XVI CHEMOMETRICS
IN ANALYTICAL CHEMISTRY

JUNE, 6-10, 2016
BARCELONA, SPAIN

BOOK OF ABSTRACTS

INDEX

INVITED SPEAKERS:

ABSTRACTS:

INVITED SPEAKERS

PLENARY SPEAKERS

UNCOMMON PENALTIES FOR COMMON PROBLEMS

Paul H. C. Eilers[1]

[1]*Department of Biostatistics, Erasmus University Medical Centre, Rotterdam, the Netherlands*
p.eilers@erasmusmc,nl

Penalties are a familiar tool in Chemometrics, going back to the early adoption of ridge regression. Penalties to achieve smoothness are also finding application in many places. My presentation is about less common penalties which can help to solve common problems. The main themes are: asymmetric penalties to enforce shape constraints, different norms to achieve sparseness and adapting ideas from differential equations. Penalties can also be used to split a signal into multiple components with desirable properties. I will present many applications to real data, to show that by creative use of penalties we can solve many practical problems.

CHEMOMETRICS IN CHEMISTRY
–
CHEMISTRY IN CHEMOMETRICS

Marcel Maeder

University of Newcastle, Newcastle, Australia
marcel.maeder@newcastle.edu.au

Many chemists are not well aware of the powerful data analysis methods that are available in chemometrics. They treat chemomemtrics methods with suspicion and hesitate to apply them. There are, of course, many exceptions. One example of very successful chemometrics in the real world is the determination of the alcohol content in wine by NIR-PLS, which is almost universally performed in the commercial wine laboratory as it is much faster than alternative redox titrations or GC.

Correspondingly, many chemometricians tend to treat their data as abstract entities, they do not introduce into the analyses any chemical knowledge about the system under investigation. Of course there are exceptions here too, think of the most commonly and very widely applied data analysis method in chemistry which is data fitting. It relies on a mathematical relationship that is based on the chemistry of the system under analysis.

This presentation attempts to demonstrate that chemometricians can improve their analyses by including the fact that they are dealing with chemical data and that chemists can profit from accepting chemometrics methods as valuable tools.

The following example will be discussed to illustrate some of the issues, it goes back to Jean Thomas Clerc, former editor of the chemometrics section of Analytica Chimica Acta: Find the reason or an algorithm that rationalises the following sequence.

5, 10, 2, 9, 4, 7, 6, 3, 1, 8, 0

Algorithms like neural networks or adaptations of QSAR or of any other nature can be developed to organise the numbers (‘sort’ is not useful as it does not give information about the reason for the order). The author will accept elegant, interesting or any other suggestions for solutions and will discuss them in the presentation.

RECENT ADVANCES IN THE ESTIMATION OF MULTIVARIATE/MULTIWAY ANALYTICAL FIGURES OF MERIT

A. C. Olivieri

Departmento de Química Analítica, Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Instituto de Química Rosario (IQUIR-CONICET), Suipacha 531, Rosario (2000), Argentina, email: olivieri@iquir-conicet.gov.ar

The estimation of analytical figures of merit in multivariate calibration has become an active research field in analytical chemistry. Some recent developments show the continuous interest in this area by the analytical community [1-4]. In particular, the derivation of important expressions have been possible: (1) for computing the sensitivity parameter in multivariate and multiway calibration scenarios [1], (2) for estimating the prediction uncertainty in first-order calibration in the presence of generalized noise structures [5], and (3) for establishing the detection limit in partial least-squares calibration in a IUPAC consistent manner [6].

In this presentation, the latter developments will be discussed in detail, including some new findings concerning the following issues:

1. The extension of the prediction uncertainty expression from first-order to multiway calibrations with data of any number of instrumental modes.
2. The estimation of detection limits in multiway calibrations.
3. The proposal of a generalized analytical sensitivity parameter, useful for the comparison of first-order calibration methdologies in the presence of different noise structures.
4. The extension of the latter parameter to multiway calibrations.

The availability of reliable figures of merit will allow for a proper comparison of the prediction ability of different analytical protocols and multivariate/multiway data processing algorithms.

Acknowledgement: Universidad Nacional de Rosario, CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas, Project No. PIP 0163), ANPCyT (Agencia Nacional de Promoción Científica y Tecnológica, Project No. PICT-2013-0136), are gratefully acknowledged for financial support.

References:

[1] A.C. Olivieri, N.M. Faber, J. Ferré, R. Boqué, J.H. Kalivas, H. Mark, Pure Appl. Chem., 78, 633 (2006).
[2] A.C. Olivieri, Chem. Rev., 114, 5358 (2014).
[3] A.C. Olivieri, G.M. Escandar, Practical three-way calibration, Elsevier, Amsterdam, 2014, Chapter 6.
[4] A.C. Olivieri, S. Bortolato, F. Allegrini, Figures of merit in multiway calibration, in Data Handling in Science and Technology, Vol. 29, Fundamentals and analytical applications of multiway calibration, A. Muñoz de la Peña, H.C. Goicoechea, G.M. Escandar, A.C. Olivieri, Editors, Elsevier, Amsterdam, 2015, Chapter 13.
[5] F. Allegrini, P.D. Wentzell, A.C. Olivieri, Anal. Chim. Acta, 903, 51 (2016).
[6] F.A. Allegrini, A.C. Olivieri, Anal. Chem., 86, 7858 (2014).

TOPOLOGICAL INSIGHT INTO CHEMICAL DATA: GENTLE INTRODUCTION

Beata Walczak

Institute of Chemistry, University of Silesia, Katowice, Poland

Topological Data Analysis (TDA) is a relatively new field, and its main goal is to study the shape of the data, which may reveal important information about a system or a phenomenon considered (e.g., [1]). Topology deals with geometrical properties of space, which are preserved under continuous deformations such, as bending and stretching (but not tearing, or gluing). It is mainly occupied counting basic topological features such, as components and holes. TDA already found very attractive applications. It can be useful in data exploration, clustering of numerous samples, and comparing different platforms. The main aim of this presentation is to introduce and describe basic terms and concepts of TDA to chemometric community. Such terms as, e.g., topological space, its approximation and properties, persistency of topological features, lenses and graphs, will be gently introduced and the main steps of TDA will be demonstrated on simulated and experimental data sets.

References:

[1] G. Carlsson, Topology and data, Bulletin of The American Mathematical Society 46, 255 (2009)

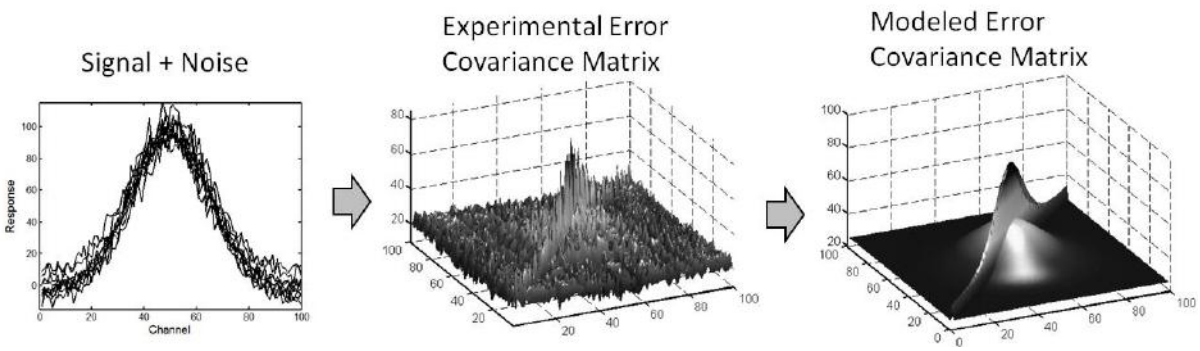
KNOWN UNKNOWNS AND UNKNOWN UNKNOWNS: UNRAVELLING MULTIVARIATE MEASUREMENT ERRORS

Peter D. Wentzell [1]

[1]*Department of Chemistry, Dalhousie University, PO Box 15000, Halifax, Nova Scotia, B3H 4R2, Canada*
peter.wentzell@dal.ca

The separation of "interesting" chemical variance from other sources of variance in a data set is the principal objective of most chemometric methods, whether they relate to exploratory analysis, classification, calibration, curve resolution, or other tools. The "other" sources of variance are generally lumped together as "noise", but relatively little attention has been paid to the characterization of these sources of variance, even though this is central to the design of optimal data analysis tools. Chemometric methods are typically based on the implicit assumption that such noise is dominated by instrumental components that are independently and identically distributed with a normal distribution (*iid* normal), even though heteroscedastic and correlated noise structures are the norm rather than the exception. Structured noise, such as chemical noise arising from the presence of secondary components, is often incorporated into the model in a deterministic way, but this distinction becomes blurred in complex mixtures with a large number of components. The *ad hoc* implementation of many preprocessing methods (*e.g.* scaling, filtering, multiplicative signal correction) is symptomatic of a limited understanding of these sources of variance. A more complete characterization of noise is inhibited by a number of factors that include: (a) a lack of understanding of what constitutes a replicate, (b) insufficient data for characterizing measurement errors, (c) limited tools for the development of measurement error models, and (d) a dearth of chemometric tools that incorporate measurement error information. There is an interdependence of these factors in that a lack of measurement error information impedes the development of tools that use this information, and the lack of tools discourages the acquisition of this information.

The objective of this presentation is to describe some of the advances and challenges in the study of multivariate measurement noise. This includes: (1) the definition of what constitutes measurement noise, (2) the experimental challenges of noise characterization, (3) advances in the development of measurement error models, (4) current chemometric approaches for implementing measurement error information, and (5) the nature of chemical noise in complex chemical mixtures.



KEYNOTE SPEAKERS

HYPERSPECTRAL IMAGE ANALYSIS. WHERE ARE WE AND WHERE SHOULD WE GO?

José Manuel Amigo

Department of Food Sciences, University of Copenhagen, Rolighedsvej 30, Frederiksberg-C, Denmark
jmar@life.ku.dk

Undoubtedly, the implementation and application of Chemometrics in the analysis of Hyperspectral Images (HSI) plays a major role in nowadays setups. The versatility of Chemometrics makes possible the finding of tailored adaptations of the analysis depending on the very diverse uses of HSI. In this presentation we will study the current state-of-the-art of the duet "Chemometrics-HSI" with special emphasis in Food, Pharmaceutical and Forensic sciences. Moreover, we will envision the possible improvements to be done in order to actually merge "Chemometrics-HSI" with digital image analysis.

MULTIVARIATE BIG DATA ANALYSIS AND ITS APPLICATION TO THE INTERNET

José Camacho Ph.D.

This speech will be devoted to discuss the application of chemometric (data analysis) methods in Big Data problems. A freeware Matlab toolbox, the MEDA Toolbox, will be introduced for this purpose. Its use will be exemplified with one of the main Big Data applications we refer to as Networkmetrics: the analysis of traffic data in Computer Networks (e.g. the Internet) Main challenges in Networkmetrics, in particular the data parameterization for multivariate analysis, will be discussed.

Acknowledgement: This work is partly supported by the Spanish Ministry of Economy and Competitiveness and FEDER funds through project TIN2014-60346-R.

MULTICOLOUR FLOW CYTOMETRY: A NEW CHALLENGE FOR CHEMOMETRICS

Jeroen Jansen[1], Gerjen Tinnevelt[1] , Rita Folcarelli[1] , Bart Hilvering [2], Selma van Staveren[2] , Oscar van den Brink [3] , Leo Koenderman[2] , Lutgarde Buydens[1]

1 *Analytical Chemistry and Chemometrics, IMM, Radboud Universiteit, Nijmegen, The Netherlands*
2 Respiratory Medicine, Laboratory of Translational Immunology (LTI), University Medical Center Utrecht, Utrecht, The Netherlands
3 TI-COAST, Science Park, 904, Amsterdam, The Netherlands

email: jj.jansen@science.ru.nl

Multicolour Flow Cytometry (MFC) is a biomedical technology in strong development. With MFC, single cells can be fluorescently marked and analysed for several specific immunological characteristics. The resulting data then contains the diversity in a large complement of cells within a specific sample (*e.g.* blood). This diversity can then be compared between samples for phenotypic biomarkers of disease and other immunological responses. The recent developments throughout all sub-disciplines of chemometrics may be highly beneficial for this comparison, specifically to solve the currently existing challenges in MFC technology.

Firstly, immune responses may only affect a very limited number of cells, whilst the majority of cells within the sample remain indistinguishable from healthy samples. Therefore, we have developed a method based on Multivariate Statistical Process Monitoring to pinpoint the response-related cells within these samples, such that they can be isolated for further study. Secondly, MFC quantifies the expression of immunological characteristics yet conclusions need to be drawn on the full diversity of cells in the same sample. To allow this, we have developed a multiset classification method that allows the direct association of an immune response to changes in expression of these characteristics on a single cell. Thirdly, current laser technology limits the number of different characteristics that can be measured on the same cell to seven; the latest developments allow up to 17 different characteristics. This is still far from a true 'omics' methodology, although more than 300 characteristics may be relevant in a comprehensive immune response. To broaden the scope of MFC technology, we have developed a data fusion strategy based on the aforementioned methods. This allows the information from different characteristics measured on the same cell to be merged into one model.

These novel methods show how chemometric developments may support MFC technology, but also how chemometric methodology may be brought forward by novel exciting application areas such as MFC.

TAKING A BIG DATA APPROACH TO LOCAL SPECTRAL CALIBRATION

J.H. Kalivas [1], **R. Emerson** [2]

[1]*Department of Chemistry, Idaho State University, 921 S. 8th, Stop 8026, Pocatello, Idaho, USA*
[2]*Idaho National Laboratory, Idaho Falls, Idaho, USA*
email (kalijohn@isu.edu)

A large global calibration set of samples (a library) spanning numerous measurement and instrument conditions often spans too much information variance and hence, prediction accuracies are not sufficient for many applications. Local modeling techniques can be used to select a subset of calibration samples from a global library specific to each new individual prediction sample in order to build models meeting accuracy requirements. Few local modeling methods use more than one spectral similarity measure at a time to compare a prediction sample spectrum to library spectra, e.g., only the distance or angle between spectral vectors for the subset selection. This work explores methods to simultaneously evaluate multiple similarity measures through data fusion in order to select calibration samples spectrally resembling each individual prediction sample. Another aspect of local modeling explored in this work is the use of "y-windowing", e.g., concentration window, to further narrow the calibration domain used in the model. The y-windowing allows sorting of the selected calibration subset samples into local ranges of the prediction property of interest. Models are then formed using these subsets or windows. Fusion rules are now applied to multiple calibration model quality measures to identify the calibration window(s) most consistent for the prediction sample. Calibration windows are further verified as best by cross-referencing calibration and predictions with other calibration windows including fusion with moving wavelength windows. The calibration y-window(s) deemed best is used as the final calibration set to form the local model. The process is termed local adaptive fusion regression (LAFR). It is local for selecting a library subset similar to each new sample, adaptive because the process adapts to different prediction properties for each new sample using the same spectrally selected library subset, and fusion is the foundation of the spectral and y-windowing analysis. While the approach presented does not utilize Big Data (petabytes to yottabytes and growing), it does embody Big Data concepts. Specifically, the library size and the number of similarity measures fused for selecting samples from the library are not limited. Likewise, the number of measures, wavelength windows, and y-windows fused to identify the final local calibration subset are not limited. Limiting the process is computer memory and speed of computations. These two aspects are becoming less restrictive and the analytical chemist needs to be thinking Big Data. Specifically, new processes to create Big Data and mine the data for the task at hand.

CLASS-MODELLING IN FOOD ANALYTICAL CHEMISTRY: WHY IT SHOULD BE OUR PREFERRED CHOICE AND WHY IT IS NOT

P. Oliveri[1]

[1]*Department of Pharmacy, University of Genoa, Via Brigata Salerno 13, Genoa, Italy*
oliveri@dictfa.unige.it

‘‘Is the product under examination compatible with the declared claim?’’. Verification of food authenticity claims is a challenging analytical task that require a qualitative answer to be addressed. The issue is conceptually very similar to quality control and, therefore, the same type of data-analysis tools should be employed. The most appropriate family of chemometric methods for addressing this type of problem goes by the name of class modelling or one-class classifiers. Such methods perform verification of compliance with a specification by defining a multivariate enclosed class space, at a predetermined confidence level, for authentic samples of the class under investigation. The first class modelling methods introduced into chemometrics were SIMCA (soft independent modeling of class analogy) [1] and UNEQ (unequal dispersed classes) [1]. These models present the advantages of describing perfectly the compliant samples and being free from the distribution of non-compliant samples in the training set.

Another important group of pattern-recognition tools is represented by the discriminant-classification techniques – also known as two-class or multiclass classifiers – which have been applied much more frequently than class modelling. All discriminant methods look for a delimiter between two or more classes, determined using a contribution from all of the classes considered. This means that all of the classes must be correctly defined and the samples included must be thoroughly representative of each class since they have a crucial influence on the decision rule to be derived. This is extremely important when the focus is on a single class like, for example, cases involving verification of a food authenticity claim. In fact, in such a case, the discriminant approach would require the collection of two sets of training samples: one representative of the product to be characterised and a second representative of the entire production of the same product that does not comply with the given claim. Such a condition is rarely realisable in practice, and collected sets of non-compliant samples are often under-representative of the non-compliance possibilities. This inevitably leads to biased decision rules, the outcomes of which are heavily dependent on those samples included in the non-compliant set. Thus, decisions regarding sample conformity based on a discriminant approach are generally less robust and altogether less suitable for practical conditions than those based on a class modelling strategy [3].

Issues related to development, optimisation and validation of suitable class models for the characterisation of food products will be critically analysed and discussed.

Acknowledgement: Financial support by the Italian Ministry of Education, Universities and Research (MIUR) is acknowledged – Research Project SIR 2014, RBSI14CJHJ (CUP: D32I15000150008).

References:

1. S. Wold, M. Sjöström, in: B.R. Kowalski (Editor), Chemometrics: Theory and Applications, ACS Symp. Ser. 52, American Chemical Society, Washington, USA, 243–282 (1977).
2. M.P. Derde, D.L. Massart, Anal. Chim. Acta, 184, 33 (1986).
3. P. Oliveri, G. Downey, TRAC, Trends in Anal. Chem., 35, 74 (2012).

SPECTROSCOPY UNDERCOVER: INVESTIGATION OF HIDDEN OBJECTS

A.L. Pomerantsev, O.Ye. Rodionova

Investigation of a sample covered by a nuisance layer is demanded in many fields, e.g. for the process control, for biochemical analysis, and for many other applications. This study is based on the analysis of spectra collected by the diffuse reflectance near infrared (DR NIR) spectroscopy. Each spectrum is a composition of the useful (target) spectrum and a spectrum of a nuisance layer. Unfortunately, DR NIR is a more complicated technique than the conventional transmission spectroscopy. Therefore, we cannot simply subtract the spectrum of the cover from the cumulative spectrum. To separate these signals we suggest using a new phenomenological approach, which employs the Multivariate Curve Resolution methods. We consider that the nuisance spectrum of the layer can be presented as a superposition of three signals $c_t s_t(\lambda) + c_a s_a(\lambda) + c_s s_s(\lambda)$, which are responsible for transmission, absorption, and scattering of the light beam inside the layer. Spectra $s(\lambda)$ are affected neither by the layer depth h, nor by the target properties. On the contrary, coefficients c depend both on h, and on the target reflectance factor, r. Applying the conventional transmission spectroscopy a researcher always seeks a method to remove both the reflectance and scattering effects from the spectra. The peculiarity of our approach is that we do not remove these components, but study them in details with the aim to use this knowledge for the target spectrum recovering.

We apply this approach to a particular system, which consists of several layers of polyethylene (PE) film and an underlayer with known spectral properties. To separate the information originated from the PE layers and the wanted target, we modify the system versus: (1) the number of the PE layers; (2) the reflectance properties of the target sample. Using the MCR approach we obtain three pure spectra $s(\lambda)$, as well as the c-profiles. The latter are described as computable functions, which depend on the PE depth, h, and the underlayer reflectance factor, r.

The concept is then utilized for obtaining the spectrum of a hidden object in the real-world systems. In particular, we succeed in reconstruction of the target using a single spectrum acquired through several PE layers.

MULTIVARIATE CURVE RESOLUTION-ALTERNATING LEAST SQUARES ANALYSIS OF HIGH RESOLUTION LIQUID CHROMATOGRAPHY-MASS SPECTROMETRY DATA

Melanie M. Sinanian[1], Daniel W. Cook[1], Sarah C. Rutan[1], Dayanjan S Wijesinghe[2]

[1]Department of Chemistry, Virginia Commonwealth University, Richmond, VA 23284-2006
[2]Department of Pharmacotherapy and Outcomes Science, Virginia Commonwealth University, Richmond, VA 23298-0533
srutan@vcu.edu

High resolution mass spectrometry (HRMS) methods are crucial for differentiating compounds with highly similar masses, a necessity when analyzing samples arising from such fields as metabolomics; however, the large size of high resolution LC-HRMS datasets can cause difficulties in the data analysis. In this work, LC-HRMS analysis was carried out and multivariate curve resolution-alternating least squares (MCR-ALS) was applied to the data to obtain mathematical separation of pure analyte chromatographic and spectral profiles. In order to minimize computational strain, a strategy was developed to minimize the number of irrelevant masses (*i.e.*, background ions) analyzed at full resolution. Data was first binned to unit mass and MCR-ALS was performed. Signals greater than a preset threshold were extracted from the component profiles and expanded to a higher level of precision. MCR-ALS was performed on this new data. This process of MCR-ALS, extraction, and expansion was repeated until 0.001 mass unit resolution was achieved, excluding more irrelevant masses at each step. This strategy allowed for the accurate recovery of all compound profiles while minimizing the size of the data analyzed in each MCR-ALS analysis – as small as 0.5 % of the masses from the original data – thereby minimizing the computational analysis time. The application of this technique to an analysis of a mixture of amphetamines and a bacterial lipidomics study will be shown. This strategy should provide a valuable tool for analyzing complex LC-HRMS data.

Acknowledgement: The authors acknowledge funding from grant CHE-517230 from the U.S. National Science Foundation.

ABSTRACTS

ORAL PRESENTATIONS

O-01

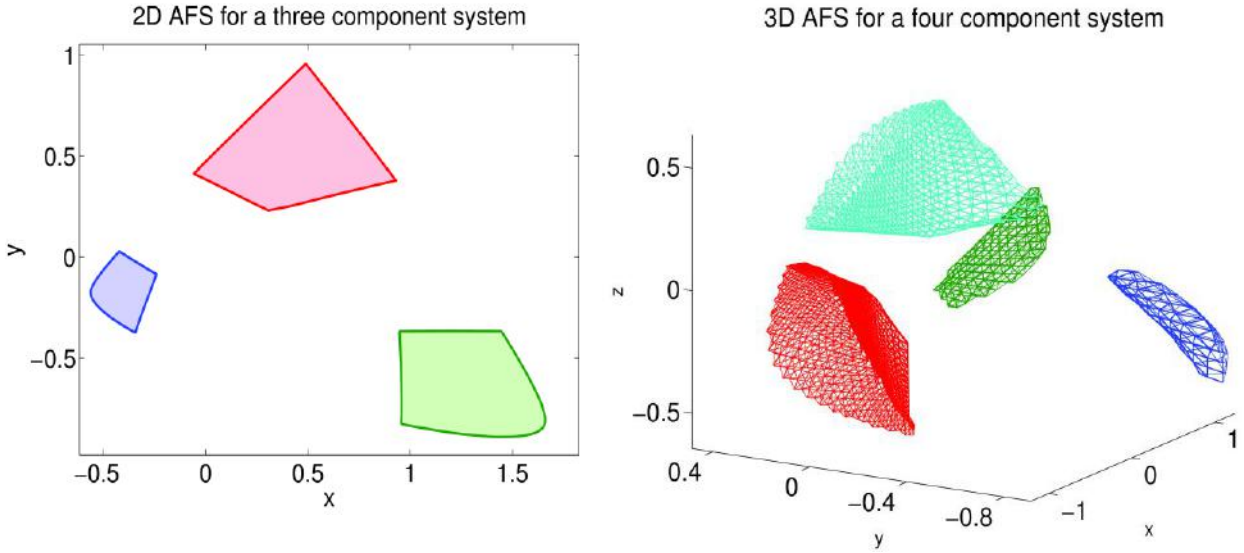
ON THE ANALYSIS AND COMPUTATION OF THE AREA OF FEASIBLE SOLUTIONS FOR THREE- AND FOUR COMPONENT SYSTEMS.

Mathias Sawall [1], **Klaus Neymeyr** [1,2]

[1]*Universität Rostock, Mathematical Institute, Rostock, Germany*
[2]*Leibniz-Institute for Catalysis (LIKAT), Rostock, Germany*
Email: mathias.sawall@uni-rostock.de

Multivariate curve resolution methods suffer from the so-called rotational ambiguity of the solution. A challenging approach to the ambiguity problem is to compute the full set of all concentration factors C and the spectral factors A in nonnegative factorizations of a given spectral data matrix. The area of feasible solutions (AFS) is the low-dimensional representation of all these solutions. The AFS analysis is a powerful methodology for the exploration of the rotational ambiguity inherent to the multivariate curve resolution problem. Up to now the AFS has been studied for two-, three- and four-component systems.

In this talk we explain the underlying concepts of the AFS theory and its contribution to a deepened understanding of the multivariate curve resolution problem. A survey is given on various methods for the computation of the AFS for three- and four-component systems. The focus is on methods which approximate the boundary of the AFS for three-component systems by inflating polygons (MatLab-toolbox FACPACK) and for four-component systems. Furthermore the reduction of the AFS by additional soft constraints is explained in order to extract chemically meaningful solutions from the set of all nonnegative solutions. Several numerical examples are discussed.



References:
[1] O.S. Borgen, B.R. Kowalski, *Anal. Chim. Acta. 174*, 1-26 (1985)
[2] R. Rajkó, K. István, *J. Chemom. 19*, 448-463 (2005)
[3] A. Golshan, H. Abdollahi, M. Maeder, *Anal. Chem. 83*, 836-841 (2011)
[4] M. Sawall, C. Kubis, D. Selent, A. Börner, K. Neymeyr, *J. Chemom. 27*, 106-116 (2013)
[5] M. Sawall, K. Neymeyr, *J. Chemom. 28*, 633-644 (2014)
[6] A. Jürß, M. Sawall, K. Neymeyr, *J. Chemom. 29*, 420-433 (2015)
[7] M. Sawall, N. Rahimdoust, C. Kubis, H. Schröder, D. Selent, D. Hess, H. Abdollahi, R. Franke, A. Börner, K. Neymeyr, *Chemom. Intell. Lab. Syst. 149*, 140-150 (2015)

O-02

ALS SCHEME USING EXTENT-BASED CONSTRAINTS FOR THE ANALYSIS OF CHEMICAL REACTION SYSTEMS

Julien Billeter, Michael Amrhein, Dominique Bonvin

Laboratoire d'Automatique,
Ecole Polytechnique Fédérale de Lausanne, Switzerland
julien.billeter@epfl.ch

Multivariate curve resolution via alternating least squares (ALS) is used to resolve the concentration profiles $\mathbf{C}$ and the pure component spectra $\mathbf{E}$ of S species from the multivariate absorbance data $\mathbf{A}$, assuming the bilinear model $\mathbf{A} = \mathbf{C}\,\mathbf{E}$. Due to the possible permutations of profiles and the presence of intensity and rotational ambiguities, soft constraints such as nonnegativity of $\mathbf{C}$ and $\mathbf{E}$ as well as unimodality, monotonicity, closure, and local rank selectivity of $\mathbf{C}$ are typically used to obtain tighter solution bounds for $\mathbf{C}$ and $\mathbf{E}$ [1].

In addition, hard constraints in the form of kinetic models are also often used. Unfortunately, these models are subject to structural plant-model mismatch and parametric uncertainty, which weakens their impact. As an alternative, this paper proposes to use constraints based on variant states called extents. The computation of these extents does not require any information on the rate processes, that is, $\mathbf{x}(t) = \mathbf{T}\,\mathbf{n}(t)$, with $\mathbf{x}(t)$ and $\mathbf{n}(t)$ the vectors of extents and numbers of moles at time t, respectively, and $\mathbf{T}$ a matrix known from the reaction stoichiometry, the inlet composition and the initial conditions [2]. Expressing the S concentrations in terms of d extents and $q = S - d$ invariants reduces the dimensionality of the problem from S to d. Each column of the extent matrix $\mathbf{X}$ describes the extent of a single rate process, for example of a reaction or an inlet flow. It turns out that the unknown concentration matrix $\mathbf{C}$ can be expressed in terms of the lower-dimensional matrix $\mathbf{X}$ as $\mathbf{C} = \mathbf{V}^{-1}\mathbf{X}\,\mathbf{T}^{-\mathrm{T}}$, where $\mathbf{V}$ is a diagonal matrix containing the volume profile. Since each column of $\mathbf{X}$ describes a single rate process, additional constraints can be enforced on $\mathbf{X}$ such as monotonicity and convexity/concavity [3]. Furthermore, the q invariant relationships can be used as constraints in the least-squares problem. As a consequence, the use of extents in ALS reduces the ambiguity between $\mathbf{C}$ and $\mathbf{E}$, yielding faster convergence and tighter solutions.

The use of extent-based constraints also opens up new perspectives for hard-soft ALS methods, since hard kinetic models can be identified individually (that is independently of the other rates) for some selected processes, while soft extent-based constraints are used for the unknown processes. Another feature involves the possibility of initializing or constraining the ALS scheme with a concentration submatrix (of dimension at least $S \times S$) estimated from multiple experiments performed under well-designed conditions and local-rank information. Together with the corresponding absorbance data, this submatrix can efficiently replace the traditional initialization via factor analysis and be used to compute a better initial estimate of $\mathbf{E}$.

After a brief review of the mathematical properties of extents for batch and open reactors, this talk will present the modified ALS scheme and illustrate it via simulated examples.

References:

[1] A. de Juan, J. Jaumot, R. Tauler, Anal. Methods 6, 4964 (2014).
[2] D. Rodrigues, S. Srinivasan, J. Billeter, D. Bonvin, Comp. Chem. Eng. 73, 23 (2015)
[3] S. Srinivasan, D.M.D. Kumar, J. Billeter, S. Narasimhan, D. Bonvin, IFAC Symposium DYCOPS (2016)

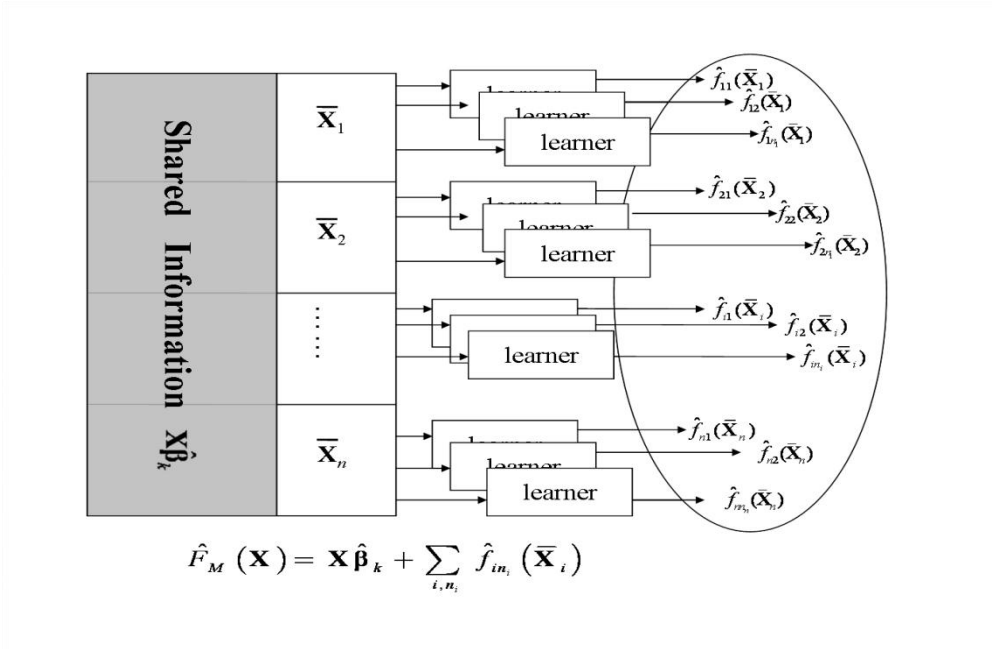
O-03

BOOSTING IN BLOCK VARIABLE SUBSPACES: AN APPROACH OF ADDITIVE MODELING FOR QSAR

Qing-Song Xu, Jian Xu

School of Mathematics and Statistics, Central South University, Changsha, 410083, China

Quantitative structure activity relationships (QSAR) and quantitative structure property relationships (QSPR) are established by a novel approach of additive modeling: boosting in block variable subspaces (BBVS). BBVS combines partial least squares regression (PLS) with a kind of gradient boosting in stepwise way. It searches for a base learner from a series of smaller block variable subspaces instead of from the whole variable space. This character allows BBVS to use the information correlating to *y* more effectively. Hence the fitting accuracy is improved, yielding a model of higher prediction ability. The use of this approach permits the development of PLS and boosting in modeling for QSAR and QSPR studies.



References:

1. J. M. Sutter, T. A. Peterson, P.C. Jurs, Prediction of gas chromatographic retention indices of alkybenzenes. *Anal. Chim. Acta*. 342 (1997) 113-122.
2. Y. Freund, R. E. Schapire, A decision-theoretic generalization of online learning and an application to boosting. *J. Comput. Syst. Sci.* 55 (1997) 119-139.
3. J. H. Friedman, Greed Function Approximation: A Gradient boosting Machine. *Ann. Stat*. 29 (2001) 1189-1232.
4. M. H. Zhang, Q. S. Xu, D. L. Massart, Boosting Partial Least Squares. *Anal. Chem.* 77 (2005) 1423-1431.
5. D.S. Cao, Q.S. Xu, Y.Z, Liang, L.X. Zhang, H.D. Li, The boosting: A new idea of building models; *Chemom. Intell. Lab. Sys.*, 100 (2010) 1–11
6. J. H. Friedman, B. E. Popescu, Gradient Directed Regularization, Available on http:// www-stat.stanford.edu /~jhf/ftp/pathlite.pdf.

O-04

An Information Based Approach To Selecting Features In High Dimensional Data.

M. F. Dorn[1], **C. H. Spiegelman**[2]

[1]*Statistics Department, Texas A&M University, Blocker-3143, College Station, Texas*
[2] *Statistics, Department, Texas A&M University, Blocker-3143, College Station, Texas*
cliff@stat.tamu.edu

In high dimensional chemical problems where there are hundreds to millions or more predictors it is often likely that information is left behind in the after variable or feature selection. While individually they may not contribute much to the model, as a collection they may. The focus of this work is to use information-based measures from statistics and machine learning to select a few features (combined predictor variables) to add increased power to the prediction model. While the features lack the same kind of interpretability that individual predictors have, sometimes increased reliability in prediction outweighs this issue. The approach borrows from PLS ideas, but the details are very different. We demonstrate the method on chemical data sets.

Oral presentation
Category: Theory and Methods
Keywords: Feature selection, Information, High-dimensional

O-05

PERMUTATION TESTING IN PCA – THEORETICAL AND PRACTICAL ASPECTS

R. Vitale [1], J.A. Westerhuis [2], A.K. Smilde [2], O.E. de Noord [3], T. Næs [4], A. Ferrer [1]

[1]Departamento de Estadística e Investigación Operativa Aplicadas y Calidad, Universitat Politècnica de València, Camino de Vera s/n, 46022, Valencia, Spain
[2]Biosystems Data Analysis, Swammerdam Institute for Life Sciences, University of Amsterdam, Science Park 904, 1098 HX, Amsterdam, The Netherlands
[3]Shell Global Solutions International B.V., Shell Technology Centre Amsterdam, PO Box 38000, 1030 BN, Amsterdam, The Netherlands
[4]Nofima AS, PO Box 210, 1431, Ås, Norway
rvitale86@gmail.com

Principal Component Analysis (PCA) is one of the most widely used techniques to describe, summarise and easily interpret data in many fields of interest, from analytical chemistry and medicine to process monitoring, environmental surveillance and computer science [1]. When building a PCA model, selecting the correct number of components is a critical step to allow only useful information to be retained and noise to be filtered out. In the last decades, much work has been devoted to methods for addressing this issue like Bartlett's chi-square test, Kaiser's eigenvalue greater than 1 rule, minimum average partial rule, scree test, parallel analysis and cross-validation [2]. However, scarce attention has been paid to the possibility of assessing the significance of the computed factors via permutation tests, which may represent a feasible approach in case the aforementioned ones cannot be applied (e.g. when the datasets under study do not fulfil specific statistical assumptions or are characterised by a not sufficiently high number of objects) [3].

The main aim of this work is to provide theoretical and practical insights for an improved understanding of this latter strategy, highlighting its pros and cons, mathematically formalising the algorithmic procedure to be followed when performing permutation tests in PCA, and possibly propose *ad hoc* solutions for optimising computational time, efficiency and robustness.

An additional interesting point to be noticed is that this methodology could be possibly combined to multi-set techniques such as Canonical Correlation Analysis (CCA) and Joint and Individual Variation Explained (JIVE) for a preliminary determination of the effective rank of the multiple handled data structures.

Acknowledgement: This research work was partially supported by the Spanish Ministry of Economy and Competitiveness under the projects DPI2011-28112-C04-02 and DPI2014-55276-C5-1R and Shell Global Solutions International B.V. (Amsterdam, The Netherlands).

References:

[1] H. Martens, J. Chemometr., 29, 563 (2015).
[2] E. Saccenti, J. Camacho, Chemometr. Intell. Lab., 149, 99 (2015).
[3] I. Endrizzi, F. Gasperi, M. Rødbotten, T. Næs, Food Qual. Prefer., 32, 198 (2014).

O-06

A NEW FORMULATION FOR ESTIMATING THE VARIANCE OF LINEAR MODEL PREDICTION

E. Fernandez-Ahumada[1], C. Gomez[2], J.M. Roger[3]

[1]Department of Animal Production, ETSIAM, University of Córdoba, 14014, Córdoba, Spain
[2]IRD, UMR LISAH (INRA-IRD-SupAgro), F-34060 Montpellier, France
[3]UMR ITAP, Irstea, F34196 Montpellier, France

Introduction:
There are two basic ways of estimating prediction uncertainty, namely, error propagation or resampling strategies. Error propagation leads to closed-form expressions where some hypotheses are made but which provide a platform for evaluating the different sources of uncertainty. Resampling is essentially a "black box" approach which, however, is often more accurate because fewer assumptions and approximations are made. Some analytical expressions can be found in the literature for error propagation method, but all consider local linearization and other important assumptions. Particularly, the errors in the predictors are assumed to be independent and to have constant variance. This latter assumption is never fulfilled in spectroscopy. So, this paper proposes a new expression for prediction uncertainty estimation based on the error propagation strategy, using as few as possible assumptions.

Theory:
Let **b** be a linear model between dependent variables **x** and a response *y*. One of the most complete existing and published expressions of prediction uncertainty is (A):

$$Var(\hat{y}) = \left(1 + \frac{1}{N}\right)\|\mathbf{b}\|^2 s_x^2 + \mathbf{z}^T\mathbf{Var}(\mathbf{b})\mathbf{z} + \frac{s_y^2}{N}$$

Where: z is the spectrum centred against calibration set and **Var** represents the variance covariance matrix. This expression uses some strong hypothesis on the error structure and dependency. Let assume only that the measurement errors have zero mean and that errors on the spectra and on the model are independent. Then, a new expression of prediction uncertainty is (B):

$$Var(\hat{y}) = \left(1 + \frac{1}{N}\right)\mathbf{b}^T\mathbf{Var(x)}\mathbf{b} + \mathbf{z}^T\mathbf{Var}(\mathbf{b})\mathbf{z} + Var(\mathbf{z}^T\mathbf{b}) + \frac{s_y^2}{N}$$

The main differences between the two expressions are: (i) The first term is more general in expression B, as it takes into account the complete variance / covariance of X; (ii) The third term of expression B is new; it represents a kind of covariance between the variations of z and those of b.

Material and methods:
This new formulation was tested on two datasets. (1) The first one consisted of NIR spectra of feed acquired in Laboratory with 10 repetitions, regressed against protein content. (2) The second one consisted of hyperspectral Vis-NIR airborne data acquired over 300 km² of cultivated area in Tunisia, regressed against clay content of soil.

Results:
On the dataset (1), the values of all terms of expressions A and B were estimated for several models, which differ by the preprocessing. It appears that expression A always overestimates the prediction variance, because it does not take into account systematic variance, due for example to the baseline shifts. It is also noticeable that the third term introduced by expression B is not at all negligible. On the dataset (2), several sources of uncertainty were simulated and introduced one the one hand in a bootstrap procedure and on the other hand in the new variance expression. This permitted to produce images of uncertainty, which were interpreted in relation to each term of the expression B.

O-07

INVESTIGATING CORRELATION VS. CAUSALITY - USING STRAW SAMPLES AND NIR AS AN EXAMPLE

Å. Rinnan[1], S. Bruun[2], J. Lindedam[2], S. B. Engelsen[1]

[1] Department of Food Science, University of Copenhagen, Rolighedsvej 26, 1958 Frederiksberg, Denmark
[2] Department of Plant and Environmental Sciences, University of Copenhagen, Thorvaldsensvej 40, 1871 Frederiksberg C, Denmark
aar@food.ku.dk

There are several applications in industry that show how slow and cumbersome reference analysis can be replaced with fast and non-invasive measurements by a spectroscopic instrument coupled with regression models. This has caused the end-user to ask for the prediction of smaller and smaller compounds within the product or sample of interest. Instrument vendors and creators of multivariate models have responded by making the desired tailor-made multivariate regression models for their customers, in order to keep them satisfied. However, this blind pursuit of customer satisfaction has come at the cost of a certain lack of real validation of the model, since for most cases, there is a high correlation between the different chemical compounds in a given sample type, especially in the food industry [1].

In this work, we will investigate whether a calibration model is performing well due to the correlation to another reference, or if the attribute of interest actually is giving a spectroscopic signature. Methods used will include the work by Eskildsen et al. [2], as well as Andersson [3].

References:

[1] Å. Rinnan and L. Munck, Journal of Chemometrics, 26, 2012, 487–495
[2] C.E. Eskildsen, M.A. Rasmussen, S.B. Engelsen, L.B. Larsen, N.A. Poulsen, T. Skov, Journal of Dairy Science, 2014, 97, 7940 – 7951
[3] C. A. Andersson, Chemometrics and Intelligent Laboratory Systems, 47, 1999, 51–63

O-08

COMDIM: FROM MULTIBLOCK DATA ANALYSIS TO PATH MODELING

E. M. Qannari [1], V. Cariou [1], E. Vigneau [1], D. N. Rutledge [2]

[1]LUNAM université, ONIRIS, Unité de Sensométrie et Chimiométrie, Nantes F-44322, France
INRA, Nantes, France
[2]UMR Ingénierie Procédés Aliments, AgroParisTech, Inra, Université Paris-Saclay, 91300, Massy, France.
elmostafa.qannari@oniris-nantes.fr

ComDim analysis, initially called Common Components and Specific Weights Analysis, has been efficiently applied to various kinds of data ranging from sensory and preference data [1] to chemometrics data [2]. It was also adapted to cope with multigroup data [3]. From a very different perspective, ComDim was extended to assess the significance of the effects of factors on multivariate data [4].

We recall that for a set of multiblock data $X_1, \ldots X_K$, supposed to be centered, ComDim aims at determining common components $q_1, q_2 \ldots q_n$, assumed to be of unit length, and saliences ($\lambda_r^{(k)}$) which reflect the importance of each block dataset in determining the underlying common components. These components and saliences are sought so as to approximate the cross-products matrices $W_k = X_k X_k^T$ by $Q \Lambda_k Q^T = \sum_{r=1}^{n} \lambda_r^{(k)} q_r q_r^T$.

We start by showing how the method of analysis can be extended in order to cope with the situation where we aim at explaining a dataset Y from K datasets $X_1, X_2, \ldots, X_k$. This new strategy of analysis follows the same pattern of analysis as ComDim by replacing the matrices $W_k = X_k X_k^T$ by $T_k = X_k X_k^T Y Y^T$.

In a second stage, we consider a multiblock setting with a specific pattern of directed relations among the blocks of variables and we show how ComDim analysis can be extended to investigate the relationships between these blocks taking account of the path diagram. For two datasets X_k and X_l linked by a causal path directed from X_k to X_l, we consider the matrix $T_{kl} = X_k X_k^T X_l X_l^T$ and we seek components t and u_{kl} supposed to be of unit length so as to minimize the quantity:

$$\sum_{l=1}^{L} \sum_{k=1}^{K} \partial_{kl} ||T_{kl} - \lambda^{(kl)} t u_{kl}^T||^2$$

Where ∂_{kl} takes the value 1 if the blocks k and l are linked and 0 otherwise. The quantity ($\lambda^{(kl)}$) highlights the importance of the link between X_k and X_l. Several algorithms can be proposed to solve this problem.

The methods of analysis are illustrated on the basis of case studies and the outcomes are compared to those of standard methods.

References:

[1] E. M. Qannari, P. Courcoux, E. Vigneau, Common components and specific weights analysis performed on preference data. Food Quality and Preference, 12 , 365-368 (2001).

[2] G. Mazerolles, M. Hanafi, E. Dufour, D. Bertrand, E. M. Qannari, Common components and specific weights analysis: A chemometric method for dealing with complexity of food products. Chemometrics and Intelligent Laboratory Systems, 81, 41-49 (2006).

[3] P. Courcoux, , M. F. Devaux , B. Bouchet. Simultaneous decomposition of multivariate images using three-way data analysis: Application to the comparison of cereal grains by confocal laser scanning microscopy. Chemometrics and Intelligent Laboratory Systems, 62 , 103-113, (2002).

[4] Jouan-Rimbaud Bouveresse, D., Climaco Pinto, R., Schmidtke, L. M., Locquet, N., Rutledge, D. N. Identification of significant factors by an extension of ANOVA-PCA based on multi-block analysis. Chemometrics and Intelligent Laboratory Systems, 106 , 173-182, (2011).

O-09

STRATEGIES FOR POWER ANALYSIS AND SAMPLE SIZE DTERMINATION FOR PCA AND PLS-DA OF MULTIVARIATE DATA

E. Saccenti, M. E. Timmerman

[1]*Laboratory of Systems and Synthetic Biology, Wageningen University, The Netherlands*
[2]*Department of Psychometrics and Statistics, University of Groningen, The Netherlands*

email: esaccenti@gmail.com

We deal with the problem of power analysis and optimal sample size determination for the analysis of multivariate data with PCA and PLS-DA[1]. The problem of determining the minimal sample size to obtain stable and reliable estimation of the model parameters is considered in the social sciences, especially in the PCA context [2], but has been seldom addressed in the chemometric field. Since *omics* data are usually analysed and explored using tools from the chemometric toolbox such as PCA and PLS-DA, determining the optimal sample size when planning an experiment becomes crucial.

To identify the optimal sample size for an ANOVA (univariate case) and MANOVA (multivariate case), approaches are available. However, for PCA and PLS-DA the matter complicates considerably and no generally applicable strategies exist. We present here relevant concepts and offer strategies for minimally required sample size estimation when planning experiments to be analysed using PCA and/or PLS-DA.

For PCA we discuss 1) the problem of determining the minimal sample size to obtain stable and reproducible component loading estimations and 2) the determination of the minimal sample size required to assess the dimensionality of a data set using arguments from Random matrix theory, which is also relevant in derived applications of PCA such as in deconvolution/curve resolution. For PLS-DA we discuss the analogies between the discrimination problem and hypothesis testing and propose a simulation strategy for sample size estimation.

These problems are treated at both the theoretical and empirical level and reviewed with both simulated and real data.

Acknowledgement: This work was partially supported by European Commission funded FP7 project INFECT (contract number: 305340).

References:

[1] Saccenti, E., Timmerman, M., *Approaches to sample size determination for multivariate data: applications to PCA and PLS-DA***.** Submitted

[2] Timmerman, M.E., H.A.L. Kiers, and A.K. Smilde, *Estimating confidence intervals for principal component loadings: A comparison between the bootstrap and asymptotic results.* British Journal of Mathematical and Statistical Psychology, 2007. **60**(2): p. 295-314.

O-10

Unsupervised Random Forest: An Introduction with Case Studies

Nelson Lee Afanador[1], Agnieszka Smolinska[2,3], Thanh Tran[4], Lionel Blanchet[2,3,5]

[1] Center for Mathematical Sciences, Merck, Sharp & Dohme, West Point, PA, USA
[2] Department of Pharmacology and Toxicology, School of Nutrition, Toxicology and Translational Research in Metabolism (NUTRIM), Maastricht University Medical Center, Maastricht, the Netherlands.
[3] Thayer school of Engineering, Dartmouth College, Hanover, New Hampshire, United States of America
[4] Center for Mathematical Sciences, Merck, Sharp & Dohme, Oss, The Netherlands
[5] Top Institute Food and Nutrition (TIFN), Wageningen, The Netherlands.

nelson.afanador@merck.com

Keywords: Random Forest, Unsupervised Methods, Explorative analysis, Visualization

Unsupervised methods, such as Principal Component Analysis, have gained popularity and wide-spread acceptance in the chemometrics and applied statistics communities. Unsupervised Random Forest is an additional method capable of discovering underlying patterns in the data.
The number of applications of unsupervised Random Forest (URF) in chemometrics has been limited. One possible cause for this is the belief that Random Forest (RF) can only be used in a supervised analysis setting.

URF makes a clever use of a simple assumption: if the data holds any structure it should be distinguishable from a randomly generated version of itself. For this reason, a *synthetic* dataset is randomly generated from the original dataset and together they form a two-class classification problem that is then modeled using classical (supervised) RF. If the subsequent analysis results in a meaningful classification model between real and *synthetic* data, then a *proximity matrix* obtained from a RF model can be used to search for trends and clusters in the real data. These *proximity scores* are used to perform a powerful unsupervised analysis such as clustering, multi-dimensional scaling (MDS), and Principal Coordinates Analysis (PCoA) for detecting meaningful structure in a data set.

This presentation introduces the basic concepts of unsupervised Random Forest and illustrates several applications in chemometrics through worked examples.

O-11

A NOVEL APPROACH TO DIAGNOSIS AND FOLLOW-UP OF INDIVIDUAL PATIENTS BY SPARSE MODELING

J. Engel [1,2], L. Blanchet [1,3], U.F.H. Engelke [4], R.A. Wevers [4] and L.M.C. Buydens [1]

[1]Radboud University Nijmegen, Institute for Molecules and Materials, P.O. Box 9010, 6500 GL Nijmegen, The Netherlands
[2]NERC Biomolecular Analysis Facility – Metabolomics Node (NBAF-B), School of Biosciences, University of Birmingham, Edgebaston, B15 2TT Birmingham, United Kingdom
[3] Department of Pharmacology and Toxicology, School of Nutrition and Translational Research in Metabolism (NUTRIM), Maastricht UniversityMedical Centre, Universiteitssingel 50, 6229 ER Maastricht, the Netherlands
[4] Laboratory of Genetic Endocrine and Metabolic Diseases at the Department of Laboratory Medicine, Raboud University Medical Centre, Geert Grooteplein 10, Nijmegen, the Netherlands
l.buydens@science.ru.nl

The –omics technologies are becoming increasingly important in health care and are expected to contribute to personalized health care. Statistical health monitoring (SHM) was recently introduced for analysis of –omics data to automatically identify the disease response in an individual patient [1]. This approach could be of use in all sorts of applications such as diagnosis of rare diseases, analysis of individual patterns in disease amnisfestation, disease monitoring, or personalized therapy.

SHM essentially combines estimation of Mahalanobis distances (MD) with principal component analysis (PCA). It is well known that the dimension reduction step via PCA can hamper reliable identification of the disease response in a patient. Therefore, sparse SHM (sSHM) is introduced in this presentation. The method combines estimation of the MD with variable selection by inclusion of an l_1-norm constraint.

Simulations are used to show that the sSHM model can identify the disease response in an individual patient more reliably compared to SHM. Subsequently, sSHM is applied to urine ^{1}H-NMR metabolomics data for diagnosis of several orphan diseases. Additionally, the method is used in combination with 1H-NMR metabolomics to obtain a very detailed assessment of the health status of five individuals and to closely monitor this status over time.

References:

[1] J. Engel, L. Blanchet, U.H.F. Engelke, R.A. Wevers, L.M.C. Buydens, PloS one, 9, 2013.

O-12

ORTHOGONALISATION METHOD FOR ROBUSTNESS IMPROVEMENT OF INLINE NIR APPLICATIONS

Lallemand J.[1], Guilment J.[2], Dubuc P[2]. , Montagnier S.[2] et Roussel S. [1]

[1] Ondalys, 4 rue Georges Besse, 34830 Clapiers, France

[2] ARKEMA - CERDATO / Laboratoire d'Étude des Matériaux (LEM) Route du Rilsan, 27470 Serquigny – France

jlallemand@ondalys.fr

Online model maintenance is a main problem for developing NIRS applications. Perturbations appearance due to environmental changes, maintenance operation or aging of the instrument, often affect model performances. Model correction with classical methods such as bias and slope correction or model redevelopment are not always satisfactory strategies.
The use of an orthogonalisation method can be an effective way to solve this problem and it is illustrated in this study with an industrial application.

Monitoring of polyamide polymerization by NIRS is a well-known subject which gives excellent results. The measurements can be made at-line on powders or granulates, but can also be performed on line on powders or in molten medium. A PLS model allows to access directly the end of polymer chains, or less indirectly at the viscosity of the product. In this study, the viscosity prediction by PLS allows real time monitoring of the process. However, after several years of operation, an unidentified perturbation appeared, leading to the failure of the PLS model during several months.

This industrial application is an ideal case for applying Dynamic Orthogonal Projection [1] (DOP). The purpose of this chemometric method is to make the model independent from perturbations.
The principle is to rebuild spectra as if they were measured without the perturbation. Only a small number of samples are needed to model the perturbation space. This is done by PCA, based on spectral differences between real spectra and reconstructed spectra. The calibration database is then projected orthogonally from this space and the model is rebuilt. The corrected model becomes independent of the presence or not of the perturbation and new spectra do not require any orthogonalisation processing before applying the model.

DOP has been applied with success to correct the PLS model of viscosity prediction with few samples, whereas model redevelopment was not entirely satisfactory. Furthermore, the study of the spectral zone affected by the perturbation and corrected by DOP, has allowed to come back to the process to identify what went wrong and then act directly on the process.

Maintenance and robustness problems of predictive models in NIRS are a real restraint for its expansion in the industrial world. DOP is an elegant mathematical solution which allow to overcome the impact of appearance and disappearance of perturbations with only few samples.

References

[1] M. Zeaiter, J.M Roger, V.Bellon-Maurel, « Dynamic Orthogonal Projection. A new method to maintain the on-line robustness of multivariate calibrations. Application to NIR based monitoring of wine fermentations ». Chemometrics and Intelligent Laboratory Systems, Vol.80, issue 2, pp. 227–235, 2004.

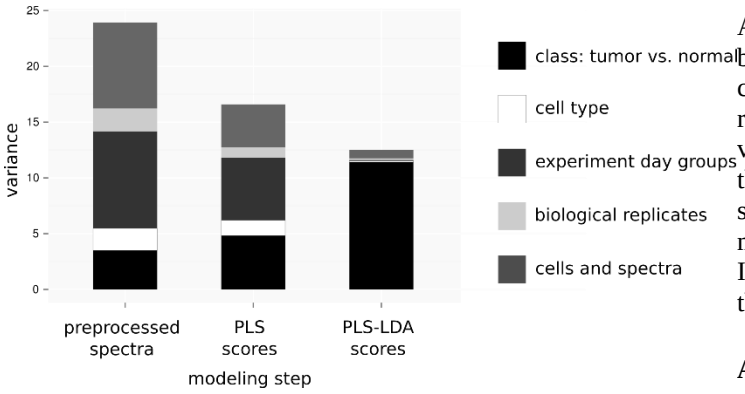
O-13

EXPERIMENTAL DESIGN CONSIDERATIONS FOR MODEL TRAINING WITH MANY CONFOUNDERS: VARIANCE IN THE COURSE OF THE MODELING PROCESS

C. Beleites [1,2], V. Tafintseva[3], C. Krafft[1], J. Popp[4], A. Kohler[3]

[1]Claudia Beleites Chemometric Consulting, Södeler Weg 19, D-61200 Wölfersheim, Germany,
[2]Leibniz Institute of Photonic Technology, Albert-Einstein-Str. 9, 0775 Jena, Germany,
[3]Norwegian University of Life Sciences, Drøbakveien 31, 1432 Ås, Norway.
[4]Institute of Physical Chemistry and Abbe Center of Photonics, Friedrich-Schiller-University Jena, Germany.
E-mail: chemometrie@beleites.de, Claudia.Beleites@leibniz-ipht.de

Biomedical vibrational spectroscopy is often challenged by a hierarchy of variability levels that occurs during sampling. When sampling for clinical studies, typically multiple specimen are taken per patient, many spectra are collected of each specimen and instrumental replicate or operator-related variability may occur. Predictive multivariate models are often challenged by these underlying and hierarchically occuring variability levels. In order to develop stable predictive models, it is desirable to estimate the variability that occurs at different levels. While guidelines for the number of samples that are required for training and testing of predictive models are readily available (e.g., [1]), guidelines for numbers of samples that are required at the different levels of variability have been missing so far.
We propose to analyze the variance structure at different variability levels and at different steps in the modeling process by ANOVA simultaneous component analysis (ASCA) [2]. This allows identification of confounding factors that require a high number of samples in order to develop stable models.
We demonstrate the proposed strategy and our findings with Raman spectra acquired in the context of circulating tumor cell identification.



ASCA of data from [1]: **bottom to top** in each stack of bars: class squred distance (tumor vs. normal cells); confounders: cell line/patient variance; biological replicates; 3 groups of measurement days; cell-to-cell variance and noise. **Left to right**: Development throughout modeling. PLS clearly reduces variance on single spectra and between biological replicates; but much less so the variation between different cell lines. In contrast, subsequent LDA leaves mainly variance on the cells/spectra.

Acknowledgement: CB thanks the BMBF for financial support via project RamanCTC (13N12685). This collaboration was also supported by a short term scientific mission within COST Action Raman4Clinics (BM1401).

References:

[1] C. Beleites *et al.*: Sample size planning for classification models., Anal Chim Acta, 2013, **760**, 25-33, DOI:10.1016/j.aca.2012.11.007.
[2] H. Hoefsloot et al: Multiset Data Analysis: ANOVA Simultaneous Component Analysis and Related Methods. In: Comprehensive Chemometrics. Elsevier, 2009, 453–472. DOI: 10.1016/B978-044452701-1.00054-5.

O-14

Fault Diagnosis: Contribution plots Vs oMEDA

Marta Fuentes-García, José Camacho, Gabriel Maciá-Fernández [1]

[1]Dept. Signal Theory, Telematics and Telecommunications, CITIC - ETS Ingenierías Informática y de Telecomunicación, University of Granada, C/ Periodista Daniel Saucedo Aranda, s/n, 18071 Spain

nmfuentes@ugr.es

Multivariate Statistical Process Control (MSPC) based on multivariate methods like Principal Component Analysis (PCA) is a well known methodology in the chemometrics domain. This methodology is based on splitting the data collected from a process into a model and a residual sub-spaces. Then, each of these sub-spaces are monitored with a different statistic, namely the D-statistic for the model space and the Q-statistic for the residual space. The D-statistic and the Q-statistic are monitored in a pair of Shewhart charts, where control limits are defined. Anomalous events are detected when these statistics exceed the control limits for one or a given number of consecutive sampling times in any of the two charts. Once an anomaly is detected, it should be diagnosed, so that problems within the process are timely identified and can be corrected for.

Contribution plots are the most accepted tools for fault diagnosis within PCA-based MSPC. They show the contribution of each original variable to the anomalous value in any of the statistics. A reconstruction-based method [1] was proposed as an alternative to traditional contribution plots. Both approaches are known to suffer from smearing [2], which affects the accuracy of the diagnosis. The smearing effect is not well understood yet, and it remains an open problem.

Fault diagnosis can also be performed by using a recently proposed method referred to as oMEDA (observation-based Missing-data method for Exploratory Data Analysis) [3]. The goal of oMEDA is to identify the contribution of the variables to a given direction in the model or residual sub-spaces. This let us discover the relationships between observations and variables, which is useful for Exploratory Data Analysis. As such, the oMEDA approach is not limited to diagnosis for a single faulty observation, and allows the comparison between groups of observations in a specific sub-space. The main difference between fault diagnosis using oMEDA and contribution plots is that the former focus on the sub-space while the latter focus on a specific statistic in that sub-space. However, in a preliminar study we found that in most faulty situations, contribution plots and oMEDA identify the same related variables.

The aim of this work is to present a comparison in fault diagnosis between contribution plots and oMEDA. The goal is to examine the common features and differences between these methods.

Acknowledgement: This work is partly supported by the Spanish Ministry of Economy and Competitiveness and FEDER funds through project TIN2014-60346-R.

References:

[1] C. F. Alcala, S. J. Qin, Analysis and generalization of fault diagnosis methods for process monitoring, Journal of Process Control 21 (3) 322-330 (2011).

[2] J. A. Westerhuis, S. P. Gurden, A. K. Smilde, Generalized contribution plots in multivariate statistical process monitoring, Chemometrics and intelligent laboratory systems 51, 95-114 (2000)

[3] J. Camacho, Observation-based missing data methods for exploratory data analysis to unveil the connection between observations and variables in latent subspace models, Journal of Chemometrics (2011).

O-15

MULTIVARIATE STATISTICAL PROCESS CONTROL FOR THE CONTINUOUS MANUFACTURING OF SOLID ORAL DOSAGE FORMS

Zhenqi (Pete) Shi [1], Joshua Hanson[1], Salvador García Muñoz [1]

[1]*Small Molecule Design and Development, Lilly Research Laboratories, Indianapolis, IN, 46285, USA*

Email: sal.garcia@lilly.com

Nowadays a commodity in the industrial software industry, the application of principal components analysis (PCA) to monitor the state of a process with Multivariate Statistical Process Control (MSPC) charts is far from a venture. Its implementation in a commercial plant with a contemporary control and/or automation system requires a modest effort, given a stable process and the availability of a historian data base with abundant data representative of the normal operating conditions (NOC) for the plant. This work presents the development of an MSPC monitoring scheme for a continuous drug product manufacturing process, using information available at the R&D stage. At the time when the process is still being developed, the data representative of the expected NOC may not be abundant. Additionally, the product development efforts may have other experimental priorities, for example running experimental trials at different mass rates and/or configuration of the feeders, mixer, and/or tablet press equipment.

The approach presented is based on the assumption that, regardless of the target set point for the main operating conditions (e.g. mass rate), the common cause variability in the process will remain the same. Hence the covariance structure for the variables in the system can be identified by a PCA model on a matrix resulting from the concatenation of data from multiple "trial runs" that were pre-centered around their mean values. An analysis of the loadings vectors reveals variable correlations that matches the expectations from a process engineering standpoint; and the corresponding movement in the scores space reveals directions of expected normal (and tolerated) variability, as well as patterns that can be matched with the startup and the shutdown of the system.

Our case study presents the analysis of a model built with the data collected from the system of screw feeders that dispense material to the unit, as well as the analysis of the model obtained with data collected from the tablet press. Furthermore, a method is presented to determine the lag time between a measurement vector from the feeders, and a measurement vector from the tablet press; such that a combined lagged model can be built. The combined model eases the holistic identification of contributions in the case of a disturbance, and the use of the three models (one per system and the combined) in tandem provides a complete monitoring approach for the system as a whole.

The online implementation of the above assumes that the mean value for the process variables in the upcoming manufacture run is known a-priori. This may not be true in the early lifecycle of the process, and as such we propose a method that provides an adaptive estimate of the mean to mean-center upcoming measurement vectors. Several approaches to achieve this are contrasted and discussed in terms of their ability to detect unwanted disturbances.

O-16

A Statitistical Process Control Framework to Assess High- Throughput Analytical Instrumentation.

Lorenzo Vega-Montoto

Idaho National Laboratory, Chemistry and Radiation Measurement Division, Idaho Falls, Idaho, USA

In the last two decades analytical instrumentation has increased the amount of measurement it can provide by a few orders of magnitude. This high-throughput revolution has been accompanied by an increment in the instrumental complexity as well as in the analysis and variability control of the measurement systems. Proteomics, metabolomics, genomics and even green bioenergy fields have benefited by the availability of big data. However, in order to draw significant chemical and biological conclusions from these highly hyphenated instrumental setups, it is imperative that we can control the sources of variability and being able to identify situations when the instrumentation is providing sub-optimal performance. These issues have being recognized and a significant number of freely available tools have been introduced that monitor different analytical techniques. Unfortunately, most of these approaches tend to produce an large number of quality control variables that in many cases are redundant, highly collinear and composed by non-linear relationships. On top of that, they are usually analyzed from an univariate quality control scenario neglecting the intrinsic correlation that they exhibit. Statistical process control (SPC) is a robust set of tools that aids in the visualization, detection, and identification of assignable causes of variation in any process that creates products, services, or information. In this work, we describe a framework that has been implemented based on diverse aspects of SPC such as: multivariate control charts and Pareto analysis among many others. It is able to identify the most informative variables and pre-process them in such a manner the quality metrics are robustified to minimize the effect instrumental tuning and calibration events. The framework can also identify out-of-control situations and suggest corrective actions to keep the instrumentation working an optimal fashion. Two study cases: one applied to a large proteomic study using LC-MS/MS and the other to the characterization of lignocellulosic biomass using Pyrolisis-GC/GC/MS will be presented to showcase the framework and its numerous possibilities.

Acknowledgement: Daniel C. Liebler, David Tabb and Gary S Groenewold

O-17

PENALIZED CANONICAL CORRELATION: CONNECTING THE MICROBIOME WITH DIARY DATA

J.J. de Rooi[1], J. Thorsen[2], S. Nørgaard[2], J. Stokholm[2], K. Bønnelykke[2], H. Bisgaard[2], A. Smilde[1]

[1]Biosystems Data Analysis Group, University of Amsterdam, The Netherlands.
[2]Copenhagen Prospective Studies on Asthma in Childhood, Health Sciences, University of Copenhagen, Denmark.
j.j.derooi@uva.nl

In recent years, there has been an increasing interest in the link between the microbiome and human health and disease. A first step towards better understanding of the interactions between host and microbiome is to investigate microbial communities in terms of their diversity and abundance. Data is most often obtained by sequencing and are summarized in a table with counts per species and subjects. The dominant way of analyzing microbiome data in a multivariate manner is by using indirect analysis i.e. transforming the data into distances and subsequently analyse those distances.

