantitatives sélectionnées est calculée. A chaque fois que dans cette matrice, SmartICA détecte un coefficient de corrélation supérieur à un seuil fixé par l'utilisateur, un arc est créé entre les deux variables quantitatives fortement corrélées. Au final, SmartICA dispose de différents graphes reliant les variables quantitatives entre elles. Il ne retient que les graphes qui ont un nombre de nœuds égal ou supérieur à un seuil fixé par l'utilisateur. L'utilisateur peut alors voir ces graphes et la liste des variables quantitatives qui les constituent (triées par nombre de connexion au sein d'un graphe)

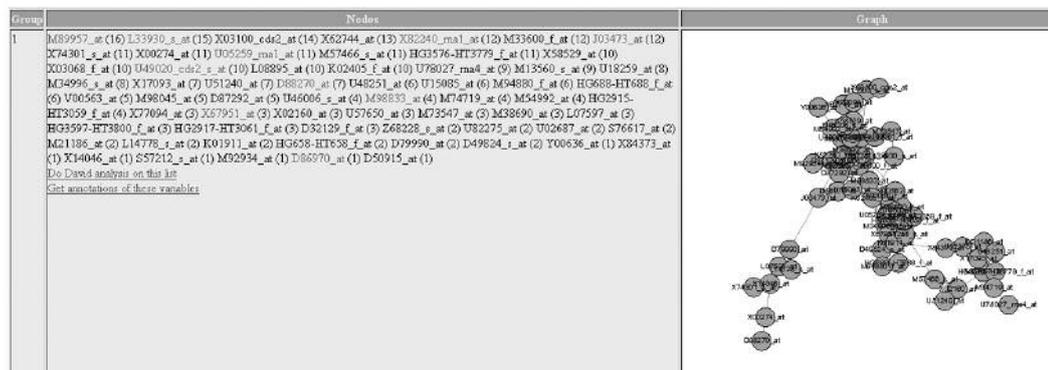


Figure 3 - Exemple de graphes reliant des variables quantitatives fortement corrélées.

CONCLUSION

SmartICA est, à notre connaissance, la seule application web qui permette à un scientifique, non initié dans la manipulation des langages R ou MatLab, de soumettre ces données à une Analyse en Composantes Indépendantes. De plus, SmartICA propose à l'utilisateur plusieurs fonctionnalités qui lui permettront d'identifier efficacement les composants indépendants rendus par cette analyse.

L'application est en libre accès à l'adresse <http://genome.jouy.inra.fr/SmartICA>.

Une documentation complète ainsi qu'un exemple d'analyse pré calculée, sont proposés à l'utilisateur.

Références

[1] Hyvärinen A. et al. (2000) Independent Component Analysis: Algorithms and Applications. Neural Networks, 13(4-5), 411-430

Etude des propriétés écotoxicologiques des solvants dans le cadre de la réglementation REACH

A. Levet¹, C. Bordes¹, J.Y. Gauvrit¹, A. Berthod¹, H. Chermette¹, P. Mignon¹, L. Chancerelle², L. Geoffroy², C. Cren-Olivé³, P. Lantéri¹

¹ Laboratoire des Sciences Analytiques, Bat Curien (CPE), 43, Bd du 11 novembre 1918, 69222 Villeurbanne Cedex,

² INERIS, Parc Technologique Alata BP2, 60 550 Verneuil en Halatte

³ CNRS Service Central d'Analyse, Echangeur de Solaize chemin du canal, 69360 Solaize
amelie.levet@etu.univ-lyon1.fr

Mots clés : Solvants, EC₅₀, classification, modélisation

1- Introduction

Le 1^{er} Juin 2007, le règlement européen REACH (Registration Evaluation Autorisation and restriction of CHemicals) est entré en vigueur. REACH fait porter à l'industrie la responsabilité d'évaluer et de gérer les risques posés par les produits chimiques et de fournir des informations de sécurité adéquates à leurs utilisateurs. Se pose alors le problème de la toxicité des substances et de leur éventuel remplacement.

Nous avons choisi de nous intéresser aux solvants industriels, substances très utilisées aussi bien pour la synthèse que pour la formulation et dont les caractéristiques attendues dépendent de leurs propriétés physico-chimiques qui peuvent en partie être modélisées pour proposer une méthodologie de substitution ^[1].

Du point de vue écotoxicologique, l'un des paramètres pertinent est l'EC₅₀ du composé (qui correspond à la concentration à laquelle 50% de la population mise en contact avec la substance décède) et que l'on doit être capable d'évaluer sur différentes espèces vivantes. Cette évaluation très coûteuse pose des difficultés expérimentales.

Dans ce contexte, nous nous intéressons à la mise en place de modèles de prédiction de l'EC₅₀ pour les solvants. Alors qu'il existe de nombreuses études dans le domaine de la toxicologie classique, peu d'études ont été réalisées dans ce domaine de l'écotoxicologie. Par exemple en 2009, Lee *et al.* ont réalisé un modèle QSAR pour les acides benzoïques reliant l'EC₅₀ d'une micro-algue verte, le log P et le nombre de groupements OH de la molécule^[2]. En 2010, Luis *et al.* ont mis en place un modèle QSAR afin d'estimer l'EC₅₀ d'une bactérie pour les liquides ioniques^[3]. En 2008, Reuschenbach *et al.* ont comparé les données expérimentales de l'EC₅₀ des poissons, algues et daphnies de produits chimiques industriels avec les résultats de modèles QSAR obtenus par le logiciel ECOSAR^[4]. En 2010, dans sa thèse sur les composés nitroaromatiques Guillaume Fayet a mis au point des modèles QSPR pour prédire une propriété physico-chimique : l'explosivité^[5].

2- Démarche

La démarche de ce travail qui nécessite une approche pluridisciplinaire, a consisté, dans un premier temps, à établir une base de données fiable de solvants comportant différents descripteurs. Pour 14 familles chimiques de solvants comme les hydrocarbures, les hydrocarbures halogénés, etc, les descripteurs retenus ont été définis suite à une étude bibliographique de la littérature sur les méthodologies de substitution des solvants. Ces descripteurs peuvent être regroupés en 3 familles : les descripteurs physico-chimiques, les descripteurs obtenus à partir de la chimie théorique et enfin les descripteurs écotoxicologiques.

Dans un deuxième temps, des techniques d'analyse de données (ACP, classification hiérarchique) ont été utilisées pour essayer de proposer une classification des solvants à partir des différents jeux de descripteurs. Le but ultime de ce travail est la mise en relation de descripteurs avec les propriétés écotoxicologiques des solvants par des techniques de modélisation. Nous présentons et discutons ici les tout premiers résultats obtenus dans un premier temps avec des techniques classiques telles que la RLM, la PLS. La figure 1 résume la démarche suivie.

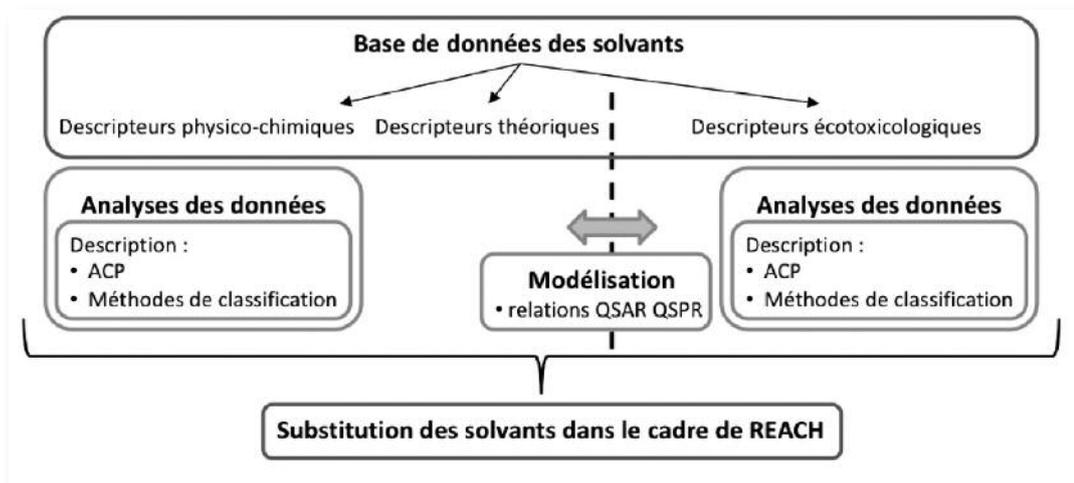


Figure 1 - Stratégie développée

3- Méthodologie et résultats

Les descripteurs physico-chimiques tels que la température d'ébullition, la masse moléculaire, l'indice de réfraction, le paramètre de Hildebrand sont considérés dans la littérature comme des descripteurs importants pour comparer des solvants. Dès 1985 Chastrette *et al.* ont utilisé entre autre le point d'ébullition, l'indice de réfraction et le paramètre de Hildebrand pour réaliser leur classification des solvants^[6]. Un autre paramètre physico-chimique important est le coefficient de partage octanol-eau (log P). Un des moyens utilisés pour l'obtenir avec précision est la chromatographie à contre courant (CCC), technique séparative originale bien développée dans notre laboratoire^[7] qui repose sur l'interaction du soluté entre deux phases liquides ; le log P qui mesure l'affinité du solvant pour l'eau et l'octanol peut donc être obtenu de cette manière avec précision. Nous avons employé cette technique pour valider ou compléter les valeurs relevées dans la littérature.

Les descripteurs théoriques ont eux été obtenus par modélisation grâce à la méthode de la Théorie de la Fonctionnelle de la Densité (DFT)^[8]. Cette dernière permet de calculer les propriétés électroniques des molécules en prenant comme propriété fondamentale la densité électronique.

Les descripteurs du comportement écotoxicologique considérés comme les plus importants par les spécialistes sont les EC₅₀, qu'ils soient mesurés chez les algues, les invertébrés ou les poissons. Compte-tenu de la diversité des méthodes de mesures ainsi que des espèces utilisées, une grande difficulté dans ce type de données est d'évaluer la validité des valeurs annoncées. Cette étape primordiale dans une démarche de modélisation a été réalisée en collaboration avec l'INERIS, la qualité de la prévision finale dépendant fortement de la validité des données utilisées en entrée.

Différentes techniques d'analyse de données (ACP, Classification, Régressions, PLS) ont été appliquées sur cette base de données. Nous présentons ici les premiers résultats obtenu pour établir des groupes de solvants et des relations structures / activité écotoxicologique.

NB : Ce travail s'inscrit dans le cadre des travaux d'Amélie Levet qui est en thèse de doctorat financée par l'ANR (projet NESOREACH 2010-2013).

Références

- [1] Bordes, C., Fréville, V., Ruffin, E., Marote, P., Gauvrit, J.Y., Briangon, S., Lanteri, P. Determination of Poly(ϵ -Caprolactone) solubility parameters: application to solvent substitution in a microencapsulation process, *Int. J. Pharm.* 383, 236-243, 2010
- [2] Lee, P. Y., Chen C. Y. Toxicity and quantitative structure-activity relationships of benzoic acids to *Pseudokirchneriella subcapita*. *J. Hazard. Mater.* 165, 156-161, 2009
- [3] Luis, P., Gare, A., Irabien, A. Quantitative structure-activity (QSARs) to estimate ionic liquids ecotoxicity EC50 (*Vibrio fischeri*). *J. Mol. Liq.* 152, 28-33, 2010
- [4] Reuschenbach, P., Silvani, M., Dammann, M., Warnecke, D., Knacker, T. ECOQAR model performance with a large test set of industrial chemicals. *Chemosphere* 71, 1986-1995, 2008
- [5] G. Fayet. Développement de modèles QSPR pour la prédiction des propriétés d'explosivité des composés nitroaromatiques. Thèse réalisé en 2010 à l'Université Pierre et Marie Curie (Paris)
- [6] Chastrette, M., Rajzmann, M., Chanon, M., Purcell, K.F. Approach to a general classification of solvents using a multivariate statistical treatment of quantitative solvent parameters. *J. Am. Chem. Soc.*, 107, 1, 1-11, 1985
- [7] Berthod, A., Carda-Broch, S., Determination of liquid-liquid coefficients by separation methods. *J. Chromatogr A*, 1037, 3-14, 2004
- [8] Chermette, H. Chemical Reactivity Indexes in Density Functional Theory. *J. Comput. Chem.* 20, 129-154, 1999

VODKA-PLSR: une famille de modèles PLSR dérivés de l'algorithme NIPALS

J.C. Boulet¹, D.Bertrand², G.Mazerolles¹, R.Sabatier³, J.M.Roger⁴

¹ INRA, UMR1083, 2 place Viala, F-34070 Montpellier, bouletjc@supagro.inra.fr

² INRA, Bioinformatique, rue de la Geraudière, F-44136 Nantes, Dominique.Bertrand@nantes.inra.fr

³ UM1, EA2415, 15 av. C.Flahault, F-34093 Montpellier, sabatier@univ-montp1.fr

⁴ CEMAGREF, UMR1201, 361 rue J.F.Breton, F-34093 Montpellier, jean-michel.roger@cemagref.fr

Mots clés : regression, PLSR, NIPALS

Un nouveau modèle de régression basée sur NIPALS est proposé. Il s'appuie sur trois entrées: un jeu d'étalonnage constitué de N spectres sur P variables donnant \mathbf{X} , les valeurs de référence pour une grandeur d'intérêt donnant \mathbf{y} de dimension $N \times 1$, enfin un vecteur \mathbf{r} de dimensions $P \times 1$ choisi arbitrairement. Ces entrées sont utilisées pour le calcul des deux paramètres Σ et \mathbf{P} , avec $\Sigma = (\mathbf{X}'\mathbf{X})^+$ la pseudo-inverse de $(\mathbf{X}'\mathbf{X})$ au sens de Moore-Penrose.

1- Théorie

1.1 Nouvelle écriture de NIPALS

NIPALS utilise (\mathbf{X}, \mathbf{y}) pour calculer itérativement les A colonnes de trois matrices contenant des scores (\mathbf{T}), des vecteurs (\mathbf{P}) et des poids (\mathbf{W}) [3]. Par exemple, soit i un indice entre 1 et A , si nous notons $\mathcal{P}_{1:i}^\perp$ le projecteur de \mathbb{R}^N orthogonal à $\{\mathbf{t}_1 \dots \mathbf{t}_i\}$, à chaque boucle les vecteurs \mathbf{p}_i sont calculés ainsi :

$$\mathbf{p}_i = \mathbf{X}' \mathcal{P}_{1:i-1}^\perp \mathbf{t}_i (\mathbf{t}_i' \mathbf{t}_i)^{-1}$$

Le projecteur $\mathcal{P}_{1:i-1}^\perp$ est inutile puisque \mathbf{t}_i est déjà orthogonal à $\{\mathbf{t}_1 \dots \mathbf{t}_{i-1}\}$, d'où :

$$\mathbf{p}_i = \mathbf{X}' \mathbf{t}_i (\mathbf{t}_i' \mathbf{t}_i)^{-1} \quad (1)$$

Multiplions de chaque coté à gauche par $\mathbf{X}\Sigma$ soit $\mathbf{X}(\mathbf{X}'\mathbf{X})^+$. Le terme $\mathbf{X}(\mathbf{X}'\mathbf{X})^+\mathbf{X}'$ est le projecteur orthogonal sur \mathbf{X} ([2]). Or \mathbf{t}_i est obtenu par combinaison linéaire des colonnes de \mathbf{X} , \mathbf{t}_i appartient obligatoirement à l'espace décrit par les colonnes de \mathbf{X} , donc sa projection sur \mathbf{X} donne \mathbf{t}_i . Ainsi :

$$\mathbf{t}_i = \mathbf{X}\Sigma \mathbf{p}_i (\mathbf{t}_i' \mathbf{t}_i) \quad (2)$$

L'expression de $\mathbf{t}_i' \mathbf{t}_i$ selon \mathbf{P} et Σ s'obtient en calculant $\mathbf{p}_i' \Sigma \mathbf{p}_i$ à partir de l'équation 1. Ce résultat est introduit dans 2 :

$$\mathbf{t}_i' \mathbf{t}_i = (\mathbf{p}_i' \Sigma \mathbf{p}_i)^{-1} \quad (3)$$

$$\mathbf{t}_i = \mathbf{X}\Sigma \mathbf{p}_i (\mathbf{p}_i' \Sigma \mathbf{p}_i)^{-1} \quad (4)$$

On démontre aussi que les \mathbf{p}_i sont strictement orthogonaux entre eux au sens de Σ , soit pour $i \neq j$, $\mathbf{p}_i' \Sigma \mathbf{p}_j = 0$. Donc $\mathbf{P}' \Sigma \mathbf{P}$ est une matrice diagonale dont le terme de la i^{eme} ligne et de la i^{eme} colonne est $\mathbf{p}_i' \Sigma \mathbf{p}_i$. En conséquence l'équation 4 conduit à :

$$\mathbf{T} = \mathbf{X}\Sigma \mathbf{P} (\mathbf{P}' \Sigma \mathbf{P})^{-1} \quad (5)$$

\mathbf{T} est la matrice des scores de la projection Σ -orthogonale de \mathbf{X} sur \mathbf{P} . Les b-coefficients du modèle sont déduits après une régression aux moindres carrés de \mathbf{y} sur \mathbf{T} .

1.2. Calcul de P par NIPALS-P

NIPALS-P est un calcul plus direct des \mathbf{p}_i puisqu'il reste dans \mathbb{R}^P . Transposons chaque terme de l'équation 1 et multiplions à gauche par \mathbf{t}_i , puis remplaçons le \mathbf{t}_i de la partie gauche par sa valeur donnée par l'équation 4 :

$$\mathbf{t}_i \mathbf{p}'_i = \mathbf{t}_i (\mathbf{t}'_i \mathbf{t}_i)^{-1} \mathbf{t}'_i \mathbf{X} \quad (6)$$

$$\mathbf{X} \Sigma \mathbf{p}_i (\mathbf{p}'_i \Sigma \mathbf{p}_i)^{-1} \mathbf{p}'_i = \mathbf{t}_i (\mathbf{t}'_i \mathbf{t}_i)^{-1} \mathbf{t}'_i \mathbf{X} \quad (7)$$

Soit $\mathcal{Q}_{1:i}^\perp = \mathbf{I}_P - \Sigma \mathbf{P}_{1:i} (\mathbf{P}'_{1:i} \Sigma \mathbf{P}_{1:i})^{-1} \mathbf{P}'_{1:i}$ le projecteur Σ -orthogonal à $\{\mathbf{p}_1 \dots \mathbf{p}_i\}$. On démontre à partir de l'équation 7 que :

$$\mathcal{P}_{1:i}^\perp \mathbf{X} = \mathbf{X} \mathcal{Q}_{1:i}^\perp \quad (8)$$

D'autre part, à partir de NIPALS, la substitution de \mathbf{w}_{i+1} dans l'expression de \mathbf{t}_{i+1} puis celle de \mathbf{t}_{i+1} dans l'expression de \mathbf{p}_{i+1} donne :

$$\mathbf{p}_{i+1} = \alpha_{i+1} \mathbf{X}' \mathcal{P}_{1:i}^\perp \mathbf{X} \mathbf{X}' \mathcal{P}_{1:i}^\perp \mathbf{y} \quad (9)$$

avec α_{i+1} un scalaire. Reprenons la formule 9 en remplaçant $\mathcal{P}_{1:i}^\perp \mathbf{X}$ par $\mathbf{X} \mathcal{Q}_{1:i}^\perp$:

$$\mathbf{p}_{i+1} = \alpha_{i+1} \mathcal{Q}'_{1:i} \mathbf{X}' \mathbf{X} \mathcal{Q}_{1:i}^\perp \mathbf{X}' \mathbf{y} \quad (10)$$

Chaque nouveau vecteur \mathbf{p}_{i+1} est à un coefficient près le produit de $\mathbf{X}' \mathbf{X}$ par $\mathbf{X}' \mathbf{y}$ après Σ -orthogonalisation par rapports aux vecteurs $\{\mathbf{p}_1 \dots \mathbf{p}_i\}$ obtenus précédemment.

1.3. Le modèle VODKA-PLSR

VODKA-PLSR propose de remplacer $\mathbf{X}' \mathbf{y}$ par tout autre vecteur de même dimension, et de normer selon Σ . L'algorithme est le suivant.

– A l'étape 1 :

$$\begin{aligned} \mathbf{p}_1 &= \mathbf{X}' \mathbf{X} \mathbf{r} \\ \mathbf{p}_1 &\leftarrow \mathbf{p}_1 (\mathbf{p}'_1 \Sigma \mathbf{p}_1)^{-0.5} \end{aligned}$$

– A l'étape $i + 1$:

$$\begin{aligned} \mathbf{p}_{i+1} &= \mathcal{Q}'_{1:i} \mathbf{X}' \mathbf{X} \mathcal{Q}_{1:i}^\perp \mathbf{r} \\ \mathbf{p}_{i+1} &\leftarrow \mathbf{p}_{i+1} (\mathbf{p}'_{i+1} \Sigma \mathbf{p}_{i+1})^{-0.5} \end{aligned}$$

Le vecteur \mathbf{r} intègre des informations supplémentaires dans le modèle de régression. D'où le nom : Vector Orientation Decided through Knowledge Assessment-Partial Least Square Regression (VODKA-PLSR). Nous retrouvons NIPALS-P lorsque $\mathbf{r} = \mathbf{X}' \mathbf{y}$. Les b-coefficients sont obtenus par une régression aux moindres carrés de \mathbf{y} sur \mathbf{T} . Après simplification :

$$\mathbf{b} = \Sigma \mathbf{P} (\mathbf{P}' \Sigma \mathbf{P})^{-1} \mathbf{P}' \Sigma \mathbf{X}' \mathbf{y} \quad (11)$$

2- Matériels et méthodes

L'objectif était la quantification de l'éthanol en cours de fermentation par spectrométrie entre 500 et 1900nm. Les spectres disponibles étaient :

- \mathbf{X}_G : 165 moûts, ne contenant pas d'éthanol ;
- \mathbf{X} : 315 moûts en fermentation ou vins, pour l'étalonnage ;

- \mathbf{X}_V : 1000 moûts en fermentation ou vins pour la validation ;
- \mathbf{k} , \mathbf{k}_w , \mathbf{k}_g , \mathbf{k}_l : spectres purs de l'éthanol, de l'eau, du glycérol et de l'acide lactique.

Les valeurs de référence \mathbf{y} et \mathbf{y}_V des individus de \mathbf{X} et \mathbf{X}_V étaient connues.

Les 4 premiers vecteurs propres d'une ACP sur \mathbf{X}_G donnent \mathbf{Q} . Une matrice \mathbf{R} est obtenue en concaténant \mathbf{Q} , \mathbf{k}_w , \mathbf{k}_g et \mathbf{k}_l . Le NAS-Net Analyte Signal [1] est la projection de \mathbf{k} orthogonalement à \mathbf{R} . Différents choix ont été faits pour \mathbf{r} : m_1 vecteur composé de la valeur 1 ; m_2 $\mathbf{X}'\mathbf{1}_N$ colinéaire au spectre moyen ; m_3 $\mathbf{X}'\mathbf{y}$ NIPALS ; m_4 \mathbf{k} spectre pur de l'éthanol ; m_5 NAS ; m_6 NIPALS sur données centrées. Les 6 modèles ont été construits sur (\mathbf{X}, \mathbf{y}) avec 1 à 20 variables latentes, leurs prédictions sur $(\mathbf{X}_V, \mathbf{y}_V)$ comparées selon l'erreur standard de prédiction *RMSEP*.

3- Résultats

Les 6 modèles sont différents. Leurs valeurs de *RMSEP* sont reportées dans le tableau 1. Deux modèles : m_2 et m_5 ont de meilleures performances que les modèles NIPALS-P centrés (m_6) ou non (m_3). Le gain de *RMSEP* est certes modeste, environ 3p.cent, l'atout majeur des modèles m_2 et m_5 est de présenter une large plage de variables latentes dans laquelle la qualité de prédiction est stable. Au contraire, pour NIPALS-P, la plage optimum est réduite à une seule variable latente.

Modèle	LV4	LV5	LV6	LV7	LV8	LV9	LV10	LV11	LV12	LV13	LV14	LV15
m_1	2.09	2.30	2.94	1.43	1.12	1.09	1.08	0.99	0.96	0.97	0.96	1.22
m_2	3.16	2.22	2.50	2.23	1.46	0.94	0.93	1.02	0.97	1.01	1.00	1.11
m_3	1.81	1.26	1.04	1.03	1.34	1.02	1.38	1.19	1.08	1.19	1.18	1.16
m_4	2.26	1.93	2.42	1.88	1.21	1.02	1.01	1.02	1.03	1.03	1.02	1.17
m_5	1.04	0.94	0.92	0.92	0.93	0.97	0.99	1.02	1.04	1.04	1.01	1.28
m_6	1.23	1.05	1.00	0.95	1.25	1.02	1.40	1.20	1.11	1.23	1.22	1.21

Tab. 1 - Erreurs standard de prediction (*RMSEP*)

DISCUSSION ET CONCLUSION

Les modèles d'étalonnage inverse utilisent uniquement de l'information expérimentale. Le modèle VODKA-PLSR offre la possibilité de compléter cette information expérimentale en introduisant de l'information experte comme le spectre pur de la grandeur d'intérêt ou le NAS *via* le paramètre \mathbf{r} . Ainsi le calcul de \mathbf{P} est orienté par \mathbf{r} de manière à identifier au mieux l'espace utile de \mathbf{X} .

Plusieurs options pour \mathbf{r} se sont révélées intéressantes, mais beaucoup d'autres choix restent encore possibles. Ainsi VODKA-PLSR ouvre une infinité de modèles de régression de type PLSR issus de NIPALS. Le problème est maintenant d'identifier le vecteur \mathbf{r} optimum. L'utilisation du NAS nous paraît une première piste intéressante puisqu'elle utilise la complémentarité entre informations expérimentales et expertes. Mais d'autres pistes sont envisageables.

Références

- [1] A.Lorber, K.Faber, and B.R.Kowalski. Net analyte signal calculation in multivariate calibration. *Analytical Chemistry*, 69(8) :1620–1626, 1997.
- [2] J.F.Durand. *Eléments de calcul matriciel et d'analyse factorielle de données*. Université Montpellier II, 2002.
- [3] J.Trygg. *Parsimonious multivariate models*. PhD thesis, Umea University, Sweden, 2001.

TRAITEMENT D'IMAGES

CHAIRMEN:

CLAUDE-ALAIN SABY

JEAN-MICHEL ROGER

Discrimination of olive varieties by digital imaging of the kernels and chemometric analysis

P. Vanloot¹, D. Bertrand², C. Pinatel³, N. Dupuy¹, J. Artaud¹

¹CNRS UMR 6263, ISM2, équipe AD2EM groupe Systèmes Chimiques Complexes, case 451, Université Paul Cézanne, Avenue Escadrille Normandie Niémen, 13397 Marseille cedex 20, France, pierre.vanloot@univ-cezanne.fr

²INRA de Nantes, unité BIA, équipe Bioinformatique et Gestion de données, Rue de la Géraudière, 44316 Nantes cedex 03, France

³AFIDOL, Maison des agriculteurs, 22 avenue Henri Pontier, 13626 Aix-en-Provence cedex, France

Keywords: Olive kernels, discrimination, digital imaging, chemometrics.

1- Introduction

The establishment of traceability in the food industry has become essential because they may have health problems and to ensure fair trade. Traceability requires the availability of tools for non-subjective knowledge of all the information needed to track the product from its origin to its use by consumers. The olive oil sector is not immune to this need for traceability.

To determine the origin of varieties of fresh olives, the near or middle infrared spectrum of the pulp or oil is acquired and then introduced into a predictive model of chemometric PLS-DA (Partial Least Squares Discriminant Analysis) [1-3]. In the case of confectionery olives, pulp having undergone a change during the preparation, it is difficult to use the infrared spectrum. The solution, innovative in our study proposed to circumvent this problem is to use the features discriminating olive kernels to get back on the original variety. These characteristics are grouped into three categories of variables: color, shape and texture. These variables are taken from digital photographs of kernels by image processing [4, 5] and used as input variables in the construction and for the routine use of a chemometric model prediction.

2- Material and methods

Kernel samples (n=605) were obtained from the French Inter-Professional Olive Oil Association (AFIDOL), Aix-en-Provence, France. Kernel samples belong to four main French varieties: "Aglandau (Ag), n=242", "Picholine (PI), n=152", "Salonenque (SA) n=92" and "Tanche (TA), n=119"

The acquisition of digital photography of kernels is made using an experimental device comprising: a color camera (Baumer TXD13C, Fujinon HF25HA-1B lens), two neon lights and computer unit with video capture software (Visilog 6.7) (Fig. 1). All olive kernels are photographed in front and profile.

Then, the photographs are processed to extract features more than 100 variables (shape, color and texture), later used for the construction of chemometric models (PLS-DA).

PLS regression can be adapted for pattern recognition, giving rise to the PLS-DA method. PLS-DA is performed using an exclusive binary coding. During the calibration process, the PLS-DA method is trained to compute the "membership values", one for each class; the sample is then assigned to one class when the value is above a specific prediction threshold.

Digital photographs processing is performed by the MATLAB software version 7.8 from MATHWORKS and chemometric applications by the UNSCRAMBLER software version 9.8 from CAMO (Computer Aided Modelling, Trondheim, Norway).

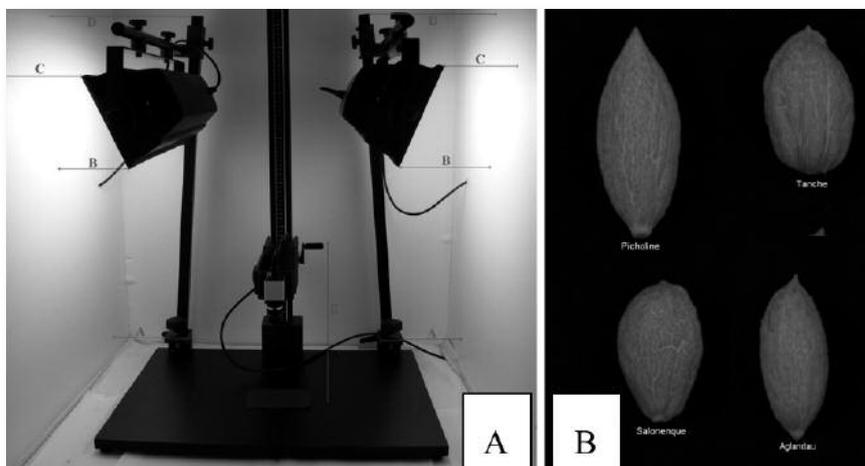


Figure 1 - Experimental device (A, color camera and neon lights) and photographs of kernels (B)

3- 3 Results and discussion

Principal Component Analysis was performed on variables derived from front and profile photographs of kernels (Fig. 2). On PC1 and PC2 axis, "Picholine" and "Tanche" clusters are well separated. "Aglandau" and "Salonenque" are partially separated on PC1 and PC2 axis but are properly separated on PC 3 and PC 4 axis.

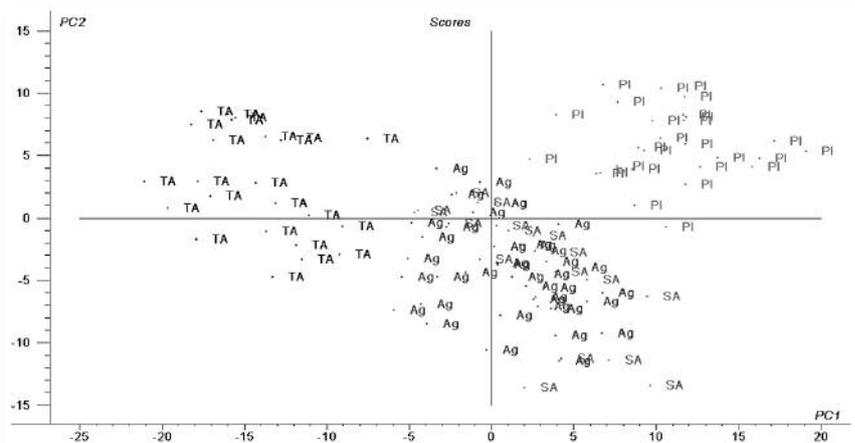


Figure 2 - ACP obtained from front and profile photographs of the kernels

The predicted origins by PLS-DA seldom lead to a binary result exactly equal 0 or 1 but to a result near 0 or 1 which justified by the natural variability of the sample constituents. In front of the difficulty of calibrating and predicting origin with binary variables, it is necessary to discriminate the results between the initial values 0 or 1. Samples with values lower than 0.5 and higher than 1.5 were identified as outside the defined origin and samples with values between 0.5 and 1.5 were identified as belong to the defined origin. For example with these thresholds, Aglandau model (Fig. 3) gives 100% of correct classification. For the three others varieties, 100% of correct classifications are obtained with very strict limits of acceptance (values between 0.6 and 1.2).

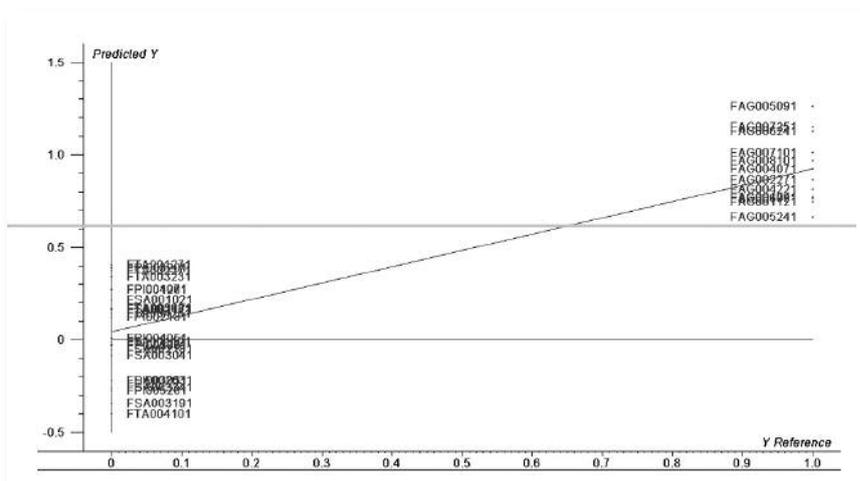


Figure 3 - Prediction of the varietal origin by the Aglandau model from front photographs of the kernels

CONCLUSION

The results of this preliminary study have shown the discriminatory capacity of olive kernels in the recognition the variety origins of olives confectionery. This study is being extended to other French and foreign varieties.

Références

- Dupuy, N., O. Galtier, D. Ollivier, P. Vanloot, and J. Artaud, Comparison between NIR, MIR, concatenated NIR and MIR analysis and hierarchical PLS model. Application to virgin olive oil analysis. *Analytica Chimica Acta*. **666**(1-2): p. 23-31.
- Dupuy, N., O. Galtier, Y.L. Dréau, C. Pinatel, J. Kister, and J. Artaud, Chemometric analysis of combined NIR and MIR spectra to characterize French olives. *European Journal of Lipid Science and Technology*, 2010. **2112**: p. 12.
- Harrington, P.B., J. Kister, J. Artaud, and N. Dupuy, Automated principal component-based orthogonal signal correction applied to fused near infrared- mid infrared spectra of French olive oils. *Analytical Chemistry*, 2009. **81**(17): p. 9.
- Chtioui, Y., D. Bertrand, and D. Barba, Reduction of the size of the learning data in a probabilistic neural network by hierarchical clustering. Application to the discrimination of seeds by artificial vision. *Chemometrics and Intelligent Laboratory Systems*, 1996. **35**(2): p. 175-186.
- Chtioui, Y., D. Bertrand, D. Barba, and Y. Dattee, Application of fuzzy C-Means clustering for seed discrimination by artificial vision. *Chemometrics and Intelligent Laboratory Systems*, 1997. **38**(1): p. 75-87.

Relating hyperspectral images having different spatial resolutions

F. Allouche¹, M.-F. Devaux¹, M. Hanafi², F. Jamme³, F. Guillon¹

1- UR1268 Biopolymères Interactions Assemblages, INRA, F-44300 Nantes, France. fallouche@nantes.inra.fr • 2- ONIRIS, site de la Géraudière. Unité Mixte de recherche Sensométrie et Chimiométrie, 44322 Nantes Cedex 03. mohamed.hanafi@oniris-nantes.fr • 3- Synchrotron SOLEIL, L'Orme des Merisiers Saint-Aubin - BP 48 91192 GIF-sur-YVETTE CEDEX. frederic.jamme@synchrotron-soleil.fr

Keywords: maize cell walls, hyperspectral imaging, fluorescence microspectroscopy, mid infrared microspectroscopy, inter-battery Tucker analysis, PARAFAC, multivariate image analysis.

1- Introduction

About 75% of plant cell walls are polysaccharides polymers that can be theoretically broken down to sugar monomers useful for the production of biofuel. Our project focuses on the investigation of the relationships between maize cell walls composition and degradability. The first step consists in exploring the heterogeneity of composition of cell walls encountered in maize stems.

Two complementary hyperspectral imaging techniques were used to investigate polysaccharides and lignins: infrared and fluorescence microspectroscopies. Coupling spectral data by multi-way analysis could be the key point to heighten differences in composition not observable using only one imaging technique.

Multiblock data treatment has been retained to jointly analyse the data acquired. To achieve the goal, data had firstly to be paired by means of spatial registration techniques. As spatial resolutions differed between the two spectral techniques, the data could not be analysed directly with a classical three-way decomposition model. In order to tackle this problem, two approaches are proposed: inter-battery Tucker and PARAFAC analysis.

2- Material and methods

2.1 Samples

Samples were taken from the middle of maize internodes, embedded into paraffin and sectioned using a microtome with a thickness of 10 μm . Paraffin and starch found into the cells were removed before spectral acquisition^[1]. A region corresponding to a vascular bundle with sclerenchyma and phloem cells near the epidermis was observed.

2.2 Hyperspectral imaging

All hyperspectral images were acquired using synchrotron facilities. Mid-infrared images were acquired using a spectrometer FT-IR Thermo Nicolet 5700 coupled to a Nicolet Continuum XL microscope. The aperture of the initial light was set to 10x10 μm and the step for spatial displacement was 5x5 μm . Fluorescence hyperspectral images of the same section were acquired by means of the synchrotron SOLEIL source on an Olympus IX71 microspectrophotometer equipped with a 40x Ultrafluar Zeiss lens in the UV energy spectral range using an excitation of 280 nm with a step size of 1 μm . Fluorescence spectra were preprocessed to remove noise and to normalize intensities. Second derivative infrared spectra were considered.

2.3 Image registration

An automatic 2D rigid registration of hyperspectral images has been implemented to determine translations, rotation and scaling changes between images. This algorithm is based on the combination of template matching technique with a multi resolution technique^[2].

After image registration, it is possible to get from each spectral image, the spectra acquired at a given location point.

2.4 Data structure

An hyperspectral image is a three way data matrix with two ways corresponding to the spatial coordinates of the pixels and a third way corresponding to the wavelength. The structure of the data was obtained by first unfolding the infrared image. Infrared and fluorescence images have different spatial resolutions: the infrared pixel was 5x5 μm large, each pixel

covering actually an area of $10 \times 10 \mu\text{m}$ and the fluorescence pixel was $1 \times 1 \mu\text{m}$ large. Each infrared spectrum corresponded in fact to a small set of fluorescence spectra (figure 1). The final data structure was a two way infrared data table with a spatial and a spectral dimension that is paired to a three way data table with two spatial dimension and a spectral dimension:

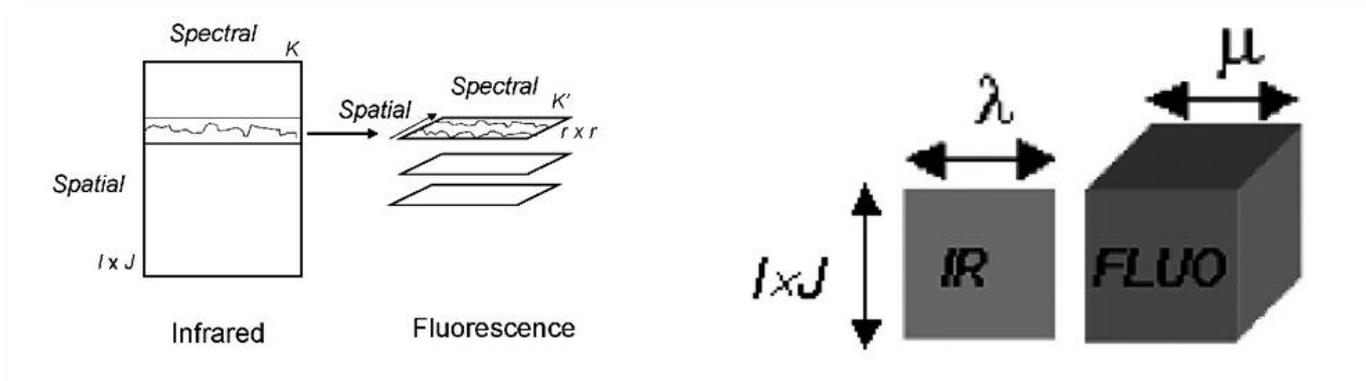


Figure 1 - Data structure

3- Multi-way data analysis

The hyperspectral image spatial dimensions describe the image scene. The third way is the spectral dimension usually considered as variables in multivariate spectral data analysis. In such a situation, it is admitted to ignore the geometry of the image temporarily. Applied on multivariate images, the score matrix can be refolded and displayed as grey-level images. The process of refolding allows interpreting spectral variations according to their spatial origin^[9]. We propose to analyse the data paired using two approaches: inter-battery Tucker and PARAFAC analysis.

3.1 Inter-battery analysis^[4]

As shown in figure 2, fluorescence spectra were averaged to obtain a two way array. The crossing operation is defined by $X \times Y$ where X is the two way infrared data table and Y is the mean of the fluorescence three way data table under the second spatial mode.

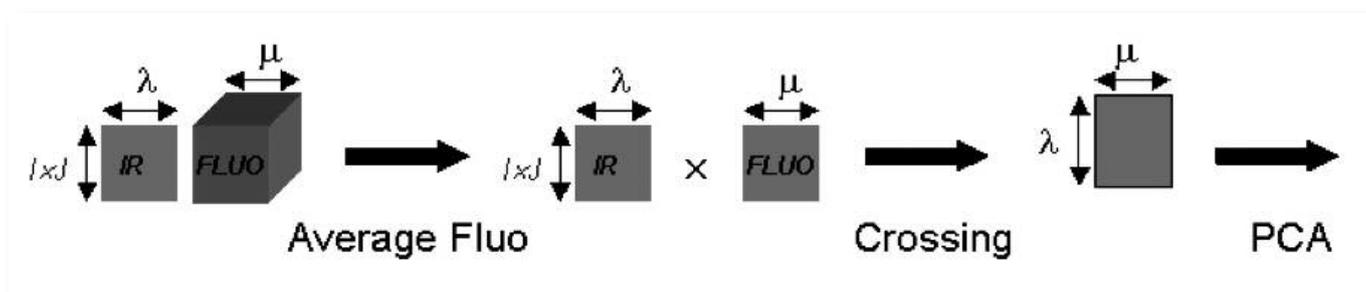


Figure 2 - inter-battery Tucker analysis

Two principal components (Figure 3) corresponding to 81.11% and 8.86% of the total covariance were examined. The first loading revealed intensity variations within the spectra. Principal component 2 could be interpreted as an opposition of pixels with higher lignin content to pixels with higher phenolic acids content. The fluorescence positive peak related to phenolic acids was correlated to the infrared positive peaks which are assigned ester-ether components. The negative peak corresponding to lignin was related to the negative infrared peak relative to xylan and lignin.

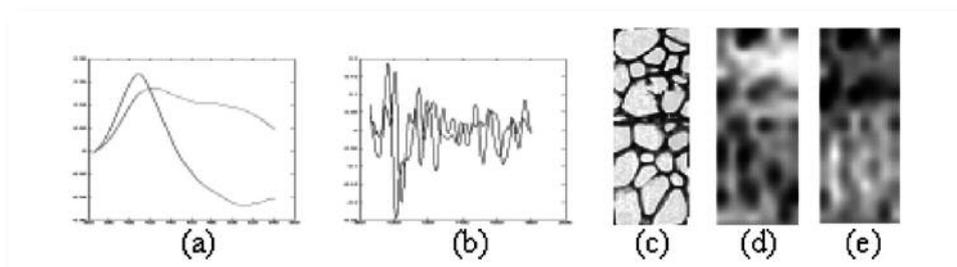


Figure 3 - Scores and loadings from inter-battery analysis: (a) fluorescence loadings (b) infrared loadings, (c) region selected (d) infrared second component score (e) fluorescence second component score.

3.2 PARAFAC

The second strategy to analyse the data was to duplicate the infrared table (figure 4). The three way array obtained after crossing the two three way data tables was decomposed using PARAFAC model^[5].

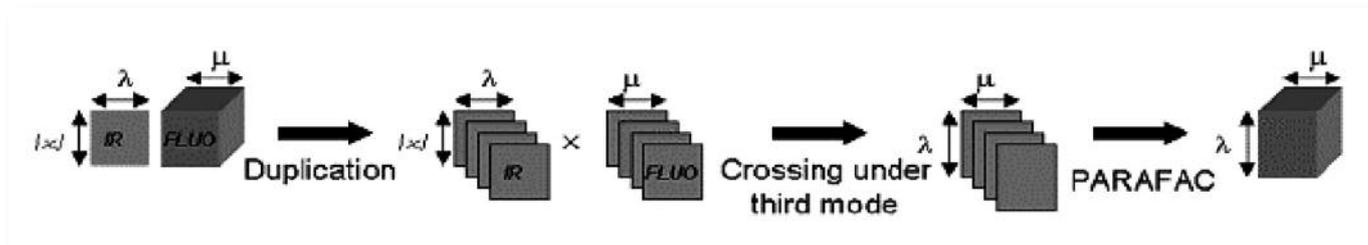


Figure 4 - Building the PARAFAC analysis

Constraints of orthogonality had to be added for the fluorescence mode to solve the problem of redundancy leading to the following results:

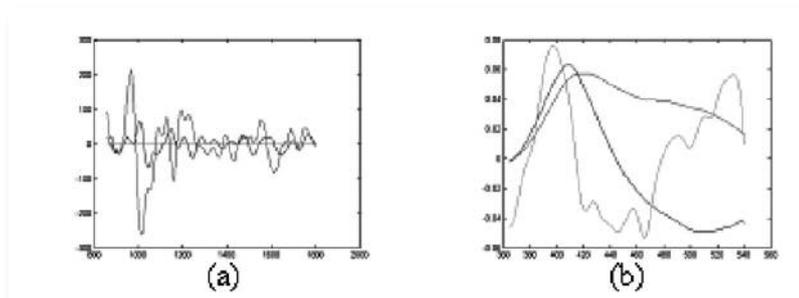


Figure 5 - Scores and loadings from PARAFAC analysis: (a) infrared loadings 1,2 and 3 (b) fluorescence loadings 1,2 and 3.

CONCLUSION

Mid-infrared and fluorescence microspectroscopies were employed in order to analyse polysaccharides, lignin and phenolic compounds found in maize cell walls. The strategy was to couple and link the data acquired without any loss of information provided by both of the spectroscopic techniques. The data were analysed using inter-battery Tucker analysis and PARAFAC model. Constraints had to be imposed to the PARAFAC model to eliminate redundancy in the loadings calculated.

The approach used is generic, and the work will be extended by analysing jointly three modalities. Our further work is to use Raman microspectroscopy as a third modality to distinguish other chemical components not observable by infrared and fluorescence microspectroscopies.

Références

- [1] Jamme F., Robert R., Bouchet B., Saulinier L., Dumas P., Guillon F., 2008, Aleurone cell walls of wheat grain: high spatial resolution investigation using synchrotron infrared microspectroscopy. *Applied Spectroscopy* 62 (8), 895-900.
- [2] Barbara zitova and Flusser J., *Image registration methods a survey*, Image and vision computing, 2003.
- [3] Pedersen F., *Interactive explorative analysis of multivariate images using principal components*, 1994.
- [4] Tucker, L. R. , *An inter-battery method of factor analysis*. *Psychometrika*, 1958, 23, 11-136.
- [5] Smilde Age K., Rasmus B., Geladi P. , *Multi-way analysis with applications in the chemical sciences*, 2004, 381 pages.

Increasing the spatial resolution in vibrational microscopic imaging using post-processing algorithms

L. Duponchel, M. Offroy

Laboratoire de Spectrochimie Infrarouge et Raman, LASIR, CNRS UMR 8516, Bât. C5, Université Lille I, 59655 Villeneuve d'Ascq Cedex, France.

Keywords: Spectroscopic imaging, Infrared, Raman, Super-resolution.

1- Introduction

Fourier Transformed Infrared Spectroscopic imaging is a valuable tool in order to obtain information concerning the molecular distribution of various compounds in complex samples. FTIR spectrometers are thus coupled with microscopes to enable the spectral analysis of physically small samples or even to focus on specific regions of these. The spatial resolution of a microscope is ideally determined by diffraction of the selected radiation. When a point source monochromatic radiation is passed through the microscope, the well-known Airy pattern is observed at the beam focus. The central circular area, the Airy disc has thus a radius r given by:

$$r=0.61.(\lambda/NA) \quad (1)$$

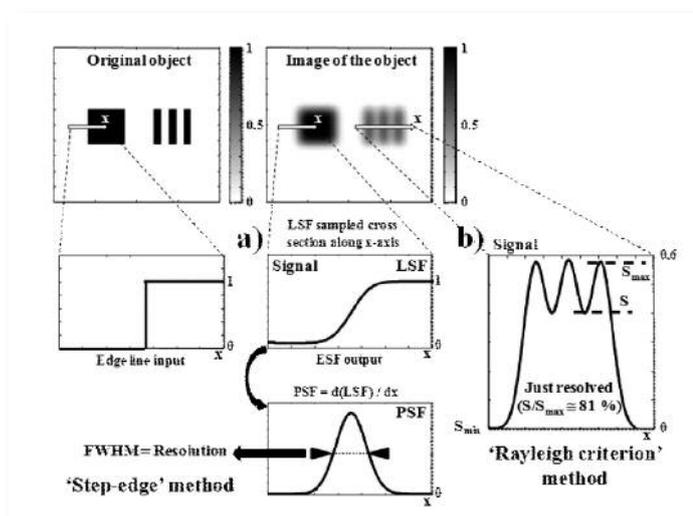
with λ the wavelength of the radiation and NA the numerical aperture of the microscope. Considering the previous equation, two objects are completely resolved if they are separated by $2r$ and hardly resolved with r (Rayleigh criterion of resolution). In this way, one can consider that the diffraction-limited spatial resolution is more or less equal to the wavelength of the light when conventional Cassegrainian objectives are used in the air. Unfortunately, this relation is never observed for many reasons. First, the detector is not a point and has a finite size. The pixel detector becomes a sort of aperture. When the pixel size of the sample plane is higher than the wavelength, the system is said to be 'pixel size-limited' [i]. In these conditions, the spatial resolution is limited by pixel size and less by diffraction phenomenon. Second, a low axial spatial resolution (z-direction, depth of the sample) can degrade the achievable optimal lateral spatial resolution. Third, optical setup of the microscope has a direct influence on the spatial resolution. The number of aperture in the optical pathway is an example [ii, iii]. For all the above reasons, it is often impossible to find and use a theoretical model in order to evaluate the spatial resolution of a micro-spectrometer particularly when Focal Plane Array detectors are used. The preferred approach is to propose a direct estimation based on real measurements for the considered instrumental setup.

Even if the evaluation of spatial resolution is not straightforward, the main issue comes from its intrinsic value observed in mid infrared spectral range (2.5-25 μm). Considering few microns-sized samples, the spatial resolution of far-field infrared imaging spectrometers is always too low in order to observe many details. The aim of the presented work is thus to propose the super-resolution concept in order to increase the spatial resolution of infrared imaging spectrometers using FPA detectors. Super-resolution is defined by the use of image processing algorithms in order to overcome the limitations of optical systems [iv, v, vi]. The main idea of super-resolution is the fusion of several low-resolution images (LR) of the same object to obtain a high-resolution image (HR) [vii, viii, ix]. The first paper dealing with recovering high resolution image using multi-frames was written by Tsai and Huang [x]. It was shown that fusing under-sampled low resolution images with relative sub-pixel motion could give a super-resolution image with few or no aliasing effects.

2- Experimental

2.1 The spatial resolution evaluation method

We use the so-called 'step-edge' or the 'knife-edge' method to evaluate the spatial resolution in the presented work [xi]. The main idea is to observe a step on the considered sample. Considering the knife-edge as the input to an optical system, then the signal along the perpendicular direction to the edge is the Line Spread Function (LSF) (Figure 1a). The derivative



of the LSF with respect to position is the Point Spread Function (PSF) in the considered direction. The PSF is the response of the system to a point source observation. The latter function is then used as a measure of resolution by calculating its full width at half maximum (FWHM).

Figure 1

2.2 Materials

In order to evaluate the spatial resolution and explore the super-resolution concept, all spectroscopic measurements were done on the 1951 USAF resolution target. This chart is a resolution test pattern conformed to the MIL-STD-150A standard, set by US Air Force in 1951. The resolution target has chrome metal coating pattern on soda-lime glass. It is an appropriated sample for the infrared imaging experiment: contrasted images are obtained due to the high reflectivity of metal and the high absorbance of glass. The pattern consists of groups of three bars with decreasing dimensions going to $0.78 \mu\text{m}$ (sub-wavelength size).

2.3 FT-IR measurements

An Equinox 55 spectrometer (Bruker, Ettlingen, Germany) coupled with a Hyperion 3000 microscope equipped with a 64×64 Mercury-Cadmium-Telluride Focal Plane Array (MCT-FPA) detector were used to acquire the IR reflectance spectra between 3900 and 900 cm^{-1} at 32 cm^{-1} resolution under N_2 purge. After signal integration at a defined wavenumber from reflectance spectra, it was thus possible to obtain an image of 64×64 pixels. The number of scans was 5 and the surface analyzed by one FPA measurement was a $270 \mu\text{m} \times 270 \mu\text{m}$ area due to the x15 Cassegrainian objective ($\text{NA}=0.45$). The pixel size of the sample plane was thus $4.22 \mu\text{m} \times 4.22 \mu\text{m}$ on raw FPA measurements. Between two FPA measurements, it was possible to move the sample with a XY-motorized stage with a repeatability of $1 \mu\text{m}$ in position.

