

**Dr Jean-Michel Roger :**

Cemagref, UMR ITAP

**Développements chimiométriques pour améliorer la robustesse des mesures spectrométriques appliquées aux agro-procédés.**

Pour répondre aux besoins de l'agriculture de précision, le Cemagref développe des capteurs, notamment à base de spectrométrie infrarouge. Une analyse théorique rapide de ce principe de mesure montre qu'un des problèmes majeurs de ces capteurs a trait à la robustesse des modèles d'étalonnage. Une analyse détaillée de cette problématique nous a conduits à isoler trois voies de recherche :

1/ La prise en compte des grandeurs d'influence. Selon qu'une grandeur d'influence, responsable d'un problème de robustesse, est mesurable ou non, différentes stratégies de correction du modèle d'étalonnage sont envisageables. Les grandes familles de méthodes ont été identifiées et comparées dans le cas de l'effet de la température sur la mesure du taux de sucres des pommes. Nous avons développé une méthode originale de correction de l'étalonnage : L'EPO (pour External Parameter Orthogonalisation) propose d'ôter de l'espace de mesure la partie influencée par la grandeur d'influence, au moyen d'une projection orthogonale. Cette opération, réalisée à partir de quelques mesures simples, permet de traiter des bases d'étalonnage existantes.

2/ La maintenance de la robustesse. Les problèmes de robustesse interviennent lors de l'utilisation répétée des capteurs. Cette maintenance est traditionnellement réalisée par ré-étalonnage ou transfert d'étalonnage. Dans les deux cas, la mesure d'échantillons "standard" est nécessaire. Nous avons développé une méthode de correction d'étalonnage en ligne, qui ne nécessite pas ces standards : DOP (pour Dynamic Orthogonal Projection) utilise le principe de projection orthogonale de l'EPO, en identifiant le sous espace des perturbations à partir de la connaissance de quelques points de recalage.

3/ La discrimination à partir des spectres. Cette opération, qui permet de réaliser des diagnostics à partir de la mesure spectrale, est nécessaire pour : détecter les mesures aberrantes, reconnaître des classes d'objets, réaliser des régressions locales, etc. Nous proposons une nouvelle méthode de discrimination, basée sur le parcours de fonctions numériques, nommées Fonctions Propres Focales. Certaines propriétés de ces fonctions permettent de parcourir une série d'espaces au pouvoir discriminant croissant. Des algorithmes de parcours classiques, avec arrêt d'apprentissage, peuvent ainsi être mis en oeuvre.

[Retour au programme](#)

---

**Pr. Douglas N. Rutledge**

Laboratoire de Chimie Analytique, UMR Ingénierie Analytique pour la Qualité des Aliments – IAQA, INRA / INA P-G

**OPA-Parafac**

Outer-product analysis (OPA) was introduced by Barros and Rutledge (1) as a method to detect co-evolutions of spectral regions in two signals. The calculation of the outer-product (OP) matrix from two signal-vectors links the two domains as a mutual weighting of the two

signals. Many statistical techniques – both univariate and multivariate – can be applied to these matrices to highlight simultaneous variations. Until now, the rectangular matrices resulting from the Outer Product for each sample were unfolded and concatenated row-wise before being statistically analysed. The results of the analyses (Loadings, B-coefficients ...) were then folded back to give again rectangular matrices that were then interpreted to reveal simultaneous variations in the two signal domains.

In this paper, a multi-way statistical analysis techniques, PARAFAC, is directly applied to the « cubic » set of rectangular OP matrices (i.e. an OP tensor). It will be shown that in this way, it is possible to determine directly the number of species contributing to the two signals, their evolution and the corresponding « pure » signals.

In this paper, the PARAFAC analysis of the Outer Products (OPA-PARAFAC) is used :

1) to characterise the retrogradation of starch by studying the OP matrices corresponding to two-by-two combinations of the following three types of signals - Time Domain Nuclear Magnetic Resonance (TD-NMR), Mid-InfraRed spectroscopy (MIR) and X-ray diffraction (XRD). It is shown that during the process of retrogradation, the starch evolves from an amorphous state to a semi-crystalline state and OPA-PARAFAC reveals how this transformation affects these three signals.

2) to determine the evolution of Near Infrared (NIR) spectra of water (1300–1600 nm) at temperatures between 6 and 80 °C. The different loadings profiles obtained by applying PARAFAC to the OP tensor calculated between the pairs of vectors corresponding to the same NIR spectrum show how temperature modifies O-H bond strengths and vibrations. Particular bands are shown to be responsible for most of the spectral variations and correspond to weakly and strongly hydrogen-bonded water molecules.