Here we wish to use direct analysis (i.e. using the raw data) to unveil relationships between microbiomes and a second block of variables. When the relations between the blocks are (assumed) to be symmetric, one can use Canonical Correlation analysis (CCA) [1]. Penalized CCA was introduced to overcome disadvantages typically encountered when analyzing high-dimensional data (see e.g. [2]). Such penalties can be posed on the estimated weights of the model, usually an L_2-norm or L_1-norm penalty is used.

In case of microbiome data we can use prior knowledge, in the form of the known phylogenetic relations, to guide the shrinkage procedure. This information can be utilized by using a Laplacian matrix containing both the topology of the phylogenetic tree and distances between species. In this research the aim is to connect the microbiome with common childhood illness. Hence, our second block of data consists of diary reports of the symptom burden related to childhood illness, measured over time. These data can be included in different forms. A first option is to summarize the measurements within a certain time window to a single value per symptom. Alternatively, when dynamics are assumed to be informative, we can include the diaries as 'signals'. In this setup we consider the different symptoms as smooth curves within the time window and estimate those using P-splines [3]. This results in a penalized CCA with the microbiome in one block and a set of functional variables as the second block.

References:

[1] H. Hotelling. (1936). Relations between two sets of variates. *Biometrika*, 28(3/4):321-377.
[2] S.Waaijenborg, P.C. Verselewel de Wit Hamer, A.H. Zwinderman. (2008). Quantifying the association between gene expressions and dna-markers by penalized canonical correlation analysis. *Statistical Applications in Genetics and Molecular Biology*, 7(1):1-29.
[3] P.H.C. Eilers, B.D. Marx. (1996). Flexible Smoothing Using B-Splines and Penalized Likelihood, *Statistical Science*, 11, 89–121.

O-18

DYNAMIC ELEMENTARY MODE MODELLING OF NON-STEADY STATE FLUX DATA

A. Folch-Fortuny[1], B. Teusink[2], H.A.L. Kiers[3], H.C.J. Hoefsloot[4,5], A.K. Smilde[4,5], A. Ferrer[1]

[1]Departamento de Estadística e IO Aplicadas y Calidad, Universitat Politècnica de València, Valencia, Spain.
[2]Systems Bioinformatics, Centre for Integrative Bioinformatics, Free University of Amsterdam, Amsterdam, The Netherlands
[3]Heymans Institute for Psychology, University of Groningen, Groningen, The Netherlands.
[4]Biosystems Data Analysis, Swammerdam Institute for Life Sciences, University of Amsterdam, Amsterdam, The Netherlands.
[5]Netherlands Metabolomics Center, Leiden, The Netherlands.
abfolfor@upv.es

Principal component analysis (PCA) and multivariate curve resolution (MCR) models have been proposed to obtain a set of key pathways in metabolic networks, assuming steady state conditions [1,2]. These pathways or modules in the network are identified using the existing relationships between metabolic fluxes, measured experimentally. Recently, a new method called principal elementary mode analysis (PEMA) [3] has been proposed to model this kind of data. The methodology is based on the flux projection to a reduced set of elementary modes (EMs) of the metabolic network. The EMs are the simplest representations of pathways crossing the metabolic network. Basically, each EM connects substrates with end-products concatenating reactions in a thermodynamically feasible way. For non-steady state cases, *e.g.* when measuring the concentrations of the metabolites at early stages after perturbation, 13C-metabolic flux analysis (MFA) [4], dynamic flux balance analysis (DFBA) [5,6], and the Goeman's Global test [7] have been proposed, among other methods.

Here we define a new framework to model non-steady state metabolic fluxes. This methodology is based on the novel concept of dynamic EMs (dynEMs), *i.e.* EMs that are used partially at each time point of the experiment. In this way, we propose dynamic elementary mode regression discriminant analysis (dynEMR-DA) to identify the set of dynEMs whose activation pattern allows discriminating between different biological conditions.

Actual [7] and simulated [8] non-steady state flux data sets from *Saccharomyces cerevisiae* are analysed using this methodology, identifying the most discriminant dynEMs when changing the initial concentrations of glucose and ethanol, and when changing from aerobic to anaerobic conditions.

Acknowledgement: Research in this study was partially supported by the Spanish Ministry of Science and Innovation and FEDER funds from the European Union through grants DPI2011-28112-C04-02 and DPI2014-55276-C5-1R

References:

[1] J.M. González-Martínez, A. Folch-Fortuny, F. Llaneras, M. Tortajada, J. Picó, A. Ferrer, Chemometr. Intell. Lab., 134, 89-99 (2014).
[2] A. Folch-Fortuny, M. Tortajada, J.M. Prats-Montalbán, F. Llaneras, J. Picó, A. Ferrer, Chemometr. Intell. Lab., 142, 293-303 (2015).
[3] A. Folch-Fortuny, R. Marques, I.A. Isidro, R. Oliveira, A. Ferrer, Mol. BioSyst., Accepted (2016)
[4] W. Wiechert, 13C metabolic flux analysis, Metab. Eng., 3(3), 195-206 (2001).
[5] R. Mahadevan, J.S. Edwards, F.J. Doyle 3rd, Biophys. J., 83, 1331-1340 (2002).
[6] A.M. Willemsen, D.M. Hendrickx, H.C.J. Hoefsloot, M.M.W.B. Hendriks, S.A. Wahl, B. Teusink, A.K. Smilde A.K., A.H.C. van Kampen, Mol. BioSyst., 11, 137-145 (2015).
[7] D.M. Hendrickx, H.C.J. Hoefsloot, M.M.W.B. Hendriks, B. Canelas, A.K., Anal. Chim. Acta, 719, 8-15 (2012).
[8] J.H. van Heerden, M.T. Wortel, F.J. Bruggeman, J.J. Heijnen, Y.J.M. Bollen, R.Planqué, J.Hulshof, T.G. O'Toole, S. Aljoscha Wahl, B. Teusink, Science, 343, 1245114 (2014).

O-19

Analyzing unbalanced multifactorial experiments with ASCA and APCA: Application in metabolomics

Bernadette Govaerts [1], Michel Thiel [2]

[1] ISBA, Université Catholique de Louvain, Voie du Roman Pays 20, 1348 Louvain-La-Neuve, Belgium
[2] Nonclinical Statistics and Computing, Janssen Pharmaceutica, Turnhoutsweg 30, 2340 Beerse, Belgium
mthiel2@ITS.JNJ.com

Metabolomics aims to analyze the entire pool of endogenous metabolites in biofluids and is then particularly indicated to extract biochemical information reflecting biological events [1,2]. Analytical methods used in metabolomics produce spectra for each experimental unit wich result in multivariate databases typically characterized by a number of correlated variables (descriptors) greater than the number of experimental units [3]. It is therefore unpossible to detect spectral differences by visual inspection and systemic differences are often masked by random variations (noise). Therefore, biomarker discovery involves the use of multivariate statistical tools to highlight descriptors which are consistently modified by the different biological states [3].

Two methods developed about ten years ago [4,5] which combine ANOVA and PCA, include the information from the design of experiment. Those methods are called ASCA (ANOVA-simultaneous component analysis) and APCA (ANOVA-principal component analysis). Their major limitation is that they provide biased estimators of the effects of factors when designs are unbalanced.

This presentation will introduce two new methods: ASCA+ and APCA+ that allow, respectively, to extend the use of ASCA and APCA to unbalanced designs by using least squares estimators rather than simple differences of means. Figures 1 and 2 illustrate the advantage of ASCA+ compared to ASCA in the highlighting of biomarkers corresponding to a factor of interest in a complex experimental design. The data correspond to urine samples of rats spiked with different concentrations of hippurate and citrate. ASCA and ASCA+ are used to analyze the effect of both products on spectral profiles. Figures 1 and 2 show respectively the loading plot of the factor citrate with ASCA and ASCA+ on an unbalanced dataset. With balanced designs, ASCA and ASCA+ give identical results, but in unbalanced cases, ASCA provides biased estimates while ASCA+ is unbiased. It can be seen in Fig. 1 that citrate is highlighted but hippurate is also detected due to errors of estimation. In ASCA+ (Fig. 2), only citrate is detected. ASCA+ is then able to correctly discriminate biomarkers associated with factors of interest also in cases of unbalanced designs.

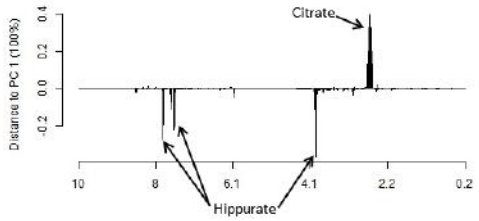


Figure 1: loading plot of the factor citrate on PC1 in ASCA.

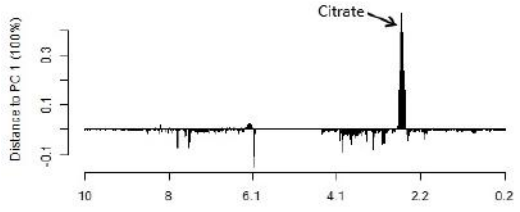


Figure 2: loading plot of the factor citrate on PC1 in ASCA+.

References:

[1] J.C. Lindon, J.K. Nicholson, E. Holmes, The Handbook of Metabonomics and Metabolomics, p. 413 (2011).
[2] D. Beyoğlu, J.R. Idle, Biochemical pharmacology Vol. 85 (1), p. 12 (2013).
[3] M. Lämmerhofer, W. Weckwerth, Metabolomics in Practice: Successful Strategies to Generate and Analyze Metabolic Data, (2013).
[4] A.K. Smilde, J.J. Jansen, H.C.J. Hoefsloot, R.J.A.N. Lamers, J. van der Greef, M.E. Timmerman, Bioinformatics Vol. 21 (13), p. 304 (2005).
[5] P.B. Harrington, N.E. Vieira, J. Espinoza, J.K. Nien, R. Romero, A.L. Yergey, Analytica Chimica Acta Vol 544 (1–2) p. 118 (2005).

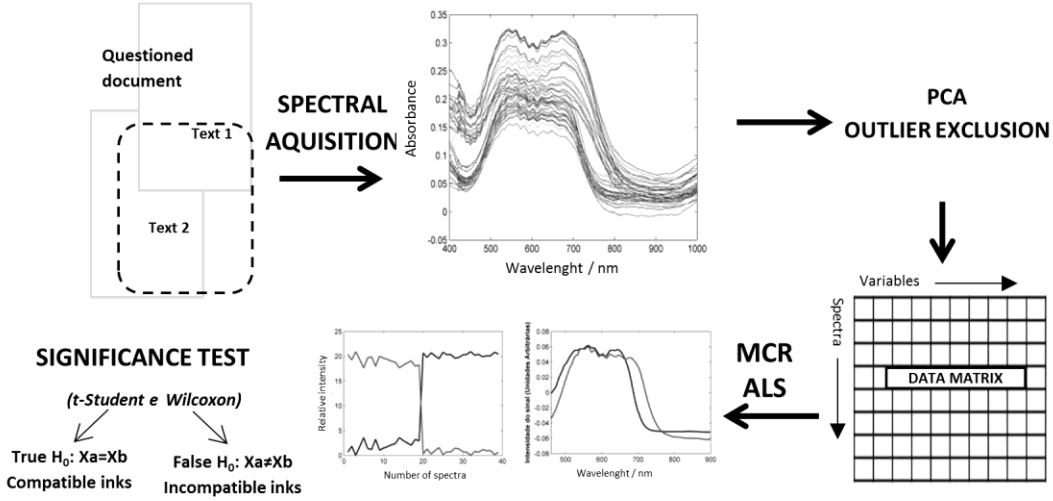
O-20

IDENTIFICATION OF FORGERY ON WRITING DOCUMENTS BY INK ANALYSIS USING VISIBLE SPECTROSCOPY AND MCR-ALS

V.A.G. Silva[1], M. Talhavini[2], J.J. Zacca[2], B.R. Trindade[2], J.W.B. Braga[1]

[1]Institute of Chemistry, University of Brasília, Brasília, Brazil.
[2]National Institute of Criminalistics, Brazilian Federal Police, Brasília, Brazil.
jez@unb.br

The analysis of inks was performed to verify the occurrence of forgery in documents. Spectroscopic methods have become an attractive technique in forensic document analysis, since they preserve the integrity of the documents [1]. In this work, it is proposed a method for pairwise analysis of blue pen inks of different brands of the same type. Five ballpoints, four rollerball and three gel pen brands were tested in standard samples produced in white sheet of paper, weight 75g/m^2. Each ink stroke was performed simulating a cursive handwriting [1]. The visible reflectance spectra were obtained by the Video Spectral Comparator 6000 and the analysis performed by Multivariate Curve Resolution – Alternating Least Squares (MCR-ALS). Twenty spectra were obtained for each ink stroke along of the trace. Figure 1 present the step by step of the analysis. The method was also validated by the analysis of a blind test and a real forensic case. In all comparisons, it was possible to correctly discriminate all pen brands tested by the relative intensities and the pure spectra obtained by the MCR-ALS. The decision about the similarity or discrimination of the inks was performed by Wilcoxon-Mann-Whitney significance tests, at the 0.05 significance level. In the blind test four volunteers selected randomly four pens, chosen from twenty-five possibilities. Ink strokes were produced with these pens, which were analyzed according the proposed method. The results showed that all inks were considered incompatible when the brands were different and compatible for the same brands. The method has proved to be accurate and robust regarding the handwriting of different individuals and pen batches in the blind test. The analysis of the real case consisted of a written contract on white paper weight of 75 g/m^2, containing signatures. The results demonstrated that the inks in the signatures were compatible with the same pen brand. This result was consistent with the result of the classical analysis performed by a forensic expert. The proposed method is non-destructive, fast, requires only a small amount of ink and is robust regarding instrumental variations of the spectrometer.



Acknowledgement: CAPES/PROFORENSE, CNPq and FAPDF.

References:

[1] V.A.G. Silva, M Talhavini, J.J. Zacca, B.R. Trindade, J.W.B. Braga, J. Braz. Chem. Soc., 25, 1552 (2014).

O-21

EXTENDING THE APPLICATION OF REGULARIZED MANOVA TO THE ANALYSIS OF MULTI-FACTOR MASS SPECTROMETRY METABOLOMICS DATA

J. Engel [1], U. Sommer [1], M.R. Viant [1]

[1]NERC Biomolecular Analysis Facility – Metabolomics Node (NBAF-B), School of Biosciences, University of Birmingham, Edgebaston, B15 2TT Birmingham, United Kingdom
j.engel@bham.ac.uk

Experimental designs of metabolomics experiments often comprise factors such as concentration of drug or toxicant, time, and/or gender or age of the study organism. Recently, regularized MANOVA (rMANOVA) was proposed for the analysis of such experiments [1]. The model is a weighted average of the ANOVA simultaneous component analysis (ASCA) and MANOVA models and contains these techniques as special cases. The optimal weight is determined in a data driven fashion employing a Stein-type shrinkage estimator of the within-group covariance matrix. It was shown that rMANOVA outperforms ASCA and MANOVA under many scenarios. However, only data with hundreds of variables such as binned ^{1}H-NMR spectra were considered.

The present work focuses on extending the application of rMANOVA to mass spectrometry data typically with thousands of variables (peaks). Efficient computation becomes necessary, and changes in the rMANOVA algorithm are introduced that considerably reduce the numerical load and speed up the calculations. Additionally, a false discovery rate procedure is introduced to assign significance to the variables. Simulated and real data examples are used to study the performance of the modified rMANOVA approach for different numbers of samples and variables. Performance is assessed with respect to proper discrimination between the groups of conditions for a specific factor and identification of associated key metabolites. We conclude that rMANOVA is a highly promising method for analysis of multi-factor MS metabolomics data.

References:

[1] J. Engel, L. Blanchet, B. Bloemen, L.P. van den Heuvel, U.H.F. Engelke, R.A. Wevers, L.M.C. Buydens, Analytica chimica acta, 899, 1 (2015).

O-22

LIKELIHOOD RATIO MODELS SUPPORTED BY CHEMOMETRIC TOOLS – RECENT MODELS OVERVIEW

G. Zadora[1,2], A. Martyna[1,3], A. Michalska[1], P. Własiuk[1,3]

[1] Institute of Forensic Research in Krakow, 9 Westerplatte, Krakow, Poland,
[2] Institute of Chemistry, University of Silesia in Katowice, 9 Szkolna, Katowice, Poland
[3] Faculty of Chemistry, Jagiellonian University in Krakow, 3 Ingardena, Krakow, Poland,
gzadora@ies.krakow.pl

The increasing complexity of new forms of crime and the need by those who administer justice for higher standards of scientific work require development of new approaches for measuring the evidential value of physicochemical data of trace evidence recorded by application of numerous analytical methods . The methods used for evaluation of these data should reveal the role of the forensic experts in the administration of justice. This means that such data (evidence; E) should be evaluated in the context of two competing propositions H1 and H2 formulated by two opposite sides in the legal proceeding, i.e. respectively prosecution and defence. Nowadays, the best method of evaluation of evidential value of physicochemical data for forensic purposes seems to be an application of likelihood ratio test ($LR=Pr(E|H1)/Pr(E|H2)$) in so-called comparison and classification problem [1].

Till now, the LR models have been applied for evidence evaluation only in situations when the number of objects in database (m) exceeded the number of the variables (p) describing them ($m >> p$) [e.g. 1]. The problem emerges in the case of highly multivariate data such as spectra stored in the numerical form, for which the situation is opposite ($m << p$). This is because it is indispensable to use huge database (so that $m>p$) to reliably estimate the relevant parameters of the population for multidimensional data using LR approach – like sources of errors (within and between object variability) and rarity of determined physicochemical features. Therefore, in the aim to cope with this inconvenience new models/strategies for data interpretation have been proposed by authors. They benefit from the best features of chemometric tools (reduction of data dimensionality and selection of the best variables) and likelihood ratio tests (expressing in a probabilistic way the support of physicochemical data for particular hypothesis and taking into account in one calculation run important factors like sources of variability and rarity of physicochemical data). The correctness and effectiveness of the proposed models could be verified by estimating the rates of false positive and false negative answers and, additionally, by using the Empirical Cross Entropy approach [e.g. 1].

The aim of this presentation is to illustrate the overview of strategies which combine LR test with information obtained from chemometric tools like principal component analysis (e.g. determination of the geographical origin of olive oil [2] or wines [3]), wavelet transformation (e.g. comparison problem of blue car paint samples analysed by Raman spectroscopy or polypropylene samples analysed by infrared spectroscopy [4]), linear discriminant analysis (e.g. determination of the geographical origin of olive oil [2] and evaluation of elemental composition of glass samples) or distance representation (e.g. comparison problem of infrared or Raman spectra as well as pyrograms of samples collected from car bodies).

Acknowledgement: The experiments were undertaken within the National Science Centre in Poland project (Preludium 6 no. 2013/11/N/ST4/01547) and the Institute of Forensic Research project VI K/2013-15 and IV K/2015-17.

References:

[1] G. Zadora, A. Martyna, D. Ramos, C. Aitken, Chichester 2014: John Wiley & Sons (ISBN 978-0470972106)
[2] P. Własiuk, A. Martyna, G. Zadora, Anal Chim Acta, 853, 187 (2015).
[3] A. Martyna, G. Zadora, I. Stanimirova, D. Ramos, Food Chem, 150, 287 (2014).
[4] A. Martyna, A. Michalska, G. Zadora, Anal Bioanal Chem, 407, 3357 (2015).

O-23

A PAIRWISE LOG-RATIO METHOD FOR THE IDENTIFICATION OF BIOMARKERS

J. Walach [1], P. Filzmoser [1], K. Hron [2], B. Walczak [3]

[1]*Institute of Statistics and Mathematical Methods in Economics, Vienna University of Technology, Wiedner Hauptstrasse 8-10, 1040 Vienna, Austria*
[2]*Department of Mathematical Analysis and Application of Mathematics, Palacky University, 17. listopadu 12, 77146 Olomouc, Czech Republic*
[3]*Department of Analytical Chemistry, Institute of Chemistry, University of Silesia, Szkolna 9, 40-006, Katowice, Poland*

jan.walach@tuwien.ac.at

One of the main goals in metabolomics is the identification of biomarkers – metabolites which are capable of distinguishing between groups of, e.g., healthy and unhealthy patients. There are various methods for identifying biomarkers in the statistical field. Difficulties arise by facing the so-called „size effect", which occurs due to different sample volume or concentration. In that case, the true signal is hidden in the data structure, and it can be revealed only after a special treatment. One possibility is to normalize the data first, other possibilities include certain transformations, see e.g. [1].

Here we propose a method that makes use of the log-ratio approach [2]. We use the elements of the variation matrix, which are defined as the variance of $\log(x_i/x_j)$, for all pairs of variables x_i and x_j . The advantage of log-ratios is that the absolute concentration is irrelevant, which is appropriate in this context. The variation matrix is computed for the joint data, as well as for the single groups separately. A statistic is then constructed, involving all three sources of information. Since the distribution of the statistic is unknown, we use the bootstrap technique; biomarkers are then considered as variables where most of their pairwise log-ratios are significantly different.

The method has been tested on simulated data as well as on real data sets. The simulations have been carried out according to the scheme outlined in [1]. In both the low-dimensional (9 variables) and the high-dimensional (500 variables) situation, the new proposal shows excellent behavior with respect to the true positives, false discovery and false negative rates. These simulations reveal slight advantages over PQN normalization, the method which turned out in [1] as the best among all considered options. The new pairwise log-ratio method has the big advantage that it can easily be robustified against outliers in the data, by simply using a robust estimator of the variance.

Acknowledgement: This work is supported by the Austrian Science Fund (FWF), project I 1910-N26.

References:

[1] P. Filzmoser, B. Walczak, What can go wrong at the data normalization step for identification of biomarkers?, Journal of Chromatography A, Vol. 1362, p. 194 (2014).
[2] V. Pawlowsky-Glahn, J.J. Egozcue, R. Tolosana-Delgado, *Modeling and Analysis of Compositional Data*, Wiley, Chichester, UK (2015).

O-24

NEW INSIGHTS IN AUTHENTICATION IDENTIFICATION OF EDIBLE OILS BASED ON METABOLOMICS AND CHEMOMETRICS

L.X. Zhang[1,2,4], P.W. Li[1,3,4,5]

1 Oil Crops Research Institute, Chinese Academy of Agricultural Sciences, Wuhan, China
2 Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture, Wuhan, China
3 Key laboratory of Detection for Mycotoxins, Ministry of Agriculture, Wuhan, China
4 Laboratory of Risk Assessment for Oilseeds Products (Wuhan), Ministry of Agriculture, Wuhan, China
5 Quality Inspection and Test Center for Oilseeds Products, Ministry of Agriculture, Wuhan, China
liangxiao_zhang@hotmail.com

Edible vegetable oils are a kind of important food in our daily life，ensuring the authenticity of food has been a problem for millennia. As the same as adulteration of olive oil in western countries, adulteration in other high-price oils is still the biggest source of agricultural fraud problems. Therefore, it is a great demand for a reliable method to detect such adulterations. The authenticity identification model was built by the one-class partial least squares (OCPLS) classifier for edible oils. Subsequently, the established model was validated by independent test sets. Moreover, counterfeit oils adulterated with different levels of other edible oils were simulated by the Monte Carlo method and employed to test the lowest adulteration level of this one-class classifier. According to MCCV with a sampling ratio of 0.8, when the number of significant OCPLS components was set to 8 ($\delta 2 = 10$), the lowest standard deviation of residual of cross validation was obtained. The samples in training set by OCPLS with the fatty profiles of pure peanut oils could be completely identified. The counterfeit oils with different adulteration levels (3%-12%) were simulated to check the lowest adulteration level of this model. As shown in Figure 1, the OCPLS model just misidentified 3 adulterated peanut oils as authentic peanut oils at the adulteration level of 3%, indicating that the accuracy rate of this model equals 92.5% (37/40) for the adulterated peanut oils with the adulteration level of 3%. When the adulteration level is higher than 3%, this OCPLS model could completely detect these adulterated peanut oils. Thus, the lowest adulteration level of this OCPLS model is set to 4%.

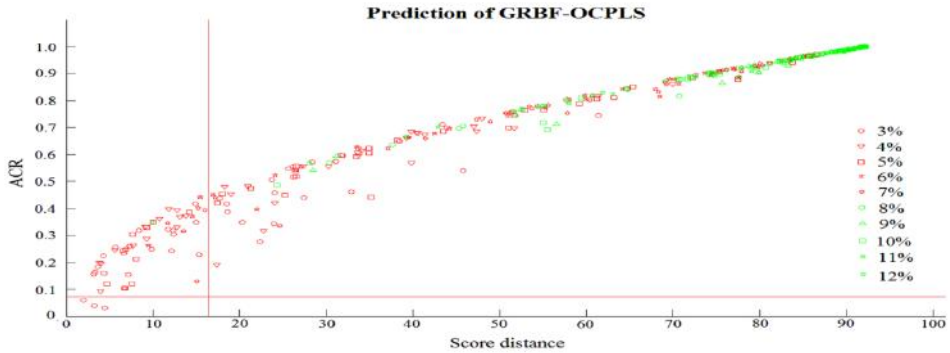


Figure 1 Prediction for adulterated peanut oils with the adulteration levels of 3%-12%

Acknowledgement: This work was supported by the Project of National Science & Technology Pillar Plan (2012BAK08B03), the National Natural Science Foundation of China (21205118), and the earmarked fund for China Agriculture research system (CARS-13).

References:

[1] Zhang LX, Li PW, Sun XM, Mao J, Ma F, Ding XX, Zhang Q, One-class classification based authentication of peanut oils by fatty acid profiles. RSC Adv. 2015, 5, 85046-85051.
[2] Zhang LX, Li PW, Sun XM, Wang XF, Xu BC, Wang XP, Ma F, Zhang Q, Ding XX. Classification and Adulteration Detection of Vegetable Oils Based on Fatty Acid Profiles. J. Agric. Food Chem.2014, 62(34):8745-8751.

O-25

PROCESSING AND ANALYSING NANOSCOPY IMAGING DATA

C. Ruckebusch[1], S. Hugelier[1], O. Devos[1], M. Sliwa[1], P.Dedecker[2], J.J. de Rooi[2], P. H. C. Eilers[2]

[1.] Université de Lille, LASIR CNRS UMR 8516, F-59000 Lille, France;
[2.] Department of Chemistry, KU Leuven, Belgium;
[3.] Erasmus MC, Department of Biostatistics, Rotterdam, the Netherlands.
Cyril Ruckebusch@univ-lille1.fr

Super-resolution wide-field fluorescence microscopy can provide structural information at the nanoscale and dynamic insight about biological processes in live cell samples. In general, the available information in super-resolution images is related to the density of emitters, with more emitters leading to more information. One of the strategies for obtaining a high spatial resolution is based on the sequential imaging and localization of sparse subsets of blinking fluorophores distributed over thousands of images (movie frames), resulting in a high-density image of their positions and intensities. However, to obtain a high spatial resolution on short time sampling, and potentially probe dynamic processes in live cells, this principle must be extended to the analysis of high-density of emitters distributed over a few tens of movies frames only. As many emitters are simultaneously active, their emissions strongly overlap and single-emitter fitting methods collapse. Thus, analyzing high-density super-resolution data, the development of new methods and image processing algorithms remains a topical and challenging issue for dynamic imaging and faster super-resolution experiments.
The cornerstone of high-density super-resolution single-molecule fluorescence imaging is data analysis and involves several steps ranging from preprocessing,[1] e.g. image background handling, to post-processing, e.g. image rendering, clustering, etc. The core of our approach is the SPIDER algorithm for SParse Image DEconvolution and Reconstruction.[2,3] This approach tackles the image deconvolution problem in a penalized regression framework by implementing an approximation of the L_0-norm.
In this talk, we present the processing and analysis approaches developed, which can be relevant to analyze other types of image time series. We illustrate the result obtained on different data, including protein-labeled mitochondria in a HEK293-T cell observed in wide-field fluorescence microscopy. On this example we demonstrate a resolution of 50 nm over a time sampling as short as 0.5 s. On overall, we show that we can work with higher densities of active emitters, for faster imaging, allowing the investigation of highly dynamic structural and morphological changes that membrane organelles undergo to accommodate the continuous processes occurring in live cells.

References:

[1] Ruckebusch, C.; Bernex, R.; Allegrini, F.; Sliwa, M.; Hofkens, J.; Dedecker, P. Anal. Chem., Anal. Chem. 87, 4675-4682, 2015. [2] De Rooi, J.; Ruckebusch, C.; Eilers, P.H.C, Anal. Chem., 86, 6291 – 6298, 2014. [3] Hugelier, S., De Rooi, J.J., Bernex, R., Duwé, S., Devos O., Sliwa M., Dedecker,P. Eilers, P.H.C., Ruckebusch, C., Sci. Reports, tbp

O-26

COMPARISON OF MULTIVARIATE METHODS ON RAMAN IMAGES OF PLANT CELL WALLS

B. Prats-Mateu[1], N. Gierlinger [1,2,3]

[1] Department of Material Sciences and Process Engineering, BOKU-University of Natural Resources and Life Sciences, Peter-Jordan-Straße 82, 1190 Vienna, Austria
[2] Institute for Building Materials, Eidgenössische Technische Hochschule Zurich, Hönggerberg, 8093 Zurich, Switzerland
[3] Applied Wood Research laboratory, Empa-Swiss Federal Laboratories for Material Testing and Research, Überlandstrasse 129, 8600 Dübendorf, Switzerland

b.prats-mateu@boku.ac.at

Spectroscopic imaging techniques become more and more powerful for biologists and material scientists as data on the molecular composition of the sample can be aqcuired in context with the (micro-)structure. Not to be overwhelmed with the huge amount of data generated by modern imaging techniques like Infrared or Raman microscopy, multivariate data analysis plays an important role by extracting the most relevant information and revealing important changes, relationships, trends and/or gradients [1]. Confocal Raman imaging offers a non-destructive and relatively quick characterization of biological samples with a spatial resolution limited by the diffraction of light [2]. Mapping a sample with the best achievable spatial resolution (~300nm) means that thousands of spectra (each representing around 3,800 wavenumbers) are acquired even if only one cell (~20 x 20 µm) is measured. With thousands of spectra and wavenumbers in the image it is necessary to compress the data to be able to describe the sample in a more simple meaningful way. Biological samples usually do not give Raman spectra with only a few sharp bands, but show many overlapping broad features representing the multicomponent nature. It is obvious that most of the pixels do not reflect pure components but mixtures of 2 or more components. Due to the overlapping bands and multicomponent nature band intensity and integration give only in a few cases good results and multivariate approaches are necessary to extract a maximum of information. Depending on the research questions and method constraints it is essential to choose the right multivariate approach, but also to compare the results based on different methods to reveal the most meaningful facts or relationships.
Different types of unmixing algorithms to reveal the most pure components independent of the „maximum explained variance constraint" have been developed in the last decade including "Vertex Component Analysis" (VCA) [3], Non-negative Matrix Factorization (NMF) [4] or Multivariate Curve Resolution Alternating Least Squares (MCR-ALS) [5]. The potential of Raman Imaging [6] and VCA has been succesfully exemplified in the mature plant cell wall [7]. We present the comparison between these different algorithms for analysing hyperspectral data generated on plant cell walls and the standard univariate approach *e. g.* band integration. We demonstrate the power of multivariate methods in combination with Raman microscopy to elucidate and unravel hidden features in the complex and highly variable plant cell wall.

Acknowledgement: The authors thank the Austrian Science Fund (FWF) (Project Y-728-B16) for financial support.

References:
[1] P. Geladi, H. Grahn, M. John Wiley & Sons, Inc., 93 (2010).
[2] W.I.O. Hollricher in: O.H. T. Dieing, J. Toporski (Eds.) (Ed.) Confocal Raman Microscopy (2010).
[3] J.M. Nascimento, J.M.B. Dias, IEEE Transactions on 43, 898 (2005).
[4] D.D. Lee, H.S. Seung. Nature 401, 788 (1999).
[5] J. Jaumot, R. Garfallo, A. De Juan, R. Tauler. Chem. Intell. Lab. Sys. 76, 101 (2005).
[6] N. Gierlinger, T. Keplinger, M. Harrington. Nat Protoc 7, 1694 (2012).
[7] N. Gierlinger. Front Plant Sci 5, 306 (2014).

O-27

MOLECULAR MONITORING OF BIOMEDICAL SAMPLES BY RAMAN IMAGING AND CHEMOMETRICS

M. Marro[1], P. Loza-Alvarez[1]

[1] ICFO-Institut de Ciencies Fotoniques, Av. Carl Friedrich Gauss, 3, 08860 Castelldefels, Barcelona – Spain
monica.marro@icfo.es

Raman spectroscopy (RS) is non-invasive, rapid and sensitive. Thus, RS represents a promising technique for studying biomedical samples. However, although RS has the maximum specificity among all optical techniques for detecting molecular changes, the interpretation of Raman spectra is complex. During the past decades applications of Raman spectroscopy have been focused to discriminate several groups of samples by means of multivariate analysis. However, little information can be obtained from those methods to extract meaningful molecular components from the Raman spectra that could be assigned to pure molecules constituting the sample and monitor them during a biochemical process. For this reason, we proposed to apply Multivariate Curve Resolution (MCR) to deconvolve pure molecular components from the Raman spectra and monitor its content in the tissue or cell over the illness or biological process under study. MCR requires minimal a priori knowledge of the system providing objective and chemically meaningful information.

We present several successful biomedical applications of our approach in different fields like neuroinflammation, cancer or food analysis. First, retinal tissue is damaged during inflammation in Multiple Sclerosis. We assess molecular changes in murine retinal cultures suffering inflammation by RS [1]. By using MCR analysis, we deconvolved 6 molecular components suffering dynamic changes along inflammatory process. Those include the increase of immune mediators, changes in molecules involved in energy production and decrease of Phosphatidylcholine. Following this work, individual retina cell lines are studied inducing different challenges to investigate which metabolites are synthetized in each cell line. RS combined with MCR allows monitoring the evolution of retina inflammation based in a number of molecular components sensitive to inflammation.

Second, we study the molecular composition of cancerous tissue and healthy tissue by Raman imaging, identifying regions that are compositionally changing during the previous steps of malignancy. This, combined with our previous work on cells studying the Epithelial to Mesenchymal transition (EMT) [2] shows that our approach permitted to deconvolve and track biomarkers for cancer aggressiveness and prognosis.

Finally, applications in food monitoring as the tomato ripening imaging will be shown.

Thus, the combination of Raman imaging of biomedical samples and MCR represents a novel methodology that will push forward the applicability of RS for non-invasive monitoring of the biochemical content in vivo.

Acknowledgement: We thank Dr. Anna de Juan, for the interesting discussions about the applications of MCR-ALS. We also thank Dr. Pablo Villoslada, Dr. Angels Sierra, Dr. Gemma Fuster and Lucio Andrea for the fruitful collaborations on the biological applications.

References:

[1] M. Marro, A. Taubes; et al, Dynamic molecular monitoring of retina inflammation by in vivo Raman spectroscopy coupled with multivariate analysis, J Biophotonics,7(9), 725-734 (2014).

[2] M. Marro, C. Nieva, R. Sanz-Pamplona, A. Sierra, Molecular monitoring of epitelial-to-mesenchymal transition in breast cancer cells by means of Raman spectroscopy, Biochimica et Biophysica Acta, 1843(2014), 1785-1795.

O-28

DETECTION OF COUNTERFEITED PHARMACEUTICAL TABLETS WITH NEAR INFRARED SPECTROSCOPY AND CHEMOMETRICS

K. Dégardin [1], Y. Roggo [2]

[1]Complaint and Counterfeit, F. Hoffmann-La Roche, Wurmisweg, Kaiseraugst, Switzerland, klara.degardin@roche.com .
[2] Complaint and Counterfeit, F. Hoffmann-La Roche, Wurmisweg, Kaiseraugst, Switzerland

Counterfeiting is a crime with dreadful consequences, especially in the case of medicines. All types of counterfeits can be found, from the ones devoid of active pharmaceutical ingredient (API) to under dosed medicines. Organised criminal networks have proven to be behind the production and selling of counterfeit medicines. Fast and reliable analyses are consequently necessary to confirm the cases and evaluate the risk encountered by the patients. Near Infrared spectroscopy (NIRS) presents many advantages for that purpose. It is indeed a fast, non-destructive method, that provides chemical and physical information about the analysed samples. Thanks to chemometric tools, the chemical and physical signature of a suspect sample can be rapidly compared to the genuine references, providing a fast yes/no answer. The difficulty in the NIRS analyses resides in the computation of the right models for a correct and easy authentication. Several techniques have been published so far for counterfeit identification of tablets and capsules. The methods that will be presented consist of the NIR identification of all the tablets produced by Roche, which represents 30 pharmaceutical products, each of them possessing at least one dosage. The novelty of these results resides especially in a huge dataset, composed of more than 7000 spectra. The spectra were first investigated by Principal Component Analysis (PCA) and then different supervised techniques, namely the Support Vector Machines (SVM) (SVC kernel linear and Radial Basis Function), the K-Nearest Neighbors (KNN) and the Discriminant Analysis (DA). The different chemometric models were validated and then tested against each other product, and additionally against 94 counterfeits, generics and placebos as challenging samples. Additionally NIR handheld spectrometers have been evaluated in this study for the analysis of counterfeited tablets. The use of handheld instruments would enable the detection of counterfeits on the field. These results will also be presented and compared with the data obtained with the lab instruments.

O-29

PROCESSING LARGE TOF-SIMS DATASETS FOR THE STUDY OF SURFACE SEGREGATION OF POLYMER ADDITIVES

G.F. Trindade, M.L. Abel, J.F. Watts

The Surface Analysis Laboratory, Department of Mechanical Engineering Sciences, University of Surrey Guildford, Surrey, GU2 7XH, UK.

g.ferraztrindade@surrey.ac.uk

All commercial polymers are influenced by the surface segregation of minor components. The presence of such molecules at the surface of a polymer can have a deleterious effect on performance, such as adhesion, biocompatibility and other surface-active processes. These layers are often only a few nanometres in thickness and are therefore not suitable for analysis by conventional chemical methods. The solution is to use a surface specific method such as time-of-flight secondary ion mass spectrometry (ToF-SIMS). ToF-SIMS is based on the detection of ionised molecules, molecular fragments or atoms generated by sputtering as a result of the bombardment of primary ions onto the surface of the sample to be analysed. Only the fragments from the first few monolayers of the sample will have sufficient energy to overcome the surface binding energy and leave the sample [1].

Over the last twenty years, the use of multivariate methods has increased significantly within the SIMS community. This is mainly due to the constant development of the technique towards better mass, lateral and depth resolutions. State-of-art spectrometers in dual beam depth profile mode will typically generate datasets distributed throughout a 3D cube containing 256x256x500+ voxels with each voxel containing 1,000,000 spectral channels (resulting in more than 3.2×10^{13} data points). Cumpson et. al. have recently proposed a method for computing the SVD of large datasets using a random vectors algorithm, enabling PCA of such huge datasets to be done in conventional computers [2]. Another way to process the data with no loss in resolution is by means of low discrepancy subsampling, since ToF-SIMS data have a fundamental volume in which voxels will be highly correlated [3].

This work presents the analysis of ToF-SIMS data of automotive grade polypropylene under heating conditions. Such copolymers are filled with particles and a number of processing aids and other additives are used to improve behaviour and performance [4]. The samples were mounted on a special sample holder that allows temperature control in ultra-high vacuum, enabling *in situ* ToF-SIMS analysis. The data is stored in a 3D cube in which each layer represents a point in time containing a 500x500 μm^2 image with a full mass spectrum in each pixel (as shown in Fig. 1). PCA and MCR results enabled insights on the diffusion mechanism of molecules that segregate to the surface under high temperatures and interfere in surface activation processes such as flame treatment [5].

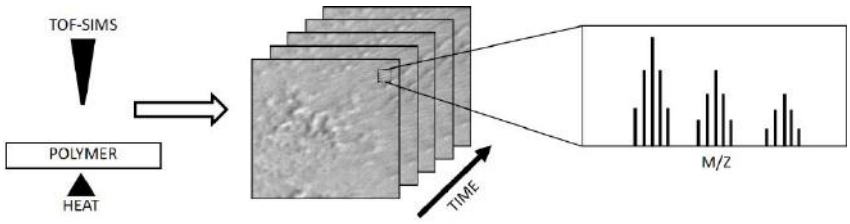


Figure 1: Experiment description.

References:

[1] J. F. Watts, J. Wolstenholme, R. Webb. Secondary Ion Mass Spectrometry. University of Surrey, 2012.
[2] P. J. Cumpson, N. Sano, I. W. Fletcher, J. F. Portoles, M. Bravo-Sanchez, A. J. Barlow. Surf. Interface Anal., 47, 986 (2015).
[3] P. J. Cumpson, I. W. Fletcher, N. Sano, J. Barlow. Surf. Interface Anal., in press (2016).
[4] S. Moritomi, T. Watanabe, S. Kanzaki. R&D Report, Sumitomo Kagaku, 1 (2010).
[5] D. Williams, M-L. Abel, E. Grant, and J. F. Watts. Int. J. Adhes. Adhes., 63, 26 (2015).

O-30

ENSEMBLE MONTE CARLO VARIABLE SELECTION FOR METABOLITE IDENTIFICATION IN NMR DATASETS

C. Esquerre, A.A. Gowen, C. O'Donnell

UCD School of Biosystems and Food Engineering, University College Dublin, Belfield, Dublin, Ireland.
carlos.esquerre@ucd.ie

The aim of this study was to investigate the potential of ensemble Monte Carlo Variable Selection (EMCVS) [1] to identify the relevant portions of ^{1}H NMR spectra suitable for use as a metabolite fingerprinting tool. Datasets from an experiment on mushroom (*Agaricus bisporus*) quality and two publically available datasets, namely urine samples of rats [2] and wine samples [3] were examined. NMR spectra of physically damaged and sound mushrooms were obtained and variable selection using EMCVS was carried out in order to (a) discriminate between mushrooms that had been subjected to low levels of damage and those that were undamaged and (b) to investigate the chemical changes induced by low levels of mechanical damage. NMR spectra from urine samples of rats fed with different concentrations of onion by-products to identify dietary biomarkers, and from a wine dataset employed to improve the accuracy of calibration models for ethanol, glycerol, lactic acid, methanol and malic acid content. EMCVS was compared to other PLS based variable selection methods as Variable Importance on Projection, Selectivity Ratio [4] and Significance Multivariate Correlation [5]. In each dataset, EMCVS selected fewer variables while maintaining a low predictions error, demonstrating its potential for metabolite fingerprinting.

Acknowledgement: Dr. Carlos Esquerre acknowledges funding from the EU FP7 under the Marie Curie Outgoing International Fellowship.

References:

[1] Esquerre, C. Gowen A.A. Downey, G. O'Donnell, C.P., et al., Journal of Near Infrared Spectroscopy, 19, 443 (2011).
[2] Winning, H. Roldan-Marin, E. Dragsted, L. O., Viereck, N., Poulsen, M., Sanchez-Moreno, C., Engelsen, S. B. Analyst, 134, 2344 (2009).
[3] Larsen, F.H. van den Berg, F. Engelsen, S.B. Journal of Chemometrics, 20, 198 (2006).
[4] Kvalheim, O.M. Journal of Chemometrics, 24, 496 (2010).
[5] Tran, T. N. Afanador, N. L. Buydens, L. M. C. Blanchet, L. Chemometrics and Intelligent Laboratory Systems, 138, 153 (2014).

O-31

Accurate anisotropy mapping of Perylene obtained after Rayleigh scattering correction

Y. Casamayou-Boucau and A.G. Ryder

Nanoscale Biophotonics Laboratory, School of Chemistry, National University of Ireland, Galway, Galway, 1reland,
y.casamayou-boucau1@nuigalway.ie

Anisotropy resolved multidimensional emission spectroscopy (ARMES) is a potentially useful analytical technology platform for the rapid, in-situ, non-contact analysis of multi-fluorophore macromolecules such as proteins [1]. The method combines anisotropy and multidimensional fluorescence measurements with factor-based chemometrics like Multivariate Curve Resolution (MCR). In order to apply ARMES to intrinsic protein fluorescence, it was critical to be able to collect polarized fluorescence data in the $\lambda_{ex/em}$=250-300 nm region and also to have a stable standard for method validation. This UV requirement made the commonly used thin film polarizers (TFP) unsuitable as they absorb below 290 nm. Instead, we used wire grid polarizers (WGP) because of their better UV transmission even if they have a poorer extinction ratio (~1:10) in this region.

Perylene, a very well know anisotropic rotor, was selected as a standard for ARMES because it is chemically stable and has two emission bands: a short wavelength ($\lambda_{ex/em}$ ~ 254 nm/440 nm) and a long wavelength ($\lambda_{ex/em}$ ~ 434 nm/440 nm). These correspond to two electronic states with orthogonal excitation dipoles, generating very different anisotropies [2]. Glycerol solutions of 1µM Perylene were analyzed in triplicate at 25°C, over a wide wavelength range by excitation-emission matrix (EEM) and total synchronous fluorescence scan (TSFS) using both WGP and TFP fittings. For the WGP, there was no spectral reshaping, and because of the non tri-linearity of TSFS data [1], we could not recover reliable components by MCR. With EEM data, the issue of the Rayleigh scattering inducing distortion of anisotropy measurements is very serious. In the case of Perylene, the scatter overlapped the most intense emission band and thus interpolation based correction was not accurate. Another solution was to use the algorithm developed by Rinnan *et al*. [3], which shifted the Rayleigh scattering into low-rank bilinear data. This allowed modelling scatter as a separate set of components after which it was subtracted from the original data to enable accurate MCR modelling of the fluorescence emission.

Using this method, Rayleigh scatter was fully removed and accurate anisotropy values were recovered. Using EEM-TFP data for the low energy band a constant anisotropy of 0.220 ± 0.002 was obtained (single measurement dispersion of 3.7%), that was comparable to the equivalent TSFS data: 0.219 ± 0.003 (single measurement dispersion of 3.2%). Those results are also very similar to the literature values (0.22) obtained under similar conditions, using Glan-Thompson polarizers [4]. With a single WGP, r = 0.181 ± 0.004 (4.5%) was obtained with EEM and 0.182 ± 0.007 (4.4%) with TSFS. The use of a double WGP improved the extinction ratio, which increased the anisotropy value to 0.204 ± 0.002 (3.5% variation) and 0.203 ± 0.001 (3.7%). This was closer to other literature sources (0.197 ± 0.002) [2] and the variances can be explained by the different relative efficiencies of the various polarizers. This now enabled the accurate recovery of anisotropy values for individual MCR factors in multi-fluorophore mixtures.

Acknowledgement: This research was supported by an Irish Research Council postgraduate award to YCB and the loan of a fluorescence spectrometer by Agilent Technologies.

References:

[1] R.C. Groza, B.Y. Li, A.G. Ryder, Analytica Chimica Acta, 886, 133 (2015).
[2] M.D. Barkley, A.A. Kowalczyk, L. Brand, The Journal of chemical physics, 75, 3581 (1981).
[3] Å. Rinnan, K.S. Booksh, R. Bro, Analytica Chimica Acta, 537, 349 (2005).
[4] J.B.-A. Johansson, Journal of the Chemical Society, Faraday Transactions, 86, 2103 (1990).

O-32

A GEOSTATISTICAL APPROACH FOR EVALUATING PREDICTION MAPS FROM HYPERSPECTRAL IMAGES

A. Herrero-Langreo[1], B. Tisseyre[2] , N. Gorretta[1] , M. Nouri[1], R. Bendoula[1], J.M. Roger[1]

[1]. UMR-ITAP, Irstea. 361 rue Jean-François Breton BP 5095, 34196 Montpellier Cedex 5, France
[2] . UMR ITAP, Montpellier SupAgro, Bât. 21, 2 Place Viala, 34060 Montpellier, France
anahelangreo@gmail.com

Introduction: Applying a calibration model onto hyperspectral (HS) images is of great interest because it produces images of chemical or physical properties. Usually, the main difficulty of this process is that such models need to be calibrated or transferred, and so requires both image spectra and corresponding response values. Furthermore, these response values, also known as "ground truth", are also necessary for evaluating the quality of the estimations on the predicted images. Measuring ground truth can be difficult, either because of time, cost or practical constraints. Moreover, in laboratory images taken at high spatial resolution, the pixel size of the HS images is usually much smaller than the minimum sample volume needed to obtain a wet chemical measurement. In such circumstances, there are two common approaches: to calibrate the model using the mean spectra of the imaged samples or to bypass the model calibration with resolution methods, such as MCR. In some cases, databases relating laboratory point spectra and the response to estimate already exist or they are easier to acquire than a database based on HS imaging. An appropriate alternative in this scenario would be to transfer those model calibrations to the HS images. Once a model calibration is available and applied to the HS images, a second problem arises: how to evaluate the performance of these models on the image when ground truth is not available for each pixel?

The objective of this work is to propose and test an approach to evaluate model estimations obtained from HS images. This procedure takes into account the spatial distribution of the estimated values as well as their mean ground truth value.

Materials and methods: A predictive model built from a laboratory database of 934 point spectra was transferred and applied to HS images of sugar beet slices. Calibration transfer was done through a model update, introducing 112 mean spectra from the HS images on the model calibration. Images of soluble solids content estimations were obtained as a result. The number of latent variables considered in the model was modified from 2 to 18 in order to provide a range of variation regarding the model performance. Finally, a two-step procedure was tested for the evaluation of the estimations response images:

1) A geostatistical approach to evaluate the structure of the image. This procedure can be applied, without using ground truth values, to test if the spatial distribution of the estimations is coherent with the nature of the measured samples,

2) The bias of the model was then tested using mean ground truth values of the whole sample on the image.

Results and conclusions: Geostatistical indexes, extracted from the semivariograms of prediction maps, were found related to the number of latent variables used in the model as well as to the standard deviation of the estimations. Particularly, the nugget effect was related to the RMSEP and the number of latent variables of the model. As an original contribution, the presented approach allowed to evaluate not only average estimations, but also the spatial structure of prediction maps obtained from hyperspectral images.

O-33

Non-Parametric Statistics to Assess Exposure to Deployed Military Personnel to PAHs and PCDDs/PCDFs

P.K. Hopke,[1] T.M. Mallon,[2] M.J. Utell[3]

[1]Center for Air Resources Engineering and Science, Clarkson University, Potsdam, NY 13699 USA, phopke@clarkson.edu
[2]Uniformed Services University of the Health Sciences, Bethesda, MD 20814
[3] Departments of Medicine and Environmental Medicine, University of Rochester Medical Center, Rochester, NY 14642

The objectives of this work were: 1) to determine if polycyclic aromatic hydrocarbons (PAHs), polychlorinated dibenzo-p-dioxins (PCDDs) and dibenzofurans (PCDFs) could be detected in very small human serum samples and 2) to determine if these PAH and PCDD/PCDF concentrations could serve as markers of exposure to environmental hazards such as burn pits and motor vehicle exhaust. Using a cloud point extraction method, 100 µL samples of serum collected from 400 persons, PAHs and PCDDs/PCDFs were then analyzed in the extracts using GC/MS methods. Of the 400 individuals, 200 had deployed to Iraq or Afghanistan (CASE) and the rest had not deployed to a combat zone (CONTROL) Samples were collected pre- and post-deployment. Naphthalene was found in 661 of the 800 samples and was statistically different in post-deployment personnel relative to pre-deployment. A 2-factor ANOVA on ranks found the CASE values were different than the CONTROL values ($p<0.001$) and pre- and post-deployment samples were different ($p<0.001$). Using paired rank sums tests, the pre- and post-deployed CASE samples were statistically different. The pre- and post-samples for the non-deployed were not different. Unpaired rank sums tests for the CASE pre-samples and CONTROL pre-samples were different while CASE post-samples were not different from the CONTROL post-samples. No other PAHs showed significant differences. For the PCDDs/PCDFs, 4 compounds (1,2,3,4,6,7,8-Heptachlorodibenzo-p-dioxin (HpCDD), Octachlorodibenzo-p-dioxin (OCDD), 1,2,3,7,8,9-Hexachlorodibenzofuran (HxCDF), and 1,2,3,4,6,7,8-Heptachlorodibenzofuran (HpCDF)) were found in 291, 330, 259, and 265 samples, respectively. The 2-factor ANOVA on ranks showed all 4 compounds having significant differences between CASE and CONTROL and between pre- and post-deployment samples. However, only 1,2,3,7,8,9-Hexachlorodibenzofuran was significantly different between pre- and post-deployment CASE samples and was not different between pre- and post-deployment CONTROL samples.

O-34

BAYESIAN SPATIAL MULTIVARIATE RECEPTOR MODELING FOR MULTI-SITE MULTI-POLLUTANT DATA

Eun Sug Park[1], Inyoung Kim[2], Shuman Tan[1], Clifford H. Spiegelman[3]

[1]Texas A&M Transportation Institute, The Texas A&M University System, College Station, TX 77843, USA
[2]Department of Statistics, Virginia Tech University, Blacksburg, VA 24061, USA.
[3]Department of Statistics, Texas A&M University, College Station, TX 77843, USA
e-park@tamu.edu

The goal of multivariate receptor modeling is to identify the major pollution sources and quantify their impacts based on ambient measurements of pollutants. This goal cannot be easily achieved because of several obstacles preceding estimation of pollution source profiles and their contributions. Often the number of major contributing sources is unknown. Also, even with the known number of sources, estimation of source profiles and contributions is hindered due to the lack of prior knowledge on appropriate identification conditions for eliminating factor indeterminacy. With the EPA Speciation Trends Network, extensive multivariate air pollution data obtained from multiple monitoring sites (multi-site multi-pollutant data) are becoming available. Although considerable research has been conducted on modeling multivariate space-time data in other contexts, there has been little research on spatial multivariate receptor models for multi-site multi-pollutant data. We present a Bayesian spatial multivairate receptor modeling (BSMRM) approach that can incorporate spatial correlations in multi-site multi-pollutant data into the estimation of source composition profiles and contributions, based on the discrete process convolution model for multivariate spatial processes. The new BSMRM approach enables predictions of source contributions at unmonitored sites as well as simultaneously dealing with model uncertainty caused by the unknown number of sources and identifiability conditions. The new approach can also provide the uncertainty estimates of predicted source contributions at any location, which was not possible in previous spatial multivariate receptor modeling. The proposed approach is applied to 24-hour ambient air concentrations of 17 Volatile Organic Compounds (VOCs) measured at nine monitoring sites in Harris County in Texas, USA, during years 2000 to 2005.

O-35

RELIABILITY ON THE PERFORMANCE IN GAS-CHROMATOGRAPHY EQUIPMENTS: LIFETIME OF LINERS

M.L. Oca [1], M.C. Ortiz [1], L.A. Sarabia [2]

[1] *Department of Chemistry, Faculty of Sciences, University of Burgos, Plaza Misael Bañuelos s/n 09001, Burgos, Spain.*

[2] *Department of Mathematics and Computation, Faculty of Sciences, University of Burgos, Plaza Misael Bañuelos s/n 09001, Burgos, Spain*

mloca@ubu.es

The analysis of any chemical substance is a complex procedure that involves a large number of steps, from the first stages of the sample pretreatment to the final measurement. Every step is in need of control in order to ensure the maintenance of the sensitivity and repeatability of the analytical equipment and the reliability of the results finally yielded. This is especially important in routine analyses, where a certain sequence for the determination of some specific compounds is regularly carried out.

Some of the variables in the analytical method can be easily controlled either by the analyst or by the software package of the equipment used; others require the design of a strategy to check if some change arises eventually, so reset will be needed. This is the case of the state of an inlet liner in a gas chromatograph. This work shows the evaluation of the performance of two different liners (one of them new, the other already used and thoroughly cleaned prior this study) over time while a migration testing of bisphenol F, bisphenol A and their corresponding diglycidyl ethers from polycarbonate glasses is conducted (bisphenol A-d_{16} has been used as internal standard). Regarding the control study, after the analysis of 216 samples (including standards at 50 µg L^{-1} of every substance, solvent blanks and system blanks), the mandatory unequivocal identification [1] of these five compounds in terms of relative ion abundances and relative retention time has been achieved thanks to PARAFAC decompositions, taking advantage of the trilinearity of GC–MS data. Next, the values of the predicted concentration of each analyte in every standard have been assessed, on one hand, by means of individual control charts, and on the other hand, using a principal component analysis carried out on the four analytes altogether. Both results have revealed the quite different behaviour of the two liners under study and the time point when the sensitivity of the equipment is drastically altered by the continuous use of the liner, so it should be replaced. This test means a greatly useful tool to monitor the validity of the estimations from a calibration model that is routinely applied in a laboratory for the quantification step.

Acknowledgement: The authors thank the financial support provided by the Ministerio de Economía y Competitividad (CTQ2014-53157-R) and FEDER funds.

References:

[1] Guidelines for performance criteria and validation procedures of analytical methods used in controls of food contact materials, EUR 24105 EN – first edition 2009.

O-36

Receptor modeling of environmental aerosol data using MLPCA-MCR-ALS

Yahya Izadmanesh [1], Jahan B. Ghasemi[2], Roma Tauler[3]

[1], Faculty of Chemistry, K. N. Toosi University of Technology, Tehran, Iran
[2], Faculty of Chemistry, University of Tehran, Tehran, Iran
[3], Institute of Environmental Assessment and Water Research, Spanish Council for Scientific Research (CSIC), Jordi Girona 18, Barcelona, 08034, Spain

Email: yizadmanesh@mail.kntu.ac.ir

Receptor models apportion the measured mass of an ambient particulate matter (PM) at a given site, called the receptor, to its emission sources by using multivariate factor analysis [1]. A general procedure for receptor modeling and source apportionment of PM data using multivariate curve resolution–alternating least squares (MCR-ALS) incorporating measurement uncertainty information is described. After pretreatment of the dataset for missing and less than detection limit values and categorization of variables based on their signal to noise ratios, maximum likelihood principal component analysis-alternating least squares (MLPCA-ALS) [2] and multivariate curve resolution-weighted alternating least squares (MCR-WALS) [3] were applied. Propagation of experimental uncertainty into MCR resolved sources was estimated by a numerical resampling method. Simulated noise was generated to be structurally similar to experimental data uncertainty. and added multiple times to the dataset [4]. Comparison among MLPCA-ALS, MCR-WALS and positive matrix factorization (PMF) results confirmed the reliability of the proposed method. Using MLPCA- ALS over MCR-WALS has the advantage of using traditional MCR-ALS after MLPCA as a preliminary data pretreatment. Because of existence of intrinsic selectivity in the measured chemical species, rotation ambiguities [4] were rather limited in this case. Results are shown for the analysis of a particulate matter (PM10) data set where main PM pollution sources were found to be secondary inorganic aerosols and traffic, as well as sea salt, fossil oil burning, biogenic and biomass burning.
Keywords: Receptor modeling; Multivariate curve resolution-alternating least squares, Measurement uncertainty.

References

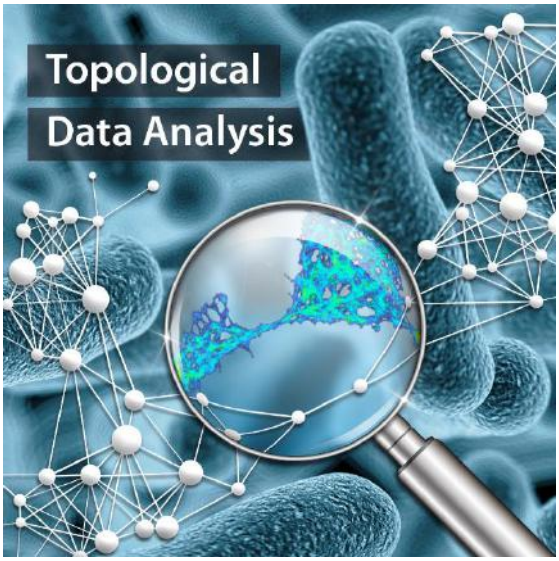
[1] B.R.L. Claudio A. Belis, Fulvio Amato, Imad El Haddad, Olivier Favez, Roy M.Harrison, Philip K. Hopke, Silvia Nava, Pentti Paatero, André Prévôt, Ulrich Quass, Roberta Vecchi, Mar Viana, EUROPEAN GUIDE ON POLLUTION SOURCE IDENTIFICATION WITH RECEPTOR MODELS.pdf, Publications Office of the European Union, 2013.
[2] M. Dadashi, H. Abdollahi, R. Tauler, J. Chemomet., 27 (2013) 34-41.
[3] M. Dadashi, H. Abdollahi, R. Tauler, Chem. Intel. Lab. Sys., 118 (2012) 33-40.
[4] J. Jaumot, R. Gargallo, R. Tauler, J.Chemomet., 18 (2004) 327-340.

O-37

EXPLORING BACTERIA STRAINS WITH RAMAN SPECTROSCOPY AND TOPOLOGICAL DATA ANALYSIS

M. Duponchel

Université de Lille, LASIR Lab, 59655 Villeneuve d'Ascq Cedex, FRANCE.
ludovic.duponchel@univ-lille1.fr



An important feature of analytical chemistry is that data of various kinds is being produced at an unprecedented rate. This is mainly due to the development of new instrumental concepts and experimental methodologies. It is also clear that the nature of the data we are acquiring is significantly different. Indeed in chemometrics, we are given data in the form of always longer vectors, where all but a few of the coordinates turn out to be irrelevant to the questions of interest, and further that we don't necessary know which coordinates are the interesting ones. "Big data" in chemometrics is a future that might be closer than any of us suppose. It is in this sense that new tools have to be developed in order to explore and valorize such datasets. Topological data analysis (TDA) is probably one of these [1]. The purpose of this work is first to introduce this method and second to explore a Raman data set of bacteria strains (4000 single bacteria). In parallel with conventional chemometrics tools (i.e. PCA and HCA), TDA will be evaluated considering various experiment conditions and data structures. We will focus on the effect of spectral pre-processing, signal to noise ratio, spectral shift and spectral resolution. Indeed it is very important to study behaviors of methods because looking at big data sets a little closer, we often note that they consist of merged experiments (and even sometimes acquired with different methodologies or platforms) which have theses data structures.

References:

[1] Extracting insights from the shape of complex data using topology, P. Y. Lum, G. Singh, A. Lehman, T. Ishkanov, M. Vejdemo-Johansson, M. Alagappan, J. Carlsson & G. Carlsson, Scientific Reports, 3, 1236, doi:10.1038/srep01236.

O-38

Objective Structural Analysis of Multicomponent Amyloid Systems by Chemometrics SAXS Data Decomposition

Fátima Herranz-Trillo[1,2], Minna Groenning[2], Andreas van Maarschalkerweerd[2], Romà Tauler[3], Bente Vestergaard[2] and Pau Bernadó[1]

[1]Centre de Biochimie Structurale. INSERM U1054, CNRS UMR 5048, Université de Montpellier, France. 29, rue de Navacelles, 34090 – Montpellier, France
[2]Department of Pharmacy and Department of Drug Design and Pharmacology, University of Copenhagen, Universitetsparken 2, 2100 Copenhagen, Denmark.
[3]Environmental Chemometrics Group, Department of Environmental Chemistry, Institute of Environmental Assessment and Water Diagnostic (IDAEA-CSIC), Barcelona, Spain.
herranztrillo@cbs.cnrs.fr

In amyloid pathologies (e.g. Parkinson's or Alzheimer's diseases) there are indications that oligomeric aggregated precursors of fibrillation, and not mature fibrils, are the main cause of cytotoxicity and neuronal damage. This observation emphasizes the importance of characterizing early stages in the fibrillation process. The structural analysis of these oligomeric species is a major challenge due to their instability, low relative concentration, the difficulties for isolation, and their interdependence with other species of very different sizes [1]. Mechanistic studies normally monitor individual species of the fibrillation process, such as mature fibrils, whereas the other species remain invisible.
Small-Angle X-ray Scattering (SAXS) is a structural biology technique that allow probing at low-resolution the structure of macromolecules in solution. SAXS is an additive technique, therefore the resulting individual scattering patterns measured at different time-points throughout the fibrillation process, are the sum of the contributions from each component of the mixture. This additive nature of SAXS data allows for probing the evolution of these mixtures of oligomeric states. However, manual decomposition of SAXS datasets into species-specific spectra and relative concentrations is burdened by ambiguity and is a very laborious task [4,5]. We present an objective SAXS data decomposition method by adapting the Multivariate Curve Resolution Alternating Least Squares (MCR-ALS [2,3]) chemometric method. The approach enables rigorous and robust decomposition of synchrotron SAXS data by simultaneously introducing these data in different representations that emphasize molecular changes at different time and structural resolution ranges. The approach has allowed the study of the amyloidogenic processes of insulin and the familial mutant E46K of the Parkinson's disease related protein □lpha-synuclein, and has allowed the structural characterization of the species present (including the oligomeric species) and the kinetic characterization of their transformations. Our approach is generally applicable to any macromolecular mixture with tunable equilibria that can be probed by SAXS.

Acknowledgement: This work has been sponsored by ANR-SPIN-HD (France), Institut national de la santé et de la recherche médicale (INSERM) – France, University of Copenhagen and Neuromed (SUDOE) and the Danish Research Council Sapere Aude programme SAFIR.

References:

[1] Vestergaard, B.; Groenning, M.; Roessle, M.; Kastrup, J. S.; van de Weert, M.; Flink, J. M.; Frokjaer, S.; Gajhede, M.; Svergun, D. I. PLoS Biol. 2007, 5, 1089.
[2] Tauler, R. Chemom. Intell. Lab. Syst. 1995, 30, 133.
[3] Blobel, J.; Bernadó, P.; Svergun, D. I.; Tauler, R.; Pons, M. J. Am. Chem. Soc. 2009, 131, 4378.
[4] Langkilde, A. E.; Vestergaard, B. In Amyloid Proteins: Methods and Protocols; Sigurdsson, E. M., Calero, M., Gasset, M., Eds.; Methods in Molecular Biology; Structural Characterization of Prefibrillar Intermediates and Amyloid Fibrils by Small-Angle X-Ray Scattering: Totowa, NJ, 2012; Vol. 849, pp 137–155.
[5] Giehm, L.; Svergun, D. I.; Otzen, D. E.; Vestergaard, B. Proc. Natl. Acad. Sci. U. S. A. 2011, 108, 3246.