2.4 The super-resolution concept

The aim of this part is to describe a global framework of the super-resolution concept. More detailed procedures can be found in numerous papers [iv,v,ix]. As in image processing papers, images are represented column-wise lexicographically

ordered for matrix notation convenience. Given are N measured low resolution images $\{\underline{Y}_k\}_{k=1}^N$, where each image is defined by $M \times M$ pixels and represented by the unfolded matrix \underline{Y}_k of size $[M^2 \times 1]$.

These LR images are different representations of a single HR image ($L \times L$ pixels) represented by the unfolded matrix \underline{X} of size $[L^2 \times 1]$, where $L > M$ for $1 \leq k \leq N$. Considering 16 FPA measurements for our experiment, $N = 16$, $M = 64$ and $L = 256$. In fact, it can be considered that each LR image measured is the result of a particular geometric warping (translations in our microspectrometric case), linear space-invariant blurring and uniform rational decimation performed on the ideal HR image \underline{X} . Additive Gaussian noise is also considered corrupting the low resolution images. It is thus possible to propose an analytical model (eq. 2), in order to express the steps described previously: $\underline{Y}_k = D_k H_k F_k \underline{X} + \underline{V}_k$

The matrix F_k of size $[L^2 \times L^2]$ corresponds to the geometric warp operation between the \underline{X} image and \underline{Y}_k . The matrix H_k of size $[L^2 \times L^2]$ is the blur matrix defined by the optical system's Point Spread Function. The matrix D_k of size $[M^2 \times L^2]$ corresponds to the decimation resulting in \underline{Y}_k . In other words, this step corresponds to the reduction of the number of observed pixels

in the measured images. The additive Gaussian noise observed in the k -th measurement is described by the vectors $\{\underline{V}_k\}_{k=1}^N$ with zero mean. In this way, super-resolution is an inverse problem since \underline{X} has to be found from known low resolution

images $\{\underline{Y}_k\}_{k=1}^N$. It is necessary to evaluate D_k , H_k and F_k matrices for all $k = 1 \dots N$ in order to resolve (eq. 4). For our

spectroscopic experiment F_k are just defined by the motion between low resolution images $\{\underline{Y}_k\}_{k=1}^N$ and one of the low resolution images (such as, say, \underline{Y}_1) chosen as a reference. These motions are micron (sub-pixelic) translations induced by the XY stage between two FPA measurements. Concerning H_k blur matrices, we consider that all low resolution images are obtained with the same optical system and thus viewed through the same PSF that is to say $\forall k, H_k = H$. H is usually estimated from a spatial resolution evaluation but even a rough guess of it like a Gaussian filter can be used. The matrix D_k depends only on the decimation ratio between the HR image and the measured LR image i.e. the ratio between the number of pixel in the LR image M^2 and the HR image L^2 . Considering a generalization of the analytical model (eq. 3) applied on all N given LR images we have:

$$\begin{bmatrix} \underline{Y}_1 \\ \vdots \\ \underline{Y}_N \end{bmatrix} = \begin{bmatrix} \underline{D}_1 \underline{H}_1 \underline{F}_1 \\ \vdots \\ \underline{D}_N \underline{H}_N \underline{F}_N \end{bmatrix} \underline{X} + \begin{bmatrix} \underline{V}_1 \\ \vdots \\ \underline{V}_N \end{bmatrix}$$

It is now possible to retrieve the HR image \underline{X} with many classical restoration methods like Maximum A Posteriori estimation [xii, xiii, xiv]. Due to the high dimension of the equation system (eq. 3) and inherent numerical instability of such inverse problems, regularization procedure are also necessary in order to find a stable solution of \underline{X} . We decide to use the "Norm 1 + Bilateral Total Variation" [iv,v] algorithm to retrieve the super-resolved image in this work.

3- Results and discussion

The super-resolved images were obtained by moving the sample with a micron step (sub-pixelic moves) in the x and y directions between FPA measurements. Figure 2 presents the application of the super-resolution concept on 16 raw FPA images at 3600 cm⁻¹. This global number of FPA images was optimized in order to obtain a better spatial resolution within a reasonable time for all spectroscopic analysis. First we notice visually the uncontested improvement of the spatial resolution between the original FPA and the super-resolved image. Moreover, the 'step-edge' method applied

to this new image gives a resolution of 6.09 μm compared to 9.13 μm at the same wavenumber. Figure 3a and 3b present the evaluation of the spatial resolution of super-resolved images at different wavenumbers in x direction compared to raw images results. We observe a 30% improvement for the spatial resolution for all wavenumbers which is a rather spectacular result. It can be also noticed that the dispersion on the evaluation of the resolution is less important for super-resolved images than for raw FPA images at a given wavenumber. This is mainly due to a higher image signal to noise ratio obtained with the super-resolution algorithm.

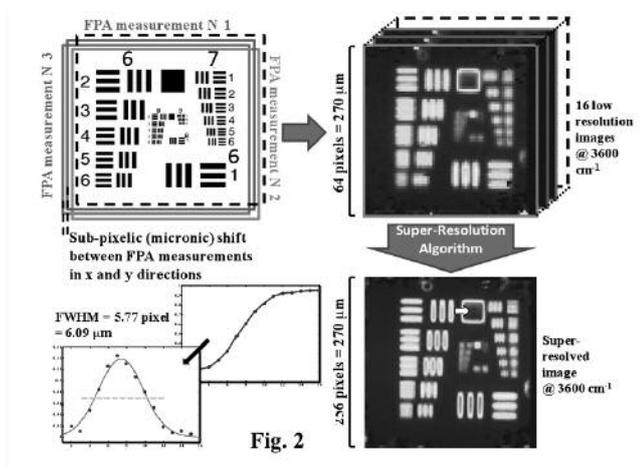


Figure 2

Figure 3a

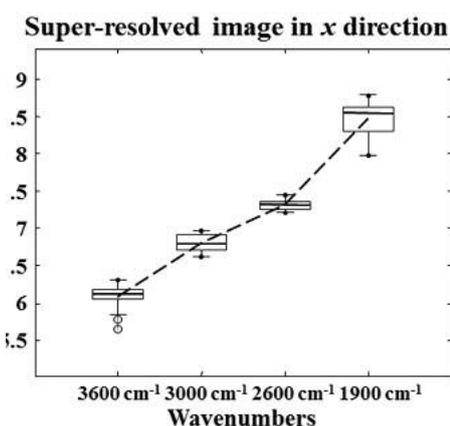
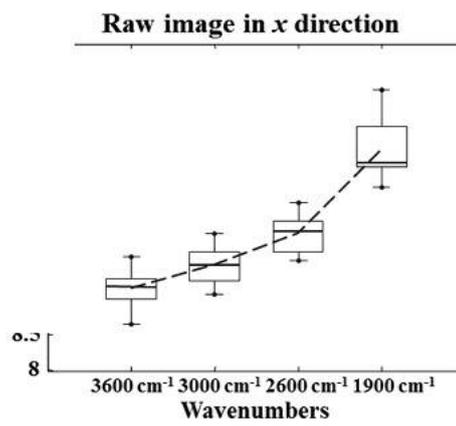


Figure 3b



CONCLUSION

This work presents the first attempt to use super-resolution methodology applied to FTIR imaging spectroscopy. Applying the super-resolution concept on a relatively low number of FPA acquisitions, it was possible to observe a 30 % improvement in the spatial resolution. It is a very important result since FPA imaging spectrometers have in general a rather low spatial resolution. Future works will be focused on the study of factors affecting the performance of multi-frame super-resolution restoration. It will be interesting to apply super-resolution on another FPA spectrometer with a higher density pixel and a higher magnification objective for which the spatial resolution is more dictated by wavelength-limited diffraction than the pixels size. Another challenge will be to use super-resolution on time resolved FPA measurements in order to propose chemical images over time with a higher spatial resolution. Finally, super-resolution has to be tested on infrared spectrometers with single element detector in order to know if better sub-wavelength imaging is possible.

Références

- [i] R. Salzer, H.W. Siesler, *Infrared and Raman spectroscopic imaging*, Wiley-VCH (2009).
- [ii] A.J. Sommer, J.E. Katon, *Appl. Spectrosc.* 45 (1991) 1633.
- [iii] N. Teramae, T. Nishikoda, *Infrared Microscopy*, IPC Publisher, Tokyo (2003).
- [iv] S. Farsiu, M.D. Robinson, M. Elad, P. Milanfar, *IEEE trans. image process.*, 13 (2004), 1327.
- [v] S. Farsiu, D. Robinson, M. Elad, P. Milanfar, *Int. j. imaging syst. technol.*, 14 (2004) 47.
- [vi] M.K. Ng, N.K. Bose, *IEEE signal process. mag.*, 20 (2003) 62.
- [vii] S.C. Park, M.K Park, M.G. Kang, *IEEE signal process. mag.*, 20 (2003) 21.
- [viii] M.M. Hadhoud, F. Abd El-Samie, S.E. El-Khamy, *Proc. IEEE Conf. 4th Workshop Photon. and Appl.*, (2004) 2.
- [ix] N. Nguyen, P. Milanfar, G. Golub, *IEEE trans. image process.*, 10 (2001) 573.
- [x] R.Y. Tsai, T.S. Huang, *Advances in Computer Vision and Image Processing*, JAI Press, 1984.
- [xi] E. Levenson, P. Lerch, M. Martin, *Infrared Phys. Techn.* 49 (2006) 45.
- [xii] M. Elad, Y. Hel-Or, *IEEE trans. image process.*, 10 (2001) 1187.
- [xiii] R.C. Hardie, K.J. Barnard, E.E. Armstrong, *IEEE trans. image process.*, 6 (1997) 1621.
- [xiv] R.R. Schultz, R.L. Stevenson, *IEEE trans. image process.*, 3 (1994) 233.

MÉTHODES SÉPARATIVES MULTIDIMENSIONNELLES

CHAIRMEN:
SERGE RUDAZ
MAX FEINBERG

Comparaison de différentes approches statistiques pour évaluer l'orthogonalité de séparations chromatographiques : application aux systèmes en phase inverse

Ramia AL BAKAIN^a, Isabelle RIVALS^b, Patrick SASSIAT^a, Didier THIEBAUT^a,
Marie-Claire HENNION^a, Guillaume EUVRARD^b, Jérôme VIAL^a

^a Laboratoire des Sciences Analytiques, Bioanalytiques et Miniaturisation UMR- CNRS –UPMC- PECSA,
ESPCI ParisTech, 10 rue Vauquelin, 75005 Paris, France

^b Équipe de Statistique Appliquée, ESPCI ParisTech, 10 rue Vauquelin, 75005 Paris, France

ramia.bakain@espci.fr, isabelle.rivals@espci.fr, patrick.sassiat@espci.fr, didier.thiebaut@espci.fr, marie-claire.hennion@espci.fr,
guillaume.euvrard@espci.fr, jerome.vial@espci.fr

Mots-clé : RPLC ; orthogonalité ; ANOVA ; théorie de l'information ; tests statistiques d'indépendance.

1- Introduction

La capacité de séparation est une préoccupation majeure lors de la conception de systèmes chromatographiques destinés à la caractérisation de mélanges complexes. Il est possible de l'augmenter en couplant en ligne deux systèmes chromatographiques dont les mécanismes de rétention sont différents, ou « orthogonaux », et en créant ainsi un système bidimensionnel (2D). En se limitant à des mécanismes de séparation en « phase inverse » (solution favorable dans l'optique d'un couplage en ligne), on peut jouer pour chacun des deux systèmes sur trois facteurs : la phase stationnaire, le pH de la phase mobile et le modificateur organique. Plusieurs critères peuvent être utilisés pour évaluer l'impact de ces facteurs sur le degré d'orthogonalité^[1-9].

Expérimentalement, 32 systèmes de chromatographie en phase liquide ont été évalués par le même gradient générique avec un ensemble de 63 solutés couvrant une large gamme de propriétés physico-chimiques (acido-basicité, pKa, poids moléculaire, hydrophobie). Chaque système chromatographique a été défini par la nature de sa phase stationnaire (huit différentes), le pH de la fraction aqueuse de la phase mobile (2.5 ou 7.0) et la nature du modificateur organique (acétonitrile ou méthanol). Les 496 couples de systèmes chromatographiques ont été classés suivant leur degré d'orthogonalité à l'aide de huit critères différents : les trois coefficients de corrélation (Pearson, Spearman et Kendall), deux critères géométriques caractérisant la zone de couverture de la séparation 2D, le pourcentage de synentropie de Slonecker, et deux statistiques de χ^2 d'indépendance entre les temps de rétention. Une analyse de variance (ANOVA) a été utilisée pour évaluer la capacité de chaque critère à mettre en évidence l'importance des trois facteurs étudiés : différence ou non des phases stationnaires des deux systèmes du couple, pH différents ou non, et modificateurs organiques différents ou non.

2- Résultats et discussion

Les huit critères portent donc sur les valeurs des temps de rétention des 63 solutés pour chaque couple de systèmes, c'est-à-dire leur répartition dans l'espace 2D. Les critères fondés sur les trois coefficients de corrélation sont notés « r_p », « r_s » et « r_k ». En approchant la densité 2D par une gaussienne, nous avons évalué l'aire d'une ellipse de confiance à 95% : plus cette aire est grande, plus le couple est orthogonal (critère que nous noterons « conf »). Les quatre autres critères demandent une discrétisation de l'espace de séparation (10x10 dans le Tableau 2). Les critères fondés sur les statistiques de χ^2 sont notés « d^2 » (statistique de Pearson), et « g^2 » (statistique du maximum de vraisemblance). On peut montrer que cette dernière conduit à une évaluation de l'orthogonalité équivalente à celle de la similarité informationnelle introduite par Slonecker^[2]. Afin de compléter la comparaison, nous avons également évalué le second indicateur de Slonecker, le pourcentage de synentropie (noté « %SS »), qui est le pourcentage de l'entropie totale apporté par la diagonale. Enfin, nous avons construit un autre critère géométrique très simple, le taux d'occupation du quadrillage, noté « %occ ».

À l'aide de chacun de ces huit critères, nous avons classé les 496 couples par ordre d'orthogonalité croissante. Chaque couple a été caractérisé par un score dans l'intervalle [0, 1], le couple le moins orthogonal ayant donc un score de 0, et le plus orthogonal un score de 1. Pour chaque critère, nous avons réalisé une ANOVA préliminaire avec comme facteur influant

sur l'orthogonalité, donc sur le score, la différence de phases stationnaires ($\varphi \neq$ ou $\varphi =$), la différence des pH ($\text{pH} \neq$ ou $\text{pH} =$), et la différence des modificateurs organiques ($\text{mo} \neq$ ou $\text{mo} =$). Seuls les effets directs et celui de l'interaction {différence de pH} x {différence de phase stationnaire} étant significatifs, nous avons reparamétrisé l'ANOVA suivant le Tableau 1. Les résultats de l'ANOVA avec cette paramétrisation sont rassemblés dans le Tableau 2.

	om =		om \neq	
	S φ =	S $\varphi \neq$	S φ =	S $\varphi \neq$
pH =	m	m + b=	m + c	m + b= + c
pH \neq	m + a	m + a + b \neq	m + a + c	m + a + b \neq + c

Tableau 1 - Paramétrisation de l'ANOVA. m : score moyen pour les trois caractéristiques identiques, a : augmentation moyenne du score due à la différence des pH, b= : l'augmentation moyenne du score due à la différence des phases stationnaires à pH identique, b \neq : l'augmentation moyenne du score due à la différence des phases stationnaires à pH différent, c : augmentation moyenne du score due à la différence des modificateurs organiques, quels que soient les phases stationnaires et les pH.

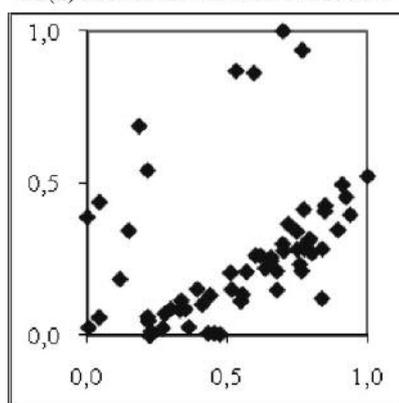
	P_a	$P_{b=}$	$P_{b\neq}$	P_c	a	b=	b \neq	c	a+b \neq +c
r_P	3.5 10⁻⁸	9.0 10⁻⁸	0.015	0.22	0.45	0.37	0.12	0.03	0.60
r_S	2.0 10⁻¹²	1.2 10⁻⁷	0.052	0.13	0.52	0.34	0.09	0.03	0.66
r_K	1.3 10⁻¹²	1.9 10⁻⁷	0.0035	0.039	0.52	0.32	0.13	0.06	0.70
conf	5.1 10⁻¹¹	8.4 10⁻⁸	0.015	0.079	0.49	0.35	0.11	0.04	0.66
%occ	9.8 10⁻⁷	5.3 10⁻⁵	0.02	0.0028	0.40	0.28	0.11	0.07	0.58
%SS	6.9 10⁻⁷	3.0 10⁻⁷	0.12	0.017	0.39	0.37	0.08	0.06	0.57
d^2	9.3 10⁻⁶	2.3 10⁻⁷	4 10⁻⁶	0.0022	0.35	0.35	0.22	0.07	0.65
g^2	6.4 10⁻⁵	9.5 10⁻⁵	0.0053	0.0076	0.33	0.28	0.14	0.07	0.54

Tableau 2 - Valeurs des paramètres de l'ANOVA et p-valeurs associées (en gras les p-valeurs < 0,05). Les critères de classement sont : r_P , r_S , r_K : coefficients de corrélation respectivement de Pearson, Spearman, et Kendall ; conf : ellipse de confiance ; %occ : pourcentage d'occupation de l'espace de séparation ; %SS : pourcentage de synentropie de Slonecker ; d^2 : statistique de χ^2 de Pearson ; g^2 : statistique de χ^2 du maximum de vraisemblance.

Le coefficient de corrélation de Kendall apparaît comme le critère le plus pertinent pour mesurer l'orthogonalité. En effet, selon les résultats de l'ANOVA, il présente la plus grande sensibilité à tous les facteurs expérimentaux. Les statistiques du χ^2 sont également pertinentes, mais à condition que la discrétisation de l'espace de séparation ait été effectuée de manière appropriée, c'est-à-dire en ajustant les frontières du quadrillage pour que le nombre de points par case ne soit pas trop faible (condition à laquelle la convergence des statistiques vers le χ^2 est atteinte). Par ailleurs, suivant tous les critères, les résultats de l'ANOVA permettent de conclure que c'est le pH de la phase mobile qui a le plus d'impact sur la sélectivité, suivi par le type de la phase stationnaire et enfin par la nature du modificateur organique.

La Figure 1 montre les deux couples de systèmes les plus orthogonaux au sens du coefficient de corrélation de Kendall ($r_K = 0.42$). Pour ces deux cas, les systèmes considérés diffèrent par leurs trois caractéristiques : leur pH, leur phase stationnaire, et leur modificateur organique.

C8(2) MeOH 2.5 vs. PEG MeCN 7.0



Kromasil C18 MeOH 2.5 vs. PEG MeCN 7.0

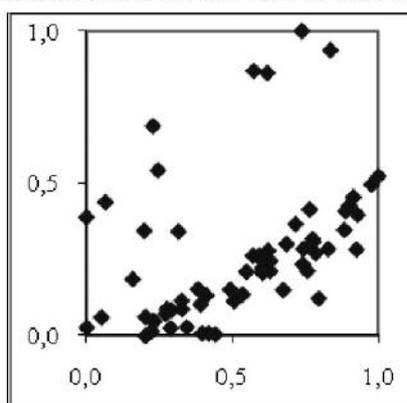


Figure 1 - Temps de rétention normalisés pour les couples les plus orthogonaux ($r_K = 0.42$) parmi les 496.

CONCLUSION

Le coefficient de corrélation de Kendall apparaît comme le critère à la fois le plus pertinent et le plus facile d'utilisation pour mesurer le degré d'orthogonalité de couples de systèmes de chromatographie en phase liquide.

En ce qui concerne les facteurs qui influencent l'orthogonalité, comme l'on pouvait s'y attendre, le plus grand degré d'orthogonalité s'obtient lorsque les trois caractéristiques des deux systèmes diffèrent. Mais l'analyse de la variance nous a permis de conclure que c'est la différence de pH des phases mobiles qui a le plus d'effet sur l'orthogonalité, suivie de celle des phases stationnaires, et enfin de celle des modificateurs organiques.

Nous avons vérifié que la hiérarchie de l'effet des facteurs est conservée même lorsque l'on modifie le jeu de composés tests, notamment en écartant les solutés ionisés.

Références

- [1] Z. Liu, D.G. Patterson, M.L. Lee, *Anal. Chem.* 67 (1995) 3840.
- [2] P.J. Slonecker, X. Li, T.H. Ridgway, J.G. Dorsey, *Anal. Chem.* 68 (1996) 682.
- [3] M. Gray, G.R. Dennis, P. Wormell, R. Andrew Shalliker, P. Slonecker, *J. Chromatogr. A* 975 (2002) 285.
- [4] M. Gilar, P. Olivova, A.E. Daly, J.C. Gebler, *Anal. Chem.* 77 (2005) 6426.
- [5] M. Gilar, P. Olivova, A.E. Daly, J.C. Gebler, *J. Sep. Sci.* 28 (2005) 1694.
- [6] E. Van Gyseghem, I. Crosiers, S. Gourvénec, D.L. Massart, Y. Vander Heyden, *J. Chromatogr. A* 1026 (2004) 117.
- [7] E. Van Gyseghem, M. Jimidar, R. Sneyers, D. Redlich, E. Verhoeven, D.L. Massart, Y. Vander Heyden, *J. Chromatogr. A* 1074 (2005) 117.
- [8] L.R. Snyder, J.W. Dolan, P.W. Carr, *J. Chromatogr. A* 1060 (2004) 77.
- [9] J.W. Dolan, L.R. Snyder, *J. Chromatogr. A* 1216 (2009) 3467.

Stratégie chimiométrique pour l'analyse de chromatogrammes bidimensionnels

Isabelle Rivals^a, Jérôme Vial^b, Didier Thiébaud^b, Patrick Sassi^b, Béatrice Teillet^c, Xavier Cahours^c

^aÉquipe de Statistique Appliquée, ESPCI ParisTech, 10 rue Vauquelin, 75005 Paris, France.

^bLaboratoire Sciences Analytiques, Bioanalytiques et Miniaturisation, ESPCI ParisTech, UMR CNRS UPMC PECSA, 10 rue Vauquelin, 75005 Paris, France.

^cSEITA, Imperial Tobacco Group, Science & Stewardship, Centre de Recherche, 4 rue A. Dessaux, 45 404 Fleury les Aubrais, France.
Adresse e-mail : isabelle.rivals@espci.fr

■ L'exploitation de chromatogrammes bidimensionnels (2D) comportant des milliers de pics ne peut raisonnablement s'envisager sans des outils chimiométriques adaptés à cette complexité. Nous avons ainsi développé une stratégie générale de comparaison automatique de chromatogrammes 2D conduisant à l'identification des composés dont la concentration diffère significativement entre les diverses classes d'échantillons de composition voisine. Après une série de prétraitements (redressement de ligne de base, alignement temporel par Dynamic Time Warping contraint) rendus nécessaires par les variations fortuites des conditions expérimentales, cette stratégie s'appuie sur la sélection des pixels les plus discriminants pour chaque classe d'échantillons grâce à l'établissement de cartes de corrélation des pixels avec les différentes classes. Les pics 2D auxquels appartiennent ces pixels sont alors identifiés par un algorithme original de reconnaissance de pics. Par recoupement avec les informations de spectrométrie de masse, il est alors envisageable non seulement d'identifier les composés à l'origine de la discrimination, mais aussi d'évaluer par ANOVA l'impact de différents facteurs (variabilité injection à injection, extraction à extraction, différence d'origine, etc.) sur leur concentration. Le potentiel de cette stratégie, sera illustré sur plusieurs jeux de données de chromatogrammes d'extraits de trois variétés de tabacs, pour lesquelles nous disposons de lots d'origines différentes.



**APPLICATIONS
INDUSTRIELLES**

CHAIRMAN:
LUDOVIC DUPONCHEL

Utilising technology to simplify process optimisation, improvement and control

Dr. Malcolm Moore, Frank Baudet

malcolm.moore@jmp.com, frank.baudet@jmp.com

Keywords: Methods, QbD, PAT, optimization, DOE,

1- Introduction

This presentation will examine ways of integrating and simplifying QbD, PAT, Six Sigma and Lean approaches to enable mainstream scientists and engineers to efficiently and effectively optimise, improve, and control new and existing processes and formulations. The effectiveness of this integrated and simplified approach will be illustrated through a variety of case studies using JMP software from SAS.

2- Material and methods

Scientists and engineers throughout the world need to be learning smarter and faster in order to enable faster delivery of new products to market, gain bigger market share and reduced costs. Many initiatives are supposed to help with achieving these objectives, but their numbers can sometimes be confusing. For example, Process Analytical Technology (PAT) is meant to encourage a shift towards learning and continuous improvement in manufacturing. In practice what has been observed is that PAT ends up being little more than end point detection and closed loop controls. Other initiatives include Quality by Design (QbD) to help develop and transfer robust processes and Six Sigma to drive quality improvements.

Using practical examples, and a pharmaceutical development case study^[1], it is demonstrated how next generation Design of Experiments (DOE) enable more scientists to apply DOE with reduced experimental effort and explain why exploratory analysis, statistical and predictive modelling of historical data simplifies and shortens whole the DOE process.

3- Results and discussion

The screening experiment shows that one of the factors initially considered, which were Active Pharmaceutical Ingredients (API) Particle Size (μm), Magnesium Stearate (%), Croscarmellose Sodium (%), Roller Pressure (bar), Mill Screen Size (in), Mill Speed (rpm), the Croscarmellose Sodium did not actually influence the multiple responses ribbon density (p-value = 0.52), Granule Surface Area (p-value = 0.83) and Dissolution of the tablet at 30 minutes (p-value = 0.53) (Figure 1).

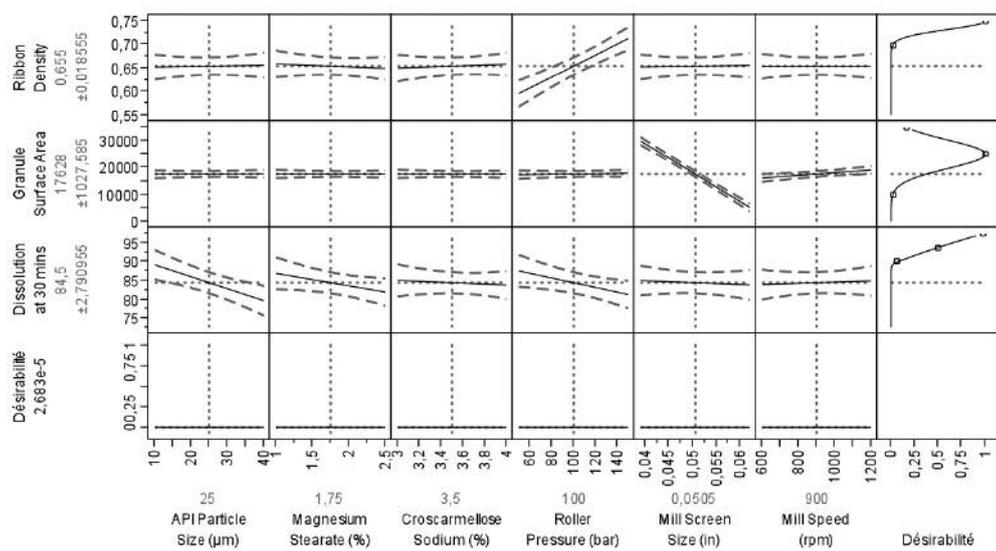


Figure 1 - Identifying statistically and visually the important factors

After having considered different numbers of additional runs to augment the initial design using the statistical power (Figure 2), a 32-run central composite design was obtained whilst making sure to use a blocking variable.

Variance relative des coefficients			
Niveau de signification			0,05
Rapport signal / bruit			1
	Coefficient	Variance	Puissance du test
	Constante	0,3	0,379
	API Particle Size (µm)	0,054	0,971
	Magnesium Stearate (%)	0,05	0,98
	Roller Pressure (bar)	0,052	0,976
	Mill Screen Size (in)	0,057	0,963
	Mill Speed (rpm)	0,051	0,978
	Bloc	0,056	0,966
	API Particle Size (µm)*API Particle Size (µm)	0,294	0,385
	API Particle Size (µm)*Magnesium Stearate (%)	0,071	0,92
	Magnesium Stearate (%)*Magnesium Stearate (%)	0,3	0,379
	API Particle Size (µm)*Roller Pressure (bar)	0,072	0,92
	Magnesium Stearate (%)*Roller Pressure (bar)	0,064	0,945
	Roller Pressure (bar)*Roller Pressure (bar)	0,321	0,358
	API Particle Size (µm)*Mill Screen Size (in)	0,071	0,922
	Magnesium Stearate (%)*Mill Screen Size (in)	0,083	0,879
	Roller Pressure (bar)*Mill Screen Size (in)	0,09	0,852
	Mill Screen Size (in)*Mill Screen Size (in)	0,252	0,437
	API Particle Size (µm)*Mill Speed (rpm)	0,058	0,961
	Magnesium Stearate (%)*Mill Speed (rpm)	0,066	0,938
	Roller Pressure (bar)*Mill Speed (rpm)	0,062	0,95
	Mill Screen Size (in)*Mill Speed (rpm)	0,062	0,95
	Mill Speed (rpm)*Mill Speed (rpm)	0,311	0,368

Figure 2 - Assessing the number of runs necessary using power

Once the 24 additional experiments have been run, the robustness of the process in terms of dissolution of the tablet, granule surface area and ribbon density could be assessed using a combination of Monte-Carlo simulation and defect profiler (Figure 3) which uses a quadratic model.

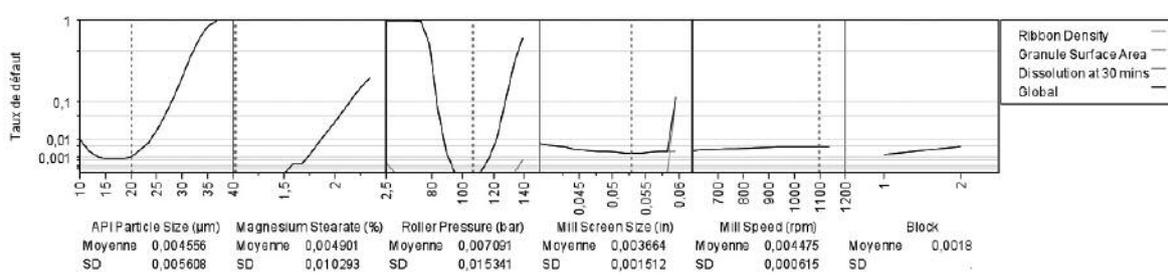


Figure 3 - Defect profiling helps defining operating range

CONCLUSION

This study demonstrates the usefulness of taking a step-by-step approach to optimisation and the power of advanced analytics in designing and analysing experiments. In addition, the authors would also like to stress the importance of utilising prior data that may be available. This could for example be achieved by using data mining techniques [2].

Références

- [1] QbD Case Study based on Pharmaceutical Development Case Study: "ACE Tablets" Prepared by CMC/IM Working Group March 13, 2008
- [2] Interactive Data Mining and Design of Experiments: the JMP® Partition and Custom Design Platforms Gaudard M., Ramsey, P. Stephens M., MS North Haven Group March 2006

Biography

Dr. Malcolm Moore, JMP Technical Manager Europe has worked with clients to integrate statistical methods and software into R&D, quality improvement, defect reduction, cycle time reduction, and corporate Six Sigma consulting activities for a variety of industries, including pharmaceutical and semiconductors. Prior to joining SAS UK he worked at Light Pharma, BBN, Astra Zeneca and lectured in Medical Statistics at Newcastle University. He is an expert in design of experiments, and received his Ph.D. in Design of Non-Linear Experiments at London University.

Dardenne à venir

MÉTHODES MULTIVOIES

CHAIRMEN:
DELPHINE BOUVERESSE
PHILIPPE GIAMARCHI



**PLANS D'EXPÉRIENCES
ET PLANS NUMÉRIQUES**

CHAIRMEN:

MICHELLE SERGENT

PIERRE LANTÉRI

Construction de plans uniformes en grande dimension

J. Santiago¹ M.Claeys-Bruno² M. Sergent³

^{1,2,3}, ISM2, UMR 6263, Laboratoire de Méthodologie de la Recherche Expérimentale, Université Paul Cézanne, Faculté des Sciences et Techniques, Avenue Escadrille Normandie Niémen, case D52, 13397 Marseille Cedex 20, France

jenny.santiago@etu.univ-cezanne.fr
m.claeys-bruno@univ-cezanne.fr
michelle.sergent@univ-cezanne.fr

Keywords: Space-filling designs, simulation numérique, plans numériques.

1- Introduction

Du fait de la croissance accrue de la performance des ordinateurs et de la complexité des codes de calculs ou du modèle, les plans d'expériences sont de plus en plus utilisés en simulation numérique, et plus particulièrement les plans uniformes, ou "space-filling designs", qui visent à répartir les points uniformément dans l'espace des variables d'entrée. Cependant, dès que des études avec un grand nombre de paramètres sont envisagées, supérieur à 20 dimensions, la construction de tels plans d'expériences s'avère délicate et les qualités de remplissage de l'espace par ces distributions de points sont difficiles à évaluer. Il apparaît donc une nécessité sur le choix de mesures d'uniformité et sur le développement de méthodes de construction de plans uniformes en grande dimension.

2- Méthodes et principes

2.1 Critères d'uniformité

Pour juger la qualité d'un plan uniforme, il est nécessaire d'utiliser des critères quantitatifs intrinsèques permettant d'évaluer l'uniformité d'une distribution de n points dans un espace à d dimensions. Dans cette étude, deux critères, les plus couramment utilisés, sont retenus :

- la mesure de *recouvrement* (Cov)^[1], qui représente une mesure de la dispersion des distances minimales entre deux points, ce qui permet de savoir si les distances minimales entre deux points sont homogènes. Une mesure de recouvrement nulle est caractéristique d'une grille régulière.

$$Cov = \frac{1}{\bar{\gamma}} \left(\frac{1}{n} \sum_{i=1}^n (\gamma_i - \bar{\gamma})^2 \right)^{1/2}$$

avec $\gamma_i = \min_{k \neq i} dist(x^i, x^k)$ and $\bar{\gamma} = \frac{1}{n} \sum_{i=1}^n \gamma_i$,

- la distance *MinDist*^[2], qui retient la valeur minimale des distances minimales entre deux points du plan.

$$MinDist = \min_{x_i \in X} \min_{\substack{x_k \in X \\ k \neq i}} dist(x^i, x^k)$$

avec $X = \{x^1, x^2, \dots, x^n\} \subset [0,1]^d$, un ensemble de n points en d dimensions.

Les plans uniformes optimaux selon ces critères doivent avoir une mesure de recouvrement Cov faible, et une distance $MinDist$ élevée.

2.2 Les plans Wootton, Sargent et Phan-Tan-Luu [3,4]

Cet algorithme de construction de plans uniformes se base sur un ensemble de N_i points candidats et sélectionne un sous-ensemble de N_f points de telle façon qu'ils soient à une distance minimale de chaque point déjà présent dans la matrice. Ainsi, la première étape de l'algorithme WSP étant la génération d'un ensemble de points candidats, nous nous interrogeons sur l'impact du choix des points candidats, quant à leur nombre et leur distribution, sur les propriétés des matrices résultantes.

3- Résultats et discussion

Les études ont été réalisées en 5, 10, 15, 20, 30 et 50 dimensions. Dans cet exposé ne sera présentée que l'étude à 20 dimensions, en évaluant la qualité des plans considérés selon les deux critères mentionnés auparavant.

3.1 Nature de l'ensemble de points candidats

Pour la construction de l'ensemble des N_i points candidats, nous avons choisi différentes matrices, classiquement utilisées en simulation numérique: une suite à faible discrédance [5], un Hypercube Latin (LHS) [6] et un plan généré aléatoirement ($N_i = 3000$ points candidats). Pour la matrice finale, nous avons retenu 20 points par dimension, soit $N_f = 400$ points.

A partir de cette étude, nous pouvons comparer les critères MinDist et Cov des plans WSP obtenus à partir des différents plans initiaux et ainsi connaître l'impact de la génération des points candidats. Les résultats sont présentés dans le tableau ci-dessous puis représentés graphiquement.

Matrices	Nombre de points	MinDist	Cov
Aléatoire	3000	0.585	0.095
Aléatoire + WSP	400	1.319	0.023
LHS	3000	0.589	0.096
LHS + WSP	400	1.320	0.023
Suite de Hammersley	3000	0.512	0.237
Hammersley + WSP	400	1.292	0.024

Tableau 1 - MinDist et Cov pour les différents plans

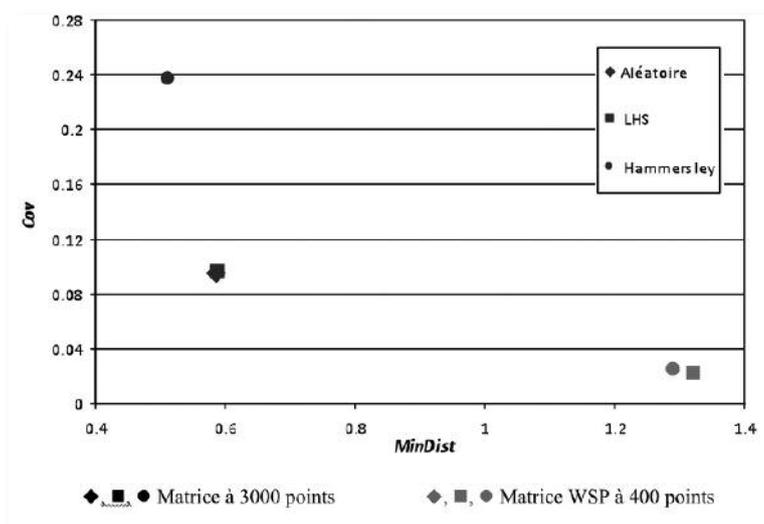


Figure 1 - MinDist et Cov des matrices initiales (N = 3000) et des matrices WSP résultantes (N = 400).

Nous pouvons observer que quelle que soit la nature du plan initial, les critères des plans finaux sont comparables, ce qui permet de conclure que la distribution initiale a peu d'impact sur les propriétés du plan final. De plus, les critères de ces matrices sont meilleurs que ceux d'une matrice SFD classique avec un même nombre de points (N = 400).

3.2 Nombre de points candidats

Comme mentionné précédemment, l'algorithme WSP sélectionne un sous-ensemble de points dans un ensemble initial et n'ajoute aucun point au plan initial. Nous pouvons alors nous demander si le nombre de points NI du plan initial joue un rôle sur les critères intrinsèques du plan résultant. Ainsi, pour un même type de plan, nous avons fait varier le nombre de points du plan initial et appliqué l'algorithme WSP pour obtenir des plans finaux avec 20 points par dimension, suivi du calcul des critères. Le tableau ci-dessous présente les résultats en dimension 20.

Matrices	Nombre de points candidats	Nombre de points de la matrice finale	MinDist	Cov
HAMMERSLEY		400	0.510	0.280
	500	399	0.670	0.172
	1000	401	1.160	0.037
	1500	399	1.220	0.028
	2000	400	1.250	0.027
	2500	399	1.284	0.025
	3000	399	1.292	0.024

Tableau 2 - Critères *MinDist* et *Cov* des plans en faisant varier le nombre de points candidats

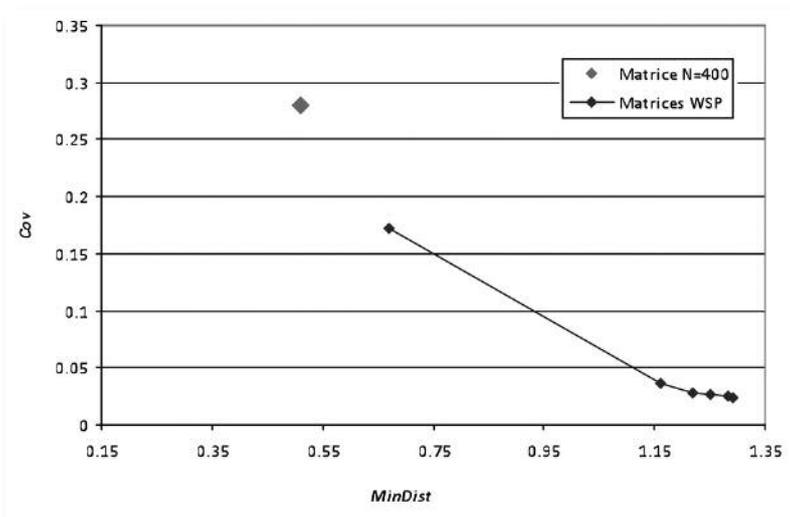


Figure 2 - *MinDist* et *Cov* de matrices WSP (N = 400) à partir de différentes suites de Hammersley avec des nombres de points candidats variables

L'étude a été également réalisée sur des Hypercubes Latins et des plans générés aléatoirement. Pour chaque étude, nous pouvons faire les mêmes observations. Tout d'abord, nous remarquons que la qualité des critères intrinsèques du plan final s'améliore avec l'augmentation du nombre de points candidats jusqu'à une valeur au-delà de laquelle l'ajout de points se révèle sans effet. Cette asymptote apparaît vers environ 5 fois le nombre de points de la matrice finale.

CONCLUSION

L'utilisation de l'algorithme WSP est simple, rapide et semble prometteur en terme de nombre de points et de dimensions.

Références

- [1] Gunzburger M, Burkardt J. (2004). *Uniformity measures for point samples in hypercubes*. www.csit.fsu.edu/~burkardt/pdf/ptmeas.pdf.
- [2] Johnson M E, Moore L M, Ylvisaker D (1990) *Minimax and maximin distance designs*. *J. of Statistical Planning and Inference* 26:131-148.
- [3] Sergent M (1989) *Contribution de la Méthodologie de la Recherche Expérimentale à l'élaboration de matrices uniformes : Application aux effets de solvants et de substituants*. Ph D Thesis, Université Aix Marseille III.
- [4] Sergent M, Phan-Tan-Luu R, Elguero J (1997) *Statistical Analysis of Solvent Scales. Part 1*, *Anales de Quimica Int. Ed.* 93:71-75.
- [5] Hammersley J.M (1960) *Monte-Carlo methods for solving multivariate problems*, *Annals of the New York Academy of Sciences* 86:844-874.
- [6] Stein M (1987) *Large Sample Properties of Simulations Using Latin Hypercube Sampling*. *Technometrics* 29:143-151.

High dimensional space filling design analysis

Adrian Azarian¹, Olivier Vasseur²

¹ Département d'optique théorique et appliquée, ONERA, Chemin de la Hunière, 91761 Palaiseau, France, adrian.azarian@onera.fr

² Département d'optique théorique et appliquée, ONERA, Chemin de la Hunière, 91761 Palaiseau, France, olivier.vasseur@onera.fr

Keywords: space filling design, radial scanning statistic, minimum spanning tree.

1- Introduction - Guidelines

In some of the most complex interference optical systems^[1,2], there is no direct analytic formula to express the relationship between the outputs and the inputs of the computer code. These systems are also highly sensitive to any variation of input parameters, especially because of the interactions with all the other parameters. So the impact of uncertainties of input parameter values and the possible interactions between these input parameters on the output must be assessed to identify the most critical parameters and interactions.

However, classical approaches, like Monte-Carlo, to study systems with a high number of parameters are still too time consuming, despite the increasing power of computers over the past decade. Moreover these approaches do not permit to exhibit the most critical interactions between input parameters. The use of numerical space filling designs is an effective method to study the outputs of a system all over the domain of variation of the inputs, enabling metamodel construction and high dimensional code exploration by a limited number of code runs.

Therefore, we propose to evaluate the intrinsic quality of numerous space filling designs, by two different methods. We will first present the different designs that we will study. We will then analyze these designs by two different methods: the radial scanning statistic^[3] and the minimum spanning tree^[4]. Finally we are able to conclude on the space filling design, which has the best intrinsic quality.

2- Material and methods

Numerical space filling designs have been actively studied and these designs are constructed using different algorithms going from random designs to deterministic sequences. We propose to study the following designs:

- Low discrepancy sequences [5, 6] in 18 dimensions (400 and 200 point) and 29 dimensions (614 points)
- Latin hypercube [7] in 18 dimensions (400 and 200 point) and 29 dimensions (614 points)
- Random designs in 18 dimensions (400 and 200 point) and 29 dimensions (614 points)
- WSP designs [8] in 18 dimensions (400 and 200 point) and 29 dimensions (614 and 598 points)

The knowledge of their quality in higher dimensional cases (here 18 and 29 dimensions) is of great importance to study complex phenomena.

At first we analyze the quality of the designs by the radial scanning statistic method using the DICE package in R-software developed in the framework of the DICE consortium. It is based on the analysis of all the 2-dimensional projection of a design. For each projection, it scans angularly the domain and gives a statistic about the uniformity of the projected points of the design. The worst projection is sorted out with an associated quality value, which is related to alignments of points in the design. A graphic representation is also given in order to show the worst projection. For example, points' alignments in the plane 1-2 lead to a less efficient design if the code is a linear combination of the parameters 1 and 2, due to loss of information.

Then we use the Minimum Spanning Tree (MST) to analyze the same designs. This method is based on a tree constructed from the set of points of the design. This tree, contains all the points, has no cycle and the sum of the length of its edges is minimal. The design is then associated by the mean m and the standard deviation σ of its edges lengths, which characterize the type of distribution of the points (cluster, ordered...). In the (m, σ) plane, the best design is characterized by a high average length of MST branches to fill the space and a small standard deviation to obtain sufficient regularity. The standard deviation should not be equal to 0, in order to have points evenly spread across the projection of the experimental space onto all subspaces.

3- Results and discussion

We compare the value given by the radial scanning statistic on table 1 in the case of 18-dimensional designs. In higher dimension, the comparison of different designs by this criterion is more difficult. The values in 29 dimensions for some designs are almost the same but a plot show a great difference between these designs. Therefore, it is not always sufficient to use the single radar value to accurately assess the quality of a design and a graphic representation is needed.

Number of points of the design	random	Sobol	Latin hypercube	WSP	Faure
200	0.0145	0.083	0.0148	0.0196	0.161
	0.0145		0.0132	0.0269	
	0.0147		0.0145		
	0.0144		0.0145		
	0.0153		0.0169		
400	0.0064	0.042	0.0062	0.0124	0.0692
	0.0068		0.0063	0.0256	
	0.0065		0.0063		
	0.0065		0.0064		
	0.0066		0.0062		

Therefore, it is not always sufficient to use the single radar value to accurately assess the quality of a design and a graphic representation is needed.

Table 1 - Radial scanning statistic of the 18-dimensional designs

Finally, with this criterion we conclude that designs made by low discrepancy sequences have the worst intrinsic quality. But it is unclear which designs (between the other types of designs) is the best.

The results of the intrinsic quality of the designs (of table 1) by the MST method are presented on figure 1. The WSP designs are the best as they have the largest mean and the smallest standard deviation.

Except for the Faure design which is clearly the worst, the other designs are very similar in terms of quality.

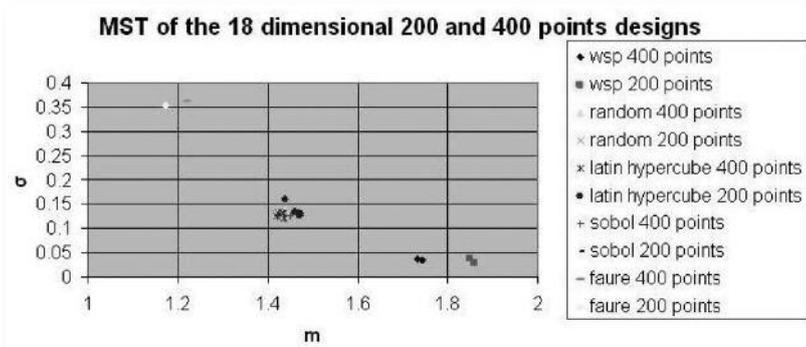


Figure 1 - MST of the 18-dimensional designs

CONCLUSION

Finally, by analyzing the intrinsic quality of the different designs, the conclusions of the two methods are slightly different. In fact, these methods are analyzing two different aspects of the points' distribution, and are complementary. Using simultaneously the two criteria, we conclude that the low discrepancy sequences are the design with the worst quality. The WSP design appears to be the best design: the best with respect to the MST method, and acceptable with respect to the radial scanning statistic.

These designs have been used on interferometric systems and the results will be subject to a future publication.

Acknowledgment

The authors wish to thank Michelle Sergent and Magalie Claeys-Bruno for their help and specially for providing us with all the WSP designs.

Références

[1] V. Jolivet, P. Bourdon, B. Bennai, L. Lombard, D. Goular, E. Pourtal, G. Canat, Y. Jaouen, B. Moreau, O. Vasseur : Beam shaping of single-mode and multimode fiber amplifier arrays for propagation through atmospheric turbulence. *IEEE journal of selected topics in quantum electronics*, 15(2):257-268, 2009.

[2] H.A. Macleod : *Thin optical filers*, Adam Hilger Ltd, 1986.

[3] O. Roustant, J. Franco, L. Carraro, A. Jourdan : a radial scanning statistic for selecting space-filling designs in computer experiments. *Proc. MODA-9, 2010*, pp.189-196.

[4] J. Franco, O. Vasseur, B. Corre, M. Sergent : Minimum Spanning Tree: A new approach to assess the quality of the design of computer experiment. *Chemometrics and intelligent laboratory systems*, 97:164-169, 2009.

[5] I.M. Sobol, uniformly distributed sequences with an additional uniform property, *USSR computational Math. And Math. Phys.* 16(5):236-242, 1976.

[6] H. Faure : Discrépance de suites associées à un système de numération (en dimension s). *Acta Arith.* 41(4):337-351, 1982.

[7] M. Stein : Large sample properties of simulations using latin hypercube sampling. *Technometrics*, 29(2):143-151, 1987.

[8] M. Sergent, R. Phan-Tan-Luu, J. Elguero : Statistical analysis of solvent scales, Part 1 *An. Quim. Int.* Ed 93, 1997.

Potentiel des plans numériques de type Space Filling Design pour l'analyse de sensibilité et la construction de métamodèles dans le domaine de l'optique interférentielle.

O. Vasseur¹, A. Azarian², V. Jolivet³, P. Bourdon⁴

¹ ONERA/DOTA, Chemin de la Hunière et des Joncherettes, 91761 Palaiseau Cedex, olivier.vasseur@onera.fr

² ONERA/DOTA, Chemin de la Hunière et des Joncherettes, 91761 Palaiseau Cedex, adrian.azarian@onera.fr

³ ONERA/DOTA, Chemin de la Hunière et des Joncherettes, 91761 Palaiseau Cedex, veronique.jolivet@onera.fr

⁴ ONERA/DOTA, Chemin de la Hunière et des Joncherettes, 91761 Palaiseau Cedex, pierre.bourdon@onera.fr

Keywords: Plans d'expériences, Space Filling Design, Laser, Combinaison cohérente, Filtres optiques.

Les codes de calcul qui modélisent ou simulent des phénomènes complexes sont de plus en plus réalistes et même si la puissance des ordinateurs augmente sans cesse, les temps de calcul demeurent importants et limitent ainsi le recours aux techniques de Monte-Carlo. Le développement de métamodèles permet alors de remplacer le simulateur par un outil plus "simple" construit à partir du simulateur complexe. Ces métamodèles ou surfaces de réponse sont en général des fonctions obtenues à l'aide de méthodes d'interpolation ou d'approximation à partir d'un nombre limité d'exécutions du simulateur sur des jeux de paramètres constituant le plan d'expériences numériques. Du fait des caractéristiques non linéaires et/ou non paramétriques des codes de modélisation ou simulation, il est nécessaire de répartir les points dans l'espace le plus uniformément possible de façon à capter au mieux le comportement du simulateur. C'est ce mode de répartition des points que cherchent à proposer les plans d'expériences Space Filling Designs (SFD).

Afin de construire un métamodèle de qualité, il est nécessaire que ce dernier soit représentatif non seulement de l'effet de chacun des paramètres d'entrée mais également de leurs possibles interactions. Dans le cas de l'optique interférentielle, la prise en compte des interactions s'avère particulièrement cruciale pour déterminer les paramètres les plus sensibles ou la construction d'un métamodèle de qualité.

Dans notre travail, différents types de plans numériques "Space Filling Designs" caractérisés par les méthodes de l'arbre de longueur minimale [1] [2] [3] et du radar [4] [3] ont été utilisés pour réaliser des analyses de sensibilité et la construction de métamodèles dans le cas des systèmes/composants interférentiels optiques suivants : un filtre interférentiel composé de 29 couches minces d'une part et la combinaison cohérente de 19 amplificateurs laser à fibre d'autre part. Les métamodèles obtenus par l'utilisation des différents plans numériques sont alors comparés au moyen de plusieurs critères afin de déterminer les plans les plus performants. En définitive, la qualité des plans Space Filling Designs sur ces deux cas d'application est alors mise en relation avec leur qualité intrinsèque [2] [3].

Références

[1] Franco J., Vasseur O., Corre B., Sergent M., *Minimum Spanning Tree : A new approach to assess the quality of the design of computer experiments*, *Chemometrics and Intelligent Laboratory Systems*, 97, 164-169, 2009.

[2] Vasseur O., Claeys-Bruno M., Franco J., Sergent M., *Analyse de la qualité intrinsèque des plans d'expériences numériques "Space Filling Design" de grande dimension par l'arbre de longueur minimale*, *Chimiométrie 2009*, 2009.

[3] Azarian A., Vasseur O., *High dimensional space filling design analysis*, *Chimiométrie 2010*, 2010.

[4] Roustant O., Franco J., Carraro L., Jourdan A. : *A radial scanning statistic for selecting space-filling designs in computer experiments*, *Proc. MODA-9*, 189-196, 2010.

[5] Macleod H.A., *Thin film optical filters*, Adam Hilger Ltd, Bristol, 1986.

[6] Jolivet V., Bourdon P., Bennai B., Lombard L., Goular G., Pourtal E., Canat G., Jaouën Y., Moreau B., Vasseur O., *Coherent Combining of Fiber Amplifiers Through Atmospheric Turbulence*, *IEEE J. Sel. Topics Quantum Electron.*, 15, 257-268, 2009.