3) to demonstrate how the effect of increasing fructose concentration (5, 10, 20, 30, 40, 50 and 60%) on Carr–Purcell–Meiboom–Gill relaxation curves acquired by TD-NMR can be related to particular NIR spectral features and help understand which NIR peaks are related to weakly and strongly hydrogen-bonded sugar molecules.

4) to extract the « pure » MIR and NIR spectra of the active principle of a pharmaceutical product from the study of a set of pills at different concentrations.

5) to determine the stoichiometry of a complexation reaction between tetraphenylporphine (TPP) and copper by combining UV-Visible spectra and TD-NMR relaxation signals. The complexation is studied using Job's Continuous Variations Method. The multi-way analysis using PARAFAC was able to extract the pure UV-Visible spectra of both the TPP and the TPP-Copper complex, as well as the corresponding « pure » relaxation curves.

1) D.N.Rutledge, A.S. Barros

"A method for detecting information in signals: application to 2-D Time Domain - NMR data"

The Analyst, (1998) 123, 551-559

[Retour au programme](#)

---

**Pr. Riccardo Leardi**

Department of Pharmaceutical and Food Chemistry and Technology – via Brigata Salerno (ponte) – I 16147 GENOVA (ITALY)

**Les Algorithmes Génétiques pour la sélection de régions spectrales.**

Genetic Algorithms (GA) are very useful in solving complex problems of optimisation, and the selection of a subset of variables is surely one of them.

After suitable modifications, genetic algorithms can be a very useful tool in the problem of wavelength selection in the case of a multivariate calibration performed by PLS. Unlike what happens with the majority of feature selection methods applied to spectral data, the variables selected by the algorithm often correspond to well-defined and characteristic spectral regions instead of being single variables scattered throughout the spectrum. This algorithm has been applied to several data sets, and the obtained results were at least as good as those obtained by human experts and better than those obtained by a commercial software embedded in a very well-known series of FT-IR instruments.

The main problem of GA is the risk of overfitting. This risk increases with the number of input variables, and it has been found that the upper limit at which it can be safely used is 200 variables. When dealing with spectra, the number of wavelengths is usually by far exceeding these values. To solve this problem, a “windowing” process is applied, by which a lower number of variables is obtained, each one being the average of  $n$  contiguous wavelengths,  $n$  being chosen in such a way that no more than 200 variables are obtained. This procedure can be safely applied to NIR data, in which the spectral features are very broad, but it can give some problems when applied to FT-IR data, on which the windowing can destroy some very narrow peaks. This problem is solved by applying an “iterative” approach, in which at each time the least relevant regions of the spectrum are removed; this leads to a lower number of retained wavelengths and therefore to a lower window size ( $n$  becomes smaller). This approach produces very good results, but it has the great disadvantage of being extremely time-consuming.

A solution to this problem has been found by combining a variation of the iPLS (interval PLS) technique with GA-PLS. When used in its backward form, instead of in its forward form as originally described in literature, iPLS can be a very good method to remove the most non-informative regions. As a result, a number of wavelengths will be retained that will be acceptable for GA. The results obtained with this approach are totally comparable with the results obtained with the iterative approach, but the computing time is much smaller, since in any case only one step of GA-PLS will be applied.

The Matlab software for GA-PLS and for backward iPLS can be freely downloaded from the site of KVL (<http://www.models.kvl.dk/source/GAPLS/index.asp> and <http://www.models.kvl.dk/source/iToolbox/index.asp>)

[Retour au programme](#)

---

### **Dr. Thomas Verron**

Centre d'Ecologie Fonctionnelle et Evolutive, CNRS

- Laboratoire de Physique Moléculaire et Structurale, Faculté de Pharmacie Université Montpellier I

### **Prétraitement et utilisation d'ondelettes pour l'amélioration des modèles PLS appliqués aux spectres NIRS**

Le prétraitement des spectres Proche Infrarouge est une étape indispensable à la modélisation d'une des caractéristiques de l'échantillon analysé. Jusqu'à la fin des années 90, les objectifs

des méthodes de correction et de calibration été traités séparément. Afin de rendre les méthodes de prétraitement plus ciblées et spécifiques à l'analyse de la variable réponse, les chimométriciens ont développé des méthodes de corrections qui prennent en compte l'information de la variable à prédire. Aujourd'hui nous proposons de passer un palier supplémentaire en intégrant la méthode de correction à la méthode de calibration, pour obtenir des corrections qui s'auto adaptent et évoluent avec le critère de modélisation à maximiser. Nous nous sommes consacrés au développement de méthodes de type PLS, basées sur l'optimisation d'un critère sous contraintes, par étape, pour réaliser simultanément la correction et la modélisation de spectres PIR : SiROPLS (Signal Reconstruction Optimized by PLS). La méthode SiROPLS a pour objectif de séparer l'information pertinente de l'information inutile en utilisant des méthodes du traitement du signal (ondelettes et transformée de Fourier), et de reconstruire des spectres les plus proches possible des spectres "idéaux" pour la modélisation de la cible.