O-39

CLASSIFICATION PROBLEM BY SPARSE PLSDA: CLASSIFICATION OF MICROORGANISMS SUCH AS MOULDS, YEASTS, BACTERIA AND POLLEN USING FTIR SPECTROSCOPY

V.Tafintseva, B.Zimmermann, V.Shapaval, A.Kohler

Institute of Mathematical Sciences and Technology,
Norwegian University of Life Sciences, 1430 Aas, Norway
Email: valeria.tafintseva@nmbu.no

FTIR spectroscopy has been proven a powerful tool for the identification of microorganisms such as bacteria, yeasts, fungi and pollen. Spectral reference libraries are used to establish hierarchical classification trees that allow identification of microorganisms at different taxonomic levels. It has been shown that identification is possible down to the species and in some cases even to strains level.
The development of modern FTIR high-throughput technology combined with high-throughput liquid handling technology allows the acquizition of spectral data of a high number of samples [1]. Even multiple fingerprints from the same microbial samples can be obtained, e.g. when microbial samples are cultivated under different conditions [2] or when samples are preprocessed differently before spectral acquisition. We have recently shown that multiple FTIR fingerprints, where each fingerprint refers to a different and defined cultivation condition applied to the microorganisms, resolve different taxa. While some specific taxa are resolved on the basis of one cultivation condition, other taxa are better resolved under other cultivation conditions [2].
Instead of building separate identification schemes that are built on one specific spectral fingerprint each, one may combine the multiple FTIR fingerprints to a multiblock data set for establishing an identification scheme. Since different spectral regions and different blocks may resolve different taxa, powerful variable selection tools are required.
Sparse Partial Least Squares Discriminant Analysis (SPLSDA) is a method that can be used for variable selection and in discrimination problems [3,4]. It allows interpretation of classification results on the basis of few distinctive variables that are selected during the modelling process. We used SPLSDA recently to establish a hierarchical identification scheme for the identification of pollen species based on FTIR spectroscopy [4]. Single grain infrared spectra of 13 different pollen types, belonging to 11 species, were obtained and analysed by the new approach and classified by SPLSDA. Robust classification models with high accuracy and interpretability were obtained.
In the paper at hand we investigated to what extaned SPLSDA can improve classification results when a multiblock data set based on multiple spectral fingerprints is availabe for building an identification scheme. To this purpose hierarchical identification schemes were established, where each node was built on variables that were selected from a spectral multiblock data set. While FTIR biochemical fingerprints in many cases reflect the known phylogenetic hierarchical taxonomic structure, we frequently encounter situations where spectral fingerprints reveal a different (FTIR) taxonomic structure. Validation by external test sets or cross-model validation was used to validate the established clssification tree.

References:

[1] J. Li, et al. Advances in Reconfigurable Mechanisms and Robots II, Mechanisms and Machine Science 36, 769 (2015).
[2] V. Shapaval, et al., Analyst, 138, 4129 (2013)
[3] I. Karaman,et al., Metabolomics, 11, 367 (2015).
[4] B. Zimmermann, et al., Analytical Chemistry, 81, 803 (2016).

O-40

SPECTRA ALIGNMENT CHARACTERIZATION AND CORRECTION IN GAS-CHROMATOGRAPHY – ION MOBILITY SPECTROMETRY

S. Oller-Moreno[1,2], J. Fonollosa[1,2], J.M. Jiménez-Soto[1,2], R. Garrido-Delgado[3], L. Arce[3], A. Pardo[2], S. Marco[1,2]

[1]Signal and Information Processing for Sensing Systems, Institute for Bioengineering of Catalonia, Barcelona, Spain
[2]Department of Engineering, Electronics. University of Barcelona, Spain
[3]Department of Analytical Chemistry, University of Córdoba, Spain
soller@ibecbarcelona.eu

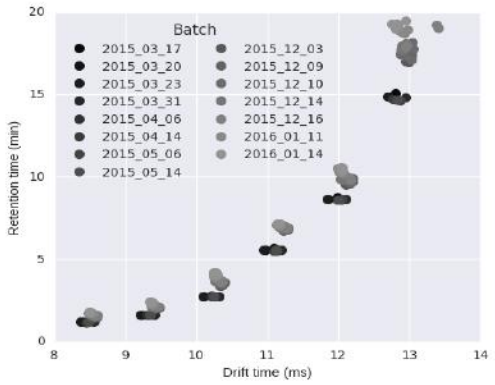
Ion Mobility Spectrometry (IMS) is an analytical technique that relies on the ionization of gas phase samples and its separation under an electric field at atmospheric pressure . To enhance the selectivity of IMS, especially when analyzing complex biological samples, a gas chromatograph can be inserted for sample pre-separation. The resulting hyphenated technique is gaining popularity in the analysis of breath or urine samples for metabolomic applications, and in quality control and fraud detection in alimentary sector (e.g. fraud prevention screening in wines or olive oil sectors).

The degradation of the chromatographic column, the variation of IMS drift and sample flow and temperature and pressure changes cause misalignments in GC-IMS samples, more relevant in long term usage of GC-IMS instrumentation. Characterizing and correcting misalignments is necessary if classification or regression models are to be built based on acquired spectra.

Previous work on this shortcoming [1] shows that major misalignments can be corrected with linear transformations of the retention time axis. However, improvements are still required, especially in short retention times where non-linear transformations are required. Commercial software available with GC-IMS instrumentation offers linear retention time corrections, but need to be manually parametrized for each sample, being a slow and time consuming task.

In this work we compare 42 standard samples of six spiked compounds measured in a 10-month time lapse. Using those samples we first characterize the misalignments in different time scales (intra-day, inter-day, inter-week, one-month, several months). Based on this characterization, we propose a required time span between consecutive standard samples. We also explored how the correction in the standard can be propagated to align olive oil samples measured in the same sample batch in an automated fashion. The final aim is to provide an easy-to-use graphical user interface for automated and (semi-)supervised GC-IMS sample alignment.

Acknowledgement: S.O, J.F, J.M.J., A.P, S.M are part of a consolidated research group recognized by the Generalitat de Catalunya (2014-SGR-1445). This work is also financially supported by the SIGVOL project (TEC2014-59229-R) funded by the Spanish Ministerio de Economía y Competitividad.



References:
[1] T. Perl, B. Bödeker, M. Jünger, J. Nolte, and W. Vautz, "Alignment of retention time obtained from multicapillary column gas chromatography used for VOC analysis with ion mobility spectrometry.," Anal. Bioanal. Chem., vol. 397, no. 6, pp. 2385–94, Jul. 2010.

O-41

ON-LINE OUTLIER/REDUNDANCY FILTERING AND SEMISUPERVISED INCREMENTAL CALIBRATION MODELING IN MELAMINE RESIN PRODUCTION USING FT-NIR SPECTRA

C. Cernuda[1,2], E. Lughofer[1], T. Reischer[3], W. Kantner[3], M. Pawliczek[4], M. Brandstetter[4]

[1]Department of Knowledge-Based Mathematical Systems, Johannes Kepler University Linz, Linz, Austria.
[2] BCAM - Basque Center for Applied Mathematics, Bilbao, Spain.
[3]MetaDynea Austria GmbH, Krems, Austria.
[4]RECENDT GmbH, Research Center for Non Destructive Testing, Linz, Austria.
carlos.cernuda@jku.at, ccernuda@bcamath.org

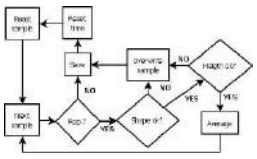
In the considered batch process of melamine resin production, the essential parameter to be monitored is the cloud point in the condensation process. It indicates the optimum moment to start the cooling process in order to stop the condensation and is of importance to assure high product quality. The current manual controlling of the condensation process is time-consuming and includes uncertainty in the monitoring of the whole condensation reaction.

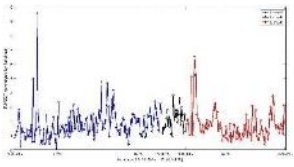
Standard linear models have been trained on an initial calibration set and applied to a separate in-line test set. However, the obtained results leave much room for improvement, especially in case of lamp changes and intensity downtrends of the halogen light source, and some dynamics which cannot be fully controlled by static models. Furthermore, some internal dynamics, mainly non-linear, such as changing compositions of the educt, are assumed.

In the data acquisition process, three consecutive spectra are recorded and matched to one single target measurement, and many other spectra, without target information, is stored. Theoretically, the latter could be considered as repetitions, thus averaged, but the presence of outliers and some other factors, produce artifacts that influence severely the posterior model accuracy. The non-linear behavior requires the usage of non-linear models, preferable those ones in which the degree of non-linearity can be adjusted on demand and on the fly.

We (i) propose an on-line outlier/redundancy filter (left figure) that detects outliers and determines whether averaging repetitions makes sense, and (ii) a dynamically self-adaptive calibration technique using a semi-supervised flexible fuzzy inference system [1] (middle figure), in which only the antecedent parts of the fuzzy rules are updated when no target information is available (unsupervised), and the whole rules are updated in case of known target values (supervised, maximal one per batch to keep operators effort low). The fuzzy system is updated by re-training cycles based on sliding windows, which contain most informative samples as selected by an on-line, single-pass active learning approach [2].

The results show robustness (non-significantly different residuals) against three lamp changes and one reactor cleaning (right figure), and a significant reduction on some dramatic error peaks compared to re-trained linear (PLS) models.





Acknowledgement: This work was funded by the Austrian research funding association (FFG) under the scope of the COMET programme within the research network "Industrial Methods for Process Analytical Chemistry" (imPACts). This research is also supported by the Basque Government through the ELKARTEK and BERC 2014-2017 programs and by the Spanish Ministry of Economy and Competitiveness MINECO: BCAM Severo Ochoa accreditation SEV-2013-032

References:

[1] E.Lughofer et al., Generalized Smart Evolving Fuzzy Systems, Evolving Systems, Vol 6, pp 269-292 (2015).
[2] C. Cernuda et al., Incremental and Decremental Active Learning for Optimized Self-Adaptive Calibration in Viscose Production, Chemometrics and Intelligent Laboratory Systems, Vol 138, pp 14-29 (2014).

O-42

INTERPRETING MULTIWAY PLS MODELS IN TERMS OF VARIABLES RELEVANCE

M. Cocchi, M. Li Vigni

[1]*Department of Chemical and Geological Science, University of Modena and Reggio Emilia, Via Campi, 103, 41125 Modena, Italy*
marina.cocchi@unimore.it

Multiway regression and discriminant models are nice because take directly into account the multiway structure of data, e.g. when dealing with hyphenated analytical techniques. However, model interpretation in term of assessing relevance of predictors is still perceived as cumbersome by most of the users. As well, variable ranking and selection techniques are abundant for two-dimensional data set and more limited for multi-way arrays. Nonetheless, assessing which are the most relevant variable both for the prediction/discrimination and to describe the phenomenon/property expressed by each dependent variable, is needed. In this respect, the examination of regression coefficients, which are a landscape, is even less straightforward than in the bilinear case, also often due to the unreadable complex patterns associated to them.

In this work, we compare different approaches, including a new proposal, which extend the target projection formulation, and the related Selectivity Ratio (SR) parameter [1] to multi-way arrays. The target component (TP) in the mutiway implementation is obtained by a one component PARAFAC model of $\underline{\mathbf{X}}$ constrained on scores obtained as in the two-mode case by $\mathbf{X}b_{NPLS}/||b_{NPLS}||$, where $\mathbf{X}$ is the matricized form of $\underline{\mathbf{X}}$. The SR is defined for all the variables/conditions modes of the $\underline{\mathbf{X}}$ array, according to two slightly different formulations.

Some data sets, mainly in the field of food authentication and processing, are shown considering regression (NPLS) as well as discrimination frameworks (NPLS-DA).

For each case of study weights, regression coefficients, VIP scores [2] are discussed together with TP and SR.

References:

[1] T. Rajalahti, R. Arneberg, F. S. Berven, K. M. Myhr, R.J. Ulvik, O.M. Kvalheim, *Chemom. Intell. Lab. Syst.*, 2009, *95*, 35–48.
[2] S. Favilla, C. Durante, M. Li Vigni, M. Cocchi, *Chemom. Intell. Lab. Syst*. 2013, *129* 76–86.

O-43

A novel Multi-block Variable Influence on Orthogonal Projections (the multi-block VIOP) for a sharper data interpretation in O2PLS and OnPLS multivariate models.

Beatriz Galindo-Prieto [1,2], Paul Geladi [3], Johan Trygg [1]

[1] Department of Chemistry, CLiC, Umeå University, Sweden
[2] Industrial Doctoral School, Umeå University, Sweden
[3] Forest Biomaterials and Technology, Swedish University of Agricultural Sciences, Umeå, Sweden
beatriz.galindo@umu.se

Keywords: variable influence on projection, VIOP, multi-block variable selection.

In multivariate and multi-block models there is a need for reducing dimensionality and increasing interpretability; variable selection is one of the most promising techniques for achieving that goal. In case of PLS and OPLS models [1,2] for both multivariate data analysis (MVA) and multivariate time series analysis (MTSA), the VIP_{OPLS} (variable influence on projection for orthogonal projections to latent structures, also called OPLS-VIP) presented in [3,4] leads to a sharper model interpretation than the traditional VIP. A new variable influence on projection approach for multi-block data analysis using O2PLS and OnPLS has been developed and it is now being presented as the multi-block VIOP.

Multi-block VIOP is a parameter that summarizes the importance of the X-variables in O2PLS and OnPLS models, taking into consideration not only the predictive/joint components (correlated variation) and the orthogonal/unique components (uncorrelated variation), but also the connectivity between the data matrices (blocks). In addition, a high-resolution data interpretation is achieved by the obtainment of four multi-block VIOP profiles representing the VIOP for the total model, the VIOP for the global variance (shared by all the data blocks), the VIOP for the local variance (shared by some, but not all, of the data blocks), and the VIOP for the unique variance (which is unique for each block). Results on both synthetic and real data sets will be presented to prove the value of this new development.

Acknowledgement: The authors would like to acknowledge the financial support from MKS Data Analytics Solutions (formerly MKS Umetrics) (BG-P) and the Industrial Doctoral School (IDS), Umeå University, Sweden.

References:

[1] J. Trygg, S. Wold, Orthogonal projections to latent structures (O-PLS), Journal of Chemometrics, 16 (2002) 119-128.
[2] M. Dumarey, **B. Galindo-Prieto**, M. Fransson, M. Josefson, J. Trygg, OPLS methods for the analysis of hyperspectral images—comparison with MCR-ALS, Journal of Chemometrics, 28 (2014) 687-696.
[3] **B. Galindo-Prieto**, L. Eriksson, J. Trygg, Variable influence on projection (VIP) for orthogonal projections to latent structures (OPLS), Journal of Chemometrics, 28 (2014) 623-632.
[4] **B. Galindo-Prieto**, L. Eriksson, J. Trygg, Variable influence on projection (VIP) for OPLS models and its applicability in multivariate time series analysis, Chemometrics and Intelligent Laboratory Systems, 146 (2015) 297-304.

O-44

NON-LINEAR MULTI-WAY MODELING

Federico Marini

Dept. of Chemistry, University of Rome "La Sapienza", P.le Aldo Moro 5, I-00185 Rome, Italy
federico.marini@uniroma1.it

Analytical chemistry is facing a growing complexity of the signals, which are recorded by modern instruments. Indeed, it is not uncommon that, instead on one-dimensional profiles, such as spectra or TIC chromatograms, 2D or even higher dimensional landscapes are recorded on each individual sample: excitation-emission fluorescence signals, or the profiles resulting from hyphenated or multi-dimensional chromatography are just a few examples. When such signals are recorded on multiple samples, the corresponding data structures are (hyper-)cubes or so-called multi-way arrays. Although multi-way data bay be analyzed with conventional 2-way algorithms after unfolding, several algorithms have been proposed in the literature to take directly into account the multi-modal structure of the arrays when dealing with exploratory (e.g. Tucker3 [1] or PARAFAC [2]) or predictive (e.g., NPLS [3], multi-way covariate regression [4] or SCREAM [5]) modeling. All these algorithm are based on the underlying hypothesis of a (multi-)linear relationship between the variables or the blocks. However, especially when predictive modeling is concerned, i.e. when one or more responses **Y** are predicted from an independent block **X,** if multiple sources of (often spurious) variation are present in the data, the linearity hypothesis may not hold any more and some kind of non-linear modeling may be needed.
Based on these considerations, in the present communication the possibility of building non-linear multi-way regression and classification models is addressed, by extending the locally weighted regression approach, originally proposed for two-way arrays [6]. In locally weighted regression, a globally non-linear problem is expressed as the combination of smaller models, which are linear in specific (local) regions of the original space. To extend this concept to the multi-way arrays, many ways of defining distances between matrices have been considered (such as the RV coefficient and its modifications). Then, NPLS was mainly used for building the locally linear models (even if some preliminary results have also been obtained by multi-way covariate regression and SCREAM). The approach has been tested with promising results on both simulated and real data sets.

References

[1] L.R. Tucker, Psychometrika, 31, 279 (1966).
[2] R.A. Harshman, UCLA Work. Pap. Phon., 16, 1 (1970).
[3] R. Bro, J. Chemometr., 10, 47 (1996).
[4] A.K. Smilde, H.A.L. Kiers, J. Chemometr., 13, 31 (1999).
[5] F. Marini, R. Bro, Chemometr. Intell. Lab. Syst., 129, 64 (2013).
[6] V. Centner, D.L. Massart, Anal. Chem., 70, 4206 (1998).

O-45

ASPECTS OF MULTIWAY CALIBRATION IN BIOANALYTICAL CHEMISTRY

Hai-Long Wu*, Xiao-Li Yin, Hui-Wen Gu, Ru-Qin Yu**

*State Key Laboratory of Chemo/Biosensing and Chemometrics, College of Chemistry and Chemical Engineering, Hunan University, Changsha 410082, China. *Email: hlwu@hnu.edu.cn*

Multiway calibration has become one of the research hot spots in the field of bioanalysis due to the following reasons: 1) instruments that easily generate multidimensional experimental data array per sample are increasingly available to analysts; 2) based on the "mathematical separation", multiway calibration methods enable one to achieve direct quantitative analysis of analyte(s) of interest in complex systems even in the presence of unknown or uncalibrated interferences, which is known as the "second-order advantage". However, there are still some fundamental issues are immature, such as higher-order advantage, figures of merits in high-order calibration, and rank-estimation for high-order data, which will be an important research fields in the future. Moreover, with the development of "-omics", bioinformatics has gone from an obscure area of biological sciences into a most mainstream technology field, often under the limelight. The explosion of computational techniques for the extracting of biological information makes the bioinformaticians under the similar problems that the chemomtricians currently faced-- tsunamis of data are flooding the scientific world. In the post-genomic era, these complex data sets not only include macromolecular sequences and structures (DNA, RNA, or protein) information, but also include chromatographic, spectral and concentration information. Therefore, the combination of chemometrics and bioinformatics may become a new growth point for the application of chemometrics in bioanalysis. In addition, for people engaged in biological analysis, they may not use multiway calibration or other chemometric methods owing to the perceived difficulties in their application. The great difficulty, therefore, is with the transfer of technology from the laboratory to intelligent analytical instruments. Once this occurs, the use of chemometric methods, including multiway calibration in bioanalysis, will become common place.

Acknowledgement: The authors gratefully acknowledge the National Natural Science Foundation of China (Grant Nos. 21175041, 21221003, 21521063 and 21575039), and the National Basic Research Program (Grant Nos. 2007CB216404 and 2012CB910602) as well as the Grants from the Ministry of Education of China for financial support.

References:

[1] H.L. Wu, M. Shibukawa, K. Oguma. Journal of Chemometrics, 1998, 12: 1-26.
[2] M. Martı´nez Galera, M.D. Gil Garcı´a, H.C. Goicoechea, Trends in Analytical Chemistry, 2007, 26 (11): 1032-1042.
[3] H.L. Wu, J.F. Nie, Y.J. Yu, R.Q. Yu. Analitica Chimica Acta, 2009, 650: 131-142.
[4] A.C. Olivieri, Chemical Reviews. 2014, 114: 5358-5378.
[5] H.L. Wu, Y. Li, R.Q. Yu. Journal of Chemometrics, 2014, 28: 476-489.
[6] X.L. Yin, H.L. Wu, H.W. Gu, Y. Hu, H. Xia, L. Wang, R.Q. Yu. Chemometrics and Intelligent Laboratory Systems, 2015, 146: 447-456.
[7] L.X. Xie, H.L. Wu, Y. Fang, C. Kang, S.X. Xiang, L. Zhu, X.L. Yin, H.W. Gu, Z. Liu, R.Q. Yu. Chemometrics and Intelligent Laboratory Systems, 2016, 148: 9-19.
[8] H.L. Wu, Y. Li, C. Kang, R.Q. Yu, Chapter 3. Multiway Calibration Based on Alternating Multilinear Decomposition. In: A. C. Olivieri, G. M. Escandar, H. C. Goicoechea, A. M. de la Peña, Eds., Fundamentals and Analytical Applications of Multiway Calibration, Elsevier, Inc., Oxford, 2015, pp. 83-166.
[9] H.L. Wu, C. Kang, Y. Li, R.Q. Yu, Chapter 4. Practical Analytical Applications of Multiway Calibration Methods Based on Alternating Multilinear Decomposition. In: A. C. Olivieri, G. M. Escandar, H. C. Goicoechea, A. M. de la Peña, Eds., Fundamentals and Analytical applications of Multiway Calibration, Elsevier, Inc., Oxford, 2015, pp. 167-246.

O-46

Parameter-Free Support Vector Machines for Calibration

Peter de Boves Harrington[1]

Center for Intelligent Chemical Instrumentation
Department of Chemistry and Biochemistry, Clippinger Laboratories
OHIO University
Athens, OH 45701-2979 USA
Peter.Harrington@OHIO.edu

Support vector machines (SVMs) are pattern recognizers that build models via quadratic programming constrained minmization[1]. The approach is to minimize the Euclidena length of the linear regression vector and the residual error. Quadratic programming offers a key advantage of efficient computation. This advantage arises from inequality constraints that remove data objects from the model construction and two modes towards the miminization (i.e., primal that occurs in the object space and dual that occurs in the variable space). However, one often does not gain an advantage without sacrificing elsewhere, because fewer objects are used to construct the model noise may also be embedded in the model.

SVMs may be used for regression and models may be constructed using L1 and L2 loss functions. L2 minimizes the sum of squared residuals to the model while L1 minimizes the sum of the absolute value of the residuals. L1 models are advantageous because they are robust and less prone to outliers in the data. L1 models compare favorably to L2 models when outliers are present in the training dataset.

SVMs have a fitting or cost parameter *C* that determines the trade-off between bias and variance in the regression model. Using internal Latin bootstrap partitions [2] and the high-speed advantage of SVM, a parameter-free SVMR calibration method is devised. This method compares favorably to a self-optimizing partial least squares regression (PLSR) method [3] with respect to speed (i.e., efficiency) and prediction accuracy (i.e., efficacy).

A dataset of 300 near-infrared bovine plasma spectra with 519 wavelength was used to construct calibration models for the glucose concentrations. The SVMR constructed a model five times faster than the PLSR algorithm, however PLSR predicted better than SVMR with an error of 0.05 mM less than the RMS standard prediction error of 1.1 mM.

Acknowledgement: OHIO University Center for Intelligent Chemical Instrumentation, USDA, and Cannaprint, Inc.

References:

[1] P.B. Harrington, Analytical Chemistry, 87, 11065 (2015) DOI: 10.1021/acs.analchem.5b03113.
[2] P.B. Harrington, Trends in Analytical Chemistry, 25, 1112 (2006), DOI: 10.1016/j.trac.2006.10.010.
[3] P.B. Harrington, J. Kister, J. Artaud, and N. Dupuy, Analytical Chemistry, 81, 7160, (2009) DOI: 10.1021/ac900538n.

O-47

HIGH-LEVEL FUSION METHODS TO CLASSIFY SAMPLES ASSOCIATED TO MULTIPLE ANALYTICAL SOURCES

D. Ballabio [1], R. Todeschini [1]

[1]Milano Chemometrics and QSAR Research Group, Department of Earth and Environmental Sciences, University of Milano-Bicocca, p.zza della Scienza 1, Milano, Italy

davide.ballabio@unimib.it

Recent advances in technology enabled the collection of huge amounts of data from multiple analytical sources. Low-level and mid-level data fusion methods jointly analyse data achieved by means of a variety of analytical techniques since their combination can enhance the identification of relevant information [1,2]. An alternative to directly merge all available analytical sources is high-level fusion approach, which combines predictions obtained by different modelling techniques. High-level methods have the advantage of giving predictions even if one of the sources is not available, as well as avoiding peculiar data pre-treatment to make the analytical sources comparable; on the other side, calibration of several multivariate models is required.

This study deals with the application and comparison of an exhaustive set of high-level fusion approaches on several analytical multiple datasets where the target was classification of samples. The aim is to show advantages of high-level predictions and how recent advances in high-level fusion modelling, such as Dempster-Shafer theory of evidence and Bayesian approaches, can reduce uncertainty in analytical data. In particular, the following five datasets were taken into account: Apple (508 samples, 2 analytical sources, 2 classes), Milk (174 samples, 3 sources, 3 classes), Plasma (94 samples, 3 sources, 2 classes), Wines (44 samples, 4 sources, 3 classes), Biodegradation (1055 samples, 3 sources, 2 classes). Both linear (LDA, PCA-LDA, PLSDA) and not-linear (QDA, PCA-QDA, kNN, N3, Support Vector Machines) multivariate classification models were evaluated on each analytical source. For each dataset, a comprehensive set of high-level methods was thus applied to combine predictions: scoring (normal and weighted) schemes, Dempster-Shafer's theory of evidence and Bayesian inference based on probability estimation (both with strict and protective versions) [3].

Comparison was performed on the basis of a Montecarlo validation protocol: models were calibrated 1000 times; each time 25% of samples was randomly selected and used as test set; classification models (as well as parameters for high-level approaches) were calculated on the training set and then validated on the test set; test predictions were then aggregated to get high-level fusion results.

Results achieved by means of high-level fusion approaches were then compared (with anova, t-test and boxplots) to single models and a mid-level fusion approach on the basis of classification performances, which were estimated with non-error rate (average of class sensitivities) and ratio of not assigned samples (if present).

Results did not indicate a single optimal fusion approach; however, the application of high-level modelling always allowed significant reduction of uncertainty, enhanced data interpretation and showed evident benefits when sources of information are jointly analysed, without compromising the quality of analytical model predictions.

References:

[1] E. Acar, M.A. Rasmussen, F. Savorani, T. Næs, R. Bro, Chemometrics and Intelligent Laboratory Systems 129, 53–63, (2013)

[2] E. Borràs, J. Ferré, R. Boqué, M. Mestres, L. Acena, O. Busto, Analytica Chimica Acta, 891, 1-14, (2015)

[3] A. Fernández, A. Lombardo, R. Rallo, A. Roncaglioni, F. Giralt, E. Benfenati, Environment International 45 51-58 (2012)

O-48

COMBINING ANOVA WITH POCHEMON FOR PATHOGEN INTERACTION AND DEVELOPMENT

B.P. Geurts[1], A.H. Neerincx[1],S. Bertrand[2], J.-L. Wolfender[3], S.M. Cristescu[1], J.J. Jansen[1] and L.M.C. Buydens[1]

[1]*Institute for Molecules and Materials, Radboud University, P.O. Box 9010, 6500 GL Nijmegen, The Netherlands*
[2]*Groupe Mer, Molécules, UFR des Sciences Pharmaceutiques et Biologiques, Université de Nantes, Santé-EA 2160, Nantes, France*
[3]*School of Pharmaceutical Sciences, EPGL, University of Geneva, University of Lausanne, Quai Ernest-Ansermet 30, CH-1211, Geneva 4, Switzerland*
mailto:b.geurts@science.ru.nl

Revealing the biochemistry associated to microbial interspecies interactions is relevant for many medical purposes. Examples are inducing the production of novel metabolites and improving diagnostics in patients with polymicrobial infections. Each pathogen has a characteristic metabolic fingerprint, that allows their identification based on their unique multivariate biochemistry. When multiple pathogens come in contact with each other, the profile of their co-culture will display a chemistry attributed to both mixing of the individual chemistry of the monocultures and competition between the pathogens. Therefore, investigating pathogen development in a polymicrobial environment requires dedicated chemometric tools.

As a first step, the overall time dynamics has to be separated from the pathogen specific chemistry to analyse the contributions of both aspects separately. Analysis of Variance (ANOVA) is an established method to separate information from different sources of variance. Analysis of the resulting effect matrices may be done by methods like ASCA [1], Principal Component Analysis (ANOVA-PCA) [2] or recently proposed alternatives such as target projection (ANOVA-TP) [3]. These methods are now in widespread use to analyse time-resolved data from experiments that include multiple treatments or multiple groups of individuals. However, none of these methods allow identification of chemistry, specific to interspecies interaction. We specifically developed the multivariate data analysis method Projected Orthogonalised Chemical Encounter Monitoring (POCHEMON) to highlight metabolites characteristic for the interaction of two pathogens in co-culture [4]. However, until now this approach was limited to non-dynamic data.

We propose to integrate ANOVA-PCA with the POCHEMON approach to disentangle the pathogen dynamics and the specific biochemistry in interspecies interactions. Two complementary case studies show great potential for both liquid and gas chromatography / mass spectrometry to reveal novel information regarding overall pathogen development as well as chemistry, specific to interspecies interaction.

References:

[1] Smilde, A.K., et al., *ANOVA-simultaneous component analysis (ASCA): a new tool for analyzing designed metabolomics data.* Bioinformatics, 2005. 21(13): p. 3043-3048.
[2] de Haan, J.R., et al., *Interpretation of ANOVA models for microarray data using PCA.* Bioinformatics, 2007. 23(2): p. 184-190.
[3] Marini, F., et al., *Analysis of variance of designed chromatographic data sets: The analysis of variance-target projection approach.* Journal of Chromatography A, 2015. 1405: p. 94-102.
[4] Jansen, J., et al., *Projected Orthogonalized CHemical Encounter MONitoring (POCHEMON) for microbial interactions in co-culture.* Metabolomics, 2014: p. 1-12.

O-49

THE MANNE-MAEDER CRITERIA AND THE COMPLEMENTARITY THEORY

K. Neymeyr[1,2], M. Sawall [1],

[1]Universität Rostock, Mathematical Institute, Rostock, Germany.
[2]Leibniz-Institute for Catalysis (LIKAT), Rostock, Germany
email: klaus.neymeyr@uni-rostock.de

The partial knowledge of the factors can considerably simplify the multivariate curve resolution problem. In 1995 Manne [1] presented criteria on the basis of results by Maeder which allow to compute the concentration profile of a component if the pure component spectra fulfill certain window conditions. The more general *complementarity theory* in [2,3] provides precise predictions for certain unknown parts of the factors on the basis of known parts. The key equation is

$$CA^T = (U\Sigma T^{-1})(TV^T)$$

with the singular value decomposition $D=U\Sigma V^T$ of the spectral data matrix and a regular matrix T. If the true factor A is known partially, then some rows of T are determined. This implies linear and affine linear restrictions on T^{-1} and results in restrictions on the concentration factor.

In this talk we present a new and simple SVD-free form of the complementarity theory [4] which starts from the sum of dyadic products

$$D=CA^T = [c_1,\ldots,c_s][a_1,\ldots,a_s]^T = \sum c_i a_i^T .$$

If D and one or more of the concentration profiles c_k and spectra a_k are known, then elementary arguments of linear algebra allow to derive linear and affine linear constraints on the remaining pure components. These conditions include the classical Manne-Maeder criteria. In its general form these constraints can be written in terms of orthogonal projection operators. This allows the easy and straightforward applicability of the SVD-free form of the complementarity theory.

The results are demonstrated and tested for experimental FT/IR spectroscopic data from the Rhodium-catalyzed hydroformylation process and for UV/Vis data from the catalytic formation of hafnacyclopentene.

References:

1. Manne, R. *Chemom. Intell. Lab. Syst*. **1995**, 27(1), 89-94
2. Sawall, M., Fischer, C., Heller, D., Neymeyr, K., *J. Chemom*. **2012**, 26(10), 526-537
3. Sawall, M., Neymeyr K., *Anal. Chim. Acta,* **2014**, 828, 17-26
4. Neymeyr, K., Sawall, M., *J. Chemom.**2016**, 30, 30-36*

O-50

CHEMOMETRIC APPROACHES TO INCREASE EFFICIENCY OF NMR SPECTROMETRY IN ANALYSIS OF FOOD PRODUCTS

Y.B. Monakhova[1], B.W.K. Diehl[2]

[1]Spectral Service AG, Emil-Hoffmann-Straße 33, 50996, Cologne, Germany
[2]Institute of Chemistry, Saratov State University, Astrakhanskaya Street 83, 410012 Saratov, Russia
monakhova@spectralservice.de

In recent years the number of NMR spectroscopic studies utilizing multivariate techniques has been dramatically increased. In this contribution, several new chemometric methodologies were developed and applied to NMR data of food products.

First, to address the data standartization problem, a protocol for calibration transfer of partial least square (PLS) regression model between high-resolution NMR spectrometers of different frequencies and equipped with different probes was established. As the test system quantitative model to predict the concentration of blended soy species in sunflower lecithin was used. The results revealed that piecewise direct standardization (PDS) showed the best performance for estimating lecithin falsification regarding its vegetable origin resulting in a significant increase decrease in root mean square error of prediction (RMSEP) from 5.0-7.3% without standardization to 2.9-3.2%. The sensitivity of instrument transfer methods with respect to the type of spectrometer, the number of samples and the subset selection will be also discussed.

Second, the major challenge facing NMR spectroscopic mixture analysis is the overlapping of signals and the arising impossibility to recover the spectral profiles for identification and to integrate separated signals for quantification. Different independent component analysis (ICA) algorithms and multivariate curve resolution-alternating least squares (MCR-ALS) were applied for simultaneous NMR spectroscopic determination of organic substances in food matrices, including honey, soft drinks, and liquids used in electronic cigarettes. Good quality spectral resolution of up to eight-component mixtures was achieved (correlation coefficients between resolved and experimental spectra were not less than 0.90). The resolved ICA and MCR-ALS concentrations match well with the results of reference analysis. In the absence of reference materials ICA can be combined with PULCON principle (pulse length based concentration determination) for quantification. In this case instead of conventional application of absolute integral intensity in case of undisturbed signals, the multiplication of resolved IC absolute integral and its relative concentration in the mixture for each component was used.

O-51

QUANTITATIVE MEASUREMENTS OF ISOMERS IN FOOD BY TERAHERTZ SPECTROSCOPY AND CHEMOMETRICS

Zhuoyong Zhang*, Xin Zhang, Congmin Du, Shaohua Lu

Department of Chemistry, Capital Normal University, Beijing 100048, China
email: gusto2008@vip.sina.com

Isomers have only tiny distinctions in chemical structure and they are difficult quantified by conventional analytical instruments. Terahertz spectroscopy is a novel spectroscopic technique and has shown a great potential as a crucial method for isomers research [1, 2].

In this work, terahertz time domain spectroscopy (THz-TDS) combined with chemometrics has been utilized for the qualitative and quantitative analysis of two amino acids (L-Glutamic acid and L-Glutamine mixtures) and two saccharide (D-(−)fructose and D-(+)galactose anhydrous). Different from previous studies, our binary amino acids mixtures samples were prepared as pressed pellets with yellow foxtail millet substituted for polyethylene (PE). Results show that L-Glutamic and L-Glutamine, D-(−)fructose and D-(+)galactose anhydrous can be discriminated after transforming the time-domain signals into absorption coefficients. After proper pretreatment of absorption spectra, quantitative analysis was achieved by partial least-squares (PLS) and interval partial least squares (iPLS) regression. A reliable and unbiased estimation for cross validation was achieved by bootstrapped Latin partition. The performance of models was evaluated according to the root mean square error of prediction (RMSEP) and correlation coefficient (R). iPLS yielded accurate results with low RMSEP and high R for the isomers. Multivariate curve resolution alternating least squares (MCR-ALS) can successfully be applied for resolution of pure THz spectra and concentration of the components in the prepared samples.

As a rapid, non-destructive technique with less or no pretreatment procedure, THz time domain spectroscopy has a potential for isomers quantitative analysis in food.

Acknowledgement: This study was funded by National Instrumentation Program (2012YQ140005) and National Natural Science Foundation of China (21275101).

References:

[1] Baxter, J. B., Guglietta, G. W. Anal. Chem. 83(12), 4342 (2011).
[2] Walther, M., Fischer, B. M. Chem. Phys. 288 (2), 288, 261(2003).

O-52

Robustness: A key parameter in the validation of multivariate sensory regression models

E.J.J. van Velzen[1], S. Lam[2], E. Saccenti[2], A.K. Smilde[2], E. Tareilus[1], J.P.M. van Duynhoven[1], D.M. Jacobs[1]

[1]Unilever Research, Microbiology & Analytical, Olivier van Noortlaan 120, Vlaardingen, The Netherlands
[2]Universiteit van Amsterdam, Biosystems Data Analysis, Science Park 904, Amsterdam, The Netherlands.

International guidelines define robustness of an analytical procedure as a measure of its capacity to remain unaffected by small variations in experimental conditions [1]. Robustness is generally not considered in validation protocols even though it provides an important indicator of the fitness of an analytical procedure during normal use. Also in the validation of multivariate sensory regression models robustness is seldom included. Particularly in sensory analysis were multivariate relationships are assessed between descriptive (consumer) data and compositional (chemical) data we state that a measure of robustness is a critical parameter for establishing model reliability and model suitability.

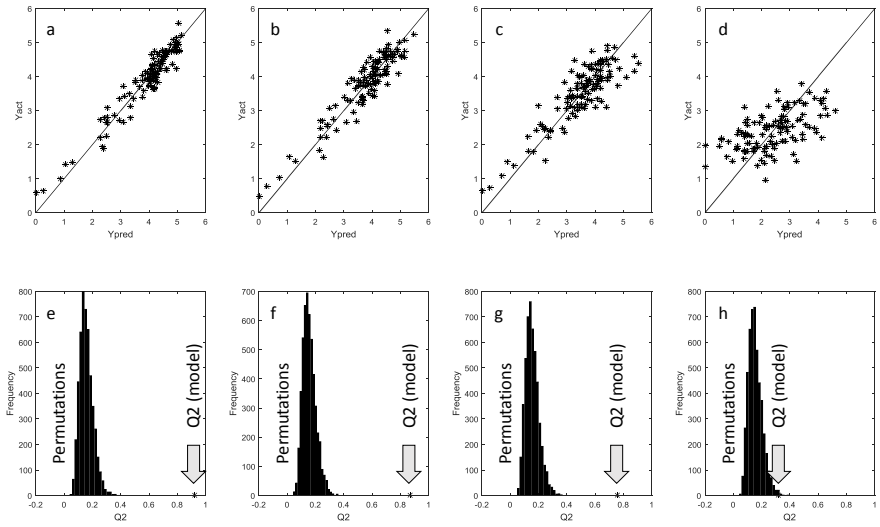


Figure 1. Actual versus predicted sensory scores obtained from a statistically significant PLS model (a-d). Increasing levels of artificial noise is added. As a result the statistical significance of the model dropped. This is demonstrated by the Q2-values in a permutation test (e-h).

We demonstrate the importance of robustness testing in a sensory study in which the chemical compositions of natural food products were associated with the descriptive scores of a trained consumer panel. This multivariate association was modelled by means of PLS. The data was rather complex as (*i*) the predictor set was compiled of various analytical datasets with different error structures, and (*ii*) the predictor set contained missing data. Furthermore, the data was sparse given the enormous sample diversity in the training set. By integration of predefined noise simulations in the double cross-validation and the permutation testing [2], estimations of the relevant model statistics (e.g. Q2-values) could be obtained (Figure 1). The aim of this approach is that the susceptibility of the multivariate models against experimental errors can be tested. We found that robustness testing is of great additional value in the validation of multivariate prediction models. An important prerequisite, however, is that the error structures of the dependent and independent variables should be well known in advance.

References:

[1] Y. vander Heyden *et al.*, J Pharm Biomed Anal., 5-6, 723-753 (2001).
[2] J.A. Westerhuis *et al.*, Metabolomics, 4, 1, 81-89 (2008).

O-53

Fat and protein determination in milk using Vis/SW-NIR spectroscopy

A. Melenteva[1], A. Bogomolov[1,2]

[1]Samara State Technical University, Molodogvardeyskaya Street 244, 443100 Samara, Russia
[2]Global Modelling, Rembrandtstrasse 1, 73433 Aalen, Germany
melenteva-anastasija@rambler.ru

The rapid growth of the food industry requires new solutions in the quality control of processes and products. The development of new methods enabling quick and accurate determination of the analyzed components is an important task especially for mass consumption products such as milk. Due to the presence of fat globules and protein micelles milk has pronounced light-scattering properties. Economically attractive visible and short-wave near infrared (Vis/SW-NIR) spectroscopy can be successfully used for the quantitative analysis of milk. The recently proposed new method is based on multiple light scatter, which strongly prevailing in the Vis/SW-NIR region, and requires the calibration to be performed on raw spectral data, without a corrective preprocessing [1, 2]. The difference in scattering spectral patterns by fat and protein particles can be utilized for their accurate quantitative determination using multivariate modelling methods, i.e. partial least squares (PLS) regression analysis.

In the first step, the dependence of intensities and shapes of diffuse transmittance spectra in the Vis/SW-NIR region has been systematically investigated using gradual homogenization of raw milk samples [3]. The homogenization results in significant spectral changes, which were explained in terms of the representative layer theory and other scattering theories. The fundamental feasibility of scatter-based quantitative analysis of milk using their low-selectivity diffuse transmission spectra in the region 400-1100 nm has been proved in a series of designed experiments [2]. This new analytical technique is resistant to an essential variability of fat globule sizes that may occur in raw milk samples. Finally, the global (i.e. resistant to seasonal, genetic, regional and other milk variations) models for determination of fat and total protein content in raw milk based on historical spectroscopic data collected during a year has been developed [4].

In the subsequent study, full-range spectroscopy was replaced by a set of light-emitting diodes (LEDs) as a light source and conventional digital camera as a detector [5]. This simplified technology has shown an acceptable determination accuracy of fat and protein content in raw natural milk. Additionally, the optimal configuration of LEDs for the optical milk sensor was calculated using a new variable selection method [6].

The reported results can be put into the basis of compact and inexpensive analyzers of raw milk quality, in particular, for in-line or field measurements.

References:

7. A. Bogomolov, S. Dietrich, B. Boldrini, R.W. Kessler, Food Chemistry Vol. 134, P. 412–418, 2012.
8. A. Bogomolov, A. Melenteva, Chemometrics and Intelligent Laboratory Systems, Vol. 126, P. 129–139, 2013.
9. A. Bogomolov, A. Melenteva, D.J. Dahm, Journal of Near Infrared Spectroscopy, Vol. 21, P. 435–440, 2013.
10. A. Melenteva, V. Galyanin, E. Savenkova, A. Bogomolov, Food Chemistry, Vol. 203, P. 190–198, 2016.
11. S. Kucheryavskiy, A. Melenteva, A. Bogomolov, Talanta, Vol. 121, P. 144-152, 2014.
12. V. Galyanin, A. Melenteva, A. Bogomolov, Sensors & Actuators B: Chemical, Vol. 218, P. 97–104, 2015.

O-54

COMPARING K-NEAREST NEIGHBOUR CLASSIFICATION AND NEURAL NETWORKS FOR DIFFERENTIATION BETWEEN FRESH AND FROZEN/THAWED TUNA USING VIS-NIR SPECTROSCOPY

M.M. Reis[1], E. Martínez[2], E. Saitua[2], R. Rodríguez[2], I Pérez[2], I Olabarrieta[2].

[1]Food Assurance and Meat Quality Team, AgResearch, Ruakura Research Centre, 10 Bisley Road, Hamilton, New Zealand. [2]AZTI-Tecnalia, Parque Tecnológico de Bizkaia, Astondo Bidea – Edif. 609, E-48160, Derio-Bizkaia.
(emartinez@azti.es)*

Fresh tuna is an expensive product sold on local and international markets. Fish sourced from locations distant from market have reduced shelf-life and there is the motivation to freeze the fillets and thaw before sale. Fillets frozen bellow -60°C (practice found in the market) do not show visual characteristics when thawed and it is difficult to differentiate between fresh and frozen/thawed fillets. This study investigates the ability of Visible-Near InfraRed Spectroscopy (Vis-NIRS) to detect non-invasively whether a sample of tuna is fresh or if it has been frozen/thawed. Samples (n=135) were scanned by Vis-NIRS (370-2350nm) and subsequently frozen at -60°C. After five, twenty one or thirty five days the samples were thawed at 4°C for 24 hours and re-scanned. Two methods were compared for classification of the samples using Vis-NIRS: k-Nearest Neighbour classification and Neural Networks with a Principal Component Step (PCANNet) [1]. Double cross validation (DCV) was applied to test the validity of the classification models to discriminate between fresh and frozen/thawed samples [2]. The goal of this comparison was to test the ability of these two distinct approaches for classification of the samples. Overall PCANNet provided better performance than KNN (Table 1). While these results suggest the ability of Vis-NIRS to detect the effect of freezing thawing, both classification models were not able to account well for the variation sample to sample in the fresh status, which bias on predicted values. These differences in performance are discussed in this study.

Table 1 - Assessment of the performance of the classification models [3].

	KNN	PCANNet
Accuracy	0.74	0.82
Kappa	0.45	0.62
Sensitivity	0.68	0.77
Specificity	0.78	0.86
Positive Predictive Value	0.69	0.79
Negative Predictive Value	0.77	0.84

Acknowledgement: The financial support from Royal Society New Zealand within the New Zealand-EU International Research Staff Exchange Scheme (IRSES)/REPLAY is fully appreciated. Ekaitz Martinez acknowledges the scholarship of the Department of Environment, Territorial Planning, Agriculture and Fisheries of the Basque Country Government for the development of this work. . The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement nº (613688) FOOD INTEGRITY Project.

References:

[1] B. D. Ripley. Pattern Recognition and Neural Networks. (1996) Cambridge.
[2] P. Filzmoser, B. Liebmann, K.Varmuza. Repeated double cross validation.
Journal of Chemometrics, 23,160, (2009).
[3] M. Kuhn, K. Johnson, Applied Predictive Modeling. (2013) Springer, New York.

O-55

TAKE YOUR BREATH AWAY: CHEMOMETRICS FOR REAL-TIME BREATH MEASUREMENTS FROM IN-VIVO AROMA-RELEASE STUDIES

E. Szymańska[1,2], P. Brown[1,3], A. Ziere[4], S. Martins[4], M. Batenburg[4], F. Harren[3], L. Buydens[2]

[1]*TI-COAST, Science Park 904, 1098 XH Amsterdam, The Netherlands*
[2]*Radboud University, Institute for Molecules and Materials (IMM), Analytical Chemistry, P.O. Box 9010, 6500 GL Nijmegen, The Netherlands*
[3]*Radboud University, Institute for Molecules and Materials (IMM), Molecule and Laser Physics, P.O. Box 9010, 6500 GL Nijmegen, The Netherlands*
[4]*Unilever R&D, Olivier van Noortlaan 120, 3133 AT Vlaardingen, The Netherlands*
E.Szymanska@science.ru.nl

Real-time measurements of many low abundant volatile organic compounds (VOCs) in breath and air samples are already feasible due to progress in analytical technologies, such as proton-transfer reaction mass spectrometry (PTR-MS). Nevertheless, the information content of real-time measurements is not fully exploited, due to the lack of suitable data handling methods.

This study develops a data scientific procedure to enhance data analysis and interpretation of longitudinal, multivariate data sets from real-time, in-vivo, aroma-release studies [1]. The developed procedure includes an automated data pre-processing and a multivariate assessment of the test panel performance. A large multifactorial PTR-MS dataset is investigated that includes four experimental protocols, two tested food products, four aroma compounds and eight panelists.

Real-time measurements are converted into standardized breath profiles by pre-processing and ten kinetic parameters are derived. Next to this, panel performance is evaluated per experimental protocol and food product. Comprehensive information about panel performance, individual panelists, studied products, aroma compounds and kinetic parameters is extracted demonstrating great value of the developed approach.

Acknowledgements: This research received funding from the Netherlands Organization for Scientific Research (NWO) in the framework of the Technology Area COAST.

References:

[1] Szymańska et al, Analytical Chemistry, 87: 10338 (2015)

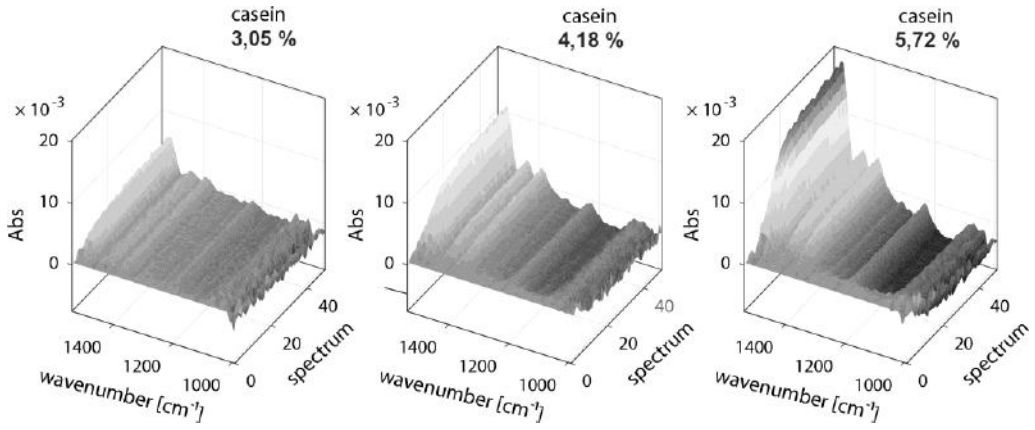
O-56

QUANTIFICATION OF CASEIN IN MILK BY ENZYMATIC PERTURBATION, FTIR AND MULTIWAY ANALYSIS

A. Baum[1], P. W. Hansen[2], L. Nørgaard[2], J. Sørensen[3], J. D. Mikkelsen[1]

[1]Center for BioProcess Engineering, Department of Chemical and Biochemical Engineering, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark
[2]FOSS Analytical, Foss Allé 1, 3400 Hillerød, Denmark
[3]Arla Foods amba, Innovation, Sønderupvej 26, 6920 Videbæk, Denmark
email: aba@kt.dtu.dk

Milk proteins are hardly distinguishable when being measured by FTIR. Although multivariate calibrations towards casein in milk exist, they break down as soon as the casein vs. total protein ratio changes significantly. Hence, the established correlations are found to be indirect and are furthermore goverened by underlying co-variance effects.
We propose a methodology which is utilizing an enzymatic perturbation to induce specific changes to casein to stand out from the interfering sample matrix. Chymosin is a protease which targets κ-casein by cleaving the Phe105-Met106 bond yielding para-κ-casein and a glycopeptide [1]. As a result of the hydrolysis all the casein proteins clot to form a creamy precipitate while the whey proteins remain in the supernatant. The reaction is measured in-situ by FTIR during the course of the reaction resulting in specific spectral evolution profiles which can be affiliated to the enzyme activity of chymosin [2]. If constant levels of chymosin are utilized the evolution intensity depends on the underlying casein concentration and is therefore a perfect measure for its quantification.



PARAFAC, PARAFAC2, N-PLS and u-PLS have been used to establish calibration models on the time-resolved data in a comparative manner. The method indicated robustness against possible interferences, i.e. from whey proteins, and was furthermore unaffected by the sample matrix since solely spectral change was utilized for calibration modeling.

Acknowledgement: This work was financed by FOSS Analytical, Hillerød, Denmark. ARLA Foods is acknow-ledged for providing milk samples and reference analysis.

References:

[1] J. L. Jensen, A. Mølgaard et al, Acta Crystallogr. Sect. D, 69, 901 (2013).
[2] A. Baum, A. S. Meyer et al, Anal. Chim. Acta, 790, 1 (2013)

POSTERS

P-001

OPTIMISATION OF MULTILINEAR GRADIENTS IN SERIALLY COUPLED COLUMNS ASSISTED BY PARETO PLOTS AND GENETIC ALGORITHMS

M.C. García-Álvarez-Coque, J.R. Torres-Lapasió, T. Álvarez-Segura

Department of Analytical Chemistry, University of Valencia, c/Dr. Moliner 50, 46100 Valencia (Spain)
celia.garcia@uv.es

The separation capability of single HPLC columns is often insufficient to resolve complex separations, due to the limited functionality. A relatively simple solution is the connection of two or more columns with different stationary phases in series (tandem columns). Each combination behaves as a totally new column, and often outperforms the results given by the individual columns. The full exploitation of this approach requires, however, the development of powerful interpretive optimisation strategies, able to scan efficiently the capabilities of the separation system. The most useful search configuration requires considering the column nature, length and order, and the profile of the gradient program, which should be preferably multi-linear. The number of candidate solutions to be examined is, however, easily so high that the calculation cannot be carried out on a systematic basis, and numerical computation is needed.

In this study, we examined the performance of five types of stationary phases (C18, pentafluorophenyl-C18, C4, cyano and phenyl) offering different selectivity, connected in series. A unique predictor system was developed in our laboratory, which implemented the different strategies with single and serially coupled columns, in both isocratic and gradient elution [1–3]. Multilinear gradients were optimised by Genetic Algorithms encoding the node position in discrete levels. A population size of 150 gradients with probabilities of 100%, (crossover), 3% (mutation), 5% (refreshing with global best) was operated. The populations generated along the optimisation were not discarded along the iterations, but saved along, so that at the end of the process, the results (gradient time and global peak purity) of the cumulative historical population gave rise to a Pareto front, from which the selection of the most convenient gradient and column set was made.

An interesting feature of the approach was that there is no need of additional experimental data: the same information used to characterise the single columns can be further used to prospect the performance of gradients with serially-coupled columns. Generally speaking, the separations found with tandem columns are closer to those achieved in two-dimensional liquid chromatography, with the advantage of requiring only a mono-dimensional chromatographic system. An outstanding agreement was found between experimental and predicted chromatograms.

Acknowledgement: This work was supported by Project CTQ2013-42558-P (Ministry of Economy and Competitivity, MINECO, Spain).

References:

[1] C. Ortiz-Bolsico, J.R. Torres-Lapasió, M.C. García-Álvarez-Coque. J. Chromatogr. A, 1350, 51 (2014).
[2] C. Ortiz-Bolsico, J.R. Torres-Lapasió, M.C. García-Álvarez-Coque. J. Chromatogr. A, 1373, 51 (2014).
[3] T. Álvarez-Segura, M.C. García-Álvarez-Coque, C. Ortiz-Bolsico, J.R. Torres-Lapasió. Current Chromatography, 2, 110 (2015).

P-002

MEASUREMENT OF RESOLUTION FOR HIGLY COMPLEX MUTI-ANALYTE SAMPLES

M.C. García-Álvarez-Coque, J.R. Torres-Lapasió, J.A. Navarro-Huerta, T. Álvarez-Segura

Department of Analytical Chemistry, University of Valencia, c/Dr. Moliner 50, 46100 Valencia (Spain)
celia.garcia@uv.es

**celia.garcia@uv.es, Tel: +34-96-354-4005, Fax: +34-96-354-4436*

The quantification of chromatograms requires finding out experimental conditions that separate each peak from the others, and is routinely carried out with the information provided by standards in a small number of runs, in a 'design of experiments' fashion. However, there are samples where the identity of the solutes is partially or fully unknown, and consequently, there are no standards support available. For these situations, the chromatographer lacks of a definite criterion to assist in the selection of the suitable experimental conditions, and computer-based optimisations are not applicable.

In this work, a resolution function valid for general situations (with or without standards) is developed and validated. The function is based on the automatic measurement of peak prominences [1,2], and is compared with the performance given by the peak purity criterion [3,4]. A Matlab application was developed to automate the analysis of chromatograms, subtracting the baselines with a modified BEADS algorithm (explained in detail in another communication), comprehensively detecting all peaks, and calculating peak properties, such as the peak prominences.

The separation of a mixture of the *o*-phthalaldehyde/N-acetyl-L-cysteine derivatives of the 19 primary proteic amino acids under gradient elution was taken as controlled case. The amino acid derivatives could be resolved only in excessively long analysis times, even using multi-linear or multi-isocratic gradients. If the analysis time was reduced, significant overlapping occurred for several compounds. These overlappings give access to interesting cases of study for the evaluation of the resolution functions. Using as training data the retention and peak width information of standards in a set of 10 isocratic conditions, the separation in around 1100 gradients, ranging from 5 to 27.5% acetonitrile, was predicted. Several levels of difficulty were considered: with or without differences in peak size, presence of unknown compounds, different noise levels, and real baselines.

The comparison is carried out based on the coincidence of the gradients chosen as Pareto-optimal by two criteria: the peak purity (which implies a comprehensive knowledge of the individual signals of each compound), and the peak prominence (whose information is only limited to the combined signal, without any prior knowledge of the number of compounds in the sample). If the two Pareto fronts coincide, the same gradients are selected as optimal. Finally, several fingerprint chromatograms of a tea extract, where the same sample was eluted with different gradients, were evaluated.

Acknowledgement: This work was supported by Project CTQ2013-42558-P (Ministry of Economy and Competitivity, MINECO, Spain).

References:

[1] T. Álvarez-Segura, A. Gómez-Díaz, C. Ortiz-Bolsico, J.R. Torres-Lapasió, M.C. García-Álvarez-Coque. J. Chromatogr. A, 1409, 79 (2015).
[2] T. Álvarez-Segura, E. Cabo-Calvet, J.R. Torres-Lapasió, M.C. García-Álvarez-Coque. J. Chromatogr. A, 1422, 178 (2015).
[3] S. Carda Broch, J.R. Torres Lapasió, M.C. García Álvarez-Coque. Anal. Chim. Acta, 396, 61 (1999).
[4] S.J. López Grío, G. Vivó Truyols, J.R. Torres Lapasió, M.C. García Álvarez-Coque. Anal. Chim. Acta, 433, 187 (2001).

P-003

MAXIMISATION OF THE INFORMATION IN CHROMATOGRAPHIC FINGERPRINTS

J.R. Torres-Lapasió, J.A. Navarro-Huerta, M.C. García-Álvarez-Coque, T. Álvarez-Segura

Department of Analytical Chemistry, University of Valencia, c/Dr. Moliner 50, 46100 Valencia (Spain)
jrtorres@uv.es

A chromatographic fingerprint (CF) is the chromatogram of a highly complex sample, where multiple peaks of unknown identity are present, but whose overall separation pattern typifies the sample and allows its classification. This kind of chromatogram has been demonstrated very useful for the characterisation of natural products, such as medicinal herbs and their extracts. The information content of a chromatographic fingerprint correlates to the number of resolved peaks.

In a recent report, we proposed the unsupervised measurement of peak prominences to quantify the separation level of real highly complex multi-analyte samples [1,2]. The optimisation of CFs is challenging. Not only there are no standards that could be used to forecast separation conditions of optimal selectivity, but the identity of peaks in chromatograms taken in different conditions can be ambiguous as well. Moreover, the peak size differences affect seriously the separation (i.e., a minor peak in a longer elution time can become undetected). In spite of the lack of knowledge of the compound identity, some major peaks can be tracked between independent training experiments, which allow building retention models for these compounds.

In this study, the local peak capacity is measured between each major consecutive peak and compared with the peak number found between them. The proposed optimisation approach changes the node positions of a multi-linear gradient program, aimed to enlarge the elution window size of the congested time subdomains (i.e., chromatogram regions with a high number of peaks with regard to its separation window), and to narrow the regions whose peak count is smaller with regard to its local peak capacity. The separation is done on a probabilistic basis.

Acknowledgement: This work was supported by Project CTQ2013-42558-P (Ministry of Economy and Competitivity, MINECO, Spain).

References:

[1] T. Alvarez-Segura, A. Gómez-Díaz, C. Ortiz-Bolsico, J.R. Torres-Lapasió, M.C. García-Alvarez-Coque. J. Chromatogr. A, 1409, 79 (2015).

[2] T. Alvarez-Segura, E. Cabo-Calvet, J.R. Torres-Lapasió, M.C. García-Álvarez-Coque. J. Chromatogr. A, 1422, 178 (2015).

P-004

THEORETICAL MODELS AND NEW SPLINE INTERPOLATION TECHNIQUE IN THE MODELING OF HILIC RETENTION BEHAVIOR ON THE EXAMPLE OF OLANZAPINE AND ITS EIGHT IMPURITIES

A. Tumpa[1], S. Mišković[2], Z. Stanimirović[3], B. Jančić-Stojanović[1], M. Medenica[4]
B.

[1] University of Belgrade – Faculty of Pharmacy, Department of Drug Analysis, Vojvode Stepe 450, 11000 Belgrade, Serbia (e-mail: jancic.stojanovic@pharmacy.bg.ac.rs)

[2] University of Belgrade – Faculty of Mathematics, Department for Computer Science, Studentski trg 16/IV, 11000 Belgrade, Serbia

[3] University of Belgrade – Faculty of Mathematics, Department for Numerical Mathematics and Optimization, Studentski trg 16/IV, 11000 Belgrade, Serbia

[4] University of Belgrade – Faculty of Pharmacy, Department of Physical Chemistry and Instrumental Methods, Vojvode Stepe 450, 11000 Belgrade, Serbia

Taking into account that hydrophilic interaction chromatography (HILIC) as analytical method is relatively young compared to the other techniques, retention modeling could still bring scientifically valuable data to the field [1]. Therefore, in this paper olanzapine and its eight impurities were selected as a test mixture, considering that they have never been analyzed in HILIC before. Their investigation on four different HILIC columns (bare silica, cyanopropyl, diol and zwitterionic) has been carried out. The mixture of nine structurally similar substances allows the examination of complex HILIC retention behavior depending on the chemical properties of the analytes, as well as of the stationary phase. In order to describe the nature of the relationship between the retention and the stronger eluent content in the mobile phase, experimentally obtained data were fitted to several theoretical (localized adsorption, non-localized partition, quadratic and mixed) models [2]. Results show that the best fit is quadratic model with the highest R^2 values, but its usage has some drawbacks. With the aim to improve the possibility to predict retention behavior in HILIC, a new empirical model was proposed. For that purpose spline interpolation technique was performed, by dividing the experimental range into several subdivisions [3]. This type of interpolation was performed for the first time in the chromatographic field. The estimation of the polynomial equations was performed using cross validated coefficient of determination (Q^2). Obtained Q^2 values pointed out the goodness of fit of the model, as well as its good predictive capabilities. At the end, the prediction capabilities were experimentally verified, under randomly chosen conditions from the experimental range. The errors in prediction were all under 10% which is satisfying for HILIC.

Acknowledgements: The authors thank the Ministry of Education, Science and Technological Development of the Republic of Serbia for supporting these investigations in Projects 172052 and 174010.

References:

[1] Y. Guo, Analyst, 140, 6452 (2015).
[2] E. Tyteca, A. Périat, S. Rudaz, G. Desmet, D. Guillarme, J. Chromatogr. A, 1337, 116 (2014).
[3] A.M. Burden, J.D. Faires, Numerical Analysis, nineth ed., Brooks/Cole, Cengage Learning, pp. 144 (2011).

P-005

QUALITY BY DESIGN APPROACH IN THE DEVEOPMENT OF HILIC METHOD FOR THE ANALYSIS OF BILASTINE AND ITS DEGRADATION IMPURITIES

J. Terzić[1], I. Popović[1], A. Tumpa[2], A. Stajić[2], B. Jančić-Stojanović[2]

[1] Medicines and Medical Devices Agency of Serbia, Vojvode Stepe 458, 11000 Belgrade, Serbia
[2] University of Belgrade – Faculty of Pharmacy, Department of Drug Analysis, Vojvode Stepe 450, 11000 Belgrade, Serbia
(e-mail: jancic.stojanovic@pharmacy.bg.ac.rs)

New trends in development of chromatographic methods impose the application of QbD approach which implies that that quality should be built in the method. Different tools could be used in order to achieve this aim and one of the most used ones is design of experiments (DoE) methodology [1]. In this paper, QbD concept is applied for the developement of hydrophilic interaction liquid chromatographic (HILIC) method for the analysis of bilastine and its degradation impurities. Taking into account the complexity of HILIC [2], special attention should be paid to method development, esspecially when substance and its degradation products are analyzed in chromatography for the first time. Also, there is only one published paper related to the application of QbD in HILIC [3]. In aim to conduct study in line with QbD rules, next steps sould be performed: 1) Definition of analytical target profile (ATP) and critical quality attributes (CQAs); 2) Quality risk assessment (QRA) and deffinition of critical process parameters (CPPs); 3) Investigation of knowledge space and Critical quality attributes modeling; 4) Optimization and design space creation; 5) Verification of optimal point and design space; 6) Robustness testing and 7) Method validation. Considering that bilastine and its degradation impurities are investigated for the first time, after carefully conducted preliminary study, DoE was employed. CPPs which have the most influence on method performance were defined as acetonitrile content in the mobile phase, pH of the aqueous phase and ammonium acetate concentration in the aqueous phase. Box-Behnken design was applied for establishing relationship between CPPs and CQAs. The defined mathematical models and Monte Carlo simulations were used to identify the design space. Next, fractional factorial design was applied for experimental robustness testing. Finally, developed HILIC method was validated and applied for the real sample analysis. Optimal and robust chromatographic conditions for bilastin and its degradation impurities determination were: the analytical column Luna® HILIC (100 mm x 4.6 mm, 5 μm particle size); mobile phase consisted of acetonitrile – aqueous phase (50 mM ammonium acetate, pH adjusted to 5.3 with glacial acetic acid – 90.5:9.5 v/v); column temperature 30 ºC, mobile phase flow rate 1 mL min^{-1} and wavelength of detection 275 nm.

Acknowledgements: The authors thank the Ministry of Education, Science and Technological Development of the Republic of Serbia for supporting these investigations in Project 172052.

References:

[1] O.S. Pinzauti, S. Furulanetto, Anal. Bioanal. Chem. 405, 443 (2013).
[2] Y. Guo, Analyst 140, 6452 (2015).
[3] M. Jovanović, T. Rakić, A. Tumpa, B. Jančić-Stojanović, J. Pharm. Biomed. Anal. 10, 42 (2015).

P-006

COMBINATION OF BEADS AND WAVELETS FOR CHROMATOGRAPHIC BASELINE DETRENDING

S. López-Ureña[1], J.R. Torres-Lapasió[2], M.C. García-Álvarez-Coque[2], R. Donat[1]

[1]Dept. Applied Mathematics, Faculty of Mathematics, C/Dr. Moliner 50, 46100-Burjassot, Valencia, Spain.
[1]Dept. Analytical Chemistry, Faculty of Chemistry, C/Dr. Moliner 50, 46100-Burjassot, Valencia, Spain.
sergio.lopez-urena@uv.es

The quantification of signals in High-Performance Liquid Chromatography (HPLC) is critically affected by the presence of background signal and noise, and its removal is a mandatory first step in any data processing treatment. State-of-the-art algorithms aimed to the suppression of baseline effects (e.g., airPLS [1] and backcor [2]) split the raw signal in two contributions, namely the combination of noise and signal, and baseline itself.

Very recently, a new algorithm called *Baseline Estimation And Decomposition with Sparsity* (BEADS) was proposed, which, as a novelty, presents the capability of performing a full decomposition in baseline, noise and neat signal. A second advantage is the speed of computation, associated to the use of a Majorization-Minimization approach, with calculation times below one second for a standard chromatogram with ten thousand points. On the other hand, wavelet transforms are a consolidated tool in the field of chromatographic data processing, which separate the raw signal in a sum of different frequency contributions, whose decomposition level can be arbitrarily set.