DOE: from empiricism to tailor-made, the story of the optimisation of red blood cells culture

L. Guyonneau-Harmand^{1,2,3}, S. Jolly², F. Baudet⁴, L. Douay^{1,2,3,5}

¹ INSERM, UMR_S938, Proliferation and Differentiation of Stem Cells, Paris, France, ² UPMC Univ Paris 06, UMR_S938, Proliferation and Differentiation of Stem Cells, Paris, France, ³ Etablissement Français du Sang Ile de France, Ivry-sur-Seine, France, ⁴ JMP France, ⁵ AP-HP, Hôpital Armand Trousseau, Service d'Hématologie biologique, Paris, France, laurence.harmand@upmc.fr

Keywords: DOE, erythrocytes, temperature, % CO₂, %O₂, pH.

1- Introduction

A number of reports describe how to produce cultured red blood cells from CD34+ cells ^[1, 2], human embryonic stem cells ^[3] or human induced pluripotent stem cells ^[4, 5] with ^[6] or without ^[7] microenvironment, but none of these focus on the environmental variables. Yet, temperature, pH and % O₂ during cell growth are key variables that must be considered in the design of the process.

Reported temperature for the growth of human cells is 37°C. The temperature of various bone marrow sites has been measured and it has been found that the temperatures vary from 34°C to 38°C. Therefore, hematopoietic activity is not necessarily associated with precise maintenance and regulation of temperature ^[8, 9].

Most studies with human cells are conducted at 5% CO₂, which regulates the pH of the medium. As the bone marrow has a non-uniform structure, and the circulating blood within this network ensures the optimal supply of oxygen and concomitant removal of CO₂ ^[10], a great disparity is evident for pH. Moreover pH has been shown to be an important modulator of proliferation and differentiation in ex vivo haematopoiesis ^[11].

Standard cell culture conditions use incubation of cells in air, i. e. about 20% O₂. A similar, but greater, disparity is evident for O₂ from the deeply hypoxic areas of bone marrow (1% O₂) suitable for the long-term maintenance of stem cells to the better oxygenated areas (8% O₂) at the edge of blood vessels which enhance the proliferation of more differentiated progenitors ^[12, 13, 14, 15].

A fractional-factorial design was chosen to optimize the environmental condition with respect to metabolic activity of the cells from CD34+ to reticulocytes.

2- Material and methods

The experimental design was run with pooled CD34+ cells selected from 7 umbilical cord blood units. The cells concentrations and the different media used were previously described ^[1].

Based on the literature, a wide range of temperature (35-39°C), % CO₂ (2.5-7.5) and % O₂ (4.9-16.2) values were screened for optimization.

Using the JMP-8TM software package experiments was designed, based on a response surface methodology ^[16], with three replicates at the center point (37°C, 5% CO₂, 10.5% O₂), representing a total of 17 experiments (central composite design). Each experiment was run for 21 days under controlled environmental conditions.

Crucial quality criteria for evaluation of red blood cell production in vitro are cell integrity, amplification rate and enucleation rate. Each of these responses was analysed at every step of the culture.

3- Results and discussion

The results of the 17 experiments were treated by regression analysis in order to fit the response function.

The optimization for multiple responses, based upon the concept of desirability, is a proven method to simultaneously determine optimum settings of input factors that achieve optimum performance levels.

The prediction model was built for amplification rate for each step and enucleation rate for the last step.

For each step of the culture, the predicted values for achieving an optimum production were used. As the amplification rate and the enucleation rate were very low respectively 1603308 cells/seeded CD34+ cell and 43%; and there was a high rate

of orthochromatic erythroblasts, the culture was run for three more days. The enucleation rate increased (79%) but not the amplification rate (1440520 cells/seeded CD34+ cell).

The result of each step of the culture was compared to the predictive model to find out where the problem laid. A big gap between the theoretical rate and the obtained rate appeared during the day 8 to day 11 phase (D8-D11). The cells were cultivated at 37.4°C for the first 8 days and at 35.7°C from the 8th to the 11th day.

As this abrupt temperature decrease may affect the cell metabolism, the cell metabolism was followed on a daily basis to investigate this aspect.

At D8, cells were seeded at 105cells/ml in several wells in order to follow the culture day by day from D10. Each day, one well was studied for each culture, cells were enumerated, and supernatant frozen to test for glucose, lactic acid and ammonium.

The new culture conditions indeed affected the metabolism of the cells. The cells, which grew under the new conditions, showed a delay for every tested parameter, and were still in exponential phase after 16 days. Moreover glucose consumption as well as ammonia and lactate acid production were 2 days delayed.

We previously demonstrated (data not shown) that an ammonia concentration higher than 1000 µM is toxic for the amplification step. This ammonia concentration was reached at the end of D11 for the normal culture and at the end of D13 for the culture under the experimental conditions.

To benefit from the new conditions and to maintain cells in exponential phase, the phase D8-D11 became the phase D8-D13.

The predictive model was run another time with a D8-D13 step and its end was delayed for 2 days due to a poor enucleation rate (40%) and a high concentration of orthochromatic erythroblasts.

In these conditions, the proliferation rate not only reached but exceeded the theoretical rate of 120%.

CONCLUSION

This study demonstrates the utility of Design of Experiment method (DoE) to optimize the key variables of cell cultures. A relatively modest number of experiments provided detailed and precise information about our cells sensitivity.

This study outlined that new culture conditions meant new metabolism, which is why new conditions cannot be applied blindly. It is where the statistics end and where the cell culture knowledge begin.

One CD34+ is able to produce 3 679 045 red blood cells. Therefore with less than 300000 CD34+ a unit of red blood cells can be processed.

In addition, what has been learned in this study could be extrapolated to future bioreactors.

Thanks to the DoE, the thought of in vitro generation of cultured red blood cells units is no more a dream and with further enhancements it could be a real alternative to donor red blood cells units.

Références

- [1] Giarratana MC, Kobari L, Lapillonne H, Chalmers D, Kiger L, Cynober T, Marden MC, Wajcman H, Douay L. Ex vivo generation of fully mature human red blood cells from hematopoietic stem cells. *Nat Biotechnol.*, 23(1):69-74, 2005.
- [2] Baek EJ, Kim HS, Kim S, Jin H, Choi TY, Kim HO. In vitro clinical-grade generation of red blood cells from human umbilical cord blood CD34+ cells. *Transfusion*, 48(10):2235-45, 2008.
- [3] Lu SJ, Feng Q, Park JS, Lanza R. Directed differentiation of red blood cells from human embryonic stem cells. *Methods Mol Biol.*, 636:105-21, 2010.
- [4] Lapillonne H, Kobari L, Mazurier C, Tropel P, Giarratana MC, Zanella-Cleon I, Kiger L, Wattenhofer-Donzé M, Puccio H, Hebert N, Francina A, Andreu G, Viville S, Douay L. Red blood cell generation from human induced pluripotent stem cells: perspectives for transfusion medicine. *Haematologica*, 95(10):1651-9, 2010.
- [5] Lengerke C, Grauer M, Niebuhr NI, Riedt T, Kanz L, Park IH, Daley GQ. Hematopoietic development from human induced pluripotent stem cells. *Ann N Y Acad Sci.*, 1176:219-27, 2009.
- [6] Fujimi A, Matsunaga T, Kobune M, Kawano Y, Nagaya T, Tanaka I, Iyama S, Hayashi T, Sato T, Miyanishi K, Sagawa T, Sato Y, Takimoto R, Takayama T, Kato J, Gasa S, Sakai H, Tsuchida E, Ikebuchi K, Hamada H, Niitsu Y. Ex vivo large-scale generation of human red blood cells from cord blood CD34+ cells by co-culturing with macrophages. *Int J Hematol*, 87(4):339-50, 2008.
- [7] Baek EJ, Kim HS, Kim JH, Kim NJ, Kim HO. Stroma-free mass production of clinical-grade red blood cells (RBCs) by using poloxamer 188 as an RBC survival enhancer. *Transfusion*, 49(11):2285-95, 2009.
- [8] Petrakis NL. Temperature of human bone marrow. *J Appl Physiol.*, 4(7):549-53, 1952.
- [9] Petrakis NL. Some physiological and developmental considerations of the temperature-gradient hypothesis of bone marrow distribution. *Am J Phys Anthropol.*, 25(2):119-29, 1966.
- [10] Adler, SS. Hemopoietic reconstitution of marrow and spleens in mice after whole-body irradiation and marrow transplantation. *Transplantation*. 37(5), 517-20, 1984.
- [11] Brown RA, Adkins D, Goodnough LT, Haug JS, Todd G, Wehde M, Hendricks D, Ehlenbeck C, Laub L, DiPersio J. Factors that influence the collection and engraftment of allogeneic peripheral-blood stem cells in patients with hematologic malignancies. *J Clin Oncol.*, 15(9):3067-74, (1997).
- [12] Cipolleschi MG, Dello Sbarba P, Olivetto M. The role of hypoxia in the maintenance of hematopoietic stem cells. *Blood*. 82(7):2031-7, 1993.
- [13] Dexter TM, Wright EG, Krizsa F, Lajtha LG. Regulation of haemopoietic stem cell proliferation in long term bone marrow cultures. *Biomedicine*, 27(9-10):344-9, 1977.
- [14] Allen TD, Dexter TM. The essential cells of the hemopoietic microenvironment. *Exp Hematol.*, 12(7):517-21, 1984.
- [15] Mizuno S, Glowacki J. Low Oxygen Tension Enhances Chondroinduction by Demineralized Bone Matrix in Human Dermal Fibroblasts in vitro. *Cells, tissues, organs*, 180(3):151-158, 2005.
- [16] Box, G. E. P. and Draper, N. R. (I). A Basis for the Selection of a Response Surface Design", *Journal of the American Statistical Association*, 54, 622-654, 1959.

MULTI-TABLEAUX

CHAIRMEN:

CHRISTOPHE CORDELLA

DOMINIQUE BERTRAND

On the study of the influence of geographical factors on skin surface lipids (SSLs): Multivariate analysis of Raman and GC-MS data

Rime Michael-Jubeli¹, Ali Tfayli¹, Jean Bleton², Pierre Chaminade¹, Michel Manfait³, Arlette Baillet-Guffroy¹

¹EA 4041 "Groupe de Chimie Analytique de Paris Sud", Faculté de Pharmacie, Université de Paris-Sud, Châtenay-Malabry, France, rime.michael@u-psud.fr, ali.tfayli@u-psud.fr, pierre.chaminade@u-psud.fr, arlette.baillet-guffroy@u-psud.fr.

²EA 4041 "Groupe de Chimie Analytique de Paris Sud", LETIAM-IUT d'Orsay, Université de Paris-Sud, Orsay, France, jean.bleton@u-psud.fr.

³Unité MEDyC UMR CNRS 6237, Faculté de Pharmacie, Université de Reims Champagne Ardenne, Reims, France, michel.manfait@univ-reims.fr.

Keywords: Skin Surface Lipids, GC-MS, Raman, Data Multivariate Analysis.

Introduction

Human skin surface lipids (SSLs) present in the hydrolipidic film are "the final frontier between the organism and the environment". They participate in maintaining a good skin barrier function and preventing aging.

The SSLs represent a complex lipid mixture produced by epidermal and sebaceous activities: free fatty acids, wax esters, squalene, glycerides, free and esterified sterols.

One can expect a physiological adaptation of SSLs composition according to food habits and to climatic factors (temperature, relative humidity and pollution) or to the skin pigmentation.

In this work, we investigated the potential variability of SSLs composition within several human groups which differ from each other by their geographic location (France and Ivory Coast) and/or their skin color (Phototype II and VI). For this reason we have developed an analytical methodology based on Raman spectroscopy and on high-temperature gas chromatography coupled to mass spectrometry. Chemometrics were performed in order to highlight the variations between the different groups and to investigate potential correlations.

Materials and methods

Sample collection: SSLs samples were collected using absorbent papers from foreheads of 18 male volunteers, divided into 3 groups: phototype II living in France (A), phototype VI living in France (B) and phototype VI living in Ivory Coast (C).

Data acquisition: collected lipids were analyzed using a high temperature gas chromatography - mass spectrometry (Thermo scientific) using "On-column" mode injection. The separation was achieved using a weakly polar stationary phase (Phenomenex). Two chromatograms were realized for each sample.

Raman spectra were collected using HR Labram (Horiba Jobin Yvon) spectrometer with a 785 nm excitation laser source.

Data treatment: for GC-MS data preprocessing, MZmine software was used. Data processing was then realized using SIMCA-P Multivariate analysis software (Umetrics AB). Principal component analysis (PCA) and partial least squares regression (PLS) after an orthogonal signal correction (OSC) were performed.

Raman data were analyzed with in-house software that operates in the Matlab environment (The Mathworks, Inc., Natick, Massachusetts). Kruskal wallis, principal component analysis (PCA) and hierarchical cluster analysis (HCA) were performed.

Results and discussion

Classical approach was shown to be not sufficient to highlight relevant variations in the SSLs composition between the different groups. In order to reveal these variations, different multivariate analysis methods were performed for processing both GC-MS and Raman data.

Firstly, for chromatographic analyses, PCA and PLS were directly performed on raw data. No discriminant variables could be determined between the groups. Contrarily, OSC-PLS showed a good separation between volunteers living in France (A+B) with those living in Ivory Coast (C), and the sapienic acid (C16:1 Δ 6) was the compound that contributes the most to the separation.

These results suppose that the climatic, geographical and alimentary factors have a more important influence than skin pigmentation on the composition of the hydrolipidic film (SSLs), particularly on the sapienic acid (C16:1 Δ 6). In fact, this acid participates in the composition of the majority of compounds in the different lipid classes of hydrolipidic film.

Secondly, Raman data enabled us to determine specific Raman signatures for the representative standards for each SSLs lipid class. The vibrational features of extracted SSLs are directly related to their structure and their composition. Kruskal Wallis enabled to identify spectral regions with high variations between the groups. Cluster analysis was then performed on these spectral regions to separate Raman signatures.

Results from Raman signal processing were finally compared to those from GC-MS data treatment.

CONCLUSION

The SSLs composition variability in terms of environmental and alimentary factors was assessed, using GC-MS and Raman methods. Chemometric data preprocessing and processing is crucial for the separation and the identification of discriminant components. Slight skin lipid variations were highlighted, which proves the potential of our approach. Therefore, this approach could be applied to detect lipid disorder related to skin disease and evaluate the efficiency of certain cutaneous treatment.

Analyse de plans d'expériences en situation multivariée : une contribution des méthodes multi-blocs

G. Mazerolles^{1,2,3}, J. Wirth^{1,2,3}, A. Verbaere^{1,2,3}, J. Boccard⁴, S. Rudaz⁴, M. Hanafi⁵, V. Cheynier^{1,2,3}

¹ INRA, UMR 1083, F-34060 Montpellier, France.

² Montpellier SupAgro, UMR1083, F-34060 Montpellier, France

³ Université Montpellier I, UMR1083, F-34060 Montpellier, France

⁴ Université of Genève, Université de Lausanne, Ecole des sciences pharmaceutiques, Switzerland.

⁵ ONIRIS, UR "Sensometrics and Chemometrics, F-44322 Nantes, France.

Keywords: ANOVA ; ASCA, plan d'expériences, technique Multi-tableau.

1- Introduction

La situation dans laquelle l'expérimentateur fait appel à la planification expérimentale et mesure un ensemble de réponses (spectre RMN, spectre de masse, ...) sur chacun de ses échantillons d'étude est aujourd'hui classiquement rencontrée. Pour effectuer le traitement des résultats obtenus, c'est à dire quantifier l'influence de chacun des facteurs du plan sur l'ensemble des réponses mesurées, plusieurs propositions sont rencontrées en littérature. Les plus abouties d'entre elles adoptent une stratégie proche de l'analyse de la variance (Anova) et peuvent être considérées comme des extensions de cette technique au cas multivarié. En parallèle à cette approche, une autre direction peut être exploitée en considérant qu'un jeu de données issu d'un plan d'expérience est structuré en blocs d'échantillons partageant les mêmes modalités pour certains des facteurs d'étude. Les techniques multi-tableaux, destinées à l'analyse simultanée de blocs d'expériences, semble donc a priori proposer une solution naturelle à la problématique abordée. Nous proposons dans ce travail une contribution originale destinée à montrer la pertinence de leur utilisation, en situant cette contribution par rapport à des méthodologies existantes.

2- Théorie

En complément des approches « naturelles » (Anova pour chaque réponse^[1], ACP suivie d'une Anova sur les scores obtenus), deux approches dédiées au traitement de plans d'expériences à réponse multivariées ont été récemment développées au sein de communautés "omiques" (protéomique et métabolomique): l'Anova-PCA^[2] et l'Anova-SCA^[3] (SCA : Simultaneous Component Analysis^[4]). Ces deux méthodes peuvent être considérées comme des extensions de l'Anova au cas multivarié et partagent le même point de départ.

Elles proposent, dans un premier temps, de diviser le jeu de données en une série de contributions (matrices), pouvant chacune être attribuée un facteur du plan ou à une de ses interactions. Une illustration de cette approche est présentée en Figure 1a.

Sur cette figure, les J mesures réalisées sur huit échantillons obtenus à partir d'un plan d'expériences croisant deux facteurs à 2 niveaux sont réparties en 4 matrices décrivant la contribution de chaque facteur, de leur interaction de la variance résiduelle. Dans certains cas, pour réduire le nombre de matrices à gérer et obtenir une analyse plus globale, la contribution d'un facteur et celle de ses interactions peuvent être assemblées dans la même matrice.

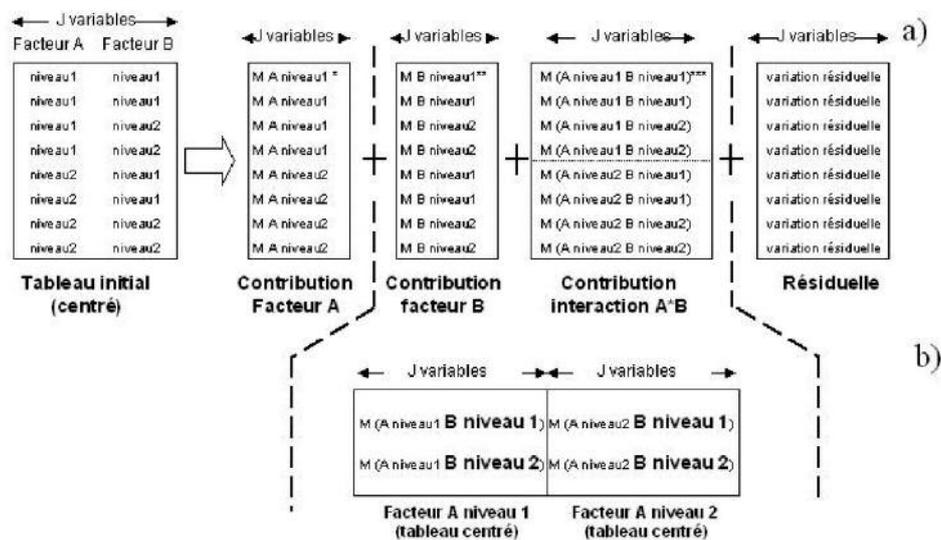


Figure 1

* = moyenne des expériences pour lesquelles A=niveau1
 ** = moyenne des expériences pour lesquelles B=niveau1; soustraction effet A
 *** = moyenne des expériences pour lesquelles A=niveau1 et B=niveau1 ; soustraction effet A, et B

A l'issue de cette première étape, pour l'Anova-PCA comme pour l'ASCA, l'effet des différents facteurs est modélisé indépendamment, chaque matrice/contribution faisant l'objet d'une ACP. Les scores des différentes modalités du facteur étudié sont alors comparés; l'ensemble des variations apportées aux mesures par un changement de modalité sont interprétées en utilisant les loadings correspondants. Si les deux méthodes sont très voisines, l'ASCA dans sa formulation, propose de détailler l'étude de l'interaction entre facteurs en faisant appel à l'analyse multi-tableaux. Ainsi, dans la Figure 1a, la matrice traduisant l'interaction A*B peut être divisée en deux sous-blocs réunissant les expériences ayant le même niveau pour l'un des deux facteurs (A ou B). Par exemple, si l'on définit, à partir de la matrice d'interaction, un premier bloc regroupant les expériences pour lesquelles A=1 et un second bloc regroupant les expériences pour lesquelles le niveau de A=2, une analyse multi-tableau de l'ensemble (réalisée par SCA) peut déterminer si l'influence du facteur B est la même quelle que soit la modalité prise par le facteur A. On peut remarquer que cette façon d'aborder l'interaction dans une situation multivariée est identique à celle utilisée pour l'Anova dans le cas univarié.

Toutefois cette idée n'est pas complètement exploitée dans la version de la méthode proposée aux utilisateurs. Nous proposons dans ce travail de la développer, d'une façon différente de celle proposée par les auteurs de l'ASCA. Notre proposition consiste en une nouvelle façon de caractériser l'interaction entre facteurs fondée sur :

- une « restructuration » de la table d'origine (restructuration présentée à titre d'illustration en figure 1b et portant sur le plan d'expérience décrit en figure 1a)
- l'utilisation de l'Analyse en co-inertie multiples (AcoM) [5].

3- Mise en œuvre

Une étude menée sur la plateforme Polyphénols de l'UMR Sciences pour l'Oenologie permet la mise en œuvre de l'approche proposée et d'en comparer les résultats à ceux obtenus en utilisant l'ASCA. Ce travail est destiné à caractériser l'influence de différents facteurs technologiques sur la composition polyphénolique de vins déterminée par spectroscopie.

Remerciements :

Les auteurs remercient :

- la Société Nomacorc pour le financement de cette étude
- D.J Vis, J.J. Jansen et D. Bertrand pour la mise à disposition du code Matlab nécessaire à la mise en œuvre des méthodes ASCA et ACOM.

Références

- [1] J. Boccard, E. Grata, A. Thiocone, J.Y. Gauvrit, P. Lanteri, P.A. Carrupt, J.L. Wolfender, S. Rudaz, *Multivariate data analysis of rapid LC-TOF/MS experiments from Arabidopsis thaliana stressed by wounding*, *Chem. and Intell. Lab. Sys.*, 86 (2007) 189.
- [2] P. de B. Harrington, N. E. Vieira, J. Espinoza, J. K. Nien, R. Romero, A. L. Yergey, *Analysis of variance-principal component analysis : A soft tool for proteomic discovery*, *Anal. Chem. Acta*, 544 (2005) 118-127.
- [3] A.K. Smilde, J.J. Jansen, H.C.J. Hoefsloot, R.J.A.N. Lamers, J. van der Greef, M.E. Timmerman, *Bioinformatics*, 21 (2005) 3043.
- [4] J.M.F. Ten Berge, H.A.M. Kiers, V. van der Stel, *Simultaneous component analysis*, *Statistica Applicata*, 4 (1992) 377.
- [5] D. Chessel, M. Hanafi, *Analyses de la co-inertie de K nuages de points*, *Rev. Statistique Appliquée*, XLIV-2 (1996) 35.

Discrimination PLS en présence de données à trois entrées

S.S.Ouertani¹ M.Hanafi¹ J.Boccard^{1,2} S.Rudaz²

¹ ONIRIS, Unité de Sensométrie et de Chimiométrie, site de la géraudière, BP 82225, 44322 Nantes cedex 3, France
samia-samar.ouertani@oniris-nantes.fr mohamed.hanafi@oniris-nantes.fr

² Ecole des sciences Pharmaceutiques, Université de Genève
Serge.Rudaz@unige.ch Julien.Boccard@unige.ch

Mots clés : :, discrimination, AD-PLS, tableaux trois entrées, Métabolomique, UPLC-TOF/MS.

1- Introduction

Décrire les différences entre des groupes prédéfinis est un objectif largement formulé. Il arrive que la phase d'acquisition des données conduit à des tableaux à trois entrées. Les exemples qui illustrent une telle situation sont nombreux. On cite, à titre d'exemple, les données à composante temporelle ou les données acquises par des instruments analytiques. Bien que les données à discriminer possèdent plus de deux entrées, le recours à l'analyse discriminante PLS² demeure toujours possible, par exemple si une matricisation (dépliage) est préalablement effectué sur le tableau à trois entrées. Cette stratégie est à l'évidence la plus indiquée et la plus simple à utiliser pour la discrimination des données à trois entrées. Cependant, elle ne prend pas en compte la nature multivoie des données et peut conduire à des variables latentes peu discriminantes.

La présente communication propose une extension de l'AD-PLS classique aux données à trois entrées. Pour se faire, la notion de variable latente associée à un tableau à trois entrées sera introduite. Cette nouvelle notion étend naturellement celle utilisée dans le cadre classique. La détermination des variables latentes tridimensionnelles passe par une nouvelle formulation de l'algorithme NIPALS basée sur la notion du produit n mode¹. Les règles d'interprétation de cette nouvelle démarche seront illustrées sur la base d'un exemple didactique en métabolomique.

1.1 AD-PLS3D

1.1.1 Notations et Définitions

Le produit n mode généralise le produit matriciel classique limité aux tableaux à deux entrées (matrices) à des tableaux à plusieurs entrées (tenseurs).

- $\underset{(n,p)}{\mathbf{A}} \bar{\mathbf{x}}_1 \underset{(n,1)}{\mathbf{v}} = \underset{(p,1)}{\mathbf{A}} \mathbf{v}$ (Produit selon le premier mode (1 mode) d'une matrice par un vecteur)
- $\underset{(n,p)}{\mathbf{A}} \bar{\mathbf{x}}_2 \underset{(p,1)}{\mathbf{v}} = \underset{(n,1)}{\mathbf{A}} \mathbf{v}$ (Produit selon le second mode (2 mode) d'une matrice par un vecteur)
- $\underset{(n,p,q)}{\mathbf{B}} \bar{\mathbf{x}}_1 \underset{(n,1)}{\mathbf{v}} = \underset{(p,q)}{\mathbf{C}}$ (Produit selon le premier mode (1 mode) d'un tenseur par un vecteur)
- $\underset{(n,p,q)}{\mathbf{B}} \bar{\mathbf{x}}_2 \underset{(p,1)}{\mathbf{v}} = \underset{(n,q)}{\mathbf{C}}$ (Produit selon le deuxième mode (2 mode) d'un tenseur par un vecteur)
- $\underset{(n,p,q)}{\mathbf{B}} \bar{\mathbf{x}}_3 \underset{(q,1)}{\mathbf{v}} = \underset{(n,p)}{\mathbf{C}}$ (Produit selon le troisième mode (3 mode) d'un tenseur par un vecteur)

1.1.2 Extension à trois entrées de l'algorithme NIPALS

La démarche méthodologique est résumée dans la figure 1 ci-dessous où Y désigne les indicatrices des groupes et X le tableau des mesures effectuées sur l'ensemble des individus.

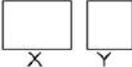
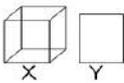
Structures des données			
Formulation	Formulation classique	Nouvelle formulation	Extension
Algorithme NIPALS	<i>initialisation</i> t_0 $w_y = Y^t t_0 \left(\ w_y\ = 1 \right)$ $t_y = Y w_y$ $w_x = X^t t_y \left(\ w_x\ = 1 \right)$ $t_x = X w_x$	<i>initialisation</i> t_0 $w_y = Y \bar{x}_1 t_0 \left(\ w_y\ = 1 \right)$ $t_y = Y \bar{x}_2 w_y$ $w_x = X \bar{x}_1 t_y \left(\ w_x\ = 1 \right)$ $t_x = X \bar{x}_2 w_x$	<i>initialisation</i> t_0 et w_0 $w_y = Y \bar{x}_1 t_0 \left(\ w_y\ = 1 \right)$ $t_y = Y \bar{x}_2 w_y$ $M = X \bar{x}_1 t_y$ $v_x = M \bar{x}_2 w_0 \left(\ v_x\ = 1 \right)$ $w_x = M \bar{x}_1 v_x \left(\ w_x\ = 1 \right)$ $t_x = X \bar{x}_2 v_x \bar{x}_2 w_x$
Déflation	$X^{(0)} = X$ $X^{(k+1)} = X^{(k)} - t_x^{(k)} t_x^{(k)T} X^{(k)}$	$X^{(0)} = X$ $X^{(k+1)} = X^{(k)} \times_1 \left(I - t_x^{(k)} t_x^{(k)T} \right)$	

Figure 1 - Reformulation et extension de l'algorithme de NIPALS d'un tableau à deux entrées à un tableau à trois entrées.

2- Application

2.1 Problématique de dopage à la testostérone

La testostérone est une hormone stéroïdienne, du groupe des androgènes. L'apport externe en testostérone a un effet bénéfique sur les sportifs dopés car c'est un psychostimulant qui augmente le potentiel de motivation du sportif. Elle permet également d'augmenter la masse musculaire et la force ainsi que la résistance à la fatigue. Elle est éliminée dans l'urine avec d'autres métabolites, dont un de ses isomères, l'épitestostérone. Interdit depuis 1984, l'apport de testostérone exogène est détectable par une analyse qui mesure le rapport testostérone/épitestostérone (T/E). Depuis 2004, le Comité International Olympique tolère ce rapport limite établi à 4, ceci dit il y a des cas pour lesquels l'interprétation du rapport T/E pose problème.

Une analyse de type UPLC-TOF/MS a été réalisé sur une étude clinique composée de 16 athlètes avec deux groupes expérimentaux correspondant à un groupe placebo (300 mg mannitol) et un groupe testostérone exogène (80mg + 115 mg mannitol). Trois fois par semaine, le lundi mercredi et vendredi, on donne du placebo à un groupe et de la testostérone à l'autre sous forme de comprimé administré par voie orale. Comme le montre figure 2, on procède à un extrait d'urine le premier jour (t00) avant de commencer l'étude clinique, et le 24^{ème} jour, on fait 4 extraits (t01, t02, t03, t04) avec 4 heures d'intervalle chacun. Le dernier extrait d'urine est effectué au bout de la 6^{ème} semaine à la date t05. On est donc en présence d'un tableau à trois entrées (16 individus repartis en deux groupes (testostérone, Placebo), 10 variables, 6 temps).

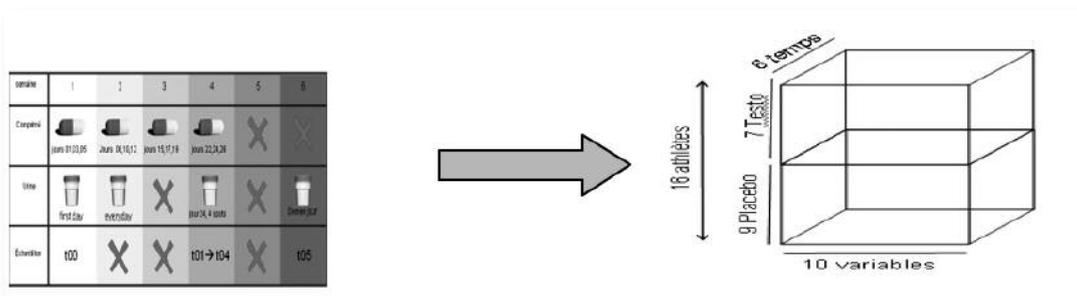


Figure 2 - Plan d'acquisition et structures des données.

L'application de l'AD-PLS3D permet de mettre en évidence une bonne discrimination des deux groupes (figure 3.a) à l'exception de l'individu 11 qui peut être comme aberrant. La figure 3.b indique clairement l'évolution de la discrimination dans le temps. Cette évolution uni modale sur la première dimension décrit trois phases : phase chronique, phase aiguë et une phase de rétablissement avec un retour à la normale. Les composantes discriminantes sont clairement identifiées sur la figure 3.c.

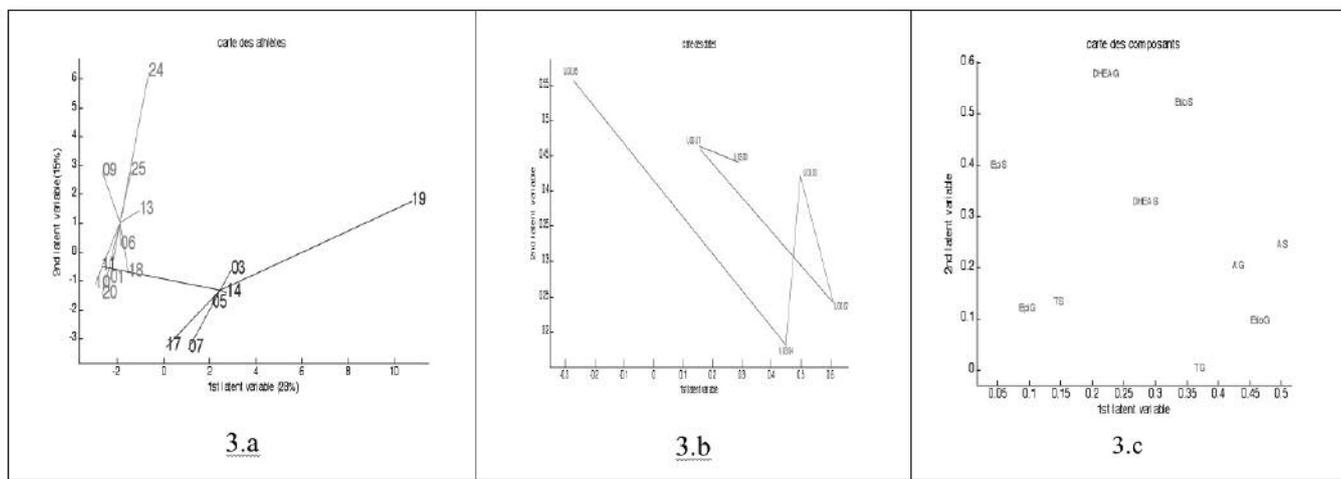


Figure 3 - Les cartes factorielle issues de l'AD-PLS3D.

CONCLUSION

L'AD-PLS3D étend d'une manière élégante AD-PLS classique. Elle permet de discriminer différents groupes de variables tout en préservant l'intégrité des données. Les premières applications sont prometteuses et méritent d'être élargies pour des données de grande dimension. L'aspect prédictif de la méthode fera l'objet des développements futurs.

Références

- [1] L.De LATHAUWER: Decomposition of a higher-order tensor in block terms, SIAM J, 2008.
- [2] Tenenhaus M. La regression PLS. Théorie et pratique, Edition Technip, Paris, 1998.

Combination of several mass spectrometry ionization modes: a multiblock analysis for a rapid characterization of the red wine polyphenolic composition

G. Mazerolles¹, S. Preys², C. Bouchut¹, E. Meudec¹, H. Fulcrand¹, J.M. Souquet¹, V. Cheynier¹

¹UMR 1083 Sciences pour l'Oenologie, INRA, 2 Place Viala, 34060 Montpellier, France, gerard.mazerolles@supagro.inra.fr

²Ondalys, 385 avenue des Baronnes, 34730 Prades-le-Lez, France, spreys@ondalys.fr

Keywords: Mass spectrometry, Fingerprinting, Polyphenols, Multiblock technique.

1- Introduction

Polyphenols are responsible for main sensorial properties of red wines (especially colour and taste). Nevertheless, due to the complexity and the diversity of the polyphenolic compounds involved, an exhaustive characterisation of the polyphenolic composition of a red wine is time consuming as it requires a large variety of standard analytical methods (High Performance Liquid Chromatography coupled with Diode Array Detection, thiolysis, ...).

In the present study, flow injection mass spectrometry was investigated as an alternative for a rapid characterisation of the polyphenolic composition of red wines. Atmospheric pressure chemical ionisation (APCI) and electrospray ionisation (ESI) (in both positive and negative ion modes) have been used for direct analysis of wines without any prior separation. In this way, four mass spectra were recorded for each wine. These four fingerprints related to the chemical composition were managed simultaneously using an appropriate multiblock technique, the common component and specific weights analysis (CCSWA), to combine the different MS measurement modes.

2- Multiblock analysis

A multiblock technique called Common Components and Specific Weights Analysis (CCSWA) can be used to deal with various types of measurements carried out on the same samples. This technique was presented in detail in previous works^[1,2] and has demonstrated its ability to identify distinctive and common information provided by different analytical techniques used simultaneously. The objective of CCSWA is to describe several data tables observed on the same n samples by recovering the maximum inertia (total variance) of each of them. In this way, the method can reveal structures that are common to the data tables and determines a common space of representation, described by common component and allowing an overall graphical representation of the samples based on the information provided by all the data sets. This representation can be seen as a compromise where the data tables can contribute at different levels. An interesting point in CCSWA is to associate a weight with each table for each common dimension. The weight corresponding to the common dimension j and the table i (noted λ_j^i) gives the contribution of the data table i in the building of the dimension j and represents similarly the part of variance in table i taken into account by the component j .

The equivalence between CCSWA and Hierarchical Principal Component Analysis (HPCA) was established recently^[3]. Considering the interpretation tools provided by the two methods, the common components in CCSWA can be considered as the super-scores in HPCA while both methods provide local scores and related loadings^[4].

In our situation, the hypothesis that underlies CCSWA is that each common dimension built using the four tables of mass fingerprints refers to one underlying latent characteristic (age of wine, vintage, variety, ...) which influences the polyphenolic composition and allows the discrimination of our groups of sample.

3- Material and methods

3.1 Wine samples

The whole database included sixty two French commercial red wines: forty of them originated from the Beaujolais region (3 different types: Beaujolais, Beaujolais Village, Beaujolais Crus) and were produced with Gamay grape variety. Twelve wines were produced with Gamay variety in other regions, including two Côteaux du Lyonnais produced in an area close to Beaujolais region, and ten wines were produced with other grape varieties in other regions.

3.2 Wine analyses

- Analysis of polyphenolic composition by reference methods:

The polyphenolic analysis of these wines is the result of a previous study^[6]. A total of 22 known polyphenolic compounds were characterized and quantified by a HPLC–DAD–MS (High Performance Liquid Chromatography–Diode Array Detection–Mass Spectrometry) method. The quantity and the mean degree of polymerization of tannins (proanthocyanidins) were determined using HPLC–DAD after thiolysis of a methanol extract.

- Analysis of wines by direct flow injection mass spectrometry:

Samples of wines were directly analyzed after bottle opening and introduced in the mass spectrometer without prior preparation. All the mass spectrometric measurements were performed on a mass spectrometer equipped with APCI and ESI interfaces.

3.3 Data pre-processing

The fingerprints were derived from the TIC (Total Ion Current) and exported for data pre-processing and data treatment. Three steps were realized, consisting in:

- (1) adjusting mass shifts from one spectrum to another by multivariate regression.
- (2) reducing the number of data points for each spectrum by selecting the maximum intensity value in each interval of one m/z unit
- (3) normalizing the intensities collected at each m/z unit for each spectrum.

4- Results and discussion

Figure 1 shows the projection of the 62 wine samples on the two first common components q_1 and q_2 . The values of the corresponding saliences for these two components are presented in Table 1. The first common component discriminates the Beaujolais Crus (the highest quality of Beaujolais) from the other Beaujolais wines. The second common component sets apart from Beaujolais wines the samples produced with a variety other than Gamay (OC group). The separation with the OC group seems even more significant for the 2003 vintage of Beaujolais than for the other vintages.

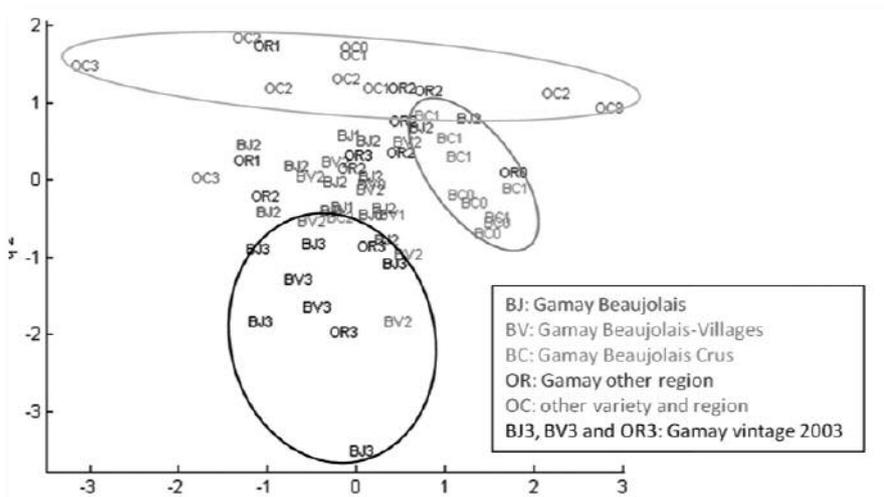


Figure 1 - Similarity map obtained for the 62 wine samples in the plane defined by the two first common components q_1 and q_2 .

	q_1	q_2
APCI-	0.40	0.07
APCI+	0.65	0.05
ESI-	0.10	0.42
ESI+	0.72	0.02

Table 1 - Saliences of the common components 1 and 2 for the CCSWA performed on the 4 MS data tables.

The corresponding loadings presented in figure 2 show that APCI+ spectra of Beaujolais Crus are characterized by a greater relative contribution of the peaks at $m/z = 289$ and 291 (fragment ions arising from the cleavage of proanthocyanidin interflavanic bonds) and a lower relative contribution of the peaks at $m/z = 493$, 535 , 639 and 331 (malvidin-3-glucoside, its acetylated and p-coumaroylated derivatives, and their aglycone fragment). The same conclusion can be obtained considering ESI+. The two methods agreed to characterize the Beaujolais Crus on the basis of an anthocyanin to tannin ratio lower than in the other Beaujolais wines. Table 1 shows that the common component q_2 mainly reflected the variability expressed in the ESI- table. The corresponding loadings show that ESI- spectra of Beaujolais wines are characterized by a greater relative contribution of the peaks at $m/z = 295$ and 311 (p-coumaroyl and caffeoyl tartaric esters respectively) and a lower relative contribution of the peaks at $m/z = 329$ and 591 . These results are consistent with some of the characteristics of the wine making in the Beaujolais region. Rapid methods can thus be established in relation to the objectives of the user. When the aim is to discriminate Beaujolais Crus, APCI+ or ESI+ ionization modes should be chosen. When the main goal is to differentiate the Gamay variety in the Beaujolais region, ESI- is the best analytical method.

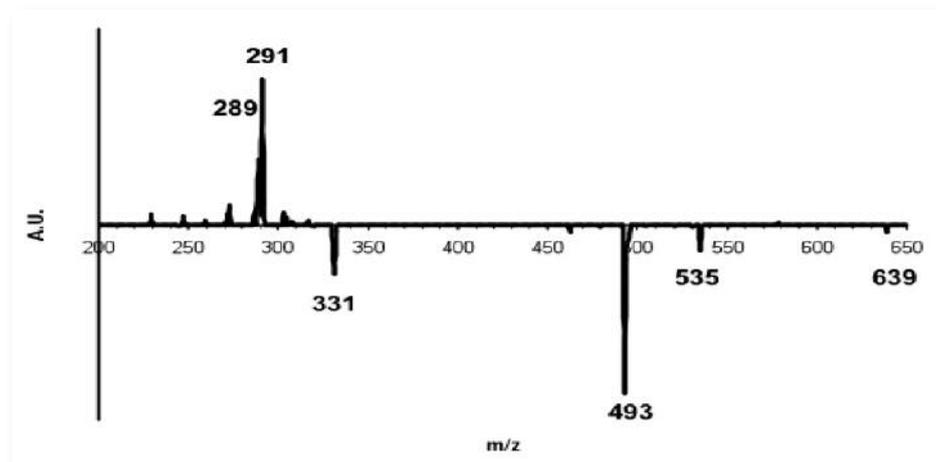


Figure 2 - Loadings associated to the first common component q_1 for the table APCI+.

CONCLUSION

A rapid method to characterize the polyphenolic profiles of red wines was developed using direct flow injection mass spectrometry. Attention was paid to validate that the information in the mass spectra was referring to polyphenolic compounds. These profiles, corresponding to different polyphenolic compositions, allowed to discriminate wines in relation with some interesting variability factors, such as the grape variety, the wine making and the age of the wines. The complementarity of information acquired on the samples by varying the ionization conditions was handled by an efficient chemometrical multiblock approach suited to this kind of situation. This technique appears as a very promising tool to deal with several data tables of multivariate signals in order to define, by combining the whole information obtained, an operating protocol according to the desired analytical objectives.

Références

- [1] E.M. Qannari, I. Wakeling, Ph. Courcoux, J.M. MacFie, *Food Quality and Preference* 11, 151, 2000.
- [2] G. Mazerolles, M. Hanafi, E. Dufour, D. Bertrand, E.M. Qannari, *Chem. and Intell. Lab. Sys.* 81, 41, 2006.
- [3] E.M. Qannari, M. Hanafi, A. Kohler, G. Mazerolles, *10^{ème} Journées Européennes Agro-industrie et Méthodes Statistiques*, Société Française de Statistiques, Louvain-la-Neuve, Belgique, 2008.
- [4] A.K. Smilde, J.A. Westerhuis, S. J. de Jong, *J. of Chemometrics* 17, 323, 2003.
- [5] S. Preys, G. Mazerolles, Ph. Courcoux, A. Samson, U. Fisher, M. Hanafi, D. Bertrand, V. Cheynier, *Anal. Chim. Acta* 563, 126, 2006.

Applications de modèles de mélanges gaussiens pour la classification supervisée de données spectroscopiques (de grande dimension)

C. Ruckebusch,¹ J. Jacques,² C. Bouveyron,³ O. Devos,¹ L. Duponchel¹ and S. Girard⁴

¹ LASIR CNRS-Université Lille1

² Laboratoire Paul Painlevé CNRS-Université Lille1

³ Laboratoire SAMM Université Paris I Panthéon-Sorbonne

⁴ MISTIS Laboratoire Jean Kuntzmann-INRIA Rhone-Alpe

Keywords: model-based classification, high-dimensional Gaussian model, generative model, vibrational spectroscopy

Abstract

Le développement de modèles de classification supervisée, pour attribuer des échantillons inconnus à des classes connues, reste une problématique d'un très grand intérêt en spectroscopie analytique. Les méthodes de classification peuvent être divisées en deux catégories. D'une part, les méthodes discriminatives (*e.g.*, SVM) pour lesquelles la règle de classification (la frontière) est estimée directement à partir des données. D'autre part, les méthodes génératives (*e.g.*, LDA, QDA), qui modélisent d'abord la densité des données complexes à l'aide de fonctions paramétriques (gaussiennes), puis construisent la règle de classification. Néanmoins, dans le cas de données de grande dimension, l'estimation d'un très grand nombre de paramètres est requise et des approches de régularisation doivent être couplées au critère de classification. Dans cette présentation, une méthode basée sur une famille de modèles génératifs gaussiens adaptés à la classification des données de grande dimension est présentée (High Dimensional Discriminant Analysis, HDDA) et appliquée à une problématique de classification multi-classes en spectroscopie proche infrarouge.

1- Introduction

Supervised classification, which aims at attributing unlabeled samples to known classes based on the knowledge of labeled learning samples, is of particular relevance in analytical spectroscopy. Classification methods can be divided into two categories. On the one hand, discriminative methods consist in estimating directly the classification rule. Support Vector Machines (SVM) is one of the most popular discriminative methods. On the other hand, generative classification methods model the complex system which has generated the observed data and then, from the modeling, build the classification rule. Linear Discriminant Analysis (LDA) is probably the most famous generative classification method. Unfortunately, conventional classification methods usually suffer from the "curse of dimensionality" in high-dimensional spaces.

An alternative and recent way for dealing with the problem of high-dimensional data classification is to model and classify the data in low-dimensional class specific subspaces. The Gaussian models for high-dimensional data, and their associated classification method HDDA (High-Dimensional Discriminant Analysis), defined in ^[1] and under study in a paper to be published soon^[2], allow to efficiently model and classify complex high-dimensional data in a parsimonious way. The key point is that the model complexity is controlled by the intrinsic dimensions of the classes. In contrast to other generative methods which incorporate dimensionality reduction or variable selection for dealing with high-dimensional data, HDDA does not reduce the dimension while modeling the data of each class in specific low-dimensional subspace.

2- Theory

Supervised classification aims to associate a new observation \mathbf{x} (a spectrum) with one of the k known classes through a learning set of labeled observations. We refer to^[12] for more details on the general classification framework. In this context, a popular approach is the use of the Gaussian mixture model which assumes that each class can be represented by a Gaussian density.

As in the classical Gaussian mixture model framework, we assume here that class conditional densities are Gaussian with means $\boldsymbol{\mu}_i$ the mean spectrum of the i -th class and covariance matrices $\boldsymbol{\Sigma}_i$ being the covariance matrix between the different wavelengths, for $i = 1, \dots, k$.

Let Q_i be the orthogonal matrix with the eigenvectors of Σ_i as columns. The class conditional covariance matrix Δ_i is therefore defined in the eigenspace of Σ_i by:

$$\Delta_i = Q_i^t \Sigma_i Q_i$$

The matrix Δ_i is thus a diagonal matrix which contains the eigenvalues of Σ_i . It is further assumed that Δ_i can be divided into two blocks as in Figure 1.

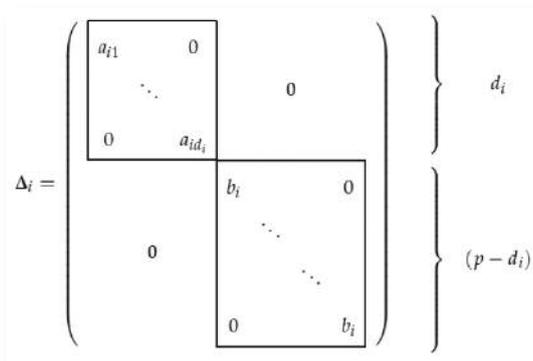
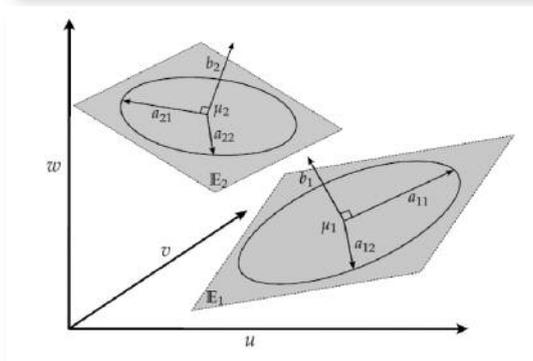


Figure 1 - The matrix Δ_i

With $a_{ij} > b_i$, $j = 1, \dots, d_i$, and where $d_i \in \{1, \dots, p-1\}$ is unknown.

Let us notice that assuming $d_i = (p-1)$ for all $i = 1, \dots, k$ leads to the classical Gaussian mixture model with full covariance matrices for each mixture component which yields in the supervised framework the well-known QDA.

Figure 2 - The parameters of the gaussian model $[a_{ij}b_iQ_i]$ in the case of two classes.



The class specific subspace E_i is defined as the affine space spanned by the d_i eigenvectors associated to the eigenvalues a_{ij} and such that $\mu_i \in E_i$. In this subspace E_i , the variance is modeled by the single parameter b_i .

By fixing some parameters to be common within or between classes, we obtain particular models which correspond to different regularizations. The family of models $[a_{ij}b_iQ_i]$ is divided into three categories: models with free orientations, common orientations and common covariance matrices.

The HDDA method is currently available through both stand-alone Matlab toolboxes and within the Mixmod software. For our study, we have used the Matlab toolboxes, available for download at <http://samm.univ-paris1.fr/~charles-bouveyron->. Alternatively, the Mixmod software provides eight of the most useful models presented in this article (available for download at <http://www-math.univ-fcomte.fr/mixmod/>).

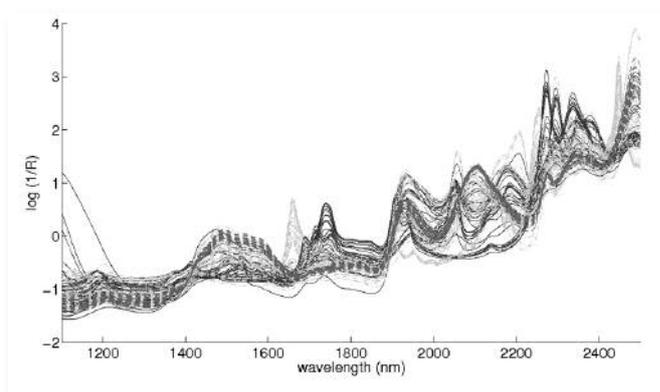
3- Materials and methods

The 3-class NIR data set contains 221 NIR spectra of manufactured textiles of various compositions, the classification problem consisting in the determination of a physical property which can take three discrete values^[7]. The samples were separated in a learning subset (130 samples) and a test subset (91 samples).

The NIR spectra were measured on a XDS rapid content analyzer instrument (FOSS) in reflectance mode in the range 1100-2500 nm at 0.5 nm apparent resolution (2800 data points per spectrum).

Prior to model development, Standard Normal Variate (SNV) was applied on the individual sample spectra as pretreatment. The SNV transformation consists of a centering and a reduction of each spectrum by its own standard deviation. Figure 3 shows the corresponding spectra.

Figure 3 - Spectra of the 3-class NIR learning set. Spectra are SNV pre-treated and colour coded according to class membership (blue solid line: class 1, green dashed line: class 2, red: class3).



4- Results and discussion

The results obtained for the fourteen HDDA models described in Section 2.2 are presented in Table 1. The choice of the model dimension (for models with fixed dimensions) and the choice of the threshold (for models with free dimensions) were done by a 5-fold cross validation.

For each model, the following results are presented: the correct classification rate on the learning subset estimated by 5-fold cross-validation (*learning CV-CCR*), the value of the BIC criterion, the correct classification rate on the test subset (*test CCR*) and the class specific subspace dimensions (d_i). Using cross-validation on the learning sample leads to select the three models with fixed dimensions and common variance outside the class specific subspaces: $[aijbQid]$, $[aibQid]$ and $[abQid]$, for which 92.3% correct classification rate was obtained.