[Retour au programme](#)

---

**Dr. Nathalie Dupuy**

UMR 6171 Systèmes chimiques complexes. Université Paul Cézanne 13397 Marseille Cedex 20

**Reconnaissance par spectroscopie proche infrarouge de l'origine d'huiles d'olive vierges françaises**

N Dupuy(1), O. Galtier(1), D. Ollivier(2), Y. Le Dreau(1), J. Artaud(3)

1Laboratoire GOAE, Université Paul Cézanne, UMR CNRS 6171, case 561, 13397 Marseille cedex 20

2Laboratoire de Marseille, Direction Générale de la Concurrence, de la Consommation et de la Répression des Fraudes, 146 traverse Charles-Susini, 13388 Marseille Cedex 13

3 Laboratoire de Chimie Analytique de l'Environnement, Université Paul Cézanne, UMR CNRS 6171, IFR PMSE 112, BP80, Aix-Marseille III.

La France est un pays producteur modeste d'huiles d'olive vierges (HOV) avec 2.500 à 3.500 T annuelle soit environ 0,15 % de la production mondiale. L'intérêt des consommateurs pour le régime méditerranéen, la forte médiatisation des bienfaits de l'huile d'olive sur la santé ont créé un renouveau de l'oléiculture française qui s'est traduit ces dernières années par la création de cinq appellations d'origine contrôlée (AOC). Le prix élevé des HOV, la demande croissante de traçabilité des produits nécessitent de disposer de méthodes d'analyses rapides, fiables et sensibles permettant de répondre à la fois aux besoins de la profession oléicole et à une demande des consommateurs. La mise en place de systèmes de traçabilité sur tout le procédé de fabrication est un impératif depuis les olives entières à l'entrée dans les moulins jusqu'au produit fini en sortie.

Le but de ce travail est de réaliser un traitement statistique des spectres proche infrarouge d'huiles d'olive vierges de plusieurs AOC différentes par une analyse discriminante PLS afin de caractériser et de déterminer l'origine des différentes d'huiles d'olive vierges. 123 huiles d'olive vierges provenant de cinq AOC de la région PACA ont été analysées par spectroscopie proche infrarouge. L'analyse en prédiction d'origine donne des résultats très satisfaisants (100% de reconnaissance) permettant d'envisager d'utiliser cette technique pour déterminer rapidement l'origine d'huiles d'olive vierges.

**Dr. Fabien Chauchard**

Ondalys

**Pourquoi rajouter une nouvelle dimension à la spectroscopie proche infrarouge ?**

**Time-Resolved ou Spatially-Resolved Spectroscopy.**

La spectroscopie proche infrarouge est reconnue depuis plusieurs années comme méthode permettant l'évaluation de la qualité interne des fruits et légumes. La prédiction de la qualité se fait grâce à des modèles multivariés basés sur les spectres enregistrés et souvent en utilisant la mesure sans contact par réflexion (nombreux avantages). Une des principales contraintes liée à l'exploitation industrielle est l'obtention de modèles de prédiction qui gardent leur capacité prédictive malgré les variations de la mesure. La diffusion de la lumière liée à la structure complexe des échantillons biologiques modifie les spectres infrarouge. Cette modification se répercute sur la qualité des modèles de prédiction. L'effet diffusion ne pouvant être correctement pris en compte par la chimiométrie, une nouvelle voie de recherche a été suivie. A partir de nouvelles méthodes de mise en oeuvre de la spectroscopie proche infrarouge il a été possible de séparer et d'évaluer les interactions chimiques (absorption) des interactions physiques (diffusion). La spectroscopie résolue temporellement (avec supercontinuum de lumière), TRS, a permis de mesurer les coefficients de diffusion et d'absorption de différents échantillons diffusants. Plusieurs méthodes de calcul de ces coefficients basés sur des signaux temporels ont été mises au point. La TRS étant une méthode surtout dédiée au laboratoire, nos efforts ont ensuite porté sur l'utilisation de la spectroscopie résolue spatialement. Dans ce cas, le concept de réinjection de la lumière a été validé à l'aide de simulations Monte Carlo avec pour objectif la réalisation d'un capteur à bas coût permettant l'évaluation de la diffusion et de l'absorption.