In this work, both techniques are combined in a single algorithm, with the purpose of exploiting their best features: the capability of wavelets for processing each frequency level isolatedly, and the fast triple decompositions of BEADS by setting cut-off frequencies. In the study, several wavelet-like families are considered, one of them non-linear. Several complex baseline examples are resolved and compared with the solution offered by the optimal BEADS decomposition. For simple cases, the performance of BEADS is excellent, with or without wavelets, but harder cases benefit from the concurrence of both techniques.

Acknowledgements: This work was supported by Projects CTQ2013-42558-P and MTM2014-54388 (Ministry of Economy and Competitivity, MINECO, Spain) and the FPU14/02216 grant (Ministry of Education and Culture and Sports, MECD, Spain)

References:

[1] Z.-M. Zhang, S. Chen, Y.-Z. Liang, Analyst, 135(5), 1138 (2010)
[2] V. Mazet, C. Carteret, D. Brie, J. Idier, B. Humbert, Chemometr. Intell. Lab. Syst., 76(2), 121 (2005)
[3] X. Ning, I.W. Selesnick, L. Duval, Chemometr. Intell. Lab. Syst., 139, 156 (2014)

P-007

MOISTURE CONTENT DETERMINATION OF FREEZE-DRIED PRODUCTS BY NEAR-INFRARED SPECTROSCOPY: A CASE STUDY FOR UNIVERSAL REGRESSION MODEL

M. Clavaud [1,2], Y. Roggo [1], K. Degardin [1], Ph. Hubert [2], E. Ziemons [2]

[1]Vibrational Spectroscopy and Forensics, Complaint and Counterfeit, Quality Control for Commercial Bulk Products, F. Hoffmann-La Roche Ltd., Wurmisweg, CH-4303 Kaiseraugst, Switzerland
[2]University of Liege (ULg), CIRM, Department of Pharmacy, Laboratory of Analytical Chemistry, Quartier Hôpital, Avenue Hippocrate 15, B36, B-4000 Liege, Belgium
klara.degardin@roche.com

Karl Fischer (KF) titration is the reference method for moisture content (MC) determination in the pharmaceutical industry. Near-infrared spectroscopy (NIRS) is considered the most suitable alternative technique [1,2]. Indeed, NIRS is a safe and fast method which does not require sample preparation. Nevertheless, the development and validation phases are time-consuming. In addition, the NIRS methods presented so far were mostly product specific. The main objective of this study is to highlight that an universal calibration model can be validated for several freeze-dried products in order to speed up the validation time. This objective was led in two steps. A universal model was first evaluated. A calibration set and a validation set were built up with three freeze-dried products. An antibody drug conjugate, a large molecule and a small molecule all packed in sealed vials were used to introduce more variability. Regression methods were then compared in order to optimize the prediction values.

In order to have a wide range of MC, the vials were either humidified, dried or kept intact. Then, NIRS spectra of each vial were acquired on two similar NIRS devices. MC was determined by conventional KF titration method. The MC range from 0.05% to 4.96% was used. The programming language Python version 3.4 and Scikit-learn module version 0.15.2 were used as chemometric tools to compute the models. The classical Partial least square (PLS) regression was employed. PLS models were developped for each of the three freeze-dried products independently. Then a single PLS model was computed using the spectra of the three freeze-dried products. With the universal regression model, the standard error of prediction (SEP) increased from 0.22% (with independant models) to 0.29%. The universal regression model had therefore to be optimized. Alternative regression models were finally evaluated: Decision Tree, Ridge regression (Bayesian), K Nearest Neighbors (KNN), and Support Vector Regression (SVR) using the Radial Basis Function (RBF). SVR RBF provided the most accurate results for an universal model with a SEP of 0.15%.

In conclusion, a database with different freeze-dried products can be used for MC using a unique regression model. This study shows the use of SVR RBF as a benefit tool for a regression model. SVR RBF provides better results compared to classical PLS. The use of a unique model developped with different freeze-dried products is an advantage in terms of time and accuracy.

References:

[1] Y. Roggo, P. Chalus, L. Maurer, C. Lema-Martinez, A. Edmond, N. Jent, Journal of Pharmaceutical and Biomedical Analysis, 44, 683 (2007).
[2] C. De Bleye, P.-F. Chavez, J. Mantanus, R. Marini, Ph. Hubert, E. Rozet, E. Ziemons, Critical review of near-infrared spectroscopic methods validations in pharmaceutical applications, Journal of Pharmaceutical and Biomedical Analysis, 69, 125 (2012).

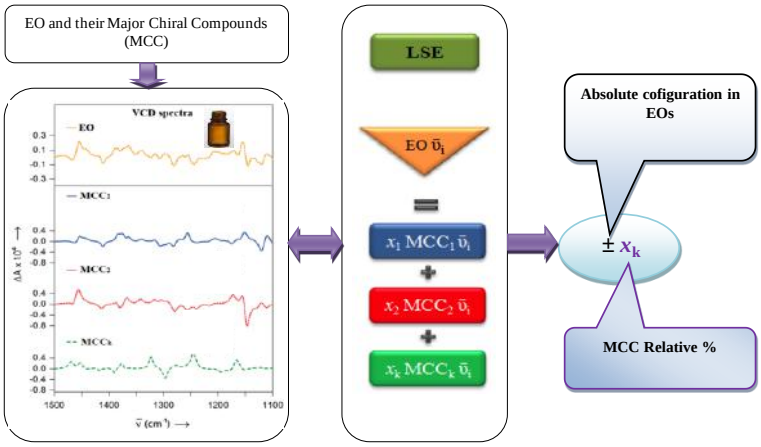
P-008

VIBRATIONAL CIRCULAR DICHROISM (VCD) AND LEAST SQUARE ESTIMATION (LSE) APPLIED TO *BUBONIUM GRAVEOLENS* ESSENTIAL OIL.

M. E. A. Said[1,3], P. Vanloot[1], I. Bombarda[1], J. V. Naubron[2], N. Dupuy[1], C. Roussel[3].

[1]Aix-Marseille Université, EA4672 LISA Equipe METICA, Case 451, Av. Escadrille Normandie Niémen, 13397 Marseille Cedex 20, France
[2]Aix-Marseille Université, Spectropole, service 511, F-13397 Marseille, France
[3]Aix-Marseille Université, Ecole Centrale, CNRS, ISM2 UMR 7313, Marseille, France
email : saidmedamin@gmail.com

The spectroscopic IR and vibrational circular dichroism (VCD) chiral signatures of essential oils (EOs) were used to obtain the configuration and the relative percentage of the major chiral compounds.[1] For this purpose, a method was developed based on the use of VCD spectra of pure enantiomers, VCD spectra of EOs and a mathematical model (least square estimation). In order to validate the method, the chemical compositions of EOs of *Bubonium graveolens*, grown in Algeria, were determined by gas chromatography coupled to mass spectrometry (GC–MS). The major chiral compounds in the EOs were *cis*-chrysanthenyl acetate (**1**), 6-oxocyclonerolidol (**2**) and the recently disclosed *cis*-8-acetoxychrysanthenyl acetate (**3**). The IR and VCD spectra of the pure molecules (**1** to **3**) and crude EOs were acquired to build the model. To obtain the configuration and relative percentage of the major chiral compounds, the VCD spectra of EOs were modelized as a linear weighted combination of the individual spectra of pure enantiomers. The value of each weighting gives the relative percentage of the chiral compounds while the signs address the correctness of the enantiomer employed for the model.



References:

[1] M. E. A. Said, P. Vanloot, I. Bombarda, J. V. Naubron, E. M. Dahmane, A. Aamouche, M. Jean, N. Vanthuyne, N. Dupuy, et C. Roussel, *Anal. Chim. Acta*, 903, 121‑130, (2016).

P-009

THE COLUMN-WISE *K*-FOLD (*CKF*) ALGORITHM FOR FAST AND EFFICIENT PCA CROSS-VALIDATION

E. Saccenti[1], J. Camacho[1]

[1]Wageningen University, Dreijenplein 10, 6703 HB, Wageningen, The Netherlands
[2]University of Granada, C/ Periodista Daniel Saucedo Aranda, s/n, 18071, Granada, Spain

email: esaccenti@gmail.com

The element-wise *k*-fold (*ekf*) algorithm [1-3] is the most used cross-validation algorithm to determine the optimal number of components in PCA but the observation-wise *k*-fold operation makes it computationally costly and hampers its application on large data set.

To overcome this limitation we propose here a modified version of the algorithm obtained by removing the observation-wise *k*-fold operation. The new algorithm, named column-wise *k*-fold (*ckf*), is computationally very efficient and bears some interesting mathematical properties that are discussed [4].

The *ekf* and *ckf* are compared both theoretically and practically on an array of 116 simulated data sets and 19 real data sets from different research areas. Advantages and possible limitations of both algorithms are also discussed.

Our results show that *ckf* is a fast and effective alternative to the classical *ekf* algorithm and opens the possibility of applying cross-validation on very large data sets.

Acknowledgement: This work was partially supported by European Commission funded FP7 project INFECT (contract number: 305340), the Spanish Ministry of Science and Innovation and FEDER funds from the European Union through grant TEC2011-22579.

References:

[1] Camacho, J. and A. Ferrer, *Cross-validation in PCA models with the element-wise k-fold (ekf) algorithm: Practical aspects.* Chemometrics and Intelligent Laboratory Systems, 2014. **131**(0): p. 37-50.

[2] Camacho, J. and A. Ferrer, *Cross-validation in PCA models with the element-wise k-fold (ekf) algorithm: theoretical aspects.* Journal of Chemometrics, 2012. **26**(7): p. 361-373.

[3] Bro, R., et al., *Cross-validation of component models: A critical look at current methods.* Analytical and bioanalytical chemistry, 2008. **390**(5): p. 1241-1251.

[4] Saccenti, E. and J. Camacho, *On the use of the observation-wise k-fold operation in PCA cross-validation.* Journal of Chemometrics, 2015. **29**: p. 467 - 478.

P-010

APPLICATION OF THE EXPONENTIALLY MODIFIED GAUSSIAN (EMG) FUNCTION TO THE MULTIVARIATE CURVE RESOLUTION OF HIGHLY ASYMMETRIC VOLTAMMOGRAMS

N. Serrano[1], S. Cavanillas[1], J.M. Díaz-Cruz[1], C. Ariño[1], M. Esteban[1]

[1]Department of Analytical Chemistry, University of Barcelona, Martí i Franquès 1-11, E-08028 Barcelona (Spain).
nuria.serrano@ub.edu

A multivariate curve resolution method is proposed for the analysis of overlapping and highly asymmetric voltammograms typically obtained in non-reversible electrochemical systems. The method is based on the least squares fitting of the well-known exponentially modified Gaussian (EMG) function, which had been previously used for the resolution of coeluted chromatographic peaks [1]. The main advantage of this function is that one of the adjustable parameters is just the area under the peak, which is a magnitude more reliable than the peak height in the case of the considered electrochemical systems.

In contrast with chromatograms, which only present queues at the right side of the maximum, voltammograms may exhibit such deformation at both sides of the peak potential. This is why two versions of the EMG function have been used:

$$I_- = \frac{a}{2d} exp\left(\frac{c^2}{2d^2} - \frac{E-b}{d}\right) \text{erfc}\left(\frac{1}{\sqrt{2}}\left(\frac{c}{d} - \frac{E-b}{c}\right)\right) \quad (1)$$

$$I_+ = \frac{a}{2d} exp\left(\frac{c^2}{2d^2} + \frac{E-b}{d}\right) \text{erfc}\left(\frac{1}{\sqrt{2}}\left(\frac{c}{d} + \frac{E-b}{c}\right)\right) \quad (2)$$

where *'I'* is the current, *'E'* is the potential, *'a'* is the peak area, *'b'* is a potential value close to the maximum, *'c'* is related to the peak width and *'d'* is related to the peak symmetry. The term 'erfc' denotes the complementary error function. When relatively symmetric peaks are present, they can be easily described by a double gaussian function [2]:

$$I = a\ exp(-\ c\ (E - b)^2) \quad (3)$$

where *'a'*, *'b'* are the peak height and position, respectively, and the width parameter *'c'* is different at both sides of the peak. In a first overview of the data, the method decides which function (according to Eqns. 1, 2 or 3) is fitted to every overlapping signal and then adjusts a set of parameters producing an optimal lack of fit. Such parameters (especially *'a'* and *'b'*) can be further used to obtain information on the concentration profiles and the nature of the electrochemical processes involved.

After a preliminary test with simulated data, the proposed method has been successfully applied to linear sweep and differential pulse voltammograms measured in the experimental system Cd(II)-Cysteine by using both mercury and carbon screen-printed electrodes. The results are compared with those attained by the previous approaches using double gaussian [2] and asymmetric logistic [3] functions.

Acknowledgement: This work is supported by the Ministry of Science and Innovation of Spain (Project CTQ2012–32863) and the Generalitat of Catalonia (Project 2014SGR269).

References:

[1] Y. Kalambet, Y. Kozmin, K. Mikhailova, I. Nagaev, P. Tikhonov, J. Chemometrics 25 (2011) 352-356.
[2] S. Cavanillas, C. Ariño, J.M. Díaz-Cruz, M. Esteban, Anal. Chim. Acta 689 (2011) 198-205.
[3] M. Kooshki, J.M. Díaz-Cruz, H. Abdollahi, C. Ariño, M. Esteban, Analyst 136 (2011) 4696-4703.

P-011

Comparison of performance of multilinear MCR-ALS, PARAFAC1 and PARAFAC2 in the analysis of comprehensive two dimentional gas chromatorgphy-mass spectrometry (GCxGC-MS) data

Yahya Izadmanesh [1], Elba Garreta [2], Roma Tauler[2], Jahan B. Ghasemi[1]

[1]*Faculty of Chemistry, K. N. Toosi University of Technology, Tehran, Iran*
[2], *Institute of Environmental Assessment and Water Research, Spanish Council for Scientific Research (CSIC), Jordi Girona 18, Barcelona, 08034, Spain*
Email: yizadmanesh@mail.kntu.ac.ir

Comprehensive two dimensional gas chromatography-mass spectrometry (GC×GC-MS) is a valuable technique for the analysis of complex biological and non-biological samples, because it provides enhanced sensitivity and boosted separation and peak detection capacity [1]. However, chemometric data analysis is often required to extract comprehensive information from the data. In GC×GC-MS data, within run and between-run retention time shifts occur frequently, which causes deviation from trilinearity. In these cases, application of the second-order chemometric methods based on the fulfillment of the trilinear model, such as parallel factor analysis, PARAFAC, is rather cumbersome [2]. PARAFAC2 [3] is a modified version of PARAFAC in which strict trilinearity fulfillment is not required, sometimes useful when retention time shifts are present. However, PARAFAC2 is computationally more complex and expensive, and it does not allow for the application of constraints like non-negativity or unimodality. Moreover, in the presence of interferences, the performance of this method becomes even worse [4]. It has already been proved that bilinear multivariate curve resolution-alternating least squares MCR-ALS is insensitive to these deviations [4,5]. In this work, we have studied the performance of the bilinear and multilinear versions of the MCR-ALS [6] method and results were compared with those obtained by PARAFAC and PARAFAC2. All these methods were applied to experimental data set from a metabolomics study of daphnia magna. The overall results favor bilinear and multi-linear MCR-ALS modelling over PARAFAC1 and PARAFAC2 modelling, especially in the presence of potential interferents.

Keywords: Chromatography; Multilinearity; Second order advantage

References

[1] H. Parastar, J.R. Radović, M. Jalali-Heravi, S. Diez, J.M. Bayona, R. Tauler, Anal. Chem, 83 (2011) 9289-9297.
[2] D.K. Pinkerton, B.A. Parsons, T.J. Anderson, R.E. Synovec, Anal. Chim. Acta, 871 (2015) 66-76.
[3] H.A. Kiers, J.M. Ten Berge, R. Bro, PARAFAC2-Part I., J. Chem., 13 (1999) 275-294.
[4] S.A. Bortolato, A.C. Olivieri, Analytica chimica acta, 842 (2014) 11-19.
[5] H. Parastar, M. Jalali-Heravi, R. Tauler, Chem. Intel. Lab. Sys., 117 (2012) 80-91.
[6] A. Malik, R. Tauler, , Anal. Chim. Acta, 794 (2013) 20-28.

P-012

Statistically Correlated Sensor Selection Under Adverse Conditions

Adam C. Knapp, Kevin J. Johnson

Chemistry Division, U.S. Naval Research Laboratory, 4555 Overlook Ave. SW, Washington DC, USA
adam.knapp.ctr@nrl.navy.mil

Minimizing the global error of a sensor array when presented with many sensor options presents a significant challenge to the practitioner due to the combinatorial explosion in number of sensor configurations. Further complicating the sensor selection process are the possible presence of multiple analytes and statistical correlations amongst the sensors all of which are subject to noise that is often concentration dependent. Despite the challenges that both numeracy of configurations and noise presents, this work shows that optimal, noise minimizing selections for sensors may be made using convex optimization to minimize the theoretical global error of a proposed sensor array via the Fisher information matrix. Applications are discussed and comparisons amongst various sensor array configurations presented.

Acknowledgement: The authors would like to thank the Office of Naval Research and the Naval Research Laboratory for funding this work. Adam C. Knapp would like to thank the NRC Research Associateship Program.

P-013

ON THE SET OF FEASIBLE SOLUTIONS IN CHEMICAL KINETICS

H. Schröder[1,2], M. Sawall[1], K. Neymeyr[1,2]

[1]*Department of Mathematics, Universität Rostock, Ulmenstraße 69, 18057 Rostock, Germany.*
[2]*Leibniz Institute for Catalysis, Albert-Einstein-Straße 29a, 18059 Rostock, Germany.*
henning.schroeder2@uni-rostock.de

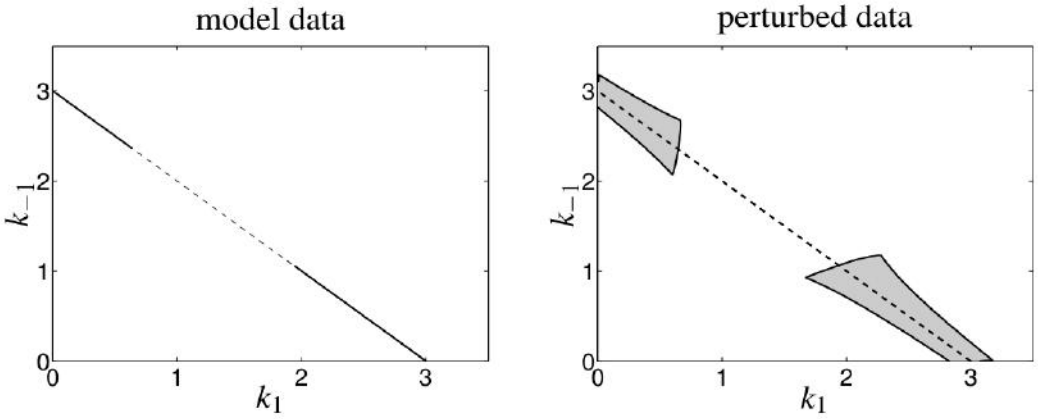
If for a chemical reaction with a known reaction mechanism the concentration profiles are accessible only for certain species, e.g. only for the main product, then often the reaction rate constants cannot uniquely be determined from the concentration data. This is a well-known fact which includes the so-called slow-fast ambiguity [1,2].

This work combines the question of unique or non-unique reaction rate constants with factor analytic methods of chemometrics. The idea is to reduce the rotational ambiguity of pure component factorizations by considering only those concentration factors which are possible solutions of the kinetic equations for a properly adapted set of reaction rate constants. The resulting set of reaction rate constants corresponds to those solutions of the rate equations which appear as feasible factors in a pure component factorization.

The new analysis of the ambiguity of reaction rate constants extends recent research activities on the Area of Feasible Solutions (AFS). The consistency with a given chemical reaction scheme is shown to be a valuable tool in order to reduce the AFS.

In this poster the new methods are presented and applied to model as well as experimental FT/IR data.

The figures show the



ambiguity of reaction rate constants for the kinetic model A ★ B without noise (left) and with noise (right).

References:

[1] N. W. Alcock, D. J. Benton, P. Moore, Trans. Faraday Soc., 66, 2210-2213 (1970).
[2] S. Vajda, H. Rabitz, J. Phys. Chem., 92, 701-707 (1988).
[3] H. Schröder, M. Sawall, C. Kubis, D. Selent, D. Hess, R. Franke, A. Börner, K. Neymeyr, Preprint Universität Rostock (2016).

P-014

AUTOMATIC BASELINE SUBTRACTION, PEAK DETECTION AND QUANTIFICATION OF COMPLEX CHROMATOGRAMS

J.R. Torres-Lapasió[1], J.A. Navarro-Huerta[1], S. López-Ureña[2], M.C. García-Álvarez-Coque[1]

[1]Dept. Analytical Chemistry, Faculty of Chemistry, C/Dr. Moliner 50, 46100-Burjassot, Valencia, Spain.
[2]Dept. Applied Mathematics, Faculty of Mathematics, C/Dr. Moliner 50, 46100-Burjassot, Valencia, Spain.
jrtorres@uv.es

Modern High-Performance Liquid Chromatography (HPLC) instruments are able to provide highly complex signals suitable for routine analysis, from which the relevant information should be extracted. Nowadays, the data processing step of such samples may constitute a bottleneck, conditioning sample throughput. Problems such as the presence of irregular baselines, noise, multiple peaks (sometimes highly overlapped), should be addressed. This should be done preferably with little (or none) user supervision, for a maximal benefit and highest speed.

In this work, several tools for the automatic analysis of highly complex signals are developed and evaluated, with the final aim of using them for the gradient optimisation of fingerprint chromatograms. The BEADS algorithm [1], using self-configured working parameters, is used for baseline correction. An objective function was developed to measure the quality of the recovered baseline. Peak detection made use of the original signal and its smoothed and first-order derivative transforms, the two latter calculated with the Savitzky-Golay algorithm adjusting the window size until obtaining uncorrelated noise [2].

The process implies operations such as the iterative elimination of false positives, peak merging of residual side signals, and recursive baseline re-evaluation for partially overlapped peaks. By default, the critical peak size is established attending to the noise, once the peak regions have been removed until self-consistency. The optimisation is based on the measurement of the peak prominence (area fraction of the emerging part of a peak with regard to the two valleys delimiting it) [3], and parameters such as peak area, height and limits. For testing the algorithm, several complex chromatograms corresponding to fingerprint extracts and gradients of Brij-35 (a non-ionic surfactant) in micellar conditions, were studied.

Acknowledgements: This work was supported by Projects CTQ2013-42558-P and MTM2014-54388 (Ministry of Economy and Competitivity, MINECO, Spain) and the FPU14/02216 grant (Ministry of Education and Culture and Sports, MECD, Spain)

References:

[1] Z.-M. Zhang, S. Chen, Y.-Z. Liang, Analyst, 135, 1138 (2010)
[2] G. Vivó-Truyols, J.R. Torres-Lapasió, A.M. van Nederkassel, Y. Vander Heyden, D.L. Massart, J. Chromatogr. A. 1096, 133 (2005)
[3] T. Alvarez-Segura, E. Cabo-Calvet, J.R. Torres-Lapasió, M.C. García-Álvarez-Coque, J. Chromatogr. A 1422, 178 (2015)

P-015

DEVELOPMENT OF MULTIPLE SELF ORGANIZING MAPS FOR CLASSIFICATION PROBLEMS

C. Krongchai, S. Funsueb, J. Jakmunee and S. Kittiwachana*

Department of Chemistry, Faculty of Science, Chiang Mai University, Chiang Mai, 50200 Thailand
E-mail: chanida511@yahoo.co.th

Multiple self organizing maps (mSOMs) are an extension of a conventional SOM where additional maps are used to represent some parts of data instead of binding them into a single map. In this presentation, mSOMs, in an unsupervised manner, were applied for classification problems. The classification rule was simply based on identification of the trained map unit which was the most similar to an unknown sample. To determine the optimum size and arrangement of the SOM maps, a growing self organizing map (GSOM) algorithm was adapted. Validated based on some datasets such as Iris flower [1], breast cancer [2], thermal analysis of polymer [3] and soil datasets, the predictive results were compared with the classification model using a single SOM, some previously established Kohonen network methods such as counter propagation network (CPN) and supervised Kohonen network (SKN) and a classical non-linear classifier such as *k*-nearest neighbors (*k*-NN). The reliability of the models was evaluated using a bootstrap methodology and this resulted in some statistical indices based on the majority vote including percentage predictive ability (%PA), percentage model stability (%MS) and percentage correctly classified (%CC). In any case, the mSOMs showed significantly better predictive ability when compared to the single SOM. The developed mSOMs could be confronted by CPN, SKN and *k*-NN when applied to the Iris flower, polymer and soil datasets. One possible explanation for the improvement could be that each of the maps was executively trained using the samples from only one specific group or class membership, and therefore each of the trained maps could efficiently learn and independently form itself to represent the characteristic variation within the class data. On the other hand, if the samples from all of the class memberships were organized onto a single map, some parts of the map were used to express the dissimilarity between the samples from the different classes and they were not useful for the classification task.

Acknowledgement: This research was financially supported by Chiang Mai University and the "Research Fund for DPST Graduate with First Placement" with Grant No. 005/2555. C.K. would like to thank the Development and Promotion of Science and Technology Talents Project (DPST). The soil data was provided by Ubon Ratchathni Rice Research Center, Ubon Ratchathani, Thailand.

References:

[1] R. A. Fisher, (1936). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml/datasets.html].
[2] W. N. Street, W. H. Wolberg, O. L. Mangasarian, (1993). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml/datasets.html]. CA: University of Wisconsin, Clinical Sciences Center, Madison
[3] G. R. Lloyd, R. G. Brereton, J. C. Duncan, Analyst, 133, 1046-1059 (2008).

P-016

DISCRIMINANT ANALYSIS IS AN INAPPROPRIATE METHOD OF AUTHENTICATION

O. Rodionova[1], A. Pomerantsev[1]

[1]Semenov Institute of Chemical Physics RAS, Kosygin str. 4, 119991, Moscow, Russia
rcs@chph.ras.ru

Authentication is the process of determining whether an object is, in fact, what it is declared to be. In some cases the answer is found by means of direct chemical analysis, which confirms that the product quality meets technical/regularity documentation. As a rule, these analyses are time and labor consuming. Another approach is to conduct some quick, relatively cheap, and often non-destructive measurements with subsequent data processing by means of chemometrics. When claiming authentication as a goal, analysts often substitute authentication task by solving discrimination problems. Pattern recognition encloses a big variety of different methods and techniques. Each type of problem demands requires an application of relevant methods. A well constructed discrimination method will perfectly classifies a new sample, only if in case this sample is a member of a one of the predefined classes. However, in case a new sample does not belong to any of such classes, the discriminant analysis is unable to properly define the membership of the sample. Thus, discrimination methods are inappropriate for solving authentication problems. Class-modeling methods intend to develop the acceptance area around the target class, and, thus, delimit the target objects from any other objects and classes. This is the reason why only the one-class classifiers should be used for authentication [1].

Our considerations are illustrated by a real-world example and comparison the results provided by two types of methods. They are Partial Least Squares- Discriminant Analysis, PLS-DA, and Data Driven Soft Independent Modeling of Class Analogy, DD-SIMCA.

References:

[1] Rodionova,O., Titova, A., Pomerantsev, *Trends Anal. Chem*. **2016,** 78(4), 17-22.

P-017

SELECTIVITY-RELAXED CLASSICAL LEAST SQUARES CALIBRATION AND SELECTIVITY MEASURES

J.H. Kalivas [1], J. Ferré [2], A.J. Tencate [1]

[1]Department of Chemistry, Idaho State University, 921 S. 8th, Stop 8026, Pocatello, Idaho, USA
[2]Chemometrics, Qualimetrics and Nanosensors Group, Department of Analytical Chemistry and Organic Chemistry, Universitat Rovira i Virgili, 43007 Tarragona, Spain
emails (kalijohn@isu.edu, joan.ferre@urv.cat)

There are many strategies to stating a multivariate calibration model. Two of the most popular approaches are often referred to as classical least squares (CLS) and inverse least squares (ILS). Underlying CLS is the concept that the analyte signal used for quantitation (the net analyte signal, NAS) has to be orthogonal to the signal of the other constituents in the sample (here regarded as interferents), to avoid analyte prediction bias from the other constituents. Although this orthogonality condition ensures full selectivity for the analyte, it may also decrease the NAS excessively and increase the mean squared error (MSE) of prediction. The magnitude of bias introduced by an interferent depends on the amount of the interferent and the sensitivity of the model for that interferent and hence, under certain circumstances, the orthogonality requisite of CLS can be relaxed to allow a small bias from the interferents if, in return, there is an increase in the NAS and hence a reduction in MSE. This presentation begins with CLS and then develops a relaxed version of CLS (rCLS) by relaxing the CLS selectivity constraints. Comments are provided relating rCLS models to ILS approaches such as partial least squares (PLS), principal component regression (PCR) and ridge regression (RR). Other works have developed modeling theories characterizing interrelationships between CLS and ILS calibrations. Different with rCLS is the reliance on directly relaxing the CLS orthogonality (selectivity) constraints to form a multitude of models depending on the relaxation parameters. From the development of rCLS, selectivity and sensitivity measures are presented that converge to the usual selectivity and sensitivity merits in the orthogonal constrained situation. A selectivity coefficient is defined that directly shows the bias introduced in a prediction. Trends in selectivity, sensitivity, and other NAS merits for CLS, *r*CLS, PLS, and RR are characterized for two near infrared spectral data sets. Observations are provided on using these trends to select respective tuning parameters.

P-018

OUTLIER DETECTION USING BOOTSTRAP ESTIMATES

S. Kaowphong, C. Krongchai and S. Kittiwachana*

Department of Chemistry, Faculty of Science, Chiang Mai University, Chiang Mai, 50200 Thailand
E-mail: silacmu@gmail.com

Outlier detection is an important part of data analysis. Strong outliers are known to be harmful to predictive ability of models and lead to misinterpretation of data. This work proposed a new strategy to detect outliers in analytical data. The decision was based on a majority vote from classifiers trained on bootstrap samples. The sample data were iteratively and randomly split into train and test sets. Classification models were established and statistical indices namely percentage predictive ability (%PA) and percentage model stability (%MS) were calculated [1]. With respect to the classification methods used, the %PA and %MS values of each sample were used to judge whether they were outliers in the dataset; samples having high %MS but low %PA [2]. The developed methodology was tested with various classification methods including one and multi-class classifications. The results from several classification methods were compared and discussed. We illustrated our methodology using both real data examples and simulated data.

Acknowledgement: C.K. would like to thank the Development and Promotion of Science and Technology Talents Project (DPST).

References:

[1] R.G. Breerton, Chemometrics for Pattern Recognition, Wiley, (2009).
[2] S. Kittiwachana, D. L. S. Ferreira, G. R. Lloyd, L. A. Fido, D. R. Thompson, R. E. A. Escott, and R. G. Brereton, J. Chemom., 24, 96–110 (2010).

P-019

Statistics Analysis on the Conductance of Single Molecular Wires with Log bin size

Ya-Hao Wang[1], Xiao-Shun Zhou[2], Jian-Feng Li[1]

[1]MOE Key Laboratory of Spectrochemical Analysis and Instrumentation, College of Chemistry and Chemical Engineering, Xiamen University, Xiamen 361005, China
[2]Zhejiang Key Laboratory for Reactive Chemistry on Solid Surfaces, Institute of Physical Chemistry, Zhejiang Normal University, Jinhua, Zhejiang 321004 China

Abstract:
Scanning tunneling microscope-break junction (STM-BJ) has been developed to one of the most successful methods to determine the conductance of single-molecule junctions with two metal electrodes[1]. Through this method, current-distance spectroscopy is recorded to measure the single-molecular conductance. However, one can not get the reasonable value from single current-distance spectroscopy for the single-molecule conductance for the complicated factor difference in every junctions. Thus, statistical analysis is needed to get a representative conductance value for single-molecule jucntions[2].
In this work, we measurement the single-molecule conductance of amine-terminated benzene containing different non-metallic element (C, O and S) in the molecule backbone at room temperature. Comparing with the conductance histogram obtained from statistical analysis from linear bin size, that from log bin size can give out obviuos results as shown in Figure 1. The results give out $10^{-3.75}$, $10^{-4.0}$, $10^{-4.2}$ G_0 (G_0 =77400 nS) for the single-molecule conductance of 4,4'-oxydianiline (ODA), 4,4'-thiodianiline (TDA) and 4, 4'-methylenedianiline (MDA) (Figure 1), respectively. Those results give out the order of $G_{ODA} > G_{TDA} > G_{MDA}$, which is consistent with the order of electronegativity of O, S and C. The current study shows that conductance histogram from statistical analysis with log bin size can easily detect the single-molecule conductance.

Keywords: Statistics analysis, log-scale histogram, Single-molecule conductance, STM-BJ

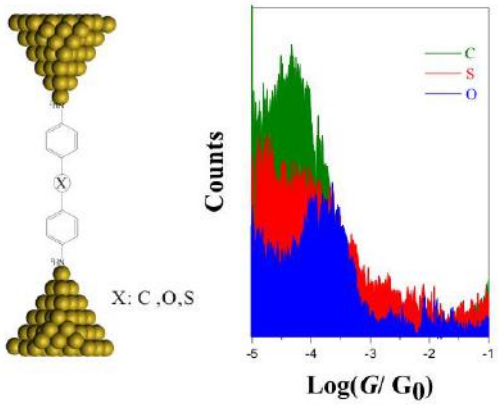


Figure 2: Schematic diagram of the STM-BJ approach for conductance measurement of single molecular junctions and One-dimensional conductance histogram of ODA, TDA and MDA constructed from thousand current-distance spectroscopy.

References:

13. Xu BQ. and Tao NJ.(2003) Measurement of Single-Molecule Resistance by Repeated Formation of Molecular Junctions, *Science*, 301, 1221-1223.
14. Wang YH., Hong ZW., Sun YY., Li DF., Han D., Zheng JF.,Niu ZJ.,Zhou XS.(2014) Tunneling Decay Constant of Alkanedicarboxylic Acids: Different Dependence on the Metal Electrodes between Air and Electrochemistry, *J. Phys. Chem. C*, 118, 18756−18761

P-020

PROBABILISTIC SIMCA – A CLASSIFICATION TOOL THAT INCORPORATES MEASUREMENT UNCERTAINTY INFORMATION

I. Stanimirova

Department of Theoretical Chemistry, Institute of Chemistry, University of Silesia, 9 Szkolna Street,40-006 Katowice, Poland
istanimi@us.edu.pl

The interest in developing chemometric approaches that incorporate *a priori* knowledge about sampling error, instrumentation noise or other possible sources of variation has been driven by the possibility of improving the extraction of information from collected multivariate chemical data. To date, several methods such as maximum likelihood principal component analysis, MLPCA [1], multivariate curve resolution-weighted alternating least-squares, MCR-WALS [2], or positive matrix factorization, PMF [3,4], which include measurement uncertainty information, have been presented in the literature. In this work, we present a probabilistic soft independent modelling of class analogy, SIMCA, approach as a counterpart to the classic SIMCA method. The goal of classic SIMCA is to create boundaries for each group separately by using classic principal component analysis. Thus, a straightforward way to create a probabilistic SIMCA approach is to use one of the following methods: probabilistic principal component analysis [5], maximum likelihood common factor analysis [6] or MLPCA instead of the classic principal component analysis. We discuss and illustrate the properties of these candidate methods in a comparative study using simulated data sets. Furthermore, several other important issues need to be addressed, namely, how to evaluate the complexity of such a probabilistic model, how to make a prediction for the independent test set and how to select representative model and test sets. All of these issues using real and simulated data sets are also discussed in this study.

References:

[1] P.D. Wentzell, D.T. Andrews, D.C. Hamilton, K. Faber, B.R. Kowalski, J. Chemometrics, 11, 339 (1997).
[2] P.D. Wentzell, T.K. Karakach, S. Roy, M.J. Martinez, C.P. Allen, M. Werner-Washburne, BMC Bioinformatcs, 7, 343 (2006).
[3] P. Paatero, U. Tapper, Environmetrics, 5, 111 (1994).
[4] I. Stanimirova, R. Tauler, B. Walczak, Environmental Science & Technology, 45, 10102 (2011).
[5] M.E. Tipping, C.M. Bishop, J. R. Statist. Soc. B, 61, 611 (1999).
[6] D.N. Lawley, A.E. Maxwell, Biometrika, 60, 331 (1973).

P-021

REGULARIZED MANOVA (rMANOVA) IN UNTARGETED METABOLOMICS

J. Engel [1], L. Blanchet [2], B. Walczak [3], L.M.C. Buydens [4]

[1]NERC Biomolecular Analysis Facility – Metabolomics Node (NBAF-B), School of Biosciences, University of Birmingham, Edgebaston, B15 2TT Birmingham, United Kingdom
[2] Department of Pharmacology and Toxicology, School of Nutrition and Translational Research in Metabolism (NUTRIM), Maastricht University Medical Centre, Universiteitssingel 50, 6229 ER Maastricht, the Netherlands
[3] Institute of Chemistry, The University of Silesia, Szkolna Street 9, 40-006 Katowice, Poland
[4] Radboud University Nijmegen, Institute for Molecules and Materials, P.O. Box 9010, 6500 GL Nijmegen, the Netherlands
j.engel@bham.ac.uk

In advanced metabolomics experiments often a large number of response variables is measured, while the levels of one or several factors are varied. Often the number of response variables vastly exceeds the sample size. Consequently, well-established techniques such as MANOVA cannot be used to analyze such data. ANOVA simultaneous component analysis (ASCA), as an alternative to MANOVA, can be used in this case, but at the price of assuming that variables are uncorrelated [1].

Here, regularized MANOVA (rMANOVA) analysis is presented [1]. The rMANOVA model is essentially a weighted average of the ASCA and MANOVA models, and comprises these two as special cases. The optimal weight is determined in a data driven fashion employing a Stein-type shrinkage estimator of the within-group covariance matrix. Compared to ASCA, rMANOVA allows for correlated variables, thus offering a more realistic view of the data. Compared to MANOVA, rMANOVA is also applicable when the number of variables is (much) larger than the sample size.

rMANOVA is compared with ASCA, PCA + MANOVA (MANOVA applied to PCA-scores), and ANOVA-PLS [2] for a number of simulated data structures. The application of rMANOVA to multi-factor untargeted metabolomics data is illustrated by analysis of NMR data and LC-MS data. We conclude that rMANOVA is a highly promising method for analysis of metabolomics data.

References:

[1] J. Engel, L. Blanchet, B. Bloemen, L.P. van den Heuvel, U.H.F. Engelke, R.A. Wevers, L.M.C. Buydens, Analytica chimica acta, 899, 1 (2015).
[2] F. Marini, D. de Beer, E. Joubert, B. Walczak, Journal of Chromatography A, 1405, 94 (2015).

P-022

NON PARAMETRIC STATISTICAL CONTROL: THE CONTENT OF BIOGENIC AMINES IN SPANISH WINES

E. Meléndez[1], L.A. Sarabia [2], M.C. Ortiz[3]

[1]Estación Enológica de Haro, Bretón de los Herreros, 4 26200 Haro, La Rioja, Spain.
[2]Mathematics and Computation, Burgos University, Pza. Misael Bañuelos s/n, 09001 Burgos, Spain.
[3]Department of Chemistry, Burgos University, Pza. Misael Bañuelos s/n, 09001 Burgos, Spain.
e-mail: lsarabia@ubu.es

Biogenic amines are formed by precursor amino acids and various microorganisms present in the wine, at any step of production, ageing or storage. The presence of biogenic amines in wines has been studied extensively since 1980 and particularly over the last 10 years as a consequence of the increasing attention to consumer health protection. Also, biogenic amines have the potential to be applied as indicators of food spoilage and/or authenticity.

In this work, 684 samples of wines from different Spanish Regions have been analysed in order to control the content of histamine, tyramine, phenylethylamine, cadaverine and putrescine during 2010, 2014 and 2015.

The β-content tolerance intervals are used to model the statistical distribution of histamine[1]. Copulas [2] are also used to obtain their multivariate confidence region between histamine and tyramine. They have been built for the first time in the oenological field. The β-content tolerance intervals were obtained with STATGRAPHICS [3]. The 'Statistics and Machine Learning Toolbox' [4] has been used to estimate the margin distribution for each BAs content and the copula for the multivariate interrelation between them. A home-made function in Matlab has been built to obtain the probability of the tolerance region in a level of the joint cumulative density function estimated.

When the statistical control is made for each BA separately and there is lack of normality, then the β-content tolerance interval is more adequate than the usual confidence interval. Using β-content tolerance intervals, for histamine content, in the case of the population of red Spanish wines in 2010, only 53.9 % of the distribution would be below 10 ppm (if we want to affirm it with a 95 % confidence). This percentage of population increases from 53.9 to 74.6 % in the year 2014 and to 90.2 % in 2015. (10 ppm was the limit in Switzerland until 2011).

Multivariate methods are necessary since independence cannot be assumed among the variables under investigation as in BAs content in wines. In this case lack of normality can occur in different ways: i) The marginal distribution of BAs may not be normal; ii) The BAs content together have a very nonlinear relationship in the wines, so the multivariate normality is not valid once more. Multivariate distributions built from copulas have been proved very useful in recent years in many applications. A copula can be used to couple different marginals (BAs content in wine) together and to build Their bivariate distributions. This method separates a bivariate distribution into two components, two marginals and a copula, providing a very flexible framework in multivariate modeling. In this work the conjoint distribution of histamine and tyramine is well modeled by a Clayton copula with margins estimated by Gaussian kernel, being this distribution similar for the three years analyzed.

Acknowledgement: The authors thank the financial support provided by the MINECO (CTQ2014-53157-R) and FEDER funds.

References:

[1] M.C. Ortiz, L.A. Sarabia, M.S. Sánchez, A. Herrero, Quality of Analytical Measurements: Statistical Methods for Internal Validation in Comprehensive Chemometrics, Vol 1. Eds Brown S.D., Tauler R. Walczak B., Elsevier, Amsterdam (2009).
[2] R, B. Nelsen, An Introduction to Copulas, Springer, Berlin Heidelberg New York (1999).
[3] STATGRAPHICS Centurion XVI. StatPoint Technologies, Inc. 2015.
[4] Statistics and Machine Learning Toolbox for Matlab R 2015b. The MathWorks, Inc.

P-023

Using chemometric methods for resolution of multi-frequency eddy current data for reliable diagnostic of conductive materials

A.V. Egorov[1], S.V. Kucheryavskiy[1], V.V. Polyakov[1,3]

[1] Department of Altai State University, Lenina str 66, Barnalul, Russia
[2]Department of Chemistry and Bioscience, Aalborg University, Niels Bohrs vej 8, Esbjerg, Denmark
[3]Institute of Strength Physics and Materials Science SB RAS, 2/4, Akademicheskii ave., Tomsk, Russia
svk@bio.aau.dk

Eddy current testing (ECT) is one of the well-known non-destructive methods for identification of conductive materials as well as for their diagnostics, including detection of flaws (both on and under surface), cracks, corrosion, evaluation of electrical conductivity, thickness and many other properties. The method can be used both with magnetic and non-magnetic objects.

The main part of any ECT device is a sensor, which usually consists of two inductance coils with a magnetic core (or just wire coils in a simple form). An alternating current (AC) source with preset frequency and amplitude is used to activate one of the coils (usually called as transmitting coil), which creates a changing magnetic field around the sensor. If the sensor is located close to a conductive sample, the magnetic field is induced to the sample and creates eddy currents, which, in their turn, produce a secondary magnetic field opposed to the primary field produced by the transmitting coil. The second (receiving) coil collects the superposition of the two fields and generates a signal for analysis

The properties of the secondary field (and, therefore, of the superposition of the two fields) depend on many parameters including type of material (mainly its electrical conductivity), its thickness, distance between the sample and the sensor, as well as the presence of any disturbances on or under the sample's surface, such as cracks, scratches, coatings, and other flaws. This actually leads to the one of the biggest disadvantage of the method — its sensitivity depends on many interfering factors and often it is very difficult to resolve them if more than just one are unknown.

One of the ways to tackle this problem is to use multi-frequency measurements, when the parameters of the magnetic fields are measured for a range of AC frequencies used to activate the induction coil. The results of such measurements are often represented in a graphical form by so called scanning hodographs — diagrams showing how resistance and reactance of the receiving coil are changing depending on the activation frequency. The shape of the hodographs reflects influence of the main factors and thorough investigation of the shape as well as comparing the shapes with measurements made for standard objects can be quite useful. At the same time such approach is rather subjective and does not allow to carry out automatic measurements. It was also found out that for many real it does not allow to resolve several factors.

In the present study we propose a multivariate approach for solving the problem of interfering factors in eddy current testing. The general idea is to represent changes in resistance and reactance of the measuring system caused by different activating frequencies for a particular sample in a form of a spectrum and use multivariate techniques for finding hidden patterns in the changes, which may mainly reflect an influence of a particular factor.

In order to test the feasibility of the proposed approach we investigated a possibility to resolve at least two competing factors, namely a conductivity of a sample and a margin between the sample and a sensor. Besides that, an analysis of how the range of used frequencies influences the resolution results has been carried out.

P-024

ENHANCING MODEL INTERPRETATION BY THE SELECTION OF APPROPRIATE PREPROCESSING

Jan Gerretzen[1,2], Ewa Szymańska[1,2], Jacob Bart[3], Tony Davies[3], Henk-Jan van Manen[3], Edwin van den Heuvel[4], Jeroen Jansen[1] and Lutgarde Buydens[1]

[1]Radboud University, Institute for Molecules and Materials, Heyendaalseweg 135, 6525 AJ Nijmegen, The Netherlands
[2]TI-COAST, Science Park 904, 1098 XH Amsterdam, The Netherlands
[3]AkzoNobel, Supply Chain, Research & Development, Zutphenseweg 10, 7418 AJ Deventer, The Netherlands
[4]Eindhoven University of Technology, Groene Loper 5, 5600 RM Eindhoven, The Netherlands

j.gerretzen@science.ru.nl

The aim of data preprocessing is to remove data artifacts—such as a baseline, scatter effects or noise—from the data and to enhance the analytically relevant information. Many preprocessing methods exist to do either or both; it is not at all clear on beforehand which preprocessing methods should optimally be used [1]. Recently, we have developed a novel, simple approach based on Design of Experiments (DoE) and Partial Least Squares (PLS), which enables to select an optimal preprocessing within reasonable time [2].

In that approach, the focus was solely on increasing model performance (Root Mean Square Error of Prediction, RMSEP). In many chemometrics applications, however, the interpretation of the model results may be just as important, if not even more relevant. A common way of interpreting chemometric models is by variable selection. Here, we extend our original approach by replacing PLS with PPRV-FCAM (Predictive Property Ranked Variable reduction using Final Complexity Adapted Models, developed by Andries et al. [3]), which iteratively integrates PLS with variable selection.

Preprocessing and variable selection are strongly related, since proper removal of data artifacts should also improve the selection of relevant variables. By analyzing several experimental data sets of which the true relevant variables are known, we show that the simultaneous optimization of preprocessing and variable selection considerably improves model performance. Moreover, a selection of variables is provided that complies more with the true relevant variables compared to individual optimization of both model aspects.

This work is part of the 'Analysis of Large data sets By Enhanced Robust Techniques' project (ALBERT), which aims to develop generic strategies and methods to facilitate better and more robust chemometric and statistical analyses of complex analytical data.

Acknowledgement: This research received funding from the Netherlands Organization for Scientific Research (NWO) in the framework of Technology Area COAST.

References:

[1] J. Engel, et al. Trends in Anal. Chem., 50, 96 (2013)
[2] J. Gerretzen, et al., Anal. Chem., 87, 12096 (2015)
[3] J. Andries, et al., Anal. Chim. Acta, 705, 292 (2013)

P-025

WEIGHT RANDOMIZATION TEST VS. CROSS-VALIDATION TO SELECT THE OPTIMAL NUMBER OF COMPONENTS IN PLS: TWO APPLICATIONS

Jan Gerretzen[1,2,*], Ewa Szymańska[1,2,*], Lutgarde Buydens[1], Nelson Lee Afanador[1,3], Lionel Blanchet[4,5,6] and Thanh Tran[1,4,7]

[1]Radboud University, Institute for Molecules and Materials, Heyendaalseweg 135, 6525 AJ Nijmegen, The Netherlands
[2]TI-COAST, Science Park 904, 1098 XH Amsterdam, The Netherlands
[3]Center for Mathematical Sciences, Merck, Sharp & Dohme, West Point, PA, USA
[4]Department of Pharmacology and Toxicology, School of Nutrition, Toxicology and Translational Research in Metabolism (NUTRIM), Maastricht University Medical Center, Maastricht, The Netherlands
[5]Thayer school of Engineering, Dartmouth College, Hanover, New Hampshire, USA
[6]Top Institute Food and Nutrition (TIFN), Wageningen, The Netherlands
[7]Center for Mathematical Sciences, Merck, Sharp & Dohme, Oss, The Netherlands

[*]: equal contribution

j.gerretzen@science.ru.nl

Selecting the optimal number of components remains a difficult and essential task in PLS model optimization. A wrong number of components may result in overfitted models with suboptimal prediction power and lead to incorrect model interpretation. Several methods including cross-validation, leverage correction and randomization tests are currently in use for this selection. Here, we present and evaluate a new randomization test: Weight Randomization Test (WRT), in light of the recently introduced underlying theory of the PLS algorithm [1]. WRT does not use the incorrect assumption that "the latent variables enter the model in a natural order" [2]. This method is less computationally demanding than standard randomization tests [2] and can thus be applied to large datasets and chemometric procedures including optimization of external (hyper-)parameters.

In this study, we focus on a practical comparison of WRT with cross-validation. We consider two different application areas of PLS models: classification and regression. In the classification example, several two-class datasets with different levels of known differences are used. Double cross-validated models with permutation tests [3] are compared with single cross-validated models with WRT. Computational time, model performance and number of components are evaluated and guidelines for practitioners are provided.

In the regression example, the influence of model complexity on the selection of the correct hyperparameters (i.e. data preprocessing [4]) is assessed, using an experimental spectroscopy data set. Computational time, model performance, number of components and variable selection of single cross-validated models are compared with WRT-based models. These examples show that WRT leads to less complex and more stable and interpretable PLS models in a fast way.

References:

[1] Tran T. et al., Chemometr. Intell. Lab., 149, 127 (2015)
[2] Wiklund et al., J. Chemometrics, 21, 427 (2007)
[3] Szymańska et al., Metabolomics, 8, S3 (2012)
[4] Engel et al., TrAC, 50, 96 (2013)

P-026

WEIGHT RANDOMIZATION TEST FOR THE SELECTION OF THE NUMBER OF LATENT VARIABLES IN PLS MODELS

Thanh Tran[1,2,*], Ewa Szymańska[2,3], Jan Gerretzen[2,3], Lutgarde Buydens[2], Nelson Lee Afanador[2,4], and Lionel Blanchet [5,6,7]

[1]Center for Mathematical Sciences, Merck, Sharp & Dohme, Oss, The Netherlands
[2]Radboud University, Institute for Molecules and Materials, Heyendaalseweg 135, 6525 AJ Nijmegen, The Netherlands
[3]TI-COAST, Science Park 904, 1098 XH Amsterdam, The Netherlands
[4]Center for Mathematical Sciences, Merck, Sharp & Dohme, West Point, PA, USA
[5] Department of Pharmacology and Toxicology, School of Nutrition, Toxicologyand Translational Research in Metabolism (NUTRIM), Maastricht University Medical Center, Maastricht, The Netherlands.
[6] Thayer school of Engineering, Dartmouth College, Hanover, New Hampshire, United States of America
[7] Top Institute Food and Nutrition (TIFN), Wageningen, The Netherlands.

thanh.tran@merck.com, or thanh.tran@science.ru.nl

The selection of the optimal number of latent variables remains a difficult, but essential, task in Partial Least Squares (PLS). Randomization tests have the advantage of being automatic and using the entire dataset, in contrary with the widely used cross-validation approaches. PLS modeling may include component(s) with large amount of irrelevant data variation. The reformulation of PLS model may or may not be affected by these irrelevant components depending on the assigned *y*-loading. This has been recently indicated by us in the basic sequence framework on the underlying theory of the PLS algorithm and presented to the chemometrics society [1]. These irrelevant latent variables, however, immediately impact the null-distribution in a randomization test for the selection of the number of latent variables. In this case, PLS models with these latent variables may result in false positive tests due to the incorrect assumption that "*the latent variables enter the model in a natural order*" [2]. In this work, we introduce a new randomization test, Weight Randomization Test (WRT), in light of the underlying theory of the PLS algorithm. In the proposed method the null-distribution is better understood and defined with a more efficient implementation. We illustrate its effectiveness in optimization of preprocessing as model hyperparameter, as well as in model selection, where results are compared with the double cross-validation procedure.

Keywords: Partial Least Squares, Number of components, Randomization test

References:

[1] T.N. Tran, L. Blanchet, N.L. Afanador, L.M.C. Buydens, Novel unified framework for latent modeling and its interpretation, Chemometrics and Intelligent Laboratory Systems, 149 (2015) 127-139.
[2] S. Wiklund, D. Nilsson, L. Eriksson, M. Sjostrom, S. Wold, K. Faber, A randomization test for PLS component selection, Journal of Chemometrics, 21 (2007) 427-439.

P-027

A NOVEL MULTIVARIATE METHOD FOR SELECTION OF IMMUNOLOGICAL CELLS WITH MULTICOLOUR FLOW CYTOMETRY

R. Folcarelli [1], L. Koenderman [2] , L. M. C. Buydens [1], J. Jansen [1]

[1] Radboud University, Institute For Molecules and Materials, Heijendaalseweg 135 6525 AJ, Nijmegen, The Netherlands
[2] Respiratory Medicine, University Medical Center Utrecht, Heidelberglaan 100, Utrecht, The Netherlands.

Flow Cytometry (FC)-based gating allows the selection of single cells based on their expression of surface markers. Currently gating mostly is done bivariately, even though many more markers may be measured on the same cell using current FC technology. Only multivariate approaches may extract all aspects of cell variability from the data [1], including those associated with co-expression of multiple surface markers. A quantitative multivariate approach dedicated to gating would lead to objective analysis of cells that are activated during the immune response. In this context we have developed a method called Elimination of Cells Laying in Patterns Similar to Endogeneity (ECLIPSE) that provides a multivariate filter for cells of which the surface markers expression is also observed in healthy individuals. With those cells that pass the filter we build a Response model that focuses only on variability present in activated cells. The objectivity of all models is warranted by validation procedures in widespread use, specifically adapted to the multiset structure of MFC data.
The resulting ECLIPSE method provides a focused view on the variability in surface marker expression observed in cells that were quantitatively determined to be response-related.

References:
[1] J.J. Jansen, B. Hilvering, A. van den Doel, P. Pickkers, L. Koenderman**,** L. M.C. Buydens**,** O.F. van den Brink, Chemometrics and Intelligent Laboratory Systems, 151, 126 (2016).

P-028

Estimating the Number of PCs and Detecting Outliers based on the Stability Information of PCA Model Using Angle Distribution of Loading Subspaces (ADLS)

Y.J. Liu[1,2], T. Tran[1,3], J. Jansen[1], G. Postma[1], H.L. Wu[2], L.M.C. Buydens[1]

[1]Radboud University, Institute for Molecules and Materials (IMM), Heyendaalseweg 135, 6525 AJ Nijmegen, The Netherlands
[2]State Key Laboratory of Chemo/Biosensing and Chemometrics, College of Chemistry and Chemical Engineering, Hunan University, 410082 Changsha, China
[3]Center for Mathematical Sciences, Merck, Sharp, & Dohme, Oss, The Netherlands
Email: yajuan@science.ru.nl

Principal Component Analysis (PCA) is widely used in analytical chemistry, to reduce the dimensionality of a multivariate data set by a few Principal Components (PCs) that summarize the predominant patterns in the data. An accurate estimate of the number of PCs is indispensable, in order to provide meaningful interpretations and extract useful information. Existing estimates for the number of PCs fall short with complex datasets with serious collinearity, noise or outliers. We present here how Angle Distribution of the Loading Subspaces (ADLS) can be used to estimate the number of PCs based on the variability of loading subspace across bootstrap resamples. Based on comprehensive comparisons with other well known methods for the determination of the number of PCs applied on datasets with complementary characteristics, we show that ADLS (1) may quantify the stability for a PCA model with several numbers of PCs; (2) better estimates the appropriate number of PCs when compared with the cross-validation and scree plot methods, specifically for coherent data, and (3) can be used to detect outliers by comparing the sample frequency across bootstrap set. This is the first time that a method integrates these tasks. We further demonstrate how the analysis of different types of real-life spectroscopic datasets may benefit from these advantages of ADLS.

References:

[1] P. Besse. Statistics & Probability Letters. 13, 405(1992)
[2] R. Bro, K. Kjeldahl, A. K.Smilde, H. A. L. Kiers, Analytical and bioanalytical chemistry, 390, 1241(2008)
[3] Y.J. Liu, J. Jansen, G. Postma, T. Tran, H.L. Wu, L.M.C.Buyden, Chemometrics and Intelligent Laboratory Systems, In press

P-029

Towards better comparing variable selection methods in near infrared spectroscopy

Yong-Huan Yun[1], Bai-Chuan Deng[1], Yi-Zeng Liang*

[1]College of Chemistry and Chemical Engineering, Central South University, Changsha, P.R. China.
email (yunyonghuan@foxmail.com)

Near-infrared (NIR) spectroscopy has increasingly been adopted as an very useful analytical tool in various fields, such as the agricultural, food, pharmaceutical, environmental, clinical and herbal in the past 20 years. Variable selection techniques have played a key role in the analysis of near infrared spectroscopy. Liang and Yun confirmed the importance and necessity of variable selection in near infrared spectroscopy [1]. The purpose and significance of variable selection can be summarized in three aspects: (1) improving the prediction performance of the predictors, (2) providing faster and more cost-effective predictors by reducing the curse of dimensionality, (3) providing a better understanding and interpretation of the underlying process that generated the data [2]. However, in the face of the large number of spectral variables, it is a NP hard optimization problem to find the optimal variable subset that satisfies the above three aspects. In the past two decades, a large number of variable selection methods have been developed to deal with this problem. They can be divided into two categories as wavelength inteval selection and wavelength point selection method. In this study, we will review these two kinds of methods and make a deep comparison on the selected variables and its prediction performance and reliability through nine NIR datasets which have different ratioes of the number of variables to samples. Fig.1. shows the size of nine NIR datasets and the variable selection methods used to compare in this study.

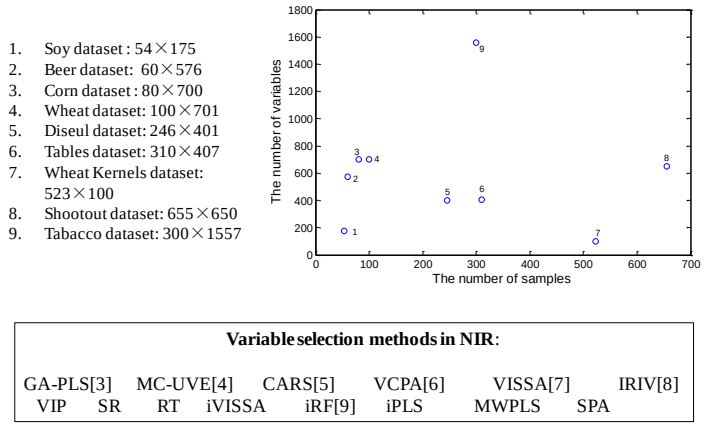


Fig. 1. Size of nine NIR datasets and the variable selection methods used to compare in this study.

Acknowledgement: This work is financially supported by the National Nature Foundation Committee of P.R. China (grants no. 21275164 and 21175157).

References:

[1] Y.H. Yun, Y.-Z. Liang, G.-X. Xie, H.-D. Li, D.-S. Cao, Q.-S. Xu, , Analyst, 138 (2013) 6412.
[2] I. Guyon, A. Elisseeff , J. Mach. Learn. Res., 3 (2003) 1157.
[3] R. Leardi, J. Chemometrics, 14 (2000) 643.
[4] W. Cai, Y. Li, X. Shao, Chemometr. Intell. Lab, 90 (2008) 188.
[5] H.-D. Li, Y.-Z. Liang, Q.-S. Xu, D.-S. Cao, Anal. Chim. Acta, 648 (2009) 77.
[6] Y.H. Yun, W.T. Wang, B.C. Deng , Y.Z. Liang, W. Fan, Q.S. Xu, Anal. Chim. Acta., 862 (2015) 14~23.
[7] B.C. Deng, Y.H. Yun, Y.Z. Liang, L.Z. Yi, Analyst, 139 (2014) 4836~4845.
[8] Y.H. Yun, W.T. Wang, Y.Z. Liang, H.D. Li, D.S. Cao, Q.S. Xu, Anal. Chim. Acta, 807 (2014) 36~43.
[9] Y.H. Yun, H.-D. Li, L.R. E. Wood, Q.-S. Xu, Y.-Z. Liang, Spectrochim.Acta.A., 111 (2013) 31.

P-030

MISSING DATA IMPUTATION (MDI) TOOLBOX FOR MATLAB

A. Folch-Fortuny[1], F. Arteaga[2], A. Ferrer[1]

[1]Departamento de Estadística e IO Aplicadas y Calidad, Universitat Politècnica de València, Valencia, Spain.
[2]Department of Biostatistics and Investigation, Universidad Católica de Valencia San Vicente Mártir, Valencia, Spain
abfolfor@upv.es

Missing Data Imputation (MDI) Toolbox [1] is presented here to impute missing values in incomplete data sets following missing completely at random (MCAR) patterns [2]. MDI Toolbox includes principal component analysis (PCA) model building methods with missing data able to reconstruct the missing values coherently with the latent structure of the available measurements.
Several methods from the literature are included in this toolbox: trimmed scores regression (TSR), known data regression (KDR), KDR with principal component regression (KDR-PCR), KDR with partial least squares (KDR-PLS), projection to the model plane (PMP) [3], iterative algorithm (IA) [4], modified nonlinear iterative partial least squares regression algorithm (NIPALS) [5] and data augmentation (DA) [6]. TSR is presented as the default method for its good performance with all data structures [3].
A graphical user-friendly interface is provided with the toolbox to ease its use. In this way, several windows guide the user step by step: from the data acquisition and settings to the results exploitation via interactive loadings and scores plots.
MDI Toolbox presents a general procedure to impute missing data, thus it can be used to infer PCA models with missing data, to estimate the covariance structure of incomplete data matrices, or to impute the missing values as a preprocessing step of other methodologies.
The MDI toolbox is freely available for academic purposes at http://mseg.webs.upv.es.

Acknowledgement: Research in this study was partially supported by the Spanish Ministry of Science and Innovation and FEDER funds from the European Union through grant DPI2011-28112-C04-02 and DPI2014-55276-C5-1R, and the Spanish Ministry of Economy and Competitiveness through grant ECO2013-43353-R.

References:
[1] A. Folch-Fortuny, F. Arteaga, A. Ferrer, submitted.
[2] F. Arteaga, A. Ferrer, in: Comprehensive Chemometrics, 3, 285–314, Elsevier (2009).
[3] A. Folch-Fortuny, F. Arteaga, A. Ferrer, Chemometr. Intell. Lab., 146, 77-88 (2015).
[4] B. Walczak, D.L. Massart, Chemometr. Intell. Lab., 58, 15-27 (2001).
[5] S. Wold, C. Albano, W.J. Dunn, K. Esbensen, S. Hellberg, E. Johansson, Food Research and Data Analysis, 183-185, Applied Science Publishers (1983).
[6] M.A. Tanner, W.H. Wong, J. Am. Stat. Assoc., 87, 528-550 (1987).

P-031

MORE RELIABLE MODELS BY CORRECTING FOR INDIVIDUAL VARIATION

M. Koeman[1], J. Jansen[1] and L. Buydens[1]
[1]Radboud University, Institute for Molecules and Materials (IMM)
Heyendaalseweg 135, 6525 AJ Nijmegen, The Netherlands
Email: m.koeman@science.ru.nl

In chemometrics, the goal of building a model is often not using the model itself but rather the interpretation of said model. Moreover, variations among individuals is a phenomenon that can greatly hinder the analysis of an experiment. These individual variations are caused by for example the subject's age, sex or diet and cannot be completely known or controlled by the researcher[1]. Individual variation can be especially hindering when it comes to interpretation, as it can lead to identification of the wrong variable(s), for this reason analyzing the data directly is often not feasible. Coping with these variations is therefore essential for a correct model interpretation.
In this work we focus on the situation where we have a relevant group of controls and a group under investigation. The goal is to find a reliable model for the differences between the two groups. This model can then be interpreted to find the mechanism responsible for whatever difference there is. The control group can be used to get an estimate of the aforementioned individual variation. We investigate a method where we first model the control group and project the data of the experimental group on this model. The residuals of this projection are expected to be free of the unwanted healthy individual variations and to contain only the effect under consideration. However, this approach is known to suffer from the so-called smearing effect[2], an effect that can cause the identification of the wrong variables. We show with a simple simulation that this effect indeed occurs because the residuals are orthogonal to the control model. This leads to a biased model.
To alleviate the smearing effect we propose to create a compromise between the residuals and the original data of the group under investigation. We show that this compromise performs better than the two extremes it is created from on simulated data as well as (quasi) real world data. This method is promising for the analysis of data with hindering individual variations and can improve the interpretation of residuals based methods in general.

References

[1] Leek, Jeffrey T., and John D. Storey. "Capturing heterogeneity in gene expression studies by surrogate variable analysis." *PLoS Genet* 3.9 (2007): e161.
[2] Van den Kerkhof, Pieter, et al. "Analysis of smearing-out in contribution plot based fault isolation for Statistical Process Control." Chemical Engineering Science 104 (2013): 285-293.

P-032

ENSEMBLING REGRESSION AND CLASSIFICATION FOR PREDICTING PROPORTIONS OF COMPLEX BLENDS IN FOOD PRODUCTS

Beatriz Carrasco[1], Elena Montañés[2], José Ramón Quevedo[2]

[1]Blendhub Corp., 30169 San Ginés, Murcia, Spain.
[2]Artificial Intelligence Center, Universidad de Oviedo, Gijón, Spain
bcarrasco@blendhub.com

Certain ingredients present in complex blends of food products play an important role in the processing industry for adding texture, enhancing tasting, etc… However, controlling their proportions is essential both to asses the legal requirements for the safety of consumers and to avoid fraud in food labelling. The samples available were processed using NIR technology to obtain their spectra and were labelled with the proportion of ingredients with which they were built. This work proposes a combination of two Machine Learning (ML) techniques for predicting such proportions of ingredients in food products from NIR spectra. These techniques have been shown good performance on NIR spectra before in the literature [1, 2].