The dimensions d_i retained for these models were the same for each class: 16. It should be noticed the good agreement between the cross-validation and the BIC criterion, retaining the model $[aijbQid]$, which is one of the models providing the best test CCR (96.7%). When compared to the results obtained with more classical chemometric methods, such as SIMCA (CCR 82.4%) and PLS-DA (CCR 87.7%), the results obtained with HDDA models show improved performances. The fact that these two methods showed poorer classification performance can be explained by the complexity of the problem where the three classes strongly overlap.

model	learning CV-CCR	BIC	test CCR	d_i
$[a_{ij}b_iQ_i d_i]$	85.4%	-1422737	83.5%	(5, 7, 6)
$[a_{ij}b_iQ_i d_i]$	91.5%	-2133807	94.5%	(16, 15, 14)
$[a_i b_i Q_i d_i]$	85.4%	-1422097	83.5%	(5, 7, 6)
$[a_i b_i Q_i d_i]$	85.4%	-1421695	83.5%	(5, 7, 6)
$[a_i b_i Q_i d_i]$	91.5%	-2128701	94.5%	(16, 15, 14)
$[a_i b_i Q_i d_i]$	91.5%	-2127724	94.5%	(16, 15, 14)
$[a_{ij}b_iQ_i d]$	85.4%	-1422205	82.4%	(8, 8, 8)
$[a_{ij}b_iQ_i d]$	92.3%	-2162226	96.7%	(16, 16, 16)
$[a_i b_i Q_i d]$	85.4%	-1420973	82.4%	(8, 8, 8)
$[a_i b_i Q_i d]$	85.4%	-1420407	82.4%	(8, 8, 8)
$[a_i b_i Q_i d]$	92.3%	-2156407	96.7%	(16, 16, 16)
$[a_i b_i Q_i d]$	92.3%	-2155267	96.7%	(16, 16, 16)
$[a_i b_i Q_i d]$	70.8%	-381399	73.6%	(3, 3, 3)
$[a_i b_i Q_i d]$	70.8%	-381367	73.6%	(3, 3, 3)
SVM	88.5%	-	91.2%	-
SIMCA	-	-	82.4%	-
PLS-DA	87.7%	-	84.7%	-

Table 1 - 3-class NIR data set: Correct classification rates on the learning sample evaluated by 5-fold cross-validation (CV-CCR on learning), BIC value and correct classification rates on the test sample (CCR on test), and dimensions of the class specific subspace for the fourteen HDDA models, SVM, SIMCA and PLS-DA.

CONCLUSION

This study^[2] has presented a generative discriminant analysis method, called HDDA, designed to the classification of high-dimensional data and applied to the classification of multi-class spectroscopic data sets.

HDDA is a generative discriminant analysis method based on a family of parsimonious Gaussian models which allow HDDA to be both flexible and efficient. For this, HDDA models and classifies the data in class specific and low-dimensional subspaces without reducing the data dimensionality. Therefore, no information loss is to be deplored due to data dimensionality reduction. This latter point is a specificity of HDDA, when compared to other generative methods, and it allows users to avoid the invasive dimension reduction step with PCA.

Références

- [1] Bouveyron C, Girard S, Schmid C. High Dimensional Discriminant Analysis. *Comm.Statist. Theory Methods* 2007; 36(14) :2607–2623.
 [2] Jacques j, Bouveyron C, Girard S, Devos O., Duponchel L., Ruckebusch C. Gaussian mixture models for the classification of high-dimensional vibrational spectroscopy data, *Journal of Chemometrics* DOI: 10.1002/cem.1355

Analyse canonique généralisée à noyau

Arthur Tenenhaus

SUPELEC, Département Signaux et Systèmes Electroniques, Plateau de Moulon, 3 rue Joliot-Curie, 91192, Gif-sur-Yvette, Cedex
Arthur.tenenhaus@supelec.fr

Keywords: Kernel, Regularization, Generalized Canonical Correlation Analysis, Partial Least Squares Introduction

A classical problem in applied statistics is to study relationships between several blocks of variables observed on the same set of individuals. Typical examples are found in large variety of fields such as bioinformatics, sensory analysis, marketing, food research where the common general objective is to extract "pattern(s)" of one block related to pattern(s) of others blocks. To study such a kind of relationships between blocks, the starting point of the paper is the regularized generalized canonical correlation analysis (RGCCA) proposed in ^[1]. RGCCA is a framework for modeling linear relationships between several blocks of variables with the constraint that blocks are not necessary fully connected. Considering a network of connections between blocks, the objective of RGCCA is to find linear combinations of block variables (block components) such that (i) block components well explain their own block and/or (ii) block components that are assumed to be connected are highly correlated. In the literature, many methods exist with the objective of finding block components verifying these properties. It appears that RGCCA constitutes a very general framework for multi-blocks data analysis and includes an astonishing large number of well-known methods as particular cases (among others, two-block case: Redundancy analysis, PLS regression, Regularized Canonical Correlation Analysis, multi-block case: SUMCOR, SABSCOR, SSQCOR, SUMCOV, SABSCOV, SSQCOV to name a few). However, RGCCA captures only linear relationships between blocks and to recover nonlinear relations, the Kernel Generalized Canonical Correlation Analysis (KGCCA) is proposed in this communication. There is no analytical solution to the KGCCA optimization problem and a monotone convergent algorithm is presented (i.e. the bounded criterion we want to maximize increase at each step of the iterative procedure) which is based on a modification of the PLS algorithm proposed by Wold ^[2]. One of the distinct advantage of KGCCA is that we this single algorithm all particular cases cited above are automatically kernelized. This communication present in details the KGCCA algorithm and the empirical usefulness of KGCCA is shown by analyzing two real dataset.

Références

[1] Tenenhaus A. and Tenenhaus M. (in press): Regularized Generalized Canonical Correlation Analysis, *Psychometrika*, in press.

[2] Wold H. (1985): Partial Least Squares, in *Encyclopedia of Statistical Sciences*, vol. 6, Kotz, S & Johnson, N.L. (Eds), John Wiley & Sons, New York, pp. 581-591.



Analyse stéroïdomique dans l'urine par UHPLC-QTOF-MS pour le dépistage anti-dopage

J. Boccard^{1,2}, F. Badoud^{2,3}, E. Grata^{2,3}, S.S. Ouertani^{4,5}, M. Hanafi^{4,5}, G. Mazerolles⁶, P. Lantéri⁷, J.-L. Veuthey^{1,2}, M. Saugy^{2,3}, S. Rudaz^{1,2}

¹School of Pharmaceutical Sciences, University of Geneva, University of Lausanne, Geneva, Switzerland

²Swiss Centre for Applied Human Toxicology, University of Geneva, CMU, Geneva, Switzerland

³Laboratoire Suisse d'Analyse du Dopage, Centre Hospitalier Universitaire Vaudois, Epalinges, Switzerland

⁴ONIRIS, Unité de Sensométrie et de Chimiométrie, Nantes, France

⁵Université Nantes Angers Le Mans, France

⁶INRA-UMR 1083 SPO, INRA, Montpellier, France

⁷Laboratory of Analytical Sciences, University Claude Bernard Lyon I, CNRS, Villeurbanne, France

Mots-clefs : UHPLC-QTOF-MS, lutte anti-dopage, stéroïdomique, N-PLS-DA, O-PLS-DA.

1- Introduction

Le profil stéroïdien est largement utilisé dans le cadre de la lutte anti-dopage pour déceler la prise exogène de produits destinés à améliorer les performances des athlètes. Cette approche qui repose sur le suivi de stéroïdes endogènes est toutefois limitée par l'utilisation d'un nombre restreint de biomarqueurs identifiés et reconnus par les autorités. Un profil étendu pourrait donc permettre d'améliorer la fenêtre de détection et d'augmenter la sensibilité des tests anti-dopage. Une méthode analytique associant la chromatographie liquide à ultra haute pression (UHPLC) et la spectrométrie de masse quadropole à temps de vol (QTOF-MS) a été développée pour l'analyse des stéroïdes endogènes dans l'urine. Cette méthode est appliquée aux échantillons d'une étude clinique ^[1] et l'ensemble de données obtenu comporte une structure à trois entrées (athlètes x stéroïdes x temps) et deux groupes (testostérone et placebo).

2- Matériel et méthodes

2.1 Echantillons d'urine

Une étude clinique a été réalisée pour évaluer les effets de l'administration de stéroïdes androgéniques anabolisants ^[1]. Plusieurs doses ont été administrées oralement durant un mois et des prélèvements de sang et d'urine ont été effectués de manière régulière. Six échantillons d'urine sont sélectionnés pour chaque athlète ayant reçu soit des pilules de placebo (300 mg de mannitol), soit des pilules contenant de la testostérone (80 mg de testostérone undecanoate et 115 mg de mannitol). Cette sélection vise à comparer un état basal (t00), un suivi cinétique sur 24H (t01 à t04) et un éventuel retour à la situation initiale (t05).

2.2 Analyse UHPLC-MS

Une extraction en phase solide est réalisée sur des cartouches Oasis HLB (Waters, Milford, MA, USA). Les échantillons sont ensuite analysés à l'aide d'un système UPLC Acquity (Waters, MA, USA) couplé à un spectromètre de masse Micromass-Q-TOF Premier (Waters, MA, USA) avec une interface électrospray (ESI). Les données sont acquises en mode ESI négatif pour des m/z entre 95 et 1000 Da.

2.3 Analyse de données

Les modèles N-PLS-DA sont calculés dans l'environnement MATLAB[®] 7 (The MathWorks, Natick, USA) à l'aide du N-Way Toolbox (version 3.1) [2]. Les modèles O-PLS-DA sont calculés avec le logiciel SIMCA-P (version 12, Umetrics, Umeå, Suède). Une validation croisée de type *leave-one-subject-out* est appliquée pour évaluer l'ajustement des modèles ainsi que le taux de bien classés, la sensibilité et la spécificité. La validité des modèles est confirmée à l'aide de tests de permutation.

3- Résultats et discussion

L'analyse de données proposée vise à évaluer l'apport d'un profil stéroïdien étendu et repose sur :

- (i) la construction d'un modèle discriminant appliqué sur la structure de données multivoies à l'aide de l'algorithme N-PLS afin d'établir la cinétique des changements observés après la prise de testostérone.
- (ii) l'analyse du profil stéroïdien traditionnel et l'extension à un profil étendu par la mise en évidence de nouveaux biomarqueurs avec la méthode O-PLS-DA et le SUS-plot.

3.1 Modèle N-PLS-DA du profil stéroïdien étendu

Une analyse stéroïdomique non-ciblée est envisagée pour évaluer l'apport d'un profil stéroïdien étendu. Les données brutes UHPLC-MS sont prétraitées à l'aide du logiciel MarkerLynx. Un filtre spécifique est appliqué aux 5'287 pics détectés automatiquement pour ne sélectionner que les m/z correspondants aux stéroïdes et à leurs conjugués (glucuronides et sulfates). Un modèle N-PLS-DA est construit sur le cube de données (16 athlètes x 260 pics x 6 temps). Deux variables latentes (LVs) sont calculées pour obtenir la meilleure capacité prédictive estimée par validation croisée. Les paramètres de performances suivants sont obtenus : taux de bien classés 63%, sensibilité 71% et spécificité 56%. Les graphiques LV1 vs. LV2 pour le mode athlètes (A) et le mode temporel (B) sont reportés sur la figure 1. Une séparation claire est observée dans le mode des athlètes, principalement sur la première LV. Un regroupement compact des athlètes du groupe ayant pris de la testostérone est obtenu, à l'exception des athlètes U03 et U11. Il est à noter que ceux-ci avaient déjà été décrits comme des profils stéroïdiens atypiques^[1]. Le mode temporel apporte quant à lui des informations précieuses concernant la cinétique du phénomène. A partir d'une situation basale, un état intermédiaire est visible à t01, une phase aigüe de la réponse est mise en évidence pour les temps de prélèvement t02 et t03 avant un retour à l'état intermédiaire à t04 puis à la situation de départ à t05.

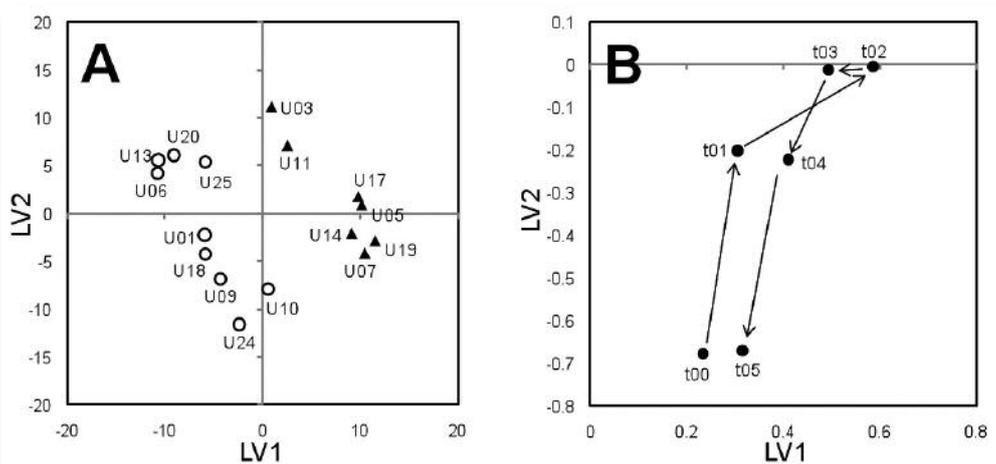


Figure 1 - Modèle N-PLS-DA du profil stéroïdien étendu. (A) Mode des athlètes et (B) mode temporel. Les athlètes du groupe placebo sont représentés par des cercles blancs (○) et les athlètes du groupe testostérone par des triangles noirs (▲). Les temps de prélèvement sont symbolisés par des cercles noirs (●).

3.2 Modèles O-PLS-DA et analyse du SUS-plot

L'analyse multivoie a permis d'obtenir des informations essentielles à la compréhension de la cinétique de métabolisation de la testostérone. Le modèle nécessite toutefois un suivi temporel précis de l'athlète, ce qui n'est pas compatible avec l'analyse anti-dopage de routine qui repose généralement sur un échantillon unique sans information temporelle. La matricisation de la structure de données est donc effectuée pour permettre la prédiction d'échantillons urinaires uniques. La contribution du profil stéroïdien étendu est évaluée en parallèle d'une quantification manuelle basée sur des standards de 10 stéroïdes conjugués connus dans chaque échantillon, c'est-à-dire un ensemble de 260+10 variables. Deux modèles O-PLS-DA sont construits pour comparer un groupe d'échantillon négatifs (classe NEG) qui comporte le groupe placebo et les temps t00 et t05 du groupe testostérone et respectivement les échantillons t01 et t04 (classe POS14, figure 2A) ou t02 et t03 (classe POS23, figure 2B) du groupe testostérone. L'O-PLS présente l'avantage de séparer la variation prédictive des différentes classes sur la première LV de la variabilité intra-classes, ce qui facilite l'interprétation biologique des contributions des variables^[3]. Les modèles obtenus présentent les performances suivantes :

- (i) modèle NEG vs. POS14: taux de bien classés 93.8%, sensibilité 61.5%, spécificité 100%.
- (ii) modèle NEG vs. POS23: taux de bien classés 93.8%, sensibilité 64.3%, spécificité 100%.

Un biais des deux modèles vers une haute spécificité est probablement dû à la proportion majoritaire d'échantillons de la classe NEG. Ceci n'est toutefois pas un problème et peut même constituer un avantage dans le contexte de l'anti-dopage où les faux positifs doivent être évités à tout prix. Un SUS-plot est ensuite construit sur la base du vecteur prédictif $\text{corr}(t_p, X)$ de chacun des modèles et la contribution des variables aux modèles est ensuite analysée au regard de leur position sur le SUS-plot (figure 2C). On peut observer l'augmentation de la concentration de 8 des 10 composés connus pour les classes POS14 et POS23 car ils sont situés dans le quadrant I (en haut à droite). Ces composés incluent les métabolites de la testostérone, de l'étiocolanalone, de l'androstérone et de la déhydroépiandrostérone. Seuls les deux métabolites de l'épitéstostérone ne présentent pas de contribution marquée. Il est à noter que ces derniers sont connus pour ne pas être affectés par la prise de testostérone et constituent une référence interne lors des tests de dépistage. Ces résultats sont donc en accord avec les connaissances biologiques du métabolisme de la testostérone. D'autre part, le SUS-plot permet de mettre en évidence des candidats biomarqueurs qui se situent dans cette même zone. Ceux-ci constitueront le sujet d'études ciblées pour leur identification. Grâce à la précision de mesure du rapport m/z obtenue par QTOF-MS (5-10 ppm), certains d'entre eux ont déjà pu être associés à des isomères probables de stéroïdes issus du métabolisme de la testostérone.

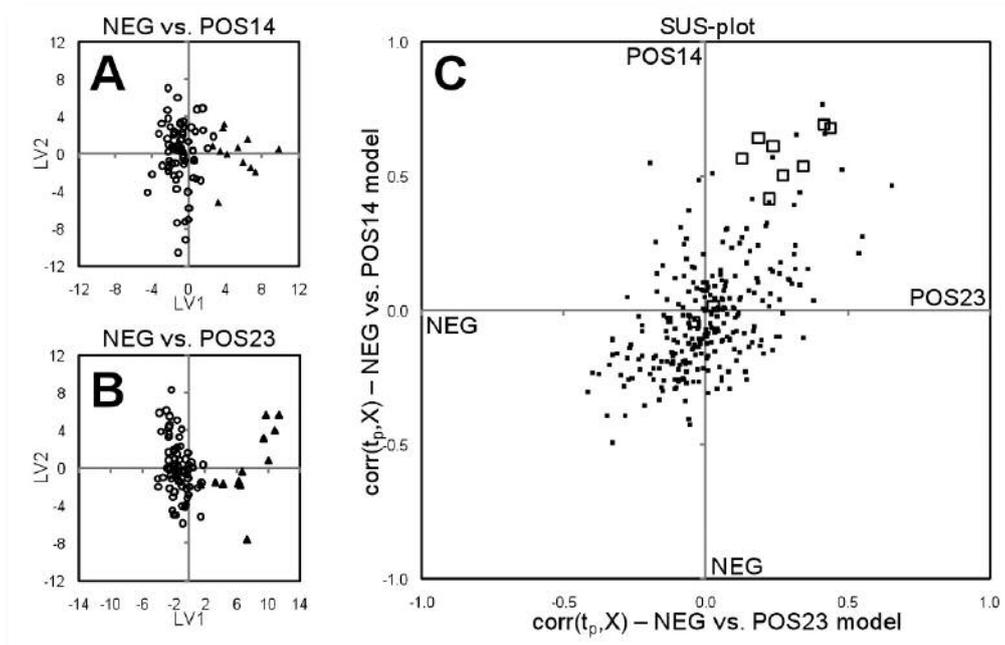


Figure 2 - Analyse O-PLS-DA (A) modèle NEG vs. POS14, (B) modèle NEG vs. POS23 et (C) SUS-plot. Les athlètes du groupe placebo sont représentés par des cercles blancs (○) et les athlètes du groupe testostérone par des triangles noirs (▲). Les métabolites de stéroïdes connus sont symbolisés par de gros carrés blancs (□) et les pics inconnus par de petits carrés noirs (■).

CONCLUSION

L'approche stéroïdomique proposée a permis de mettre en évidence la cinétique du métabolisme de la testostérone dans l'urine et de détecter de nouveaux candidats biomarqueurs stéroïdiens dans le cadre de la lutte anti-dopage.

Références

- [1] Baume N. et al. Effect of multiple oral doses of androgenic anabolic steroids on endurance performance and serum indices of physical stress in healthy male subjects. *Eur.J.Appl.Physiol.* 98, 329-340, 2006.
- [2] Andersson, C. A. et al. The N-way Toolbox for MATLAB. *Chemometr.Intell.Lab.* 52, 1-4, 2000.
- [3] Trygg, J. et al. Orthogonal projections to latent structures (O-PLS). *J.Chemometr.* 16, 119-128, 2002.

Influence de la levure, de la glucose oxydase et des farines de fève ou de soja sur les caractéristiques biochimiques de la pâte de farine de blé. Analyse en Composantes Principales et Distance de Malahanobis

Auteurs : A. Boussard^{1,3}, C.B.Y. Cordella^{2,3}, L. Rakotozafy^{1,3}, G. Moulin^{1,3}, F. Buche^{1,3}, J. Potus^{1,3} et J. Nicolas^{1,3}

¹ CNAM, Biochimie Industrielle et Agro-Alimentaire, case 306, 292, rue Saint-Martin, 75141 Cedex 03 Paris.

² AgroParisTech, UMR-A 1145 Ingénierie Procédés Aliments, F75231 PARIS Cedex 05–

³ INRA, UMR1145 INRA/AgroParisTech, F91300 MASSY–

aline.boussard@cnam.fr

Keywords: pétrissage, lipoxygénase, glucose-oxydase, levure, AGPI, TAG, oxylipides, pigments, ACP, Malahanobis.

1- Introduction

La discrimination de différentes farines de blé par l'étude de la consommation d'oxygène au cours du pétrissage a déjà été réalisée¹⁻³. Au cours du pétrissage, les réactions d'oxydoréduction qui se déroulent dans la pâte pendant sa formation ont un impact sur ses propriétés rhéologiques et organoleptiques. En particulier, la lipoxygénase (LOX) du blé catalyse l'oxydation des acides gras polyinsaturés (AGPI) libres de la farine par l'oxygène et génère des radicaux très réactifs susceptibles de cooxyder les pigments caroténoïdes et les protéines² (figure 1). Ces radicaux donnent naissance aux oxylipides constitués, notamment par les hydroperoxydes (ROOH), produits primaires formés lors de l'oxydation de l'acide linoléique en présence d'oxygène et en absence de réactions de cooxydation. On y place également des produits secondaires tels que les hydroxyacides (ROH) et les oxodiènes (RO) qui sont issus des réactions de cooxydation impliquant notamment les fonctions thiols portées par les protéines⁴ dans le cas des ROH.

D'autres composés, utilisés de façon systématique ou occasionnelle dans la formulation de la pâte à pain, entrent en compétition avec la LOX de blé vis-à-vis de la consommation d'oxygène. C'est le cas de la levure (figure 1), de la glucose oxydase (GOX) (figure 1) et des farines de fève ou de soja qui catalysent l'oxydation de certains constituants de la pâte, en particulier les oses et les AGPI libres ou liés aux triglycérides (TAG) et entraînent ainsi des modifications des propriétés de la pâte. Des pâtes de farine de blé ont été pétries en absence ou en présence de farines de fève ou de soja, en absence et / ou en présence de levure et / ou de GOX. Nous avons dosé dans la farine initiale et dans l'ensemble de ces pâtes, les composants biochimiques liés à l'activité LOX : les AGPI libres ou estérifiés aux TAG (TAG AGPI), les oxylipides et les pigments caroténoïdes.

Les objectifs de cette étude sont :

- d'explorer les relations existantes entre les différents échantillons de pâte préparés selon différentes formulations,
- de déterminer quels sont les facteurs biochimiques qui différencient ces pâtes et
- d'évaluer l'impact du type de LOX, de la levure et de la GOX sur ces facteurs biochimiques.

Nous utiliserons l'analyse en composantes principales (ACP) et la distance de Malahanobis intergroupes pour répondre à ces questions.

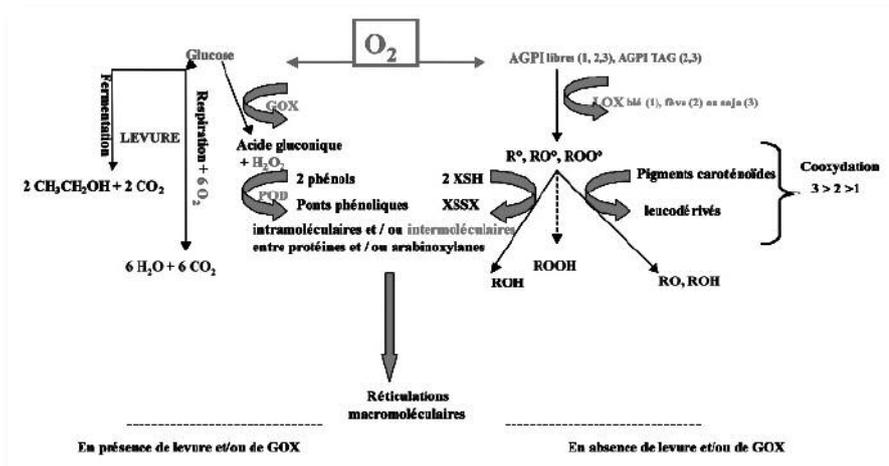


Figure 1 - Réactions biochimiques mises en jeu au cours du pétrissage d'une pâte de farine de blé

2- Matériels et méthodes

Echantillons et analyses biochimiques

La farine de blé BPMF (blés pour la meunerie française) provient des Moulins Soufflet (France), les farines de fève et de soja, des moulins AIT (France). La levure et la glucose oxydase (GOX) proviennent de la société PURATOS. La pâte de farine de blé, seule ou en mélange avec de la farine de fève (1 %) ou de soja (0,5 %) en absence et / ou en présence de levure (1 %) et / ou de GOX (40 ppm) est pétrie dans un consistographe (Chopin) pendant 7 minutes avec une hydratation de 60 %. Les échantillons de pâtes sont immédiatement congelés à - 20 °C.

Les AGPI, les TAG et les pigments caroténoïdes sont extraits des farines et des pâtes et quantifiés selon les méthodes décrites par Castello *et al.*⁵, les oxylipides, (hydroperoxyde (ROOH), hydroxyacide (ROH), oxodiène (RO)), selon⁶. Les teneurs en AGPI libres et estérifiés aux TAG ainsi qu'en oxylipides sont exprimées en $\mu\text{mol}\cdot\text{g}^{-1}$ de matière sèche (ms). Les teneurs en pigments caroténoïdes sont exprimées en unités d'absorbance $\cdot\text{g}^{-1}\cdot\text{ms}$. Deux extractions par pétrissage et deux analyses par extrait ($n = 4$) ont été réalisées. Les résultats sont analysés en termes de quantités consommées (teneur dans la farine – teneur dans la pâte) ou générées (teneur dans la pâte – teneur dans la farine).

Traitement des données

Les données ont été traitées sous Statistica, Statsoft, Inc 2001. La **matrices de données** est constituée de (12*4) échantillons avec 7 variables correspondant aux différentes pâtes de farine de blé pétries en absence (**T**) ou en présence de farines de fève (**TF**) ou de soja (**TS**), en absence et / ou en présence de levure (**TL**, **TSL**, **TFL**) et / ou de GOX (**TGOX**, **TLGOX**, **TFGOX**, **TSGOX**, **TFLGOX** et **TSLGOX**).

Les 7 variables correspondent aux paramètres biochimiques liés à l'activité LOX : les AGPI consommés libres (**AGPI cons**) ou estérifiés aux TAG (**TAG AGPI cons**), les TAG totaux (**TAG cons**), les oxylipides générés (**ROOH gen**, **ROH gen** et **RO gen**) et les pigments caroténoïdes oxydés (pigments cons).

De la matrice globale comprenant la totalité des données, trois matrices secondaires ont été judicieusement construites afin de comparer certaines formulations entre elles (Tableau 1).

Matrice	Contient	Modalité étudiée
M1	[T ; TL ; TF ; TFL ; TS ; TSL, TGOX]	Impact comparé de L, F, F+L, S, S+L et GOX
M2	[T ; TGOX ; TLGOX ; TSGOX ; TFGOX]	Impact comparé de GOX, L+GOX, S+GOX, F+GOX
M3	[T ; TFLGOX ; TSLGOX ; TGOX]	Impact comparé de F+L+GOX, S+L+GOX, GOX

L : levure ; F : fève ; S : soja

Tableau 1 - Matrices secondaires constituées à partir de la matrice de données initiales. La comparaison des effets se fait toujours par rapport au témoin (la farine non supplémentée) noté T.

Une ACP a été réalisée sur chacune de ces matrices afin de mettre en évidence le lien entre les échantillons et les mesures biochimiques réalisées.

Parallèlement, le calcul de la distance de Mahalanobis intergroupes ($D_{i,j}$), c'est-à-dire entre les échantillons de chaque type de formulation (7 groupes pour la matrice 1, 5 pour la matrice 2 et 4 pour la matrice 3) est réalisé après une analyse factorielle discriminante (AFD). Cela permet d'évaluer selon la méthode décrite ci-dessous, l'impact d'une formulation sur la pâte à pain obtenue. En effet, l'AFD maximise la variance intergroupe et minimise la variance intragroupe. Ceci a donc un impact sur la distance calculée entre les centres des groupes. Par conséquent, les distances intergroupes évoluent en fonction des groupes sélectionnés lors du calcul.

La méthode suivie se déroule en 4 étapes (figure 2 - exemple sur la matrice 1 (M1)):

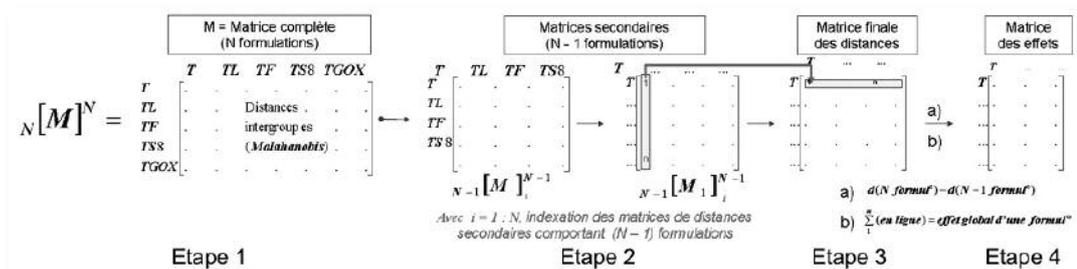


Figure 2 - Démarche en 4 étapes pour le calcul des effets (distance de Mahalanobis)

Etape 1 : La matrice des distances intergroupes est calculée à partir des 3 matrices initiales M1, M2, M3 comportant chacune des formulations différentes selon la formule classique :

$$D_{ij}^2 = (x_i - x_j) C^{-1} (x_i - x_j)^T$$

Avec i, j les index des objets i et j ; C matrice de covariance

Etape 2 : de façon itérative, une formulation de chaque matrice est retirée puis la distance intergroupes est recalculée (après une nouvelle AFD) afin d'évaluer l'impact de cette formulation (ou pâte) sur le positionnement des autres formulations.

Etape 3 : l'effet d'une formulation est calculé en retranchant la distance intergroupes obtenue pour une formulation donnée dans la matrice complète (N formulations) et la distance intergroupes calculée à partir des matrices comportant (N - 1) formulations.

Etape 4 : les distances ainsi obtenues correspondent aux effets de chaque formulation sur toutes les autres. La somme en ligne pour chaque formulation donne l'effet global d'une formulation.

Un graphe en histogramme permet la comparaison.

■ 3- Résultats et discussions

Analyse en composantes principales

La constitution d'une première matrice (M1 : T et TGOX, TL, TF, TFL et TS et TSL) de données a permis de différencier les pâtes en fonction de leur formulation. La composante principale 1 (CP1, 43 % de la variance totale) et la CP2 (31 % de la variance totale) montrent clairement que les facteurs **AGPI cons** et **ROOH gen** caractérisent principalement les échantillons de pâte avec levure seule (TL). La levure n'entre pas en compétition avec la LOX de blé pour la consommation d'oxygène. La levure apporte des thiols, il y a un effet protecteur de la levure envers les ROOH (figure 1). Les échantillons additionnés de farine de soja avec ou sans levure sont caractérisés par une forte valeur en **TAG AGPI cons** et en **TAG cons**. La LOX de soja oxyde les AGPI portés par les TAG⁸. Les pâtes additionnées de fève seule (TF) sont essentiellement expliquées sur CP2 par la teneur en **RO** et **ROH gen**. La farine de fève apporte des composés thiols de bas poids moléculaires (XSH) et notamment le γ -glu-cys qui n'existe pas dans la farine de soja (communication personnelle). Enfin, aucun effet GOX⁹ n'est observé (figure 3).

Une seconde matrice (M2 : T, TGOX, TSGOX, TFGOX et TLGOX) montre aussi une différenciation selon la formulation pour 4 groupes sur 5 confirmant l'absence d'effet de la GOX seule (graphe non présenté). La CP1 (63 % de la variance totale) indique que T (et TGOX confondu) et TLGOX sont surtout caractérisées par une forte valeur en **AGPI cons** puis dans une moindre mesure pour TLGOX par des teneurs élevées en **ROH** et **ROOH gen**. Les pâtes TSGOX et TFGOX se distinguent par des valeurs élevées en **TAG AGPI** et pigments cons et en **RO gen**. Les LOX de fève ou de soja ont un pouvoir de cooxydation des pigments supérieur à celui de la LOX de blé¹⁰. Les LOX de fève ou de soja ont une plus grande capacité à oxyder les AGPI estérifiés aux TAG que la LOX de blé¹¹.

Enfin, une troisième matrice (M3 : T et TGOX, TSLGOX, TFLGOX) permet aussi une bonne distinction des pâtes en 3 groupes, T et TGOX toujours confondus, TSLGOX et TFLGOX (graphe non présenté). La CP1 (50 % de la variance totale) indique que T et TGOX sont principalement associés aux **AGPI cons**, tandis que les TFLGOX le sont plutôt aux **ROOH** et **ROH gen** ainsi qu'aux **pigments cons**.

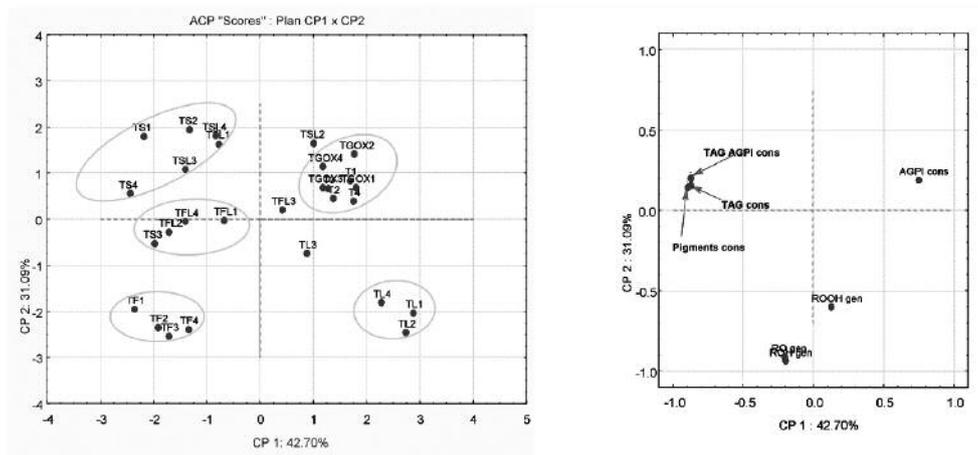
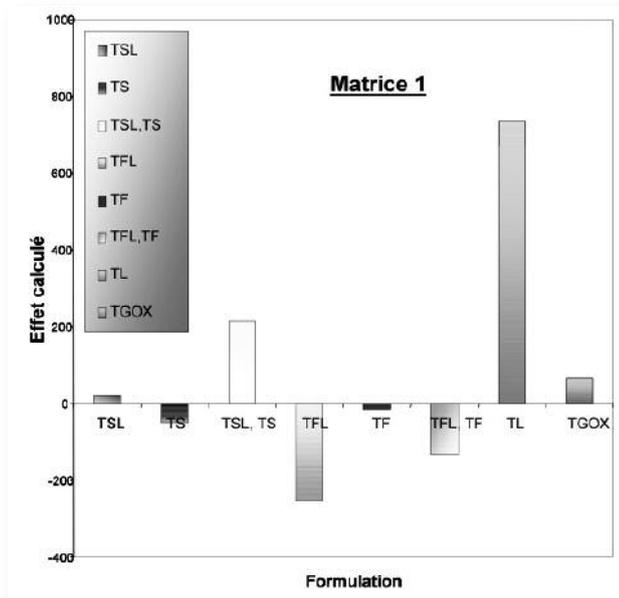


Figure 3 - Représentation des scores (échantillons) et des loadings (variables biochimiques) sur le plan CP1 x CP2 de M1.



Calcul de la distance de Malahanobis

Le calcul itératif de la distance intergroupes de Malahanobis appliqué séparément aux trois matrices complètes précédentes permet d'évaluer l'impact de chaque formulation sur la pâte. La figure 4 ne présente que le résultat obtenu pour la matrice 1.

Figure 4 - Histogramme représentant les effets des formulations de M1.

Les effets importants ressortant de cette méthodologie sont :

- 1) l'effet « **levure** » (matrice 1),
- 2) en présence de GOX, l'effet « **soja** » est le plus marqué (matrice 2),
- 3) en présence de levure et de GOX, un **effet antagoniste** entre les farines de soja et de fève est observé (matrice 3), ainsi qu'une synergie entre la farine de fève et la levure (matrice 3).

CONCLUSION

L'ACP a permis de discriminer les pâtes selon leur formulation. L'importance de l'origine végétale de la LOX est mise en évidence. L'oxydation des AGPI libres est anti-corrélée à l'oxydation des AGPI portés par des TAG. L'effet de l'ajout de la levure est mis en évidence par la variable ROOH. L'ensemble de ces résultats permet une meilleure compréhension des imbrications entre les mécanismes réactionnels mis en jeu par les systèmes enzymatiques endogènes de la farine (LOX du blé) et exogènes (LOX de fève ou de soja, GOX et levure) ainsi que leurs conséquences sur les caractéristiques biochimiques (et rhéologiques) de la pâte liées à l'activité LOX (figure 1).

Références

1. Levavasseur, L.; Rakotozafy, L.; Manceau, E.; Louarne, L.; Robert, H.; Baret, J. L.; Potus, J.; Nicolas, J., *Discrimination of wheat varieties by simultaneous measurements of oxygen consumption and consistency of flour dough during mixing*. *Journal of the Science of Food and Agriculture* 2006, 86, (11), 1688-1698.
2. Levavasseur, L. *Suivi simultané de la consommation d'oxygène et de la consistance des pâtes de farine de blé à l'aide d'un pétrin instrumenté (le sitoxygraphe) : tentative d'explication biochimique et rhéologique. Application à l'ajout de laccases*. *Biochimie, Option Sciences alimentaires, Paris 7, Paris*, 2007.
3. Eyoum, A.; Celhay, F.; Néron, S.; El Amrani, F.; A., B.; Poiffait, A.; Potus, J.; Baret, J.-L.; Nicolas, J., *Biochemical factors of importance in the oxygen consumption of unyeasted and yeasted wheat flours during dough mixing*. In *Recent Advances in Enzymes in Grain Processing*, Leuven, K. U., Ed. Courtin C.M., Veraverbeke, W.S. Delcour J.A.: Leuven, 2002; Vol. 46, pp 303-309.
4. Belitz, H.-D.; Grosch, W.; Schieberle, P.: *Food Chemistry*, Springer, Berlin, Third Edition 2004.
5. Castello, P.; Jollet, S.; Potus, J.; Baret, J.-L.; Nicolas, J., *Effect of Exogenous Lipase on Dough Lipids During Mixing of Wheat Flours*. *Cereal Chemistry* 1998, 75, (5), 595-601.
6. Bousard, A.; Buche, F.; Potus, J.; Nicolas, J., *Effet de l'addition de levure et / ou de glucose oxydases sur le devenir des oxylipides dans une pâte de farine de blé*. *Industrie des Céréales* 2009, 164, 24-25.
7. Lee, J.; Kaltsikes, P. J., *The application of Mahalanobis's generalized distance to measure genetic divergence in Durum wheat*. *Euphytica* 1973, 22, (1), 124-131.
8. Feussner, I.; Kühn, H.; Wasternack, C., *Lipoxygenase-dependent degradation of storage lipids*. *Trends in Plant Science* 2001, 6, (6), 268-273.
9. Moulin, G.; Buche, F.; Potus, J.; Nicolas, J., *Influence de la formulation des pâtes (levure, glucose oxydase) sur la consommation d'oxygène et sur l'oxydation des acides gras oxydables*. *Industries des Céréales* 2009, 164, 10-14.
10. Nicolas, J.; Autran, M.; Drapron, R., *Purification and some properties of wheat germ lipoxygenase*. *J. Sci. Food Agric.* 1982, 33, 365-372.
11. Nicolas, J.; Potus, J.; Garcia, R.; Davidou, S., *Oxidoreductases and Their Inhibitors*. Chap 11: *Enzymes and Enzyme Inhibitors Endogenous to Wheat* In *Wheat : Chemistry and Technology*, AACC, International Press, Eds. Khan K., Shewry P.R., St Paul, Fourth edition, 2009, pp 401-435.

Validation des modèles ICA

Delphine Jouan-Rimbaud-Bouveresse^{1,2}

Douglas N. Rutledge^{1,2}

¹ INRA, UMR 1145 "Ingénierie Procédés Aliments", 16 rue Claude Bernard, F-75005 Paris, delphine.bouveresse@agroparistech.fr
² AgroParisTech, UMR 1145 "Ingénierie Procédés Aliments", 16 rue Claude Bernard, F-75005 Paris, douglas.rutledge@agroparistech.fr

Keywords: Analyse en Composantes Indépendantes

La construction de modèles multivariés basés sur des variables latentes requiert une étape de *validation*, pendant laquelle on essaie de déterminer de façon objective le nombre de variables latentes optimal pour le modèle. La détermination de la dimensionnalité du modèle est une étape cruciale: si trop peu de variables latentes (ou facteurs) sont introduites dans le modèle, une partie de l'information significative des données n'est pas prise en compte, on parle alors de *sous-ajustement* ("*underfitting*") ; si, au contraire, trop de facteurs sont utilisés dans le modèle, on y introduit non seulement toute la variation significative dans les données, mais également de la variation aléatoire liée au bruit, et qui perturbe le modèle final: on dit alors que le modèle est *sur-ajusté* ("*overfitting*"). Le modèle optimal est celui dans lequel toute la variation importante est prise en compte, avec aussi peu de variation aléatoire que possible.

Les modèles multivariés utilisés traditionnellement en chimiométrie (Analyse en Composantes Principales, PCA; Régression PLS) sont souvent associés à la méthode de validation appelée *Validation Croisée* (CV). Pour la mettre en œuvre, la matrice de données \mathbf{X} ($N \times P$) est segmentée en G groupes, ($G \leq N$), et chaque groupe est tour à tour enlevé de \mathbf{X} , et A modèles sont construits à partir des $G-1$ groupes restant (avec de 1 à A facteurs), et les échantillons du groupe supprimé sont projetés sur chacun de ces modèles. Le nombre de facteurs conduisant à la meilleure projection des échantillons n'ayant pas été utilisés dans la construction du modèle correspond au nombre optimal à utiliser pour la construction du modèle final. Toutefois, cette méthode est plus difficile à mettre en œuvre avec l'Analyse en Composantes Indépendantes (ICA): En effet, lorsque des groupes différents sont enlevés de \mathbf{X} pendant la validation croisée, les ICs calculées pour un nombre donné de facteurs A peuvent être différents, ou dans un ordre différent. Il pourra donc en être de même pour le modèle final construit à partir de toutes les données dans \mathbf{X} . De plus, pour un même ensemble de groupes, la structure, l'ordre et les signes des ICs peuvent changer en fonction du nombre de facteurs.

Notre poster proposera des méthodes pour prendre en compte ces complications.

Adaptation and optimisation of a method of nanomolar Iron (III) determination on a deep-sea chemical analyser CHEMINI.

K. Bucas^{1,2} P. Giamarchi³

¹ Département Recherche Développement Technologique, IFREMER Centre de Brest, 29280 Plouzané, France
karenn.bucas@ifremer.fr

² Student, Master Optimisation des Protocoles Expérimentaux (OPEX) / Master's degree in the optimization of experimental protocols, département chimie UBO, 29238 Brest, France

³ UFR Sciences et Techniques, Université de Bretagne Occidentale, 29238 Brest, France
philippe.giamarchi@univ-brest.fr

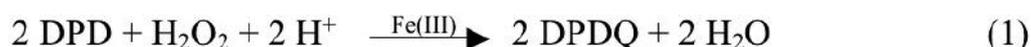
Keywords: flow injection analysis (FIA), Iron, in situ measurements, seawater, deep-sea chemical analyser, Hadamard screening design, D-Optimal experimental design, optimisation.

Introduction

The Iron, the fourth element the most abounding in the earth's crust with 5.6 % in mass^[1] is present in the state of track in the marine environment. Indeed, in the opened ocean, the concentrations of dissolved iron vary from 0.03-0.5 nmol.L⁻¹ in waters of surface, to 0.6-1.4 nmol.L⁻¹ in deep sea waters. They reach values more raised in coastal waters (from 1 to 100 nmol.L⁻¹)^[2], even very high in hydrothermal environment (from 2 μmol.L⁻¹ to 24 mmol.L⁻¹)^[3]. It is only recently that the iron contribution of the hydrothermal vents was considered as influential in the ocean^{[4],[5]}. The iron is the major metallic element of the hydrothermal fluid. This metallic load can then produce a particular flow important as well as a potential effect on the bodies of this extreme environment. Thus the Iron seems to be an element-key in the hydrothermal environment and it can serve as indicator in the chemical reactions of the hydrothermal ecosystem^[3]. There is a method of dosage of the iron by ferrozine using the technique of the FIA (Flow Injection Analysis) coupled with a photodetector^[3]. This method adapted on the deep-sea CHEMINI analyzer^[6], is used to characterize the warm fluids and their plumes. The search for characterization of the zone between the deep sea water and the plumes of the warm fluids, requires to have a method of dosage the range of measurement of which would from 0 to 200 nmol.L⁻¹ with a limit of detection at the level of the "track" (nanomole) ; what does not allow the current method the ferrozine. The method of dosage in situ by DPD developed for waters of surface, answers the criteria of popular concentrations^[7]. The objective of this study thus is to adapt and to optimize this method of dosage in a deep-sea CHEMINI analyzer.

Material and methods

The method of dosage of iron used is described by Hirayama and Unohara^[8] then successively adapted for FIA with spectrophotometric detection^[9], then for the in situ measurement^[7] adapted on a in situ analyzer. It's based on the reaction of oxidation of an aromatic amine, N, Ndimethyl-p-phenylenediamine (4-amino-N, N-diméthylaniline) dihydrochloride (DPD) by the peroxide of hydrogen in weakly acid environment, and catalysed by the Fe III (Equation 1). Both products of the oxidized amine are of pink color: N, Ndiméthyl-p-phenylenediamine semiquinones (DPDQ).



The CHEMINI analyzer used during the study is a deep sea version. The analyzer is built around two main modules : the hydraulic and the detection modules. The hydraulic box is equipped with three 2-way peristaltic pumps and with eight 3-port solenoid valves. The hydraulic system is consists of tube in Tygon[®] of internal diameter of 0,8 mm. The photodetector consists of a emission LED (wavelength : 570 nm), a detector photodiode, a cell of measurement in quartz.

All the results and graphs were calculated using software Nemrodw^[10].

Results and discussion

Screening design

To limit the number of runs and to determine the major effects, a screening study is necessary. A Hadamard screening design was used with twelve runs and eight factors. The answer was the absorbance (Abs). After a analysis of variance and a study of effects to valid the model, the selected factors were the six most influents (Figure 1) : SF (Stop flow), DPD (concentration of DPD), VPA, VPB, VPC (speed of pumps A, B and C), BI (length of injection loop).

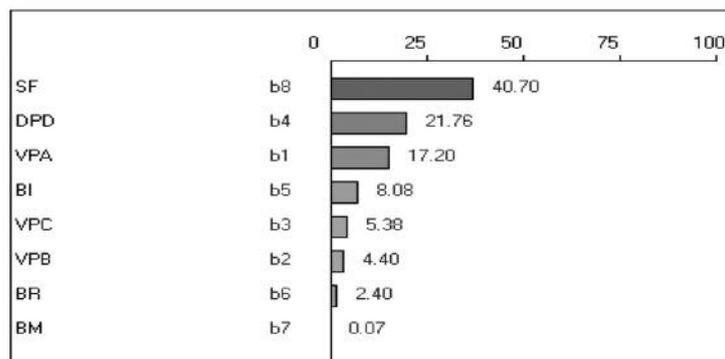


Figure 1 - Graph of Pareto for the screening design

Optimization

According to the several levels of factors (2 to 7), the limitation of the experimental domain and 2 postulated models (linear and quadratic), a D-Optimal matrix was selected, with 42 runs and 3 center points. Two answers was studied : Abs (absorbance) and Bruit (noise corresponding to the variation of the based-line).

For the answer Abs, the postulated model was quadratic. Two Fisher tests compose a complete analysis of variance : the first test ANOVA 1 compares regression variance with residual variance, the second test ANOVA 2 is a comparison between lack-of-fit and pure error variances. In our case for the Abs answer with a risk of 5%, the probability for ANOVA 1 is smaller than 5% and the probability for ANOVA 2 is bigger than 5% (Table 1). These 2 conditions permit to consider that the model was validated after verified a good repartition of residual plots.

Source of variation	Sum of squared	Degrees of freedom	Mean square	Ratio	Signif	
Regression	1.3685	26	0.0526 (Vreg)	9.7795	< 0.01 ***	ANOVA 1
Residual	0.0969	18	0.0054 (Vres)			
Lack of fit	0.0936	16	0.0059 (Vlof)	3.5877	24.0	ANOVA 2
Pure error	0.0033	2	0.0016 (Vpe)			
Total	1.4654	44				

Table 1 - Multiple regression analysis for Abs (95% confidence level)

For the second answer Bruit, the postulated model was linear with interactions. With a risk of 5 the analysis of variance was made. For ANOVA1, the probability is smaller than 5% and for the ANOVA 2, it's bigger than 5%. Thus, the model was validated after verified a good repartition of residual plots (Figure 2).

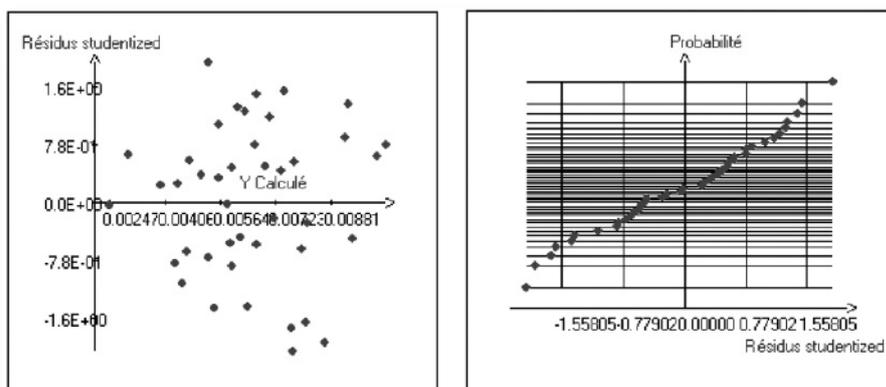


Figure 2 - Overview of residual (Bruit) : Residual plot and normal probability

CONCLUSION

The aims of this study were to adapt the nanomolar iron method on the in situ CHEMINI analyzer, and to optimize this method for the in situ applications and in the same time to simplify the using conditions. Chemical limits of the method were not obtained because the values of factors are in high level of experimental domain. Several physical and chemical limits of apparatus were not permit to extend it ; a new study must be made to permit it. Having determined the optimized values of the 6 factors by the tool of Nemrodw software, a study of linearity and detection limits were made in MilliQ water and in natural seawater. We obtained the similar results (range from 0 to 100 nM, detection limit of 1.9 nM, precision of 5% at 5 nM) than the method of reference [7] with worse conditions to simplify the implementation.

The obtained results by optimization using chemiometric tools, permit to use this method in hydrothermal environment but not in opened ocean where the iron concentration is in picomole.

Acknowledgment

A. Laës-Huon, P.M. Sarradin, C. Le Gall for their support.

Références

- [1] Taylor, S.R. Abundance of chemical elements in the continental crust : a new table. *Geochim. Cosmochim. Acta*, **1967**, 28, pp 1273-1285.
- [2] de Baar, H.J.W.; de Jong, J.T.M. Distributions, sources and sinks of iron in seawater. In *Biogeochemistry of Fe in Seawater* ; Turner, D.R., Hunter, K.A., Eds ; SCOR/IUPAC, **2001** ; Vol. Chapter 5, pp 123-253.
- [3] Sarradin, P.M.; Le Bris, N.; Le Gall, C.; Rodier, P. Fe analysis by the ferrozine method : Adaptation to FIA towards in situ analysis in hydrothermal environment. *Talanta*, **2005**, 66, pp 1131-1138.
- [4] Boyle, E.A.; Bergquist, B.A.; Kayser, R.A.; Mahowald N. Iron, manganese, and lead at Hawaii Ocean Time-series station ALOHA: Temporal variability and an intermediate water hydrothermal plume. *Geochim. Cosmochim. Acta*, **2005**, 69, pp 933-952.
- [5] Tagliabue, A.; Bopp, L.; Dutay, J.-C.; Bowie, A.R.; Chever, F.; Jean-Baptiste, P.; Bucciarelli, E.; Lannuzel, D.; Remenyi, T.; Sarthou, G.; Aumont, O.; Gehlen, M.; Jeandel, C. Hydrothermal contribution to the oceanic dissolved iron inventory. *Nature Geoscience*, **2010**, 3, pp 252-256.
- [6] Vuillemin, R.; Le Roux, D.; Dorval, P.; Bucas, K.; Sudreau, J.P.; Hamon, M.; Le Gall, C.; Sarradin, P.M. CHEMINI: a new in situ CHEMical MINIaturized analyser. *Deep-Sea Res. Part I*, **2009**, 56-8, pp 1391-1399.
- [7] Laës, A.; Vuillemin, R.; Leilde, B.; Sarthou, G.; Bournot-Marec, C.; Blain, S. Impact of environmental factors on in situ determination of iron in seawater by flow injection analysis. *Mar. Chem.*, **2005**, 97, pp 347-356.
- [8] Hirayama, K.; Unohara, N. Spectrophotometric catalytic determination of an ultra trace amount of iron (III) in water based on the oxidation of N,N-dimethyl-p-phenylenediamine by hydrogen peroxide. *Anal. Chem.*, **1988**, 60, pp 2573-2577.
- [9] Measures, C. I.; Yuan, J.; Resing, J.A. Determination of iron in seawater by flow injection analysis using in-line preconcentration and spectrophotometric detection. *Mar. Chem.*, 1995, 50, pp 3-12.
- [10] NEMROD-W Software, in: D. Matthieu, R. Phan Tan Luu (Eds.), Univ. Aix-Marseille III, **1999**.