The goal of this work is twofold: On one hand, the aim is to predict the proportion of ingredients as much as closer to the real one. On the other hand, the aim is to avoid predicting a high number of low proportions of ingredients that in fact are not present in the real sample. Regression is the technique taken to reach the first goal. Simple regression models rather than other more complex models will be chosen in order to avoid overfitting in high dimensional and noisy problems, as those that involves NIR spectroscopy. In this sense, and taking into account the second goal, linear models are more suitable in this case. This is so, since in addition to being simple, they are able to predict zero or negative values avoiding the so-called long tail; in contrast, for instance, to other simple models, as sigmoidal models, which could always predict positive values. In certain sense, linear models in regression techniques can handle the second goal. However, they are not focused on detect the presence or absence of ingredients in a sample. Hence, classification techniques are more adequate to this purpose. Therefore, combining both techniques leads to a more promising method that provides good proportion predictions that taking just one technique on their own. The combination takes places as follows: i) Only if the classification model predicts the presence of the ingredient and the regression model offers a positive value and the classification model predicts the presence of the ingredient, then, that positive value is taken for this ingredient. ii) A null proportion is taken otherwise. The well-known Support Vector Regression (SVR) for regression and Support Vector Machine (SVM) for classification were taken, since both techniques have shown promising performance with high dimensional and noisy data [3]. Experiments carried out show an error of 0.32% in detecting the presence of ingredients and of 7.22% in predicting the proportions of ingredients whose presence is predicted.

References:

[1] V. Fernández-Ibáñez, T. Fearn, A. Soldado, B. de la Roza-Delgado, Development and validation of near infrared microscopy spectral libraries of ingredients in animal feed as a first step to adopting traceability and authenticity as guarantors of food safety, Food Chemistry, Volume 121, Issue 3, 1, Pages 871-877, (2010).
[2] Fernández-Ibáñez V, Fearn T, Montañés E, Quevedo JR, Soldado A, de la Roza-Delgado B. Improving the discriminatory power of a near-infrared microscopy spectral library with a support vector machine classifier.Applied Spectroscopy. 2010 Jan;64(1):66-72, (2010).
[3] R. E. Fan, K. W. Chang, C. J. Hsieh, X. R. Wang, and C.-J. Lin. LIBLINEAR: A library for large linear classification *Journal of Machine Learning Research* 9, 1871-1874, (2008).

Acknowledgments
This research has been supported by Spanish Ministerio de Economía y Competitividad (grant TIN2015-65069-C2-2-R).

P-033

Sampling calibration sets: optimal size, samples and methods

Antoine Stevens[1], L. Ramirez-Lopez[2]

[1]Georges Lemaître Centre for Earth and Climate Research, Université Catholique de Louvain, 3, place Louis Pasteur B-1348, Louvain-la-Neuve, Belgium.
[2]BUCHI Labortechnik AG, Meierseggstrasse 40, Flawil, Switzerland

ramirez.lopez.leo@gmail.com

We investigated the effect of both the calibration set size (number of samples) and the calibration sampling method on the performance of visible- and near-infrared (vis–NIR) models to predict clay content. We evaluated the following calibration sampling algorithms: Kenard–Stone (KSS), fuzzy *c*-means (FCMS) and the conditioned Latin hypercube (cLHS). These algorithms were tested on a continental vis–NIR library of Europan soils which comprises about 19.000 samples. A validation set of 2.000 samples was randomly selected from the vis–NIR library and the remaining samples were used as potential candidates for calibration. From this set of candidates, we sampled substes of different sizes starting with 30 samples up to 2.000 samples in steps of 10. This process was done for each sampling algorithm. Each calibration subset was used to build a vis–NIR model to predict clay content in the validation set. These models were calibrated using the support vector regression machine (SVM) algorithm. The root mean square error (RMSE) of prediction was used to evaluate the sensitivity of the models to both the sampling algorithm and the calibration set size.

In this presentation, we provide an overview of the different calibration sampling algorithms in terms of predictive performance and representativeness of the selected samples. In addition, we propose a simple method to optimize the calibration set size based only the vis–NIR data (i.e. without prior knowledge of the response variables). The method presented here is an extension of the ones recently introduced in [1]. It is based on the comparisons of the probability density functions of the vis–NIR data of the calibration subset against their equivalents in the population. Finally, we present some software tools [2] that we have developed to aid in the selection of optimal calibration sets.

[1] Ramirez-Lopez, L., Schmidt, K., Behrens, T., van Wesemael, B., Dematte, J. A., Scholten, T.. Sampling optimal calibration sets in soil infrared spectroscopy. *Geoderma*, *226*, 140-150. (2014).

[2]Stevens, A., Ramirez-Lopez, L. An introduction to the prospectr package. R package Vignette R package version 0.1.3. (2013).

P-034

MULTIVARIATE ANALYSIS BASED ON HS-SPME/GC-MS FINGERPRINT AND VOLATILE COMPOSITION FOR THE CHARACTERIZATION OF EXHALED HUMAN BREATH

M. Pôssas Abreu[1], E. Descours [2], L. Rousseau[1], F. Ghassemi[1], N. Vallet[2], G. Lissorgues[1]

[1]ESYCOM ESIEE Paris, Noisy le Grand, France.
[2]ISIPCA, 34-36 Rue du Parc de Clagny, Versailles, France.
maira.possasabreu@esiee.fr

This work aims to exhaustively identify the volatile organic compounds (VOCs) that can appear in exhaled human breath and determine their relative composition using Gas Chromatography/Mass Spectrometry (GC-MS) in combination with Solid-Phase Microextraction (SPME)[1]. Later, this study will contribute to choose the appropriate coating to fonctionnalise an array of sensors for electronic nose[2].

Tedlar gas bags were chosen to collection of human breath. A dozen of healthy subjects (smoker and non-smokers of all ages and gender) who had not ingested coffee or alcohol for at least 24h have been asked to inhale moderately and then to exhale as much as possible. After collection, SPME fiber have been exposed for one hour in the collected breath and ambient air was collected at same time[3].

A number of 2000 volatile organic compounds was detected on breath samples like hydrocarbons alcohols, aldehydes, ketones, amides, furanes, aromatic compounds, ester, thiols, etc (Figure 1). Multivariable PCA over total ion chromatogram (TIC) and over mass ion spectrum have been done. The statistical method using mass spectrum seems to be the most discriminant and allowed to distinguish two groups : smoker and non smoker subjects (Figure 2).

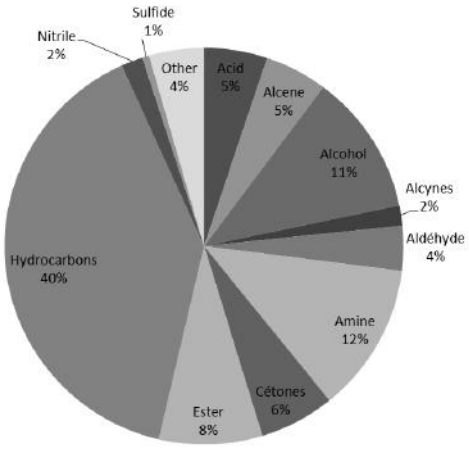


Figure 1 : Volatile compounds family repartition

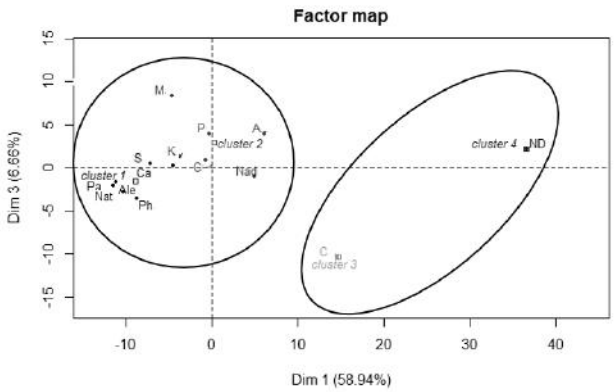


Figure 2 : Principal component analysis (PCA) results using mass ion spectrum dataset

References:

[1] Ouyang, G., & Pawliszyn, J. (2006). SPME in environmental analysis. Analytical and bioanalytical chemistry, 386(4), 1059-1073.

[2] Possas, M., Rousseau, L., Ghassemi, F., Lissorgues, G., Scorsone, E., Manai, R., & Bergonzo, P. (2015, April). Frequency profile measurement system for microcantilever-array based gas sensor. In Design, Test, Integration and Packaging of MEMS/MOEMS (DTIP), 2015 Symposium on (pp. 1-5). IEEE.

[3] Deng, C., et al. (2004). Determination of acetone in human breath by gas chromatography–mass spectrometry and solid-phase microextraction with on-fiber derivatization. Journal of Chromatography B, 810(2), 269-275.

P-035

TRANSFER OF CALIBRATION MODELS BETWEEN DIFFERENT ANALYTICAL METHODS IN LATENT VARIABLE SPACE

V.V. Panchuk [1], D.O. Kirsanov [1,2], E. S. Oleneva[1], M.M. Khaydukova[1,2], A.V. Legin[1,2]

[1]Institute of Chemistry, St.Petersburg State University, Universitetskaya nab. 7/9, 199034, St.Petersburg, Russia
[2]Laboratory of artificial sensory systems, ITMO University, Kronverkskiy pr. 49, 197101, St.Petersburg, Russia
d.kirsanov@gmail.com

A problem of calibration transfer is well-known in analytical chemistry. Most of the works addressing this problem are coming from spectroscopy domain. The problem appears every time when one wants to use calibration model developed for one analytical instrument (e.g. near infrared (NIR) spectrometer) with the data obtained by another instrument of the same type (another NIR spectrometer). This is especially relevant to multivariate calibration models since they typically require large number of samples to establish reliable calibration. This large number may be hard/long to measure again on another instrument. Due to the fact that two spectrometers are not identical copies and they may differ in spectral shape, sensitivity, operating conditions, etc. direct application of a calibration model to the data from another spectrometer is normally not possible, since it leads to unacceptable growth of analytical errors in quantification of target sample parameters. Numerous methods were developed to perform calibration transfer and most of them are reviewed in the work [1]. The calibration transfer procedure can be based on correction of regression model parameters (slope and bias) or it can employ correction (conversion) of the second instrument response. The latter case can be considered in a broader context as a calibration transfer between two absolutely different analytical methods.

In this work we suggest the procedure for transfer of calibration models between different analytical methods. It is based on the conversion of analytical signals from one instrument into the format of another instrument. This conversion is performed in latent variable space as follows:

$X_1 = T_1 P_1^t$ (1), $P_2 = T_{1tr}^{+} X_{2tr}$ (2), $T_2 = X_{2data} P_2^{+}$ (3), $X_{2cor} = T_2 P_1^t$ (4).

First we decompose the data X_1 from the first instrument using SVD (singular value decomposition) procedure, then from the score matrix T_1 we choose only the lines corresponding to the samples available in X_{2tr} which yields T_{1tr}. Using T_{1tr} matrix and X_{2tr} we compute P_2 – the loadings for the second instrument, where T_{1tr}^{+} refers to matrix pseudo-inverse. Then we calculate corresponding score matrix for the data from the second instrument (X_{2data}) which we plan to use for prediction with calibration model from the first instrument and finally we compute corrected response from the second instrument (X_{2cor}) which can be further used with the model from the first instrument.

The performance of the suggested method was tested with simulated data sets and with real data. Three real datasets were addressed: calibration models constructed for quantification of Co, Ni and Cu in their triple mixtures with energy-dispersive X-ray fluorescence measurements were successfully employed for prediction of these three metals from UV-Vis spectrometry data; calibration models for sugar quantification in marzipan samples by two NIR instruments operating in different wavelength ranges; calibration models for API quantification in tabletsby two similar NIR instruments. It appears that the method can have a broad range of possible applications in analytical chemistry.

References:

[1] S. D. Brown, Transfer of Multivariate Calibration Models in Comprehensive Chemometrics, Elsevier, vol.3, p.345 (2009).

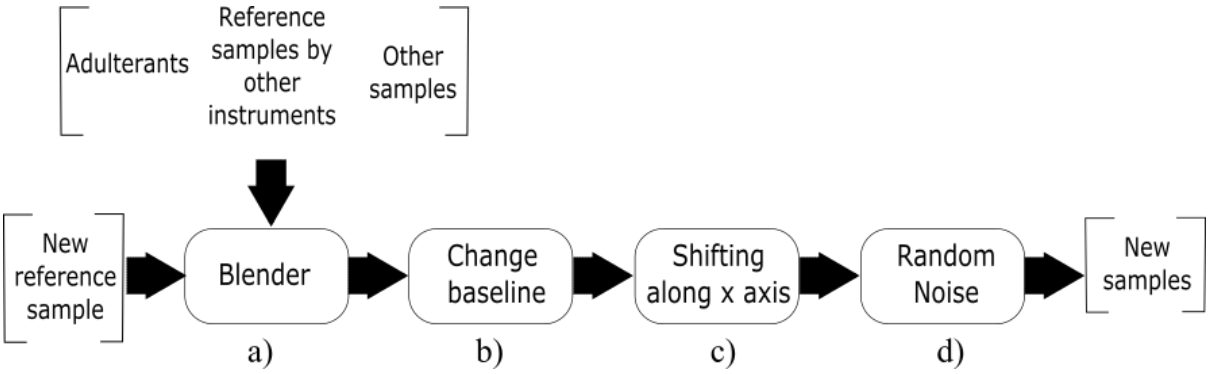
P-036

GENERATING DATA AUGMENTED SPECTROSCOPIC DATA FOR PERFORMANCE ENHANCEMENT

Konstantia Georgouli [1], Jesus Martinez Del Rincon [2], Anastasios Koidis[1]

[1] Institute for Global Food Security, Queens University Belfast, 18-30 Malone Road, Belfast, Northern Ireland, UK
[2] Institute of Electronics, Communications and Information Technology, Queens University Belfast, Northern Ireland Science Park Queen's Road, Belfast, Northern Ireland, UK
kgeorgouli01@qub.ac.uk

The application of chemometrics in food science has revolutionized the field by allowing the creation of models able to automate a broad range of applications such as food authenticity and food fraud detection. In order to create effective and general models able to address the complexity of real life problems, a vast amount of varied training samples are required. Training dataset has to cover all possible types of sample and instrument variability [1]. However, acquiring a varied amount of samples is a time consuming and costly process, in which collecting samples representative of the real world variation is not always possible, specially in some application fields. To address this problem, a novel framework for the application of data augmentation [2] techniques to spectroscopic data has been designed and implemented. This is a carefully designed pipeline of four complementary and independent blocks which can be finely tuned depending on the desired variance for enhancing model's robustness: a) blending spectra, b) changing baseline, c) shifting along x axis, and d) adding random noise.



This novel data augmentation solution has been tested in order to obtain highly efficient generalised classification model based on spectroscopic data. Fourier transform mid-infrared (FT-IR) spectroscopic data of eleven pure vegetable oils (106 admixtures) for the rapid identification of vegetable oil species in mixtures of oils have been used as a case study to demonstrate the influence of this pioneering approach in chemometrics, obtaining a 10% improvement in classification which is crucial in some applications of food adulteration.

Acknowledgement: This research was supported with funding from The Department Learning and Employment Northern Ireland (DELNI).

References:

[1] M. Sáiz-Abajo, B.-H. Mevik, V. Segtnan, T. Næs, Ensemble methods and data augmentation by noise addition applied to the analysis of spectroscopic data, Analytica chimica acta 533 (2) (2005) 147-159.
[2] D. A. Van Dyk, X. -L. Meng, The art of data augmentation, Journal of Computational and Graphical Statistics 10 (1) (2001).

P-037

COMPARISON OF O2PLS-VIP WITH i-PLS AND OTHER VARIABLE SELECTION METHODS.

Beatriz Galindo-Prieto [1,2], Johan Trygg [1], Paul Geladi [3]

[1] *Department of Chemistry, CLiC, Umeå University, Sweden*
[2] *Industrial Doctoral School, Umeå University, Sweden*
[3] *Forest Biomaterials and Technology, Swedish University of Agricultural Sciences, Umeå, Sweden*
beatriz.galindo@umu.se

Keywords: variable importance, interval PLS, O2PLS-VIP.

In the age of the *internet of things*, when increasingly computerized instrumentation can produce data sets with thousands of variables, a need for reducing the number of variables keeping high interpretability has emerged as high priority in multivariate statistics. For this reason, several variable selection techniques, like interval partial least squares (i-PLS) [1] and variable influence on projection (VIP) for latent models [2], inter alia, have been developed [3].

Since the choice of the variable selection method depends on the characteristics of the data and the purpose of the multivariate analysis, fair comparisons of different variable selection methodologies need to be done in order to select the most convenient one. Some interesting comparisons have been published [4]; and here, we want to contribute presenting a fair comparison between a new VIP approach for O2PLS models [5] (which performs a variable sorting by importance in both ways X↔Y) and other variable selection methods, e.g. i-PLS, using real data.

Acknowledgement: The authors would like to acknowledge the financial support from MKS Data Analytics Solutions (formerly MKS Umetrics) (BG-P) and the Industrial Doctoral School (IDS), Umeå University, Sweden.

References:

[1] F. Lindgren, P. Geladi, S. Rännar, S. Wold, Interactive variable selection (IVS) for PLS. Part 1: Theory and algorithms, Journal of Chemometrics, 8, p. 349 (1994).
[2] B. Galindo-Prieto, L. Eriksson, J. Trygg, Variable influence on projection (VIP) for orthogonal projections to latent structures (OPLS), Journal of Chemometrics, 28, p. 623 (2014).
[3] C.M. Andersen, R. Bro, Variable selection in regression-a tutorial, Journal of Chemometrics, 24, p. 728 (2010).
[4] M. Farrés, S. Platikanov, S. Tsakovski, R. Tauler, Comparison of the variable importance in projection (VIP) and of the selectivity ratio (SR) methods for variable selection and interpretation, Journal of Chemometrics, 29, p. 528 (2015).
[5] J. Trygg, O2-PLS for qualitative and quantitative analysis in multivariate calibration, Journal of Chemometrics, 16, p. 283 (2002).

P-038

ACHIEVING ACCURATE REPRODUCIBLE RESULTS AND COMPUTATIONAL CHALLENGES IN LARGE SCALE UNTARGETED METABOLOMIC STUDIES

A. Pasamontes [1], V. Koval [1], A.C. Dubbelman [1], M.S. van Vliet [1], A.C. Harms [1], T. Hankemeier [1]

[1]Leiden Academic Center for Drug Research, Leiden University, Leiden, The Netherlands.

Untargeted feature extraction in metabolomics has been a long standing challenge in HPLC/MS [1]. Another level of complexity is added for mapping features between batches in large scale multi-batch (sometimes spread in time) studies. Apart from the complicated task to map features detected in different samples and batches and merging them in a unified output, we also have to deal with analytical drift and possible changes in instrumentation over long running projects.

In this study, we propose a workflow for untargeted metabolomics including several additional techniques to assure consistent results in large studies. This includes library building from pooled samples, inclusion of standards to correct for detector response and retention index correction to facilitate accurate alignment. Previously, using the retention time index was applied to GCMS to convert the measured retention time to system independent constants using Kovats retention index [2]. For HPLC, the offset in retention times are less constant over time than in GC, especially when working with less robust separations such as HILIC where the shifts may be unpredictable. We show that with appropriate controls in place, that rugged metabolomics profiles can be obtained and accurate relative quantitation is possible for diverse samples measured in large scale studies.

Acknowledgement:
PhenoMeNal is funded by European Commission's Horizon2020 programme, grant agreement number 654241

References:

[1] R. Di Guida, J. Engel, et al “Non-targeted UHPLC-MS metabolomic data processing methods: a comparative investigation of normalisation, missing value imputation, transformation and scaling” *Metabolomics*. (2016) 12, 93.

[2] Kind T, Wohlgemuth G, Lee DY, et al. FiehnLib – mass spectral and retention index libraries for metabolomics based on quadrupole and time-of-flight gas chromatography/mass spectrometry. *Analytical chemistry*. (2009) 81(24),10038

P-040

NEAR INFRARED ANALYSIS OF A PHARMACEUTICAL PRODUCT WITH TENDENCY TO SEGREGATION: IMPROVEMENT OF SAMPLE REPRESENTATIVENESS

Vanessa Cárdenas[1], Judit Brassier[1,2], Manel Alcalà[1], Marcelo Blanco[1], Josep Mª González[2]

[1]Applied Chemometrics Research Group, Departament de Química, Facultad de Ciències, Universitat Autònoma de Barcelona, 08193, Bellaterra, Spain.
[2]Grupo Menarini España, Alfons XII 587, 08918, Badalona, Spain
judit.brassier@uab.cat

In the last years Near Infrared spectroscopy (NIR) has become a very useful analytical technique for the pharmaceutical industry, since it provides chemical and physical information of samples in a non-destructive manner, allowing on-line qualitative and quantitative analysis. Also Chemometrics plays a fundamental role for the extraction and analysis of the information contained in the NIR spectra.
Segregation is known as the process in which the components of a powder mixture are separated by effect of an external stimulus, resulting in the spatial heterogeneity. Sifting is the most common segregation mechanism in powdered pharmaceuticals and it occurs due to the difference between particle sizes. If a pharmaceutical product undergoes to segregation during manufacturing, several problems that affect directly the quality of products and processes arise; eventually leading to batch failures related with uniformity of content and dosage units.
In this study two spectra acquisition modes were evaluated –dynamic and static- in which the spectra were acquired while the sample holder accessory was rotating. In the dynamic mode the sample area analyzed was approximately 6 times bigger than in the static mode; therefore the influence of the scanned area with the performance of the calibration models was assed. Two PLS calibrations able to quantify an API present in a formulation in 16.7% w/w were calculated after spectral acquisition using the two recording modes. Both models showed a good predictive ability RSEP dynamic mode 2.9% and static mode 5.0%. Moreover the two proposed methodologies were fully validated according the ICH & EMA guidelines.
The obtained results showed a better performance of the model using the dynamic mode in terms of accuracy, precision and robustness; in this manner the influence and importance of sample representativeness analysis In NIR analysis was confirmed.
This work represents an easy and effective alternative for the analysis of samples with tendency to segregation, and provides a contribution for the optimization of quality control methods based on NIR spectroscopy.

P-041

VISUAL STEERING IN MULTIVARIATE EXPLORATORY DATA ANALYSIS

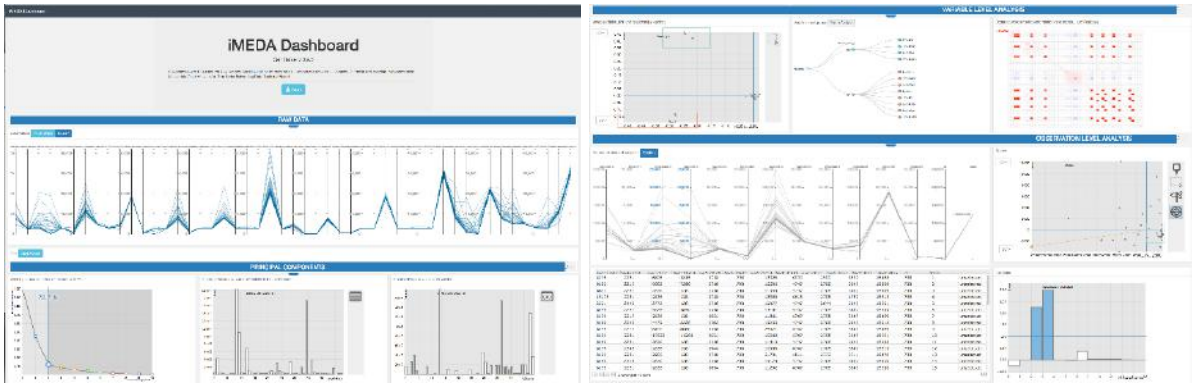
R. Therón [1], J. Camacho [2]

[1]*Department of Computer Science and Automation, University of Salamanca, Salamanca, Spain.*
[2]*Departmento de Teoría de la Señal, Telemática y Com., Universidad de Granada, Granada, Spain*
theron@usal.es

The Exploratory Data Analysis (EDA) approach has been very popular in many domains. The term "exploratory" refers to an explicit combination of the notions of visualization and interactive data analysis. Although data visualization has been in the core of EDA from its beginning decades ago, in many fields the visual aid is limited to the integration of typical EDA graphical techniques (box plots, scatter plots, histograms, Pareto charts, etc.), but many available tools fail to fully exploit the interaction aspect that any visualization may offer. This also happens with Multivariate EDA visualizations like score and loading plots.

The development of (interactive) data analysis tools has been central in chemometrics with such success that the area has been capable of exporting its algorithms and ideas to other areas of knowledge. However, there is a certain level of agreement that chemometric visualizations are somehow limited. To this regard, chemometrics may benefit from the achievements of the visual analytics realm, focused on the study of the interaction expert-visualization that leads to knowledge discovering beyond the formal modelling or hypothesis testing tasks. Bioinformatics is one example of an early adoption of a visual analytics approach to the analysis of multivariate data [1]. In this approach, artificial intelligence is combined with natural intelligence (enhanced by visualization and interaction) which allows for an intelligible human interpretation. Indeed, by properly mixing the interpretation from human beings and the high computing capabilities of machines it is possible to discover relevant (even unexpected) patterns within big data sets in an effective way [2].

In this work we integrate an interactive visual analysis approach with the processing power of the MEDA Toolbox [3] for data analysis in the SPC context.



Acknowledgement: This work is partly supported by the Spanish Ministry of Economy and Competitiveness and FEDER funds through project TIN2014-60346-R.

References:

[1] Santamaría, R., Therón, R., Quintales, L. "A visual analytics approach for understanding biclustering results from microarray data". BMC bioinformatics, 9(1), 247 (2008)
[2] Alvarado-Pérez, J. C., Peluffo-Ordóñez, D. H., Therón, R. "Bridging the gap between human knowledge and machine learning". Advances in Distributed Computing and Artificial Intelligence Journal, 4(1), 54-64 (2015)
[3] J. Camacho, A. Pérez-Villegas, R. A. Rodríguez-Gómez, & E. Jiménez-Mañas, "Multivariate exploratory data analysis (MEDA) toolbox for Matlab." Chemometrics and Intelligent Laboratory Systems, 143 49-57 (2015

P-042

MULTIVARIATE DATA ANALYSIS TO REAL-TIME MONITORING OF DISTILLATION PROCESSES BY NEAR-INFRARED SPECTROSCOPY

R.R. de Oliveira [1, 2, 3], R.H.P. Pedroza [2, 3], K.M.G. Lima [2, 3], A.O. Sousa [3], A.L. da Silva [3], A. de Juan [1]

[1] Dept. Analytical Chemistry, Universitat de Barcelona, Diagonal, 645, 08028 Barcelona, Spain.
[2] GPQBQ, UFRN, Instituto de Química, Av. Senador Salgado Filho, 3000, CEP 59078-970, Natal, Brazil
[3] LabPVT, UFRN, Av. Senador Salgado Filho, 3000, CEP 59078-970, Natal, Brazil
email: rodrirochad@gmail.com

Distillation is a separation technique employed as standard for quality assessment of petroleum derivatives. The ASTM D 86 is the standard for distillation of petroleum products at atmospheric pressure and requires temperature readings by an automatic distillation apparatus [1]. The only information obtained and required from classical distillation curves of complex mixtures is a graph of boiling temperature versus percentage of distilled volume fraction. However, relevant additional information can be obtained by in-line monitoring of the distillation process using a spectroscopic technique.

The current work aims at the application of multivariate data analysis methods to real-time monitoring of distillation by near-infrared spectroscopy (NIRS) using a device incorporating synchronized temperature readings, percentage of distilled fraction and NIRS measurements, as described in [2]. From each distilled sample, a data matrix containing the in-line NIRS absorption spectra from condensate every 5 s along the distillation and synchronized arrays of column head vapor temperature and relative condensate recovered percentage are obtained. A simple model system formed by a binary mixture (hexane + toluene, 1:1) was used to check the device performance. Commercial Brazilian vehicle fuel formed by the usual mixture of gasoline and ethanol, 27 %(v/v), from a local gas station was also analyzed.

Multivariate curve resolution alternating least squares (MCR-ALS) [3] was applied to the NIRS data providing the evolution of the relative concentration profile of the distilled compounds (either pure solvents in the hexane/toluene case or mixtures of compounds with similar boiling point in gasolines) and its spectral signatures along the distillation process. Good agreement of MCR-ALS recovered profiles as a function of the percentage of distilled volume was obtained among different replicate distillations. In fuel samples, the recovered components are related to mixtures of light hydrocarbons, ethanol and mixtures of heavy aromatic hydrocarbons. NIRS spectra of different fuel batches at 10, 50 and 90 %(w/w) recovered distilled fractions, control points representing the start, middle and end of a gasoline distillation, were also analyzed to assess differences of composition among samples. MSPC models based on the use of distilled concentration profiles and spectral signatures are envisioned as a complete fuel control approach, not only based on distillation temperatures.

The chemometric analysis of NIRS from real-monitoring of distillations may allow us to (a) understand the evolution of the distillation process identifying and interpreting the spectral signatures linked to pure compounds or group of compounds resolved and (b) to establish quality control models that allow quality assessment and detection of adulterations in final commercial fuels.

References:

[1] Standard Test Method for Distillation of Petroleum Products at Atmospheric Pressure, ASTM Standard D 86 - 12, Book of Standards, West Conshohocken, PA, 2012.
[2] C. Pasquini, S.H.F. Scafi, Anal. Chem., 75, 2270 (2003).
[3] A. de Juan, J. Jaumot, R. Tauler, Anal. Methods, 6, 4964 (2014).

P-043

EXTENDED ITERATIVE OPTIMIZATION TECHNIQUE (EIOT) FOR MIXTURE COMPOSITION ESTIMATION IN MOVING POWDERS

Zhenqi (Pete) Shi [1], Salvador García Muñoz [1]

[1]*Small Molecule Design and Development, Lilly Research Laboratories, Indianapolis, IN, 46285, USA*

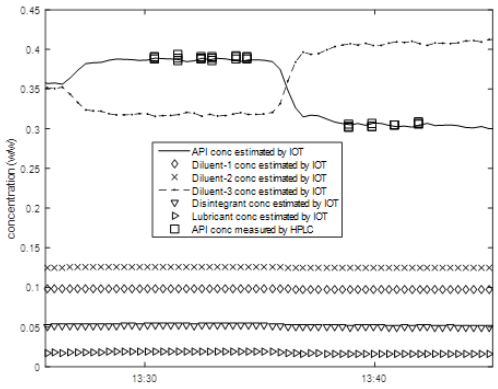
Email: shi_zhenqi@lilly.com

The calibration and maintenance effort for a near infra-red sensor and calibration based solution can be a time, labor and material-intensive exercise. A calibration-free method (iterative optimization technique or IOT) was recently published and applied onto mixture component prediction.[1]. IOT estimates fractions of individual components via optimization by rigorously imposing Beer-Lambert's law and mixture constraints to the measured spectra, given the known spectrum for the pure materials. Advances in IOT theory propose the use of massive blind-search evolutionary algorithms to select spectral regions that can improve the performance of IOT in a real-time scenario.

The application of IOT onto powder flows is severely affected by scattering and uneven sample presentation (among others) which results in large deviations from Beer-Lambert in the mixture spectra. Other application-specific artifacts (like peak overlap) can also influence the performance of IOT, resulting in predictions with trends and biases that can range from mild, to completely unreasonable and inaccurate. This work presents an extension to IOT that preserves the calibration-free nature of the method while improving its real-time performance in the estimation of fractions of individual components in the powder stream of a continuous drug product manufacturing line.

Although still calibration free, EIOT requires at the very least measured spectra at a minimum of two levels of concentration for the active ingredient – this data set is readily available if a step change study is done in the line to determine mixing behavior of the system. The power of the method resides in the spectral differentiation across materials, as such, a data set consisting of spectral measurements at several levels of the individual components is much preferred [2].

EIOT fundamentally challenges the application of Beer-Lambert in relating the fractions of each material through the mixture spectra in the process and pure spectra collected in a static sample. As such, EIOT starts with the estimation of the dynamic pure component spectra that is mathematically necessary such that, the concentrations and the dynamic mixture spectra in the "training set" is explained by Beer's law. Multiple methods are presented to produce this estimate. It has been observed that the dynamic pure spectra estimated differ only in small features from the measured pure spectra. Once these pure spectra are estimated, the procedure continues as IOT would. These differences, although small play an important role in the performance of the method. Several examples are presented, contrasting the performance of a PLS model, IOT and EIOT in estimating the fractions of ingredients from spectra measured with a NIR instrument installed in the feed frame of a tablet press, at the end of a continuous drug product manufacturing line. The trends produced by EIOT lack undesired trends that are attributed to process effects onto the spectra, and represent a useful metric to monitor the real-time fraction of the materials in the feed-frame. Such an estimate can be used in combination with a deterministic model of the system to confirm mixture composition and confirm the state of control of the process.



Graphic files: Predicted concentrations of six components of a powder mixture vs. time profile via the use of EIOT

References:

[1] Muteki, K., Blackwood, D.O., Maranzano, B., Zhou, Y., Liu, Y.A., Leeman, K.R. and Reid, G.L. Industrial & Engineering Chemistry Research, 52, 12258 (2013).

[2] Eriksson, L.; Johansson, E.; Wikstrom, C. Mixture design--design generation, PLS analysis, and model usage. Chemometrics and Intelligent Laboratory Systems, 43, 1-2, 1-24. (**1998)**

P-044

A portable and low-cost near-infrared spectroscopy (NIR) spectrometer that safeguards the solid pharmaceutical industry

D. Sun, M. Alcalà, M. Blanco

Applied Chemometrics Research Group, Department of Chemistry, Faculty of Sciences, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain
Dong.sun@uab.cat

15. Introduction

Near infrared spectroscopy (NIRS) is an important tool for Process Analytical Technology (PAT)[1] and the instrument miniaturization makes NIRS a practical solution for pharmaceutical industry. In order to reduce the cost for changes in production line and simplify the modeling process for analyzing different active pharmaceutical ingredients (API), the portable, low-cost and miniaturized spectrometer needs to be adopted to reach this goal.

16. Objective

The chief aim of this research is to investigate a quantitative method which is suitable for application of NIRS in solid pharmaceutical industry.

17. Materials and methods

Near infrared spectrum of two different pharmaceutical formulations (A and B), containing only one active principle ingredient (API 1 or 2, aprox concentration 85 and 50 mg/g, were recorded by the portable Micro-NIR (Viavi Solutions Inc., formerly JDSU, unit NO.38, USA). It´s dimension is (diameter x height) 45 x 42mm, weight<60g, and spectral region is from 1158.800nm to 2153.100nm, and resolution <1.25% of center wavelength, for example, at 1000nm the resolution is 12.5nm.

Granulate industrial samples of the two pharmaceutical formulations from several production batches were obtained from Laboratorios Menarini S.A (Badalona, Spain) and a number of laboratory powder samples were prepared. Laboratory samples consisting of accurately weighed amounts of the powder ingredients spanning a concentration range ±20% around the nominal API content were prepared according to the ICH guidelines. The sample set was established with a design which can minimize the correlation between pairs of concentrations. Quantitative partial least squares regression (PLSR) models were calculated with Unscrambler X (10.3 Trondheim, Norway).

18. Results and discussion

The calibration model of API 1 was calculated by PLSR with 7 factors. The pretreatment was median filter smoothing (segment size 3) + standard normal variate (SNV) + Gap-segment Derivatives (order 1, gap size 5, segment size 2)and the spectral region was 1558.15-1851.55nm. The calibration model of API 2 was calculated by PLSR with 7 factors. The pretreatment was Median filter smoothing (segment size 7) + Norris 1st derivate (gap size 7) and the spectral region was 1158.80-2153.100 nm.

According to the ICH guidelines, the linearity, range, accuracy, robustness, repeatability, and intermediate precision of PLSR models have been validated external sample sets. These items were assessed by relative standard deviation (RSD), paired t-test ($p<0.05$), and analysis of variance (ANOVA). The results showed excellent performance for both methods..

19. Conclusion

The Micro-NIR has a good ability to be a PAT monitor of solid pharmaceutical formulations.

Acknowledgement: Laboratorios Menarini S.A. (Badalona, Spain) for kindly providing samples and (Viavi Solutions Inc., formerly JDSU, USA) for the kindly providing the Micro-NIR spectrometer.

References:

[1] PAT – a framework for innovative pharmaceutical development, manufacturing, and quality assurance, http://www.fda.gov/cder/guidance (2004).

P-045

MULTIVARIATE CONTROL CHARTS BASED ON NEAR INFRARED SPECTROSCOPY FOR MONITORING TRANSESTERIFICATION REACTIONS FOR BIODIESEL PRODUCTION

R.F. Sales [1], R. Vitale [2], S.M. Lima [3], M.F. Pimentel [1], L. Stragevitch [1], A. Ferrer [2]

[1]Department of Chemical Engineering, Universidade Federal de Pernambuco, Av. Prof. Arthur de Sá, Recife, Brazil.
[2]Departamento de Estadística e Investigación Operativa Aplicadas y Calidad, Universitat Politècnica de València, Camino de Vera, 46022 Valencia, Spain.
[3] Instituto Federal de Pernambuco, Av. Prof. Luís Freire, Recife, Brazil.
rafaelladfs@gmail.com

Nowadays, biodiesel is looked at as a potential alternative to fossil fuels. It is mainly produced by transesterification of vegetable oils with an alcohol. This reaction, frequently carried out in batch reactors, can be affected by several parameters such as temperature, catalyst content and type, stirring speed, alcohol type and feedstock variability [1,2]. Due to its dynamic nature, this type of process requires an accurate monitoring and control, oriented to ensure quality uniformity and fulfil the commercialization requirements of the final product. In this context, the present work describes an application of multivariate statistical process control to monitor methanol-mediated soybean oil transesterification.

For the development of multivariate control charts, near infrared spectroscopy (NIRS) data were acquired in-line during the evolution of ten batches produced under normal operating conditions (0.75 w/w% of catalyst with respect to the amount of oil, temperature of 55ºC and stirring speed of 500 rpm). They were then organized in a three-way array (batch x spectral variable x time). The three-dimensional structure was then analyzed by the two procedures described in [3] and [4], respectively. The first approach, which will be referred to as NM, unfolds the data so that the batch direction is preserved. The resulting matrix is subsequently subjected to Principal Component Analysis (PCA). Conversely, the second, which will be referred to as WKFH, initially unfolds the three-way array so that the variable direction is preserved. Afterwards, a Partial Least Squares (PLS) regression model is built between the resulting two-dimensional array and the so-called *local batch time*, a measure of the batch maturity index. This approach also involves a second modelling step in which the PLS-scores are rearranged batch-wise and analyzed by PCA.

To evaluate the performance of the two techniques in terms of off-line (*end-of-batch*) and on-line (*real-time*) fault detection and diagnosis capability, eight test batches, characterized by deviations in the temperature and stirring conditions as well as in the catalyst concentration, were manufactured. In general, all the constructed off-line monitoring schemes were found to correctly point out and identify most of such induced failures. On the other hand, concerning on-line monitoring, the NM approach showed a better performance, because, unlike WKFH, it was able to detect and diagnose small variations in the stirring speed.

Acknowledgement: PRH-28, ANP, Petrobras, FACEPE-NUQAAPE, INCTAA, CNPq.

References:

[1] A.E. Atabani, A.S. Silitonga, I.A. Badruddin, T.M.I. Mahlia, H.H. Masjuki, S. Mekhilef, Renewable and Sustainable Energy Reviews, 16, 2070 (2012).
[2] S.M. de Lima, B.F.A. Silva, D. V. Pontes, C.F. Pereira, L. Stragevitch, M.F. Pimentel, Fuel, 115, 46 (2014).
[3] P. Nomikos, J.F. MacGregor, Technometrics, 37, 41 (1995).
[4] S. Wold, N. Kettaneh, N. Fridén, A. Holmberg, Chemometrics and Intelligent Laboratory Systems, 44, 331 (1998).

P-046

LATENT-VARIABLE MODEL INVERSION AND DESIGN OF EXPERIMENTS IN THE LATENT SPACE: A COMPARISON ORIENTED TO PROCESS/PRODUCT DESIGN

D. Palací-López [1], P. Facco [2], M. Barolo[2], A. Ferrer[1]

[1]Multivariate Statistical Engineering Group, Departament of Applied Statistics and operational Research, and Quality, Universitat Politècnica de València, Camino de Vera s/n, 7A, 46022, Valencia, Spain
[2]CAPE-Lab – Computer-Aided Process Engineering Laboratory, Department of Industrial Engineering, University of Padova, via Marzolo 9, 35131 Padova PD, Italy
email: dapalpe@gmail.com

A common approach when dealing with a production process consists in building a model from historical data – or data from a Design of Experiments (DOE) – in order to explore, better understand and/or optimize it. To do this, methods based on projection to latent structures (PLS) have been proposed, especially when a large number of potentially correlated variables influence the process outputs. This way the prediction of the final product properties from the process conditions is made possible, with a certain uncertainty. However, in most cases the required quality attributes of the final product have already been defined, and the problem is finding the conditions under which these attributes are fulfilled.

On the other hand, as pointed in [1], in fields such as the pharmaceutical industry, strict regulations prevent performing any changes in the process conditions that will bring them outside of its so-called 'design space', wherein the desired product properties can be guaranteed. Therefore, establishing the appropriate procedure to guarantee the desired results while meeting these restrictions is of utmost importance.

In this study, a latent variable model inversion (LVM-I) approach as proposed in [2], together with design space bracketing [1], will be applied and compared to using DOE in the latent space as a proposed alternative. Then, the effectiveness of combining both approaches will be studied. The main goal will be to illustrate the performance of each methodology in terms of achieving the desired output for a specific process, depending on the quality of the initial dataset from which the first model is built, the available budget for experimentation, and the complexity of the correlation structure among explanatory variables, outputs or both.

Lastly, this methodologies will be evaluated when applied to a mixture design problem, for which the ratios of the different components in a blend are at least of as much relevance as their absolute quantities. Some algorithms have been proposed in the literature to simultaneously take into account process conditions, raw material rates and their properties [3,4] when building a PLS-based regression model, which makes them relevant for fields such as the pharmaceutical, chemical and bioprocess industry. However, little research has been carried out up until now regarding the application of LVM-I and DOE in this case.

References:

[1] P. Facco, F. Dal Pastro, N. Meneghetti, F. Bezzo, M. Barolo, Industrial and Engineering Chemistry Research, 54, 5128 (2015)
[2] E. Tomba, M. Barolo, S. García-muñoz, Industrial and Engineering Chemistry Research, 51, 12886 (2012)
[3] K. Muteki, J.F. MacGregor, Chemometrics and Intelligent Laboratory Systems, 85, 186 (2007).
[4] S. García-Muñoz, M.A. Polizzi, Chemometrics and Intelligent Laboratory Systems, 114, 116 (2012)

P-047

NEW INSIGHTS IN THE ULTRASONIC EXTRACTION OF ARSENIC FROM SOILS AND SEDIMENTS USING DESIGN OF EXPERIMENTS

E. Castillo S.[1], R. Pérez R. [1]

[1]Department of Chemistry, National University of Colombia, Carrera 45 # 26-85, Bogotá, Colombia
rperezr@unal.edu.co

The Design of experiments (DoE) has been used as a strategy for the optimization of Arsenic ultrasound-assisted extraction (UAE) in environmental matrices[1]. The high influence of the flask position inside the ultrasonic bath has been recognized. However, the combined effect of variables such as the extractant and matrix nature, the extraction time and matrix-extractant ratio has been less studied. In this work, we used the screening Plackett-Burman design (PB) and the optimal central composite design (CCD) to determine the main effects of these variables and develop a model for the optimization of the arsenic ultrasound assisted extraction from soils and sediments. Certified soil and sediments (MATSD0105, MATSL0105, CRM016050, CRM052050, NIST 1646a, NIST 2709a y NIST 2711a) and sediments from the Vetas-California mine district (Santander, Colombia)[2] were used. The statistical analysis and arsenic extraction were carried out by using the R language and hydride generation continuum source atomic absorption spectroscopy (Analytica Jena contrAA 700/HS55A) respectively. From the model obtained ($R^2 \geq 0.95$), the soil-sediment nature and the extractant concentration were found as the most important variables in the extraction process. The optimized methodology showed recoveries from 92 to 98% and a detection limit of 0.10 $\mu g.kg^{-1}$. The conditions reached were extractant: diluted HNO_3 of 0.12 M and extraction time: of 2 min. These experimental conditions were more efficient than the traditional methodology (EPA 3050b) which suggests an extraction time of 2 hours and an extraction medium more aggressive (aqua regia 1:3 HNO_3, HCl).

Keywords: DoE, Arsenic, UAE.

References

1. Paula, J.F.R., R.E.S. Froes-Silva, and V.S.T. Ciminelli, *Arsenic determination in complex mining residues by ICP OES after ultrasonic extraction.* Microchemical Journal, 2012. **104**: p. 12-16.
2. Alonso, D.L. and E. Castillo, *Arsenic availability in a contaminated area of Vetas-California gold mining district - Santander, Colombia*, in *One Century of the Discovery in Latin America (1914-2014).* 2014. p. 242-244.

P-048

VISUALIZATION OF THE PARETO-OPTIMAL FRONT VIA PARALLEL COORDINATES PLOT FOR INTERPRETING RESULTS OF EXPERIMENTAL DESIGNS FOR MIXTURES

I. Pérez [1], M.C. Ortiz [1], L.A. Sarabia [2], M.S.Sánchez [2]

[1]Department of Chemistry (Analytical Chemistry), University of Burgos, Plaza Misael Bañuelos s/n, Burgos, Spain
[2]Department of Mathematics and Computation, University of Burgos, Plaza Misael Bañuelos s/n, Burgos, Spain
ssanchez@ubu.es

One of the distinctive characteristics of experimental designs for mixtures is that their factors are always linearly dependent, because in each experiment the variables should be positive and add up to one. The usual mixture plots in the simplex mixture space, allows representing the experimental domain up to three factors. However, with more than three factors and/or more than one experimental response to study, it is difficult to interpret the results and more even if the responses have to be simultaneously optimized.

We present here, through some case-studies, a way to handle this situation, to more easily interpret experimental results via the fitted model, and also to find optimal experimental conditions that provide a compromising solution among several conflicting responses.

For example, a problem is to determine the proportions of three colorants (Indigo Carmine, E-132, Ponceau 4R, E-124, and Tartrazine, E-102) to achieve a specific colour, by using the CIELab parameters L^*, brightness, a^*, red/green chromaticity, and b^*, yellow/blue chromaticity. The scheme is to perform an experimental design with mixtures of the three colorants and then use the mathematical models fitted to the three colour parameters to approach each value (L^*, a^* and b^*) of the problem sample. By adapting the procedure explained elsewhere [1], the Pareto-optimal front for the three distances to the targeted colour is computed. A variation [2] to improve visualization of the parallel coordinates plot allows representing together (no matter the number of factors and responses) the mixtures and distances to the values of L^*, a^* and b^* and, thus, allows studying the simultaneous behaviour of the colorants and their expected CIELab values. The valid mixtures are those for which each distance is less than 0.577, which is the difference that is no perceptible by the human eye.

The procedure is also applied with other case-studies that include mixtures with more than four factors and with sensory responses to be optimized.

Acknowledgement: The authors acknowledge the financial support of Spanish Ministerio de Economía y Competitividad through project CTQ2014-53157-R, co-financed with European FEDER funds.

References:

[1] M.C. Ortiz, L.A. Sarabia, A.Herrero, M.S. Sánchez, Chemom and Intell Lab Systems, 83, 157 (2006).
[2] M.C. Ortiz, L.A. Sarabia, M.S. Sánchez, D. Arroyo, Anal.Chim.Acta, 687, 129 (2011).

P-049

D-OPTIMAL EXPERIMENTAL DESIGN METHODOLOGY IN THE DETERMINATION OF BIOGENIC AMINES IN FISH BY HPLC-FLD

A. Herrero [1], S. Sanllorente [1], C. Reguera [1], M.C. Ortiz [1], L.A. Sarabia [2]

[1]Department of Chemistry, University of Burgos, Pza. Misael Bañuelos s/n, Burgos, Spain
[2]Department of Mathematics and Computation, University of Burgos, Pza. Misael Bañuelos s/n, Burgos, Spain
aherrero@ubu.es

Biogenic amines are organic compounds which can cause health problems if ingested over certain levels, for that reason their presence is regulated in some foodstuff [1]. In addition, they are frequently related to the quality of some foods as fish or fish products since they are a sing of the unfreshness or unadequate hygienic storage conditions or of degradation of processed or fermented foods [2]. So the development of analytical procedures for determining these compounds in such complex matrices is an issue of analytical interest.

The determination of biogenic amines in this kind of matrices requires steps previous to the analysis, such as extraction, derivatization, clean-up, etc. The optimization of these steps implies a considerable number of experiments since many experimental factors are usually involved and interactions among some of them can be expected. It is desirable to apply a strategy for reducing the experimental effort needed.

In this work, the use of a D-optimal design [3] for optimizing the pretreatment steps of the determination of some aliphatic and aromatic biogenic amines in fish by high liquid chromatography with fluorescence detection (HPLC-FLD) is reported. Cadaverine, putrescine, spermidine, spermine, histamine, phenylethylamine, tyramine and tryptamine are determined in swordfish (*Xiphias gladius*).

The extraction with an acid (trichloroacetic or perchloric acid) from the solid matrix and the derivatization with dansyl chloride of the acidic extract are optimized. Seven experimental factors at two levels and three factors at three levels which are involved in both steps are considered. A model with 19 coefficientes, which includes those corresponding to the principal effects and two interactions is fitted for each amine. A full factorial design would require 3456 experiments to estimate the coefficients of this model, whereas using the D-optimal experimental desing strategy the number of experiments is reduced to only 23 (plus some replicates of one of the points of the design). The quality of the estimates is guaranteed since the variance inflation factors (VIFs) of the coefficients of the model range from 1.08 to 1.78, which means precise estimates of the coefficients. Once the analyses are performed, a multiobjective optimization technique, the Derringer's desirability function [4], is used to simultaneously considering the response obtained for all the amines.

Acknowledgement: The authors thank the financial support provided by the MINECO (CTQ2014-53157-R) and FEDER funds.

References:

[1] Commission Regulation (EU) No 1019/2013 of 23 October 2013 amending Annex I to Regulation (EC) No 2073/2005 as regards histamine in fishery products, Off. J. Eur. Commun. L282 of 24 October 2013
[2] C. Ruiz-Capillas, F. Jiménez-Colmenero, Biogenic amines in seafood products, in: L.M.L. Nollet, F. Toldrá (Eds.), Safety Analysis of Foods of Animal Origin, CRC Press, Boca Raton, FL, 2010, pp. 743-760
[3] P.F. de Aguiar, B. Bourguignon, M.S. Khots, D.L. Massart, R. Phan-Than-Luu, Chemom. Intell. Lab. Systems, 30, 199 (1995)
[4] G. Derringer, R. Suich, J. Qual. Technol. 12, 214 (1980)

P-050

CALIBRATION TRANSFER OF NEAR-INFRARED MODELS: ADDITIVE PLS MODELS DEMONSTRATED WITH TWO INDUSTRIAL CASE STUDIES

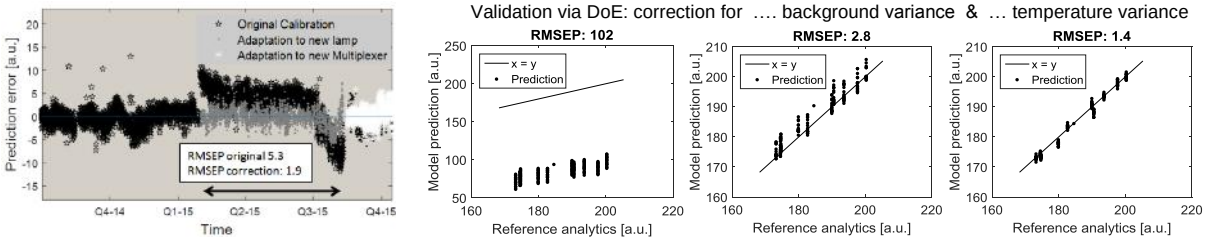
Pekka Luoma[1], Birgit Malli[2], Thomas Natschläger[2], Marcin Pawliczek[1], Markus Brandstetter[1]

[1]RECENDT – Research Center for Non-Destructive Testing, Altenbergerstrasse 69, Linz, Austria
[2]SCCH – Software Competence Center Hagenberg, Softwarepark 21, Hagenberg, Austria
pekka.luoma@recendt.at

In recent years, the requirements for Near-infrared spectroscopy have grown from passive regression vector calculations to include model performance analysis and to provide life-cycle management aspects, e.g. ease of recalibration, with the motivation to increase model confidence and to save resources. As in a typical industrial workflow Partial Least Squares (PLS) regression models are initially generated with off-line and on-line NIR data, followed by laborious model recalibration in case of changing light intensity, process conditions or raw materials, a great deal of expert knowledge is required also on a short notice.

In this contribution each new source of variance is handled independently simplifying the the chemometric tasks and thus enabling partial automation. Additive correction models based on Generalized Additive Model (GAM) approach, where the calibration models are left unadjusted are applied: After calibrating an initial PLS model, its performance in terms of prediction error is regularly checked using reference analysis. The resulting prediction residuals are used as target values and the corresponding spectra as the input variables for generating a PLS model that corrects these prediction errors. This workflow avoids elaborate re-modelling of the initial model. Two industrial case studies are presented:

Case-Study 1: In an industrial process monitored by NIR for several years a light source is changed. Calibration model off-set after the lamp change is used as an input for calculation of a correction model. The procedure is repeated after the change of an optical multiplexer. In the figure below (left) the prediction residuals of the models are presented: Black dots correspond to the original calibration performance, grey dots adaptation to new lamp and white dots adaptation to new multiplexer. The acquired error of prediction (RMSEP) is only slightly over the accuracy of the reference analytics.



Case-Study 2: In most real-life scenarios process conditions cannot be achieved in a laboratory environment. In this case-study Design of Experiment (DoE) approach measurements using stock analytes were conducted for multi-analyte feasibility study in the laboratory. Initially, a calibration model for the undisturbed laboratory experiment was established. Additive models were then generated using spiked process samples to indepentently introduce process background and temperature variance for applicability in industrial environment. The resulting RMSEP is well suitable for the targeted in-line process monitoring. The evolution of the prediction errors of a spiked process samples through the three models are presented in the figure above (right).

Acknowledgement: Financial support was provided by the Austrian research funding association (FFG) under the scope of the COMET program within the research project "Industrial Methods for Process Analytical Chemistry - From Measurement Technologies to Information Systems (imPACts) (contract #843546).

P-051

MODEL LIFE-CYCLE MANAGEMENT FOR COMPLEX PAT-SYSTEMS BASED ON NIR-SPECTROSCOPY

Marcin Pawliczek[1], Birgit Malli[2], Thomas Natschläger[2], Thomas Reischer[3], Wolfgang Kantner[3], Markus Brandstetter[1]

[1] RECENDT - Research Center for Non-Destructive Testing GmbH, Linz – Austria [2] Software Competence Center Hagenberg GmbH, Hagenberg – Austria [3] Metadynea Austria GmbH, Krems - Austria
marcin.pawliczek@recendt.at

A comprehensive concept and first corresponding tools for an efficient management of process analytical technology (PAT) systems based on NIR-spectroscopy and chemometrics were developed and applied. Nowadays, a variety of complex PAT systems are applied in chemical industry. An increasing number of manual measurement steps are being replaced by spectroscopic analysis and the corresponding chemometric models, aiming on a reduction of material and personnel costs.
In situations where a large number of chemometric models are in use, e.g. at different process stages and for different products, model maintenance efforts increase significantly. Each of these tasks requires the development of a specific chemometric model as well as regular model maintenance operations. Instead of following the common approach of manual model updates, we suggest a more efficient and automated approach based on tools of Computational Model Life-Cycle Management (CMLCM): These tools shall guarantee the viability and transferability of the implemented chemometric models, permit in-house model updating without excessive resources from external experts and enable the use of existing models for comparable applications. In this contribution we focus on techniques for the automation of model recalibration and suggest a suitable workflow for this purpose. The following basic steps are covered: data selection, data preparation, model calibration as well as validation of the developed model. Model recalibration is then triggered via the use of reliability measures, computed directly on new spectroscopic data.
A practical example for the application of several CMLCM tools in an industrial process is shown for a PAT installation at a resin production plant (Figure 1). The main purpose there is to monitor and control batch condensation processes of several melamine formaldehyde (MF) and phenolic formaldehyde (PF) resin recipes. Furthermore, control of raw substance dosage prior to production, as well as post-process quality control is shown.

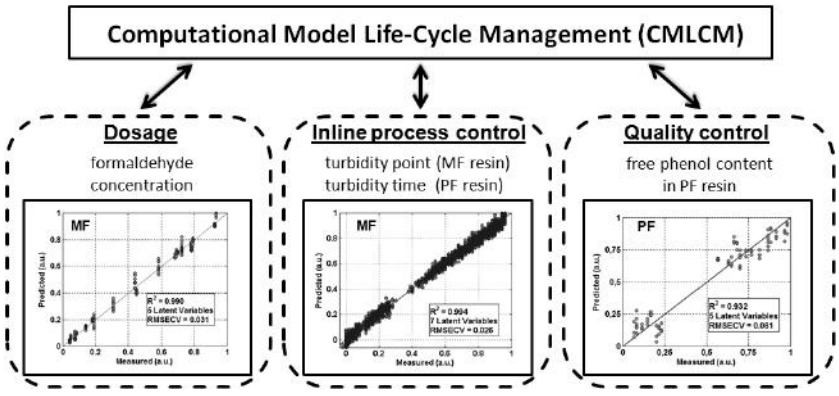


Figure 1: Three process steps at a resin production site serving as practical examples for comprehensive process modelling and integration of CMLCM concepts

Acknowledgement: Financial support was provided by the Austrian research funding association (FFG) under the scope of the COMET program within the research project "Industrial Methods for Process Analytical Chemistry - From Measurement Technologies to Information Systems (imPACts) (contract #843546)

P-052

OPTIMIZATION OF HS-SPME TO QUANTIFY VICINAL DIKETONES IN BEER BASED ON AN OPTIMAL DESIGN OF EXPERIMENTS

A. C. Pereira [1,2], J. M. Leça [1], A. C. Vieira[1], M. S. Reis [2], J. C. Marques [1,2]

[1] Centre of Exact Sciences and Engineering, University of Madeira, Funchal, Portugal,
[2] CIEPQPF, Department of Chemical Engineering, University of Coimbra, Pólo II - Rua Sílvio Lima, 3030-790 Coimbra, Portugal
acpereira@uma.pt

Application of Design of Experiments (DoE) methodologies in the development and optimization of analytical procedures is currently a crucial step. The advantages are clear: the lowest cost, since it implies lower number of trials and consequently, the consumption of reagents, standards and samples is reduced, guarantees statistically meaningful conclusions and safeguards against erroneous and biased procedures which can lead to wrong or suboptimal conclusions (usually obtained by univariate optimization, one-at-a-time manipulation of factors) [1, 2].

In this study, we focus in the optimization of HS-SPME to quantify Vicinal Diketones, namely diacetyl (DC) and pentanedione (PN) in beer. Six factors were analysed, namely the type of fiber (DVB/PDMS, Car/PDMS and DVB/Car/PDMS), the sample volume, the pre-incubation time, time and temperature of extraction and the effect of agitation. Optimal Design of Experiments (O-DoE) was the approach chosen to define the testing conditions based on the D-optimality criterion. Accordingly, thirty six experiments were planned in order to estimate the main effects and second order interactions for variables analyzed and according to trade-off established between the material and time resources available and the accuracy needed to build the final model.

The fiber type and sample volume were found to be the factors playing a major role in the extraction process. Also, several interaction effects stand out, such as fiber coating x sample volume, incubation time x extraction temperature and extraction time x sample volume. According to the results, the following conditions were established as the optimal to quantify VDKs by HS-SPME priori to GCMS: CAR-PDMS, 5 min of pre-incubation followed by an extraction of 30 min at 30ºC and considering 5 ml of sample in a 20 ml vial with agitation.

The validation of the final analytical methodology was performed using a matrix-matched calibration and the following key features were obtained: linearity (R^2 > 0.999, both for diacetyl and 2,3-pentanedione), high sensitivity (LOD of 0.92 g L−1 and 2.80 g L−1, and LOQ of 3.30 g L−1 and 10.01 g L−1, for diacetyl and 2,3-pentanedione, respectively), recoveries of approximately 100% and suitable precision (repeatability and reproducibility lower than 3% and 7.5%, respectively). The applicability of the methodology was also confirmed for several beer samples [3].

Acknowledgement: The authors acknowledge *Empresa de Cerveja da Madeira* for kindly supplying the samples used in this work. João M. Leça and Ana C. Pereira acknowledge Agência Regional para o Desenvolvimento da Investigação Tecnologia e Inovação (ARDITI) for the financial support through their grants, 0002376/2014/131 and 002248/2013/131, respectively.

References:

[1] D.B. Hibbert, J. of Chromatography B, 910, 2 (2012).
[2] R. Leardi, Analytica Chimica Acta 652, 161 (2009).
[3] J.M. Leça, A.C. Pereira, A.C. Vieira, M.S. Reis, J.C. Marques, Analytica Chimica Acta 887, 101 (2015).

P-053

DEFINITIVE SCREENING DESIGN APPLIED TO THE OPTIMIZATION OF SOLID-PHASE MICROEXTRACTION FOR VOLATILE FATTY ACIDS ANALYSIS IN WINES

A. C. Pereira [1,2], P. M. Rodrigues [1], J. M. Leça [1], M. S. Reis [2], J. C. Marques [1,2]

[1] Centre of Exact Sciences and Engineering, University of Madeira, Funchal, Portugal,
[2] CIEPQPF, Department of Chemical Engineering, University of Coimbra, Pólo II - Rua Sílvio Lima, 3030-790 Coimbra, Portugal
acpereira@uma.pt

During the last decade, we have been witnessing a shift of paradigm regarding the optimization of analytical methodologies. The one-factor-at-a-time approach have been replaced by design of experiment methodologies, which are much more efficient, faster, require fewer resources and produce results that are more precise and statistically sound. Definitive Screening Designs (DSDs) are usually a good starting point for factor screening. Comparing with standard screening designs, they present additional advantages, namely a fairly small experiment (small number of runs), avoid model ambiguity, identify important factors more quickly and efficiently, identify factors having a nonlinear effect on the response and estimate quadratic effects (in models containing only main and quadratic effects) [1, 2].
In the present study, DSDs was applied to the optimization of solid phase micro-extraction (SPME) procedure to quantify nine volatile fatty acids (VFA) in wines. Seven factors were analyzed, namely type of fiber (PA and DVB/Car/PDMS), the sample volume and its dilution, the pre-incubation time, the time and temperature of extraction and the effect of agitation. Accordingly to DSD plan and parameters under studied, eighteen experiments were carried out and the values for the response variable (peak areas) were recorded and used to estimate the effects associated with all factors considered, as well as their interactions.
The data analysis was done individually for each one of the nine compounds studied. It was found that the extraction temperature, the dilution and the sample volume were the factors playing a major role in SPME, regardless the VFA analyzed. Accordingly, the following conditions were established as the optimal: fiber type DVB/Car/PDMS, no pre-incubation time, extraction during 40 minutes at 40^0C, considering 10 ml of sample without any dilution in a 20 ml vial and apply agitation.
Regarding the performance of the analytical procedure, the results were good in terms of linearity, sensitivity, selectivity, precision and accuracy. The method was then applied to several wines, namely samples collected during the fermentation process (about twenty samples) as well as commercial wines (twenty four samples).

Acknowledgement: Ana C. Pereira and João M. Leça acknowledge Agência Regional para o Desenvolvimento da Investigação Tecnologia e Inovação (ARDITI) for the financial support through their grants.

References:
[1] B. Jones, C.J. Nachtsheim, Journal of Quality Technology 43, 1 (2011).
[2] B. Jones, C.J. Nachtsheim, Technometrics 53, 62 (2011).

P-054

OPTIMIZATION OF A QUECHERS-BASED METHOD FOR THE DETERMINATION OF SOTOLON IN FORTIFIED WINES USING FULL FACTORY DESIGN

A. I. Freitas [1], J. M. Leça [1], V. Pereira [1,2], A. C. Pereira [1,3], J. C. Marques [1,2]

[1] *Faculty of Exact Sciences and Engineering, University of Madeira, Campus da Penteada, 9020-105 Funchal, Portugal.*
[2] *Institute of Nanostructures, Nanomodelling and Nanofabrication (I3N), University of Aveiro, 3810-193 Aveiro, Portugal*
[3] *CIEPQPF, Department of Chemical Engineering, University of Coimbra, Rua Sílvio Lima, 3030-790, Coimbra, Portugal*
vpereira@uma.pt

Sotolon (3-hydroxy-4,5-dimethyl-2(5H)-furanone) is a chiral lactone and a well-known powerful odorant, which can impart a nutty/caramel/curry/rancind odour to wines, depending on its concentration and enantiomeric distribution [1, 2]. Lately, researcher's attention has been directed for its off-flavour character, associated to the premature oxidative ageing of young dry white wines, overlapping its freshness. Conversely, sotolon has been pointed out as a key odorant of aged fortified wines, such as Sherry, Port and Madeira wines [2-4], being quantified above its odour threshold (8-10 µg/L).
Considering that sotolon is associated with high-quality fortified wines, and, therefore, essential for its valorisation (quality and price), an accurate quantitative method to follow this key-odorant during wine production and in the final product was planned in an optimal way and the final settings were fully validated. For the QuEChERS-based optimization full factorial design was employed. The optimization procedure took into account three experimental factors, which included sample volume (3 levels), solvent volume (3 levels) and concentration of the extract (2 levels). Nineteen experiments were carried out in duplicate. The DoE results were then evaluated leading to the following optimal extraction conditions for the quantification of sotolon: 8 mL of wine sample, 5 mL of dichloromethane and concentration of 10-fold. Extractions and injections of each sample were carried out in duplicate.
The method validation was performed using a matrix-matched calibration, in order to minimize matrix effects. The optimized method showed good linearity ($R^2 > 0.999$), high sensitivity (LOD and LOQ of 2.3 and 6.8 µg/L, respectively), recoveries about 105% and suitable precision (repeatability and reproducibility lower than and 8%). The method applicability was confirmed through the analysis of 24 fortified wines and the concentrations found ranged between 8 to 294 g/L.

Acknowledgement: Vanda Pereira and Ana Cristina Pereira acknowledge *Agência Regional para o Desenvolvimento da Investigação Tecnologia e Inovação* (ARDITI) for the financial support of their grants.