Etude de sensibilité par la méthode de Morris a posteriori

Improved Sensitivity THrough Morris Extension

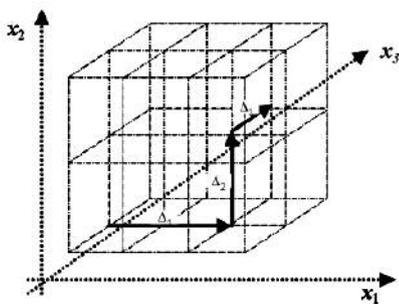
J. Santiago¹ B. Corre² M. Claeys-Bruno³ M. Sergent⁴

^{1, 3, 4} ISM2, UMR 6263, Laboratoire de Méthodologie de la Recherche Expérimentale, Université Paul Cézanne, Faculté des Sciences et Techniques, Avenue Escadrille Normandie Niémen, case D52, 13397 Marseille Cedex 20,
² TOTAL, EP/GSR/CORP, CSTJF Avenue Larribau, 64018 PAU Cedex

Keywords: Etude de sensibilité, Space-filling design, Méthode de Morris

1- Introduction

Parmi les méthodes de sensibilité, la méthode de Morris est souvent utilisée quand un nombre raisonnablement important d'expériences (ou de simulations) peut être envisagé (> 5k, avec k = nombre de facteurs). La méthode de Morris a principalement deux avantages : dans la plupart des cas, tous les facteurs influents sont détectés et en outre, elle révèle la présence de non linéarités ou d'interactions. Cette méthode est basée sur la construction d'un ensemble de points expérimentaux selon une disposition géométrique qui permet de calculer pour chaque facteur x_j un effet élémentaire $d_j(y)$

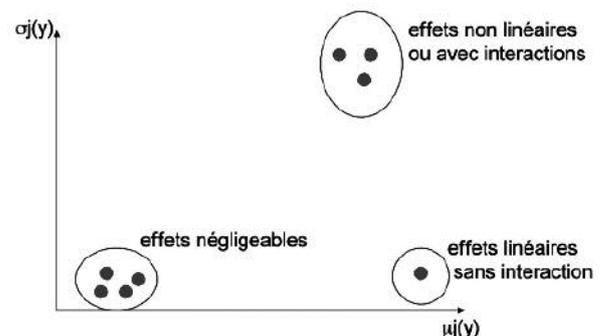


$$d_j(y) = \frac{y(x_1, \dots, x_{j-1}, x_j + \Delta_j, x_{j+1}, \dots, x_j) - y(x)}{\Delta_j}$$

Si l'on considère L trajectoires aléatoires, on obtient pour chaque facteur x_j :

- une moyenne absolue $\mu_j(y)$: $\mu_j(y) = \frac{1}{L} \sum_{\ell=1}^L |d_j^\ell(y)|$
- un écart type $\sigma_j(y)$: $\sigma_j(y) = \sqrt{\frac{1}{L} \sum_{\ell=1}^L (d_j^\ell(y) - \mu_j(y))^2}$

En fonction des valeurs obtenues pour μ_j et σ_j , les facteurs étudiés peuvent être classés en trois groupes selon leurs effets : effets négligeables, effets linéaires sans interaction, effets non linéaires ou avec interactions. Pour faciliter l'interprétation, les valeurs de $\mu_j(y)$ et de $\sigma_j(y)$ sont représentées graphiquement pour chaque facteur X_j . On peut alors définir trois zones bien distinctes :



2- Méthode a posteriori

2.1 Introduction

La méthode a posteriori permet de réaliser une étude de sensibilité selon le principe de la méthode de Morris, à partir d'un plan d'expériences déjà réalisé (si possible un plan uniforme). Pour cela, on cherche dans un espace à N dimensions toutes les constellations de N+1 points qui constituent un dièdre quasi orthogonal avec une distribution de longueurs de branches fixée.

2.2 Calcul des effets

Soit une fonction de deux variables, linéaires avec interaction, dans un espace bidimensionnel.

$$f(x_1, x_2) = ax_1 + bx_2 + cx_1x_2 \quad \text{Avec : } x_1 \in [n_1, m_1] \text{ et } x_2 \in [n_2, m_2]$$

Pour séparer les effets, il faut chercher une décomposition de f sous la forme :

$$f(x_1, x_2) = f_0 + f_1(x_1) + f_2(x_2) + f_{12}(x_1, x_2)$$

Où :

f_0 représente l'effet moyen

$f_1(x_1)$ donne l'effet principal de x_1

$f_2(x_2)$ donne l'effet principal de x_2

$f_{12}(x_1, x_2)$ donne l'interaction entre x_1 et x_2 .

L'effet moyen f_0 est donné par la moyenne de f, à savoir :

$$f_0 = \frac{1}{(m_1 - n_1)(m_2 - n_2)} \int_{n_2}^{m_2} \int_{n_1}^{m_1} f(x_1, x_2) dx_1 dx_2 = \frac{1}{(m_1 - n_1)(m_2 - n_2)} \int_{n_2}^{m_2} \left[\int_{n_1}^{m_1} f(x_1, x_2) dx_1 \right] dx_2$$

L'effet principal de x_1 est obtenu en retranchant de $f(x_1, x_2)$ l'effet moyen f_0 et en intégrant la fonction obtenue par rapport à x_2 , ce qui "annule" l'effet de cette variable.

$$\text{Ainsi : } f_1(x_1) = \frac{1}{(m_2 - n_2)} \int_{n_2}^{m_2} (f(x_1, x_2) - f_0) dx_2 \quad \text{et} \quad f_2(x_2) = \frac{1}{(m_1 - n_1)} \int_{n_1}^{m_1} (f(x_1, x_2) - f_0) dx_1$$

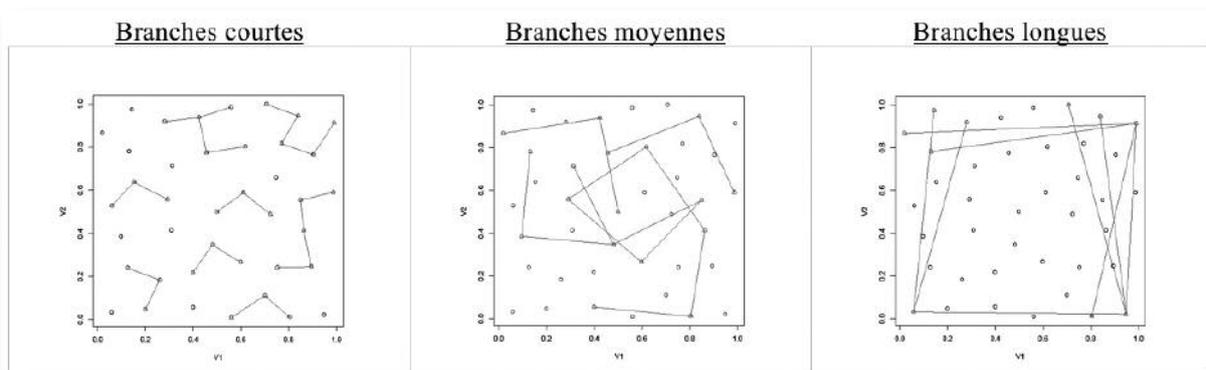
On peut ainsi calculer les effets principaux de X_1 et X_2 pour chaque dièdre et calculer finalement, les deux paramètres $\mu_j(y)$ et $\sigma_j(y)$.

Une étude exhaustive des paramètres des constellations (angle entre les branches et longueur des branches) a été réalisée sur un exemple à 2 et à 5 dimensions et montre la puissance de cette nouvelle méthode.

3- Résultats et discussion

3.1. Etude de la longueur des branches et de l'angle des dièdres

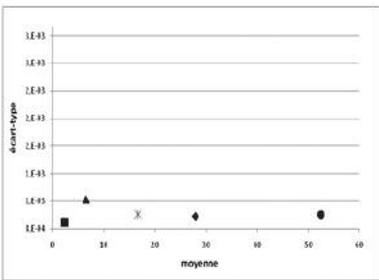
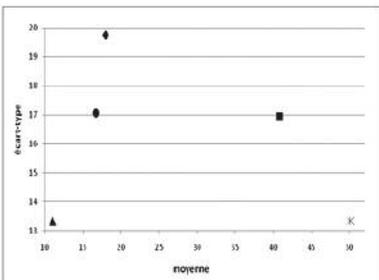
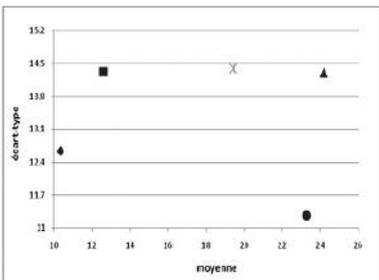
Les figures ci-dessous illustrent cette étude des longueurs de branches en deux dimensions :



On peut facilement observer qu'en fonction des longueurs de branches considérées et de l'angle entre deux branches, le nombre de dièdres varie fortement. Afin d'obtenir une étude statistique des deux paramètres $\mu_j(y)$ et $\sigma_j(y)$, nous avons considéré des longueurs de branches petites et des angles variant entre 70° et 110°.

3.2. Etude de cas à 5 dimensions par la méthode ISTHME

Les résultats présentés ci-dessous ont été réalisés à 5 dimensions à partir d'un plan uniforme WSP à 100 points avec 3 simulations différentes : modèle **degré 1** (Y_1), un modèle **avec interactions** (Y_2), un modèle avec interaction et terme carré (Y_3).

	Modèle Y_1	Modèle Y_2	Modèle Y_3		
Résultats					
	◆ facteur X_1 ■ facteur X_2 ▲ facteur X_3 ● facteur X_4 * facteur X_5				
Conclusion	X_1 : effet linéaire X_2 : effet négligeable X_3 : effet négligeable X_4 : effet linéaire X_5 : effet linéaire	X_1 : effet non linéaire ou interaction X_2 : effet non linéaire ou interaction X_3 : effet négligeable X_4 : effet non linéaire ou interaction X_5 : effet linéaire	X_1 : effet négligeable X_2 : effet non linéaire ou interaction X_3 : effet non linéaire ou interaction X_4 : effet linéaire X_5 : effet non linéaire ou interaction		
Coeff. du modèle simulé	$Y_1 = 240.4 + 27.9 X_1 + 3.0 X_2 - 6.5 X_3 - 52.4 X_4 + 16.6 X_5$	$Y_2 = 278.5 + 10.7 X_1 + 41.4 X_2 + 3.7 X_3 + 10.4 X_4 - 52.9 X_5 + 38.7 X_1 X_2 - 31.9 X_1 X_4$	$Y_3 = 109.2 + 4.1 X_1 - 7.2 X_2 + 18.7 X_3 - 23.3 X_4 + 15.9 X_5 - 26.5 X_2 X_5 + 28 X_3^2$		

Pour chacune des simulations, les effets (négligeables, linéaires, non linéaires ou interactions) sont clairement et correctement identifiés sur les graphes par rapport au "vrai" modèle simulé.

CONCLUSION

Cette nouvelle approche, ISTHME, donne de bons résultats en analyse de sensibilité. L'avantage de cette méthode par rapport à la méthode "a priori" proposée par Morris réside dans l'utilisation d'un ensemble de points existant (et plus particulièrement un plan uniforme) pouvant donner lieu à d'autres études (modélisation, krigeage...) sans simulation supplémentaire.

Références

Morris MD: Factorial sampling plans for preliminary computational experiments. *Technometrics* 33, 161-174 (1991)
 Saltelli A, Scott EM, Chen K, editors: *Sensitivity analysis*. Wiley, 2000
 Santner TJ, Williams BJ, Notz WI: *The design and analysis of computer experiments*. Springer, 2003

Etude cinétique et chimiométrie multivoies pour l'évaluation de la stabilité thermique d'huiles : comparaison de méthodes

Auteurs : C. B.Y. Cordella^{1,2,*}, T. Tekye¹, D.N. Rutledge^{2,1}, Riccardo Leardi³

¹ INRA/AgroParisTech UMR1145 Génial, Laboratoire de Chimie Analytique, 16 rue Claude Bernard, F-75005 PARIS,

² AgroParisTech/INRA UMR1145 Génial, Laboratoire de Chimie Analytique, 16 rue Claude Bernard, F-75005 PARIS

³ Department of Chemistry and Food and Pharmaceutical Technologies, University of Genoa, Via Brigata Salerno (ponte), I-16147 Genoa, Italy.

* e-mail : christophe.cordella@paris.inra.fr

Mots clés : Keywords: huiles végétales, stabilité thermique, cinétique, ACP, PARAFAC, Tucker3, 4-voies.

1- Introduction

Cette étude porte sur le chauffage de trois types d'huiles alimentaires (colza, tournesol, olive) soumises à des températures élevées. Diverses réactions chimiques participent au vieillissement chimique et physique de ces huiles. Pendant la friture, l'huile se détériore et ses caractéristiques fonctionnelles et organoleptiques sont significativement modifiées. Le chauffage induit un ensemble complexe de réactions chimiques telles que l'oxydation, la polymérisation, l'hydrolyse, ou encore l'isomérisation cis/trans, produisant des composés nocifs pour la santé ^[1, 2]. La cause la plus importante de détérioration des huiles est l'oxydation, celle-ci réduit sensiblement la durée de vie et les valeurs nutritionnelles des huiles. Parmi les produits d'oxydation des huiles, notre intérêt s'est porté plus particulièrement sur les produits d'oxydation secondaires tels que les aldéhydes car ceux-ci sont rarement présents dans l'huile naturelles non chauffées ^[3].

Les objectifs de cette étude sont :

- évaluer la pertinence d'un nouveau critère de stabilité thermique des huiles chauffées, fondé sur le suivi cinétique du chauffage des huiles par résonance magnétique nucléaire du proton,
- comparer les résultats de l'approche cinétique à un traitement chimiométrique classique (ACP) et à une approche multivoies (PARAFAC et Tucker3).

2- Matériels et méthodes

2.1 Echantillons

Trois types d'huiles végétales ont été analysés. Des huiles de colza, de tournesol et d'olive ont été achetées au supermarché et utilisées dans une étude de thermo-oxydation. Environ 12 ml d'huile ont été placés dans des boîtes de pétri en verre de 10 cm de diamètre et soumis à un chauffage contrôlé en étuve. Chaque type d'huile a été chauffé à 170 °C, 190 °C et 210 °C correspondant à des températures proches des températures de friture domestiques. Des prélèvements de 1g ont été réalisés toutes les 30 min jusqu'à la fin du chauffage fixée à 180 minutes, entraînant un total de 189 échantillons à analyser. Les échantillons ont été refroidis au bain-marie glacé pendant 4 minutes afin de stopper ou à tout le moins de réduire le plus possible les réactions d'oxydation thermique avant de procéder à l'analyse des échantillons.

2.2 Analyse RMN-1H

Une masse comprise entre 0,3 à 0,5 g d'huile a été insérée dans un tube de RMN (5 mm de diamètre) avec 700 µl de chloroforme deutéré. Le spectre de résonance de l'hydrogène a été obtenu sur un spectromètre Bruker Advance 300 MHz UltraShield avec un champ magnétique de 7,05 T. Une séquence écho de spin de base a été appliquée pour obtenir une précession libre capable de produire les spectres RMN-1H des huiles étudiées. L'expérience a été réalisée à 25 °C. Les spectres ont été acquis régulièrement tout au long du processus d'oxydation.

2.3 Traitement cinétique

Le premier traitement appliqué sur les spectres a été l'intégration de la surface de chaque pic d'intérêt entre 9 à 9,8 ppm correspondant aux protons aldéhydiques de composés chimiques produits au cours du chauffage. Le rapport entre cette

valeur de surface et celle correspondant au pic du solvant à 7,25 ppm (CDCl₃) a été utilisée comme donnée de base dans les calculs de cinétiques. La valeur calculée des rapports a été corrigée en tenant compte de la masse pesée de CDCl₃ et d'huile pour supprimer l'effet lié aux petites variations dues à la pesée.

2.4 Pré-traitement des données

Les spectres ont été recalés en utilisant l'algorithme COW (*correlation optimize warping*) et une correction de la ligne de base a été réalisée par un ajustement polynomial. Après avoir calculé la moyenne des trois reliquats et un centrage des données, la matrice initiale est constituée de 63 (objets) x 1001 (variables).

3- Résultats et discussion

A. Approche cinétique

L'hypothèse initiale est de considérer que les produits d'oxydation tels que les aldéhydes (composés B) apparaissent selon une loi de vitesse d'ordre 1. L'apparition des aldéhydes suite au chauffage conduit à une augmentation de leur concentration B_t , au cours du temps que l'on exprime par les relations habituelles et notamment :

$$\ln\left(\frac{B_\infty}{B_\infty - B_t}\right) = \frac{1}{K} t$$

B_t = rapport d'aires entre le pic aldéhydique à 9,4ppm et le pic du solvant (CDCl₃) à l'instant t. $B_\infty = B$ à t=3h.

K est la constante de stabilité thermique de l'huile étudiée. $K = 1/k$, si k est petit (constante cinétique faible), alors K est grand, l'huile a donc une plus grande stabilité thermique. Les résultats pour les trois huiles sont donnés dans le tableau 1, ci-dessous.

Huile	K (min ⁻¹)		
	170°C	190°C	210°C
Colza	41.3	37.3	37.3
Olive	138.9	68.5	45.7
Tournesol	83.3	64.9	30.4

Tableau 1 - Valeurs de K pour les trois huiles en fonction du pic RMN choisi pour le calcul d'aire ($\delta = 9.4$ ppm)

Le tableau 2 présente l'ordre de stabilité que l'on peut proposer en tenant compte de cette constante K .

Pics aldéhydiques à 9,4 ppm	Température		
	170°C	190°C	210°C
	HO >> HT > HC	HO >= HT > HC	HO > HC >= HT

Tableau 2 - Ordre de stabilité thermique des huiles étudiées, selon K .
HO : huile d'olive, HC : huile de colza, HT : huile de tournesol.

L'ordre de stabilité déterminé par cette méthode est en bon accord avec les données bibliographiques concernant la stabilité thermique des huiles de colza, olive et tournesol [4-6] réalisées par RMN ou IR. En effet, il a été observé, dans ces études, que l'huile d'olive est toujours plus stable que les autres types d'huile grâce à sa forte teneur en antioxydants de type tocophérols. Et d'autre part, entre l'huile de colza et l'huile de tournesol, les études en conditions oxydantes montrent que l'huile de tournesol est légèrement plus stable que l'huile de colza.

B. Approche chimométrique

L'ACP classique sur cette matrice montre les trajectoires en fonction du temps et de la température des trois huiles et près de 80% de la variance totale est contenue dans les 3 premières composantes. L'interprétation conjointe des « loadings » indique que la dégradation thermique est relativement similaire pour les huiles de colza et de tournesol mais qu'elle n'emprunte pas les mêmes voies chimiques pour l'huile d'olive.

Les données étant au départ à 4 voies (3 huiles x 1001 variables x 3 T° x 7 tps de chauffage), deux réarrangements supplémentaires ont été réalisés pour un traitement multivoies : (Huiles x tps) x variables x T°, (21 x 1001 x 3) et (Huiles x T°) x variables x tps, (9 x 1001 x 7). Dans les deux cas, un modèle multivoies [2 2 1] est suffisant (Cf. figure 2) et

récupère près de 40% de la variance totale et permet une interprétation simultanée des 3 modes dans lesquels on retrouve de façon claire et cohérente les résultats de la cinétique chimique classique et de l'ACP. Les résultats principaux obtenus sont que l'huile d'olive est beaucoup plus stable que les huiles de colza et de tournesol et que l'écart observé entre un chauffage à 190°C et 210°C est plus important que celui entre 170°C et 190°C.

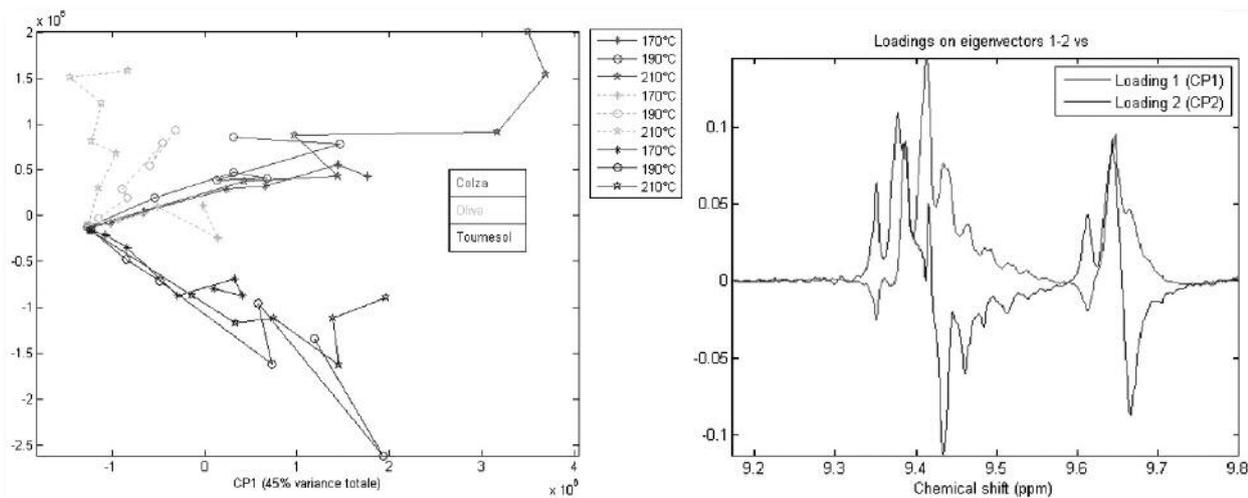
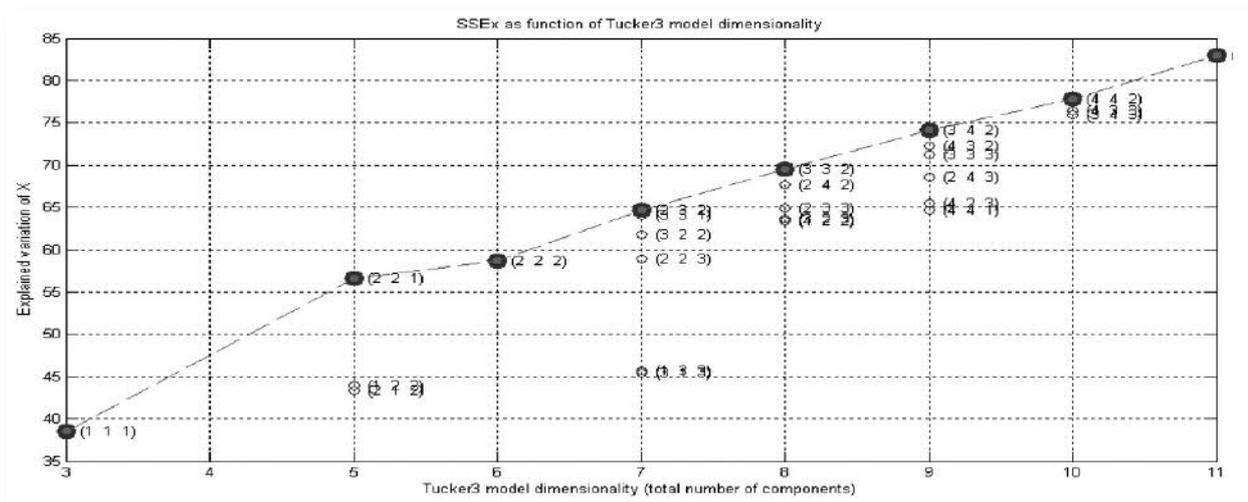


Figure 1 - ACP matrice globale (63 x 1001) des huiles chauffées. Plan factoriel CP1 vs CP2.

Modèles (21x1001x3) : (Huiles x tps) x variables x T°



Modèles (9x1001x7) : (Huiles x T°) x variables x tps

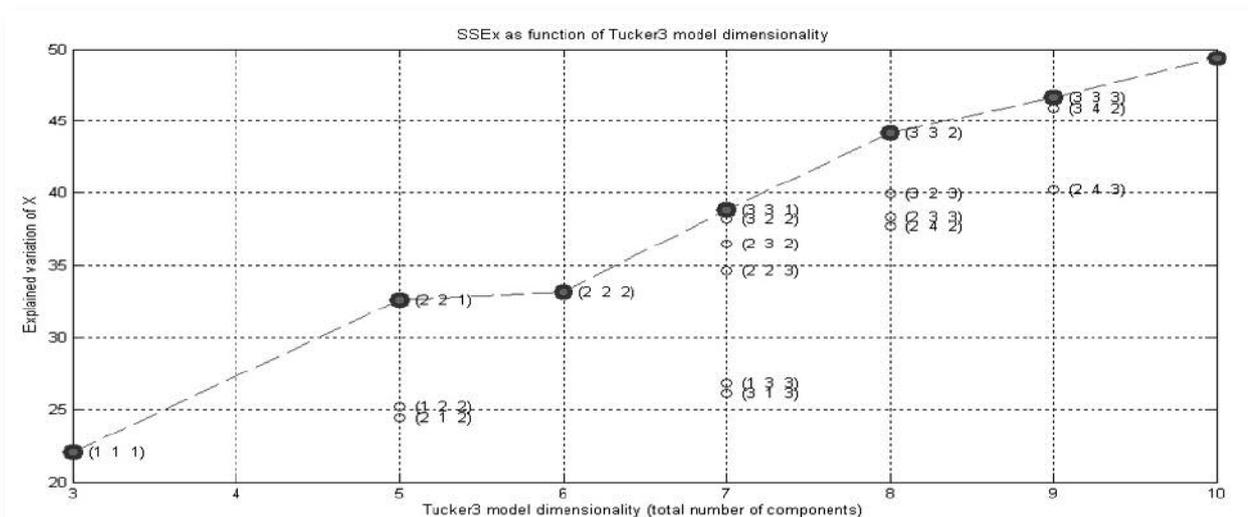


Figure 2 - Choix du modèle 3-voies optimal pour Tucker3 par la procédure TuckTest de la N-Way Toolbox v3.1

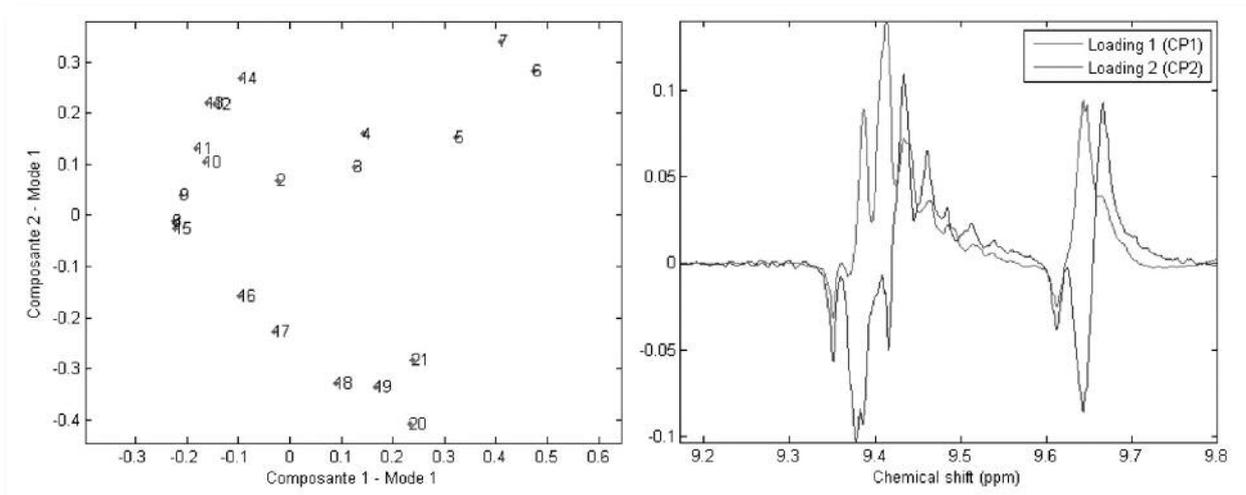


Figure 3 - **A gauche**, Tucker3- scores des composantes 1 et 2 du mode 1 (échantillons). [1-7] : Colza ; [8-14] : Olive ; [15-21] : Tournesol. **A droite**, Tucker3-loadings des composantes 1 et 2 du mode 2 (variables=déplacements chimiques).

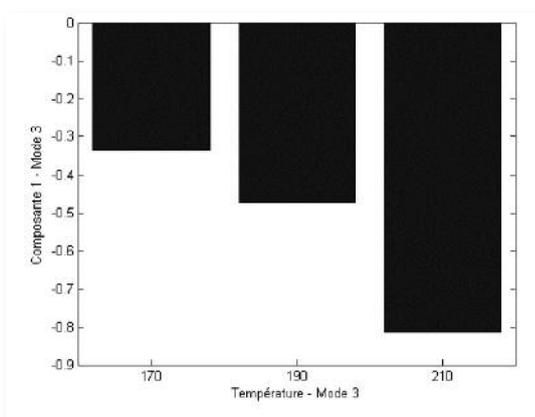


Figure 4 - Tucker3 - Scores composante 1 du mode 3 (Températures)

L'application de PARAFAC à ce même jeu de données n'apporte par un gain intéressant car il introduit une seconde composante sur le mode 3 qui n'est pas nécessaire et n'a pas de sens physique. En revanche, un traitement 4-voies avec un modèle [2 2 1 1] permet de retrouver l'ensemble des résultats précédents, que nous présentons partiellement ici, tout en récupérant près de 50% de la variance totale (figure 5). L'approche à 4 voies permet notamment de visualiser aisément l'allure de la cinétique globale des échantillons en fonction de la durée de chauffage (figure 6).

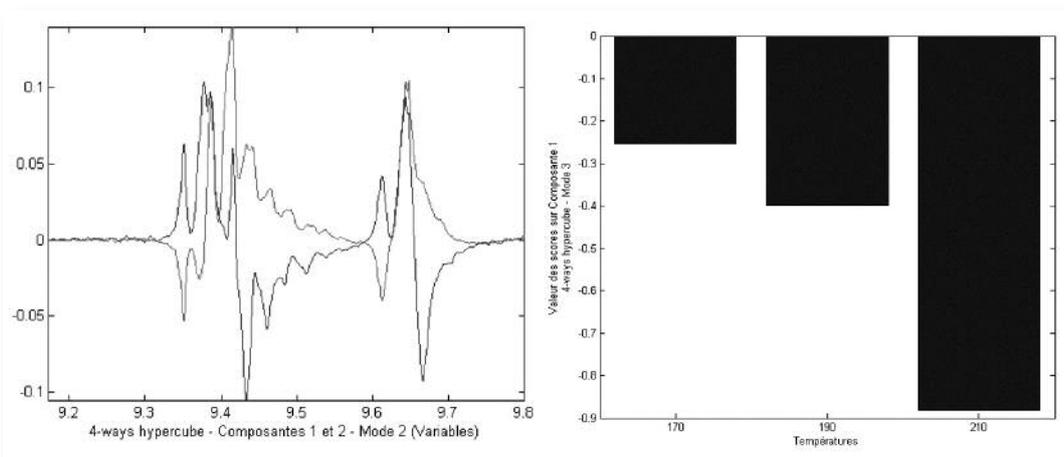


Figure 5 - Approche 4-voies Tucker3 - **A gauche**, loadings des composantes 1 et 2 (superposées) du mode 2 (variables). **A droite**, valeur des loadings de la composante 1 du mode 3 (Températures)

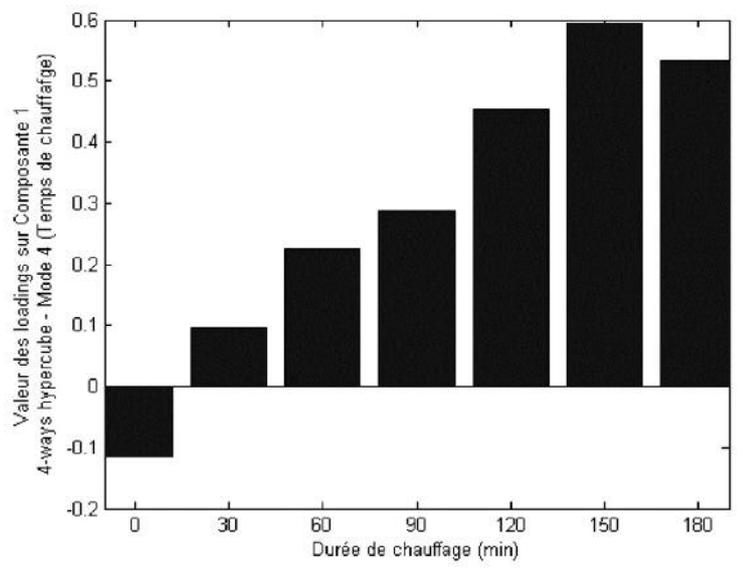


Figure 6 - valeurs des loadings sur la composante 1 du mode 4 correspondant au "Temps de chauffage" – Equivalent d'une composante cinétique

CONCLUSION

Une démarche d'analyse fondée sur des mesures de RMN-1H a permis de calculer un nouveau critère quantitatif de stabilité thermique des huiles. S'appuyant sur l'hypothèse d'une cinétique d'ordre 1, un modèle simple permet de suivre avec une bonne approximation l'apparition des aldéhydes dans les huiles chauffées. Son calcul permet de proposer un ordre de stabilité pour les trois types d'huiles en bon accord avec la bibliographique et montrent que ce critère reflète bien, dans les conditions de cette étude, la chimie des réactions d'oxydation sous-jacente. Parallèlement, une approche chimiométrique multivoies complémentaire a été mise en œuvre en utilisant l'analyse en composantes principales et l'analyse de Tucker3. Les résultats obtenus par les outils multivoies sont cohérents avec la méthode cinétique. Ces approches multivoies supposent au départ que l'effet de la température est le même pour tous les temps et les trois températures de chauffage (ce qui n'est pas exact) mais amène un gain en simplicité d'interprétation.

Bibliographie

- [1] D.B. Min, C.C. Akoh, C.R.C.P. Inc, *Food Lipids: Chemistry, Nutrition, and Biotechnology*, 2008.
- [2] E. Choe, D.B. Min, *Chemistry of deep-fat frying oils*, *J.Food Sci.*, 72 (2007) R77-R86.
- [3] E. Choe, *Mechanisms and Factors for Edible Oil Oxidation. Comprehensive Reviews, Food Science and Food Safety*, 5 (2006) 169-186.
- [4] M.D. Guillèn, N. Cabo, *Fourier transform infrared spectra data versus peroxide and anisidine values to determine oxidative stability of edible oils*, *Food Chemistry*, 77 (2002) 503-510.
- [5] M.D. Guillèn, A. Ruiz, *Study by means of 1H nuclear magnetic resonance of the oxidation process undergone by edible oils of different natures submitted to microwave action*, *Food Chemistry*, 96 (2006) 665-674.
- [6] V. Joaquin, D. Carmen, *Oxidative stability of virgin olive oil*, *European Journal of Lipid Science and Technology*, 104 (2002) 661-676.

Asymmetric supersaturated design application to a robustness study on an organic synthesis

Alice Croguennoc¹, Marc Frédéric², Magalie Claeys-Bruno³ and Roger Phan-Tan-Luu⁴

¹ Université de Bretagne Occidentale - 3 rue des Archives - CS 93837 - F29238 Brest cedex 3, a.croguennoc@laposte.net (student)

² Sanofi-aventis R&D, Site de Vitry-Alfortville, 13 quai Jules Guesde, 94400 Vitry-sur-Seine, France, marc.frederic@sanofi-aventis.com

³ Aix-Marseille Université, Institut des Sciences Moléculaires de Marseille, ISM2-UMR-CNRS-6263, Equipe AD2EM, Campus St Jérôme, Service D 52, 13397 Marseille Cedex 20, France, m.claeys-bruno@univ-cezanne.fr

⁴ LPRAI, 40, boulevard Icard, 13010, Marseille, France, roger.phan-tan-luu@univ-cezanne.fr

Keywords: Mixed-level supersaturated designs; process robustness; complementary D-optimal matrix

1- Introduction

Screening strategies are effective methods to identify active factors, i.e. influent parameters, in a process, and allow the evaluation of the robustness of a process^[1]. In some cases, however, the number of factors is important and classical screening designs require too many experiments. In such cases, if the probability for a factor to be influent is very low (< 10%: "sparsity effect"), a supersaturated design may be considered. In this design the number of effects to estimate is higher than the number of experiments.

In this project, supervised by sanofi-aventis, a mixed-level supersaturated (asymmetric) design was carried out in order to estimate the robustness of an organic synthesis. A 12-run supersaturated design was constructed, allowing us to screen 17 two-level factors and 5 three-level factors. As this mixed-level supersaturated design is not conventional and has seldom been studied, a 40-run classic matrix was built in order to compare the results.

2- Material and methods

2.1 Reference protocol

2,2'-Dithiopyridine (1.5 eq) was dissolved in a solvent (a) (5 V/Dithiopyridine) and stirred for 10 minutes. The solution was cooled down to $0 \pm 3^\circ\text{C}$, and a catalyst was added (0.02 eq). A solution of acid-thiol (1 eq) was then added at a temperature below 3°C . The reaction mixture was stirred for 5 hours at $0 \pm 3^\circ\text{C}$ and then concentrated under reduced pressure to half its initial volume V_0 . The concentrated mixture was stirred for 1 hour at $0 \pm 3^\circ\text{C}$ and the precipitated by-product (Thiopyridone) was isolated by filtration. Dicyclohexylamine (1.2 eq) was then added to the mother liquor at $15 \pm 3^\circ\text{C}$, and the mixture cooled down to $0 \pm 5^\circ\text{C}$. The mixture was then stirred for 5 hours at $0 \pm 3^\circ\text{C}$. The solid product (Dicyclohexylamine salt) was isolated by filtration, washed with solvent (a) (2 x 12.5V) and dried under reduced pressure at 40°C . The purity of the dried product was determined by HPLC.

2.2 Material

All experiments of the supersaturated design and some of the D-optimal design were carried out using a Polyblock H.E.L.[®], piloted by Winlso software, and reactors volumes where either of 150 mL or 100 mL capacities. Some of the experiments of the 40-run were done on an Auto-MATE H.E.L.[®], also piloted by Winlso, using 100 mL and 50 mL reactors.

2.3 Material

2-3-1. Factors and domain of interest

In this study, 22 factors were studied, and one response (the purity of the isolated salt) was analysed. The factors were then chosen based on the reference protocol: each of its step was modified, and the levels were established, if possible, around the nominal value in the reference protocol. Three different suppliers were used for each of the reagents and three different catalysts were tested. Factors and levels are shown in Table 1, in which eq is the molar equivalent in acid-thiol, V is volume in mL per g of Dithiopyridine and V_0 is the initial volume.

	Factors	Number of Levels	Level 1	Level 2	Level 3
U1	Dithiopyridine quantity	2	1.47 eq	1.53 eq	
U2	Solvent quantity	2	4.5 V	5.5 V	
U3	Stirring time 1	2	5 min	15 min	
U4	Temperature 1	2	-3°C	+3°C	
U5	Acid-thiol quantity	2	0.98 eq	1.02 eq	
U6	Acid-thiol adding time	2	5.5 min	16.5 min	
U7	Catalyst quantity	2	0.01 eq	0.03 eq	
U8	Stirring time 2	2	4.5 h	5.5 h	
U9	Concentration : end volume	2	0.4 V ₀	0.6 V ₀	
U10	Rinsing 1	2	0.14 V	0.16 V	
U11	Washing 1	2	0.48 V	0.53 V	
U12	Rinsing 2	2	0.14 V	0.16 V	
U13	Temperature 2	2	12 °C	18°C	
U14	Dicyclohexylamine quantity	2	1.17 eq	1.22 eq	
U15	Dicyclohexylamine dosing time	2	7.5 min	22.5 min	
U16	Temperature 3	2	-3°C	+3°C	
U17	Washing 2	2	1.43 V	1.58 V	
U18	Dithiopyridine supplier	3	Alfa Aesar® 98%	Aldrich® 98%	Acros® 98%
U19	Solvent supplier	3	SDS® 99.8%	Acros® 99.5%	Prolabo® 99.9%
U20	Catalyst type	3	Catalyst 1	Catalyst 2	Catalyst 3
U21	Dicyclohexylamine supplier	3	Acros® 99%	Alfa Aesar® 98%	Aldrich® 99%
U22	Stirring time 3	3	15 h	18 h	23 h

Table 1 - Factors

2-3-2. Construction of the designs

The mixed-level supersaturated design (N=12) was constructed according to Yamada [2] using two and three-level supersaturated designs. The complementary 40-run matrix is a traditional D-optimal matrix which includes the supersaturated design experiments [3].

3- Results and discussion

Data analysis of the supersaturated design was carried out in 2 steps: first a stepwise regression was performed, and then all subset regressions were used for a number of factors from two to six. Four criteria were studied in order to determine the number of necessary variables in the model: R², AIC (Akaike Information Criterion [4]), BIC (Bayesian Information Criterion [5]) and s² (residual variance). R² should be maximised, and the other criteria minimised. Figure 1 gives the progress of these criteria based on the number of variables regressed. Figure 2 (mapping of the variables [6]) gives the selected variables (in black) for regression with 3 variables. These figures indicate that 2 variables seem to be a reasonable choice to explain the variation of the response. Thus, 2 factors are detected as influent, U18 (Dithiopyridine supplier) and U1 (Dithiopyridine quantity).

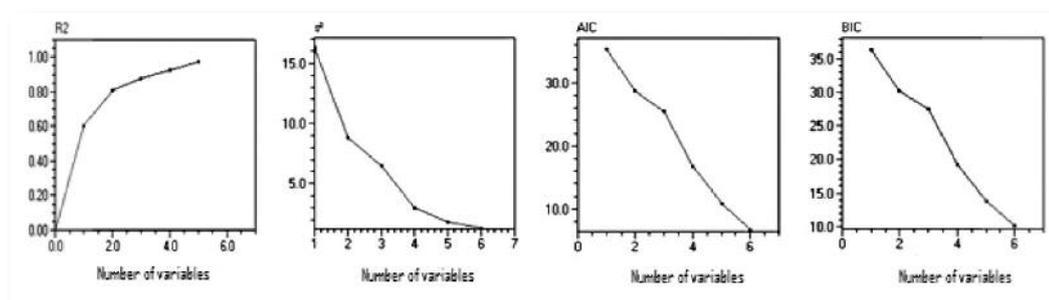


Figure 1 - Evolution of the criteria in function of the number of variables

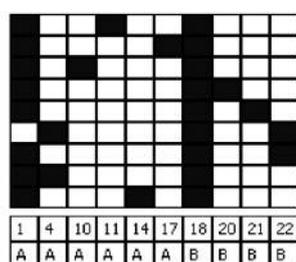


Figure 2 - Mapping with 3 variables

Data analysis of the D-optimal is conventional: residues are low, distribution is normal, and factor effect plot (Figure 3) shows that 4 factors are actually active, especially U18, and 3 factors to a lesser extent: U4, U7 and U21.

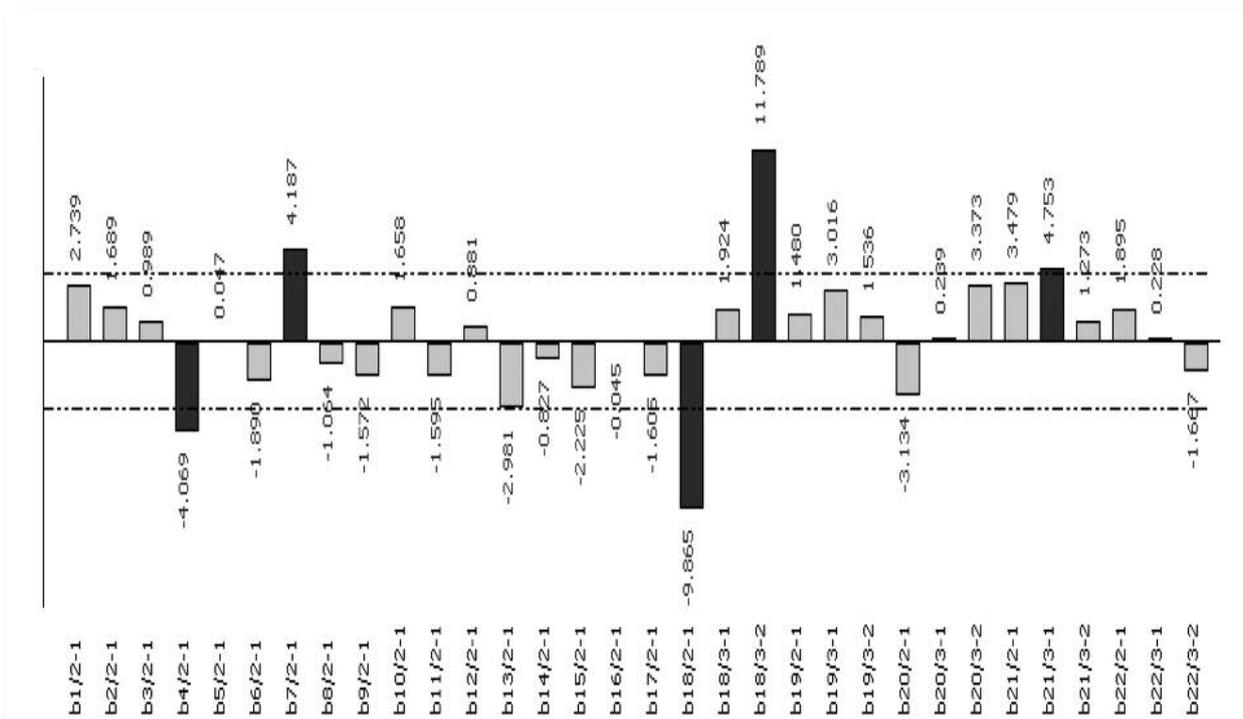


Figure 3 - Factor effect plot of the D Optimal design

CONCLUSION

The results of this study are quite encouraging. Indeed, the same very active parameter (U18) was found with both matrices. In the analysis of the 40-run matrix, a few less influent factors were found, but this difference can be explained by the loss of information due to the supersaturated design, entailing the reduction of experiments. As supersaturated designs are used as preliminary studies, further analyses of the process with conventional matrices should be considered.

Références

- 1 - Rais F, Kamoun A., Chaabouni M., Claeys-Bruno M., Phan-Tan-Luu R., Sergent M., *Supersaturated design for screening factors influencing the preparation of sulphated amides of olive pomace oil fatty acids*, *Chemometrics and Intelligent Laboratory Systems* 99 (2009) 71-78
- 2 - Yamada S. et al, *A general construction method for mixed-level supersaturated design*, *Computational Statistics & Data Analysis* 50 (2006) 254 – 265
- 3 - Phan-Tan-Luu R., Sergent M., *Non classical experimental design*, *Comprehensive Chemometrics* (Ed Stephen Bown, Roma Tauler and Beata Walczak, Elsevier) 2009 Chapter 1.14, Pages 453-499
- 4 - Akaike, H., *A new look at the statistical identification model*, *IEEE Transactions on Automatic Control* 19 (1974) 716 – 723
- 5 - Schwarz G., *Estimating the dimension of a model*, *Annals of Statistics* 6 (1978) 461-464
- 6 - D.Mathieu, J. Nony et R. Phan-Tan-Luu, *Logiciel NEMRODW*, 2009 LPRAI, Marseille

Spatial and spectral normalisation of hyperspectral images

M.-F. Devaux¹, F. Allouche¹, P. Robert¹, F. Jamme², F. Guillon¹

¹ UR 1268. Biopolymères Interactions Assemblages. INRA. F-44300 Nantes. marie-francoise.devaux@nantes.inra.fr

² Synchrotron SOLEIL. BP 48. F-91192 Gif-sur-Yvette Cedex. frederic.jamme@synchrotron-soleil.fr

Keywords: preprocessing, multivariate image analysis, fluorescence spectra.

1- Introduction

Imaging in various hyperspectral ranges – infrared, fluorescence or Raman – is a powerful tool to reveal at few micron scale chemical heterogeneity within sections of biological material. Spectra are acquired by scanning adjacent points to form three-way arrays with one spectral and two spatial dimensions. Data are analysed as sets of spectra assembled in two-way tables ignoring the spatial dimensions^[1]. The first step of the analysis is a data preprocessing that aims at removing noise and baseline deformations observed along the spectral signal. In the context of spectral image acquisition, some spectra can be acquired in regions where no matter is present. The resulting spectra are therefore flat. When spectra are acquired on the edge of the material, the overall intensity is intermediate. From the image point of view, such intensity variations depend on the morphology of the sample. As in classical microscopic techniques, some shading caused by variations in sample thickness or non-uniform illumination can be observed. Because of such artefacts, spectral intensities are normalised and only the relative amount of chemical compounds is studied. Spectral intensity variations within a small image neighbourhood could however be considered as relevant. On contrary, it is not relevant to normalise spectra corresponding to weak or flat signals. In the present work, it is proposed to combine spectral and spatial preprocessing in order to enhance both the spectral and spatial structures of the data. Methods have been developed for a UV fluorescence data set acquired at synchrotron SOLEIL from a maize stem section.

2- Material and methods

2.1 Fluorescence hyperspectral imaging of maize stem section

A 10- μm thick section was obtained from maize stem samples harvested at the female flowering stage. Three adjacent images were acquired within a vascular bundle near the epidermis to observe sclerenchyma and phloem cells. Fluorescence hyperspectral images were acquired using the synchrotron SOLEIL source on an Olympus IX71 microspectrophotometer equipped with a 40x Ultrafluar Zeiss lens^[2]. A 280 nm excitation wavelength was used to record emission spectra between 293.7 and 564.9 nm with a spectral resolution of 0.265 nm. Fields of view were 36 x 67 μm with a spatial resolution of 1 μm .

2.2 Spectral preprocessing: spike and noise filtering

Spikes are intense and narrow peaks caused by the occurrence of cosmic rays in the synchrotron source. They have been removed by applying a 1 dimensional top-hat transform^[3]. The principle consists in identifying peaks narrower than a window of a given size (21 points=5.57 nm) and higher than a given threshold (set to 40). Noise has been removed by Fourier transform filtering according to the procedure described in^[4]. A Gaussian filter of size 21 was applied to the Fourier transform.

2.3 Spatial preprocessing: background and shading correction

The background and shading corrected images were obtained as follows^[5]:

$$\mathbf{Ic}_{\lambda}(x,y)=(\mathbf{I}_{\lambda}(x,y)-\mathbf{B}(x,y))/\mathbf{S}(x,y) \quad (1)$$

where x and y represent the spatial coordinates, \mathbf{I}_{λ} is the raw intensity image at wavelength λ , \mathbf{B} is the background image, \mathbf{S} is the shading image, \mathbf{Ic}_{λ} is the corrected intensity at wavelength λ . For a given spectra, the background and shading normalisation factor were constant for all wavelength.

The background image was built by considering the minimum value of each spectrum. The shading image was estimated from the image of the sum of intensities assessed for each spectrum. Maize stems are made of cells that look like holes surrounded by walls. Inside the holes, no signal is observed and the sum of intensity is small. In order to avoid normalising the spectra recorded in holes, the usual shading normalisation factor was modified taking spectra within

a spatial neighbourhood into account: the image procedure consisting in filling holes was applied as described in [3] and the resulting image was smoothed by a 2D moving average of size 11 x 11 pixels.

2.4 Principal component analysis

Principal component analysis was applied on the normalised images as described in [1]. All softwares were developed and adapted using Matlab 2009a [6].

3- Results and discussion

Figure 1 shows a set of raw fluorescence spectra after acquisition. Spikes and noise were removed and the region between 365 and 540 nm was retained for the analysis. Fluorescence intensity varied strongly between spectra depending on the region of data acquisition.

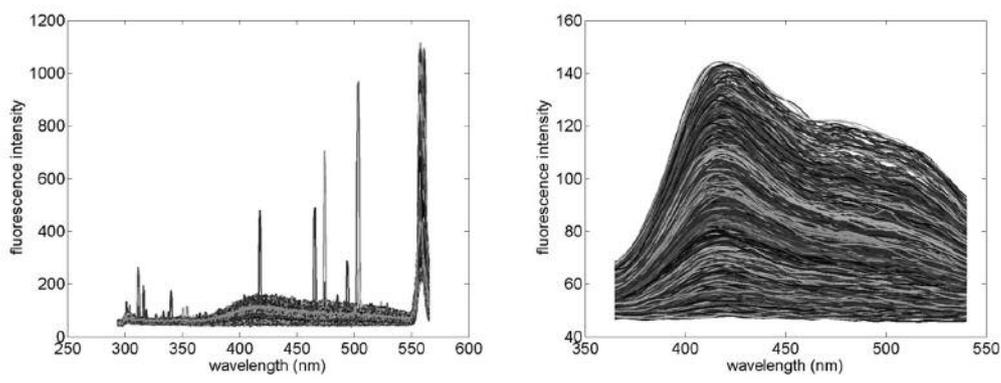


Figure 1 - spectral correction. Left: raw spectra, right: after removing spike and noise.

Figure 2 shows the spatial structure of the cells and some steps of shading correction. Figure 2b corresponds to the sum of intensities assessed for each spectrum. The image reveals the cellular structure of the maize stem section and the slow intensity variations on the walls from the top to the bottom of the image. The shading image was obtained after filling the holes and smoothing the result (Figure 2c). Smoothing removed details smaller than $11 \times 11 \mu\text{m}^2$. It was considered that within such a $11 \times 11 \mu\text{m}^2$ area, variations of intensities could be considered as relevant. After normalization (Figure 2d), sum of intensities on walls were enhanced and the higher intensities on cell junctions were preserved. One drawback of the method is the enhancement of noise in regions of weak intensities.



Figure 2 - a) bright field image, b) image of the sum of intensities, c) estimated shading image, d) sum of intensities after shading correction.

Principal Component Analysis was applied on three adjacent spectral images (Figure 3). Components 1 and 2 described 97.5 and 2.3 % of the variance. Component 1 described variations of the general intensities, thus revealing the spatial structure of cell walls and cell lumen. Component 2 opposed the sclerenchyma cell region in dark with a higher content in lignin to the cell walls in phloem with higher content in phenolic acids.

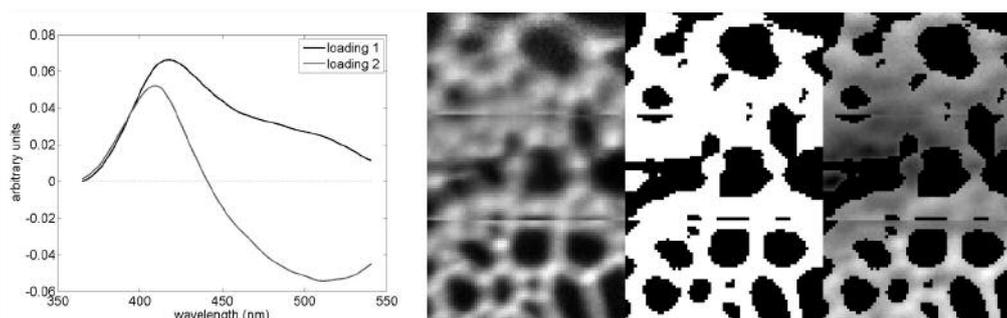


Figure 3 - a) loadings 1 and 2, b) component 1 scores, c) component 1 after image segmentation, d) component 2 scores for the pixels of cell walls selected from component 1.