References:
[1] A. Pons, V. Lavigne, Y. Landais, P. Darriet, D. Dubourdieu, J. Agric. Food Chem., 56, 1606 (2008).
[2] A.C.S. Ferreira, J.C. Barbe, and A. Bertrand, J. Agric. Food Chem., 51, 4356 (2003).
[3] B. Martin, P.X. Etiévant, J.L. Le Quéré, P. Schlich, J. Agric. Food Chem., 40, 475 (1992).
[4] J.S. Câmara, J. Agric. Food Chem., 52, 6765 (2004).

P-055

MULTIVARIATE OPTIMIZATION OF THE SYNTHESIS OF POLYMERIC MICROPARTICLES OF NUTRIENT CARRIERS TO PLANTS

J.O. Chagas[1], J.M. Gomes[1], I.C.M. Cunha[1], N.F.S. de Melo[2], G.A. da Silva[1], F.A. Lobo[1]

[1]Departamento de Química, Universidade Federal de Ouro Preto (UFOP), MG, Brazil.
[2]Departamento de Imunologia e Biologia Molecular, Universidade Estadual de Campinas (UNICAMP), SP, Brazil.
fabiana@iceb.ufop.br

Improper use of fertilizers can lead to waste, cause losses in agricultural productivity and possible environmental contamination problems [1]. Thus, there have been the need for developing alternative fertilization methods that provide the minimization and/or elimination of such problems. Thus, controlled release systems for fertilizing gained evidence for providing the abovementioned aspects [2]. In this work, it was used poly(ε-caprolactone) (PLC) and a polymer obtained by polymerization of an industrial waste (PLX) for the preparing of the carrier particles. This preparation was performed according to the nanoprecipitation method [2]. The encapsulation rate (%EE) of the nutrients nitrogen, phosphorus and potassium, the main commercial nutrients studied in this work, was evaluated by measuring them in the particles. Phosphorus quantitation was performed by UV-Vis spectrophotometry; nitrogen was determined by the Kjeldahl method; and potassium by flame atomic absorption spectroscopy. To optimize the particle synthesis, multivariate experimental designs for screening and response surface construction were adopted. In order to propose the work operating ranges for setting the levels and variables which were studied in the screening phase, it was conducted preliminary tests in the laboratory. Once defined, it was performed a fractional factorial design 2^{5-1} with triplicate at the central point [3] for the screening, which identified the variables that really have significant effect on the %EE. The studied variables were: PLC mass, PLX mass, volume of chloroform, concentration of the polyvinyl alcohol solution (PVA) and volume of water. Spreadsheets [3] were used for coding the experiments to be performed taking into consideration the relationship among the proposed levels. By means of the parameter *p* at the level of significance of 0.05, it was found that only the water volume is not a significant variable for the system. The central composite design (CCD) [3] (in triplicate at the central point) was used in this study for the optimization of %EE. All significant variables in screening were studied in the response surface methodology. Subsequently, by means of the Statistica® software, it was used the desirability function for obtaining simultaneously the optimal synthesis condition. Thus, the best conditions for preparation of the microparticles was: 100.00 mg of PLC, 825.00 mg of PLX, 9.25 mL of chloroform and 0.9% w/v of PVA, which yielded an average of encapsulation rate of 94.23% for nitrogen, 99.80% for phosphorus and 65.00% for potassium. The %EE values showed to be satisfactory, since in the literature does not exist this evaluation for the active substances proposed here [2]. With the use of multivariate experimental designs and the desirability function it was possible to obtain a significant simultaneous encapsulation rate, whereas the %EE of systems, prior to the multivariate approaches, were 60.00%; 66.66% and 12.11% for nitrogen, phosphorus and potassium, respectively.

Acknowledgement: CAPES, CNPq, FAPEMIG, PROPP/UFOP, RQ-MG UFOP.

References:
[1] Manolikar, M. K.; Sawant, M. R., 51, 811 (2003).
[2] Franchetti, S. M.; Marconato, J. C., 29, 811 (2006).
[3] Teófilo, R. F., Ferreira, M. M. C., 29, 338 (2006).

P-056

DEVELOPMENT OF METHODOLOGY FOR THE DIGESTION OF METALLIC CATION CARRIERS POLYMERIC MICROPARTICLES BY MEANS OF MULTIVARIATE EXPERIMENTAL DESIGNS

J.O. Chagas, I.C.M. Cunha, G.A. da Silva, F.A. Lobo

Departamento de Química, Universidade Federal de Ouro Preto (UFOP), MG, Brazil.
fabiana@iceb.ufop.br

Alternatives that are been highlighted to mitigate some environmental problems are the controlled release systems. To evaluate the efficiency of these systems are adopted analytical techniques that require digestion procedures of these samples [1]. However, there are not reports in the literature of digestion processes of micropolimeric controlled release systems carriers of metallic cations, for plant fertilization purposes. Thus, with this work, multivariate experimental designs were used to ensure an appropriate sample preparation, with the aim to obtain an optimum condition of digestion for further analysis by atomic absorption spectroscopy. For the digestion process was used a heater and shaker plate and for evaluation of the polimeric degradation it was used an analyzer of total organic carbon (TOC). For the screening process, it was performed a full factorial design 2^4 with triplicate at the central point [2], having been studied the variables: sample volume, volume of acid, temperature and digestion time. It was used spreadsheets [2] to generate the encoded experiments to be performed taking into account the combinations among the proposed levels. Of all the studied variables, only the sample volume had a significant effect on the evaluation system of total organic carbon and total carbon, reducing the carbon content of the samples with the lowest level studied. Evaluating the normal distribution of the total carbon effects it was observed that the variables temperature and digestion time did not distribute normally in relation to the other variables and their interactions. The behavior was similar for the total organic carbon effects with the addition of a third-order effect not normally distributed: volume of sample × volume of acid × temperature. The results for the inorganic carbon measurements were negligible. Due to these behaviors, it was decided to study all variables evalutated in screening phase in the response surface analysis. The central composite design (CCD) [2] was used in this study for the optimization phase of the microparticles digestion process. The results of the response surface methodology for total organic carbon concentration confirmed the significance of sample volume, which yielded a decrease in the carbon content of the samples at the lowest studied level. It was also observed that the system has quadratic behavior, since the interaction of the acid volume with itself was significant to the system. Through the TOC analysis results it was possible to obtain a quadratic model by multiple linear regression at a significance level of 0.05, without lack of fit. The investigated variables and their levels that provided the best particle digestion condition were: 5.00 mL of sample, 10.00 mL of nitric acid, 60 ºC and 90 min. With the use of multivariate experimental designs was possible to obtain a significant polimeric degradation, as the initial carbon concentration was 4.60 mg/L and became 0.55 mg/L.

Acknowledgement: CAPES, CNPq, FAPEMIG, PROPP/UFOP, RQ-MG, UFOP.

References:

[1] Franchetti, S. M.; Marconato, J. C., 29, 811 (2006).
[2] Teófilo, R. F., Ferreira, M. M. C., 29, 338 (2006).

P-057

PRACTICAL IMPLEMENTATION OF MULTIVARIATE STATISTICAL PROCESS CONTROL FOR MONITORING OF THE WATER INTAKE PROCESS AT THE BARCELONA DWTP

S. Platikanov[1,2], D. Baquero[2], S. González[2], J.L. Cortina[2], J. Martin[3], R. Tauler[1]

[1]Department of Environmental Chemistry, IDAEA-CSIC, Jordi Girona, 18-26, 08026 Barcelona, Spain
[2] CETAQUA Water Technological Center, Cornellà de Llobregat, Spain
[3]Aigües de Barcelona (AGBAR) Laboratory, Barcelona, Spain
stefan.platikanov@idaea.csic.es

The Llobregat River is the main surface water source for more than 3 million habitants of Barcelona. Permanent control of the raw water quality is performed at the entrance of the Sant Joan Despi Drinking Water Treatment Plant (SJD DWTP). Several water quality parameters, such as conductivity, ammonium-N, TOC, pH, temperature, UV254nm and turbidity are permanently monitored every hour. As result, a huge volume of information has been stored in the recent years and allowed for multivariate analysis.

The goal of our study was to switch the established actual water quality control process using monitoring of parameters one by one (univariate statistics) to global analysis, based on the simultaneous analysis of multiple parameters (multivariate statistics). Reduction of time and effort is expected. Multivariate analysis was applied on data about the incoming raw water quality of the Llobregat River during the period of 2013-2014.

Principal Component Analysis of the recorded historical data allowed for analysis of the hydrological regime and the investigation of accidental industrial and weather events in the water quality of the Llobregat River. As result, the variability of water quality in the intake was determined due to three factors: seasonal changes of the organic matter concentrations; seasonal changes in inorganic content; and temporal changes of pH. Moreover, it was detected that the first two factors were responsible of a great part of the short term events, especially when rainfall episodes produced an unusual variability in water quality.

Current thresholds for individually controlled parameters in the raw river water were used as criteria to expand the study to the application of the theory and practices of Multivariate Statistical Process Control (MSPC) using PCA. Two models with different operational criteria were developed - one defining the ALARM state of the process and another one, defining the ABORT state of the process. In external validation, low percentages of false positive (6% for the ALARM state and 5% for the ABORT state) and of false negative water intake observations (2% for the ALARM state and 0% for the ABORT state) were reported.

Acknowledgements: This work has been financially supported by R+ I Alliance (http://www.ri-alliance.com/site/RIA/home). The authors acknowledge the staff of the Laboratory of Aigües de Barcelona for the help provided with analysis and information management.

P-058

CALIBRATION TRANSFER IN PORTABLE NIR INSTRUMENTS FOR QUALITY CONTROL OF POLYMORPHS IN PHARMACEUTICAL RAW MATERIALS

Vitor Silva[1], Jailson Silva[2], Claudete Pereira[1]

[1]Department of Fundamental Chemistry, Federal Univesity of Pernambuco, Recife, Brazil.
[2]Department of Chemical Engineering, Federal Univesity of Pernambuco, Recife, Brazil.
Email: vitorhugo.quimica@gmail.com

Near infrared (NIR) spectroscopy to quantify pharmaceutical polymorphs using multivariate calibration methods is well established in the literature, usually employing benchtop instruments [1]. Nowadays, the miniaturization of with NIR instruments allows new ways to use these devices in pharmaceutical industries for the quality control of incoming materials with Active Pharmaceutical Ingredients (APIs) [2]. These instruments can provide a simple, fast and easy means for quality control in pharmaceutical networks. This values the calibration transfer of models developed in benchtop instruments, offering high spectral resolution, signal-to-noise ratio and better wavelength reproducibility. Then, this work evaluated the transfer of multivariate models between one benchtop and two portable instruments to determine mebendazole (MBZ) polymorphs (forms A, B and C) in raw materials. The primary (P) instrument was a Frontier FT-NIR (PerkinElmer) spectrophotometer and the secondaries were (a) MicroNIR 1700 spectrophotometer, JDSU (S1) and (b) NIRscan Nano spectrophotometer (S2), Texas Instruments. The spectral range from 1040 nm to 1590 nm was used in all instruments. The sample set was composed of thirty ternary mixtures of MBZ polymorphs varying from 0 to 100% (w/w) to polymorphs A and C, and from 0 to 30% (w/w) to polymorph B. Partial Least Squares (PLS) regression models were built for each polymorph using preprocessed spectra. The SPXY (sample set partitioning based on joint x–y distances) algorithm [3] was applied to divide the sample set into calibration (70%) and validation (30%) subsets. The predictive ability of the models was evaluated by RMSEP and R^2. For calibration transfer, the Direct Standardization (DS) method was used, with five to ten calibration transfer samples, were chosen by Kennard-Stone (KS) algorithm. For P instrument, the PLS models show RMSEP of 2.7% w/w, 1.5% w/w and 2.0% w/w for polymorphs A (Standard Normal Variate), B (derivative) and C (Multiplicative Spectra Correction), respectively. Using the DS transfer method, the RMSEPs obtained for instrument S1 were 3.8% w/w, 1.8% w/w and 3.3% w/w for polymorphs A, B and C, respectively. For instrument S2, the RMSEPs obtained were 6.0% w/w, 4.4% w/w and 4.9% w/w for polymorphs A, B and C, respectively. These results show that DS method is effective for calibration transfer between P and S1 instruments because the RMSEPs obtained were statistically similar at a confidence level of 95%, according to *F*-test. On the other hand, worse results were obtained between the P and S2 instruments (RMSEPs statistically different). A complete recalibration procedure for secondary instruments, which may not be convenient in practice, was adopted for comparison purpose only. With the exception of polymorph C for instrument S2, all the results were statistically similar to those obtained for the P instrument. This work demonstrates that portable NIR instruments and calibration transfer are suitable tools for quality control of incoming materials in industries.

Acknowledgments: FACEPE/NUQAAPE, CNPq/FAPESP/INCTAA and LAC/UFPE for financial support.

References:

[1] Silva, VH *et al*. Journal of Pharmaceutical and Biomedical Analysis. 115, 587-593 (2015)
[2] Alcalà, M. *et al*. Journal of Near Infrared Spectroscopy. 21, 445-457 (2013)
[3] Galvão RKH, *et al*. Talanta 67, 736 – 740 (2005).

P-059

CORRELATION OF ATR-FTIR TO EDS TO DETERMINE THE PRECISION OF A NOVEL FILM-THICKNESS PAT FOR MEDICAL DEVICES

T. Hou [1], B.A. Weinstock [1]

[1]Teleflex Inc., 1 Kendall Square, B14101, Cambridge, MA, USA
andre.weinstock@teleflex.com

Hydrophilic zwitterions grafted onto a polymer surface offer great potential for endowing medical devices with non-fouling and non-thrombotic properties [1,2]. However, several challenges need to be met in order to mass-produce and affordably market such technology. One challenge is to design a process analytical technology (PAT) capable of rapidly and non-invasively qualifying the presence of a thin grafted layer (100 – 1000nm) on the surfaces of a macroscopic device, such as a catheter, in a manufacturing setting. Here we describe a proto-chemometric PAT technique [3] that estimates "film" thickness using ATR-FTIR (Attenuated Total Reflectance – Fourier Transform Infrared Spectrometry) and statistical fitting to synthetic spectra constructed using Harrick's equations. Lacking a chromatographic method as a primary metric, we use EDS (Energy Disperse Spectrometry) and show a good correlation between modelled estimates of thicknesses.

Acknowledgement: SEM/EDS was performed in collaboration with the Harvard University Center for Nanoscale Science in Cambridge, MA, USA via their ClubNano for industrial and academic collaborations

References:

[1] Z. Zhang, M. Zhang, S. Chen, T Horbett, B. Ratner, S. Jiang. "Blood compatibility of surfaces with super low protein adsorption". Biomat. 2008. 29; 4285-4291
[2] R.S. Smith, Z. Zhang, M. Bouchard., J. Li, H.S. Lapp, G.R. Brotske, D.L. Lucchino, D. Weaver, L.A. Roth, A. Coury, J. Biggerstaff, S. Sukavaneshvar, R. Langer, C. Loose. "Vascular catheters with a non-leaching sulfobetaine surface modification have significantly reduced thrombus formation and microbial attachment". Science Trans Med. 2012, 4; 15301-15310
[3] B.A. Weinstock, L.M. Guiney, C. Loose. "Rapid, nondestructive estimation of surface polymer layer thickness using attenuated total reflection Fourier transform infrared (ATR FT-IR) spectroscopy and synthetic spectra derived from optical prinicples". 2012, 11; 1311-1319

P-060

IN-DEPTH UNDERSTANDING OF AN API SYNTHESIS: TECHNOLOGIES AND CHEMOMETRICS TO SUPPORT CONTINUOUS IMPROVEMENT AND INNOVATION

Francisca F Gouveia[1,3], Jesper P Rahbek[2], Asmus R Mortensen[2], Mette T Pedersen[2], Pedro G Felizardo[1], Rasmus Bro[3]

[1]4Tune Engineering Ltd, Lisbon, Portugal
[2]Chemical Prod. Dev., Chemical Production, Denmark, H. Lundbeck A/S , Nykoebing Sj., Denmark
[3] Department of Food Science, Faculty of Science, University of Copenhagen, Denmark
email (ff@4tuneengineering.com)

A comprehensive regulatory framework endorsing the use of Quality by Design in pharmaceutical manufacturing is now in place [1-2]. These documents promote a science-based approach supported by prior knowledge and enhanced process understanding obtained through Process Analytical Technologies (PAT). To fully realize the QbD vision, PAT tools need to be used in-situ, enabling process state estimation and enhanced understanding of the manufacturing requirements [3]. In the present study, in-line IR spectra collected from a complex, multi-phase reaction system were combined with chemometrics to enhance the understanding of the reaction mechanism. Different modeling strategies were applied such as, multivariate projection methods, partial least squares regression and multivariate curve resolution to (1) describe the stoichiometries between reactants and products, (2) develop a real-time monitoring system able to detect variations derived from process inputs manipulation and (3) identify improvement opportunities in the current manufacturing process.
A systematic procedure for exploiting the information provided by IR spectroscopy is highlighted, demonstrating how these [PAT] tools can support continuous improvement and innovation of commercial processes.

Acknowledgement: The authors would like to acknowledge all members of the Process Optimization Laboratory of H. Lundbeck A/S.

References:

[1] Pharmaceutical cGMPs for the 21st Century – a Risk-Based Approach, FDA, 2002
[2] FDA, Guidance for Industry: PAT—a Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance, 2004
[3] Menezes J., Gouveia, F., Felizardo, P (2014). Science-based data-driven technologies in Pharma: development and manufacturing with QbD 2.0. European Pharmaceutical Review 19(3) 50-55.

P-061

ORTHOGONALISATION METHOD FOR ROBUSTNESS IMPROVEMENT OF INLINE NIR APPLICATIONS

Lallemand J.[1], Guilment J.[2], Dubuc P[2]. , Montagnier S.[2] et Roussel S. [1]

[1] Ondalys, 4 rue Georges Besse, 34830 Clapiers, France

[2] ARKEMA - CERDATO / Laboratoire d'Étude des Matériaux (LEM) Route du Rilsan, 27470 Serquigny – France

jlallemand@ondalys.fr

Online model maintenance is a main problem for developing NIRS applications. Perturbations appearance due to environmental changes, maintenance operation or aging of the instrument, often affect model performances. Model correction with classical methods such as bias and slope correction or model redevelopment are not always satisfactory strategies.
The use of an orthogonalisation method can be an effective way to solve this problem and it is illustrated in this study with an industrial application.

Monitoring of polyamide polymerization by NIRS is a well-known subject which gives excellent results. The measurements can be made at-line on powders or granulates, but can also be performed on line on powders or in molten medium. A PLS model allows to access directly the end of polymer chains, or less indirectly at the viscosity of the product. In this study, the viscosity prediction by PLS allows real time monitoring of the process. However, after several years of operation, an unidentified perturbation appeared, leading to the failure of the PLS model during several months.

This industrial application is an ideal case for applying Dynamic Orthogonal Projection [1] (DOP). The purpose of this chemometric method is to make the model independent from perturbations.
The principle is to rebuild spectra as if they were measured without the perturbation. Only a small number of samples are needed to model the perturbation space. This is done by PCA, based on spectral differences between real spectra and reconstructed spectra. The calibration database is then projected orthogonally from this space and the model is rebuilt. The corrected model becomes independent of the presence or not of the perturbation and new spectra do not require any orthogonalisation processing before applying the model.

DOP has been applied with success to correct the PLS model of viscosity prediction with few samples, whereas model redevelopment was not entirely satisfactory. Furthermore, the study of the spectral zone affected by the perturbation and corrected by DOP, has allowed to come back to the process to identify what went wrong and then act directly on the process.

Maintenance and robustness problems of predictive models in NIRS are a real restraint for its expansion in the industrial world. DOP is an elegant mathematical solution which allow to overcome the impact of appearance and disappearance of perturbations with only few samples.

References

[1] M. Zeaiter, J.M Roger, V.Bellon-Maurel, « Dynamic Orthogonal Projection. A new method to maintain the on-line robustness of multivariate calibrations. Application to NIR based monitoring of wine fermentations ». Chemometrics and Intelligent Laboratory Systems, Vol.80, issue 2, pp. 227–235, 2004.

P-062

KERNEL PARTIAL LEAST SQUARES REGRESSION COUPLED TO PSEUDO-SAMPLE PROJECTION FOR THE ANALYSIS OF MIXTURE DESIGNS OF EXPERIMENTS

H.H.M. Kerkenaar [1], R. Vitale [2], D. Palací-López [2], G.J. Postma [1], L.M.C. Buydens [1], A. Ferrer [2]

[1]Radboud University Nijmegen, Institute for Molecules and Materials, Analytical Chemistry, P.O. Box 9010, 6500 GL, Nijmegen, The Netherlands
[2]Departamento de Estadística e Investigación Operativa Aplicadas y Calidad, Universitat Politècnica de València, Camino de Vera s/n, 46022, Valencia, Spain
aferrer@eio.upv.es

Mixture experiments are those in which the proportions x_i of the I different components of a blend are at least of as much relevance as their absolute quantities, and the sum of these proportions must be a fixed value (usually one, or 100%).

Due to such restriction, when coping with designs of mixture experiments, using classical polynomial fitting by traditional methods like Ordinary or Generalized Least Squares (OLS/GLS) is unfeasible. Therefore, alternative approaches, namely the Scheffé models and their re-parametrisation, the Cox models, are needed in these circumstances. However, they both show several limitations: the former lack a constant term, cannot handle possible additional constraints, and their coefficients are non-intuitively interpretable. The latter always requires a specific component blend to be set as reference, and their parameters cannot be directly estimated by OLS/GLS [1]. In order to solve such issues, Partial Least Squares (PLS) regression-based techniques can be resorted to: they have proved to guarantee satisfactory performance even when highly restricted mixture spaces have been dealt with and allow variables of different nature (e.g. component proportions and physicochemical properties as well as manufacturing process conditions) to be fused and simultaneously analysed [1,2].

Nevertheless, if the mixture data under study are affected by strong non-linear relationships (which is rather common in e.g. industrial scenarios), applying classical PLS (even taking into account additional interaction and/or higher-degree terms) may not represent an appropriate modelling strategy. Kernel-Partial Least Squares (K-PLS) regression could constitute a valid alternative in these cases [3]. Unfortunately, kernel-based methodologies suffer from a specific drawback: the information about the importance of the original variables cannot be retrieved by simply plotting the loadings or the weights of the resulting models. To then enable their interpretation, Gower's idea of non-linear biplots and pseudo-sample projection can be exploited [4,5,6]. The main aim of this work is to extend the utilisation of K-PLS coupled to pseudo-sample projection for the analysis of mixture designs of experiments. The feasibility of their combination and its high potential will be demonstrated via simulated and real examples.

Acknowledgement: This research work was partially supported by the Spanish Ministry of Economy and Competitiveness under the projects DPI2011-28112-C04-02 and DPI2014-55276-C5-1R and Shell Global Solutions International B.V. (Amsterdam, The Netherlands).

References:

[1] N. Kettaneh-Wold, Chemometr. Intell. Lab., 14, 57 (1992).
[2] L. Eriksson, E. Johansson, C. Wikström, Chemometr. Intell. Lab., 43, 1 (1998).
[3] B. Schölkopf, A. Smola, Learning with Kernels, 1st Edition, MIT Press, Cambridge, UK, 2002.
[4] J. Gower, S. Hardings, Biometrika, 75, 445 (1988).
[5] G.J. Postma, P.W.T. Krooshof, L.M.C. Buydens, Anal. Chim. Acta, 705, 123 (2011).
[6] R. Vitale, O.E. de Noord, A. Ferrer, J. Chemometr., 28, 697 (2014).

P-064

CERTIFICATION OF A NEW REFERENCE MATERIAL : A UOX GLASS

M. Crozet [1], C. Rivier [1], C. Roche [2], T. Cozzika [2], D. Roudil [1]

[1] CEA-DEN, DRCP - Marcoule
[2] CEA, DEN, DTCD - Marcoule
BP 17171, 30 207 Bagnols-sur-Cèze, France
marielle.crozet@cea.fr

In 2013, on request of the Joint Vitrification Laboratory (LCV) between CEA and AREVA, through actions for improvement of analytical methods on the determination of the chemical composition of glass, CETAMA (Commission d'ETAblissement des Méthodes d'Analyses) certified the mass concentration of 24 oxydes, components of a simulated UOx glass standard. An interlaboratory comparison test was also organized to evaluate performances of the main analytical methods used by laboratories for glass analysis: X-ray Fluorescence (XRF), Inductively Coupled Plasma - Mass Spectrometry (ICP-MS) or Inductively Coupled Plasma - Atomic Emission Spectrometry (ICP-AES). This reference glass is the first certified glass with 24 oxides

Matrix reference materials are dedicated to the analytical control of trueness and precision of analytical methods and are highly useful for the validation of these methods.

The manufacturing and packaging of 140 bottles of 20 grams of crushed glass were made by the LCV in 2014 and analytical control of the homogeneity of the material was entrusted to AREVA Marcoule laboratory. In 2015, in collaboration with LCV teams, CETAMA organized an interlaboratory comparison test involving seven laboratories working in differents fields of materials characterization.

Certification of mass concentration values, according to the ISO 35 guide [1], is based on the chemical formulation of the glass and on weight values of each of the oxyde precursors, except for ruthenium and palladium, for which mass concentration is certified by an expert laboratory. Assessed uncertainty associated with certified values takes into account glass charaterization uncertainty, heterogeneity and stability impacts on trueness and precision.

Determination of the performance characteristics of methods is based on ISO 13528 [2] and ISO 5725-5 [3] standards considering as the assigned values the previously certified reference values. Analysis of the results of the interlaboratory comparison test allow to quantify and compare the performances of different analytical methods of oxides in a glass.

Editing a certificate of this new CRM (Certified Reference Material) in CETAMA catalog will finalize this work done for 2 years in close collaboration between the teams at CEA Marcoule.

Acknowledgement: The authors thank AREVA for financial support and everyone who helped to carry out this project and particularly Sylvain Mure and Jean-Christophe Laugier.

References:

[1] Guide ISO REMCO WG16 N84, 2015-03-01, 2nd CD GUIDE 35, Reference Materials – general guidance for the assignment of property values.
[2] Norme NF ISO 13528, décembre 2005, Méthodes statistiques utilisées dans les essais d'aptitude par comparaisons interlaboratoires
[3] 15. Norme NF ISO 5725-5, octobre 1998, Exactitude (justesse et fidélité) des résultats et méthodes de mesure – Partie 5 : Méthodes alternatives pour la détermination de la fidélité d'une méthode de mesure normalisée

P-065

APPLYING DISTANCE REPRESENTATION WITHIN THE LIKELIHOOD RATIO FRAMEWORK FOR REPORTING THE EVIDENTIAL VALUE OF INFRARED AND RAMAN SPECTRA

A. Martyna[1,2], G. Zadora[1,3], T. Neocleous[4], A. Michalska[1], N. Dean[4]

[1] Institute of Forensic Research in Krakow, 9 Westerplatte, Krakow, Poland,
[2] Faculty of Chemistry, Jagiellonian University in Krakow, 3 Ingardena, Krakow, Poland
[3] Institute of Chemistry, University of Silesia in Katowice, 9 Szkolna, Katowice, Poland
[4] School of Mathematics and Statistics, University of Glasgow, 15 University Gardens, Glasgow, UK
gzadora@ies.krakow.pl

Many scientific fields consider spectroscopic data as the subject of comparative analysis. This is also the case in the forensic sciences, where spectroscopy is employed for characterising, for example, plastics used for car body element production (e.g. bumpers, headlamp lenses) as well as blue automotive paints collected from the scenes of hit-and-run car accidents. Fourier transform infrared spectrometry (FTIR) is applied for characterising polymers, while Raman spectroscopy (RS) is utilised for pigment identification in car paints.

For making inference about the connections between the scene of a car accident and the suspected car, the spectra of the material collected from the car accident scenario (so-called recovered samples, whose source is unknown) are compared with the spectra of the known-source control material collected e.g. from the suspected car.

In the forensic context, analytical results must be interpreted and reported according to the standards of the interpretation schemes acknowledged in forensic sciences using the likelihood ratio (LR) approach [1]. However, for proper construction of LR models for highly multivariate data, such as spectra, chemometric tools must be employed for substantial data dimension reduction. Using a sequence of chemometric techniques addressing various aspects of hidden data structure and adopting their outcome as the input for LR models was the objective of this research.

The research presented herein was aimed at verifying the suitability of combining chemometric tools for generating lower dimensional data without ignoring relevant data features with constructing LR models within the comparison problem of FTIR spectra obtained for 30 polypropylene samples and Raman spectra of solid and metallic car paints (30 samples of each). Details of analytical conditions can be found in [2, 3]. Conversion from classical feature representation to distance representation was proposed for revealing hidden data features/peculiarities, and linear discriminant analysis (LDA) was further applied for minimising the within-sample variability while maximising the between-sample variability. Both techniques enabled massive reduction of data dimensionality.

Next, univariate and multivariate likelihood ratio models were proposed for data obtained from LDA. They were evaluated by estimating the rates of false positive (FP) and false negative (FN) answers, and additionally by using the Empirical Cross Entropy (ECE) approach [e.g. 1].

Low levels of FP and FN as well as acceptable ECE plots proved that the combination of chemometric tools and the likelihood ratio approach is adept at solving the comparison problem of highly multivariate and correlated data after proper extraction of the most relevant features and variance information hidden in the data structure.

Acknowledgement: The experiments were undertaken within the National Science Centre in Poland project (Preludium 6 no. 2013/11/N/ST4/01547) and the Institute of Forensic Research project IV K/2015-17.

References:

[1] G. Zadora, A. Martyna, D. Ramos, C. Aitken, Chichester 2014: John Wiley & Sons (ISBN 978-0470972106)
[2] A. Martyna, A. Michalska, G. Zadora, Anal Bioanal Chem, 407, 3357 (2015).
[3] A. Michalska, A. Martyna, J. Zięba-Palus, G. Zadora, J Raman Spec, 46, 735 (2015).

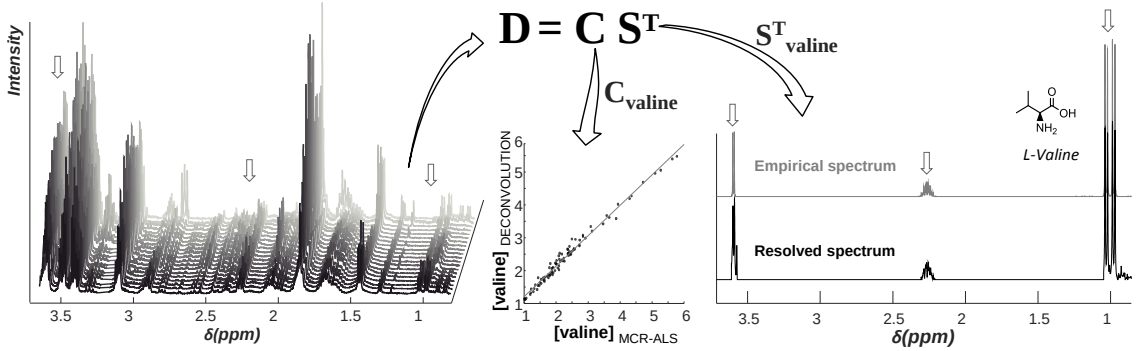
P-066

CHEMOMETRIC STRATEGY FOR UNTARGETED ^{1}H NMR METABOLIC PROFILING OF YEAST EXTRACTS USING MCR-ALS

F. Puig-Castellví [1], I. Alfonso [2], B. Piña [1], R. Tauler [1]

[1] Department of Environmental Chemistry, Institute of Environmental Assessment and Water Research (IDAEA-CSIC), Jordi Girona 18-26, 08034 Barcelona, Spain
[2] Department of Biological Chemistry and Molecular Modelling, Department of Environmental Chemistry, Institute of Advanced Chemistry of Catalonia (IQAC-CSIC), Jordi Girona 18-26, 08034 Barcelona, Spain
francesc.puig@idaea.csic.es

Metabolomics is a field of 'omics' research that is primarily focused on the identification and characterization of small molecule metabolites in cells, tissues, organs and organisms. Commonly, Mass Spectrometry (MS) and Nuclear Magnetic Resonance (NMR) are used on these studies.
Despite the fact that there are alternatives to automatize the NMR analyses[1], they are restricted to identify profile matches on their spectral libraries.
Other statistic approaches, such as STOCSY[2], may aid to stablish correlations among intramolecular protons of unknown compounds, thus leading to the elucidation of its chemical structure. However, in a ^{1}H NMR matrix containing many different spectra, STOCSY does not perform efficiently with highly overlapped signals and it does not provide multiplicity information of these signals.
In this work, an approach based on Multivariate Curve Resolution – Alternating Least Squares (MCR-ALS)[3] of restricted windows on the spectral dimension has been used to obtain the pure concentration and spectral profiles of the metabolites present in yeast extracts from a time-course experiment.
Results from this approach revealed not only that this untargeted approach allows to recover the ^{1}H NMR profile (matrix S), but also that the obtained concentration values (matrix C) are comparable with those obtained using traditional spectral deconvolution methods[4].



Acknowledgement:
The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n. 320737.

References:
[1] S. Ravanbakhsh, P. Liu, et al., PLoS One, 10, DOI:10.1371/journal.pone.0124219 (2015).
[2] O. Cloarec, M.E. Dumas, et al., Anal. Chem., 77, 1282 (2005).
[3] J. Jaumot, A. De Juan, R. Tauler, Chemometrics Intell. Lab. Syst., 140, 1 (2015).
[4] J. Hao, M. Liebeke, et al., Nat. Protoc., 9, 1416 (2014).

P-067

LIPIDOMIC PROFILES OF BRAIN, GONADS AND INTESTINAL TRACTS OF ZEBRAFISH (*Danio rerio*) EXPOSED TO CARBON-BASED NANOPARTICLES BY LC-MS AND CHEMOMETRIC ANALYSIS

E. Gorrochategui[1], J. Li[2], N.J. Fullwood[3], G.G. Ying[4], M. Tian[5], L. Cui[5], H. Shen[5], S. Lacorte[1], F.L. Martin[2] and R. Tauler[1]

[1]Department of Environmental Chemistry, Institute of Environmental Assessment and Water Research (IDAEA), Consejo Superior de Investigaciones Científicas (CSIC), Barcelona, 08034, Catalonia, Spain; [2]Centre for Biophotonics, Lancaster Environment Centre, Lancaster University, Lancaster LA1 4YQ, UK; [3]Biomedical and Life Sciences Division, Lancaster Environment Centre, Lancaster University, Lancaster LA1 4YQ, UK; [4]State Key Laboratory of Organic Geochemistry, Guangzhou Institute of Geochemistry, Chinese Academy of Sciences, Guangzhou 510640, China; [5]Key Lab of Urban Environment and Health, Institute of Urban Environment, Chinese Academy of Sciences, Xiamen 361021, China

With rising environmental levels of carbon-based nanoparticles (CBNs) there is an urgent need to develop an understanding of their biological effects in order to generate appropriate risk assessment strategies [1]. Given the multiplicity of the effects that they might pose on organisms, "omic" procedures might be a better approach towards determining endpoint alterations, which can be associated to metabolite or lipid disruption. The study of the latter effects on lipids is of crucial interest, since lipids play essential roles in energy production and storage, and cell membrane development, and thus, their study can aid in understanding the pathogenesis of many disease states. Liquid chromatography coupled to mass spectrometry (LC-MS) based methods have the ability for the analysis of low molecular weight compounds in biological systems, such as complex lipid mixtures or biomolecules [2]. However, the large amounts of data generated with this technique require extensive processing to appropriately assess sample classification/discrimination and biomarker detection [3]. In this study we exposed zebrafish *via* their diet to one of four different CBNs; C_{60} fullerene (C_{60}), single-walled carbon nanotubes (SWCNT), short multiwalled carbon nanotubes (MWCNTs) or long MWCNTs. Lipid alterations were studied in three target tissues (brain, gonads and gastro-intestinal tracts) of male and female zebrafish by LC-MS followed by chemometric analysis. The employed chemometric methodology covered different stages of data analysis including data conversion and import, data compression, data normalization, peak resolution and biomarker detection and identification. In this study, data compression was performed by searching the regions of interest (ROI) [4] and peak resolution was possible using multivariate curve resolution-alternating least squares (MCR-ALS) [5]. Overall, this study aimed to demonstrate the potential of LC-MS followed by chemometric analysis to study unkown effects of environmental stressors in exposed organisms.

Acknowledgement: The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n. 320737. Presenting author acknowledges the Spanish Government (Ministerio de Educación, Cultura y Deporte) for a predoctoral FPU scholarship (FPU13/04384).

References:

[1] J.G. Bundy, M.P. Davey, M.R. Viant, Environmental metabolomics: a critical review and future perspectives, Metabolomics. 5 (2008) 3–21. doi:10.1007/s11306-008-0152-0.

[2] K. Dettmer, P.A. Aronov, B.D. Hammock, Mass spectrometry-based metabolomics., Mass Spectrom. Rev. 26 (2007) 51–78. doi:10.1002/mas.20108.

[3] J. Boccard, S. Rudaz, Harnessing the complexity of metabolomic data with chemometrics, J. Chemom. 28 (2014) 1–9. doi:10.1002/cem.2567.

[4] E. Gorrochategui, J. Jaumot, R. Tauler, A protocol for LC-MS metabolomic data processing using chemometric tools, Protoc. Exch. (2015). doi:10.1038/protex.2015.102.

[5] R. Tauler, Multivariate curve resolution applied to second order data, Chemom. Intell. Lab. Syst. 30 (1995) 133–146. doi:10.1016/0169-7439(95)00047-X.

P-068

MULTIBLOCK CE-MS AND LC-MS METABOLIC PROFILING DATA ANALYSIS

Elena Ortiz-Villanueva, Joaquim Jaumot, Romà Tauler

Department of Environmental Chemistry, IDAEA-CSIC, Barcelona, Spain.
elena.ortiz@idaea.csic.es

In the framework of *omic* sciences, there is a current trend towards the application of chemometric methods to analyze the large amount of data generated in chemical and biological studies. In the last few years, advanced data analysis tools and novel approaches have been developed to enhance the acquisition of new knowledge and improve the understanding of biological processes. Use of multiblock based methods allows the simultaneous study of data sets from different instrumental platforms. For instance, the same samples can be investigated using CE-MS and LC-MS analytical platforms in an attempt to gather metabolic information with lower uncertainty and greater accuracy. However, data from different sources can be heterogeneous, and the extraction of common information is challenging [1, 2].

Data fusion strategies are promising tools for fundamental metabolomics studies and they provide improved biomarker detection and identification, thus allowing a better characterization of metabolic responses. Data fusion strategies can be implemented at different levels [3].

In this work, two different data fusion strategies are used for the simultaneous analysis of untargeted capillary electrophoresis-mass spectrometry (CE-MS) and liquid chromatography-mass spectrometry (LC-MS) data. A low-level data fusion strategy has been implemented merging CE-MS and LC-MS data (taking advantage of the common number of m/z values). The resulting augmented data matrix has been analyzed by means of MCR-ALS and most relevant common features were extracted. On the other hand, a high-level data fusion strategy has been performed by combining features previously obtained by independent MCR-ALS analysis of each individual data set, i.e. combining the elution profiles of the resolved components. The usefulness and the advantages of the two data fusion strategies are demonstrated in a comparative study of the metabolic profiles from baker's yeast (*Saccharomyces cerevisiae)* samples grown in two carbon sources, non-fermentable carbon source (acetate) and fermentable carbon source (glucose).

Acknowledgement: This research has been funded by the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n. 320737.

References:

[1] A. K. Smilde, J. A. Westerhuis, S. de Jong, *J. Chemometrics* 17, 323-337 (2003).
[2] M. Shouteden, K. Van Deun, I. Van Mechelen, *Behav. Res.* 45, 822-833 (2013).
[3] T. G. Doeswijk, A. K. Smilde, J. A. Hageman, *Anal.Chim. Acta* 705, 41-47 (2011).

P-069

EXPERIMENTAL DESIGN FOR *DAPHNIA MAGNA* METABOLOMICS: AN UNTARGETED GC-MS STUDY

Elba Garreta, Bruno Campos, Carlos Barata, Silvia Lacorte, Romà Tauler

Department of Environmental Chemistry, IDAEA-CSIC, Barcelona, Spain.
email: elba.garreta@idaea.csic.es

Increasing worldwide contamination and global climate change are of current concern demanding novel information about the effects of environmental stressors on living organisms. Metabolomics tries to characterize the most relevant metabolites which have suffered changes in response to a chemical or physical pressure. *Daphnia magna* is a model organism widely used for toxicological assessment in aquatic media [1].

The aim of this work is to study the metabolic variation of *D. magna* individuals exposed to changes of three different abiotic factors linked to global climate change: salinity, temperature and oxygen levels. A two-level full factorial experiment with 3 factors and 5 replicates was designed (DOE) [2]. Information about main effects and interactions were obtained and a tentative identification of metabolites affected by the three previously mentioned physical agents was performed.

Polar metabolites of *D. magna* individuals were extracted, derivatized and analysed by GC-MS in full scan mode. MCR-ALS [3] data analysis enabled the discrimintaion among elution profiles of *Daphnia* metabolites and those from the large number of undesired derivatized compounds, and the identification of the more relevant metabolites, by comparison of their MS spectra with those from NIST 2014 database (http://www.nist.gov/srd/nist1a.cfm).

Peak areas of MCR-ALS resolved components were analyzed by ANOVA-simultaneous component analysis (ASCA) [4], to assess the effects of the different factors of the experimental design. Metabolite identification allowed the description and proposal of the altered metabolite pools caused by the studied effects in *Daphnia* metabolome.

Acknowledgement: This research has received funding from European Research Council under the European Union's Seventh Framework Program (FP/2007-2013) / ERC Grant Agreement n. 320737.

References:

[1] A. Wojtal-Frankiewicz. Aquatic Ecology 46:37-53, 2012.
[2] D. B. Hibbert. Journal of Chromatography B 910: 2-13, 2012.
[3] A. de Juan, J. Jaumot & R. Tauler. Analytical Methods, 6 (14), 4964-4976, 2014.
[4] A.K. Smilde J.J. Jansen, H. C. J. Hoefsloot, R. -J. A. N. Lamers, J. van der Greed & M. E. Timmerman. Bioinformatics, 21 (13), 3043-3048, 2005.

P-070

METHODOLOGY FOR THE COMPARISON OF COMPLEX MALDI-TOF MASS SPECTRA. A CASE OF STUDY: STABILITY OF INFLIXIMAB

R. Pérez-Robles[1], L. Cuadros-Rodríguez[1], S. Medina-Rodríguez[2], N. Navas[1]

[1] Department of Analytical Chemistry, Faculty of Science, University of Granada, C/ Fuente Nueva s/n,Granada, Spain.
[2]Department of Signal Theory, Telematics and Communications, CITIC-UGR, University of Granada, C/Rafael Gómez Montero n/2, Granada, Spain.
email: rprrpr@correo.ugr.es

Infliximab (IFX) is a monoclonal antibody broadly used in hospitals for the treatment of Crohn's Disease, Ulcerative Colitis, Rheumatoid Arthritis, etc. The changes that likely could take place in the chemical structure and in consequence, in its biological activity after their aperture and during their storage have not been sufficiently studied. This fact implies that large quantities are daily discarded in hospitals with the subsequent economic lost, since the drug is expensive.

To get knowledge about the stability of this biopharmaceutical drug, INF samples were prepared as used in hospital, i.e. reconstituted with water at concentration of 10 mg/ml. INF aliquots of these pharmaceutical preparations were stored at three temperatures: room temperature, refrigerated (4ºC) and frozen (-20º) throughout 7 days. Each control day (day 0, 1, 3, 4 and 7), the samples were analysed by submitting them to enzymatic digestion with trypsin and recording the mass spectra of the resulting digested solutions. Then, the peptide mass fingerprints (PMFs) were obtained by matrix assisted laser desorption and ionization (MALDI) time-of-flight mass spectrometry (TOF-MS).

Due to the complexity of these PMFs, and the lack of repeatability of the MALDI technique, visual analysis to track changes in the PMFs was not possible was unable. Therefore, a robust mathematical comparison methodology is proposed herein. The PMF was extracted in a data vector which denotes intensities *vs m/q*. Next, the development of an ad hoc MATLAB function which transforms the intensities of PMFs in binary data was carried out. The value "1" indicates the presence of a fragment to a certain *m/q*, and the "0" the absence of it. Once the data from the PMFs were converted to binary vectors, two exploratory multivariate data methods were applied, i.e. principal component analysis (PCA) and multivariate analysis of variance (MANOVA), in order to detect groupings among PMFs. Subsequently, different similarity indexes between the two vectors were calculated based on the: (i) cosine of the angle; (ii), normalized Euclidean distances, and (iii) determination coefficient.

In this communication, the found results will be shown. It is particularly noted that during the first storage day, significant changes took place in the PMFs for all the storage temperatures checked.

P-071

DISCRIMINATION AND QUANTIFICATION OF COCAINE AND ADULTERANTS IN SEIZED DRUG SAMPLES BY FTIR AND PLSR

T.S. Grobério[1], J.J. Zacca[2], E.D. Botelho[2], M. Talhavini[2], J.W.B. Braga[1].

[1]Institute of Chemistry, University of Brasília, Brasília, Brazil.
[2]National Institute of Criminalistics, Brazilian Federal Police, Brasília, Brazil.
jez@unb.br

Infrared spectroscopy (FTIR) and Partial Least Squares (PLS) have been applied for the development of a method to perform both quantitative and qualitative analysis of real cocaine samples seized by the Brazilian Police Federal (BPF). Currently, quantification of cocaine and determination of adulterants in seizures is performed using gas chromatography. However, this technique requires a relatively complex sample preparation and a high cost of analysis. In this context, thid work presents a simpler method for simultaneously discrimination between free base and cocaine hydrochloride, to determine the oxidation degree (OXI) and quantify cocaine (COC) and its major adulterants (caffeine (CAF), phenacetin (PHE), benzocaine (BEN), aminopyrine (AMI) and lidocaine (LIN)) in seized drugs. A total of 1085 samples were analyzed, in which 500 were selected for the calibration set and 585 for the validation set. This large dataset enable a wide characterization of cocaine samples seized in Brazil, which is representative of the illicit cocaine in the country. Based on the Hotteling T^2 and Q residuals, 7.0 % of the validation samples were excluded. These samples were analyzed in detail to verify the reasons for their exclusion. Over 60% of these samples were part of seizures made at least three years before the model development, suggesting that there must have been some changes in the chemical characteristics of these samples. Table 1 present the figures of merit for quantification of all analytes. The method was able to perfectly discriminate between cocaine hydrochloride and free base samples, to quantify cocaine content as well as to estimate the oxidation degree and the concentration of the main adulterants. The minimum dedectable concentration (MDC) for cocaine indicated that the method can be applyied for samples presenting concentrations above of 10.8 %, which represent most of the cocaine seized by the BFP. Furthermore, the enable the determination of the false positive and negative rates of detect the analytes, which presented small errors, excepting for lidocaine. The results indicated that the method can be helpful in terms of both time and cost reduction in routine analysis in forensic laboratories across the country, such as airports and police border posts.

Table1 Analytical figures of merit of the method.

Figures of Merit	OXI	COC	PHE	BEN	AMI	LID	CAF
RMSEP (%)	1.1	2.8	0.7	0.2	0.6	0.7	0.6
R^2_{val}	0.947	0.946	0.994	0.926	0.940	0.849	0.834
MDC (%)	3.9	10.8	3.5	1.8	2.9	1.1	2.6
Average Uncertainty, (95% confidence)	2.3	6.7	2.2	0.9	1.8	2.7	2.0
False positive rate (%)	--	0.0	0.8	0.0	0.4	7.9	0.6
False negative rate (%)	--	0.0	2.4	0.0	0.3	0.3	2.1

Acknowledgement: FINEP, CAPES/PROFORENSE, CNPq and FAPDF.

References:

[1] T.S. Grobério, J.J. Zacca, E.D. Botelho, M. Talhavini, J.W.B. Braga, Forensic Sci. Int., 257, 297 (2015).

P-072

CLASSIFICATION APPROACHES IN FORENSICS: DETECTING SEMEN STAINS ON FABRICS

C. S. Silva[1], M. F. Pimentel[2], J. M. Amigo[3], R. H. Saldanha[4], A. Batista[5], C. Pasquini[6]

[1]Department of Fundamenta1Chemistry, Federal University of Pernambuco, Av. Prof. Moraes Rego, 1235 - Cidade Universitária, Recife, Brazil; [2]Department of Chemical Engineering, Federal University of Pernambuco, Av. Prof. Moraes Rego, 1235 - Cidade Universitária, Recife, Brazil; [3] Department of Food Science, University of Copenhagen, Rolighedsvej 30 - Frederiksberg C, Copenhagen, Denmark; [4] Department of Federal Police, Superintendência Regional em Pernambuco, Av. Cais do Apolo, 321, Bairro do Recife - Recife, PE, Brazil;5 Department of Veterinary Medicine, Federal Rural University of Pernambuco, Rua Dom Manoel de Medeiros, s/n, Dois Irmãos, Recife, Brazil; [6]Chemistry Institute, Department of Analytical Chemistry, Universidade Estadual de Campinas, Cidade Universitária - Campinas, SP, Brazil
carolina.santossilva@ufpe.br

In 2014 Widner and coworkers published a paper in Chemical & Engineering News journal about the need for realiable analytical methods in Forensics [1]. The paper emphasized that many analytical methods used by forensic laboratories lack of scientific underpinnings, and therefore it is important to produce statistically reliable results to avoid missleading evidences. In addition, DNA tests have spawned a huge impact on the field of forensic science. With an incredible sensitivity and a high power of discrimination, DNA analysis has been a powerful tool for human identification and criminal investigations.The use of non destructive analytical methods to identify traces of body fluids is especially important when those residues may contain DNA information, such as semen, blood and saliva [2,3]. In this sense, Near Infrared-Hyperspectral Imaging (HSI-NIR) becomes a plausible analytical methodology. Therefore, the aim of this work is to compare two different approaches to distinguish not only the stain, but also its nature: Partial Least Squares – Discriminant Analysis (PLS-DA) as a multivariate classification methodology, and Multivariate Curve Resolution – Alternating Least Squares (MCR-ALS) as curve resolution method. Six different pieces of cotton fabrics (white, black, blue, red, green and yellow) were used as substrates to deposit semen and common substances that are known for being false positives in current foresic tests. Four different brands of lubricants (L1-L4) and 4 samples of human semen were acquired and placed on the fabrics creating a stain. The hyperspectral images were acquired with 50 mm lens with pixel size of 156x156 µm, in the wavelength range of 928-2524 nm. The spectral sampling per pixel was 6.3 nm and the spectral resolution of 10 nm. Savitzky-Golay (SG) 1st derivative (2nd order polynomial and window's width of 11) was applied to preprocess the spectra. The MCR-ALS model was built using augmented matrix, with references of semen, lubricants (on white fabric) and fabric. The MCR-ALS distribution maps using 4 components (one for the semen, two for the lubricants and one for the fabric) showed consistent predictions, not only for the white fabric, but also for the colored ones. Only pixels from one brand of lubricant (L3) were modelled as semen for the colored fabrics. This particular lubricant has the most different composition, having a substantial amount of amine compounds, and, therefore, being confounded with the aminoacids contained in the semen. The PLS-DA probability images showed similar results to MCR-ALS and the non error rate for the cross-validation (NER_{CV}) for the three modeled classes (semen, lubricant and fabric) were 0.99, 0.98 and 0.97, respectively. Therefore, HSI-NIR associated with MCR-ALS and PLS-DA shows potential to be applied in semen detection on different fabrics.

Acknowledgement: INCTAA, CAPES, CNPq, FACEPE

References:

[1] A. Widener, C. Drahl, C&EN, 92 (19), 10-14 (2014).
[2] K. Virkler, I. K. Lednev, For. Sci. Int., 188, 1–17 (2009).
[3] C. K. Muro, K. C. Doty, J. Bueno, L. Halámková, I. K. Lednev, Anal. Chem., 87, 306−327 (2015).

P-073

NMR-BASED METABOLOMICS COMBINING MULTIVARIATE CURVERESOLUTION FOR IDENTIFICATION OF BIOMARKERS OF MYOCARDIAL ISCHEMIA

Xin Zhang, Zhongfeng Li, Zhuoyong Zhang*

Department of Chemistry, Capital Normal University, Beijing 100048, China
email: xin.kevin.zhang@gmail.com

Recognition of myocardial ischemia is critical both for the diagnosis of coronary artery disease and the selection and evaluation of therapy [1]. The identification of biomarkers that could predict the existence of disease, but no such biomarker has been described so far [2]. Metabolomics has refreshed interest in metabolism across biology and medicine, particularly in the areas of biomarker discovery.
Nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS) constitute the leading profiling technologies in metabolomics. The main advantage of using NMR is due to its nondestructive nature, which requires minimal sample preparation [3].
In this work, blood samples were obtained from animal models of myocardial ischemia and their control group. A method for the linear decomposition of NMR-based metabolomics data implemented via multivariate curve resolution alternating least squares (MCR-ALS), which has been used elsewhere, is introduced. By using constraints, the NMR spectra information relating with myocardial ischemia were resolved.

Acknowledgement: This study was funded by National Instrumentation Program (2012YQ140005) and National Natural Science Foundation of China (21275101).

References:
[1] Bodi V., Sanchis J., Morales J M., et al., J. Am. Coll. Cardiol., 59(18): 1629 (2012).
[2] Deng L., Cheng K. K., Dong J., et al., Chemometr. Intell. Lab., 118: 51 (2012).
[3] Duarte I. F., Diaz S. O., Gil A. M., J. Pharmaceut. Biomed., 93: 17 (2014).

P-074

EVALUATION OF DIRECT AND INDIRECT BIOMARKERS FOR ETHANOL CONSUMPTION FOR FORENSIC PURPOSES - A LIKELIHOOD RATIO APPROACH TO IDENTIFY CHRONIC ALCOHOL MISUSERS

E. Alladio[1,2], A. Martyna[3,4], A. Salomone[2], V. Pirro[4], M. Vincenti[1,2], G. Zadora[5,6]

[1]Chemistry Department, University of Turin, Via P. Giuria 7, 10125 Torino, Italy
[2]Centro Regionale Antidoping "A. Bertinaria", Regione Gonzole 10/1, 10043 Orbassano, Torino, Italy.
[3]Faculty of Chemistry, Jagiellonian University, Ingardena 3, 30-060 Krakow, Poland.
[4]Institute of Forensic Research, Westerplatte 9, 31-033 Krakow, Poland.
[5]Department of Chemistry, Purdue University, 560 Oval Drive, West Lafayette, 47907 Indiana, USA.
[6]Department of Analytical Chemistry, Chemometric Research Group, Institute of Chemistry, The University of Silesia, Szkolna 9, 40-006 Katowice, Poland.
mail: ealladio@unito.it

The determination of direct ethanol metabolites – such as ethyl glucuronide (EtG) and fatty acid ethyl esters (FAEEs) – to be quantified in the keratin matrix is currently remarked as the optimal strategy to effectively recognize chronic alcohol misuse conditions [1]. Even if cut-off values have been established by the Society of Hair Testing (SoHT) in order to interpret EtG and FAEEs results, several factors may strongly affect the whole interpretative process and lead analysts to infer misleading conclusions. In fact, the correlation between alcohol consumption and biomarkers concentration in hair may be altered, providing dissimilar results with respect to the conventional cut-off values and making the interpretation process extremely biased. Due to the fact that likelihood ratio (LR) models overcome the drawbacks of the traditional univariate approaches, as no cut-off values are involved during the process of evidence evaluation, several LR models were developed and tested. In the practice, LR values reveal the support to be delivered to the evaluated propositions and LR results can be expressed by means of verbal scales. LR approach evaluates the evidence (E) in case of two different, and mutually exclusive, hypotheses by examining the collected data. In the present case, the first hypothesis (H_1) was that the individual under examination is a non-chronic alcohol consumer. Otherwise, the second hypothesis (H_2) stated that the examined subject is a chronic alcohol misuser.
Collected data consisted in direct (FAEEs, EtG) and indirect [1] biomarkers of alcohol consumption from 125 scalp hair and blood samples of 125 different individuals, representing both the chronic and the non-chronic alcohol drinkers categories. Different LR models were evaluated and their capability of discriminating chronic alcohol misusers from non-chronic alcohol consumers was examined. The performance of each model was evaluated in terms of rates of correct classification (%) and empirical cross entropy parameters [2]. Since satisfactory reduction of information loss was observed, together with correct classification rates close to 100%, LR validated models proved to be capable of discriminating non-chronic from chronic alcohol consumers. As a result, the adoption of LR seems to facilitate the decision process of effectively detecting alcohol misuse conditions. In our opinion, the use of the LR might represent an efficient way to corroborate a diagnosis of chronic abuse, providing knowledge about the strength of the support to be delivered to the selected proposition, in contrast with the traditional approach involving cut-off values evaluation.

References:

[1] P. Kintz, A. Salomone, M. Vincenti, Hair Analysis in Clinical and Forensic Toxicology, First Edit., Academic Press, Elsevier, (2015).
[2] G. Zadora, A. Martyna, D. Ramos, C. Aitken, Statistical Analysis in Forensic Science - Evidential Value of Multivariate Physicochemical Data, First Edit., John Wiley & Sons, Ltd., (2014).

P-075

APPLICATION OF MULTIVARIATE STATISTICS AND LIKELIHOOD RATIO TO THE STEROIDAL MODULE OF THE ATHLETE BIOLOGICAL PASSPORT: A PROOF OF CONCEPT STUDY

E. Alladio[1,2], R. Caruso[1], E. Gerace[2], E. Amante[1], A. Salomone[2], G. Zadora[3,4], P. Wlasiuk[3,5], A. Martyna[3,5], M. Vincenti[1,2]

[1]Chemistry Department, University of Turin, Via P. Giuria 7, 10125 Torino, Italy
[2]Centro Regionale Antidoping "A. Bertinaria", Regione Gonzole 10/1, 10043 Orbassano, Torino, Italy. [3]Institute of Forensic Research, Westerplatte 9, 31-033 Krakow, Poland.
[4]Department of Analytical Chemistry, Chemometric Research Group, Institute of Chemistry, The University of Silesia, Szkolna 9, 40-006 Katowice, Poland.
[5]Faculty of Chemistry, Jagiellonian University, Ingardena 3, 30-060 Krakow, Poland.
ealladio@unito.it

The Technical Document TD2014EAAS was drafted by the World Anti-Doping Agency (WADA) in order to fight the spread of endogenous anabolic androgenic steroids (EAAS) misuse in several sport disciplines[1]. In particular, adoption of the so-called Athlete Biological Passport (ABP) – Steroidal Module allowed control laboratories to identify anomalous EAAS concentrations within the athletes' physiological urinary steroidal profile. Gas chromatography (GC) combined with mass spectrometry (MS), indicated by WADA as an appropriate technique to detect urinary EAAS, was utilized in the present study to develop and fully-validate an analytical method for the determination of all EAAS markers specified in TD2014EAAS, plus two further markers hypothetically useful to reveal microbial degradation of the sample. In particular, testosterone, epitestosterone, androsterone, etiocholanolone, 5α-androstane-3α,17β-diol, 5β-androstane-3α,17β-diol, dehydroepiandrosterone, 5α-dihydrotestosterone, were included in the analytical method. Afterwards, the multi-parametric feature of ABP profile was exploited to develop a robust approach for the detection of EAAS misuse, based on multivariate statistical analysis. In particular, Principal Component Analysis (PCA) was combined with Hotelling T^2 tests to explore the EAAS data obtained from 60 sequential urine samples collected from six volunteers, in comparison with a reference population of single urine samples collected from 96 volunteers. The new approach proved capable of identifying anomalous results, including (i) the recognition of samples extraneous to each of the individual urine series and (ii) the discrimination of the urine samples collected from individuals to whom "endogenous" steroids had been administrated with respect to the rest of the samples population. Simultaneously, a multivariate likelihood ratio (LR) approach was tested to discriminate individuals taking endogenous steroids for doping purposes from the ones suffering from hormonal imbalances or diseases related to their urinary steroidal profile. Involving the previous reference population of healthy subjects, this multiclass classification LR approach seemed to be capable of recognizing whether the anomalous values are caused by an anti-doping rule violation or a pathological condition of the investigated individuals. The proof-of-concept results presented in this study will need further extension and validation on a population of sport professionals.

Acknowledgments: The experiments were undertaken thanks to the collaboration of the hospital San Giovanni Bosco of Torino (Italy) and the help of prof. Massimo Giusti.

References:

[1] World Anti-Doping Agency, WADA Technical Document – TD2014EAAS Endogenous Anabolic Androgenic Steroids, 2014.

P-076

TRACING ILLEGAL REMOVAL OF FISCAL MARKERS FROM DIESEL FUEL BASED ON CHROMATOGRAPHIC FINGERPRINTS AND THEIR CHEMOMETRIC MODELING

M. Daszykowski [1], B. Krakowska [1], J. Orzeł [1], I. Stanimirova [1], M. Szajder [2], G. Zaleszczyk [2], I. Grabowski [2]

[1]Department of Theoretical Chemistry, Institute of Chemistry, The University of Silesia, 9 Szkolna Street, 40-006 Katowice, Poland
[2]Customs Laboratory of Customs Chamber in Biala Podlaska, 21 Celnikow Polskich Street, 21-500 Biala Podlaska, Poland
email: michal.daszykowski@us.edu.pl

In EU countries, specific chemical components such as a marker (Solvent Yellow 124) and dyes (e.g. Solvent Red 19 or Solvent Red 164) are deliberately added in diesel fuel in order to indicate its tax rate. These duty components are meant to be stable at certain conditions and difficult to remove. Unfortunately, these requirements are often violated, which lead to various legal issues.
In our previous study, which was described in reference [1], we have focused on uncovering the differences between genuine samples and their illegal variants. Analyzing chromatographic fingerprints obtained from gas chromatography with a flame ionization detector (GC-FIC), we found out that these differences were due to other specific components than duty components residuals (excise duty components are not stable and degrade under the high temperature set during the separation).
In the present work, we report the results of a more comprehensive experiment. A total of 36 genuine diesel fuel samples were processed in laboratory conditions using two illegal fuel laundering methods. Bearing in mind chemical structures of the excise duty components of interest (azo- and diazo-compounds), gas chromatography with nitrogen chemiluminescence detector (GC-NCD) was adopted to trace the counterfeiting methodology. The effect of sample extraction using methanol as a solvent was also examined. After chemometric preprocessing of obtained GC-NCD chromatographic fingerprints (for raw samples and their methanol extracts), class modelling techniques such as SIMCA and one-class partial least squares (OC-PLS) were used in the identification of method applied to remove the selected excise duty components. Constructed discriminant and classification models were extended with variable selection schemes and were validated using the Monte-Carlo framework described in reference [2].

Acknowledgement: MD, JO and IS acknowledge the support of the National Science Centre, Poland (research grant no. 2014/13/B/ST4/05007).

References:
[1] B. Krakowska, I. Stanimirova, J. Orzel, M. Daszykowski, I. Grabowski, G. Zaleszczyk, M. Sznajder, Analytical and Bioanalytical Chemistry, 407, 1159 (2015).
[2] B. Krakowska, D. Custers, E. Deconinck, M. Daszykowski, The Analyst, 141, 1060 (2016).

P-077

A BAYESIAN APPROACH FOR DATA ANALYSIS OF CHOMATOGRAPHY – MASS SPECTROMETRY

Michael Woldegebriel, Gabriel Vivó-Truyols

Analytical Chemistry Group, Van't Hoff Institute for Molecular Sciences

University of Amsterdam PO Box 94720, 1090 GE Amsterdam, The Netherlands

M.T.Woldegebriel@uva.nl

The majority of data analysis methods applied to LC-MS (MS/MS) are based on frequentist statistics. For example, (frequentist-based) peak detection will provide a binary answer on the presence/absence of a chromatographic peak; a screening method (aimed to check the presence/absence of a set of compounds) will provide a list of the compounds that have been detected, etc. None of these methods calculates the probability of each of these hypotheses being true. Only the most likely answer is given, without knowing the probability of every possible case scenario. When applied sequentially at different steps of the data-analysis, frequentist-based methods posses a tremendous drawback in features classification (e.g. "peak" or "noise"), making the method vulnerable to prematurely filter away data. In contrast, a Bayesian approach estimates the probability of each of the hypotheses being true, allowing us to incorporate prior knowledge. Additionally, this framework can elegantly combine different pieces of evidence obtained from our data. As a result, the data is not pre-filtered in the pre-processing steps, but the probability is further propagated into other pre-processing steps (e.g. peak-alignment, confirmation etc.) to obtain a final result.

In this communication, a novel methodology of LC-MS (MS/MS) data analysis based on Bayesian framework will be presented. The first part concerns probabilistic untargeted peak detection [1]. The second part concerns the application of Bayesian framework in forensic (food-safety) toxicology context, which tackles the problem of compound screening in a probabilistic way [2]. Methods for taking into account further evidence (e.g. fragment ion information) when available, and application of probabilistic machine learning will also be discussed.

REFERENCES:

[1] Woldegebriel, M.; Vivó-Truyols, G. Anal. Chem. 2015, 87, 7345−55.

[2] Woldegebriel, M.; Gonsalves, J.; Asten, A.V; Vivó-Truyols, G. Anal. Chem. 2016. DOI: 10.1021/acs.analchem.5b04484

P-078

OPTIMIZATION AND VALIDATION OF AN ALTERNATIVE PROCEDURE FOR THE UNTARGETED LC-MS DATA ANALYSIS

Núria Dalmau, Carmen Bedia, Romà Tauler

Department of Environmental Chemistry, IDAEA-CSIC, Jordi Girona 18-26, Barcelona, Spain

nuria.dalmau@idaea.csic.es

Untargeted liquid chromatography coupled to mass spectrometry (LC-MS) analysis generates massive amounts of information-rich mass data and presents storage problems. In this work, the optimization and validation of an alternative procedure for the untargeted analysis of LC-MS data sets in the omics field are presented to resolve these disadvantages. This method consists of a preliminary pre-processing of data sets based on the selection of the Regions of Interest (ROI) [1, 2, 3] coupled to multivariate curve resolution-alternating least squares (MCR-ALS) [4]. ROI data selection pre-treatment allows for an important reduction of the data size, without any loss of spectral resolution or of accuracy on m/z measures. ROI selection is based on the search of significant mass traces regions with high mass densities. The adjustment of ROI parameters has been performed by the use of different dilutions of standard stock solutions mixtures injected in an UHPLC-ToF-MS instrument. The optimization of these ROIs parameters resulted in regular data matrices containing the same m/z accuracy as the original data sets, with much lower storage requirements. These LC-MS data matrices were subjected to MCR-ALS analysis for a proper resolution of chromatographic and mass spectra profiles. Experimental data coming from different standard mixtures were used to validate the proposed ROI-MCR-ALS strategy. Additionally, LC-MS data from different lipidomic samples were also analysed with satisfactory results. Altogether, in this presentation, the ROI-MCR-ALS strategy was optimized and validated, and it is proposed as a general data pre-treatment method in LC-MS omic studies. This strategy is confirmed to provide a reliable method that enables a significant reduction of analysis times in untargeted LC-MS analysis, without any loss of relevant information.