CONCLUSION

A preprocessing procedure taking into account the spatial and the spectral structure of hyperspectral images has been proposed. Spectral noise and spikes were removed considering each spectrum individually. Intensity normalization was performed by taking into account the spatial information. The principle is similar to the classical normalization used in spectroscopy except that the normalization factor applied to each spectrum was assessed by considering neighbor pixels. It was assumed that within a certain neighborhood, intensity variations could be interpreted. For example, high intensity on cell junctions was identified. This has to be checked by complementary techniques such as immunolocalisation of polymers.

Références

- [1] P. Geladi & H. Grahn. *Multivariate image analysis*. John Wiley and Sons Ltd. Chichester. 1996.
- [2] <http://www.synchrotron-soleil.fr/Recherche/LignesLumiere/DISCO/InformationsUtilisateurs/Polypheme>.
- [3] P. Soille. *Morphological Image Analysis*. Springer-Verlag, New York, 1999.
- [4] R. C. Gonzalez, R. E. Woods & S. L. Eddins. *Digital Image Processing using Matlab*. Gatesmark Publishing. 2009
- [5] D. Tomaževič, B. Likar, & F. Pernuš. *Comparative evaluation of retrospective shading correction methods*. *J. of Microsc.* 2002.
- [6] © 1994-2010 The MathWorks, Inc. www.mathworks.fr

Etude comparative des plans composites sphérique et cubique dans l'optimisation de la minéralisation dans l'étude de la contamination des poissons par les métaux lourds

W. Skaff, N. Estephan*, N. Ouaini, V. Camel

Département de Chimie et Sciences de la Vie, Faculté des Sciences et de Génie Informatique, Université Saint-Esprit de Kaslik (USEK),
B.P. 446 Jounieh, Liban, Email : nathalieestephan@usek.edu.lb

Keywords: poissons, contamination, métaux lourds, optimisation, plan composite.

1- Introduction

Les zones côtières sont l'objet d'une urbanisation intense, à laquelle il faut ajouter une activité économique et touristique importantes. Les poissons vivants dans ces zones sont menacés par la pollution à ces endroits.

Dans la présente étude certains contaminants inorganiques (Cd, Cr, Cu, Pb, Zn) sont analysés dans deux espèces de poissons (*Siganus rivulatus* et *Diplodus sargus*) répandus au long de la côte libanaise tout en essayant de relier la nature de ces contaminants à la source de pollution.

La recherche des métaux lourds dans ces poissons nécessite une minéralisation des échantillons avant analyse par absorption atomique. La qualité de cette analyse dépend en grande partie de la préparation de l'échantillon. Une optimisation de la minéralisation se révèle donc nécessaire. A cette fin, un plan composite fut adopté. Deux conceptions de plan composite ont été testées, l'une par rapport à l'autre : plan composite cubique et plan composite sphérique. Les paramètres à optimiser sont la température de chauffage, temps de chauffage et volume d'acide ajouté.

Une légère différence a été décelée par rapport à l'optimum fourni par les deux plans. Une étude comparative a été réalisée dans le but de choisir la meilleure combinaison de paramètres pour l'extraction.

2- Matériel et méthodes

2.1 Dates et fréquences de prélèvement

Dans chaque site, selon la disponibilité des espèces étudiées, un prélèvement saisonnier de deux kilos de chacune des deux espèces choisies est effectué. Le prélèvement des poissons est effectué sur trois sites choisis autour de la région de Beyrouth qui sont : Tabarja, Dora et Ouzai. Pour s'assurer de la fiabilité de la matière première, l'on attendait les pêcheurs tôt le matin au port de pêche avant même leur arrivée. Afin de réduire les risques de contamination, une fois achetés, les poissons destinés pour l'analyse des métaux lourds sont emballés par espèce dans des sacs en polyéthylène et ceux destinés pour l'analyse des HAP dans des films en aluminium. Après emballage les poissons sont conservés au froid dans un conteneur frigorifique contenant des glaçons puis emmenés au laboratoire.

2.2 Traitement des échantillons

Une fois au labo, les poissons sont lavés individuellement avec de l'eau distillée ; la taille et le poids de chaque individu sont notés. Ensuite les échantillons sont placés dans une chambre frigorifique dans l'attente du broyage. Chaque poisson prélevé est découpé sur une plaque en polyéthylène, avec un couteau en vanadium. Les morceaux découpés sont introduits dans un hachoir de viande de marque Moulinex à lame en inox. Le broyat obtenu est placé dans un pilulier en verre préalablement lavé, taré (pesé vide et sec) et congelé en attendant la lyophilisation. On relève, pour chaque échantillon, le poids des piluliers chargés de tissus humides, de manière à pouvoir déterminer le pourcentage d'humidité après séchage.

2.3 Optimisation de la minéralisation

Afin d'optimiser les conditions expérimentales de la minéralisation, nous avons opté à la construction de deux plans composites centrés de modèle polynomial du deuxième degré : le domaine expérimental du premier est cubique et celui du second est sphérique. Les facteurs fixés dans les deux plans furent : le volume du mélange d'acide-oxydant, le temps et la température de chauffage – paliers de température.

Le logiciel utilisé pour l'analyse et le traitement des données ainsi que la détermination de la méthode de minéralisation optimale est le NemrodW (version 2007 – © LPRAI S.A.R.L.).

3- 3 Résultats et discussion

3.1 Plan composite cubique

L'étude graphique des effets et la surface d'isoréponses a permis de déduire des conditions optimales pour la minéralisation à une température fixe de 210°C (vu que le facteur température n'a pas révélé d'effet significatif sur le rendement d'extraction) : volume de 12 ml pour le mélange d'acides avec une durée de chauffage de 25 minutes. Le volume d'acide nitrique concentré (HNO_3 65%) utilisé dans la minéralisation est de l'ordre 10 ml et celui du peroxyde d'hydrogène (H_2O_2 30%) est de l'ordre de 2 ml.

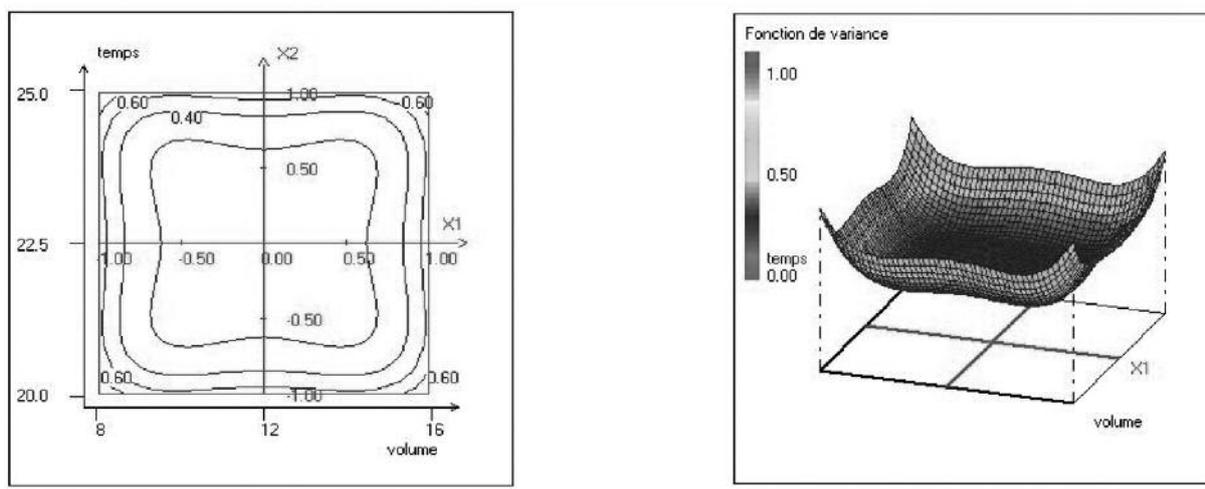


Figure 1 - Surfaces d'isovariance pour le plan composite cubique.

3.2 Plan composite sphérique

L'étude graphique des effets et la surface d'isoréponses a permis de déduire des conditions optimales pour la minéralisation à une température fixe de 220°C (vu que le facteur température n'a pas révélé d'effet significatif sur le rendement d'extraction) : volume de 15 ml pour le mélange d'acides avec une durée de chauffage de 23 minutes. Les volumes d'acide nitrique concentré (HNO_3 65%) et du peroxyde d'hydrogène (H_2O_2 30%) utilisés dans la minéralisation sont respectivement de l'ordre 13 ml et de 2 ml.



Figure 2 - Surfaces d'isovariance pour le plan composite sphérique.

3.3 La comparaison des deux optimaux tirés des deux plans composites

Afin de pouvoir déterminer la méthode optimale pour la minéralisation des poissons broyés et lyophilisés, l'on a procédé à un dosage du Zinc dans un échantillon de poissons. Le même protocole d'analyse adopté pour le dosage des métaux fut adopté dans cette étude comparative réalisé à deux reprises.

La comparaison intra site, intra espèce et intra saison

En premier lieu l'on a procédé à une minéralisation des *Diplodus Sargus* du site Ouzai prélevés en hiver 2010 selon les deux méthodes optimales résultantes respectivement des deux plans d'expérimentation cubique et sphérique ainsi que la méthode de référence proposée par la compagnie qui a développée le micro-onde. L'analyse en SAA-mode flamme des échantillons préparés a révélé un rendement maximal pour la méthode de référence proposé par la société « ETHOS PRO » productrice du micro-onde.

La comparaison inter sites, inter espèces et inter saisons

Afin d'affiner la comparaison des trois méthodes de minéralisation déjà cités, une deuxième série d'analyse a été effectuée sur des échantillons provenant de sites différents (Tabarja et Ouzai) et collectés à des saisons différentes. L'analyse en SAA-mode flamme des échantillons préparés est venue confirmer le résultat obtenu lors de la première comparaison et a révélé un rendement maximal pour la méthode de référence proposé par la société « ETHOS PRO » productrice du micro-onde.

CONCLUSION

Dans cette étude deux plans composites ont été étudiés: plan composite cubique et plan composite sphérique, pour optimiser le processus de minéralisation des échantillons de poissons lyophilisés destinés pour être analysés par absorption atomique.

Les paramètres à optimiser étant la température de chauffage, le temps de chauffage et le volume d'acide, la minéralisation était entreprise en adoptant les optimaux obtenus des deux plans composites : (210°C, 12 ml, 25 min) pour le domaine cubique et (220°C, 15 ml, 23 min) pour le domaine sphérique.

La comparaison des résultats SAA obtenus après minéralisation par les deux méthodes optimales suggérées à partir des deux plans composites a révélé des similarités avec une légère amélioration du rendement d'extraction avec le plan sphérique et ceci à cause de la structure de ce dernier.

Application des techniques multivariées dans la caractérisation des miels Libanais par Infrarouge et par Absorption Atomique

N. Ghadban, N. Estephan*, N. Ouaini, D. N. Rutledge

Département de Chimie et Sciences de la Vie, Faculté des Sciences et de Génie Informatique, Université Saint-Esprit de Kaslik (USEK), B.P. 446 Jounieh, Liban, Email : nathalieestephan@usek.edu.lb

Keywords: miel, authenticité, métaux lourds, absorption atomique

1- Introduction

Dans le but de caractériser les miels Libanais en fonction de leur origine géographique et d'étudier leur authenticité, des échantillons de miel ont été collectés à travers plusieurs régions Libanaises suivant la répartition des zones agricoles homogènes donnée par le ministère de l'agriculture Libanais et ceci durant deux récoltes consécutives en été 2007 et en été 2008.

L'analyse des échantillons est faite par Spectroscopie Moyen Infrarouge (MIR) et par Absorption Atomique. Les données spectrales collectées dans une matrice numérique sont ensuite analysées par des méthodes chimiométriques multivariées après avoir subi une série de prétraitements d'amélioration du signal et sélection des variables.

Après prétraitements, l'Analyse en Composantes Principales (ACP) et l'Analyse Factorielle Discriminante (AFD) ont été appliquées dans le but d'inspecter les différences entre les groupes de miels en fonction de chaque critère étudié et de pouvoir expliquer ces divergences.

2- Matériel et méthodes

2.1 Echantillonnage

D'après les informations données par le ministère de l'agriculture Libanais, le pays est divisé en 7 zones agricoles homogènes comme présenté dans la **figure 1**. Une enquête a été menée auprès des apiculteurs tout au long du territoire libanais, permettant de localiser les ruches, de connaître leur nombre et de savoir la saison de récolte pour chaque apiculteur afin de bien aménager l'échantillonnage. Pour les saisons d'été 2007 et 2008, 56 échantillons de miel ont été choisis en fonction des résultats de l'enquête faite auprès des apiculteurs. Pour chaque échantillon choisi, 1 à 3 prélèvements ont été faits, ce qui rend un total de 127 échantillons à analyser.

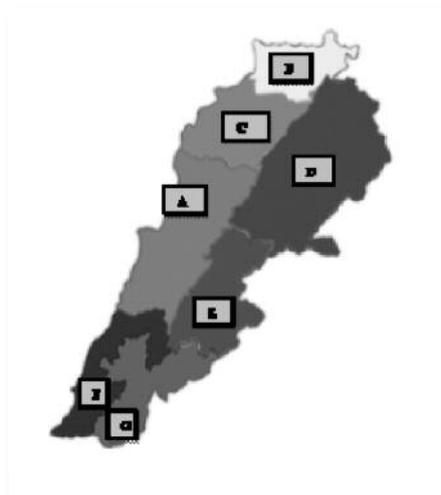


Figure 1 - Les sept zones agricoles homogènes du Liban selon le ministère de l'agriculture Libanais.

2.2 Analyse spectrale

Les métaux lourds qui ont été analysés sont le zinc, le manganèse, le cuivre, le nickel, le cadmium, le chrome, le fer et le magnésium. Les mesures ont été réalisées en utilisant un spectromètre d'absorption atomique de type ZEE nit 700.

L'analyse IR des échantillons de miel a été effectuée en utilisant un appareil FT-IR de type Nicolet IR200 et les paramètres d'acquisition des spectres dans le moyen infrarouge sont résumés dans le **tableau 1**.

Gamme Spectrale	Gamme de travail 500 à 4000 cm^{-1}
Détecteur	DTGS
Résolution	4 cm^{-1}
Balayage	64 scans
Sensibilité	Rapport Signal/ Bruit $< 2,2 \times 10^{-5}$
ATR	Thunderdome Foundation Cristal : Germanium mono-reflexion Lentille : ZnSe Enclume inclinable dynamométrique
Logiciel	EZ Omnic 7,1

Tableau 1 - Paramètres d'acquisition des spectres MIR.

3- Résultats et discussion

3.1 Le traitement des résultats du moyen IR

Les données numériques des spectres MIR ont été collectées dans une matrice de taille (378, 1816), avec 378 échantillons et 1816 variables correspondant aux nombres d'ondes. Les individus de la matrice sont classés suivant la couleur et l'origine florale. La matrice ainsi conçue a subi un lissage, avec 9 points de fenêtre de convolution, suivi d'un prétraitement par SNV sur la totalité des spectres afin de réduire l'effet de la dispersion de l'échantillonnage au moment de la mesure. Ensuite, les données ont été traitées par ANOVA-pp, ACP et AFD.

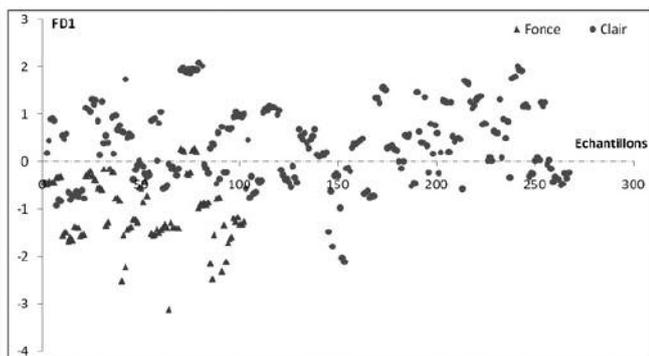


Figure 1 - Projection des individus sur l'axe FD1 après AFD en fonction de la couleur

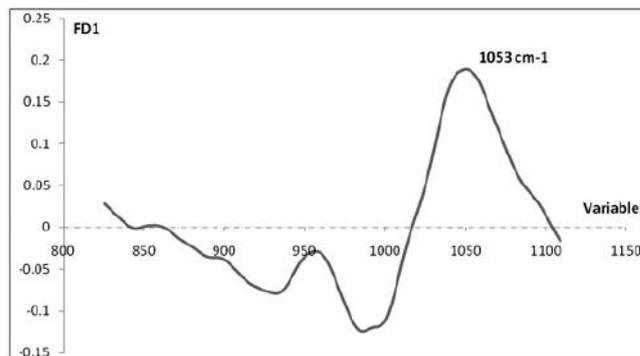


Figure 2 - Le profil des facteurs discriminants issus de l'AFD suivant la couleur.

La distribution des individus sur l'axe FD1 (**Figure 1**), montre une séparation en deux groupes correspondant aux deux couleurs du miel : clair et foncé. En observant le profil des contributions factorielles des variables de départ sur FD1, nous remarquons que la variable responsable de cette séparation correspond au pic MIR situé à 1053 cm^{-1} attribué au saccharose. Puisque cette variable contribue positivement sur FD1, ceci nous permet de déduire que les miels de couleur clair, situés du côté positif de l'axe FD1 renferment une teneur en saccharose supérieure à celle contenue dans les miels foncés, placés négativement suivant FD1.

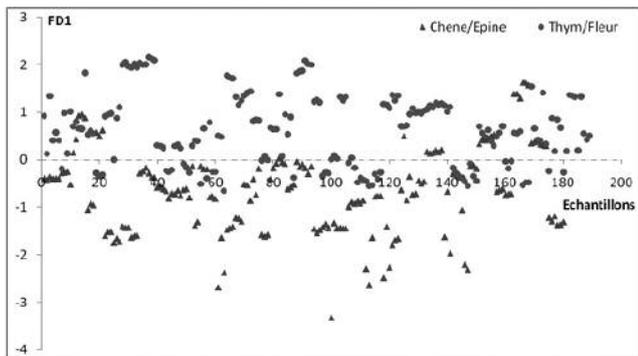


Figure 3 - Projection des individus sur l'axe FD1 après AFD en fonction de la végétation.

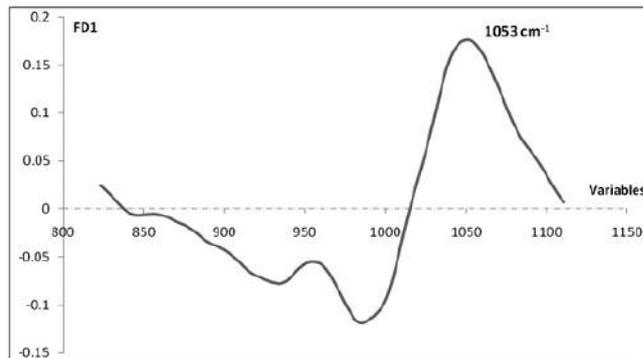


Figure 4 - Le profil des facteurs discriminants issus de l'AFD suivant la végétation.

La répartition des individus sur l'axe FD1 (**Figure 2**), montre une séparation en deux groupes suivant la végétation : « Chêne/Epine » et « Thym/Fleur ». Le profil des contributions factorielles des variables de départ sur FD1 montre que la variable qui sépare ces deux groupes d'échantillons correspond au pic MIR situé à 1053 cm^{-1} contribuant positivement sur FD1. Cette variable étant caractéristique du saccharose. Ceci mène à la déduction que les miels appartenant au groupe « Chêne/Epine » contiennent moins de saccharose que ceux qui appartiennent à la végétation « Thym/Fleur ».

3.2 Le traitement des résultats des métaux lourds par SAA

Les résultats obtenus en mg/L par absorption atomique ont été collectés dans une matrice de taille (378, 7) ; les 7 variables correspondant aux différents métaux lourds étudiés à l'exception du Nickel qui a donné des valeurs au-dessus des limites detectales pour la totalité des échantillons. L'ACP a été calculée avec sept composantes principales correspondant au rang de la matrice. Le plan PC₁-PC₂ a permis l'observation de la séparation des échantillons en deux groupes suivant PC₁ (**Figure 5**). Le profil des contributions factorielles des variables de départ sur PC₁ (**Figure 6**) montre que la discrimination entre les deux groupes de miels est basée sur la quantité de magnésium.

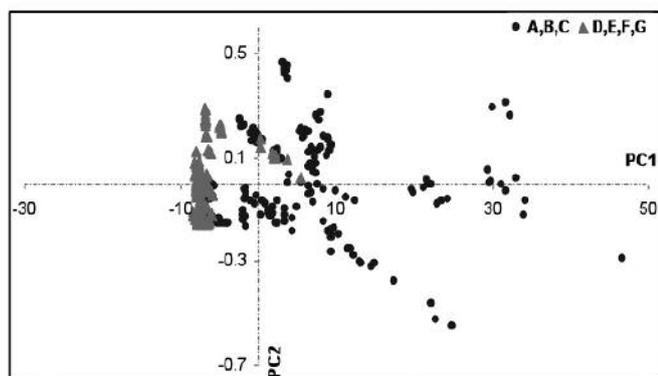


Figure 5 - Projection des échantillons dans le plan PC₁-PC₂ pour les résultats des métaux lourds.

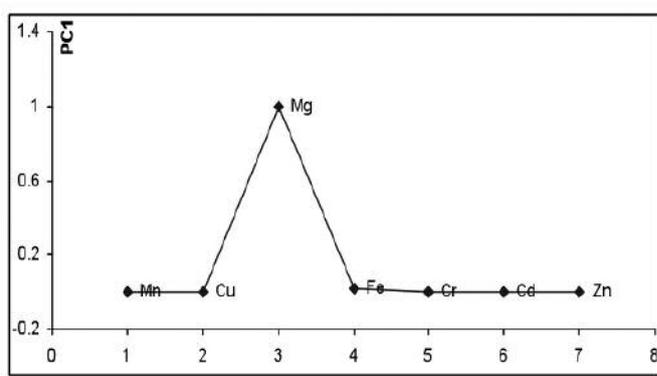


Figure 6 - Le profil des contributions factorielles des variables de départ sur PC₁.

CONCLUSION

Dans le but de fonder une appellation d'origine contrôlée, des échantillons de miel, récoltés sur tout le territoire Libanais, ont été analysés par spectroscopie moyen infrarouge et par absorption atomique. L'ACP et l'AFD ont été appliqués dans le but de visualiser les différences entre les groupes d'échantillons de miel.

L'AFD appliquée sur les spectres IR pour les critères « Couleur » et « Origine Florale » a montré que le saccharose est à la base de cette séparation ce qui a mené à la déduction que les échantillons de couleur « Clair » contiennent plus de saccharose que ceux de couleurs « Foncé » et ceux d'origine florale « Thym/Fleur » contiennent plus de saccharose que ceux provenant d'une végétation majoritaire « Chêne/Epine ». L'ACP calculée sur la matrice regroupant les concentrations des métaux lourds obtenues par SAA a montré que les échantillons de miel sont divisés en deux groupes en se basant sur leur contenu en magnésium.

Étude de la qualité de farine de blé destinée à la panification en fonction de la variété de blé par infrarouge et par tests physico-chimiques et rhéologiques couplés à la chimométrie

N. Estephan* **N. Ouaini**

Département de Chimie et Sciences de la Vie, Faculté des Sciences et de Génie Informatique, Université Saint-Esprit de Kaslik (USEK), B.P. 446 Jounieh, Liban, Email : nathalieestephan@usek.edu.lb

Keywords: farine de blé, contrôle qualité, moyenne infrarouge, ACP.

1- Introduction

Le pain produit à partir de la farine de blé est un élément essentiel dans la ration alimentaire presque partout dans le monde. La qualité nutritionnelle et organoleptique du pain varie en fonction de la qualité physico-chimique et rhéologique de la farine dont il est issu.

Le présent travail vise à comparer quatre qualités de farines obtenues à partir de quatre variétés des blés : Kazakh Sibirisky, Russe E.N, Kazakh E.N et Allemande. Ces variétés sont importées de l'étranger sur le marché Libanais pour être utilisées dans la fabrication du pain Libanais et Européen, afin de déduire laquelle est la plus convenable en termes de qualité technologique.

Les quatre variétés de blé ainsi que les farines correspondantes ont été analysées par des méthodes physico-chimiques et rhéologiques classiques utilisées pour le contrôle qualité de ces produits. La spectroscopie moyenne infrarouge (MIR) a été également utilisée pour étudier la composition des farines. Les résultats physico-chimiques et rhéologiques, ainsi que les données spectrales ont été traités par chimométrie.

L'Analyse en Composantes Principales, calculée après prétraitement des données, a montré que les deux variétés Kazakh Sibirisky et Allemande ont une qualité technologique meilleure que celle des deux variétés Russe E.N et Kazakh E.N car elles possèdent une force boulangère plus grande et des teneurs plus élevées en gluten et en protéines autres que le gluten.

2- Matériel et méthodes

2.1 Le matériel végétal

Le matériel végétal est constitué de quatre variétés de blé qui sont importées au Liban par les minoteries pour la production de farine destinée à la préparation du pain Libanais et les pains de type Européen. Ce sont les variétés Kazakh Sibirisky (V1), Russe E.N (V2), Kazakh E.N (V3) et Allemande (V4).

L'échantillonnage des quatre variétés de blé est réalisé sur les cales de blé du bateau à la réception selon GAFTA 124, ICC 101/1 ou AACC 64/70A.

2.2 L'étude physicochimique et rhéologique du blé et de la farine de blé

Les tests de qualité du blé entrepris sont les suivants : Le poids à l'hectolitre, l'indice de sédimentation de Zélény, le taux de cendre du blé, l'humidité du blé, le temps de chute du blé, le gluten du blé et les protéines du blé.

La farine de blé ayant subi une agression lors de la mouture, doit être laissée se reposer durant 15 jours dans une chambre à 24°C avant son contrôle qualité, sous peine de voir ses pâtes se relâcher au pétrissage. Les tests réalisés pour étudier la qualité de la farine sont les suivants : Le gluten de la farine de blé, l'humidité, les cendres, les protéines et le temps de chute de la farine de blé, l'alvéographe, le farinographe, l'extensographe.

2.3 L'analyse de la farine de blé par spectroscopie IR

L'appareil FT-IR utilisé pour étudier la qualité des farines issues des quatre variétés de blé est le Nicolet IR200 et les paramètres d'acquisition des spectres dans le moyen infrarouge sont résumés dans le **tableau 1**. Cinq spectres ont été acquis pour chacun des 12 sous-échantillons de farine correspondant aux quatre variétés de blé.

Gamme spectrale	• Gamme de travail entre 4000 et 500 cm^{-1}
Détecteur	• DTGS (Deuteriated Triglycine Sulfate)
Résolution	• 4 cm^{-1}
Balayage	• jusqu'à 64 scan.s^{-1}
Sensibilité	• rapport Signal/Bruit < $2,2 \times 10^{-5}$ à 4 cm^{-1} de résolution
Précision	• supérieure à 0.01 cm^{-1} à 2000 cm^{-1}
ATR	<ul style="list-style-type: none"> • Thunderdome Foundation • Cristal : Germanium mono-réflexion • Lentille ZnSe : 5000 – 650 cm^{-1} • Enclume inclinable dynamométrique
Logiciel	• EZ OMNIC 7,1

Tableau 1 - Paramètres d'acquisition des spectres de farine.

3- Résultats et discussion

3.1 Le traitement des résultats physico-chimiques du blé

Selon l'analyse en composantes principales (ACP), 3 composantes principales ont pu expliquer jusqu'à 99,99% de la variabilité des échantillons. D'après les résultats obtenus, nous avons remarqué que les deux variétés Russe E.N (V_2) et Kazakh E.N (V_3) sont très semblables. Nous les avons donc réunies en un seul groupe désigné par V_{23} . Ce qui permet de distinguer ainsi 3 groupes de blé et de farine de blé : V_1 , V_{23} et V_4 . Sur le plan PC_1 - PC_2 nous observons une séparation de deux groupes de blé selon l'axe PC_1 , où la variété V_{23} se situe dans la partie positive alors que les variétés V_1 et V_4 se trouvent dans la partie négative (**Figure 1**).

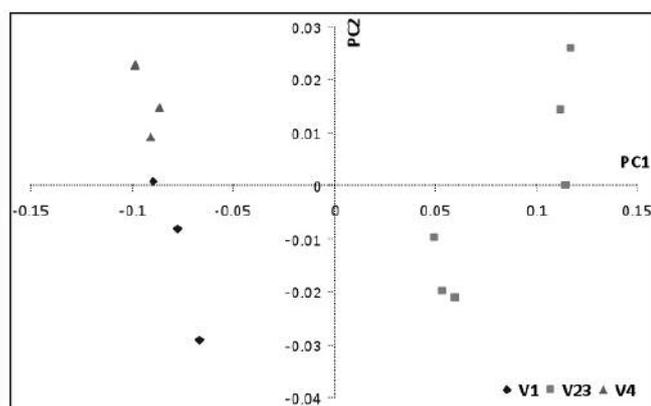


Figure 1 - Projection des échantillons dans le plan PC_1 - PC_2 pour les tests physico-chimiques du blé.

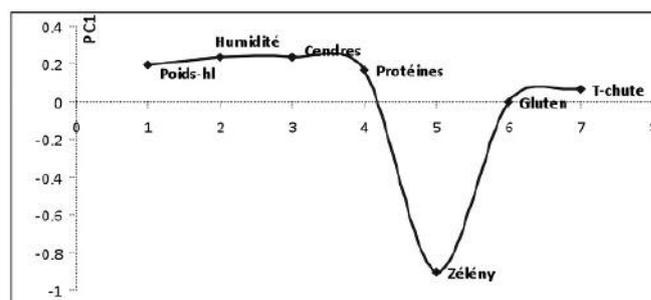


Figure 2 - Profil des contributions factorielles des variables de départ sur PC_1 , pour les tests physico-chimiques du blé.

En examinant le profil des contributions factorielles des variables de départ sur PC_1 , nous remarquons que l'indice de sédimentation de Zélény présente un pic fortement négatif permettant ainsi de séparer entre les deux groupes (**Figure 2**). Nous déduisons donc que cet indice est plus élevé chez les deux variétés V_1 et V_4 vue leur localisation dans la zone négative suivant PC_1 .

3.2 Le traitement des données spectrales MIR de la farine de blé

Les données numériques des spectres des farines de blé acquis en MIR ont été collectés dans une matrice MIR de taille (60, 1816). Après prétraitement de la matrice, l'ACP a été calculée avec six composantes principales récupérant jusqu'à 99.99% de l'information initiale. Le plan PC₁-PC₃ a permis l'observation de la séparation des échantillons en groupes de variété (Figure 6). Pour comprendre cette séparation, nous avons analysé les profils des contributions factorielles des variables de départ sur PC₁ (Figure 7) et sur PC₃ (Figure 8). Selon Un pic fortement négatif correspondant au nombre d'onde 1005 cm⁻¹, contribue négativement sur PC₁. Le pic MIR situé à ce nombre d'onde est attribué à l'amidon ce qui permet de déduire que les farines du groupe V₄ correspondant à la variété Allemande renferment plus d'amidon que les farines de la variété Kazakh Sibirisky relative au groupe V₁. Selon le profil des contributions factorielles des variables initiales sur PC₃, deux pics positifs, correspondant aux nombres d'ondes 1529 et 1631 cm⁻¹, sont observés ainsi qu'un pic relatif au nombre d'onde à 1722 cm⁻¹, contribuant négativement sur PC₃. Suivant les attributions des bandes en moyen infrarouge, le pic situé à 1529 cm⁻¹ correspond à la lysine, celui de 1631 cm⁻¹ est caractéristique des protéines du gluten alors que le pic à 1722 cm⁻¹ est relatif à l'acide glutamique.

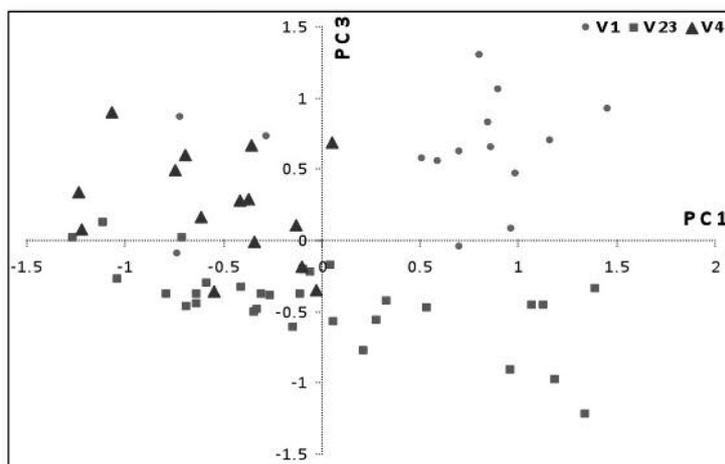


Figure 9 - Projection des échantillons sur le plan PC1-PC3 après ACP sur la matrice MIR des farines de blé.

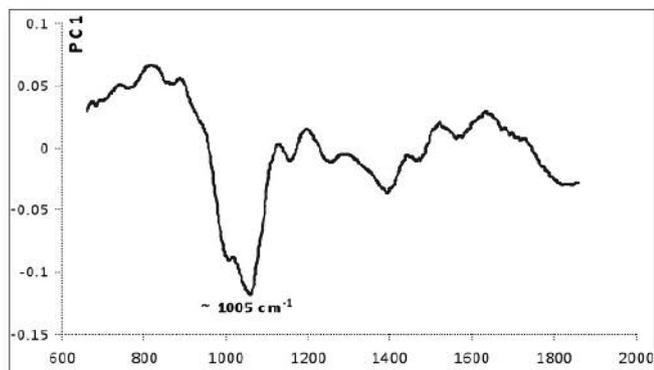


Figure 10 - Le profil des contributions factorielles des variables de départ sur PC1.

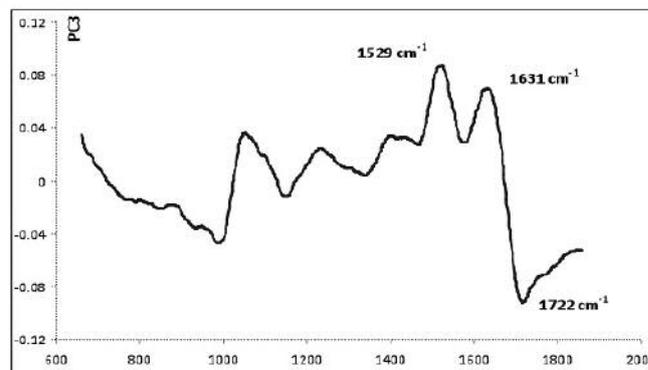


Figure 11 - Le profil des contributions factorielles des variables de départ sur PC3.

CONCLUSION

Cette étude s'est basée sur la comparaison de quatre qualités de farine issues de quatre variétés de blé étrangères utilisées dans la fabrication du pain Libanais et Européen au Liban (Kazakh Sibirisky, Russe E.N, Kazakh E.N et Allemande).

Les résultats obtenus montrent que les deux variétés Kazakh Sibirisky et Allemande présentent des taux élevés en gluten et en protéines autres que le gluten et, une force boulangère W supérieure par rapport aux deux variétés Russe E.N et Kazakh E.N. Par suite, les variétés Kazakh Sibirisky et Allemande présentent un remarquable rapprochement en termes de forte qualité technologique. En parallèle, les variétés Russe E.N et Kazakh E.N. semblent avoir une similarité par rapport à leur qualité technologique.

Une comparaison plus poussée révèle que la variété Allemande renferme un taux d'amidon plus élevé que celui de la variété Kazakh Sibirisky alors que cette dernière présente un taux protéique plus élevé (gluten et protéines autres que le gluten) et une force boulangère W supérieure.

Evaluation de la robustesse d'une source d'ionisation par approche chimiométrique reposant sur les plans d'expériences

C. Hubert¹, D. Vailhen¹, X. Machuron-Mandard¹, J. C. Tabet²

¹ Commissariat à l'Energie Atomique, Centre DAM Ile-de-France, Bruyères-le-Châtel, 91297 Arpajon Cedex, France, cecile.hubert@cea.fr

² Laboratoire de Chimie Organique Structurale et Biologique, Université Pierre et Marie Curie – Paris VI, 4 place Jussieu, 75252 Paris cedex 05, France, jean-claude.tabet@upmc.fr

Mots-clés : Electrospray, spectrométrie de masse, plan de Doehlert, Plackett-Burman, surfaces de réponse.

1- Introduction

L'électrospray est une technique majeure d'ionisation à pression atmosphérique utilisée en spectrométrie de masse. Elle n'a cessé de se démocratiser depuis son développement à la fin des années 1980. Elle est aujourd'hui utilisée dans des domaines aussi divers que la pharmaceutique, l'environnement ou la criminalistique.

Bien que largement employé, l'électrospray est encore considéré comme une « boîte noire » par de nombreux utilisateurs. Les causes d'instabilité du signal, de compétition de désorption ou de suppression ionique sont encore mal comprises [1]. De plus, l'abondance des ions obtenus dépend de nombreux paramètres d'ionisation et de transmission vers l'analyseur. Ces paramètres peuvent être aussi bien géométriques (position du spray) que physiques (tensions appliquées) ou chimiques (nature de l'analyte). Des exemples de ce type d'étude sont disponibles dans la littérature [2,3] mais ne sont pas transposables à d'autres sources compte tenu de la spécificité de leur géométrie.

Le but de cette étude est donc d'évaluer la robustesse du fonctionnement de la source disponible au laboratoire. Une méthodologie reposant sur les plans d'expériences a été appliquée. Plus complexe qu'initialement prévue de par l'existence de nombreux facteurs significatifs, cette étude a nécessité leur utilisation via une approche spécifique.

2- Matériel et méthodes

2.1 Conditions expérimentales

L'étude chimiométrique a été réalisée sur le 2,4,6-trinitrotoluène en solution dans un mélange eau / méthanol. Les analyses ont été effectuées avec un spectromètre de masse triple quadripôle (1200L, Varian) équipé d'une source électrospray en mode négatif. L'introduction des échantillons dans la source d'ionisation s'est faite au moyen d'un pousse-seringue à débit variable (Figure 1).

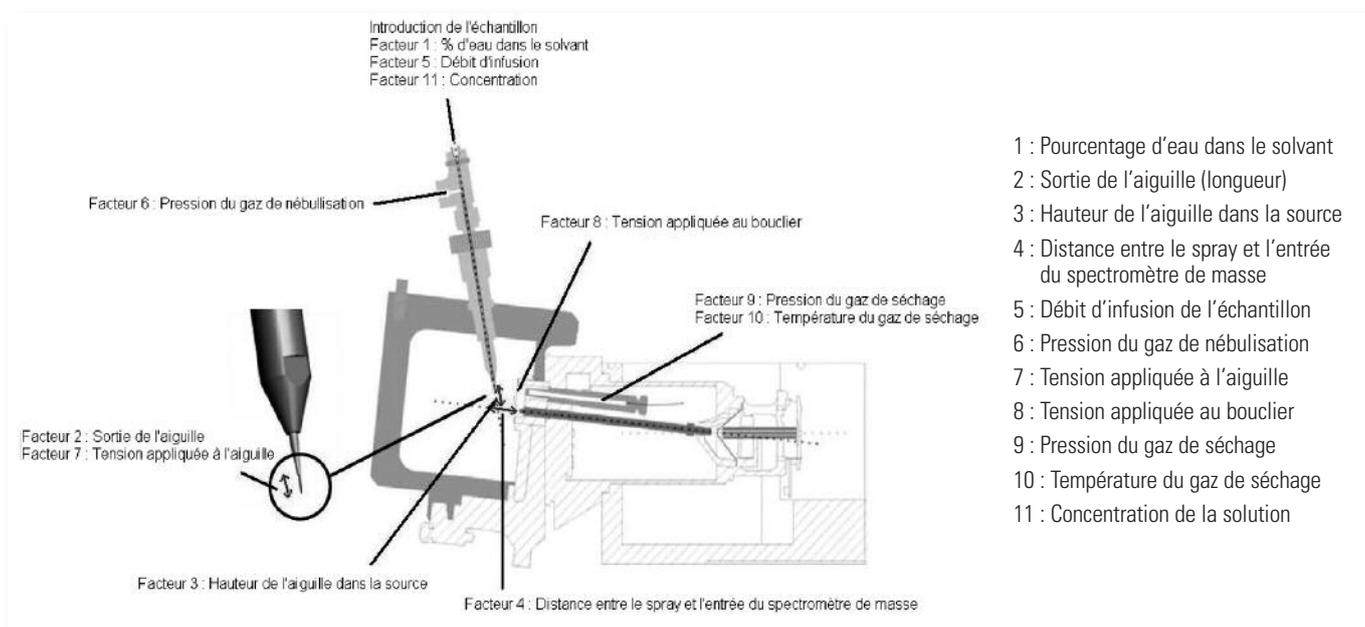


Figure 1 - Schéma de la source électrospray utilisée et définition des facteurs.

2.2 Analyse des données

Quatre réponses ont été mesurées : l'intensité et la stabilité du signal de l'ion de rapport masse sur charge (m/z) 226 ([M-H]⁻) ainsi que l'intensité et la stabilité du signal de l'ion m/z 227 (M⁻). La stabilité est définie comme étant le coefficient de variation de l'intensité du signal mesuré pendant 85 balayages en m/z (ou scans). Au total, 11 facteurs pouvant influencer sur ces réponses ont été identifiés. Tous les résultats ont été traités avec le logiciel Statistica 9.1 (StatSoft, 2010).

3- Résultats et discussion

L'étude a été menée de manière séquentielle. Dans un premier temps, un plan de criblage a été utilisé dans l'espoir de réduire le nombre de facteurs significatifs (de 11 à quelques uns). La plupart se sont avérés être significatifs. L'approche chimiométrique initialement envisagée a donc été modulée pour appliquer une stratégie de plans de Doehlert imbriqués.

3.1 Plan de criblage - Plan de Plackett-Burman

Ces plans présentent l'avantage de permettre l'étude d'un nombre élevé de facteurs dans l'objectif de discriminer ceux qui sont influents. Le modèle mathématique utilisé est un modèle sans interactions. Le principal risque sera donc d'éliminer une variable qui, seule, n'a que peu d'influence mais qui possède une forte interaction avec une autre des variables étudiées. Les résultats sont analysés à l'aide des tableaux des effets estimés et des diagrammes de Pareto. Il apparaît que 10 facteurs sur les 11 étudiés sont significatifs pour l'intensité de m/z 226 et 8 pour l'intensité de m/z 227. Aucun des facteurs n'a d'impact significatif sur la réponse « stabilité ».

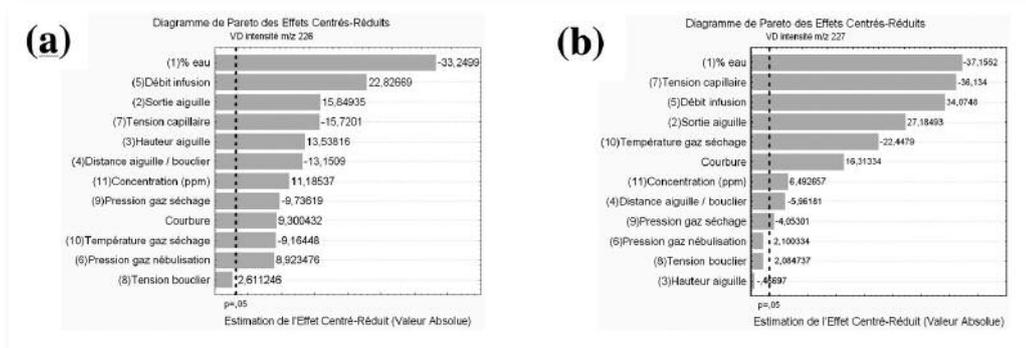


Figure 2 - Diagrammes de Pareto pour les réponses (a) intensité m/z 226, (b) intensité m/z 227.

3.2 Plan de Doehlert

Malgré le résultat inattendu d'un nombre élevé de facteurs significatifs, nous avons conservé notre objectif d'approfondissement de l'effet des facteurs par l'étude des surfaces de réponse. Les plans de Doehlert ont été choisis car 1) ils sont économes en nombre d'expériences, 2) ils permettent une approche séquentielle avec l'introduction successive des variables d'intérêt.

Un plan de Doehlert à trois facteurs a été réalisé (D_3) avec les facteurs les plus influents identifiés précédemment (pourcentage d'eau dans le solvant, débit d'infusion et tension de l'aiguille). Les facteurs non étudiés ont été maintenus à un niveau intermédiaire pendant l'étude.

La stabilité du signal est peu impactée par les modifications de la source électrospray. Quelles que soient les conditions expérimentales, le coefficient de variation calculé sur 85 scans est compris entre 1% et 5%. Concernant l'intensité des ions, les surfaces de réponse obtenues ont globalement la même allure, qu'il s'agisse de l'ion m/z 226 ou m/z 227.

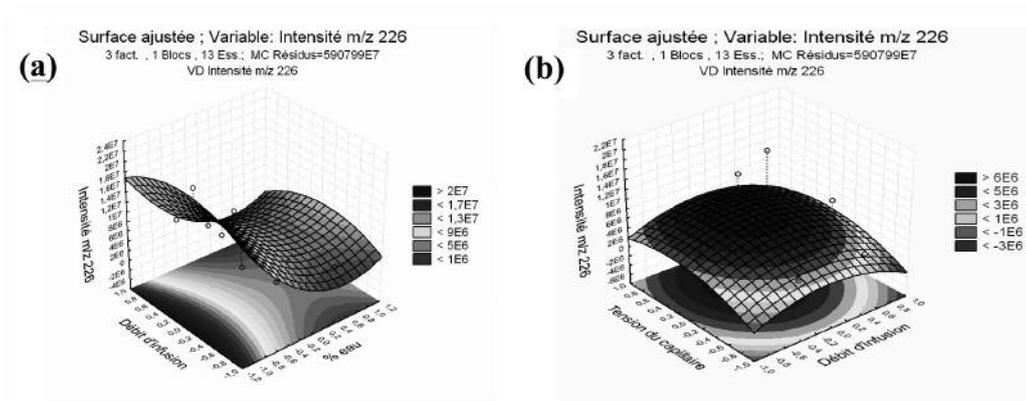


Figure 3 - Surface de réponse de l'intensité de l'ion m/z 226 en fonction (a) du débit d'infusion et du pourcentage d'eau dans le solvant, (b) du débit d'infusion et de la tension du capillaire.

Des contraintes opérationnelles ne nous ont pas permis de réaliser la succession classique des Doehlert de dimension croissante (D_3 , D_4 , D_5 ...). Nous avons choisi une approche originale permettant, en un nombre réduit d'expériences, de confirmer l'influence des facteurs et d'effectuer une première estimation de leur éventuel effet quadratique. A partir du plan D_3 , nous réalisons deux points du plan D_4 (un avec le facteur au niveau haut, un au niveau bas) et comparons les valeurs expérimentales à celles du modèle D_3 qui ignore la quatrième variable. Si la différence est significative, cette nouvelle variable est influente. Nous avons observé cette influence pour plusieurs des facteurs testés. Un complément du plan est actuellement envisagé.

Les premiers résultats apportent dès à présent des informations majeures par rapport à l'objectif de ce travail. Une zone étendue, dans laquelle les conditions expérimentales sont très robustes, a été mise en évidence. De plus, travailler avec un pourcentage d'eau plus faible présente un intérêt pour augmenter l'intensité du signal des ions m/z 226 et m/z 227. Une étude complémentaire a cependant montré que le bruit de fond du signal ionique augmente dans les mêmes proportions. Ainsi, même si l'intensité du signal augmente d'un facteur 4 environ en réduisant la teneur en eau de 50% à 10%, le rapport signal sur bruit reste quasiment constant. Il n'est donc pas possible de diminuer la limite de détection par la seule action sur le pourcentage de solvant organique. Quelle que soit la composition du solvant, la limite de détection de la méthode reste donc équivalente.

CONCLUSION

Par le biais d'une méthode de plans d'expériences, nous avons mis en évidence l'existence d'une large zone de stabilité en un minimum d'essais. La présence de cette zone illustre la robustesse de la méthode d'ionisation employée. Ces résultats pourront être affinés par la suite grâce à l'utilisation de manière séquentielle des plans de Doehlert.

Références

- [1] Cech, N. B. & Enke, C. G. *Mass Spectrom. Rev.*, 20(6): 362-387, 2001.
- [2] Raji, M. A. & Schug, K. A. *Int. J. Mass Spectrom.*, 279(2-3): 100-106, 2009.
- [3] Kruve, A., Herodes, K. & Leito, I. *Rapid Commun. Mass Spectrom.*, 24(7): 919-926, 2010.

Utilisation de l'imagerie multivariée pour estimer la teneur en amylose du grain de blé

B. Jaillais¹, E. Perrin¹, F. Balfourier², P. Berbezy³, D. Bertrand¹

¹UR 1268 BIA, INRA-Nantes, Benoit.Jaillais@nantes.inra.fr

²UMR GDEC-UBP, INRA Clermont-Ferrand, Francois.Balfourier@clermont.inra.fr

³Ulice, Clermont-Ferrand, Pierre.Berbezy@limagrain.com

Keywords: amylose, Imagerie multivariée, blé, régression pas-à-pas

L'amidon est constitué de deux biopolymères, dont l'amylose. Ce polymère de glucose a des propriétés technologiques et nutritionnelles élevées. Le projet ANR NOMAC* (*Nouvelles ressources pour maîtriser le devenir digestif des nutriments des produits céréaliers*) porte sur la qualité nutritionnelle du blé. Il a pour premier objectif de montrer l'effet de l'amylose sur la viscosité du bol alimentaire et donc sur l'index glycémique. Dans le cadre de ce projet, nous cherchons également à développer une méthode de caractérisation grain à grain de la variabilité naturelle de la teneur en amylose du blé. Cette méthode est basée sur la technique d'imagerie multivariée (Figure 1) couplée à un traitement chimiométrique des images.

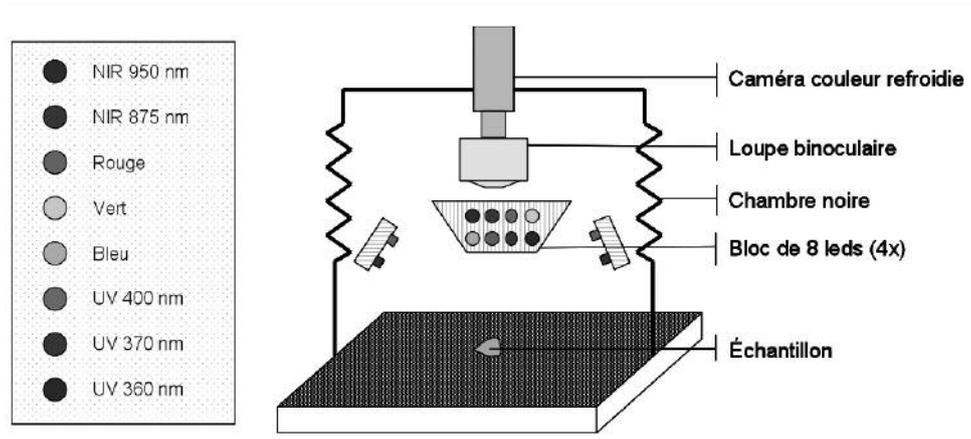


Figure 1 - Système d'imagerie multi-variée

Les grains, sectionnés à mi-hauteur, sont placés sous un éclairage à DEL (diodes émettrices de lumière) couvrant différentes régions du spectre de la lumière: UV, visible et infrarouge, et une acquisition séquentielle d'image est réalisée. Un échantillon est caractérisé par 8 images RGB (Rouge, Vert, Bleu) acquises pour chacune des 8 conditions d'éclairage (Figure 2). L'information spectrale peut être mise sous la forme d'un « cube-image » de 24 plans.

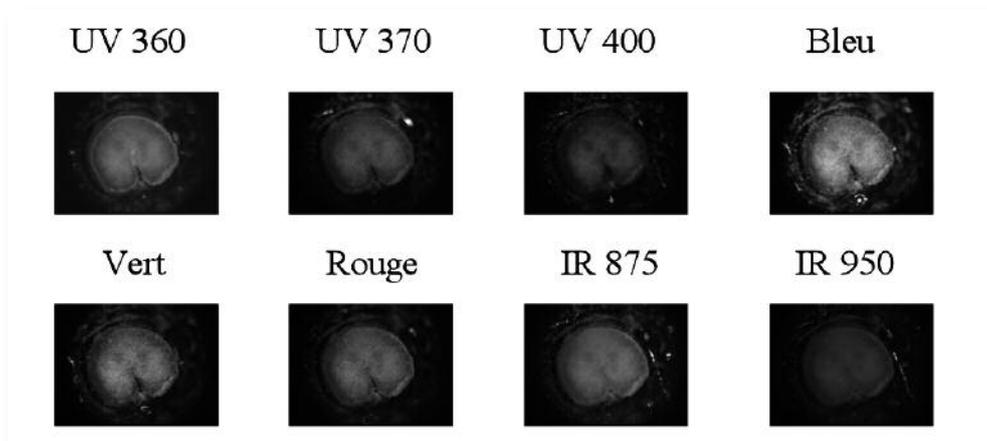


Figure 2 - Images brutes de grains de blé

Une phase d'étalonnage est réalisée par l'acquisition d'images sur des grains à teneur connue en amylose. La phase de tri est ensuite effectuée sur 100 variétés de blé, avec 10 répétitions par variété. L'établissement du modèle de prédiction est basé sur la technique de régression pas-à-pas. Le modèle de prédiction final comprend trois variables prédictives (Figure 3).

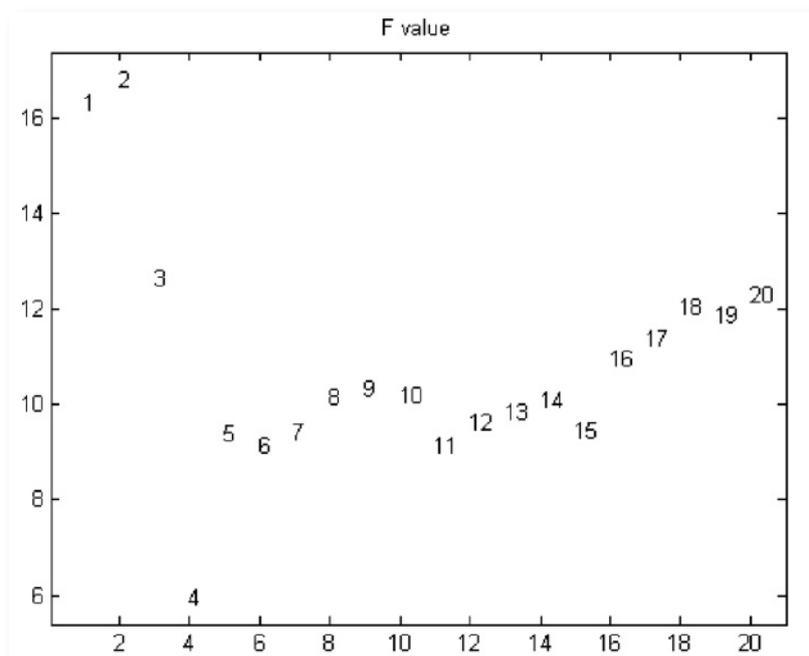


Figure 3 - ANOVA du vecteur de teneur en amylose prédit (facteur = variété) En abscisse sont représentées le nombre de variables dans le modèle de régression pas à pas et en ordonnée la valeur F du test de Fischer.

Parmi ces 100 variétés, 20 d'entre-elles sont sélectionnées pour effectuer des essais agronomiques et nutritionnels complémentaires. Deux critères sont retenus pour les sélectionner. Le premier porte sur la variabilité de paramètres extraits des images des sections de grains : les variétés présentant la variabilité la plus faible sont retenues. Le deuxième porte sur la concentration en amylose estimée : on a cherché à sélectionner les variétés dans une plage de concentration la plus étendue possible.

* Le projet « Nouvelles ressources pour maîtriser le devenir digestif des nutriments des produits céréaliers » (NOMAC) est un projet de recherche fondamentale sur l'apport nutritionnel de produits céréaliers. Le projet a été sélectionné et reçoit une aide financière de l'Agence Nationale de la Recherche, dans le cadre du programme ALIA. D'une durée de 3 ans, du 1er janvier 2009 au 31 décembre 2011, il est soutenu par le pôle de compétitivité Céréales Vallée.

Optimisation de la séparation énantiométrique de trois bêta-bloquants par CLHP chirale à phase inversée

Z.I. Harzallah, F. Farhat, F. Safta, M. Kallel

Laboratoire de chimie analytique ; UR contrôle de qualité des médicaments
Faculté de Pharmacie de Monastir, Monastir 5000, Tunisia
Mohamed.Kallel@fphm.rnu.tn

Mots-clés : Optimisation, énantiomères, plan d'expérience, bêta-bloquant, CLHP chirale

1- Introduction

La chromatographie liquide à haute performance sur colonne chirale est la méthode de choix pour l'analyse des énantiomères des molécules médicamenteuses et les colonnes chirales les plus utilisées sont remplies par des phases stationnaires à base de polysaccharides [1]. Mais la séparation des énantiomères est influencée par les conditions chromatographiques. L'objectif de ce travail est d'optimiser la séparation énantiométrique de trois bêta-bloquants par CLHP à phase inversée sur une phase chirale de cellulose 3-5 diméthyl phényl carbamate .

L'optimisation va consister à étudier l'influence de différents facteurs sur l'énantio-résolution grâce à une approche univariée puis à optimiser la réponse en utilisant l'approche chimiométrique des plans d'expérience. [2]

2- Matériel et Méthodes

L'alprénolol, le métoprolol et l'oxprénolol sont des produits Sigma (Steinheim, Allemagne). Le perchlorate de sodium monohydrate (NaClO₄) a été fourni par Fluka Chemica (Buchs, Suisse). L'acétonitrile a été acquis de Sigma-Aldrich et l'eau bidistillée a été préparée au sein de notre unité de recherche.

La chaîne chromatographique est composée d'une pompe Waters 501, d'un injecteur de type Rheodyne 7725i muni d'une boucle d'injection de 20 µl d'un détecteur UV/Visible Waters 486 et d'un intégrateur SHIMADZU C-R6A chromatopac. La colonne chirale utilisée est une colonne Chiralcel OD-R en acier inoxydable (25cm x 4,6 mm) remplie avec de la cellulose tris (3-5, diméthyl-phényl carbamate) adsorbée sur un gel de silice de granulométrie 10 µm (Daicel, Tokyo, Japon). La phase mobile choisie consiste en un mélange eau / acétonitrile (ACN) 60/40. Le débit est fixé à 0,5 ml/min. La longueur d'onde de détection est de 220 nm. Pour la construction de la matrice d'expériences et l'analyse des résultats, nous avons eu recours au logiciel statistique MODDE 8.0.

3- Résultats et discussion

3.1 Etude Univariée :

L'influence des trois facteurs suivants sur la résolution des énantiomères a été étudiée, en faisant varier à chaque fois un facteur en maintenant les deux autres fixes : la concentration en contre-ion, la température et le pourcentage du modificateur organique.

L'étude univariée a permis de délimiter les intervalles de travail pour les trois facteurs : la concentration en contre-ion a été étudiée dans un intervalle compris entre 0,5 à 1,5 M, la température entre 5 à 25°C et le pourcentage d'ACN entre 25% à 40%.

Mais cette étude ne permet pas d'estimer les interactions entre les différents facteurs.

3.2 Optimisation chimiométrique :

Dans le but de définir la relation entre la réponse et les facteurs, nous avons appliqué un modèle de régression quadratique, comme indiqué dans la formule :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_{11} X_1^2 + \beta_{22} X_2^2 + \beta_{33} X_3^2 + \beta_{12} X_1 X_2 + \beta_{13} X_1 X_3 + \beta_{23} X_2 X_3 + \epsilon$$

Dans cette équation, Y est la résolution énantiomérique, X_1 représente la température de la colonne, X_2 la concentration de NaClO_4 dans la phase mobile, X_3 la proportion de l'ACN dans la phase mobile et ϵ l'erreur résiduelle. Béta représente les différents effets.

Le plan expérimental à trois niveaux de Box-Behnken a été appliqué et 15 points expérimentaux ont été exécutés dans un ordre randomisé et répartis en 12 expériences, et 3 répétitions au niveau central.

3.2.1 Etude de l'effet des différents facteurs

Les différents coefficients du modèle et les valeurs de p correspondantes sont représentées sur le tableau 1. Les différents effets seront discutés pour chaque molécule. Par ailleurs, l'observation des surfaces de réponse (figure 2) nous indique que le maximum de résolution est obtenu pour une concentration de NaClO_4 comprise entre 1,33 et 1,45 M, et un pourcentage en ACN compris entre 25,35 et 25,87 %.

3.2.2 Optimisation :

Après avoir déterminé l'impact des trois facteurs envisagés sur la résolution énantiomérique, nous avons cherché les conditions qui permettent d'obtenir la meilleure séparation des énantiomères étudiés en fixant le temps de rétention maximal pour le deuxième pic. Ainsi, on a pu déterminer les conditions opératoires optimales pour chaque analyte, les valeurs de la résolution ainsi que les temps de rétention prédits et observés.

CONCLUSION

A partir d'un modèle polynomial du second degré, nous avons pu déterminer les effets significatifs sur la résolution des énantiomères pour chaque analyte.

A partir des résultats obtenus et après avoir fixé le temps de rétention maximum désiré du second pic, nous avons pu déduire les conditions optimales pour chaque analyte.

Ainsi, il apparaît que les conditions optimales ne sont pas identiques pour les trois molécules ce qui permet d'affirmer que chaque molécule possède des interactions différentes avec le système chromatographique et que l'optimisation est spécifique de chaque molécule.

Références

- (1) C. Perrin, N. Matthijs, D. Mangelings, C. Granier-Loyaux, M. Maftouh, D.L. Massart, Y. Vander Hayden, *J. Chromatogr. A*, 966 (2002) 119-134
(2) Vladimir Vsol, Anthony F. Fell, *J. Biochem. Biophys. Methods*, 54 (2002) 377-390

	Alprénolol		Oxprénolol		Métoprolol	
	Coef	p-value	Coef	p-value	Coef	p-value
β_1	-0,199	0,003	-0,274	<0,0001	-0,130	0,003
β_2	0,201	0,002	0,324	0,0056	0,222	<0,0001
β_3	-0,547	<0,0001	-1,085	0,0027	-0,630	<0,0001
β_{11}	-0,085	0,167	-1,153	<0,0001	-0,003	0,940
β_{22}	-0,010	0,862	-0,327	0,1378	-0,117	0,023
β_{33}	0,008	0,886	-0,229	0,0129	0,242	0,001
β_{12}	0,017	0,742	-0,030	0,0452	-0,025	0,504
β_{13}	-0,030	0,577	-0,030	0,7143	0,016	0,669
β_{23}	0,165	0,022	0,057	0,5204	0,130	0,013

Tableau 1 - Coefficients du modèle et leur p-values obtenus pour la résolution des énantiomères

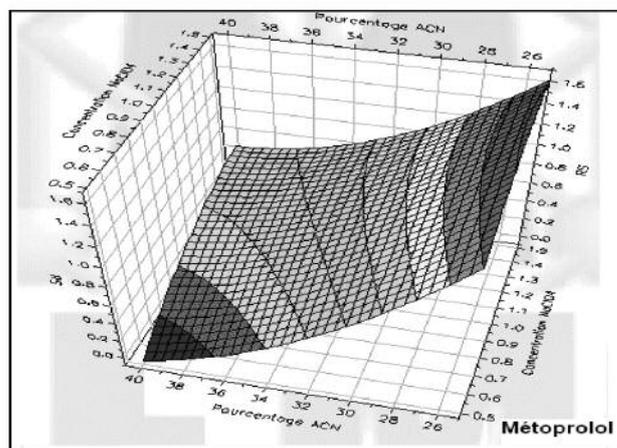


Figure 1 - Surface de réponse du Métoprolol

Analytes	Conditions optimales	Rs		Tr ₂ (min)	
		Prédite	Observée	Prédit	Observé
Alprénolol	NaClO ₄ 1,5M 34,78% d'ACN	1,39	1,29	24,11	22,9
Oxprénolol	NaClO ₄ 1,46M 34,47% d'ACN	1,82	1,62	18,08	20,72
Métoprolol	NaClO ₄ 0,5M 25% d'ACN	1,46	1,32	24,92	21,99

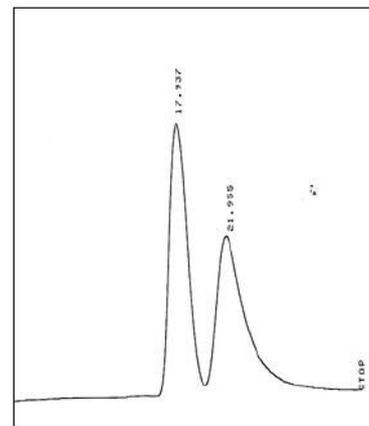


Figure 2 - Chromatogramme du Métoprolol dans les conditions optimales

Analyse quantitative par PLS multitableaux. Application au dosage des triglycérides dans les huiles d'olive vierges françaises

N. Dupuy¹, R. Korifi¹, Y. Le Dréau¹, D. Ollivier², J. Artaud¹

(1) CNRS UMR 6263, ISM², équipe AD²EM, groupe Systèmes Chimiques Complexes, Case 451, Université Paul Cézanne, 13397 Marseille Cedex 20, France.
nathalie.dupuy@univ-cezanne.fr, jacques.artaud@univ-cezanne.fr, yveline.le-dreau@univ-cezanne.fr, rabia.korifi@etu.univ-cezanne.fr

(2) Service Commun des laboratoires du Ministère des Finances, Laboratoire de Marseille, 146 traverse Charles-Susini, 13388 Marseille Cedex 13.

Mots Clés : Huiles d'olive vierges, H-PLS, Raman

1- Introduction

La détermination de la traçabilité des produits de l'industrie agroalimentaire est devenue essentielle en raison des problèmes de santé publique qu'ils peuvent présenter ainsi que pour assurer la loyauté des échanges commerciaux. La traçabilité nécessite de disposer d'outils non subjectifs pour connaître l'ensemble des éléments permettant de suivre le cheminement d'un produit de son origine jusqu'à son utilisation par les consommateurs. La filière oléicole n'échappe pas à cette nécessité de traçabilité. Bien que comportant environ 200 variétés différentes, la majeure partie de l'huile d'olive française est produite à partir d'une dizaine de variétés principales d'olives. Diverses approches ont été proposées pour répondre à la demande de traçabilité des HOV en utilisant des caractéristiques chimique ou organoleptique. Toutefois, la détermination de la traçabilité constitue un véritable défi analytique qui nécessite la mise en œuvre de méthodes permettant la détermination de paramètres multiples pour répondre à cette demande émergente. La chimiométrie permet de faire ressortir les informations les plus pertinentes d'un ensemble de données analytiques. Il est de plus en plus fréquent d'analyser un échantillon par plusieurs méthodes analytiques ce qui génère plusieurs matrices de données. Diverses méthodes d'analyse de données ont été développées afin de résoudre ce type de problème. Généralement, elles sont basées sur la méthode de régression PLS^[1]. Les méthodes multiblock, telles que la H-PLS et la S-PLS permettant d'aborder dans une même expérience, plusieurs régressions consécutives. La H-PLS consiste à utiliser les composantes principales d'une matrice plutôt que les variables spectroscopiques. Les méthodes spectroscopiques (infrarouge et RAMAN) sont depuis longtemps utilisées pour le contrôle qualité des Huiles d'Olive Vierges (HOV). Le but de ce travail a été d'étudier le potentiel des spectroscopies proche (PIR) et Raman confocal associées à un traitement chimiométrique en tant que méthode alternative plus rapide pour les déterminations en triglycérides des HOV.

2- Matériel et méthodes

2.1 Echantillons

Les échantillons proviennent du Centre Technique de l'Olivier, (CTO), Aix-en-Provence, France et sont représentatifs de lots commerciaux produits dans les moulins français. Ils sont issus de trois campagnes oléicoles successives (de 2004/2005 à 2006/2007). Les AOC Haute-Provence (n=47), Nice (n=55), Nyons (n=41) sont constituées d'une seule variété principale. Les AOC Aix-en-Provence (n=103) et Vallée des Baux de Provence (n=111) peuvent être constituées de trois ou quatre variétés principales dont deux sont obligatoires sans que leurs proportions ne soient imposées. L'AOC Nîmes (n=37) doit être constituée d'au moins 60% de la variété Picholine. Parmi ces échantillons, 200 ont été analysés par spectrométrie proche infrarouge et par spectrométrie Raman confocal. Les taux de triglycérides de référence ont été obtenus par chromatographie liquide (HPLC).

2.2 Analyses spectrales

Les spectres Raman sont enregistrés à l'aide d'un spectromètre ALMEGA (Thermo Nicolet, Madison, WI, USA) dans le mode macroscopique. Le signal Raman est obtenu après une excitation à 532 nm. Les spectres sont enregistrés avec une résolution de 2 cm⁻¹ dans la zone spectrale. Les huiles d'olive vierges sont analysées dans des tubes de 2 mm de diamètre. Les spectres proche infrarouge sont enregistrés à l'aide d'un spectromètre AntarisTM FT-Near Infrared. Les conditions d'acquisition des spectres sont les suivantes : acquisition en transmission dans des cuves en quartz de 2mm, gamme spectrale de 10 000 cm⁻¹ à 4 000 cm⁻¹, résolution de 4 cm⁻¹, 16 accumulations de scans.

2.3 Chimiométrie

Le but de cette étude a été de chercher à prédire les taux de triglycérides d'échantillons inconnus à partir de leurs spectres proche infrarouge et Raman confocal et de comparer les résultats obtenus par différentes méthodes d'analyse de données. La prédiction a été réalisée par régression PLS à partir des spectres proche infrarouge d'une part et à partir des spectres Raman confocal d'autre part, puis en utilisant les deux domaines spectraux par la méthode multi block H-PLS. La méthode H-PLS s'intéresse aux variables latentes (scores) des ACP. Elle consiste à réaliser dans un premier temps une ACP avec chacune des séries x_{ij} disponibles. Ensuite, les matrices de scores obtenues sont récupérées puis concaténées ensemble pour former la matrice des scores ACP [2]. La matrice des scores ACP est alors utilisée pour réaliser une régression PLS classique et cette méthode est évaluée à l'aide des indices calculés par validation croisée, au cours de la régression PLS. Les algorithmes et la description de la méthode présentés dans ce document peuvent être retrouvés dans une publication de Westerhuis et Coenegracht [3]. L'évaluation de l'erreur du modèle est estimée par le calcul de l'erreur standard de calibration (SEC) et de l'erreur standard de prédiction (SEP) (1 et 2). Ces paramètres donnent une estimation des performances du modèle de régression construit :

$$SEC = \sqrt{\left(\frac{\sum_{i=1}^N (C_i - C'_i)^2}{N - 1 - p} \right)} \quad (1)$$

$$SEP = \sqrt{\left(\frac{\sum_{i=1}^M (C_i - C'_i)^2}{M} \right)} \quad (2)$$

Avec C_i comme valeur de référence et C'_i la valeur calculée, N le nombre d'échantillon, p le nombre de variables indépendantes de la régression et M le nombre d'échantillons utilisés pour la prédiction. D'autres paramètres permettant l'évaluation des modèles de régressions sont aussi calculés tels que : R^2 : coefficient de régression (calibration), Q^2 : coefficient de détermination (prédiction) et le Biais : différence systématique entre la valeur prédite et la valeur mesurée. Les applications chimométriques sont réalisées à l'aide du logiciel Unscrambler 9.6, distribué par la société CAMO (Computer Aided Modelling, Trondheim, Norway).

3- Résultats et discussions

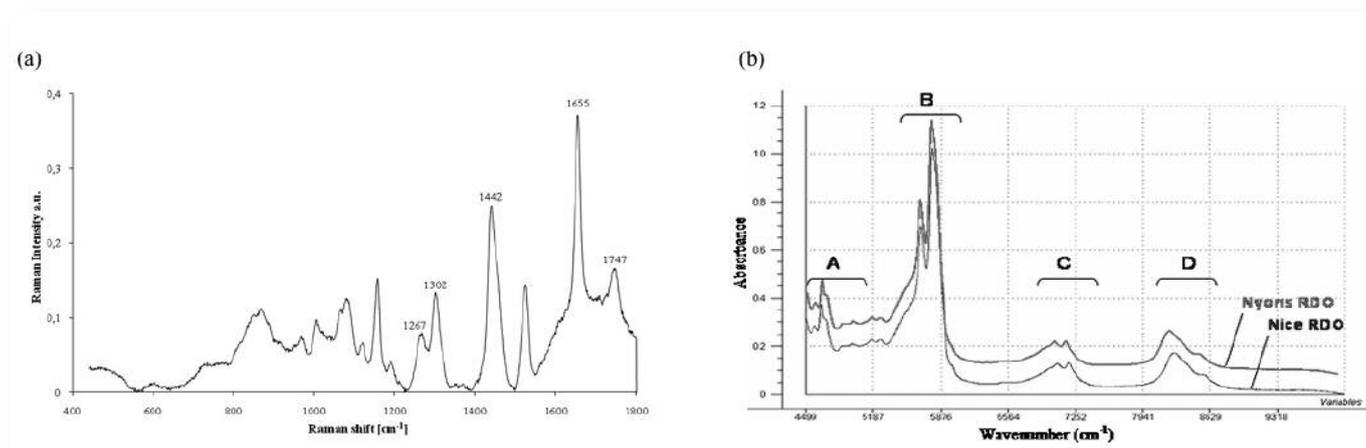


Figure 1 - Spectre Raman (a) et Proche Infrarouge (b) d'une HOV

La figure 1a montre les bandes caractéristiques des huiles d'olive vierges en spectrométrie Raman. Les bandes les plus intenses sont à 1267 cm^{-1} (déformation dans le plan $\delta\text{CHsp}2$ pour les doubles liaisons cis non conjuguées), à 1302 cm^{-1} (CH_2 twisting), à 1442 cm^{-1} (δCH_2), à 1655 cm^{-1} ($\nu(\text{C}=\text{C})$) et à 1747 cm^{-1} ($\nu(\text{C}=\text{O})$) des esters [4]. La figure 1b représente le spectre infrarouge d'huiles d'olive vierges AOC. L'interprétation des bandes est la suivante : A: Combination of the -CH stretching vibration with other vibrational modes, B: First overtone of -CH stretching vibration (methyl, methylene and ethylene groups), C: Combination of the -CH stretching vibration, D: Second overtone of -CH stretching vibration (methyl, methylene and ethylene groups). Les informations spectrales présentes dans les deux domaines sont complémentaires mais très différents.

Il paraît donc intéressant de mixer les deux informations afin de tester le potentiel de la méthode HPLS appliquée à la spectroscopie.

	Range (%)	PLS Raman			PLS Proche Infrarouge			H-PLS Raman+Proche Infrarouge		
		SEC	SEP	REP (%) (LV)	SEC	SEP	REP (%) (LV)	SEC	SEP	REP (%) (LV)
LOL	0.79-4.11	0.158	0.373	16.0 (11)	0.111	0.234	10.0 (6)	0.175	0.277	11.9 (14)
LOO	10.47-17.12	0.246	0.827	6.0 (14)	0.153	0.572	4.1 (5)	0.317	0.563	4.1 (14)
PLO	2.99-10.57	0.670	0.870	12.9 (8)	0.096	0.343	5.0 (6)	0.288	0.605	8.8 (12)
OOO	26.46-55.81	2.630	3.110	7.8 (8)	0.649	1.031	2.6 (6)	0.984	1.894	4.8 (17)
POO	16.13-23.47	0.948	0.994	4.7 (7)	0.079	0.658	3.1 (5)	0.179	0.766	3.7 (16)
POP	1.88-4.48	0.249	0.431	12.9 (9)	0.046	0.207	6.3 (7)	0.063	0.302	9.2 (17)
PoOO	0.92-2.17	0.118	0.241	12.9 (13)	0.030	0.232	12.4 (10)	0.040	0.223	11.9 (18)
SOO	2.78-4.61	0.142	0.377	10.7 (14)	0.045	0.364	10.5 (7)	0.052	0.338	9.7 (18)

Range (%) donné par la chromatographie, SEC: erreur standard de calibration, SEP: erreur standard de prédiction, REP: % d'erreur, LV : variables latentes.

Tableau 1 - Détermination des proportions en triglycérides par traitement PLS et H-PLS

Les résultats obtenus en proche infrarouge sont bons (de 4% à 8% d'erreur relative de prédiction) pour les triglycérides majoritaires (>5% de l'huile), mais médiocres pour les triglycérides minoritaires (< 2%). Les résultats en Raman sont du même ordre de grandeur. Ils sont améliorés par prédiction H-PLS pour les plus faiblement représentés (REP = 8-10%).

CONCLUSION

La prédiction par H-PLS ne permet pas une amélioration conséquente des erreurs relatives de prédiction. Il est connu que le premier coefficient de régression est une bonne approximation du composé pur discriminant^[5] dans le cas de régression PLS. Les coefficients de régression des modèles AOC Vallée des Baux et Nyons présentent, respectivement, une grande similitude avec les coefficients de régression des modèles triglycérides PoOO et SOO. L'avantage de la prédiction par H-PLS est qu'elle permet d'améliorer les erreurs sur le dosage de ces triglycérides faiblement dosés, qui s'avèrent discriminants pour la détermination de l'origine des huiles.

Références

- [1] Wold, H. (1966). Estimation of principal component and related models by iterative least squares. *Multivariate Analysis*, ed. P.R. Krishnaiah, New York : Academic Press.
- [2] J. F. MacGregor, T. Kourti, *Statistical process control of multivariate processes*, *Control Engineering Practice*, 3 (1995) 403-414.
- [3] J. Westerhuis, P. Coenegracht, *Multivariate modelling of the pharmaceutical two-step process of wet granulation and tableting with multiblock PLS*, *J. Chemom.*, 11 (1997) 379-392.
- [4] Barbara Muik, Bernhard Lendl, Antonio Molina-Díaz, María José Ayora-Cañada, *Direct, reagent-free determination of free fatty acid content in olive oil and olives by Fourier transform Raman spectrometry*. *Analytica Chimica Acta*, Volume 487, Issue 2, 8 July 2003, Pages 211-220
- [5] Haaland, D. M., Thomas, E. V. (1988). *Partial Least-Squares Methods for Spectral Analyses . 1. Relation to Other Quantitative Calibration Methods and the Extraction of Qualitative Information*. *Anal. Chem.*, 60(11), 1193-1202.

Apport de traitements chimométriques dans le suivi par IRTF de la dégradation thermique et photochimique d'huiles d'olive vierges.

Y. LE DRÉAU, J. PLARD; R. KORIFI, N. DUPUY, J. ARTAUD, C. RÉBUFA

ISM², UMR 6263, Equipe AD²EM, équipe systèmes chimiques complexes Université Paul Cézanne, case 451, 13397 Marseille cedex 20, France, yveline.le-dreau@univ-cezanne.fr; jerome.plard@etu.univ-cezanne.fr; rabia.korifi@etu.univ-cezanne.fr; nathalie.dupuy@univ-cezanne.fr; jacques.artaud@univ-cezanne.fr; c.rebufa@univ-cezanne.fr;

Keywords: Olive oils, thermo-oxidation, photo oxidation,

1- Introduction

La dégradation des huiles est due à des phénomènes d'oxydation (rancissement) qui réduisent leur durée de conservation et leur valeur nutritionnelle en générant en parallèle des produits toxiques. Ce rancissement est accéléré par la chaleur, la lumière, l'humidité ou la présence de métaux de transition [1-3]. La plupart des tests d'oxydation d'huiles [4,5] conduisent rarement à une image globale du processus d'oxydation et prennent beaucoup de temps. Parmi les méthodes d'analyse utilisées, le moyen infrarouge à transformée de Fourier (IRTF) s'est avéré très performant [1,6] en apportant une information globale. Le but de ce travail a été de développer par IRTF une méthode simple et rapide pour suivre en continu l'évolution de la composition d'une huile d'olive vierge française ayant subi une oxydation contrôlée (thermique et/ou irradiation UV) et évaluer l'impact de ces deux traitements. Des indices spectroscopiques ont été utilisés pour comparer les cinétiques obtenues. Une Analyse en Composantes Principales (ACP) a été utilisée pour différencier les différents stades de l'oxydation. Une méthode multivariée (SIMPLISMA (SIMPLe-to-use Interactive Self Modeling Analysis)) a été employée pour extraire des informations spectrales sur les composants "purs" du système ainsi que leur profil de concentration associé [7]. Pour déterminer le potentiel de la méthode pour différencier la réactivité des huiles, 7 huiles d'origines différentes ont été oxydées par la lumière et de chaleur.

2- Matériel et méthode

Huit tests de vieillissement ont été menés (résumés dans le tableau 1) sur une huile d'olive (AOC Aix en Provence) fournie par le Centre Technique de L'Olivier, Aix en Provence. 5 huiles représentatives des principales variétés françaises (Aglандаu, Salonenque, Tanche, Cailletier et Olivière) et 2 huiles d'AOP Aix en provence (Aix et Coudoux) ont été oxydées par la lumière et de chaleur.

Vieillissement thermique: 3 g d'huile sont chauffés à 180°C (bain d'huile silicone) sous agitation et sous un flux d'oxygène (0,8 bar) dans une cuve en quartz.

Vieillissement photochimique: 3 g d'huile maintenus sous agitation et barbotage d'oxygène (0,8 bar) sont irradiés à l'aide d'une lampe à arc de mercure (ORIEL 6286, 250 watts).

Analyse IRTF: Une goutte d'huile est déposée sur l'accessoire ATR monoréflexion (Golden Gate, Specac) d'un spectromètre AVATAR 370 FT-IR Thermo Nicolet. Les spectres sont enregistrés à l'aide du logiciel Omnic, de 4000 à 600 cm⁻¹ avec une résolution de 4 cm⁻¹ et 64 accumulations. Une référence sur l'air est enregistrée dans les mêmes conditions avant chaque mesure.

Essais	Traitement	Essais	Traitement
1	60 min. à 180°C puis 750 min. sous UV	5	300 min. à 180°C / 180 min. sous UV
2	180 min. à 180°C puis 180 min. sous UV	6	345 min. à 180°C sous UV
3	390 min. à 180°C puis 390 min. sous UV	7	810 min. sous UV
4	180 min. sous UV puis 360 min. à 180°C	8	810 min. à 180°C

Tableau 1 - Récapitulatif des tests d'oxydation menés

3- Résultats et discussion

Les bandes IR subissent des changements au cours du temps de chauffage ou de l'irradiation (figure 1a). La diminution de la bande à 3006 cm⁻¹, attribuée à la liaison C-H en position "cis" sur une double liaison C=C d'un acide gras ou d'un ester d'acide gras insaturé, correspond à la diminution des insaturations cis. L'apparition de la bande vers 3500 cm⁻¹ correspond à la formation de produits primaires et secondaires d'oxydation renfermant des groupements OH tel que les hydroperoxydes, les alcools, les éthers, les acides. Les élargissements des bandes vers 1740 cm⁻¹ attribuée aux liaisons C=O et entre 1800 et 1000 cm⁻¹ correspondent respectivement à la formation de produits secondaires d'oxydation oxygénés (peroxydes, aldéhydes, cétones, acides...). Un accroissement de la bande à 960 cm⁻¹, due à la liaison C-H en position "trans" d'une double liaison C=C (acide gras ou d'un ester d'acide gras insaturés, produits d'oxydation) est attribué à la formation d'insaturations trans caractéristiques d'un vieillissement.

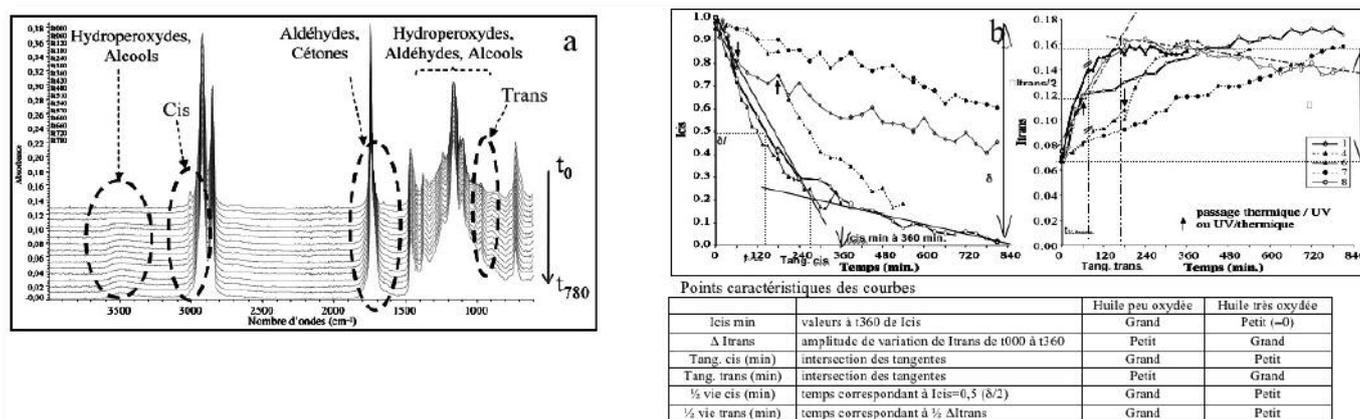


Figure 1 - (a) Changements des bandes lors d'une oxydation, (b) évolution des indices "cis" et "trans".

Une analyse en composantes principales a été réalisée sur l'ensemble des données spectrales obtenues lors du test 7 et 8 (centrage des données, validation croisée, retrait de la zone d'absorption du CO₂). La première composante principale (99% de la variance totale) discrimine dans la partie négative tous les spectres obtenus après une irradiation UV et ceux obtenus après chauffage entre 0 et 180 min. et dans la partie positive les derniers temps du vieillissement thermique (T>180 min.). Le loading associé montre que les spectres de la partie négative se différencient par des bandes à 3006, 2920, 2850, 1750 et 717 cm⁻¹. Les derniers temps thermiques se différencient par un massif à 3500 cm⁻¹, la bande à 1724 et la zone entre 800 et 1400 cm⁻¹ caractéristiques des produits d'oxydation. Le deuxième axe discrimine dans la partie négative, les premiers temps des deux vieillissements ainsi que les derniers temps thermiques. Les bandes qui les différencient (3006, 1740, 1710, 1157, 707 cm⁻¹) caractérisent les structures "cis". Les temps intermédiaires du vieillissement thermique et les derniers spectres de l'huile irradiée se démarquent par des absorptions identifiant une isomérisation "trans" (2930, 2850 et 970 cm⁻¹).

La figure 1b montre l'évolution au cours du temps des indices Icis (relatif à la bande à 3006 cm⁻¹) et Itrans (relatif à la bande à 960 cm⁻¹) calculés à partir de rapport d'aires de bandes [8] pour les différents tests ainsi que les points caractéristiques des courbes permettant la comparaison de différentes huiles. La chaleur dégrade d'avantage d'insaturations cis (diminution de l'indice "cis") et induit d'avantage de liaisons trans (augmentation de l'indice "trans") et plus rapidement que l'irradiation. Cependant lorsque l'échantillon est irradié longtemps, l'indice "trans" atteint la même valeur que lorsque l'échantillon est chauffé. Par contre les valeurs de l'indice "cis" restent supérieures. Ceci suppose que, soit l'irradiation favorise la formation d'insaturations trans, soit le chauffage entraîne dans un deuxième temps la dégradation des insaturations trans après les avoir induites. Il ne semble pas y avoir d'effet de synergie entre les deux traitements.

Le traitement des données (manipulations 1 à 8) par la méthode SIMPLISMA a permis d'extraire mathématiquement 3 spectres (figure 2). Le premier décrit des triglycérides ayant des insaturations cis (3006, 2920, 2851, 1745, 1450, 1377, 1237, 1160, 1118, 1094 et 724 cm⁻¹). Son profil de concentration montre une diminution de ces espèces au cours du temps qui est beaucoup plus importante après un stress thermique qu'après une irradiation UV. Le deuxième spectre peut être attribué aux hydroperoxydes des triglycérides présentant une isomérisation trans (3500, 2920, 2855, 1742, 1464, 1148, 1112, 970 et 891 cm⁻¹). Ces produits se forment au cours des différents traitements et sembleraient se dégrader après une longue période de chauffage (manipulation 8). Le troisième spectre est relatif aux composés d'oxydation avec de longues chaînes (724 cm⁻¹) et des fonctions acides (3500 et 1733 cm⁻¹). Ces derniers constituants se forment en plus grande quantité lors d'une oxydation thermique. Le rayonnement UV apparaît moins agressif en engendrant peu de ces structures.

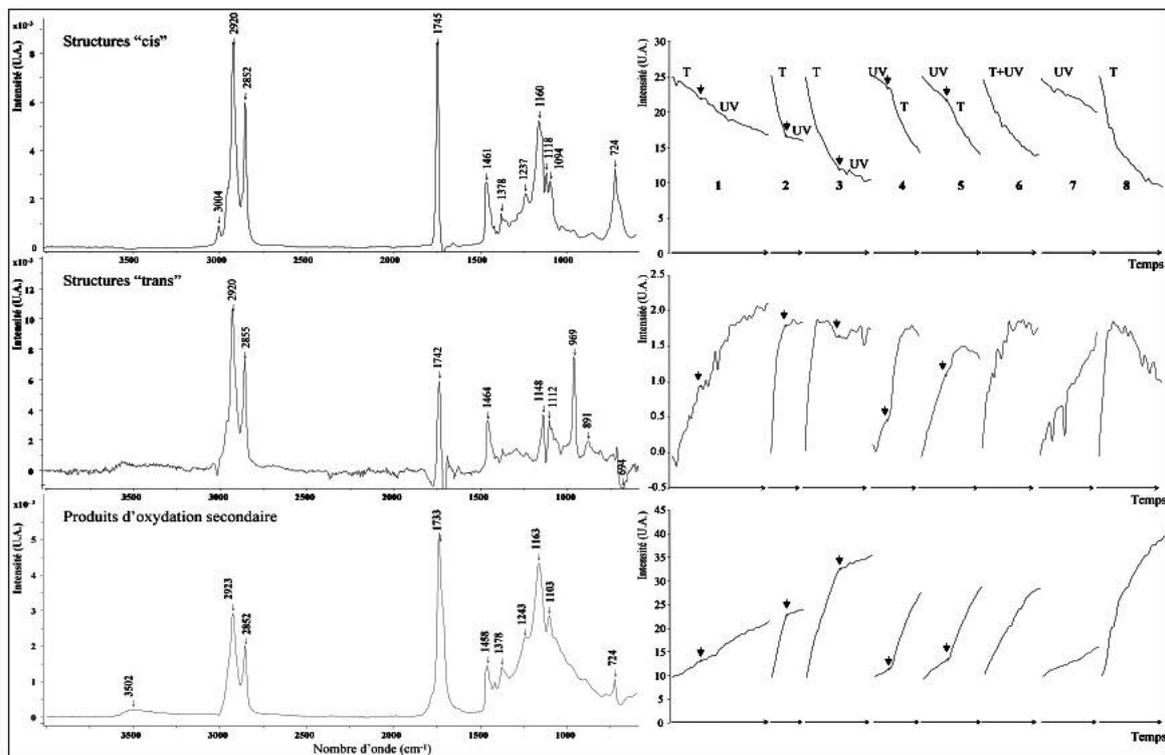


Figure 2 - Spectres extraits mathématiquement et leur profil de concentration associé.

CONCLUSION

Les deux modes d'interprétation des données spectrales sont complémentaires. La méthode des indices et l'ACP permettent de comparer simplement différents échantillons ou différents traitements. L'utilisation de SIMPLISMA met en évidence la présence de familles de composés et de suivre leur répartition relative au sein de l'huile évoluant au cours du temps.

Références

- [1] M.D. Guillén, N. Cabo. Fourier transform infrared spectra data versus peroxide and anisidine values to determine oxidative stability of edible oils. *Food Chem.*, 77(4): 503-510, 2002.
- [2] T.A. Russin, F.R. Van de Voort, J.Sedman Novel method for rapid monitoring of lipid oxidation by FTIR spectroscopy using disposable IR cards. *J. J. Am. Oil Chem. Soc.*, 80(7): 635-641, 2003.
- [3] F.R. Van de Voort, S J. edman, S.T.H. Sherazi. Correcting for underlying absorption interferences in Fourier transform infrared trans analysis of edible oils using two-dimensional correlation techniques *J. Agric. Food Chem.* 56(5): 1532-1537, 2008
- [4] T.A. Isbell, T.P. Abbott, K.D. Carlson. Oxidative stability index of vegetable oils in binary mixtures with meadowfoam oil. *Ind. Crop. Prod.* 9: 115-123, 1999.
- [5] European standard NF EN ISO 6886, AFNOR Paris (1996).
- [6] J.B. Russel. Rancity of foods. Ed Allen JC and Hamilton RJ, Elsevier Applied Science Publishers (London), 1989.
- [7] W. Windig, N.B. Gallagher, J.M. Shaver, B.M. Wise. New Approach for Interactive Self-Modeling Mixture Analysis. *Chemometr. Intell. Lab.*, 77(1-2): 85-96, 2005.
- [8] Y. Le Dréau, N. Dupuy, V.Gaydou, J. Joachim, J. Kister. Study of jojoba oil aging by FTIR. *Anal. Chim. Acta.*, 642: 163-170, 2009.

Interaction emballage(PET)-huile : mise en évidence de la contamination de certaines huiles végétales par croisement des données spectroscopiques

M.Bitar , J.Maalouly¹, C.O.Saliba, et H.Chebib

¹ Laboratoire des analyses chimiques, Université Libanaise, Faculté des Sciences II, Jdeideth el Matn 90656, Fanar, Liban
j_maalouly@hotmail.com

Mots-clefs : Huile végétale, PET, 2DCOS.

1- Introduction et problématique

Par sa présence quotidienne dans la vie du consommateur, l'emballage alimentaire ne doit pas nuire à sa santé. Cependant, les plastiques, étant des polymères organiques synthétiques, contiennent des monomères et des additifs susceptibles de migrer vers l'aliment et avoir des conséquences organoleptiques et toxicologiques.

Dans le cas des huiles végétales, il s'avère important de vérifier si la présence des contaminants migrants potentiels (benzène, éthylbenzène, bisphénol A, diméthylphtalate et dibutylphtalate) influence sur l'évolution des antioxydants (alfa-tocophérol et composés phénoliques) desquels dépend la qualité de ces huiles.

2- Matériel et méthodes

Il y avait 4 types d'échantillons : Huile d'olive dans du verre (benzène, éthylbenzène, diméthylphtalate, dibutylphtalate, bisphénol A : échantillons de références) ; Huile d'olive dans du PET ; Huile de tournesol (2 dates d'expiration différentes) ; huile de tournesol placée dans le verre.

Les échantillons ont été conservés à 40°C dans l'étuve pour subir un vieillissement accéléré sans altération de l'huile, puis analysés tous les 10 jours par deux méthodes spectroscopiques : spectrofluorimétrie et spectrophotométrie UV-Vis. Les résultats ont été ensuite traités par la spectroscopie de corrélation en deux dimensions(2DCOS).

3- Résultats et discussion

La collection spectrale obtenue est de 144 spectres pour les 28 échantillons d'huile d'olive (dans le verre avec et sans dopage et dans le PET) avec 3 répétitions pour chaque échantillon. Et les spectres collectés correspondant à l'huile de tournesol sont au nombre de 51 (17 échantillons avec 3 répétitions chacun), 76 longueurs d'onde ont été balayées sur UV-Vis, et 81 longueurs d'onde avec le spectrofluorimètre. Les tailles des matrices seront donc (144,76) et (144,81) respectivement pour UV-Vis et spectrofluorimétrie appliquées aux échantillons d'huile d'olive, et (51,76), (51,81) pour les huiles de tournesol. Après ANNOVA les matrices des huiles en UV-Vis seront (144,51) et (51,51).

3.1 Etude cinétique par la corrélation en deux dimensions 2DCOS des spectres UV-Visible et de fluorescence appliquée à l'huile de tournesol.

L'intérêt surtout de cette méthode dans le cas de l'huile de tournesol est de suivre l'évolution de ces huiles en fonction du temps. Nous avons des espèces (migrants) qui apparaissent dans le milieu. Nous pouvons par l'intermédiaire de cette méthode et suivant les signes des pics se trouvant dans les spectres synchrones et asynchrones (figure 1) savoir l'ordre chronologique d'apparition des composés dans le milieu étudié.

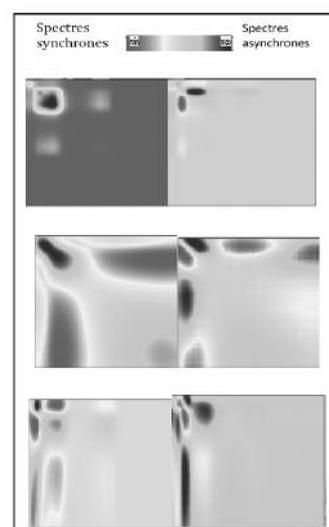


Figure 1 - spectres synchrones et asynchrones de 2DCOS fluorescence-UV des huiles de tournesol

Le tableau 1 montre quelques événements avec leur ordre. Notons que les matrices ont été arrangées suivant l'ordre du vieillissement. Commençons par la corrélation UV-UV, nous remarquons que le composé à 212 nm apparaît avant celui à 226 nm. A 212 nm l'éthylbenzène et les composés phénoliques de l'huile se trouvent avant le DMP et le BPA (226nm). Le benzène (204nm) apparaît dans le milieu après les composés phénoliques déjà présents (210 nm), il a donc probablement migré à partir du PET. Nous constatons aussi l'apparition du DMP, et du bisphénol A (226 nm) après l'alfa tocophérol (270 nm) qui était déjà dans le milieu. Nous passons ensuite aux spectres de corrélation fluorescence-fluorescence, le pic à 360 nm apparaît après celui à 312 nm donc la vitamine E est déjà dans le milieu, les phtalates (360nm) migrent ultérieurement. Enfin la combinaison entre les 2 méthodes spectroscopiques (X=longueur d'onde de fluorescence, Y=longueur d'onde d'absorption UV) est aussi informative. Le composé fluorescent à 292 nm (composés phénoliques) est présent avant le composé qui absorbe à 230 nm dans UV-Vis (DBP). D'autre part, le composé fluorescent à 360 nm (les phtalates) migre après l'éthylbenzène (212 nm). Enfin nous pouvons voir que le benzène (204 nm) apparaît après la vitamine E (312 nm).

De la même manière, nous dressons le tableau de cinétique de l'huile d'olive (tableau 2) et on obtient la même ordre chronologique. Donc, cette corrélation en deux dimensions des spectres nous a permis de mettre en évidence la migration à partir du PET des contaminants potentiels tels, le DMP, le DBP, le BPA, le benzène et l'éthylbenzène vers l'huile.

	Corrélation	Synchrone	asynchrone	Ordre des événements
UV-UV	(226nm-212nm)	+	-	226 ← 212
	(204nm-210nm)	-	+	204 ← 210
	(270nm-226nm)	-	-	270 → 226
Fluorescence-fluorescence	(312nm-360nm)	-	-	312 → 360
Fluorescence-UV	(292nm-226nm)	-	-	292 → 230
	(360nm-212nm)	-	+	360 ← 212
	(312nm-204nm)	-	-	312 → 204

Tableau 1 - tableau de cinétique ; corrélation en 2 dimensions des spectres UV-Visible et fluorescence de l'huile de tournesol.

	Corrélation	Synchrone	asynchrone	Ordre des événements
UV-UV	(226nm-212nm)	+	-	226 ← 212
	(204nm-210nm)	-	+	204 ← 210
	(270nm-226nm)	-	-	270 → 226
Fluorescence-fluorescence	(312nm-360nm)	-	-	312 → 360
Fluorescence-UV	(292nm-226nm)	-	-	292 → 230
	(360nm-212nm)	-	+	360 ← 212
	(312nm-204nm)	-	-	312 → 204

Tableau 2- tableau de cinétique ; corrélation en 2 dimensions des spectres UV-Visible et fluorescence de l'huile d'olive.

CONCLUSION

Cette étude a permis de mettre en évidence des différences entre les échantillons d'huiles d'olive et de tournesol conservés dans le PET et dans le verre et d'identifier les composés à l'origine de ces différences (alpha-tocophérol, composés phénoliques et contaminants potentiels) par croisement des méthodes spectroscopiques et chimiométriques. Ce travail peut être suivi ultérieurement par la quantification des composés qui ont été à la base des différences discernées. Il est aussi possible d'étendre l'étude pour englober différentes marques de ces huiles et même d'autres types d'huiles végétales.

Application des outils chimiométriques d'analyse multivariée à l'évaluation des impacts écotoxicologiques des boues d'épuration contaminées par les pesticides et le nonylphénol sur les sols agricoles.

Aline Ghanem¹, Jacqueline Maalouly¹, Viviane Huteau², Yves Levi², Christian Mougin³

¹Université Libanaise, Faculté des Sciences II, Département de Chimie, P.O.Box 90-656 Jdeidet-El-Matn, LIBAN. alghanem@ul.edu.lb

²Laboratoire Santé publique et Environnement, Université Paris XI, Faculté de Pharmacie, 5 rue J.B. Clément, 92296 Chatenay-Malabry, France.

³Physicochimie et Ecotoxicologie des Sols d'Agrosystèmes Contaminés, INRA, Route de Saint-Cyr, 78026 Versailles Cedex, France. mougin@versailles.inra.fr

Mots-clefs : pesticides, analyse en composante principale (ACP), écotoxicologie, boues d'épuration

1- Introduction

La valorisation agricole des boues d'épuration urbaines par épandage sur les sols agricoles est une pratique très répandue de nos jours pour des raisons socio-économiques. Bien que la valeur fertilisante de ces déchets soit reconnue, les impacts écotoxicologiques de ces épandages sur le milieu terrestre sont encore mal connus. Ceci est particulièrement dû au fait que de nombreux polluants organiques, dont les pesticides ou des perturbateurs endocriniens, ne sont pas pris en compte, notamment par la réglementation Française (Décrets du 8/12/97 et du 8/01/98). L'évaluation des impacts écotoxicologiques de trois types de boues (séchée, compostée, chaulée) appliquées à deux taux d'épandage 6 t/ha et 30 t/ha, a été réalisée en cosmes sur des espèces représentatives des différents niveaux écologiques des sols. Des tests écotoxicologiques ont été menés sur plantes (germination, élongation des racines ou des tiges), collemboles, bactéries et champignons. L'impact estrogénique de lixiviats du sol a également été évalué sur cultures cellulaires.

Les outils chimiométriques d'analyse multivariée en particulier l'ACP a été utilisée pour déterminer les effets des différents paramètres (type de boue, taux d'épandage, contamination...) sur l'ensemble des points finaux mesurés. Les résultats obtenus montrent des effets remarquables des boues apportées au taux d'épandage élevé sur différentes espèces du sol.

2- Matériel et méthodes

Le système expérimental comprenait des cosmes cylindriques ouverts (diamètre interne = 50 cm, hauteur = 40 cm), contenant 40 kg de sol. L'incubation était pendant 3 mois en chambre climatique sous alternance jour/nuit (jour 16 h à 25 °C, nuit 8 h à 18 °C). Une humidité de 80 % de la capacité de rétention du sol a été également maintenue pour le bon développement de la microflore du sol et contrôlée par une sonde.

Les boues ont été appliquées à deux doses : 6 t.ha⁻¹ correspondant à la dose annuelle autorisée, et 30 t.ha⁻¹ c'est-à-dire la quantité de boue maximale autorisée par la réglementation française sur une période de 10 ans, mais apportée ici en une seule fois. Les boues sont contaminées par les trois composés (glyphosate, diuron et 4-n-nonylphénol) en mélange. Les échantillons dopés (<6-17> µg.g⁻¹ matière sèche) ont été mélangés à l'horizon supérieur (0-10 cm) des cosmes.

Des cosmes supplémentaires ont été également conduits en présence d'échantillons de boues n'ayant pas subi de dopage, pour évaluer l'influence des différentes formes de boues sur les paramètres testés. Deux cosmes témoins ont été également mis en place : un sol non dopé et un autre dopé avec la dose de pesticides apportée par la boue à 30 t.ha⁻¹.

Trois types de boues ont été testés : une boue sèche de la station d'épuration de Plaisir, une boue compostée de la station de Gazeran et une boue chaulée (30% g/g) de la station d'Elancourt.

Les effets de l'épandage de la boue aux deux doses d'applications ont été évalués dans les cosmes sans dopage en pesticides. Par ailleurs, les effets des pesticides et du nonylphénol ont été mis en évidence en comparant les effets dans les cosmes dopés par rapport à ceux impliquant des boues non dopées.

3- Résultats et Discussion

Les effets sur la décomposition de la matière organique ont été évalués par le test du « litterbag ». L'analyse en composante principale a permis de mettre en évidence un effet positif (stimulateur) des différents types de boues appliquées sans dopage à 6 t.ha⁻¹ au sol. Cependant, un effet négatif (inhibiteur) est remarquable à forte dose d'application (30 t.ha⁻¹). Le dopage des boues inverse les effets sur la décomposition de la matière organique du sol, comme le montre la figure 1.

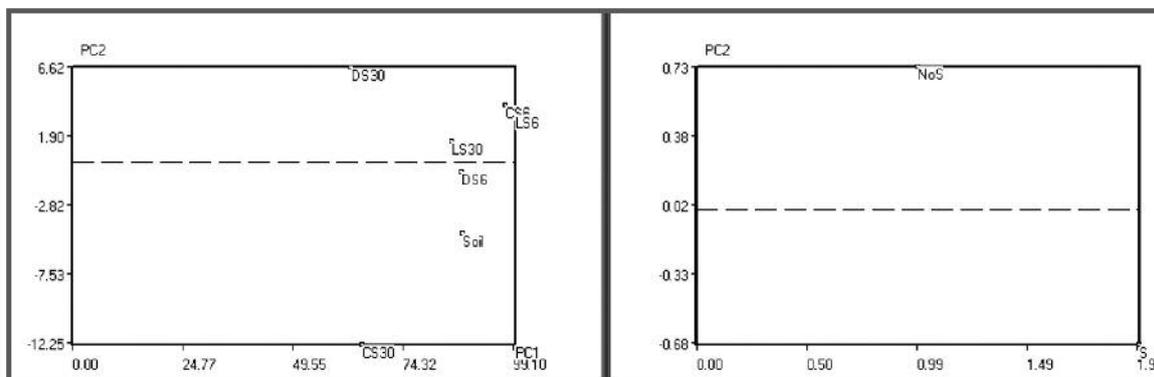


Figure 1 - ACP (4 composantes principales) sur les résultats du test du « litterbag » : à gauche scores PC2/PC1, à droite loading de PC2.

L'analyse des acides gras phospholipidiques (PLFA) est utilisée pour estimer, de façon semi-quantitative, l'effet des boues d'épuration sur l'évolution des communautés microbiennes du sol. L'ACP montre une augmentation de la biomasse bactérienne et fongique suite à l'apport de la boue séchée et compostée. Les effets stimulateurs peuvent être liés à l'apport des nutriments par la boue qui contribue aussi au développement de certaines communautés, qui se traduit par une augmentation des teneurs des PLFA correspondants. Des effets toxiques de la boue peuvent toutefois se manifester, comme c'est le cas avec la boue chaulée qui réduit les teneurs du sol en bactéries. En effet, l'apport de la chaux est réalisé dans les STEP pour réduire la teneur des boues en microorganismes. Le taux des champignons n'est toutefois pas affecté par la chaux. Il est même augmenté, et ce avec les trois formes de boue. Le dopage en pesticides et nonylphénol accroît l'effet négatif sur les champignons par rapport au non dopage dans le cas des boues chaulées.

Les effets sur les producteurs de biomasse ont été évalués en incluant les paramètres suivants : taux de germination (n=10), indice de délai de germination et quantité de biomasse produite (g) pour deux types de plantes : radis et blé. La boue chaulée inhibait la germination en raison d'un fort développement fongique dans les cosmes. Dans le cas du non dopage, l'ACP montre qu'à forte dose d'application de 30 t.ha⁻¹, toutes les boues avaient un effet inhibiteur du taux de germination. L'indice de délai de germination était faiblement affecté. La production de biomasse était généralement améliorée à faible dose d'application. Dans le cas où les boues sont dopées, le taux de germination semble être réduit (effet négatif du mélange pesticides+nonylphénol). L'indice du délai de germination augmente montrant un effet toxique du dopage (figure 2).