Keywords: Regions of interest (ROI), multivariate curve resolution-alternating least squares (MCR-ALS), untargeted omic analysis.

Acknowledgement: European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement 35 no. 320737.

References:

[1] Stolt, R., Torgrip, R., Lindberg, J, Csenki, L., Kolmert, J., Schuppe-Koistinen I., Jacobsson, S. Second-Order Peak Detection for Multicomponent High-Resolution LC-MS Data. Anal. Chem. 78, 975 (2006).

[2] Tautenhahn, R., Böttcher, C., Neumann, S. Highly sensitive feature detection for high resolution LC-MS. BMC Bioinf., 9, 504 (2008).

[3] Gorrochategui, E., Jaumot, J., Tauler, R. A protocol for LC-MS metabolomic data processing using chemometric tools. Protocol Exchange, Doi:10.1038/potex.2015.102 (2015).

[4] Tauler, R., Multivariate curve resolution applied to second order data. Chemometrics and Intelligent Laboratory Systems, 30(1), 133 (1995).

P-079

BATCH EFFECT CORRECTION IN HIGH RESOLUTION LCMS USING QUALITY CONTROL SAMPLES AND SUPPORT VECTOR REGRESSION (QC-SVRC)

D. Sanjuan[1], J. Kuligowski[2], A. Sánchez-Illana[2], Máximo Vento[2,3], G. Quintás[1,4*]

[1]Safety and sustainability Division, Leitat Technological Center, Barcelona
[2]Neonatal Research Centre, Health Research Institute La Fe, Valencia
[3]Division of Neonatology, University & Polytechnic Hospital La Fe, Valencia
[4]Analytical Unit, Health Research Institute La Fe, Valencia
gquintas@leitat.org

Ultra performance liquid chromatography – high resolution mass spectrometry is increasingly being used in metabolomics. This technique provides outstanding levels of sensitivity and selectivity with broad detection capabilities. Nonetheless, gradual changes in the instrumental response as a function of time (i.e. intra-batch effect) are common, often unavoidable, reducing the repeatability and reproducibility, as well as the power to detect underlying biological effects. Because of that, the post-acquisition chemometric correction of batch effects is a promising approach to overcome this potential pitfall. In this study, the use of a new method based on quality control (QC) samples and support vector regression (QC-SVRC) [1] is proposed for the elimination of intra-batch effects. The performance of the method and the criteria for a fast selection of the SVRC parameters (ε-insensitive loss, gamma and C, using a radial basis function kernel) was tested using a model example, the repeated analysis of a plasma sample and data from clinical studies involving untarget metabolomic profiling of plasma and urine samples. Results obtained were compared with the reference approach based on QC samples and robust cubic smoothing splines (QC-RSC) [2]. The QC-SVRC allowed a straightforward and effective fitting of the SVRC parameters to the instrument performance that significantly improved analytical precision.

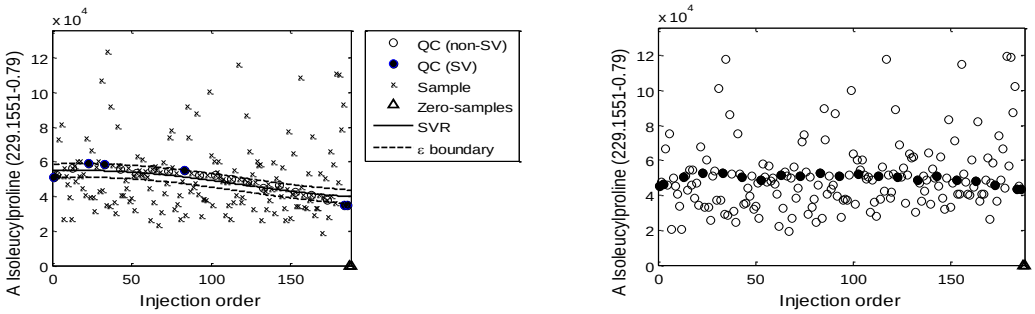


Figure 1. Left) Intensity of Isoleucylproline in plasma as a function of the injection order in a single batch (180 samples). The solid line depicts the calculated SVR function. ε boundaries are given with the dotted line; Right) Intensity after intra-batch effect correction.

Acknowledgement: JK acknowledge the 'Sara Borrell' grant (CD12/00667) from the *Instituto Carlos III* (Ministry of Economy and Competitiveness, Spain). ASI acknowledges the support of the RETICS funded by the PN 2018-2011 (Spain), ISCIII- Sub-Directorate General for Research Assessment and Promotion and the European Regional Development Fund (FEDER), reference RD12/0026. MV acknowledges the FISPI14/0433 grant from the *Instituto Carlos III* (Ministry of Economy and Competitiveness, Spain). GQ acknowledges financial support from the Spanish Ministry of Economy and Competitiveness (SAF2012-39948) and *Fundació Marató-TV3* (1007/C/2013).

References:

[1] J. Kuligowski, A. Sánchez-Illana, D. Sanjuan-Herráez, M. Vento, G. Quintás, Analyst, 140, 7810 (2015).
[2] J. A. Kirwan, D. I. Broadhurst, R. L. Davidson, M. R. Viant, Anal. Bioanal. Chem., 405, 5147 (2013).

P-080

UNIVARIATE AND MULTIVARIATE CHEMOMETRICS APPROACHES APPLIED TO LC-QTOF-MS BASED TARGETED METABOLOMICS DATA FOR THE IDENTIFICATION OF POTENTIAL BIOMARKERS IN PLASMA IN PEDIATRICS WITH CHRONIC KIDNEY DISEASE

S. Benito[1], A. Sánchez[2], N. Unceta[1], J.J. Jansen[3], G. Postma[3], F. Andrade[4], L. Aldámiz-Echevarria[4], L.M.C. Buydens[3], M.A. Goicolea[1], R.J. Barrio[1]

[1]Department of Analytical Chemistry, University of the Basque Country (UPV/EHU), Faculty of Pharmacy, Paseo de la Universidad 7, 01006 Vitoria-Gasteiz, Spain.
(sandra.benito@ehu.eus)
[2]Central Service of Analysis (SGiker), University of the Basque Country (UPV/EHU), Paseo de la Universidad 7, 01006 Vitoria-Gasteiz, Spain
[3]Radboud University, Institute for Molecules and Materials (Analytical Chemistry-Chemometrics). P.O. Box 9010, 6500 GL Nijmegen, The Netherlands.
[4]Group of Metabolism, BioCruces Health Research Institute, CIBER de Enfermedades Raras (CIBERER), Plaza de Cruces 12, 48903 Barakaldo, Spain

Chronic kidney disease (CKD) is considered a major worldwide public health problem which causes several disturbances in adults and in pediatrics due to the irreversible kidney damage which can further progress to renal hypofunction. Nevertheless, information available for CKD in pediatric population is limited and, even in adults, CKD is difficult to diagnose, to follow in progression and to evaluate response to therapy[1]. In clinical practice, creatinine is considered the classic biomarker for the assessment of renal function[2]. However, creatinine lacks sensitivity and reveals kidney damage when an important nephronic loss has already taken place[3]. For that reason, with the aim of finding new potential biomarkers, a targeted metabolomics approach has been performed.

An ion-pairing LC-QTOF-MS methodology, developed and optimized for the quantification of 16 metabolites from the arginine-creatinine metabolic pathway, arginine methylation and urea cycle, has been applied to quantify these compounds in plasma from thirty-two patients with different degrees of chronic kidney disease (aged 3-17 years) and twenty-four control patients not suffering from chronic kidney disease (aged 6-18 years)[4]. Then, univariate statistical analysis has been performed by using Student's T test for normal variables and U test in accordance to Mann and Whitney for non-normal variables to find significant differences in concentration between both groups. Finally, these results have been compared with a multivariate approach, which involved scaling the data using different methods prior to Principal Component Analysis, reconstruction of the original data matrix and subsequent classification of the data aiming at building different predictive models for classifying future observations.

Acknowledgement: The authors thank for technical and human support provided by Alava Central Service of Analysis–SGIker- of UPV/EHU as well as the Division of Metabolism belonging to Cruces University Hospital (Barakaldo, Spain) for supplying real samples for this study. This work was funded by the Department of Industry, Innovation, Commerce and Tourism of the Basque Government (SAI12/25 Project) and by the Basque Government, Research Groups of the Basque University System (Project No. IT338-10). The Basque Government is also gratefully acknowledged for a predoctoral grant (PRE_2013_1_899) for Sandra Benito (Department of Education, Language Policy and Culture).

References:

[1] V. Fanos, C. Fanni, G. Ottonello, A. Noto, A. Dessì, M. Mussap. Molecules. 18, 4644 (2013).
[2] M.E. Wasung, L.S. Chawla, M. Madero. Clinica Chimica Acta. 438, 350 (2015).
[3] R.H. Weiss, K. Kim. Nature Reviews Nephrology. 8, 22 (2012).
[4] S. Benito, A. Sanchez, N. Unceta, F. Andrade, L. Aldamiz-Echevarria, M.A. Goicolea, R.J. Barrio. Analytical and Bioanalytical Chemistry. 408 (2016)

P-081

Manganese: changes in brain revealed by a combination of metallomic and metabolome investigation.

Lucio Marianna[1], Katharina Neth[1], Alesia Walker[1], Bernhard Michalke[1] Philippe Schmitt Kopplin[12]

[1]Research Unit Analytical BioGeoChemistry, German Research Center for Environmental Health, Neuherberg, Germany
[2] Chair of Analytical Food Chemistry, Technische Univers ität München, Alte Akademie 10, D- 85354, Freising-Weihenstephan, Germany

marianna.lucio@helmholtz-muenchen.de

Occupational and environmental exposure to increased concentrations of manganese (Mn) can lead to an accumulation of this element in the brain. The consequence is an irreversible damage of dopaminergic neurons leading to a disease called manganism with a clinical presentation similar to the one observed in Parkinson's disease. Human as well as animal studies indicate that Mn is mainly bound to low molecular mass (LMM) compounds such as Mn-citrate when crossing neural barriers. Two different techniques have been combined in order to study the formation of Mn-species and their possible influence on brain metabolism: size exclusion chromatography-inductively coupled plasma mass spectrometry (SEC-ICP-MS) and electrospray ionization ion cyclotron resonance Fourier transform mass spectrometry (ESI-ICR/FT-MS). In one study we investigated Mn-species pattern in serum in two different animal models of Mn-exposure (chronic vs. subacute). Results from Mn-speciation were correlated to the brain metabolome, measurements coming from ESI-ICR/FT-MS. The powerful combination of Mn- speciation in serum with metabolomics of the brain underlined the need for Mn-speciation in exposure scenarios instead of the determination of whole Mn concentrations in blood. The second study was done applying a single low dose $MnCl_2$ injection in rats, where we observed alterations in Mn-species pattern within the brain by analysis of aqueous brain extracts. Additionally, ESI-ICR/FT-MS measurement of methanolic brain extracts revealed a comprehensive analysis of changes in brain metabolisms after the single $MnCl_2$ injection. Major alterations were observed for amino acid, fatty acid, glutathione, glucose and purine/pyrimidine metabolism. Again, results from the metallomic investigations (Mn concentrations and Mn-species in brain) were correlated with the findings from metabolomics. This first attempt/approach could serve for detecting Mn-species, which play a leading part in the cascade of neuronal injury during Mn-exposure.

P-082

APPLICATION OF CHEMOMETRIC TOOLS TO THE METABOLOMIC PROFILING OF PATIENTS WITH ASTHMA

A.Checa[1], J. Jaumot[2], H. Gallart-Ayala[1], S. Naz[1], S.N. Reinke[1], K. Alving[3], C.E. Wheelock[1]

[1]Department of Medical Biochemistry and Biophysics, Division of Physiological Chemistry II, Karolinska Institute, Scheeles väg 2, Stockholm, Sweden
[2]Department of Environmental Chemistry, IDAEA-CSIC, Jordi Girona 18-26, Barcelona, Spain
[3]Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden

e-mail: antonio.checa@ki.se

Asthma is a chronic disease of the airways that affects around 300 million people worldwide. Among patients with asthma, atopy is an important risk factor accounting for up to 70% of the asthmatics in Sweden. Metabolomic profiling of patients with asthma appears as a new tool that can help unraveling the molecular determinants of the disease. Though analyzed cohorts may be well balanced, factors such as age or gender within a class may confound the results when analyzing human biofluids such as plasma. The aim of the present work is to explore phenotypic differences at the small molecule level between a cohort of young atopic and non-atopic asthmatics relative to controls (age range = 21 years) with the aid of chemometric tools.

Four datasets were acquired using complementary chromatographic separations, reversed phase and HILIC, coupled to high resolution mass spectrometry in both positive and negative ionization modes. For each matrix, Multivariate Curve Resolution – Alternating Least Squares (MCR-ALS) allowed to extract relevant components characterized by chromatographic profile and its mass spectra. The effect of age and gender on the metabolomic profile was evaluated by means of ANOVA Simultaneous Component Analysis (ASCA). Finally, discrimination between groups was performed using partial least squares – discriminant analysis (PLS-DA) and Variables Important on Projection (VIP) were used to extract the most relevant features for class discrimination. These obtained features will be later tentatively identified with the help of their simultaneously acquired fragmentation pattern and online databases.

P-083

USE OF MULTIVARIATE ANALYSIS TO DIFFERENTIATION OF GLASSES AS FORENSIC EVIDENCE BY UV-VIS

K. Leiva[1], P. Sáez[2], L. Bustamante[2], J. Gárate[2], F. Torres[2], P. Richter[1], E. Fuentes[1].

[1]Departamento de Química Inorgánica y Analítica, Facultad de Ciencias Químicas y Farmacéuticas, Universidad de Chile, Sergio Livingstone Polhamer 1007, Independencia, Santiago, Chile.
[2]Sección Microanálisis, Laboratorio de Criminalística Central, Policía de Investigaciones, Carlos Silva Vildosola 9783, La Reina, Santiago, Chile
karlita.leiva gmail.com

Glass is an inorganic amorphous material of crystalline and translucent appearance that can be found in several crime scene, such a robbery or traffic accidents. The glass fragments are collected at the crime scene, in the belongings of a suspect or the body of a victim, which are subsequently analyzed, being the most important forensic objectives, the study of its classification, comparison and differentiation, in order to be able to determinate their origin [1]. When the evidence found fall within category of trace, is essential to develop a non-destructive, non-invasive method of analysis and that requiring a minimum preparation. . Consistent with the above, was developed and implemented an analytical methodology based of spectrophotometry UV visisble (UV-vis), technique to allow characterize and differentiate different types of glass with a same brand through multivariate analysis of the reflectance spectra. The equipment used for the spectra is a video spectral comparator 8000 (VSC8000) of Foster and Freeman.
The chemometric tools used were principal component analysis (PCA), linear discriminant analysis (LDA), soft independent modeling of class analogies (SIMCA) and partial least square discriminant analysis (PLS-DA).
Were analized 160 samples, that include 20 types of glass, of which 120 are occupied as calibration set and 40 were used as validation set was analyzed. PCA was used as a screening, seeing if we can differentiate glasses samples by reflectance spectra. The classification model LDA gave a 100% of the samples classified in the correct group, by the auto-predicction, to put through the model to the validation test, the 40 samples were well classified. The same result were obteined by PLS-DA, that is to say the 100% of validation was well classified. Finally, the SIMCA model resulted only 1 sample classified in a wrog group, whereas the validation set 3 samples were misclassified.
In conslusion, using a simple and inexpensive technique applying chemometric tools allow differentiate a large number of samples of glasses as a forensic evidence of similar characteristics, obtained very good results in their classification, achieve with the objetives of execute a rapid and non-destructive analysis, besides being a technique with expectation for analyzing this type of evidence.

Acknowledgement: The authors thank International Atomic Energy Agency, Project N° CHI-0017.

References:
[1] Zadora G., J. Chemometr. 21: 174 – 186 (2007)

P-084

MULTIVARIATE DATA ANALYSIS BASED ON TRANSFERRIN GLYCOPEPTIDE LC-MS FINGERPRINTS FOR CLASSIFICATION OF CONGENITAL DISORDERS OF GLYCOSYLATION

A. Barroso[1], E. Giménez[1], F. Benavente[1], J. Barbosa[1], V. Sanz-Nebot[1]

[1]Departament de Química Analítica, Facultat de Química, Universitat de Barcelona, Barcelona, Spain.
email: albertbarroso@ub.edu

Protein glycosylation plays an important role in many domains of life and is well known to be one of the most important and complex post-translational modifications in humans [1]. In bottom-up protein glycosylation analysis, targeting the glycopeptides is usually the most enticing option because these glycoprotein fragments provide information about the structure and composition of the glycans, as well as the glycosylation sites and their degree of occupancy. However, analysis of glycopeptides of an enzymatic protein digest is a challenge, due to the complexity of the digests, which are mixtures of peptides and glycopeptides, and the microheterogeneity of the glycopeptides, which present several glycoforms. In this regard, multivariate data analysis may be very useful for data processing, exploration and classification of the complex and massive data sets generated by high performance separation techniques coupled to mass spectrometry in glycoproteomic studies.

In this work, two different multivariate data analysis approaches are used to study the alteration of human transferrin (Tf) N-glycopeptides in patients with congenital disorders of glycosylation (CDG), a family of rare genetic, metabolic disorders that affect the biosynthesis or remodelling of the oligosaccharide moieties of glycoconjugates [2]. Tf from healthy individuals and two types of CDG patients (CDGI and CDGII) is purified by immunoextraction from serum samples before trypsin digestion and separation by capillary liquid chromatography mass spectrometry (CapLC-MS) [3]. First, following a conventional targeted approach, partial least squares discriminant analysis (PLS-DA) is applied using the relative abundance of Tf glycopeptide glycoforms obtained after integration of the extracted ion chromatograms of the different samples. Afterwards, as a novel alternative, multivariate curve resolution alternating least squares (MCR-ALS) is evaluated to automatically resolve the chromatographic profiles and the mass spectra of the different sample components, especially glycopeptide glycoforms, before PLS-DA. The performance of both approaches for classification of the different samples and for providing a novel insight into Tf glycopeptide glycoforms alteration in CDGs is demonstrated and compared.

References:

[1] A.P. Corfield, M. Berry, Trends Biochem. Sci., 40, 351 (2015).
[2] K. Scott, T. Gadomski, T. Kozicz, E. Morava, J. Inherit. Metab. Dis., 37, 609 (2014).
[3] A. Barroso, E. Giménez, F. Benavente, J. Barbosa, V. Sanz-Nebot, Anal. Chim. Acta,, 804, 167 (2013).

.

P-085

ON-LINE SPE-CE-MS COMBINED WITH ADVANCED DATA ANALYSIS TOOLS FOR THE IDENTIFICATION OF HUNTINGTON BIOMARKERS IN MICE PLASMA

F. Benavente[1], L. Pont[1], J. Jaumot[2], R. Tauler[2], J. Alberch[3], S. Ginés[3], J. Barbosa[1], V. Sanz-Nebot[1]

[1]Departament de Química Analítica, Facultat de Química, Universitat de Barcelona, Barcelona, Spain.
[2]Department of Environmental Chemistry, IDAEA-CSIC, Jordi Girona 18-26, 08034 Barcelona, Spain.
[3]Departament de Biologia Cel·lular, Immunologia i Neurociències, Facultat de Medicina, Universitat de Barcelona, Spain.
fbenavente@ub.edu

Huntington's disease (HD) is a fatal neurodegenerative disorder, which is characterized by progressive motor and cognitive disturbances [1]. In this work, an untargeted metabolomic approach based on sensitive analysis by on-line solid-phase extraction capillary electrophoresis mass spectrometry (SPE-CE-MS) [2] in combination with multivariate data analysis is proposed as an efficient method for the identification of biomarkers of HD progression in plasma.

For this purpose, plasma samples from wild type (wt) and HD (R6/1) mice of different ages (8, 12 and 30 weeks) were analyzed by C_{18}-SPE-CE-MS in order to obtain the characteristic electrophoretic profiles of low molecular mass compounds. Then, the combination of multivariate curve resolution alternating least squares (MCR-ALS) with other chemometric tools, such as partial least squared discriminant analysis (PLS-DA) [3], allowed the comprehensive analysis of the C_{18}-SPE-CE-MS metabolomic data, resolving electrophoretic peaks and mass spectra of a large number of metabolites.

A total number of 29 compounds were relevant to discriminate between wt and HD plasma samples, as well as to follow-up the HD progression. Although different pathways were found altered in HD, the intracellular signaling was observed to be the most affected, especially after 12 weeks of birth, thus suggesting that the pathology involves dysfunction of specific neurons, altered expression of several types of receptors and changed expression of neurotransmitters.

References:

[1] D. Zielonka, M. Mielcarek, G.B. Landwehrmeyer, Parkinsonism Relat. Disord., 21, 169 (2015).
[2] L. Pont, F. Benavente, J. Barbosa, V. Sanz-Nebot, J Sep Sci. 36, 3896 (2013).
[3] E. Ortiz-Villanueva, J. Jaumot, F. Benavente, B. Piña, V. Sanz-Nebot, R. Tauler, Electrophoresis, 36, 2324 (2015).

P-086

INVESTIGATION OF HEAT-TREATED CULTIVATION MEDIUM FOR MAMMALIAN CELLS WITH NEAR-INFRARED SPECTROSCOPY

É. Szabó[1], L. Párta[1,2], D. Zalai[2], S. Gergely[1] and A. Salgó[1]

[1]Department of Applied Biotechnology and Food Science, Budapest University of Technology and Economics, Szent Gellért tér 4, H-1111 Budapest, Hungary.
e-mail: szabo.eva@mail.bme.hu
[2]Department of Biotechnology, Gedeon Richter Plc., Gyömrői út 19–21, H-1103 Budapest, Hungary.

Nowadays qualification and control of medium formulations is performed based on simple methods (e.g. pH and osmolality measurement of media solutions), expensive and time consuming cell culture tests, and quantification of some critical compounds by liquid chromatography. Besides the traditional medium qualification tools, relatively new spectroscopic techniques, such as fluorescence spectroscopy, nuclear magnetic resonance, Raman and NIR spectroscopies or the combination of these techniques are increasingly being applied for medium powder investigation.[1,2]
A chemically defined medium powder for Chinese hamster ovary (CHO) cell cultivation was investigated in this study, regarding its response to heat treatments with different exposure times (1, 7 and 13 hours) and temperatures (30, 50 and 70 °C). The heat treatments were performed according to a design of experiments (DoE) approach. Spectra of the control and the treated powders were collected to compare the sample groups using a dispersive and a Fourier-transform (FT) near-infrared (NIR) spectrometer. Multivariate data analysis including unsupervised and supervised classification methods (CA, PCA, SIMCA etc.) were employed to identify the treatment-induced variations in the samples. Samples were separated according to the temperature setpoints of heat treatments and the control samples were successfully discriminated based on second derivative NIR spectra.

References:

[1] A. Gilbert,Y. Huang, T. Ryll, Pharm. Bioprocess. 2, 519 (2014)
[2] A. O. Kirdar, G. Chen, J. Weidner, A. S. Rathore, Biotechnol. Prog. 26, 527 (2010)

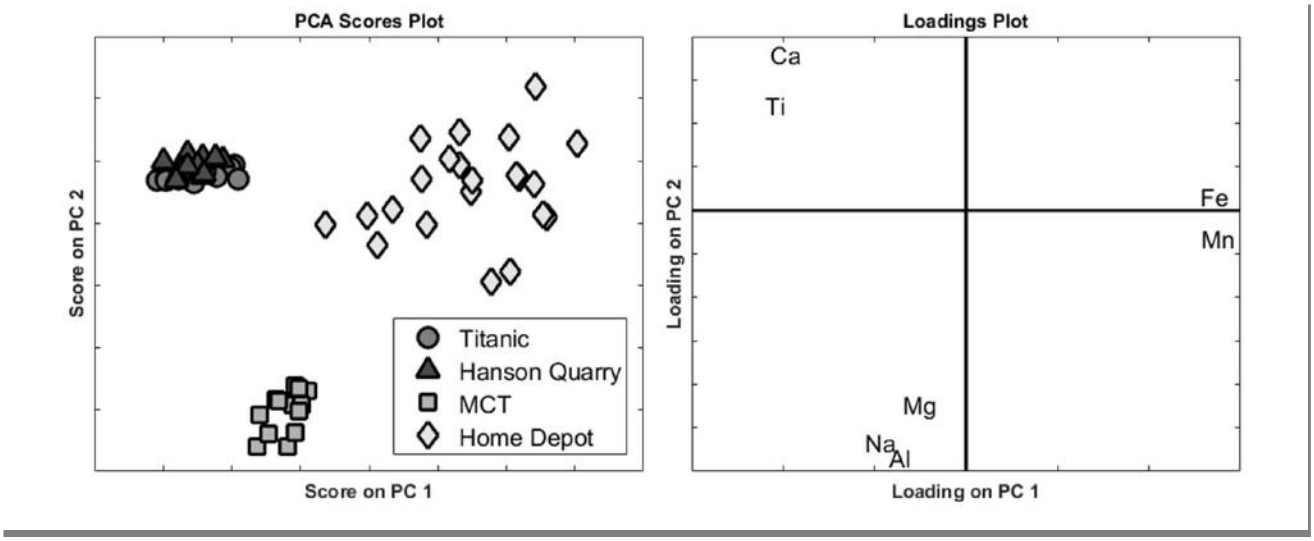
P-087

EXPLORATORY ANALYSIS OF ICP-MS DATA TO EVALUATE THE ORIGIN OF GRANITE HEADSTONES FROM TITANIC GRAVES

C. Wicks [1], P.D. Wentzell [1], D.B. Clarke [2]

[1]Department of Chemistry, Dalhousie University, Halifax, NS, B3H 4R2, Canada, ch565235@dal.ca
[2]Department of Earth Sciences, Dalhousie University, Halifax, NS, B3H 4R2, Canada

The tragic sinking of the RMS Titanic in 1912 saw the deaths of more than 1500 people. Of those victims, 149 were retrieved and buried in Halifax, Nova Scotia. The headstones are all composed of a black granite and were all provided by the same supplier, however there is no historical documentation describing the origin of the granite itself. This became an issue when one of the gravestones suffered frost damage and became in need of replacement. ICP-MS was used to measure the elemental composition of the granite from the headstone as well as three other granite samples with the objective of finding the most authentic source of the replacement stone, ideally the exact source used in 1912. One of the samples came from the Charles Hanson Quarry in south-western New Brunswick, another from Atwood's Brook [1], Nova Scotia, and the third was a tile sold at a commercial building supplies store with unknown origin. In the figure below it can be seen that principal component analysis (PCA) finds a projection with three groupings where the titanic samples are clearly grouped with those from the Charles Hanson Quarry. This should be an indication that of the three possible locations sampled, the Hanson Quarry as a source would provide the most authentic replacement for the damaged Titanic gravestone.



References:
[1] M.C. Tate, D.B. Clarke, Can. J. Earth Sci., 32,1883 (1995).

P-088

Manganese: changes in brain revealed by a combination of metallomic and metabolome investigation.

Lucio Marianna[1], Katharina Neth[1], Alesia Walker[1], Bernhard Michalke[1] Philippe Schmitt Kopplin[12]

[1]Research Unit Analytical BioGeoChemistry, German Research Center for Environmental Health, Neuherberg, Germany
[2] Chair of Analytical Food Chemistry, Technische Univers ität München, Alte Akademie 10, D- 85354, Freising-Weihenstephan, Germany

marianna.lucio@helmholtz-muenchen.de

Occupational and environmental exposure to increased concentrations of manganese (Mn) can lead to an accumulation of this element in the brain. The consequence is an irreversible damage of dopaminergic neurons leading to a disease called manganism with a clinical presentation similar to the one observed in Parkinson's disease. Human as well as animal studies indicate that Mn is mainly bound to low molecular mass (LMM) compounds such as Mn-citrate when crossing neural barriers. Two different techniques have been combined in order to study the formation of Mn-species and their possible influence on brain metabolism: size exclusion chromatography-inductively coupled plasma mass spectrometry (SEC-ICP-MS) and electrospray ionization ion cyclotron resonance Fourier transform mass spectrometry (ESI-ICR/FT-MS). In one study we investigated Mn-species pattern in serum in two different animal models of Mn-exposure (chronic vs. subacute). Results from Mn-speciation were correlated to the brain metabolome, measurements coming from ESI-ICR/FT-MS. The powerful combination of Mn- speciation in serum with metabolomics of the brain underlined the need for Mn-speciation in exposure scenarios instead of the determination of whole Mn concentrations in blood. The second study was done applying a single low dose $MnCl_2$ injection in rats, where we observed alterations in Mn-species pattern within the brain by analysis of aqueous brain extracts. Additionally, ESI-ICR/FT-MS measurement of methanolic brain extracts revealed a comprehensive analysis of changes in brain metabolisms after the single $MnCl_2$ injection. Major alterations were observed for amino acid, fatty acid, glutathione, glucose and purine/pyrimidine metabolism. Again, results from the metallomic investigations (Mn concentrations and Mn-species in brain) were correlated with the findings from metabolomics. This first attempt/approach could serve for detecting Mn-species, which play a leading part in the cascade of neuronal injury during Mn-exposure.

P-089

ANALYSIS OF DATA FROM NON-TARGETED METABOLOMICS USING DIRECT INFUSION FOURIER TRANSFORM ION CYCLOTRON RESONANCE MASS SPECTROMETRY

K.S. Smirnov [1], T. Lee [2,3], T. Clavel [4], A. Schmidt [5], I. Lagkouvardos [4], A. Walker [1], M. Lucio [1], B. Michalke [1], P. Schmitt-Kopplin [1,4], R. Fedorak [2], D. Haller [4,5]

[1]Research Unit Analytical BioGeoChemistry, Helmholtz Zentrum München, German Research Center for Environmental Health, Ingolstädter Landstraße 1, 85764 Neuherberg, Germany
[2]Division of Gastroenterology, Department of Medicine, University of Alberta, T6G 2X8, Edmonton, Canada
[3]Department of Gastroenterology, Wollongong Hospital, Loftus St, NSW 2500 Wollongong, Australia
[4]ZIEL Institute for Food and Health, Technische Universität München, Weihenstephaner Berg 1, 85354 Freising, Germany
[5]Chair of Nutrition and Immunology, Technische Universität München, Gregor-Mendel-Str 2, D-85350 Freising, Germany
kirill.smirnov@helmholtz-muenchen.de

Metabolome analysis of biological samples via high resolution Fourier ion cyclotron resonance mass spectrometry (FT-ICR-MS) provides the opportunity to screen thousands of signals simultaneously that gives the potential to obtain much explicit picture of interactions hidden in the corresponding datasets. We examined the potential of using direct infusion FT-ICR-MS (12 T) in a non-targeted study on the influence of iron replacement therapy in iron deficient patients with inflammatory bowel disease. Fecal metabolome of individuals divided into control (NON), Crohn's disease (CD), and ulcerative colitis (UC) group was analyzed. In addition, stool samples were subjected to 16S sequencing in order to track changes in gut microbiota composition during the iron replacement therapy. Thus, connections within taxonomic and metabolome datasets can be revealed, thereby providing insights into "host – gut microbiota – metabolome" axis.

Analysis of multivariate data, coming from non-targeted metabolomics studies, is a complex task. The search of metabolites of interest often involves the necessity to deal with large variable to sample ratio, huge proportion of unknown, missing, and noise signals. As a consequence, choosing appropriate preprocessing and statistical methods is of crucial importance. Through the application of different multivariate techniques to metabolome dataset, the differences between NON and CD groups before the replacement therapy were detected. This discrimination was observed to vanish after the treatment. Analysis applied to NON and UC groups showed only slight differences before the treatment. In addition, type of the iron replacement therapy (intravenous or oral) was studied with respect to changes in fecal metabolome. Features, responsible for group discrimination in all the models, were assigned to putative metabolites via database search or by in-house written algorithm for constructing a mass difference network [1]. This network allows to assign unique molecular formulas to mass spectrometric signals in the metabolome dataset and hypothesize possible interactions between compounds since every connection between two metabolites represents a biochemical transformation. Significant differences were also observed in fecal microbiota composition in regards to disease group or the type of the iron replacement therapy. Thereby, the shifts in bacterial communty can be directly linked to changes in metabolic landscape. In addition, taking into account the mass difference network, it is possible to identify and reconstruct metabolic pathways leading to changes in bacterial metabolism and its composition.

References:

[1] D. Tziotis, N. Hertkorn, P. Schmitt-Kopplin, Eur J Mass Spectrom, 17(4), 415 (2011).

P-090

eRah: A Computational Tool Integrating Spectral Deconvolution and Alignment with Quantification and Identification of Metabolites in GC-MS-Based Metabolomics

X. Domingo-Almenara [1, 2], J. Brezmes [1, 2], M. Vinaixa [1, 2, 3], S. Samino [3], M. Díaz [4], L. Ibáñez [4], X. Correig [1, 2], A. Perera [5] and O. Yanes [1, 2, 3]

[1]Metabolomics Platform - IISPV, Dept. of Electronic Engineering (DEEEA), Universitat Rovira i Virgili, Tarragona, Catalonia, Spain.
[2]Biomedical Research Centre in Diabetes and Associated Metabolic Disorders (CIBERDEM), Madrid, Spain.
[3]Centre for Omic Sciences, Rovira i Virgili University, Reus, Catalonia, Spain.
[4]Institut de Recerca Pediàtrica, Hospital Sant Joan de Déu, University of Barcelona, Barcelona, Catalonia, Spain.
[5]B2SLAB, Department of ESAII - Center for Biomedical Engineering Research (CREB), Universitat Politècnica de Catalunya, Barcelona, Catalonia, Spain.

email: xavier.domingo@urv.cat

GC/MS-based metabolomics produce large and complex datasets characterized by co-eluting compounds and fragmentation of molecular ions caused by the hard electron ionization (EI). Free and open-source integrated worksflows allowing for the identification and quantification of the corresponding metabolites are neededHere we introduce eRah, a free computational tool written in the open language R composed of five core functions: (i) noise filtering and baseline removal of GC-MS chromatograms, (ii) an innovative compound deconvolution process using multivariate analysis techniques based on compound match by local covariance (CMLC) and orthogonal signal deconvolution (OSD) [1,2], (iii) alignment of mass spectra across samples, (iv) missing compound recovery, and (v) identification of metabolites by spectral library matching using publicly available mass spectra. eRah outputs a table with compound names, matching scores and the integrated area of compounds for each sample. The automated capabilities of eRah were demonstrated by the analysis of GC-qTOF MS data from plasma samples of adolescents with hyperinsulinaemic androgen excess (HIAE) and healthy controls. The quantitative results of eRah were compared to centWave, the peak-picking algorithm implemented in the widely used XCMS package, and further validated using pure standards and targeted analysis by GC-QqQ MS.

Acknowledgement: This research was partially funded by MINECO grant TEC2012-31074 and TEC2015-69076-P (to XC), TEC2013-44666-R and TEC2014-60337-R (to AP) and SAF2011-30578 and BFU2014-57466 (to OY). CIBER-BBN and CIBER-DEM are initiatives of the Spanish Instituto de Salud Carlos III (ISCIII).

References:

[1] Domingo-Almenara X, Perera A, Ramirez N, Cañellas N, Correig X, Brezmes J. Journal of Chromatography A, 1409, 226 (2015).
[2] Domingo-Almenara X, Perera A, Ramirez N, Brezmes J. Computer Methods and Programs in Biomedicine, 130, 135 (2016).

P-091

POTENTIAL METABOLOMIC MARKERS IN ORGANIC FARMING OBTAINED USING ASCA AND PLS-DA METHODS. PLUM FRUIT: A CASE STUDY.

F.J. Cuevas, G. Pereira-Caro, Ruiz-Moreno M.J., Moreno-Rojas J.M.

Postharvest Technology and Agrifood Industry Area, Andalusian Institute of Agricultural and Fishery Research and Training (IFAPA), Alameda del Obispo, Córdoba, Spain, fjulian.cuevas@juntadeandalucia.es

Prunus genus is an important genus crop in the Mediterranean basin including peach, almond, apricot, cherry and plum. Spain produces almost 10 % of the total of European plum production but it is the main exporter (seasonal market) compared to the rest of countries. Plum fruits has been highlighted to be health promoters due to the contribution of their non-volatile compounds profile. The concentration of those different compounds presented in the fruits is influenced by a number of factors, such as variety, culture system or ripeness. Cultivation system (organic vs conventional) impact greatly on the characteristics of the plants grown under both systems affecting the whole metabolism and the profiles of compounds, increasing or decreasing their concentrations. The sampling was carried out during two consecutive years (2012 and 2013) in two similar experimental orchards located in Seville (Spain). Two plum varieties, 'Showtime' and 'Black Amber' (*Prunus salicina* Lindl.), grown under conventional and organic conditions were selected to evaluate the influence of the culture system in the metabolomic profile of the non-volatile compounds. Liquid chromatography high resolution mass spectrometry method coupled to differential expression analysis software (Sieve) was used. Identification and quantification of nontargeted compounds were performed in the samples. We studied the discrimination ability of chemometrics applied to this example. Several pre-processing signaling techniques (Autoscaling, mean center, orthogonalizing, Pareto scaling and log correction) were evaluated for the databases obtained. ANOVA-Simultaneous Component Analysis (ASCA) and Partial Least Squares Discriminant Analyisis (PLS-DA) were used to obtain the main contributors explaining 'cultivar' and 'management' factors.

P-092

CARSPLS FOR PREDICTING BASIC NITROGEN AND AROMATICS CONTENTS IN CRUDE OILS USING LDI(+)FT-ICR MS

L. A. Terra[1], P. R. Filgueiras[1,2], L. V. Tose[2], W. Romão[2,3], E. V. R. de Castro[2], L. M. S. L. de Oliveira[4], J. C. M. Dias[4], B. G. Vaz[2,5], R. J. Poppi[1]

[1]Institute of Chemistry, State University of Campinas, Campinas-SP, Brazil
[2] Chemistry Department, Federal University of Espírito Santo, Vitória, Espírito Santo, Brazil
[3] Federal Institute of Education, Science and Technology of Espírito Santo, 29106-010 Vila Velha, ES, Brazil
[4]CENPES/PETROBRAS, Rio de Janeiro, RJ, Brazil
[5] Chemistry Institute, Federal University of Goiás, Goiania, GO 74001-970, Brazil
luciana.terra@iqm.unicamp.br

The aim of this work [1] was to use positive ion mode laser desorption ionization coupled to Fourier transform ion cyclotron resonance mass spectrometry (LDI(+)FT-ICR MS) and partial least squares (PLS) regression with variable selection based on competitive adaptive reweighted sampling (CARS) for predicting basic nitrogen and aromatics contents in crude oil. The basic nitrogen and aromatic content of the samples studied ranged from 0.016 to 0.151 and 8.4 to 35.1 wt %, and their content were determined using UOP Method 269-10 and ASTM D5443-14, respectively. In this study, 70 samples of Brazilian crude oil were analysed and the mass spectra have profiles that ranged from m/z 200 to 1000 Da, with an average molar mass (Mw) distribution centered between 554 and 636 Da. Most of the identified organic species are analogues of pyridine and polyaromatic hydrocarbon compounds that belong in the Nx, Nx[H], HC, and HC[H] classes. The number of variables decreased from 47,873 initially to 48 for basic nitrogen and 10 for aromatic compounds. The prediction models based on CARSPLS are presented in the Figure 1, where it is possible to observe low prediction errors (RMSEP of 0.012 for basic nitrogen and 3.73 for aromatics) and superior performance in comparison to model using all the variables. This methodology facilitates the interpretation, indicating the molecules that are responsible for the change in value of the property and it shows an improvement over the standard approaches for determination of these parameters which are laborious and time-consuming.

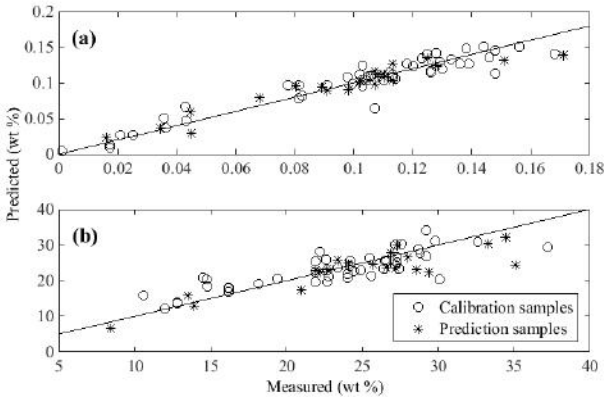


Figure 1. Predicted and measured values for (a) Basic Nitrogen and (b) Aromatics.

Acknowledgement: This research was generously funded by PETROBRAS/CENPES, FAPESP, CNPq, CAPES, FAPEG, FAPESP, and FINEP.

References:

[1] L.A. Terra; P.R. Filgueiras; L.V. Tose; W. Romao; E.V.R. de Castro; L.M.S.L. de Oliveira; J.C.M. Dias; B.G. Vaz; R.J. Poppi. Fuel, vol 160, 274 (2015).

P-093

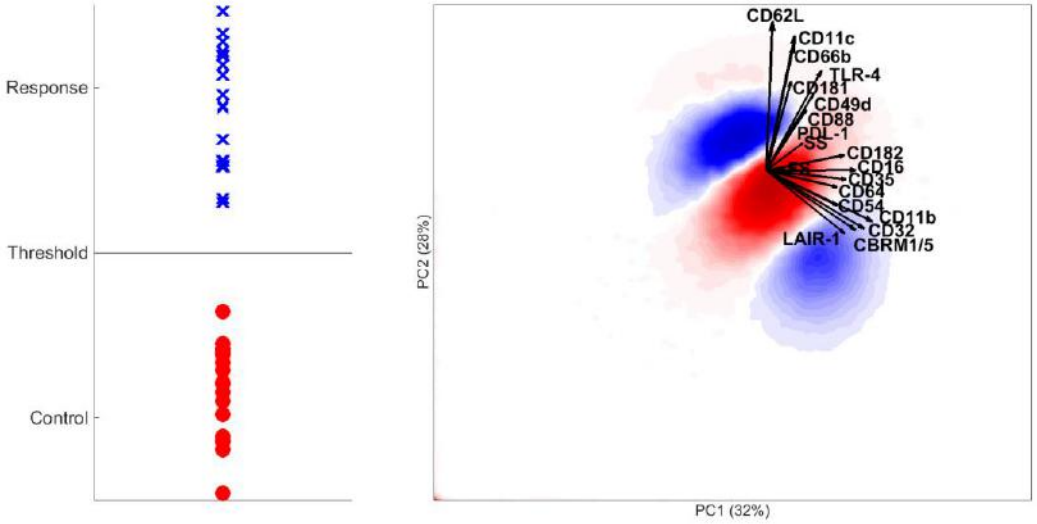
TOWARDS SINGLE-CELL OMICS WITH NOVEL DATA FUSION METHODS FOR FLOW CYTOMETRY DATA

Gerjen H. Tinnevelt[1,2], Lutgarde M.C. Buydens[1], Jeroen J. Jansen[1]

[1]Radboud University, Institute for Molecules and Materials, (Analytical Chemistry/Chemometrics), P.O. Box 9010, 6500 GL Nijmegen, The Netherlands
[2]TI-COAST, Science Park 904, 1098 XH Amsterdam, The Netherlands
Email: gtinnevelt@science.ru.nl

Multicolour flow cytometry (MFC) is a powerful analytical technique that is used to measure multiple surface markers at the single-cell level. A typical MFC sample may contain a very large number of cells (>10,000) [1]. If an individual has an immune response, certain cells with specific marker expression will be either over or underexpressed. This is why flow cytometry is used to study blood cells and immunology, including immune responses to drugs and tumour progression. However, the number of markers per measurement is technologically limited, because of the spectral overlap between fluorescent dyes. This is unfortunate, as many high-impact studies reveal that simultaneously combining more surface markers leads to a more detailed view of the immune system.

Currently, more surface markers can only be measured in multiple 'tubes'. The data from these separate tubes cannot be concatenated, as each tube contains different single cells. We propose two data fusion techniques that uses blocks of surface markers that were measured in different tubes in order to predict an immune response. One method predicts the expression of markers that were not measured on every cell. This method is based on a dedicated partial least squares (PLS)-based technique to impute this missing information. The final result can be seen in the Figure below, where all markers are in one model. The other method combines cellular distributions of each tube in order to predict the immune response. Both methods of data fusion brings flow cytometry considerably closer to the single-cell omics.



Acknowledgements: This research received funding from the Netherlands Organization for Scientific Research (NWO) in the framework of the Technology Area COAST of the Fund New Chemical Innovations.

References:

[1] H.G. Fienberg, G.P. Nolan, Vol. 337. (2014)

P-094

APPLICATION OF RAPID ANALYTICAL METHODS AND MULTIVARIATE DATA ANALYSIS FOR EVALUATING COLOR, TEXTURE AND FLAVOUR STABILITY OF YOGURT.

M. Casale[1], B. Aliakbarian[2], L. Bagnasco[1], P. Perego[2], R. Leardi[1]

[1]Department of Pharmacy, University of Genoa, Via Brigata Salerno 13-16147 Genoa, Italy
[1]Civil, Chemical and Environmental Engineering Department, University of Genoa, Via Opera Pia, 15- 16145 Genoa, Italy
monica@difar.unige.it

Color, texture and flavour are key elements of a consumer's buying decisions, thus, monitoring the stability of these features throughout the entire period of yogurt validity is fundamental for diary product producers. Color, flavour and texture deteriorations are due to changes in the physical, chemical and microbiological composition of yogurt but especially microbiological analysis of yogurt is expensive and time consuming.

In this study, UV-VIS spectroscopy was applied as a rapid and alternative technique to traditional analytical methods, to monitor the stability of yogurt up to 49 days of storage at 4°C.

UV-VIS spectroscopy was employed with an integrating sphere for diffuse reflectance measurements and, for each yogurt, color stability during storage time was evaluated in terms of CIELAB color space values [1].

In order to evaluate the texture and flavour changes, rheological curves and pH values of yogurt during storage were determined once a week for the entire period.

The information contained in the 3-way UV-VIS and rheological data sets was extracted using multivariate data analysis and specifically Tucker 3 [2,3] as a multi-way decomposition method.

It was interesting to note that the time-related information contained in the UV-VIS and rheological data was not visible by simply comparing the profiles of signals, partially visible in the 2 way Principal Component space, and very clear in the Tucker 3 models.

Color, texture and flavour of yogurt samples were also evaluated by a consumer acceptance test. The scores of the assessors were in good agreement with the results of 3-way PCA performed on the rheological measurements and the UV-VIS spectra.

Finally, GA-PLS [4,5] was performed as a very effective method for selecting the UV-VIS bands most informative in defining the temporal direction and for predicting the age of the yogurt samples; the results of GA-PLS confirmed that the band most informative in defining the temporal direction was the same individuated in the second mode of Tucker 3 and that it was possible to predict the age of the yogurt samples from their UV-VIS spectra.

Acknowledgement: This study was developed with funds from University of Genoa (CUP D31J13000000005).
The authors wish to thank the local producers (Caseificio Val d'Aveto, Centro Latte Rapallo- Latte Tigullio, Ars Food s.r.l, VIRTUS - L.Y.L.A.G. Lylag Sas) that generously made yogurt samples available for analysis.

References:

[1] CIE (Commission Internationale de l'eclairage) (1978). Supplement No. 2 to CIE Publ. No. 15 (E-1.3.L) 1971/9TC-1-3, CIE, Paris.
[2] LR. Tucker, Psychometrika 31, 279(1966).
[3] P. Geladi, Chemometrics Intell. Lab. Syst. 7, 11, (1989).
[4] R. Leardi and A. Lupiáñez González, Chemom. Intell. Lab. Syst., 41, 195 (1998).
[5] Z. Xiaobo, Z. Jiewen, M. J. W. Povey, M. Holmes and M. Hanpin, Anal. Chim. Acta, 667, 14 (2010)

P-095

ASSESMENT OF THE TASTE – MASKING EFFICIENCY OF SWEETENERS IN PHARMACEUTICAL SAMPLES BY ELECTRONIC TONGUE BASED ON VARIOUS CHEMOMETRIC TOOLS

M.Wesoły[1], P. Ciosek[1]

[1] *Department of Microbioanalytics, Warsaw University of Technology, Noakowskiego 3, 00 – 664 Warsaw, Poland*

Electronic tongues (ETs) are multisensor devices combined with various chemometric tools that are able to distinguish complex liquid samples and to recognize their characteristic properties. Nowadays electronic tongues are used for the recognition of variety of samples, but most popular application area is pharmaceutical analysis, including mainly taste masking effects detection and estimation [1].
The majority of Active Pharmaceutical Ingredients (APIs) present in oral drug products exhibit unpleasant bitter taste resulting in poor therapy adherence and patient inconvenience, especially in children and elderly. Thus taste masking has become an important part of drug formulation development. Various techniques for masking of a bitter taste of drugs have been reported. However, it is noticeable, that among various pharmaceutical formulations still simple masking with the use of sweeteners plays the key role in such investigation [2].
The main goal of this study was to compare the performance of various chemometric methods in the assessment of the taste – masking efficiency of various sweeteners using electronic tongue. Ibuprofen sodium salt was chosen as a model bitter drug and sucrose, acesulfame K, sucralose, aspartame and sodium saccharin added at 3 different concentration levels played a role of model bitter taste masking agents. Pure API and pure sweeteners were also analyzed. Sensor array composed of standard ion-selective electrodes (ISEs) was used for the acquisition of signals that formed chemical images of the studied samples.
Various pattern recognition procedures such as: Principal Components Analysis (PCA), Partial Least Squares - Discriminant Analysis (PLS-DA), Clusters Analysis and Soft Independent Modelling of Class Analogy (SIMCA) were employed in order to distinguish pharmaceutical samples. Comparison of performance of these methods to study taste – masking effects in pharmaceutical formulations measured with potentiometric sensor array was showed.

References:
[1] P. Ciosek, W. Wróbewski, Analyst vol.132, 963 (2007).
[2] M. Jańczyk, A. Kutyła, K. Sołłohub, H. Wosicka, K. Cal, P. Ciosek, Bioelectrochemistry vol. 80, 94 (2010).

Acknowledgement: This work has been supported from the financial resources for science in years 2013 – 2017 as a research project within a framework of "Diamond Grant" programme (DI 2012 0019 42) and by National Science Centre within a framework of OPUS project "Sensor arrays for the study of the release process of active pharmaceutical ingredients and excipients from pharmaceuticals" (DEC-2013/09/B/ST4/00957).

P-096

A rapid method to determine quality attributes of tomato products

D. Sun [1], J. Cruz [2], J. Casals [3], J. Simó [3], J. Sabate [4], M. Alcalà [1]

[1]Applied Chemometrics Research Group, Department of Chemistry, Faculty of Sciences, Universitat Autònoma de Barcelona, Bellaterra, Spain
[2]Escola Universitària Salesiana de Sarrià (EUSS), Barcelona, Spain
[3]Fundació Miquel Agustí, Sabadell, Spain
[4]Departament Enginyeria Agroalimentàira i Biotecnologia, Universitat Politècnica de Catalunya, Castelldefels, Spain
Dong.sun@uab.cat

1. Introduction

The production of tomato in Europe is around 22 million tons every year. [1] A main part of this crop is thermally processed and concentrated into tomato paste and the other part is purchased by consumers in raw. The quality of tomato product is controlled by several factors which can be classified into two groups, the chemical and sensory parameters. Near infrared spectroscopy (NIRS) is sensitive to both the chemical and physical changes in samples so it offers us an opportunity to correlate not only chemical but also sensory parameters with the NIRS spectra. Reference analysis is expensive and time consuming. Moreover, sensory analysis depends on the analyst experience and it would be necessary to develop more robust analytical methods.

2. Objective

The objective of this research is to develop a multiparametric NIR method to analyze the quality attributes (chemical and organoleptic) of tomato products.

3. Materials and methods

Two types of samples were analyzed (juice and puree). Near infrared spectra were acquired with a Foss NIR spectrometer system model 5000 (wavelength range is 1100-2498nm and resolution is 2nm, acquisiton modes include Rapid Content Analyzer and OptiProbe Analyzer). RCA accessory in transflectance mode with quartz cubbete and gold reflector was used for juice. Transmittance inmersion optical probe was used for puree. The chemical parameters (glucose, fructose, soluble solids and dry matter) were analyzed using reference methods. Sensory characterization (sweetness, taste intensity, aroma intensity, mealiness, acidity, crunchiness, skin perception, explosiveness and juiciness) were obtained by standarized trained tasting panel.

Partial least squares regression (PLSR) models were calculated with Unscrambler X (10.3 Trondheim, Norway).

4. Results and discussion

The models for fructose and glucose showed correlation of 0.95, RMSEC/RMSEP lower than 0.5g/L. The model for soluble solids showed correlation of 0.98, RMSEC/RMSEP lower than 0.3°Bx The model for dry matter content showed correlation of 0.95, RMSEC/RMSEP lower than 0.4g ms/100g mf.

The range for sensory characterizacion is 0-10 (arbitrary units).The models for explosiveness and juiciness showed correlation of 0.95, RMSEC/RMSEP lower than 0.2. The models for sweetness, aroma intensity and mealiness showed correlation of 0.8, RMSEC/RMSEP lower than 0.97. The models for taste intensity and crunchiness showed the correlation of 0.6, RMSEC/RMSEP lower than 0.97. At last, the models for skin perception and acidity showed correlation of 0.45, RMSEC/RMSEP lower than 1.0.

5. Conclusion

Multiparametric NIR method has been developed to quantify both chemical and organoleptic parameters of tomato products showing an excellent predictive performance.

References:

[1] Food and agriculture organization of United Nations, FAOSTAT database (2003-2013)

P-097

GRAPES TO WINE: FINDING CONTRIBUTORS TO WINE STYLE WITH MULTIBLOCK ANALYSIS

Katja Šuklje, Guillaume Antalick, Campbell Meeks, John W. Blackman, Alain Deloire and Leigh M. Schmidtke.

National Wine and Grape Industry Centre,Charles Sturt University, Locked Bag 588, Wagga Wagga, New South Wales 2678, AUSTRALIA.
email: lschmidtke@csu.edu.au

There is a complex association between grape composition and perceived quality with subsequent wine composition and style. Grape composition results from numerous interactions between cultivar, growing conditions, water, temperature, and the level of berry ripeness. A temporal and spatial investigation of berry composition and subsequent wine style was undertaken for two consecutive vintages in climatically diverse regions of Australia. Controlled triplicate fermentations of Shiraz or Cabernet Sauvignon grapes harvested at a measured berry maturity based upon a sugar accumulation model [1] to target wine styles from *Fresh* to *Mature* was undertaken. Sensory descriptive analysis of wines enabled different styles to be readily identified. Comprehensive metabolomic profiling of the grape composition (amino acids, carbohydrates, organics acids, volatile compounds, anthocyanins), wine chemical and volatile composition, winemaking inputs and wine sensory scores was undertaken. Using multiblock decomposition based upon Common Components and Specific Weights Analysis and PARAFAC the nexus between berry metabolites, wine making inputs and wine style were determined. As expected, clear markers associated with each variety were apparent, with the impact of vintage the predominant influence of within-variety differences. The effect of site was also evident, however differences associated with grape maturity and subsequent wine style were less obvious.

Acknowledgements: This work was funded by the grape growers and wine makers of Australia through the Australian Grape and Wine Authority, the Australian Federal Government and Charles Sturt University.

References:

[1] K. Šuklje, G. Antalick, C. B. Meeks, J. W. Blackman, A. Deloire, L. M. Schmidtke In *Optimising harvest date through use of an integrated grape compositional and sensory model : a proposed approach for a better understanding of ripening of autochthonous varieties ?* Exploitation of autochthonous and more common vine varieties Genetic pedigree and phenotyping, tolerance and stress, diseases to control, rootstocks: 2nd Oenoviti International Symposium, Geisenheim, 2014; Teissedre, P.-L., Ed. Oenologues de Bordeaux: Geisenheim, 2014; pp 44-49.

P-098

QUANTIFICATION OF POLYCYCLIC AROMATIC HYDROCARBONS IN PAPRIKA SAMPLES BY MEANS OF HPLC-FLD IN COMBINATION WITH MRC-ALS

O. Monago-Maraña[1], R. L. Perez[2], G. M. Escandar[2], A. Muñoz de la Peña[1] and T. Galeano-Díaz[1]

[1]Department of Analytical Chemistry and IACYS, University of Extremadura, Avenida de Elvas, s/n, Badajoz 06006, Spain
[2]Institute of Chemistry of Rosario (CONICET-UNIR), National University of Rosario, Suipacha 531, Rosario, Argentina
email: arsenio@unex.es

Polycyclic aromatic hydrocarbons (PAHs) are a large group of compounds containing two or more fused aromatic rings which are formed from proccesing, packaging and thermal processes (smoking, baking, roasting...) [1, 2, 3]. It is known that these compounds present carcinogenic, mutagenic and bioacumulative capacities and human can be exposed to them through three main routes: inhalation, skin contact and ingestion [4, 5]. Smoked process is commonly used in foods to obtain characteristic properties. One of these smoked products is the paprika. In Spain, La Vera (Extremadura) is one of the main geographical areas where paprika is produced. They employ a characteristic system for drying peppers to obtain paprika consisting in smoked-dried system (oak or holm wood fire).

In this work, several strategies were followed in order to quantify PAHs in paprika samples. Firstly, a solid extraction with Sep-Pack Silica Cartridges was optimized to concentrate and isolate the PAHs present in the samples. Secondly, a HPLC-FLD method was optimized to quantify eight PAHs (Fluorene, Phenantrene, Anthracene, Pyrene, Chrisene, Benzo(a)anthracene, Benzo(b)fluorantene and Benzo(a)pyrene). The chromatographic conditions were the following: isocratic elution: 65:35 ACN:H_2O; flow rate: 0.8 mL min^{-1}; C18 Zorbax Eclipse column and λ_{exc}= 260 nm and λ_{em}= 352 and 420 nm. With these conditions, some analytes co-eluted with matrix interferences, so it was decided to resolve and quantify these analytes with multivariate curve resolution-alternating least squares (MCR-ALS). The co-eluted analytes, quantified by this way, were fluorene, pyrene and benzo(b)fluoranthene. Good accuracy results were obtained in validation samples (standard solutions and spiked samples). Thus, these analytes were determined in a group of real samples (beloging or not to the Protected Designation of Origin, PDO, Pimentón de La Vera) without fortification. High contents of these analytes were found in PDO paprika samples compared with other dried spices regulated by the EC regulation 2015/1933 [6].

Acknowledgement: Financial support was provided by the MEyC of Spain (Project CTQ2014-52309-P) and the Gobierno de Extremadura (GR15090-Research Group FQM003), both co-financed by the European FEDER funds. Olga Monago Maraña is grateful to the MECyD of Spain for a FPU grant (Resolución de 18 de noviembre de 2013, BOE nº 279, de 21/11/13, reference number FPU13/02249). Rocío L. Pérez is grateful to Universidad Nacional de Rosario and Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET).

References:

[1] M. Ahmadvand, H. Sereshti, H. Parastar, J. Chromatography A, 1413, 117 (2015).
[2] A. Ishizaki, K. Saito, N. Hanioka, S. Narimatsu, H. Kataoka, J. Chromatography A, 1217, 5555 (2016).
[3] L. Singh, J.G. Varshney, T. Agarwal, Food Chemistry, 199, 768 (2016).
[4] E. Węgrzyn, S. Grześkiewicz, W. Poplawska, B.K. Glod. Acta Chromatographica, 17, 233 (2006).
[5] G. Purcaro, S. Moret, L.S. Conte, Talanta, 105, 292 (2013).
[6] EC (European Comission), 2015. Comission regulation (EU) Nº 2015/1933, of 27 October 2015, amending regulation (EC) Nº 1881/2006.

P-099

FISHER RATIO AND SENSORIAL ANALYSIS TO IDENTIFY GCXGC IMPORTANT VARIABLES RELATED TO CHOCOLATE FLAVOR

L. F. Oliveira, S. C. G. N. Braga, F. Augusto, R. J. Poppi

Institute of Chemistry, State University of Campinas (Unicamp),CP 6169 – 13083-971,Campinas, São Paulo, Brazil
lufontesoliveira@gmail.com

Comprehensive Two-dimensional gas chromatography coupled with quadrupole mass analyzer (GCxGC-qMS) and descriptive sensory analysis are powerful tools to study chocolate characteristics. By relating the GCxGC-qMS data with sensory data, it is possible to determine which compounds are responsible for specific quality of chocolates [1]. In this work descriptive sensory test (Optimized Descriptive Profile) [2] was applied in commercial chocolate samples to describe the chocolate flavor attribute. Furthermore, GCxGC-qMS chromatograms of the same chocolate samples were obtained after experimental optimization of extraction procedures. Fisher ratio [3] analysis in four-way data (retention time 1^{st} column x retention time 2^{nd} column x mass spectra x samples) was performed and the samples with high and low values of chocolate flavor attribute were used to define two classes. The 2D Fisher ratio plot, used as filter, selected the most important variables to the separation between the two classes. Different threshold values were used and a PCA model was applied in selected variables (for this propose the four-way data was unfolded, using the sum of all variables in mass spectra) to verify the separation between the two classes. The best threshold was $1x10^7$ and a PCA model with 3 component explained 78% of data variance, presenting excellent class separation. By using the loadings from PCA model, the 2D Fisher ratio plots and the mass spectra it was possible to identify the important compounds responsible by class separation. In this way, the proposed methodology can be useful for compounds identification related to a sensorial property.

Acknowledgement: The authors thank Fundação de Amparo à Pesquisa do Estado de São Paulo – FAPESP (process number: 2012/14695-8) for the financial support.

References:

[1] A.E. Croissant, D.M. Watson & M.A. Drake, *Annu. Rev. Food Sci. Technol.* 2, 395 (2011).
[2] R.C.S.N. Silva, V.P.R. Minim, A.A. Simiqueli, L.E.S. Moraes, A.I. Gomide & L.A. Minim, *Food Qual. Prefer.* 24, 190 (2012).
[3] K.M. Pierce, J.C. Hoggard, J.L. Hope, P.M. Rainey, A.N. Hoofnagle, R.M. Jack, B.W. Wright & R.E. Synovec, *Anal. Chem.* 78, 5068 (2006).

P-100

DETECTION OF ROASTED COFFEE ADULTERATION BY MEANS OF SYNCHRONOUS FLUORESCENCE SPECTROSCOPY WITH MULTIVATIATE DATA ANALYSIS

A.Dankowska [1], A. Kotwica

Poznań University of Economics, Department of Food Commodity Science, al. Niepodległości 10, 61-875 Poznań, Poland; anna.dankowska@ue.poznan.pl

[1] *Department of Food Commodity Science, Poznań University of Economics, al. Niepodległości 10, 61-875 Poznań, Poland*

Coffee as one of the most popular drinks consumed for its refreshing, stimulating taste and health benefits. Among about the 40 known different varieties, only two are of major importance for worldwide commercial coffee production: Coffea arabica and Coffea canephora var. robusta. On account of high price of arabica compared to robusta adulteration of coffee is practiced for economical purposes and therefore the detection of cheaper varieties of coffee is a real issue. Various instrumental methods have been proposed to establish the authenticity of Coffea arabica and to detect the level of its adulteration. Synchronous fluorescence spectroscopy is an alternative technique which is quick and avoids all sample preparation steps except for dilution and therefore it is simpler, less costly and quicker than other most widely used techniques.

The objective of this research was to investigate the potential of synchronous fluorescence spectroscopy followed by chemometric analysis for rapid detection of Coffea arabica adulteration with cheaper Coffea robusta. A total of 33 arabica and robusta samples from different countries and producers were acquired in supermarkets in Poland. The model adulterant mixtures were constructed by spiking the arabica with robusta samples at levels ranging from 0 to 100 %, at 10 % intervals (w/w). 6% (w/v) water (95˚C) extracts were prepared and after cooling were diluted 1:120 (v:v) with distilled water. The synchronous fluorescence spectra of the samples were acquired in a 10 mm fused-quartz cuvette within the excitation wavelength range of 240 to 700 nm for wavelength intervals of 60 and 80 nm.

In this experiment, Principal Component Analysis (PCA) was applied to reduce the number of variables and Successive Projections Algorithm (SPA) was applied to retain the most informative wavelengths from the spectra for further chemometric analysis. The number of wavelengths to be selected was given as input information for all wavelength intervals ($\Delta\lambda$= 60 and 80 nm). Multiple Regression Analysis (MLR) models were built separately for the data acquired at each wavelength interval ($\Delta\lambda$ = 60 and 80 nm). The root mean square errors of made it possible to asses and confirm the prediction ability of the models. Linear Discriminate Analysis (LDA) was applied to classify samples of arabica, robusta samples and their mixtures. The lowest errors of MLR calibration and prediction models did not exceed 4% [w/w] while classification errors for LDA models did not exceed 5% [w/w].