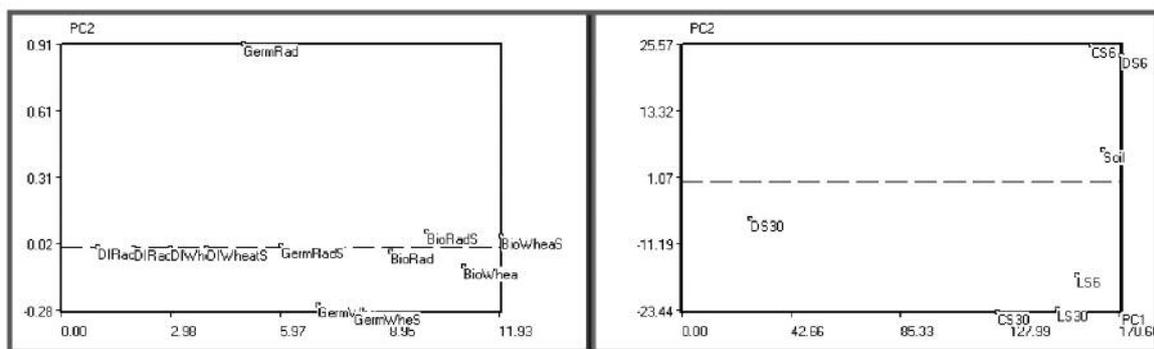


Figure 2 - ACP (4 composantes principales) sur les effets sur les plantes : à droite scores PC2/PC1, à gauche loading de PC2.

Les effets sur les consommateurs ont été évalués sur des collamboles *F. candida*. L'impact final sur le développement de la population résulte d'effets sur la reproduction (toxicité chronique) et d'effets sur la mortalité des adultes et juvéniles (toxicité aiguë). Cependant, la reproduction de *F. candida* est plus sensible aux polluants que la mortalité et fait donc l'objet d'un test normalisé. Les effets de la boue sur la reproduction des collamboles sont positifs : stimulation après apport de boue sèche à faible dose. Une augmentation de la dose de boue à 30 t.ha⁻¹ a des effets négatifs sur *F. candida* du fait du développement des champignons. L'amendement du sol avec la boue à forte dose provoque dans certains cas la mortalité des collamboles. Cet effet est indirect par piégeage dans le mycélium des champignons qui ont proliféré. La boue compostée réduit la survie de ces collamboles mais avec un effet non significatif. Par ailleurs, la diminution de la dose de boue diminue son impact néfaste sur la survie.

Des effets de ces boues se manifestent également sur la reproduction. A forte dose, les boues réduisent le taux de reproduction de cette espèce selon l'ordre suivant : boue compostée < boue chaulée < boue séchée. Les effets de cette dernière sont indirects liés à la prolifération des champignons. En revanche, à faible dose, les effets de ces boues sont stimulateurs de la reproduction, par apports de nutriments. Cet effet positif est considérable avec la boue chaulée et séchée, mais non significatif avec la boue compostée. L'analyse par ACP seul n'est pas concluante par rapport aux études statistiques classiques dans le cas de cette étude sur *F. candida*. Des comparaisons avec d'autres techniques multivariées doivent être effectuées par la suite.

L'estrogénicité des lixiviats a été évaluée par le test MELN. La toxicité des lixiviats a été testée à trois concentrations : la concentration initiale de la solution collectée, des concentrations 10 fois et 25 fois supérieures. Ces effets montrent que les boues, quelque soit leur forme, contiendraient des composés perturbateurs endocriniens qui peuvent migrer dans le sol pour contaminer les nappes phréatiques. Des effets oestrogéniques liés aux pesticides ont été observés pour les deux voies d'application au sol : directement ou via les boues plus accentué que dans les cosmes sans dopage Cette toxicité des pesticides semble plus importante après 2 mois d'incubation qu'après 45 jours (boue séchée).

CONCLUSION

L'ACP s'apprête bien à l'évaluation des effets écotoxicologiques dans cette étude. Elle permet une mise en évidence d'effets particuliers (positifs ou négatifs) sur un large éventail de paramètres. Ces effets sont souvent difficilement exploitables en raison de la complexité et de la multitude de résultats obtenus. Cependant, pour certains tests, sur les collamboles en particulier, une confirmation des effets avec les méthodes statistiques classiques est souvent nécessaire. Des interprétations plus concluantes peuvent être discutées suite à d'autres types d'analyses multivariées.

Application of chemometric tools on olive stones acid hydrolysates

***Jeanne Andary^{a,b}, Jaqueline Maalouly^b, Rosette Ouaini^c, Hanna Chebib^c, Douglas N. Rutledge^a, Naim Ouaini^b**

^a AgroParisTech, Laboratoire de Chimie Analytique, 16 rue Claude Bernard, Bernard, 75005 Paris, France, (Kaslik, 446 Jounieh, Lebanon Tel :+ 9619 600900; jeanne_andary@hotmail.com)

^b Departement of Chemistry and life Science, Université Saint-Esprit de Kaslik, 446 Jounieh, Lebanon

^c Departement of Chemistry, Faculty of Science, Lebanese University, 90656, Jdaideth El Matn, Fanar, Lebanon.

Keywords: Olive stones, diluted acid hydrolysate, xylose, furanes, 2DCOS

1- Introduction

Olive stones are mainly subjected to biotechnological or chemical modifications in order to be transformed into animal feed [1]. Dilute-acid hydrolysis (DAH) has proven to be a fast and cheap method to produce sugar from lignocellulosic materials [2, 3]. Hemicellulose constitutes 20-35% dry weight in lignocellulosic biomass, with D-xylose as major sugar after hydrolysis. However, dilute-acid hydrolysis leads to contains toxic compounds such as furans, aliphatic acids, and phenolic compounds [4]. Larsson (1999), found that overliming is the most cost effective method for detoxifying hydrolysates [5]. A potential drawback of overliming is sugar degradation reactions [6].

2- Materials and methods

Hydrolysis of olive stones is carried out with a sulphuric acid solution, while overliming is performed with $\text{Ca}(\text{OH})_2$ under different conditions: pH (10 or 12), temperature 25 or 60°C and time (15, 30 and 60 min). An experimental design including $2^2 \times 3^1 = 12$ experiments is prepared using Nemrodw software [7]. High-performance liquid chromatography (HPLC) is used to analyze samples. Sugars are analyzed using an Aminex column and a refractive index detector (RI-150 Spectra System, Thermo Finnigan). The eluent is composed of acetonitrile and water (95, 5: v/v) at a flow-rate of 1 mL.min⁻¹. Furfural (FF), and Hydroxymethylfurfural (HMF) are analyzed on a C₁₈ column with a UV detector at 280 nm (UV 1000-Thermo Finnigan). The mobile phase is composed of water, methanol and acetic acid (80, 10, 3: v/v/v). The flow is programmed as follows: 0.5 mL.min⁻¹ from 0 to 10 min, 1 mL.min⁻¹ from 10 to 11 min (hold 9 min).

In our study we applied 2D correlation spectroscopy to three combinations of data, first by combining chromatograms from HPLC-UV with themselves, second by combining chromatograms from HPLC-RI with themselves and finally by combining chromatograms from HPLC-UV and HPLC-RI in order to reveal variations due to overliming process. Matrix resulting from assembling of HPLC-RI chromatograms (xylose) (39*4090) was made of 39 samples (12*3 + 3 hydrolysate) and of 4090 variables (retention time), thus for HPLC-UV (furanes) was made of (39*3413).

3- Results and discussion

The benefits of the 2DCOS method are that small features that may be overlooked when visually examining complex data sets can be more easily visualised [8]. The order in which the constituents vary can be deduced from the sign of peaks in the synchronous and asynchronous matrices, facilitating the interpretation of kinetic studies [9]. In order to clarify the nature of the changes occurring during overliming, the retention times and signs of the cross-peaks are listed in Table 1, with the order of events.

	Correlation	Synchronous	Asynchronous	Order of events
HPLC-UV / HPLC-UV	(250 sec. - 103 sec.)	-	+	250 sec. → 103 sec.
	(250 sec. - 111 sec.)	-	+	250 sec. → 111 sec.
	(250 sec. - 127 sec.)	-	-	250 sec. → 127 sec.
	(103 sec. - 250 sec.)	+	-	250 sec. → 103 sec.
	(111 sec. - 250 sec.)	+	-	250 sec. → 111 sec.
HPLC-RI / HPLC-RI	(625 sec. - 580 sec.)	-	-	580 sec. → 625 sec.
	(580 sec. - 625 sec.)	-	+	580 sec. → 625 sec.
HPLC-UV / HPLC-RI	(625 sec. - 103 sec.)	-	+	625 sec. → 103 sec.
	(625 sec. - 111 sec.)	-	+	625 sec. → 111 sec.
	(580 sec. - 250 sec.)	-	-	580 sec. → 250 sec.
	(625 sec. - 250 sec.)	+	+	625 sec. → 250 sec.

^cX sec. → Y sec.: the event of X occurs before that of Y

Table 1: Synchronous and asynchronous correlations, signs and order of events obtained from the correlation of overlimed chromatograms.

We detail below the correlation between HPLC-UV and HPLC-RI chromatograms. The examination of the synchronous and asynchronous matrix highlights the presence of peaks in the same places (Figure 1 A and B). This observation can be explained by the fact that carbohydrates (xylose) and furans (FF and HMF) are not the only products present in olive stones DAH. The presence of positive peaks (625 sec. - 250 sec.) in the synchronous and asynchronous matrix between xylose (650 sec.) and FF (250 sec.) was expected, this latter being a breakdown product of the former. The peak at (625 sec. - 103, 111 sec.) presents a negative sign in the synchronous matrix but a positive one in the asynchronous matrix; which means that xylose (625 sec.) starts to vary before the group of peaks appearing towards 103 and 111 sec. (sugar degradation products), as a function of the severity of overliming treatment (higher pH, temperature and time).

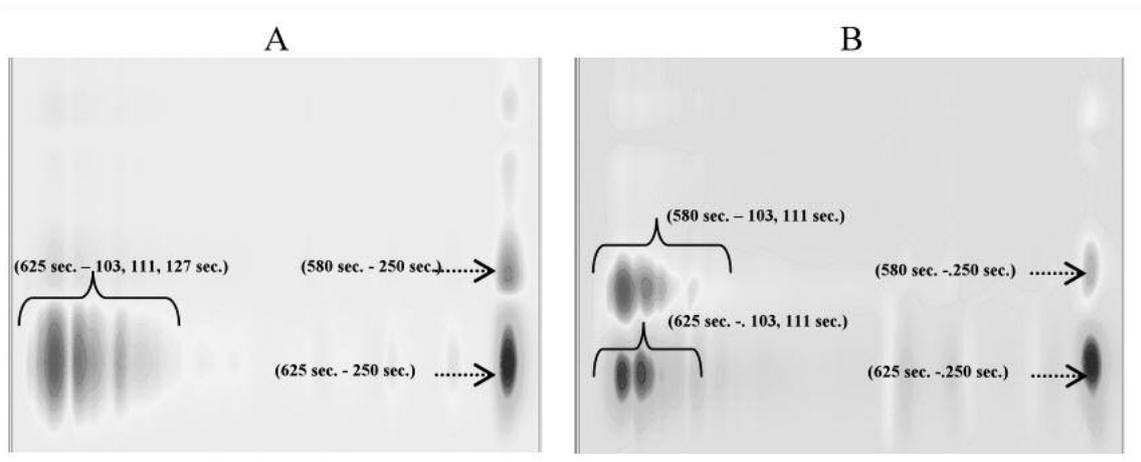


Figure 1: Matrix resulting from the hetero-correlation between the HPLC-RI and HPLC-UV chromatograms: (A) Synchronous, (B) Asynchronous.

When overliming treatments becomes more severe, FF undergoes degradation and causes an increase in the concentration of components appearing at 103, 111 and 127 sec. Moreover, the peak (580 sec. - 250 sec.) is negative in the synchronous matrix and positive in the asynchronous matrix, indicating that the diminution of the component appearing at 580 sec. (probably a C4 carbohydrate), took place before FF degradation. To complete the study we suggest that during the overliming process the chronological modifications occur first of all for the 580 sec. component then xylose, FF and finally the group of peaks appearing at 103, 111, 127 sec.

CONCLUSION

The originality of this work is not only due to the study of the principal DAH components (xylose, FF and HMF), but also in the study of the overliming effects on diluted acid hydrolysis of olive residue using 2DCOS, in order to better understand this process. In this study we have shown that the 580 sec. component varies before xylose, which varies before FF, while FF undergoes degradation to increase the peaks at 103, 111 and 127 sec. The presence of other peaks in the synchronous and asynchronous matrices indicates that other components than furans and xylose are also to be found in the DAH and are involved in the overliming process.

Acknowledgements

The authors would like to thank the Lebanese CNRS (Centre Nationale de Recherche Scientifique Libanais) for their financial support.

Références

1. Nefzaoui A, Dhia MB: *Mise au point sur les expériences réalisées sur l'utilisation des grignons d'olive en alimentation animale. In: L'olivier et autres plantes oléagineuses cultivées en Tunisie: 1978; Tunisie; 1978.*
2. Sivers MV, Zacchi G: *Ethanol from lignocellulosics: a review of the economy. Bioresource Technology 1996, 56:131–140.*
3. Palmqvist E, Hahn-Hagerdal B: *Fermentation of lignocellulosic hydrolysates. II inhibitors and mechanism of inhibition. Bioresource Technology 2000, 74:25-33.*
4. Palmqvist E, Hahn-Hagerdal B: *Fermentation of lignocellulosic hydrolysates. I: inhibition and detoxification. Bioresource Technology 2000, 74(1):17-24.*
5. Larsson S, Reimann A, Nilvebrant N, Jonsson J: *Comparison of different methods for the detoxification of lignocellulose hydrolyzates of spruce. Applied biochemistry and biotechnology 1999, 77-79:91-103.*
6. Mohagheghi A, Ruth M, Schell DJ: *Conditioning hemicellulose hydrolysates for fermentation: Effects of overliming pH on sugar and ethanol yields. Process Biochemistry 2006, 41:1806-1811.*
7. Lewis G, Mathieu D, Phan-Tan-Luu R: *Pharmaceutical Experimental Design, vol. 92. N-Y: Marcel Dekker Inc.; 1999.*
8. Noda I: *Progress in two-dimensional (2D) correlation spectroscopy. Molecular Structure 2006, 799:2-15.*
9. Izawa K, Ogasawara T, Masuda H, Okabayashi H, Noda I: *Two-dimensional correlation gel permeation chromatography (2D correlation GPC) study of the sol-gel polymerisation of octyltriethoxysilane. HCl-concentration dependance. PhysChemComm 2002, 5:12-16.*

Optimisation de la concentration (SPE /SPME) des contaminants des eaux minérales et de leur séparation par GC-FID en utilisant les outils chimométriques

A.Ghanem*, R.Yamine, S. Akl, J.Maalouly et C.O.Saliba**

UMR, Emballages alimentaires, Université Libanaise, Faculté des Sciences II, P.O.Box 90-656 Jdeidet-EI-Matn, LIBAN.
*alghanem@ul.edu.lb; **csaliba56@hotmail.com

Mots-clés : COV, BTEX, phthalates, SPME, GC-FID, analyse

Introduction

Les composés organiques volatils et semi-volatils ainsi que les aldéhydes sont connus pour leur caractère très cancérigène et très néfaste pour la santé humaine. Ces composés sont susceptibles de migrer depuis l'emballage en polyéthylène tétraphthalate (PET) et jusqu'à l'eau minérale embouteillée.

Ce projet consiste à développer une méthode analytique performante pour la concentration et le dosage des composés organiques volatils (COV) tels que le benzène, le toluène, les isomères du xylène : o, p, et m-xylène, le styrène et le chlorobenzène, un aldéhydes tel que le benzaldéhyde et les phthalates (COSV) tels que le diéthylhexylphthalates (DEHP), le phthalate de benzylbutyle (BBP), le di-n-butyle phthalate (DBP), le di-isobutylphthalate (DiBP) et le phthalate de diméthyle (DMP) présents à l'état de traces dans les eaux minérales.

La méthode analytique inclut une étape de concentration, soit par SPE (Solid Phase Extraction) ou par SPME (Solid Phase Microextraction), avant l'analyse instrumentale par GC-FID. L'optimisation de la concentration a été réalisée en ayant recours aux surfaces de réponses.

Le but de cette étude est de mettre au point une méthode d'analyse multi-résidu pour l'ensemble des contaminants cités ci-dessus, en acceptant de faire des compromis quant aux résultats de l'étude des paramètres analytiques. Par suite, lors du choix de la technique de concentration, celle-ci doit présenter les meilleures caractéristiques pour l'ensemble des contaminants: un rendement d'extraction le plus élevé possible pour la gamme la plus large des contaminants, une bonne répétabilité et des limites de quantification au-dessous des MCLs (Maximum Contaminants Levels).

Matériels et Méthodes

Tous les solvants utilisés sont de qualité HPLC. Tous les étalons sont de pureté $\geq 90\%$. L'eau utilisée est deux fois distillée dans nos laboratoires avant son utilisation. L'extraction par SPE a été réalisée une cartouche « SampliQ C18 » (500 mg, 6mL) fournie par Agilent. L'extraction par SMPE a été réalisée sur une fibre 75 μ g carboxen/polydimethylsiloxane fournie de Supelco (Sigma Aldrich). Pour l'analyse chromatographique, une colonne capillaire (Agilent) de type DB-5, 5%phényle et 95% diméthyle polysiloxane (30m \times 0.25mm ; 1 μ m) a été utilisée.

Les solutions mères d'étalons ont été préparées à 500 ppm dans de l'acétonitrile. Les solutions de travail ont été préparées par dilution dans de l'eau distillée, fraîchement avant utilisation. Les échantillons destinés à la SPE ont été dopés avec le mélange des étalons à 250 ppb; alors que les dopages avant SPME étaient de 1 ppb.

Le conditionnement de la phase SPE se fait avec 10ml de méthanol puis 10ml d'eau bidistillée sous vide. Après conditionnement de la cartouche SPE, l'échantillon (V<250 -500ml) a été déposé à un débit de 3mL/min. La cartouche a été ensuite rincée avec 3ml d'eau bidistillée. Un léger séchage est appliqué avant d'appliquer le mélange éluant. L'éluat est ensuite reconcentré par évaporation sous flux d'azote à un débit de 5 bar.

En SPME, avant la première utilisation, la fibre a été conditionnée dans l'injecteur du GC pendant 3 heures à 290 °C. Après chaque injection, la fibre a été reconditionnée pendant 15 min à la température de l'injection. Les expériences d'optimisation des paramètres de la SPME ont été réalisées comme suit : l'échantillon a été placé dans un vial SMPE sous les conditions désirées de pH, de salinité et de volume. L'échantillon est laissé pendant un temps d'équilibre dans le vial à la température d'extraction. Ensuite, la fibre SPME est insérée dans le vial pour extraire les analytes soit en mode immersion ou en mode Head-space (HS) avec des rapports de volumes échantillon/vial variables. L'extraction est laissée, avec ou sans agitation,

pendant un temps allant de 5 à 30 min, avant le retrait de la fibre pour la désorber dans l'injecteur du GC pendant un temps et à une température bien définis.

Plusieurs plans d'expériences ont été utilisés pour chacune des techniques d'extraction. Ainsi en SPE un plan d'expérience basé sur les surfaces de réponses (domaine sphérique de degré 2) a été adopté et a permis d'accéder au mode opératoire optimisé pour cette concentration ; alors qu'en SPME, un criblage (9 facteurs à deux niveaux) a permis d'identifier les facteurs influents avant de passer à la détermination des conditions optimales par un second plan. A la suite de ces analyses, les surfaces des pics sont introduites dans Nemrodw pour déterminer les effets de chaque paramètre.

■ Résultats et Discussion

Une bonne séparation chromatographique des différents contaminants appartenant aux trois familles différentes a été achevée, avec des limites de quantification allant de 0.21 ppm (benzène) à 14.62 ppm (DMP), en utilisant le fluorobenzène et la benzophénone comme étalons internes. Grâce à la concentration SPME, les limites de quantification estimées sont de 0.03 ppb (benzène) jusqu'à 2.03 ppb (benzaldéhyde).

Bien que les phtalates aient pu être analysés simultanément avec les autres contaminants par GC-FID, leur concentration simultanée par SPME n'a pas pu être réalisée sur la fibre testée (75µg carboxen/polydimethylsiloxane). Cela peut revenir à : i) les conditions ne sont pas optimales pour leur extraction, ii) la fibre SPME utilisée n'est pas capable d'extraire ces molécules semi-volatiles. Des réponses ont été toutefois obtenues aux temps de rétention du DMP mais ces réponses étaient également dans les injections de la fibre seule et/ou du blanc.

Suite aux résultats obtenus lors du criblage, l'optimisation de la SPME à réaliser doit tenir compte de tous les facteurs déjà cités (pH, salinité, etc.) surtout du mode d'extraction, du temps et de la température de désorption. En SPE, les conditions optimales ont été déterminées et sont les suivants : volume d'échantillon de 320 mL, volume d'éluat de 14 mL avec un mélange de 34/66 v/v acétonitrile/dichlorométhane. 0.17 mL de n-dodécane doivent être ajoutés lors de l'éluat pour réduire les pertes des COVs par volatilisation. La re-concentration de l'éluat sous flux d'azote permet de réduire à 60 % le volume de ce dernier.

CONCLUSION

Les résultats préliminaires mettent en évidence que la SPME a une meilleure sensibilité que la SPE. Cependant, la fibre utilisée n'est efficace que pour l'extraction des COVs et du benzaldéhyde. Les phtalates ne sont extractibles que sur la cartouche C18 en SPE. Les méthodes impliquant les deux techniques de concentration doivent être validées. Après validation, les performances de la SPME seront comparées à celles de la SPE en termes de limites de quantification et de précision (RSD%) en particulier. La méthode permettant des limites au-dessous des MCLs sera adoptée à condition de fournir des résultats suffisamment répétables.

Apport des outils chimométriques au suivi spectroscopique de la dégradation radiolytique d'une résine échangeuse d'ions

C. Rébufa^a, S. Laborie^a, A. Traboulsi^b, V. Labeled^b, M. Sergent^c et N. Dupuy^a

^a ISM², UMR 6263, Equipe SCC, Université Paul Cézanne, case 451, 13397 Marseille cedex 20, France

^b DEN/DTCD/SPDE, Laboratoire d'Etude de l'Enrobage des Déchets, CEA Marcoule, BP 1717, 130207 Bagnols-sur-Cèze.

^c ISM², UMR 6263, Université Paul Cézanne, case D52, 13397 Marseille cedex 20, France

Mots clés : FTIR, ACP, SIMPLISMA, Plans d'expériences, REI

1- Introduction

Les résines échangeuse d'ions, largement utilisées dans l'industrie du nucléaire, font partie des déchets technologiques faiblement et moyennement radioactifs. Leur dégradation est liée à la présence de radioéléments piégés dans leur squelette polymérique. Une étude a été menée sur une résine anionique dont le squelette est issu d'une polymérisation entre le styrène et le divinylbenzène qui a ensuite été greffé par la triméthylamine afin d'obtenir un sel d'ammonium quaternaire sous la forme ionique OH⁻. L'objectif a été de caractériser les modifications structurales induites par la radiolyse, à partir de l'Infrarouge à Transformée de Fourier (IRTF). Selon cet axe, peu de travaux ont été publiés sur la caractérisation IRTF de ces matrices solides car elles existent sous un grand nombre de structures différentes de part la position du greffon sur les noyaux aromatiques et parce que l'objectif premier est surtout d'estimer leur capacité d'échange après radiolyse. Cependant, à titre qualitatif, Narebska et al. [1] ont utilisé l'infrarouge pour identifier les modifications des groupements chimiques d'une résine greffée en position para avec la triméthylamine (sous une forme ionique chlorée) après différentes doses d'irradiation. Plus récemment, Selemenev [2] et Zagorodni [3] ont publié des tables d'interprétation partielle des bandes spectrales de résines anioniques et cationiques sous différentes formes ioniques. Nous avons pu confronter nos conclusions sur la formation des différents groupements chimiques après irradiation en utilisant les mécanismes réactionnels décrits par Swyler [4] et Alian [5].

Initialement un plan d'expériences a permis de choisir au mieux les différentes irradiations à réaliser pour mettre en évidence les facteurs influents. Les paramètres qui ont été retenus sont la dose absorbée (0,1, 0,5, 1,0 et 4,0 MGy), le débit de dose (énergie absorbée par unité de masse de matériau, sous l'action des rayonnements ionisants par unité de temps et exprimée en Gy.h⁻¹), la nature de l'atmosphère d'irradiation (Aérobie, Anaérobie ou Eau/Anaérobie).

2- Matériel et méthodes

L'échantillon étudié ici est une résine échangeuse d'ions de type anionique dont le squelette polystyrénique est greffé avec la triméthylamine. Elle a été fournie par Rohm & Hass, sous une forme ionique OH⁻ et présente une porosité macroporeuse. Livrée sous forme de microbilles, cette résine a dû être cryobroyée pour être étudiée avec la même granulométrie que celle rencontrée dans ses conditions d'utilisation en lit mélangé. Elle est ensuite conditionnée dans une ampoule scellée puis irradiée sous différentes conditions expérimentales relevant du plan d'expériences (tableau 1).

Les irradiateurs sont tous deux munis d'une source gamma au cobalt (⁶⁰Co). La température de la chambre de l'irradiateur Cigal est régulée à 20°C alors que celle d'Isotron est voisine de 37°C. Les doses intégrées de ces deux irradiateurs sont calculées à partir de mesures de débit de dose. Afin d'obtenir une dose à 4 MGy en aérobie, les ampoules sont maintenues ouvertes jusqu'à une dose intégrée de 3,9 MGy puis sont scellées durant le temps restant.

N° Expérience	N° Echantillons	Milieu d'irradiation	Dose intégrée (MGy)	Débit de dose 2 irradiateurs différents
1	67	Aérobie	0,1	Isotron
	68			
2	69	Anaérobie	0,1	Isotron
	70			
3	71	Aérobie reconditionnée	0,5	Isotron
	72			
4	73	Anaérobie	0,5	Isotron
	74			
5	75	Eau/Anaérobie	0,5	Isotron
	76			
6	77	Aérobie reconditionnée	1	Isotron
	78			
7	79	Anaérobie	1	Isotron
	80			
8	81	Eau/Anaérobie	1	Isotron
	82			
9	83	Aérobie reconditionnée	4	Isotron
	84			
10	85	Anaérobie reconditionnée	4	Isotron
	86			
11	87	Eau/Anaérobie reconditionné	4	Isotron
	88			
12	89	Anaérobie non reconditionnée	4	Isotron
	90			
13	91	Eau/Anaérobie non reconditionnée	4	Isotron
	92			
14	93	Anaérobie sous He à 40°C	0,1	Cigal
	94			
15	95	Anaérobie	0,1	Cigal
	96			
16	97	Anaérobie	1	Cigal
	98			

Tableau 1 - Plan d'expériences

Après irradiation, les gaz sont récupérés puis analysés alors que la matrice solide résiduelle au fond de l'ampoule est séchée à 25° dans une étuve jusqu'à que la masse reste constante. Ce solide est ensuite analysé par les techniques spectroscopiques.

En infrarouge, l'échantillon séché est directement déposé sur le cristal diamant d'un accessoire à réflexion totale atténuée. Trois spectres par échantillon sont enregistrés sur un spectromètre Nicolet IS 10 muni d'un détecteur MCT entre 650 et 4000 cm^{-1} , avec une résolution de 4 cm^{-1} , 100 scans et une vitesse optique de 1,89. Les spectres sont enregistrés à l'aide du logiciel OMNIC (version 8.1).

Le traitement mathématique des données spectrales est effectué en transférant les spectres dans le logiciel UNSCRAMBLER (version 9.2, CAMO) pour réaliser l'Analyse en Composante Principale (ACP) après différents prétraitement des données (élimination de la zone spectrale d'absorption du CO_2 (entre 1900 et 2400 cm^{-1}), correction de la ligne de base et correction de la diffraction des particules grâce à une MSC (Multiplicative Scatter Correction). Puis les données brutes matricielles d'Unscrambler sont exportées dans le logiciel MATLAB (version R2007a) pour réaliser l'analyse par SIMPLISMA (SIMPLe-to-use Interactive Self-modeling Mixture Analysis). Avant extraction des spectres par Simplisma, les données sont dérivées deux fois selon la méthode de Savitzky Golay.

3- Résultats et discussion

En comparant les spectres IRTF du squelette polystyrénique (Fig.1a) et celui de la résine greffée (Fig.1b), on s'aperçoit que la présence des groupements ammonium quaternaires sous leur forme ionique (OH^-) génère des déformations spectrales importantes des bandes du squelette dues aux liaisons hydrogènes. De nouvelles bandes sont aussi apparues, caractérisant la disubstitution des noyaux aromatiques. La superposition des spectres IRTF de tous les échantillons de résines irradiées (Fig.1c) montre des modifications dans différentes zones, qui vont être interprétées en utilisant des outils chimiométriques.

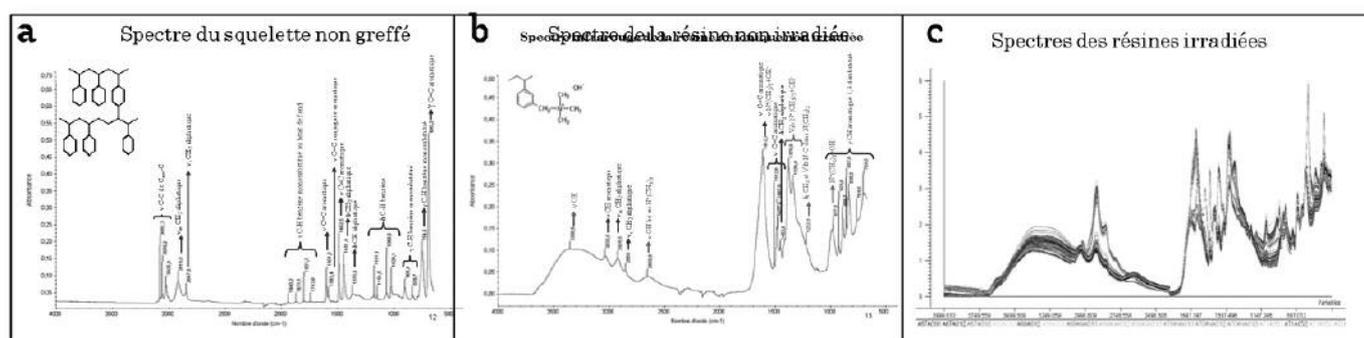


Figure 1 - Spectres FTIR-ATR de squelette polystyrénique (a), de la résine aniotique (b) et de tous les échantillons irradiés

L'Analyse en Composantes Principales des données spectrales (Fig.2) a conduit à des représentations graphiques ne permettant que de différencier les doses à 4 MGy par rapport aux autres doses. La première composante (expliquant 94% de la variance) montre que seule une forte dose semble avoir un impact significatif. Cependant les résines soumises à 4 MGy dans des conditions aérobies (cercle rouge) semblent moins dégradées que celles soumises à la même dose en milieu anaérobie et eau/ anaérobie. La deuxième composante sépare les milieux dans le cas des doses à 4 et 1 MGy. Les spectres des doublons ne montrent pas ici de grandes différences, ce qui laisse supposer une bonne reproductibilité des conditions opératoires. On note aussi que le reconditionnement et le débit de dose (irradiateur Cigal ou Isotron) ne semblent pas influencer le mode de dégradation de la résine.

De part les faibles modifications structurales engendrées aux faibles doses (inférieures ou égales à 1 MGy), seules les doses à 4 MGy seront alors prises en compte dans la présentation des résultats suivants.

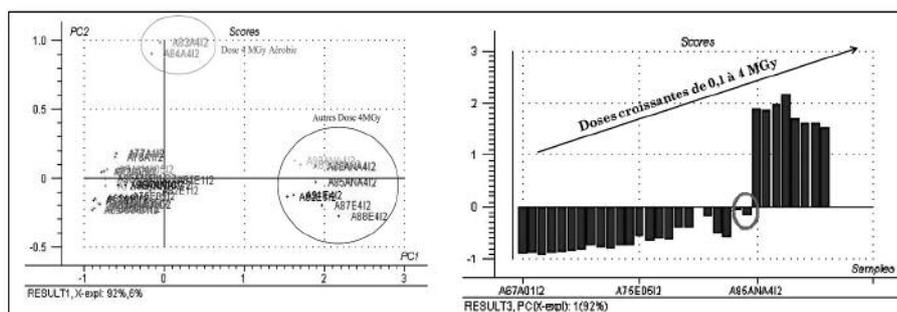


Figure 2 - ACP de tous les échantillons de résines irradiées

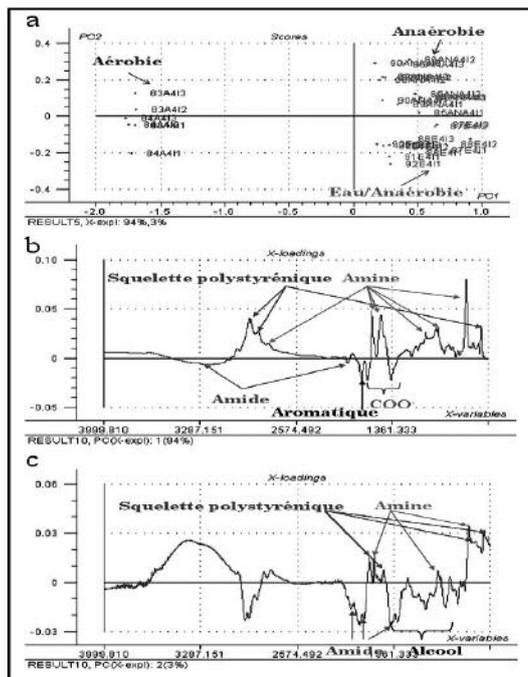
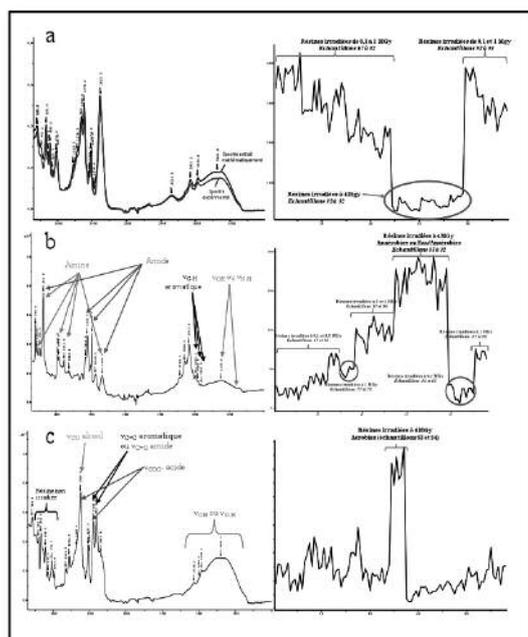


Figure 3 - ACP des résines irradiées à 4 MGy



La distinction des différents milieux est alors bien nette dans le plan (PC1,PC2) de l'ACP des spectres des résines irradiées à 4 MGy (Fig.3a).

D'après la figure 3b, la partie positive de PC1 caractérise les échantillons peu dégradés par la radiolyse (en milieux Anaérobie ou Eau/Anaérobie) qui n'auraient subi qu'une déméthylation des groupements ammonium quaternaire. Une radiolyse en milieu Aérobie semble avoir été plus agressive en donnant naissance à des groupements amide et acide.

La composante 2 montre que dans un milieu Anaérobie, il y aurait formation amines après déméthylation, et voir même coupure du greffon. Alors que la radiolyse dans des conditions Eau/Anaérobie pourrait être à l'origine de la formation de groupements amide et peut-être alcool.

L'utilisation d'une méthode de résolution de courbes sur l'ensemble des données spectrales (toutes doses confondues) nous a permis d'extraire des familles de composés issus de la radiolyse suivant les conditions opératoires. Leurs spectres et leurs profils de concentration associés sont représentés sur la figure 4. Le premier spectre extrait est celui de la résine de départ non irradiée. On retrouve ce profil dans toutes les résines mais sa contribution est la plus faible dans le cas des résines dégradées à 4 MGy. Le deuxième spectre correspondrait à des structures sur lesquelles le greffon aurait subi des transformations en groupements amide et/ou amine. C'est surtout le cas des résines à 4 MGy en milieux Anaérobie et Eau/Anaérobie. Le dernier spectre extrait laisse supposer une attaque de la résine donnant naissance à des groupements amide, alcool et/ou acide dans le cas où les échantillons sont placés dans un milieu Aérobie.

Figure 4 - Spectres extraits mathématiquement par SIMPLISMA et leurs profils de concentration associés

CONCLUSION

L'analyse par infrarouge a permis d'avoir une première approche pour appréhender les modifications structurales et fonctionnelle d'une résine anionique soumise à une radiolyse sous différentes conditions expérimentales. L'interprétation spectrale à l'aide d'outils chimiométrique montre que la dégradation de cet échantillon diffère selon la dose intégrée et le milieu auquel il est soumis en conduisant à la formation de groupements amine, amide, alcool, voire acide dans certains cas. Ces résultats doivent maintenant être confrontés à ceux obtenus après une analyse en microscopie Raman et en RMN du carbone 13.

Références

- [1] A. Narebska & al. *Nukleonika*, 15, 177-187, 1970
- [2] V.F. Selemenev & al. *Reactive & Functional Polymers*, 39, 53-62, 1999.
- [3] A.A. Zagorodni & al. *Reactive & Functional Polymers*, 53, 157-171, 2002.
- [4] K.J. Swyler & al. *NUREG/CR-3383; BNL-NUREG-51691*, 1983.
- [5] A. Alian & al. *Radiat. Phys. Chem.*, 23, 4, 441-444, 1984.

Design of experiments, a useful tool to optimize analytical techniques in view of validation

L Renaudin^{1,2}, L Pierry¹, P Fichet¹, J Vial², A Jardy², D Thiébaud²

¹Laboratoire d'Analyses Nucléaires, Isotopiques et Élémentaires du CEA de SACLAY, DPC/SECR/LANIE, Bât 459, 91191 GIF SUR YVETTE, FRANCE ; laetitia.renaudin@cea.fr

²Laboratoire des Sciences Analytiques, Bioanalytiques et Miniaturisation de l'ESPCI, CNRS – UMR PECSA 7195, 10 Rue Vauquelin, 75231 PARIS CEDEX 05, FRANCE ; jerome.vial@espci.fr

Keywords: ICP-MS, toxic elements, HPLC-ESI-MS, organic complexes

1- Introduction

In the framework of nuclear applications, the validation of analytical techniques is essential. One of the most important parameter that needs to be optimized and thus characterized is the sensitivity of the analytical technique. This parameter is of course directly in relation with the detection limits that must be generally improved. To highlight the interest of chemometric tools, two analytical techniques have been investigated. These different researches correspond to different engagements of the laboratory to analyze toxic elements and organic complexes in nuclear low and intermediate level wastes. These analyses are performed with a liquid coming from a liquid waste or a digestion of solid particularly for toxic investigations.

2- The ICP-MS technique

The ICP-MS (Inductively Coupled Plasma-Mass Spectrometry) technique, devoted to inorganic analysis, was optimized for arsenic measurements. This technique is used for ultra trace analysis whereas ICP-AES (Inductively Coupled Plasma-Atomic Emission Spectrometry) is of particular importance when high salt matrices must be investigated^[1]. Analysis by ICP-MS is particularly appropriate for waste investigations because it allows trace and even ultra trace determinations. However a great number of parameters, i.e. voltages, has to be investigated to provide as sensitive result as possible. It was proven experimentally that the use of experimental designs is more powerful than the standard one to one process proposed by the manufacturer. For ICP-MS a full factorial design of experiments has been processed to optimize four factors known as QRO (Quadrupole Rod Offset), CRO (Cell Rod Offset), CPV (Cell Path Voltage) and AFT (Axial Field Technology). These parameters must be applied to the DRC (Dynamic Reaction Cell) to improve the ion efficiency measurement. The different voltages do not act independently (see Figure 1). This result was proved by the design of experiments that only required 16 tries. The software used to provide the results was NemRodW.

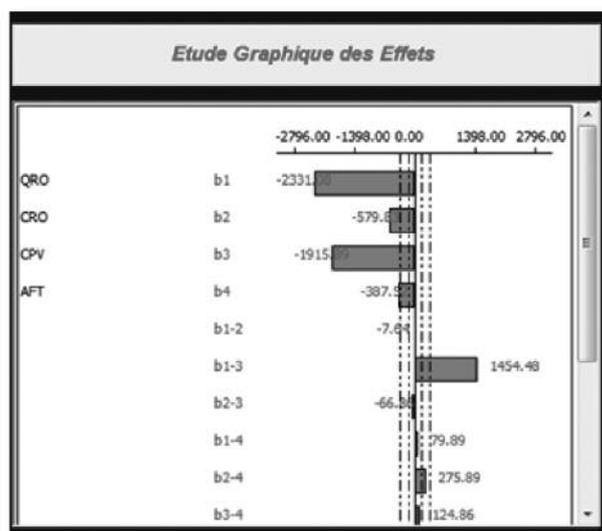


Figure 1 - Effects measurements for ICP-MS signal

3- The HPLC-ESI-MS technique

Another technique devoted to organic analysis, the HPLC-ESI-MS (High Performance Liquid Chromatography-Electrospray Ionization-Mass Spectrometry) has also been optimized by the use of experimental designs. For waste disposals, the French authority wants to control as many parameters as possible. Particularly, some organic complexes, such as EDTA (Ethylene diamine tetraacetic acid) or DTPA (Diethylene triamine pentaacetic acid) are nowadays under investigation by the laboratory. The validation of the analytical technique is essential and requires the previous optimization of numerous parameters. Two different ways of investigation were studied for this analytical technique. The first one concerned the separation process and the second one was focused on the detection process, i.e. the mass spectrometer. For HPLC-ESI-MS two full factorial designs of experiments have been processed to optimize three factors concerning the separation (the composition of the mobile phase and the concentration of tributylamine and acetic acid in the mobile phase) and four factors concerning the detection (the desolvation capillary voltage (CDL voltage), the ESI voltage (interface voltage), the flow of nebulisation and dried gases). The studied signals are the intensities of the Fe-EDTA ($m/z = 344$) and Fe-DTPA ($m/z = 445$) complexes (see Figure 2). Even if the full factorial design requires a large number of experiments it appears to be very efficient to evaluate the sensitivity of the HPLC-ESI-MS. The JMP software [2] provided calculations for the different designs.

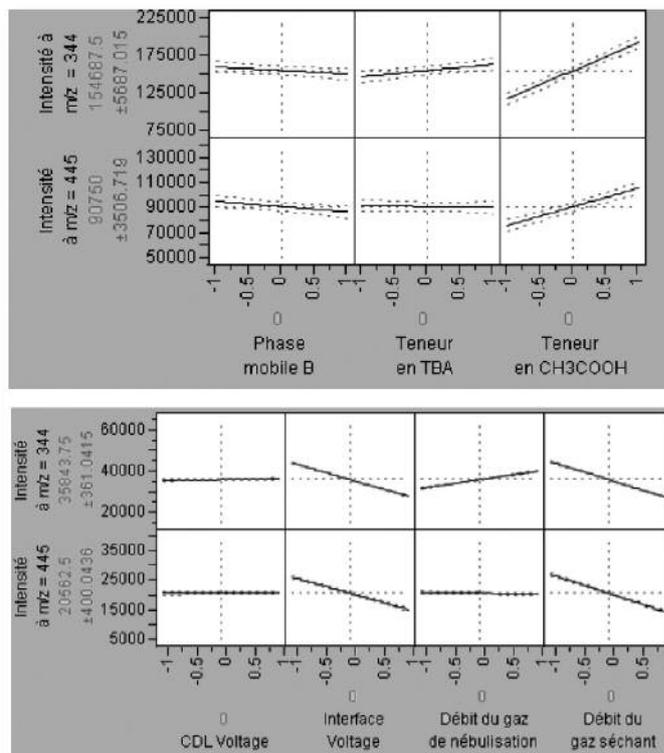


Figure 2 - Influence of the different factors on the studied signals for HPLC-ESI-MS

CONCLUSION

Full factorial designs have been systematically used to validate the analytical techniques. The aim of these developments was to obtain not only the effect of each parameter investigated but also a mathematical model for both analytical techniques. Of course the one by one procedure of researches would be too long for the investigation of all the parameters that are of importance for analytical techniques.

Références

- [1] Uranium Speciation in Drinking water from drill wells in southern Finland and its potential links to health effect, *Environ Sci Technol*, O Pratt, T Vercoeur, E Ansoborlo, P Fichet, P Perret, P Kurtion, L Salonen, 43, 3941-3946, 2009.
- [2] J. Goupy, *Pratiquer les plans d'expériences*, Dunod, 2005.

Chemometric analysis of ultrafast transient absorption spectroscopy data

C. Ruckebusch, N. Mouton² Z. Pawlowska, S. ALoise, O. Devos and M. Sliwa

¹ LASIR CNRS-Université Lille1

Keywords: ICP-MS, toxic elements, HPLC-ESI-MS, organic complexes

Abstract

Time-resolved femtosecond transient absorption spectroscopy data are two-way data containing a large amount of potential information regarding the kinetics and the nature of the photoinduced processes underlying the observable absorption change. It is now a mature technique that enables to measure ultrafast photoinduced processes occurring in the excited electronic states. Nevertheless, straightforward analysis of ultrafast time-resolved spectroscopy data using classical chemometric methods is difficult. The reasons for that are manifold and have to do with the nature of the data ii) the existence of several physical effects affecting the ideal two-way data structure, drifts or curvatures and iome severe interference of artefacts. To these difficulties should be added anotherlast is due to be considered when characterizing short-lived species which is the convolution of the pure profiles with the instrument response function (IRF).

1- Introduction

Femtosecond transient absorption spectroscopy is now a mature technique technique that enables to measure ultrafast photoinduced processes occurring in the excited electronic states. It provides UV-Vis time-resolved difference spectra in the femtosecond to picosecond time domain. These two-way data contain a large amount of potential information on the nature and on the kinetics of the ultrafast photoinduced processes underlying the observable absorption change. Nevertheless, sub-picosecond time-resolved spectroscopy data are difficult to resolve using classical chemometric approach since they may be affected by many potential physical effects, resulting in drifts, curvature of the two-way data, or so-called femtosecond artefacts. Convolution of the pure profiles with the instrumental response function (IRF) should also be considered. Among these multiple effects, some like light dispersion can be corrected by data pre-treatment whereas others require specific chemometric strategies. Regarding this, the proposed approaches are manifold in the litterature.

This presentation aims at detailing the progress made during the last couple of years in the development and applications of chemeomtric field of tandem specific data analysis / chemometric methods to be implemented. to correct for instrumental drifts, to handle femtosecond artefacts and to account for instrument response function. Data analysis thus plays a central role in femtosecond transient absorption spectroscopy in a remiable and meaningful way. Data analysis ùust be considered in the tandem Advanced spectroscopy / Tailored data analysis, including data pre-processing, data investigation and soft-modeliing- Development of hybrid hard and soft models incorporating information linked to the measurement of instrumental response function . and to exemplify them on different systems studied.

The first example will focus on the elucidation of the photophysics of pyridinium betaine derivatives that are compounds possessing unique properties of changing their dipolar moment during photoexcitation.

The second example will show how the proposed strategy helps to unravel detailed knowledge the photochromism of salicylidene anilines (SA)

Both examples are of great relevance for photonics, the development of tailored photoswithable materials, or analytical applications such as high resolution fluoresecence microscopy.

2- Theory

In this work presents, a variant of the hybrid method mixing hard- and soft-modelling into multivariate curve resolution by alternating least squares (HS-MCR)^[1] to take into account the specificities of time-resolved femtosecond spectrokinetic data. HS-MCR combines the benefits of soft-modelling to recover the pure spectral component and time dependent

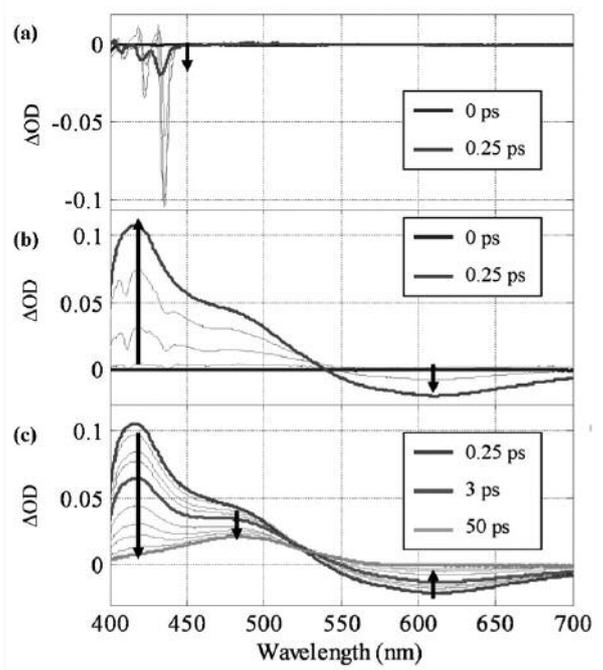
concentration profiles with the ones of hard-modelling to fit kinetic models and provide rationalized description of the processes investigated.

To overcome the influence of the specific problems linked to transient absorption spectroscopy, HS-MCR is modified so that convolution of the calculated time-dependent concentration profiles by the IRF can be considered in the hard-modelling step.^[2] The time-dependent model to fit is thus defined by convoluting the pure kinetic model with a Gaussian parametric description of the IRF. Furthermore, femtosecond artefacts, which can be independently measured in the solvent, are introduced in the process analysis via a multi-experiment data analysis scheme, where model-free (solvent experiment) and model-based (process experiment) are analysed together.

3- Data

The performance of the method will be validated on synthetic data and illustrated on experimental data with the characterization of photoinduced proton transfer, a fundamental process found in nature and new organic materials occurring at femtosecond time-scale.

Figure 1 - Transient absorption spectra of (a) acetonitrile between 0 and 250 fs (solvent) and of SA in acetonitrile observed (b) between 0 and 250 fs and (c) between 250 fs and 50 ps.



4- Results and discussion

The evolution of the time-dependent profiles in Figure 2 shows four consecutive species, which will be called A, B, C and D following the emerging sequence. Contributions B and C present transient spectra with a large absorption band showing a maximum near 420 nm with a shoulder between 460 and 550 nm and a negative band with a maximum at 610 nm, related to stimulated emission. According to the knowledge of the photochemistry of SA, contributions B and C could be associated with the excited cis-keto form of SA. The interpretation is detailed in ref^[3].

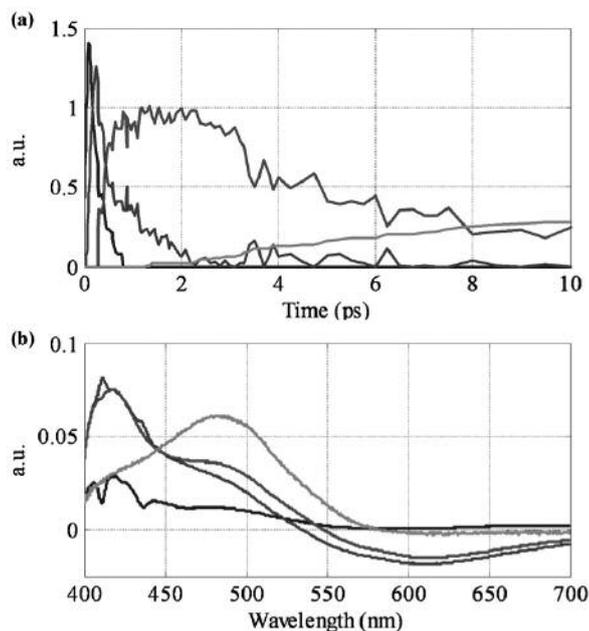


Figure 2 - Results obtained by MCR-ALS for the four contributions corresponding to the chemical process performed on column-wise augmented data: (a) time-dependent concentration profiles and (b) associated spectra. The blue, green, red and cyan profiles are associated to the contributions A, B, C and D respectively.

CONCLUSION

The distortion induced by data convolution with the IRF in femtosecond transient absorption spectroscopy can be treated in a hybrid method mixing hard- and soft-modeling HS-MCR. For this purpose, the kinetic model used in the hard-modeling step of the procedure is adapted by convolution with a Gaussian function describing the IRF.

The method enabled to recover true rate constants of photoinduced processes occurring below the instrumental time resolution. It was demonstrated on experimental data probing the photochromism of SA in the fs-ps time scale. A photo-dynamical scheme for SA was drawn with a cascade of first-order kinetic process which is in agreement with published recent results: more specifically, a 82 fs characteristic time was recovered for the excited-state intramolecular proton transfer using a femtosecond transient absorption spectroscopy setup, for which the time resolution was measured to be 150 fs.

Acknowledgment

The authors acknowledge funding in the frame of a PICS international program for scientific research between the French Centre National de la Recherche Scientifique (CNRS) and the Spanish Consejo Superior de Investigaciones Científicas (CSIC). N. M. acknowledges a PhD scholarship from the Ministère de l'Enseignement Supérieur et de la Recherche and an international research grant from the Région Nord- Pas de Calais. A .J. acknowledges the Spanish government by financial help linked to the project CTQ2009-11572.

Références

- [1] A. de Juan, M. Maeder, M. Martinez, R. Tauler, *Combining hard- and soft-modelling to solve kinetic problems*, *Chemometrics Intell. Lab. Syst.*, 54 (2000) 123-141.
- [2] N. Mouton, M. Sliwa, A. de Juan, C. Ruckebusch, *Hybrid hard- and soft-modeling approach for the resolution of convoluted femtosecond spectrokinetic data to be published in Chemometrics Intell. Lab. Syst.*
- [3] M. Sliwa, N. Mouton, C. Ruckebusch, L. Poisson, A. Idrissi, S. Aloise, L. Potier, O. Poizat, G. Buntinx, *Investigation of ultrafast photoinduced processes for saliylidene aniline in solution and gas phase: toward a general photo-dynamical scheme.*, *Photochem. Photobiol. Sci.*, 9 (2010) 661-669.