P-101

EXPLORATORY ANALYSIS AND IDENTIFICATION OF PARAMETERS RELATED TO THE ORGANOLEPTIC QUALITY OF CALÇOTS (*Allium cepa* L.) USING MULTIVARIATE METHODS

E. Simón [12], M. Alcalà [1], J. Casals [2], J. Simó [2]

[1] Applied Chemometrics Research Group, Department of Chemistry, Faculty of Sciences, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain
[2] Miquel Agustí Foundation, Polytechnic University of Catalonia, Campus del Baix Llobregat, Carrer Esteve Terrades 8, Edifici D4, 08860 Castelldefels
estela.simon@e-campus.uab.cat

'*Calçots*' are the floral stems of second-year onion (*Allium cepa* L.) resprouts of the '*Blanca Tardana de Lleida*' landrace. They are a typical product of Catalonia (NE Spain), recognized by the European Union with a Protected Geographic Quality (PGI) label. Since 2005, a plant breeding program has been conducted in order to improve the organoleptic profile and agronomic behavior of the cultivated varieties. Large amounts of multi-year and multi-parametric data were obtained, with a correlation structure that must be carefully analyzed. The objective of this work is to perform an exploratory analysis of the multivariate data sets obtained by plant breeders, in order to highlight the patterns within and between different variables. Furthermore, we want to generate efficient calibration models to predict chemical and sensory parameters to facilitate the reference laboratory analysis, which are expensive and time consuming. The data set consisted in 147 samples of calçots phenotyped in diverse locations in Spain in a period of 5 years. The number of variables was 12, including morphologic (5), chemical (5) and sensory (4) parameters. The data was analyzed using multivariate data analysis methods, such as Principal Component Analysis (PCA), Multiple Lineal Regression (MLR), Principal Component Regression (PCR) and Partial Least Squares (PLS). Generally, significant correlations ($p<0.05$) were found between the morphologic, chemical and sensory traits. From the obtained results, it stands out a) the positive correlation between the dry matter and the soluble solids content (ºBrix), b) the positive correlation of the sensory attribute overall assessment with ºBrix and dry matter, and c) the negative correlation with the traits length of the edible part, weight and titratable acidity. The sensory attribute off flavor had an opposite behavior than overall assessment. In general, the samples of the years 2013 and 2014 present higher values of ºBrix and dry matter, lower values for ash content and titratable acidity and a better sensory profile, according to panelist ratings (higher sweetness and overall assessment, and lower fiber perception, and off flavor). Lineal least squares regression model between dry matter and ºBrix was performed to develop a prediction model for the dry matter, and a significant prediction model was obtained ($R^2=0.88$). The relationship between sensory traits and chemical and morphologic parameters was modeled by partial least squares regression (PLSR). Reasonable models were obtained for the prediction of the overall assessment and off flavor. The goodness of the model fit was tested using the determination coefficient (R^2) and the root mean square error (in sensory trait units), which was termed RMSEC for calibration and RMSEP for prediction. For overall assessment attribute, R^2calibration model=62%, R^2validation model=54%, RMSEC=1.45, RMSEP=1.09. For the off flavor attribute, R^2calibration model=61%, R^2validation model=42%, RMSEC=1.04, RMSECV=1.18, RMSEP=1.22. With this work we pretend to guide the plant breeders of calçots to make a preliminary screening to multivariate data sets obtained in their experiments. And also, these methods allow to predict sensory attributes of the calçots, in an easy and inexpensive way.

P-102

DETERMINATION OF BIOACTIVE COMPOSITION AND PROPERTIES OF RED CABBAGE EXTRACTS BY INFRARED SPECTROSCOPY AND CHEMOMETRIC METHODS

I.R.N. de Oliveira[1], J.V. Roque[2], M.P. Maia[3], R.F. Teófilo[2], P.C. Stringheta[3]

[1]Instituto de Ciências Agrárias, Universidade Federal de Viçosa, Rio Paranaíba, Brazil.
[2]Departamento de Química, Universidade Federal de Viçosa, Viçosa, Brazil. email (rteofilo@gmail.com)
[3]Departamento de Tecnologia de Alimentos, Universidade Federal de Viçosa, Viçosa, Brazil.

Red cabbage (*Brassica oleracea*) is a vegetable rich in a large number of bioactive substances, including anthocyanins, that are natural pigments which belongs to the phenolic compounds class. Regular consumption of foods containing these compounds is associated with health benefits such as inhibition of blood cholesterol accumulation, prevention of some cancers and diabetes, among other benefits [1,2]. Thus, the quantification of bioactive substances and determination of red cabbage properties is very interesting. However, conventional methods are time consuming, destructive and consume large amounts of chemical reagents. So, the development of a rapid method for the determination of total and monomeric anthocyanins, total polyphenols and antioxidant capacity in liquid extract of red cabbage using near infrared (NIR) and medium (MID) spectroscopy, allied with partial least squares regression (PLSR) was the main of this work. Red cabbage extracts were obtained with ethanol solution 70% (v/v) and concentrated to 9º Brix. Then, the concentrated extract was diluted for determination of total and monomeric anthocyanins, total polyphenols and antioxidant capacity by the 2,2-diphenyl-1-picryl hydrazine (DPPH) and 2,2'-azinobis-3-ethyl-benzothiazoline-6-sulfonate (ABTS) methods. NIR (10000-4000 cm^{-1}) and MID (4000-650 cm^{-1}) spectra were obtained from the diluited extracts using the transflectance tecnique. These solutions were used to build multivariate calibration models using PLSR and a method of variable selection, the ordered predictors selection (OPS) [3]. The diluted extracts presented 21.95 to 595.72 mg L^{-1} of total anthocyanin, 23.08 to 588.24 mg L^{-1} of monomeric anthocyanin, 42.94 to 1073.39 mg L^{-1} of total polyphenols, and the antioxidant activity by ABTS and DPPH ranged from 0.20 to 5.90 and 0.23 to 4.90 µM Trolox mL^{-1}, respectively. Altogether, PLS-OPS models showed the best prediction results with correlation coefficients (*Rc*) > 0.99, ratio performance deviation (*RPD*) > 5.80 and mean relative error (% *E*) < 10%. The NIR spectra promoted the best models for monomeric anthocyanin and antioxidant capacity, by ABTS and DPPH methods, with root mean square error of prediction (*RMSEP*) of 32.23, 0.21 and 0.13, respectively. Meanwhile, the MID spectra were more predictive for total anthocyanins and total polyphenols, with *RMSEP* of 14.31 and 34.62, respectively. These results indicate that the models obtained are reliable for predictions of bioactive substances and antioxidant capacity quickly and non-destructively, and not spending chemical reagents.

Acknowledgement: CNPq, CAPES, UFV, FAPEMIG

References:

[1] J. Lachman and K. Hamouz, Plant, Soil Environ., 51, 11, 477 (2005).
[2] M.J. Kruger, N. Davies, K.H. Myburgh, S. Lecour, Food Res. Int., 59, 41 (2014).
[3] R.F. Teófilo, J.P.A Martins, M.M.C. Ferreira, J. Chemom. 23, 32 (2009).

P-103

UNEQUIVOCAL IDENTIFICATION AND QUANTIFICATION OF BISPHENOL-A IN A MIGRATION TEST FROM FOOD POLYCARBONATE CONTAINERS USING FLUORESCENCE SPECTROSCOPY AND PARAFAC

M.L. Spagnuolo[1], F. Marini[1], L.A.Sarabia[2], M.C. Ortiz[3]

[1]Department of Chemistry, Sapienza University of Rome, Ple. Aldo Moro 5, 00185 Rome, Italy.
[2]Mathematics and Computation, University of Burgos, Pza. Misael Bañuelos s/n, 09001 Burgos, Spain.
[3]Department of Chemistry, University of Burgos, Pza. Misael Bañuelos s/n, 09001 Burgos, Spain.
e-mail: mcortiz@ubu.es

Bisphenol A (BPA) is a chemical commonly used in manufacturing polycarbonate and epoxi resins, materials that are used in the production of food and beverage containers. BPA may be transferred to food and beverages by migration processes from these containers. Several regulations ban the use of BPA in baby bottles and a specific migration limit of 600 µg/kg has been established for other polycarbonate containers[1].

To determine the quantity of BPA as potencial migrant from polycarbonate cups, in this work, a fast and inexpensive procedure have been implemented by using Parallel Factor Analysis (PARAFAC) to model excitation-emission fluorescence signals. Requeriments in european regulations [2] have been followed for the unequivocal identification of BPA.

First, the figures of merit have been established: linearity ranged to 0-640 µL-1, repeatability equal to 2.04 % (at 320 μL^{-1}) and detection capability, CCbeta, equal to 40.6 μL^{-1}, when probabilities of false positive and false negative are fixed at 0.05.

Later, the effect of four factors on several experimental responses was analized. Factors were: i) temperature set during the experiments (20 or 30 ºC), ii) number of calibration standards (6 or 10), iii) excitation spectra recording mode (every 2 or 5 nanometers) and iv) spectra acquisition speed (500, 1000 or 1500 nm/min). In addition, it has been considered that one interaction between the last two factors could exist. By means of a D-optimal design the experimental effort has reduced from 24 to 9 experiments keeping the VIF below 1.8 for the seven coefficients of the proposed model.

Four answers have been considered in this experiment: analytical sensitivity, CCbeta and correlation between spectra obtained with PARAFAC and the true spectra of BPA (unequivocal identification of excitation and emission profiles). To evaluate these responses, a calibration for each experimental conditions has been built. The whole procedure is robust for the four responses and nor significant factors neither interation exist.

Finally, in the experimental conditions that gave the best responses (20ºC, 10 calibration standard, recording each 2 nm and 1000 nm/min) the migration tests were realized to a couple of polycarbonate cups. BPA was extracted from these cups into simulant B (3% w/v acetic acid in water) at 85ºC for 10 hours. Data tensor size for this analysis is (31x20x56) being 31 the number of samples, 20 the excitation wavelengths and 56 emission wavelengths. The concentration of BPA found was: 688.7μL^{-1} (n=4) and 706.1 μL^{-1} (n=5) with SDR equal to 5 y 2% respectively. The appearance of chemicals in the migration process, makes it is essential the use of PARAFAC to obtain the complete especificity and as consequence allowing the unequivocal identification an quantification of BPA migrated.

Acknowledgement: The authors thank the financial support provided by the MINECO (CTQ2014-53157-R) and FEDER funds.

References:

[1] Regulation (EU) No 10/2011 on plastic materials and articles intended to come into contact with food.
[2] Guidelines for Performance Criteria and Validation Procedures of Analytical Methods Used in Controls of Food Contact Materials, first ed., 2009 (EUR 24105 EN).

P-104

SREENING METHODS FOR TESTING THE MIGRATION OF FORMALDEHYDE FROM MELAMINE KITCHENWARE USING MULTIVARIATE CALIBRATION

Z. Cabello[1], L.A.Sarabia[2], M.C. Ortiz[1]

[1]Department of Chemistry, University of Burgos, Pza. Misael Bañuelos s/n, 09001 Burgos, Spain.
[2]Mathematics and Computation, University of Burgos, Pza. Misael Bañuelos s/n, 09001 Burgos, Spain.
e-mail: mcortiz@ubu.es

The European legislation has stablished a specific migration limit (SML) at 15 mg/kg for formaldehide [1]. Formaldehyde resins are used in the fabrication of melamine kitchenware and has become well know that formaldehide can lead to cancer. To determine the quantity of formaldehyde, as potencial migrant,from melamine glasses a fast and inexpensive screening method have been implemented by using mutivariatle calibrarion and UV-visible data. This method is a variant of the one proposed in the technical guidelines (EUR 24815 EN 2011) [2] to determine formaldehyde by UV-visible spectroscopy. This method uses a soft calibration based on partial least squares (PLS) using the full spectrum in the range from 350 nm to 510 nm instead of a single wavelength to build the calibration model. The advantage of using a multivariate calibration as PLS is (through Q and T^2 indices) to detect the presence of interferents, whose absorbances are taking place in the same spectral range as the formaldehyde. Therefore the applicability of the method is guaranteed since the method enables to detect a possible lack of specificity.

Eleven standards of calibration, ranged 0-5 mg/L, has been used. By means of a cross-validation step it was concluded that two latent variables were required, explaining 99.97% of the variance of Y-block. The accuracy line $y_{cal}=0.00075+ 0.99968x_{true}$ (PLS calculated concentration versus true concentration) let us conclude that the method is umbiased. Efficiency of UV-visible-PLS procedure in terms of the decision limit (CCα) and detection capability (CCβ) evaluating the probabilities of false non-compliance (α) and false compliance (β) were 0.065 and 0.127 mg/L respectively when α and β were fixed at 0.05.

The migration tests were conducted of with simulant B 3% (w/v) acetic acid:water (70ºC during 2 hours). The quantity found after 3th migration testing, as is recomended in the guidelines used in controls of food contact materials, in a glass of melamine was 0.715 mg/L. A recovery study was performed by spiking 1 and 2.5 mg/L concetration in two replicates of melamine glass acetic-migrate; the results were 93.6% and 103%. Migration of formaldehyde from the glass, after seven migration tests and carried out with fresh simulant each time; provided an average content of 0.66 mg/L (n=21). All results were above the decision limit of the method (0.065 mg/L) but did not exceed the specific migration limit (15 mg/kg).

A migration cinetic curve was built, with the data of seven consecutive migrations (70ºC, 2 hours), being the explained R^2 larger than 0.99 in all the cases.

Acknowledgement: The authors thank the financial support provided by the MINECO (CTQ2014-53157-R) and FEDER funds.

References:

[1] Regulation (EU) No 10/2011 on plastic materials and articles intended to come into contact with food.
[2] Guidelines for Performance Criteria and Validation Procedures of Analytical Methods Used in Controls of Food Contact Materials, 2011 (EUR 24815 EN).

P-105

DEVELOPMET OF DIFFERENT EXTRACTION METHODS OF PHENOLIC COMPOUNDS FROM RICE (*Oryza sativa*) GRAINS

W. Setyaningsih [1.2], M. Palma [2], C.G. Barroso [2]

[1]*Department of Food and Agricultural Product Technology, Faculty of Agricultural Technology, Gadjah Mada University, Jalan Flora, 55281, Yogyakarta, Indonesia.*
[2]*Department of Analytical Chemistry, Faculty of Sciences, University of Cadiz, IVAGRO, Campus de Excelencia Internacional Agroalimentario (CeiA3), Campus del Rio San Pedro, 11510, Puerto Real, Cádiz, Spain.*
widiastuti.setyaningsih@ugm.ac.id

Besides being a notable source of calories for more than half of the world's population, cultivated rice (*Oryza sativa*) provides beneficial effects for human health through several bioactive compounds, including phenolics. In general, most of these compounds are sensitive and certainly susceptible to degradation. Therefore, novel analytical extraction techniques i.e. Microwave-assisted Extraction (MAE), Ultrasound-assisted Extraction (UAE) and Pressurized Liquid Extraction (PLE) are essential to have proper sample preparation techniques and these approaches have been used to increase the extraction efficiency for various bioactive compounds.

The development of a single optimized extraction process for phenolic compounds from foods is complicated due to their structural diversity. Additionally, there were different number of factors as well as different working range in the Design of Experiments (DOE) domain when optimizing different extraction techniques. For this reason, chemometric technique appears to be a viable option for efficiently optimizing the multi factors that affect the extraction performance. In this paper, several different chemometric techniques including Fractional Factorial Design (FFD), Central Composite Design (CCD) and Box-Behnken Design (BBD) were used for the development of extraction of phenolic compounds from rice grains.

The optimised and validated methods were then applied to assess the level of phenolic compounds in a wide range of rice varieties including pigmented and non-pigmented rice. The levels of individual phenolic compounds were comparable applying different extraction techniques. However, a moderate difference in the composition of phenolic compounds in rice extracts recovered by different extraction techniques was observed. This is reasonable when taking into account the different condition of extractions in addition to distinctive rice origins as real samples. It can be concluded from the results that MAE; UAE and PLE under optimum conditions can be considered as a powerful tool for the determination of phenolic compounds from a wide variety of rice grains.

Acknowledgement: W. S. is grateful to the CIMB Foundation for a Ph.D. studentship through the CIMB Regional Scholarships 2012.

P-106

SYNTHESIS AND CHARACTERIZATION OF CARBOXY METHYL CELLULOSE (CMC) FROM SNAKE FRUIT (*Salaca edulis* Reinw) "PONDOH SUPER" KERNEL

Sri Anggrahini *, Djagal Wiseso Marseno*, Agus Setiyoko*, Amalia Wahyuningtyas*

**Department of Food and Agricultural Product Technology, Faculty of Agricultural Technology, Universitas Gadjah Mada, Jalan Flora No. 1 55281, Yogyakarta, Indonesia*

**Corresponding author, e-mail : sri_anggrahini2006@yahoo.com*

The aim of this study was to determine the optimum conditions of synthesis CMC from "Salak Pondoh Super" kernel. Some factors that likely influence the synthesis were concentration of NaOH solution , NaMCA addition, and the reaction temperature based on the degree of substitution (DS) as theresponses.

Synthesis of CMC was optimized using completely randomized design. The result then was characterized by several parameters including water content, viscosity, purity, Water Holding Capacity (WHC), Oil Holding Capacity (OHC), lightness, crystallinity, and FT-IR spectra.

Optimization was achieved by the use of 15% NaOH solution, 5 gram NaMCA per 5 gram cellulose and reaction temperature of 55ºC. The characteristics of the optimized CMC were DS 0.825, purity 90.86%, water content of 7.16 (% wb), viscosity 3.86 cps, 142.72 yield (% db), WHC 2.37 (g/g), OHC 2.31 (g/g), lightness 78.48, and crystanillity 32.69%. The FT-IR spectra MC was similar with the CMC Standard.

Keywords: Cellulose, CMC, "Salak Pondoh Super", Kernel

P-107

Pandan Leaves Extract (*Pandanous amaryllifolius Roxb.*) as Natural Antioxidant in Selected Vegetable Oils during Accelerated Storage

Andriati Ningrum[a], Matthias Scheiner[b]

[a]Department of Food Science and Agricultural Product Technology, Faculty of Agricultural Technology, Gadjah Mada University, Yogyakarta, 55281
[b]Institute of Food Science, BOKU, Vienna, Austria 1190
Email : andriati_ningrum@ugm.ac.id

The application of pandan leaves extract (PLE) were observed as a natural antioxidant in red palm oil (RPO) and soybean oil (SO) using accelerated oxidation at 70ºC for 3 days. DPPH (1,1-Diphenyl-2-picrylhydrazyl) scavenging activities, β-carotene linolenic assay, rancimat analysis were performed to analyze the effect of PLE and time incubation (0 day and 3 days storage) to the antioxidant activities in RPO and SO. During the accelerated storage, antioxidant activities of PLE in RPO and SO are commonly significantly difference ($p < 0.05$). The PLE extract (optimum concentration is 0.18%) gave the best result to retard oil oxidation in RPO and SO. These results gave a promising application of pandan leaves extract as a good natural alternative to be applied in several vegetable oils.

Acknowledgment : The authors wish to thank to OeAD for Technology Grant Scholarship

Abbreviation: PLE : Pandan leave extract, RPO: Red Palm Oil; SO: Soybean Oil

P-108

ANTIOXIDANTS OF BEVERAGE PRODUCT FROM MANDARIN BY-PRODUCTS

M. Guamán-Balcázar [1], M. Ludeña [1], J. Maldonado [1], W. Setyaningsih [2]

[1]Department of Agricultural Sciences and Food, Universidad Técnica Particular de Loja, Loja, Ecuador
[2]Department of Food and Agricultural Product Technology, Faculty of Agricultural Technology, Gadjah Mada University, Jalan Flora, 55281, Yogyakarta, Indonesia.
widiastuti.setyaningsih@ugm.ac.id

Food products from fruit and vegetables contain a significant amount of antioxidants. These compounds are of special interest due to their capacity to reduce the hazard caused by free radicals, and have been associated with lowered risks of cardio vascular diseases and other illnesses related to oxidative stress. Practically the natural antioxidants are attained through consumption of food products derived from fruits, vegetables. Therefore, the objective of this study was to determine the stability of antioxidants during the production of beverage from Mandarin liquid by-product.

The level of antioxidants of the beverage was evaluated at different stages of production processes i.e. pasteurization and storage. The beverage was pasteurized at 93 °C for 30 s, then was stored for 2 months at a temperature ranging from 0 to 5 ºC. The total phenolic compounds were determined by the Folin Ciocalteau method, while the antioxidant activity was assessed using ABTS, DPPH and FRAP methods.

The measurement result of phenolic compounds and antioxidant capacity by DPPH method showed that the compounds were stable during pasteurization process. In contrast, the antioxidant capacity of the studied compounds measured by ABTS and FRAP methods showed a slight reduction due to pasteurization. Similar trend was also observed after being stored for two months wherein the antioxidant capacity was reduced up to 6.3, 11.9 and 19.7 % by DPPH, ABTS and FRAP respectively. The degradation of phenolic compounds as much as 9.3% was also observed during the beverage production and storage. However, the amount of total phenolic compounds and their antioxidant capacity was higher compare with the beverages sold at the market in Loja, Ecuador.

P-109

CHARACTERIZATION AND CLASSIFICATION OF NATURAL PRODUCTS AND PHARMACEUTICALS WITH CHEMOMETRIC TECHNIQUES USING LC-HRMS METABOLOMIC AND POLYPHENOLIC FINGERPRINTING

M. Hidalgo [1], S. Barbosa [1], O. Núñez [1,2,3], J. Saurina [1,2], S. Hernández-Cassou [1,2], L. Puignou [1,2]

[1]Department of Analytical Chemistry, University of Barcelona. Martí i Franquès 1-11, E-08028, Barcelona, Spain.
[2] Research Institute in Food Nutrition and Food Safety, University of Barcelona, Recinte Torribera, Av. Prat de la Riba 171, Edifici de Recerca (Gaudí), E-08921 Santa Coloma de Gramenet, Barcelona, Spain.
[3] Serra Húnter Fellow, Generalitat de Catalunya, Spain.
mhidalse8@gmail.com

American cranberries (*Vaccinum macrocarpon*) and its derived products have shown some beneficial health effects associated to their polyphenolic content such as their capacity to help prevent infections of the urinary tract [1]. That is the reason why many cranberry-based extracts have recently appeared in the market as pharmaceutical preparations and, lately, it has been suspected that some of these preparations do not contain the necessary bioactive polyphenols (i.e., A-type proanthocyanidins, PACs) and actually come from other fruit extracts like grapes or blueberries. The fact that only A-type PACs have the necessary bioactive capacity shows the importance of developing analytical methods to study and characterize fruit-based extracts to achieve correct identification regarding the fruit of origin.

The aim of this work was to find a suitable method to characterize, classify and authenticate natural and pharmaceutical berry-based products employing untargeted (metabolomic fingerprints) and targeted (polyphenolic profiles) analytical approaches.

For that purpose, more than 100 cranberry-, grape-, blueberry- and raspberry-based natural products as well as cranberry-based pharmaceutical products were analyzed by UHPLC-HRMS/MS (Q-exactive Orbitrap, Thermo Fisher Scientific). Separation was performed by C18 reversed-phase chromatography in a Supelco Ascentix Express (150x2.1 mm, 2.7 µm) porous-shell column using 0.1% formic acid water and acetonitrile solutions as mobile phase. Full scan MS (m/z 100-1,500) at a resolution of 70,000 FWHM (full-width at half-maximum) and data dependent MS/MS product ion spectra at a resolution of 17,500 FWHM were used to obtain the metabolic fingerprints and the polyphenolic profiles. Blank acetonitrile samples and a mixture of all the product samples were employed as quality controls. On an untargeted approach, full scan MS raw data were considered as metabolic fingerprints to be treated by chemometric techniques such as principal component analysis (PCA). After correcting and improving MS signal quality by using specific filters, full scan MS metabolic fingerprints allowed the classification of the products depending on the fruit of origin. On a targeted approach, polyphenolic profiles were employed for characterization. MS data was processed by ExactFinder v2.0 software (Thermo Fisher Scientific) by applying a customized target database list of polyphenols. The most remarkable polyphenols were identified and selected to achieve characterization of natural products and pharmaceuticals by chemometric techniques.

Acknowledgement: Projects CTQ2014-65324, 2014 SGR-377 and 2014 SGR-539.

References:

[1] I. Tarascou, J. Mazauric, E. Meudec, J. Souquet, D. Cunningham, S. Nojeim, V. Cheynier, H. Fulcrand, Food Chem., 128, 802 (2011).

P-110

LC-HRMS TARGETED AND UNTERGETED APPROACHES IN THE CHARACTERIZATION AND CLASSIFICATION OF VEGETABLE OILS

D. Filatova [1], O. Núñez [1,2,3], J. Saurina [1,2], E. Moyano [1]

[1]Department of Analytical Chemistry, University of Barcelona. Martí i Franquès 1-11, E-08028, Barcelona, Spain.
[2] Research Institute in Food Nutrition and Food Safety, University of Barcelona, Recinte Torribera, Av. Prat de la Riba 171, Edifici de Recerca (Gaudí), E-08921 Santa Coloma de Gramenet, Barcelona, Spain.
[3] Serra Húnter Fellow, Generalitat de Catalunya, Spain.
daria.filatova@gmail.com

Olive oil has been produced for about 6,000 years, but in the last thirty years there has been a growing interest in the use of vegetable oils in cooking because of a greater knowledge of Mediterranean food and an awareness of the healthy virtues of a Mediterranean diet, and particularly olive oil. The increasing popularity of olive oils is mainly attributed to its high content of oleic acid, which may affect the plasma lipid/lipoprotein profiles, and its richness in phenolic compounds, which act as natural antioxidants and may contribute to the prevention of several human diseases. Nowadays there are some concerns that some of vegetable oil products sold in the market labeled as derived only from olives could contain vegetable oils which are coming from other fruits or seeds such as sun-flower, soy or corn. Therefore it is important to develop analytical method for the correct authentification of vegetable oils.

This work was aimed at exploring suitable methods to characterize and classify vegetable oils according to the fruit/seed of origin based on untargeted (metabolomic fingerprints) and targeted (polyphenolic profiles) analytical approaches.

For that purpose, 70 Spanish vegetable oils (olive, sun-flower, soy, corn) were analyzed by UHPLC-HRMS/MS (Q-Exactive Orbitrap, Thermo Fisher Scientific). Separetion was performed by standard gradient elution in a Supelco Ascentix C18 (150x2.1 mm, 2.7µm) porous-shell column using 0.1% formic acid water and acetonitrile solutions as mobile phase. Full scan MS (*m/z* 100-1,500) at a resolution of 70,000 FWHM (full-width half-maximum) and data dependent MS/MS product ion spectra at a resolution of 17,500 FWHM were used for both metabolomic fingerprint and polyphenolic profile acquisition. Samples were processed by liquid-liquid extraction with a mixture of ethanol/water, and a clean-up step with hexane. Data obtained was treated by means of chemometrics following both untargeted and targeted approaches.

On untargeted approach, full scan MS raw data was considered as metabolomic fingerprint to be treated by principal component analysis (PCA). After correcting and improving MS signal quality by using specific filters, full scan MS metabolomic fingerprints allowed oil classification depending on its type of fruit or seed.

On the targeted approach, polyphenolic signals were employed for oil characterization. For that purpose, MS data was processed by ExactFinder v2.0 software (Thermo Fisher Scientific) by applying a customized target database list of polyphenols. Retention time, accurate mass errors, isotopic pattern matches and product ion scan spectra were used to confirm the identity of compounds when necessary. After further processing of polyphenolic profiles by removing signals with peak scores lower than 0.65 the most remarkable polyphenols of each oil type were identified and selected to achieve characterization of Spanish vegetable oils by PCA.

Acknowledgement: Projects CTQ2014-65324, 2014 SGR-377 and 2014 SGR-539.

P-111

NIR detection of foreign traces in powder foods. An aproach independent from the spectral equipment

T. R. Cuadrado [1], P. Barreiro Elorza [2], L. Lleó García[2], Facundo Ruiz[1], Pablo Delgado Sánchez[1]

[1] Facultad de Agronomía y Veterinaria, UASLP, San Luis de Potosí, México
[2] LPF_TAGRALIA Departamento de Ingeniería Agroforestal, UPM CEI Moncloa, Madrid, España
e mail (pilar.barreiro@upm.es)

The effective detection of the presence of trace amounts of allergens in processed foods is of crucial importance in the food industry. Traces of allergens can cause adverse reactions, and may become very serious in certain sectors of the population. Previous work based on NIR spectroscopy (considering local spectra measurements and also hyperspectral images) propose models capable of detecting the presence of peanut and another dried fruits. Mentioned models include the band 1200 nm corresponding to the fats.
The objective of this work is to detect different traces of foreign and allergen foods (i.e. gluten, fats, nuts …etc) and to carry out a transfer of calibration between NIR spectroscopy equipments belonging to different laboratories in different countries (Spain-Mexico), With this work is expected to be able to apply and develop identification algorithms useful for all tested instrumentation.
Two NIR equipments have been used: Hamamatsu PMA11, C8147-34, in the range between 896 and 1686 nm (LPF-Tagralia UPM, Madrid), and an Ocean Optics model NIR512, in the range of infrared near to the 900 - 1700 nm equipment (Faculty of Sciences of the UASLP; México). For Hamamatsu, five different varieties of peanut from different geographic origin and treatment were analyzed as well as milk powder, cocoa, wheat flour, and peanut from different Spanish trademarks. The samples measured with the Ocean Optics spectrophotometer included commercial samples from Mexico (, milk powder, cocoa, wheat flour, chocolate and peanut), as well as samples from Spain (24 different kind of flour, 33 types of nuts, 2 different types of almonds sesame, 2 types of peanut, pinion, corn seed, pistachio, Hazelnut).
A comparative analysis of the sets of spectra (raw and normalized) has been faced as an initial step; different preprocessing procedures and multivariate models (PCA, PLS ...etc) in both spectra sets have been defined and tested. In addition a piecewise direct calibration transfer procedure is to be performed considering Hamamatsu as master and Ocean Optics as slave. A specific analysis on the equipment platform is faced regarding spectral sensitivity.

Acknowledgement: Funded by Comunidad de Madrid, S2013/ABI-2747 (TAVS-CM) and Fondos Estructurales de la Unión Europea

References:
[1] P. Mishra, A. Herrero Langreo, P. Barreiro Elorza, J.M. Roger, B. Diezma Iglesias, N. Gorretta, L, Lleó García, (2015). Hyperspectral to multispectral imaging for detection of tree nuts and peanut traces in wheat flour. Journal of Spectral Imaging, Vol 4 (1), Initial page. 1 (2015).
A.B. Smith, C.D. Brown, J. Good Results, Vol, Initial page (Year).
[2] S. Ghosh, T. Cuadrado Domínguez, B. Diezma Iglesias, L. Lleó García, P. Barreiro Elorza, T. García Lacarra, J.M. Roger, VIS/NIR spectral signature for the identification of peanut contamination of powder foods. CIGR Journal, Vol 17 (2); Initial page 310 (2015).

P-112

CHEMOMETRICS AND UNTARGETED METABOLITE PROFILING OF DIFFERENT BEER BRANDS

E. Szymańska[1,2], A. Suppers[1], M. Schoot[1];
E. Koussissi[3]; E.R. Brouwer[3]; L.M.C. Buydens[1]

[1]Radboud University, Institute for Molecules and Materials, P.O. Box 9010, 6500 GL Nijmegen, The Netherlands
[2]TI-COAST, Science Park 904, 1098 XH Amsterdam, The Netherlands
[3]Heineken Supply Chain BV, P.O. Box 510, 2380 BB Zoeterwoude, The Netherlands
E.Szymanska@science.ru.nl

Beer is a highly complex beverage which contains more than 800 organic compounds. Different brands of beers have distinctive tastes and quality control parameters which originate from the mineral content of the water and the types of ingredients used as well as differences in the brewing methods. In modern analytical chemistry labs extensive metabolite profiles of beer samples can be produced. However, so far, due to lack of fast and reliable data analysis procedures, these profiles are not employed in the assessment of beer parameters.

In this study, we have applied and compared different chemometric tools in the assessment of beer parameters. Untargeted metabolite profiling of 67 different beer brands was obtained with gas chromatography – mass spectrometry (GC-MS) and liquid chromatography – mass spectrometry (LC-MS) including both positive and negative ion mode measurements. Beer parameters included quality control parameters such as bitterness, foam stability and sulphur dioxide content. GC-MS and LC-MS data sets were successfully preprocessed, using alignment, baseline correction and clustering of molecular features belonging to the same metabolite. In the exploratory analysis, chemometric methods as PCA, sparse-PCA, Multidimensional Scaling (MDS) and t-distributed Stochastic Neighbor Embedding (t-SNE) were compared to visualize differences between beer brands in three different data sets: GC-MS, LC-MS-positive and LC-MS-negative. Furthermore, prediction models of beer parameters, i.e. bitterness, were obtained using either one of available data sets or their combinations. Chemometric tools as PLS, sparse-PLS and different data fusion approaches: low-level, mid-level, high-level were used and their performance compared.

This study demonstrates a great potential of chemometric analysis in the assessment of beer parameters. It yields an optimized data analysis procedure to fully exploit and benefit from a plethora of information in untargeted metabolite profiling of beer.

Acknowledgements: This research received funding from the Netherlands Organization for Scientific Research (NWO) in the framework of the Technology Area COAST.

P-113

SPECTROSCOPIC FINGERPRINT OF PALE BEERS LINKED TO SENSORY ANALYSIS AND CONSUMERS PREFERENCES

N. Cavallini[1,2], M. Cocchi[1], R. Bro[2], H. da Silva Friis[2], F. Savorani[3]

[1]Dipartimento di Scienze Chimiche e Geologiche, Università di Modena e Reggio Emilia, Via Campi 103 – 41125 Modena
[2]Department of Food Science, University of Copenhagen, Rolighedsvej 30- DK-1958 Frederiksberg C.
[3]Department of Applied Science and Technology, Politecnico di Torino, Corso Duca degli Abruzzi, 24 - 10129 Torino (TO)

During the last decade the awareness of consumers and society in general towards all aspects that concern food consumption has strongly increased. Ethics, sustainability, health, safety, quality, tradition are now everyday words, and communication and marketing are following more and more the trends that these terms represent. Research in food chemistry area has mainly focused on chemical analysis and characterization to contribute to fundamental issues such as food safety and quality, nutritional and health requirements.
The present work is part of a larger project, which is aimed to take a step beyond the aforementioned approach. The fundamental idea is to use our analytical chemistry expertise to build new tools to aid consumers when choosing foodstuff (and have proper knowledge of it) and producers to meet consumer expectations, using food quality as a driver.
To this aim the proposal is to use analytical spectroscopy to capture salient features of foodstuff (fingerprint) and build a reference database that can be efficiently searched through multivariate data analysis tools and, be likely in the near future, linked to applications for mobile smart devices implemented for consumers inquires. At the same time consumers' choice may be oriented by showing how products of similar categories cluster according to different criteria.

As a first benchmark to develop these ideas a survey on beer is presented. One hundred samples of light beer (i.e. no stout or dark beers have been considered) differing by brewery, alcohol content, yeast, brew type, etc., were collected. This work is focused on Vis-NIR, NMR and sensory data, with the aim of establishing a link between the "objective" information of the spectroscopic fingerprint and the "subjective" world of consumers' assessments. The latter is represented by online beer ratings and for a reduced number of samples by sensory reports made by a panel of experts.
Different combinations of decomposition methods, e.g. PCA, ICA, MCR and clustering (both linear and non-linear methods) were used to extract relevant information and as basis for data-fusion techniques to integrate the chemical information.
PLS regression allowed establishing a link between spectral fingerprint/information and consumer preferences as expressed by ratings (www.ratebeer.com) aroma, appearance, taste, palate, overall. Furthermore, also the mostly used descriptive words used by consumer were codified, based on the approach described in [1]. This part of the work can contribute to build a "beer vocabulary", which can be very useful to develop further real-life applications.

References:

[1] M. Bevilacqua, Visualization of comprehensive data mining problems in gastronomy, chap. 10, *Development of chemometric approaches for ensuring food quality and safety*, Doctoral Thesis, Roma La Sapienza (2014)

P-114

DETERMINATION OF ALLURA RED DYE IN HARD CANDIES BY USING DIGITAL IMAGES OBTAINED WITH A CELL PHONE CAMERA AND N-PLS

B.G. Botelho, K.C.F. Dantas, M.M. Sena

Chemistry Department, ICEx,Universidade Federal de Minas Gerais, 31270-901 Belo Horizonte, MG, Brazil
marcsen@ufmg.br

Allura red (AR) is an azo dye used as adittive in beverages and candies. Artificial food dyes may cause some health problems in children, increasing the importance of their quality control. Previously, we developed a simple method for determing sunset yellow in beverages using a table scanner, RGB histograms and PLS [1]. The present work aims to determine AR in hard candies. However, it is difficult to get a fixed position for measuring solid samples in the scanner. Thus, we built a homemade equipment based on a cell phone camera. Other problem was the presence of textural variation in the images of solid samples, which leaded to large prediction errors when the former image analysis strategy [1] was applied. The color of the candies is homogeneous, but their surfaces were not perfectly smooth. In RGB decomposition, each pixel is individually decomposed, without considering its neighborhood. The adopted alternative was a more complex methodology [2], in which multi-way data were generated from the intensity of each pixel as function of its position (x,y) and treated with a Fast Fourier-Transform, in order to obtain congruent/trilinear data, a prerequisite for the subsequent calibration with N-PLS. 240 hard candies of 4 different flavors, containing only AR as dye, were analysed (23–79 $mg.kg^{-1}$). Reference method was based on sample extraction and UV/Vis spectroscopy [3]. Images were obtained with a Motorola RAZR XT910 cell phone, in ”jpg” and processed using MATLAB 7.13, PLS Toolbox 6.5 and Image Processing Toolbox 8.0. Samples were splitted in calibration (160) and validation sets (80) using Kennard Stone algorithm. The best model was obtained with 6 latente variables, with RMSEC and RMSEP of 3.8 and 5.8 $mg.kg^{-1}$, respectively. Fig. 1 shows the model fit. A complete multivariate validation was performed in the development of this simple, rapid, low cost and nondestructive method.

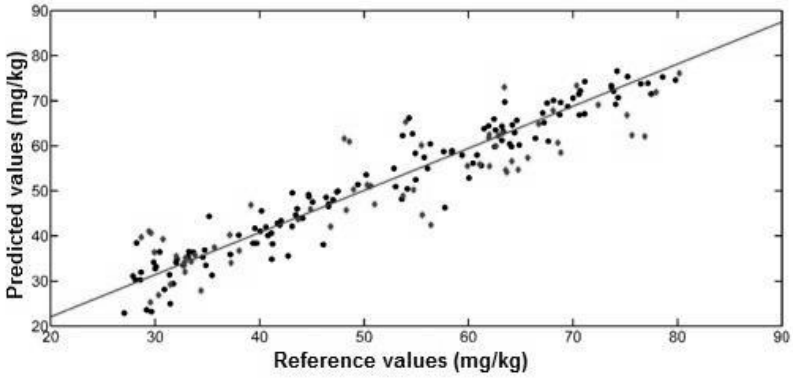


Figure 1. Plot of the reference *versus* predicted values

Acknowledgement: CNPq and FAPESP for financial support.

References:

[1] B.G. Botelho, L.P. Assis, M.M. Sena, Food Chem., 159, 175 (2014).
[2] J. Huang, H. Wium, K.B. Qvist, K.H. Esbensen, Chemom. Intell. Lab. Systems, 66, 141 (2003).
[3] IAL. Métodos Físicos-Químicos para Análise de Alimentos. IAL: São Paulo (2005).

P-115

PLS-REGRESSION AND ROC CURVES ON LIPID FINGERPRINTING DATA TO SET DETECTION THRESHOLDS OF OIL ADULTERATION

Tres A[1], Guardiola F[1], Codony R[1], Romero A[2], Caixach J[3], Vichi S[1]

[1]Department of Food Science and Nutrition-XaRTA-INSA, Faculty of Pharmacy, University of Barcelona, Barcelona, Spain
[2]Olive Production, Oil Processing and Nut Trees, IRTA, Constantí (Tarragona), Spain
[3]Mass Spectrometry Laboratory, IDAEA-CSIC, Barcelona, Spain.
Email: atres@ub.edu

Lipid fingerprints, for instance those based on acylglycerol profiles, are promising candidates to detect olive oil adulteration. Partial Least Squares Regression on the lipid fingerprinting data is a suitable strategy to develop models to predict the adulteration levels of suspicious samples. However, the lowest adulteration level detected as well as the values predicted for non-adulterated samples are to be known to avoid false negatives and false positives in future predictions. Here, we have calculated ROC curves with values predicted from PLS-Regression models to set thresholds for a decision criteria (adulteration or authentic) for olive oil adulteration.

The lipid fingerprints used have been obtained from a classical chromatographic technique (HPLC-RID analysis) and from a state-of-the-art high-throughput fast method (direct ESI-UHRMS). Sampling included genuine olive oils of three different commercial categories that were blended with soybean, sunflower and hazelnut oils at 5% and 10%, following a Latin Squares Design. Partial Least Squares Regression was calculated for both lipid fingerprints (HPLC-RID and ESI-UHRMS) using SIMCA sofware (v13.0, Umetrics AB, Umea, Sweden). Models were internally validated by leave-10% out cross-validation, and RMSEcv was used to optimize models. Models were developed independently for the adulteration with hazelnut and for the adulteration with soybean or sunflower.

ROC curves were calculated with IBM SPSS Statistics software (v 20.0). Prediction values corresponding to specificity =1 in ROC curves were set as the threshold above which no false positives were encountered, and thus, samples could be considered as adulterated. Prediction values corresponding to sensitivity =1 in ROC curves were considered the threshold below which no false negatives were encountered, and thus, samples could be considered as genuine (or at the most, adulterated at levels below 5%). When values for these two thresholds did not agree, intermediate values corresponded to an uncertainty range in which no assumptions on adulteration could be raised without the risk of facing false positives or false negatives. According to this decision criteria, the lipid fingerprinting from HPLC-RID analysis offered slightly better results as authentication tool than the UHRMS technique; however, it requires using the raw RID signal while the UHRMS uses discrete data.

P-116

APPLICATION OF FLUORESCENCE AND MULTIVARIATE CALIBRATION FOR SCREENING ANTIOXIDANT PROPERTIES OF APPLE JUICES

K. Włodarska, K. Pawlak-Lemańska, E. Sikorska

Faculty of Commodity Science, Poznań University of Economics and Business, al. Niepodległości 10, 60-967 Poznań Poland;
ewa.sikorska@ue.poznan.pl

We studied the feasibility to use the front-face fluorescence of the apple juices for the prediction of their antioxidant properties.

Phenolic compounds account for the metabolic activity and antioxidant properties of plant-based foods and for the putative health benefits to humans [1]. Moreover, polyphenols affect organoleptic properties of foods: their color, flavor, astringency, and hardness. The assessment of polyphenolic compounds and their antioxidant acitivity is one of the important aspects of quality studies of many foods.

Commercial apple juices were selected for study to cover the various ranges of juices available on the market: clear reconstituted from concentrate, cloudy produced from concentrate with added apple pulp, pasteurized naturally cloudy produced not from concentrate, and freshly squeezed juices.

The antioxidant properties of juices were evaluated on the basis of total phenolics, flavonoids, and antioxidant capacity.

The overall emission of complex, multi-constituent samples is usually characterized by the total fluorescence spectra (TFS), called also the excitation-emission matrices or total synchronous fluorescence spectra (TSyFS). The TFS are obtained by measurements of the emission spectra for a series of excitation wavelengths, thus providing a comprehensive characteristics of the absorption and fluorescence properties of all of the emitting components in the samples studied. The synchronous fluorescence spectroscopy is another technique used for the complex food samples. Here, fluorescence intensity is measured as a function of the simultaneously scanned emission and excitation wavelengths, with a constant offset between them ($\Delta\lambda = \lambda_{em}-\lambda_{ex}$). The emission of the apple juices studied was characterized by the TFS, the TSyFS and single-offset synchronous fluorescence spectra.

The regression analysis was performed using partial least squares methods (PLS1 and PLS2) on the unfolded TFS, TSyFS and on the single-offset synchronous fluorescence spectra. The *N*-way partial least squares regression (*N*PLS1 and *N*PLS2) was used for analysis of TFS and TSyFS arranged in three-way arrays.

The best regression models among parameters studied were obtained for prediction of the total flavonoids. For the particular analytical parameters the predictive ability of all the calibration models studied was generally similar. Therefore, the single-offset synchronous fluorescence spectra seems to be preferable for the screening analysis, instead of the entire TSF or TSyF spectra. Their application may significantly shorten the recording time and simplify the analysis, due to the reduced amount of spectral data.

References:

[1] V. I. Candrawinata, J. B. Golding, P. D. Roach, C. E. Stathopoulos, Food Reviews International, 29, 276 (2013)

P-117

DIFFERENTIATION OF SPIRIT BEVERAGE FROM AGAVE BY SOME VOLATILE COMPOUNDS USING GC-FID, DISCRIMINANT AND PCA ANALYSIS

J. Rodriguez-Campos[1], S.D. Bravo[1], O.Y. Lugo-Melchor[1], A. Escobedo-Reyes[1], M. Cedeño[2], F. Soltero[2]

[1]Centro de Investigación y Asistencia en Tecnología y Diseño del Estado de Jalisco A. C., Av. Normalistas 800 Guadalajara, Jalisco, México.
[2]Cámara Nacional de la Industria Tequilera, Calz. Lázaro Cárdenas 3289-5o piso, Guadalajara, Jalisco, México
email: ylugo@ciatej.mx; jarodriguez@ciatej.mx

The discriminant and principal component analysis (PCA) are important tools to classify and discriminate data. These techniques have been used to evaluate the generation of volatile organic compounds (VOCs) during different stages of Tequila production process [1] and to identify aging markers in Tequila [2]. Tequila 100%, Mezcal 100% and Bacanora 100% should be made from 100% agave sugars while Tequila and Mezcal can be made with agave sugars mixed with other sources of sugars such as sugarcane in accordance with the Mexican law. All these spirit beverages are easily adulterated although their label indicates that are made with agave. The Mexican regulations just evaluate the major compounds that can be identify and quantify by GC-FID, therefore is important to find minor compounds that can be classifiers from agave origin with simple techniques and robust analysis such GC-FID.
Ninety five samples of different spirit beverages (Tequila, Mezcal and Bacanora) were analyzed to determine the compounds regulated by Mexican law and four minor volatile compounds that could be classifiers from agave spirit beverages. The PCA separated Mezcal of other spirit beverages and discriminant analysis with the D1 (p=0.0000) classified the beverages into three groups: the first one Bacanora, second one Mezcal and Mezcal 100% and the last one Tequila and Tequila 100%.

[1] Prado-Jaramillo, N., Estarrón-Espinosa, M., Escalona-Buendía, H., Cosío-Ramírez, R. and Martín-del-Campo, S. (2015). Volatile compounds generation during different stages of the Tequila production process. A preliminary study. LWT - Food Science and Technology. 61 (2), 471-483.
[2] Muñoz-Muñoz,A., Grenier, A., Gutiérrez-Pulido, H and Jesús Cervantes-Martínez. (2008). Development and validation of a High Performance Liquid Chromatography-Diode Array Detection method for the determination of aging markers in Tequila. Journal of Chromatography A. 1213, 218-223.

P-118

Discrimination of Corsican honeys by FTIR spectroscopy and chemometrics

Y Le Dréau[1], Ying Yang[2], J. Costa[2], J. Paolini[2], N. Dupuy[1]

[1] Aix Marseille Université, LISA EA 4672, 13397, Marseille, France
[2] Université de Corse, UMR CNRS 6134 SPE, Laboratoire de Chimie des Produits Naturels, 20250, Corte, France

Honeys' traceability is required by consumers and food control institutions. Melisso-palynological routines use pollens to discriminate the botanical origin and / or the geographical origin of honeys. Honey is a very complex matrix with several chemical classes of compounds present in a very large range of concentrations. Therefore, to improve the reliability of assignments of honeys to classes, some alternative analytical methods (chromatographic methods especially) have been developed for involve for example the phenolic acids and polyphenols[2], or the volatile compounds[3, 4]. All of these methods require considerable sample preparation, are time-consuming, and their insufficient ability to correctly classify honeys has been shown. To overcome these limitations, various spectroscopic techniques such as MNR[4], near-infrared[5], mid-infrared[6,7] or Raman spectroscopy[8] have been proposed as alternative methods since they have been used successfully especially for discrimination of unifloral honeys.
The Corsican honeys may be certified by two official designations of origin: the national "Appellation d'Origine Contrôlée" (AOC) and the European "Protected Designation of Origin" (PDO), both labeled "Miel de Corse–Mele di Corsica". They are multifloral honeys classified into six ranges: "spring", "spring maquis", "honeydew maquis", "chestnut grove", "summer maquis" and "autumn maquis" according to the geographic location of the apiaries and the harvest season [9]. Therefore, the aim of this study was to investigate the potential of MIR spectroscopy coupled with chemometric analysis as fast and effective analytical technique to discriminate the six ranges. The ATR-FTIR analysis of 216 Corsican honey samples (PDO) associated with chemometric analysis allowed to highlight differences between the six origins even if it requires adjustments depending on the samples (especially for "chestnut maquis" and "spring maquis" honeys) and, its application is limited for the "honeydew maquis" honeys. However, this method is reliable for "autumn maquis" honeys and "summer maquis" honeys (98%of correct classification).

Keywords: Chemometrics, Honey, MIR

[1] Yang, Y. et al., Melissopalynological origin determination and volatile composition analysis of Corsican "chestnut grove" honeys. *Food Chemistry*, **2012**, 132 (4), 2144-2154.
[2] Escriche, I., et al., Suitability of antioxidant capacity, flavonoids and phenolic acids for floral authentication of honey. Impact of industrial thermal treatment. *Food Chemistry*, **2014,** 142, 135-143.
[3] Perez, R.A. et al., Analysis of volatiles from Spanish honeys by solid-phase microextraction and gas chromatography–mass spectrometry. *Journal of Agricultural and Food Chemistry*, **2002**, 50, 2633-2637.
[4] Consonni, R. et al, Recent Developments in Honey Characterization. *Rsc Advances*, **2015**, 73, 59696-714.
[5] Liang, X. Y. et al, Classification of Floral Origins of Honey by NIR and Chemometrics. *Advanced Materials Research.* **2012**, *605-607*, 905-909.
[6] Ruoff, K. et al, Authentification of the botanical and geographical origin of Honey by MID-infrared spectroscopy. *Food Chemistry*, **2006,** 54 (18), 6873-6880.
[7] Einar Etzold E. Et al., Determination of the botanical origin of honey by Fourier-transformed infrared spectroscopy: an approach for routine analysis.·Europeen Food Research Technology, **2008**, 227,579-586
[8] Fernández Pierna J.A. et al., Discrimination of Corsican honey by FT-Raman spectroscopy and chemometrics. *Biotechnology, Agronomy, Society and Environment*, **2011** 15(1), 75-84
[9] Battesti, M. J. et al., *Définition du périmètre de production – Rapport des experts en vue de la mise à l'enquête*. Demande de reconnaissance en A.O.C. "Miel de Corse–Mele di Corsica" **1997.**

P-119

DETECTION OF OFF-ODORS IN WINES BY INFRARED SPECTROSCOPY

Cláudia A. Teixeira dos Santos[1], Ricardo N.M.J. Páscoa[1], N. Pérez-del-Notario[2], J. M. González-Sáiz[2], C. Pizarro[2], João A. Lopes[3]

[1]LAQV/REQUIMTE, Laboratório de Química Aplicada, Departamento de Ciências Químicas, Faculdade de Farmácia, Universidade do Porto, Rua de Jorge Viterbo Ferreira, 228, 4050-313 Porto, Portugal
[2]Departmento de Química, Universidad de La Rioja, C/ Madre de Dios 51, 26006 Logroño, La Rioja, España
[3]Departmento de Farmácia Galénica e Tecnologia Farmacêutica, Faculdade de Farmácia, Universidade de Lisboa, Av. Prof. Gama Pinto, 1649-003 Lisbon, Portugal
claudixteixeira@gmail.com

Over the past decades, a wide number of analytical methods based on Fourier-transform infrared (FTIR) spectroscopy, were developed aiming the control of several wine parameters. However, this technique has still a lot to offer in what concerns the improvement of wine quality. In this work, FTIR spectroscopy and Partial Least Squares (PLS) regression, were combined for the development of new analytical methods able to assess the content of some of the most common off-odors in wines: isoamyl alcohol, isobutanol, 1-hexanol, butyric acid, isobutyric acid, decanoic acid, ethyl acetate, furfural and acetoin. These compounds are potential indicators of wine defects when present in concentrations above their odor thresholds [1]. Analytical methods currently employed on their determination are based on slow, expensive and rather complicated procedures [2]. Additionally, the sensorial analysis commonly employed to detect the presence of these compounds, are usually highly dependent on a panel of trained sensory assessors, which can also be time-consuming and expensive. In the worst scenario, the above mentioned compounds are only detected when their presence in wine already represents a defect. Fourier-transform infrared (FTIR) spectroscopy, represents an alternative solution for the early detection of these compounds, whether individually or simultaneously, in a fast and cost effective way.
The accuracy and precision of this newly developed methods (R_p^2>0.88 and range error ratio>9.23) proved the worthy ability of this instrumental technique to quantify all the aforementioned compounds within their tested concentration ranges.

Acknowledgement: C.A. Teixeira dos Santos and Ricardo N.M.J. Páscoa, acknowledge FCT-MEC (Fundação para a Ciência e Tecnologia) for the financial support through the grants SFRH/BD/91419/2012 and SFRH/BPD/81384/2011, respectively. J. Lopes thanks FSE (Fundo Social Europeu) and MEC (Ministério da Educação e Ciência) for the financial support through the POPH-QREN program.

References:
[1] M. V. Moreno-Arribas, M. C. Polo, Wine Chemistry and Biochemistry, Springer, Vol. 378 (2009).
[2] J. Torrens, M. Riu-Aumatell, E. López-Tamames, S. Buxaderas, J. Chromatogr. Sci., 42, 310 (2004).

P-120

THE USE OF A "DUMMY" CLASS WHEN PERFORMING (PSEUDO) TWO CLASS MULTIVARIATE CLASSIFICATIONS

A.M. Jiménez Carvelo, L. Valverde Som, C. Ruiz Samblás, E. Pérez Castaño, A. González Casado, L. Cuadros-Rodríguez

Department of Analytical Chemistry, University of Granada, C/ Fuente Nueva s/n, Granada, Spain.
email: amariajc@ugr.es

Classification methods have been defined as the tool which divides each one of the objects of interest into one or several series of exhaustive and exclusive categories, known as classes. Therefore, the term 'class' is referred to a division of the population sharing at least a common characteristic, attribute, quality, or property. Ideally these ones should be mutually exhaustive and exclusive, i.e. the whole population should be covered by both classes and common characteristics cannot be shared, respectively.

In general, when performing a classification method which is aimed to allocate the target objects between two classes, it is required the development of a model which has to be trained with standards that belong to both input classes, for instance, class A and class B. If A and B are exhaustive classes, the class B is equivalent to the no-A class. For example, in the differentiation of olive oil from other edible vegetable oils, the classes would be olive oil (A) and another vegetable oil (B); this last one could be also denoted as non-olive oil (no-A). In this case, it is also possible to build a classification model training the model with only one input class, for instance, the class A (or the olive oil class). The approach of working with only one input has a big advantage since it only requires experimental data from this input class. This strategy has significant advantages in the authentication of food, due to the fact that the model is developed from authentic foodstuffs and it is not necessary to use other kinds.

Since not all the classification algorithms accept models which are trained with only one input class, it is necessary to establish a fictitious (dummy) class. However, the classification model that uses a second dummy class is not exactly an one input-class classification but (pseudo) two input-classes classification, since two classes are being used. The approach of an only one input-class could be easily performed by applying a class-modelling method as SIMCA or UNEQ. However, in principle, regression-based discriminant methods as PLS-DA or SVM-C need to feed the model with two input-classes. Nevertheless, there is some recent reported proposals which make suitable the PLS method to perform only one-input class (OCPLS).

In this communication, it is presented the methodology to perform the multivariate classification by applying the (pseudo) two input-classes strategy, that is to say, the model is truly trained with standards from an only target class. As real example, the discrimination between olive oils and non-olive vegetable oils is showed. For this, the fast-liquid chromatography fingerprint on the methyl-transesterified fraction of each vegetable oil is used as characteristic data vector.

P-121

MODELING AND PREDICTING SECOND and THIRD-ORDER FLUORESCENCE SPECTROSCOPY DATA AS A NOVEL QUALITY CONTROL STRATEGY ON MAYONNAISE

S.M. Azcarate [1], C. Teglia[2], M. Montemurro[2], G. Siano[2], J.M. Camiña[1], H.C. Goicoechea[2].

[1]Facultad de Ciencias Exactas y Naturales, Universidad Nacional de La Pampa, and Instituto de Ciencias de la Tierra y Ambientales de La Pampa (INCITAP), Av. Uruguay 151 (6300) Santa Rosa, La Pampa, Argentina.
[2]Laboratorio de Desarrollo Analítico y Quimiometría (LADAQ), Cátedra de Química Analítica I, Facultad de Bioquímica y Ciencias Biológicas, Universidad Nacional del Litoral-CONICET, Ciudad Universitaria, Santa Fe (S3000ZAA).
silvanaazcarate@gmail.com

Fluorescence spectroscopy has found in food analysis a large use due to that it is fast, gives direct measurement, is not destructive and noninvasive [1].
In this work, the potential of excitation emission fluorescence matrix (EEM) data recorded with a front face system along with chemometric methods was investigated for the determination of spoilage on mayonnaise and for the rapid prediction of changes of microbial flora (counts of bacteria).
Made in home mayonnaise and commercial mayonnaises samples were maintained at 25°C during four days, and stored at 37ºC for three days. The microbial load on mayonnaise samples was determined per day. Total viable count (TVC), aerobic mesophilic bacterias, Staphylococcus aureus, Escherichia coli and Salmonella were determined on mayonnaise at each step using culture methods.
In parallel, mayonnaise samples were analyzed by 3D front-face fluorescence spectroscopy, without prior preparation, at 8 and 6 different times for the samples stored 25 °C, and 37°C, respectively. The data matrices were recorded varying the excitation wavelength between 230 and 400 nm each 10 nm, and registering the emission spectra from 300 to 600.5 nm each 0.5 nm. Thus, the EEMs were of size 18×602. PARAFAC analysis allowed capturing the changes occurring in the fluorescence spectral data. The best PARAFAC models showed 3 components for data recorded both temperatures. Profiles of main compounds were extracted with this algorithm describing quality evolution on the time. In order to meet those compounds that their concentration decreases and those that are produced (specifically amino acids) a chromatographic analysis was performed [2]. The chromatographic analysis confirms the decrease of tyrosine and the production of tryptophan in the time.
Partial least squared discriminant analysis (PLSDA) was applied to data set formed by concatenating of all the fluorescence spectra at same temperature for testing the allocation of the spectra of the individual samples within the five and four groups corresponding to the five and four investigated storage times. The results showed that 100% of good classifications were obtained using 3 PLS factors. These results allowed the classification of samples as a function of stored time. In addition, this information can be related to the microbial counts.
In order to evaluate other quality parameters, a study with optical fibers is being performed. EEMs were obtained through the degradation kinetics of mayonnaise produced by irradiation the samples each two minutes for a total of 20 minutes. Third order data analysis will be directed to the fat content evaluation as well as at the content of different dressing of mayonnaise for obtaining information about the quality mayonnaises.

Acknowledgement: Universidad Nacional Del Litoral, Universidad Nacional de La Pampa, Consejo Nacional de Investigaciones Científicas y Técnicas, Agencia Nacional de Promoción Científica y Tecnológica

References:
[1] A. Cifuentes, D. Rutledge, TrAC-Trend Anal. Chem., 52, 1 (2013)
[2] J.R. Lakowicz, Principles of *Fluorescence Spectroscopy*, Springer, 3, 63(2006)

P-122

Investigation of Synergetic Antioxidant Effects among Four Herbal Extracts

Linqi Liu, Chenxi Zhao, Xiangxue Fu, Xiangyu Tan, Lei Ren

Department of Bioengineering and Environmental Science, Changsha University, Changsha,P.R. China email (cxzh003@163.com)

Traditional Chinese medicine extracts are basic materials for the treatment of many diseases, such as cardiovascular disease. In this paper, the extracts from four heat-clearing herbs were obtained by ultrasonic extraction method, using ethanol-water mixed solvents, and DPPH free radical-scavenging and Ferric reducing antioxidant power (FRAP) assays were chosen to evaluate their antioxidant activities. Uniform design method was used to investigate the synergistic antioxygenation from the four kinds of extracts. Results showed that yields from the Pulsatillae Radix, Smilacis Glabrae Rhizoma, Isatidis Folium and Taraxaci Herba extracts were 18.0%, 8.33%, 8.83% and 3.80%, respectively. Results obtained by DPPH and FRAP assays showed that extracts from Smilacis Glabrae Rhizoma had the highest antioxidant activity, followed by Pulsatillae Radix and Isatidis Folium extracts. Extract from Taraxaci Herba had the lowest antioxidant activity at the same concentration. Their IC50 values were 28.8, 31.6, 49.2 and 74.5μg/mL, respectively and a dose-effect relationship appeared in each sample. The results for synergistic antioxidant effect investigation showed that the maximum positive synergetic antioxidant effect was exhibited when the mass ratio of extracts from the Pulsatillae Radix and Smilacis Glabrae Rhizoma was equal to 2:8, with a 24 μg/mL concentration. The achieved DPPH free radical scavenging rate was 54.3%, which was close to BHT 24 μg/mL (60.5%). These research results have thus established a foundation for development of novel natural antioxidants.

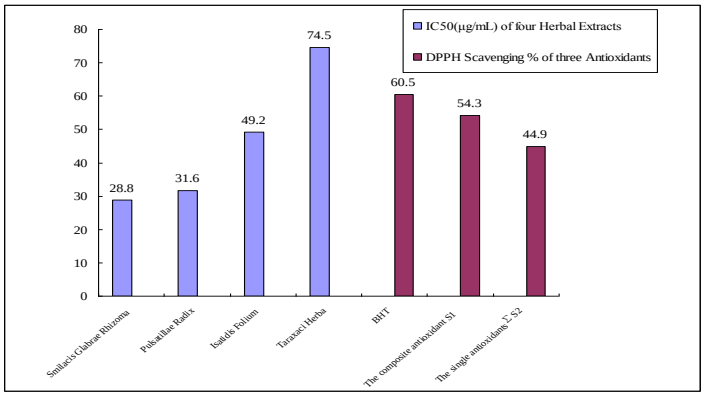


Fig. 1. The Synergetic Antioxidant Effects of the Herbal Extracts.

Acknowledgement: This work is financially supported by the National Nature Foundation Committee China (grants no. 81173533) and Scientific Research Fund of Hunan Provincial Education Department (Grant No. 15C0118) and Keystone Project of Changsha Science and Technology Agency（Grant No. K150703-11).

References:

[1] L. M. Zhang, Z. X. Lu, X. L. Li, Y. Deng, F. Q. Zhang, C. Ma, and N. Y. He, Polym. Chem. 3 (2012)1958.
[2] A. K. Babahedari, E. H. Soureshjani, M. K. Shamsabadi, and H. Kabiri, J. Bionanosci. 7 (2013)288 .
[3] L. M. Zhang, Z. X. Lu, Y. Y. Bai,T. Wang, Z. F. Wang, J. Chen,Y. Ding, F. Yang, Z. D. Xiao, S. H. Ju, J. J. Zhu, and N. Y. He, J. Mater. Chem. B. 1 (2013) 1289.
[4]T. H. Kim, H. Choi, G. S. Yu, and J. S. Choi, J. Nanosci. Nanotechnol. 13 (2013) 7325.
[5] F. Wang, S. C. Zhao, F. Li, B. Zhang, Y. Qu, T. L. Sun, T. Luo, and D. P. Li. Plos One 9(2014) e87221 .

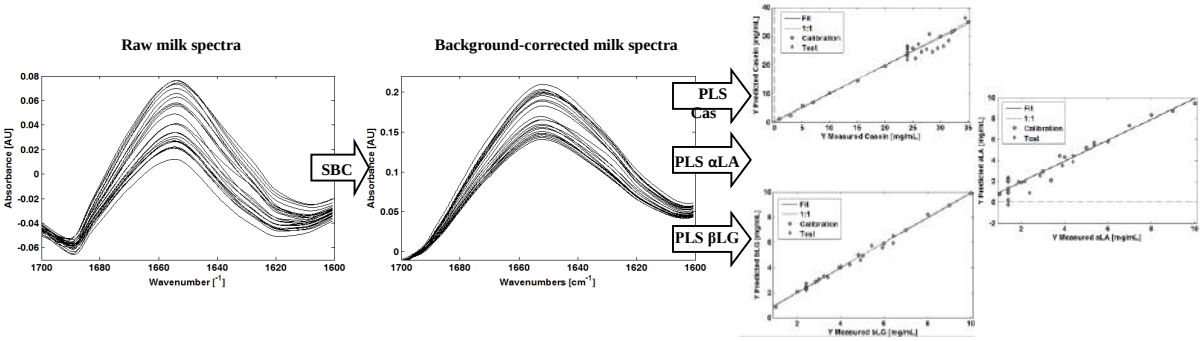
P-123

EXTERNAL CAVITY-QUANTUM CASCADE LASER (EC-QCL) SPECTROSCOPY AND CHEMOMETRICS FOR PROTEIN ANALYSIS IN COW'S MILK

J. Kuligowski[1,2], A. Schwaighofer[2], M.R. Alcaraz[3], M. Vento[1,4], B. Lendl[2]

[1]*Neonatal Research Unit, Health Research Institute La Fe, Avenida Fernando Abril Martorell 106, 46026 Valencia, Spain*
[2]*Institute of Chemical Technologies and Analytics, Vienna University of Technology, Getreidemarkt 9/164-UPA, 1060 Vienna, Austria*
[3]*Laboratorio de Desarrollo Analítico y Quimiometría, FBCB, Universidad Nacional del Litoral-CONICET, Ciudad Universitaria, 3000 Santa Fe, Argentina*
[4]*Division of Neonatology, University & Polytechnic Hospital La Fe, Avenida Fernando Abril Martorell 106, 46026 Valencia, Spain*
julia.kuligowski@uv.es

The determination of total protein content in cow's milk is a routine application of mid-infrared (IR) transmission spectroscopy. However, the quantitation of protein in this kind of samples demands laborious and time consuming experimental work. We report an analytical method based on the direct spectroscopic determination of casein (Cas), α-lactalbumin (αLA) and β-lactoglobulin (βLG) in cow milk samples. For spectra acquisition, a custom-made setup equipped with a tunable EC-QCL light source, operated in pulsed mode at room temperature, and a temperature stabilized flow cell with an optical path length of 38 µm was used. An advanced data processing protocol was applied to compensate for fluctuations in the fine structure of the emission curve that are inherent to the employed light source due to mechanical instabilities [1]. Mid-IR transmission spectra of aqueous protein standard mixtures, as well as cow's milk and spiked cow's milk samples were acquired in the 1600-1700 cm^{-1} region (amide I band). Science based calibration (SBC) algorithm was employed for compensation of changes in the background signal attributed to the milk matrix. After background correction, partial least squares (PLS) was employed to build regression models for each of the three proteins by comprising spectra of aqueous calibration standard solutions and background-corrected milk samples. Finally, predictions of individual protein concentrations in commercial and spiked cow's milk were carried out. Protein concentrations determined in commercially available milk samples were 24.5±1.4, 1.4±0.9 and 2.4±0.2 mg mL^{-1} for Cas, αLA and βLG, respectively. Recovery values for spiked samples were ranging between 95±8%, 101±22% and 99±4% for Cas, αLA and βLG, respectively. In light of the results obtained, this high throughout method has the potential to be employed as a standard tool for quality control of milk, without requiring any sample preparation.



Acknowledgement: The authors acknowledge the Austrian Research Promotion Agency (FFG) within the K-project imPACts (contract no. 843546) for funding. JK acknowledges a Sara Borrell grant (CD12/00667) from *Instituto Carlos III* (Spanish Ministry of Economy and Competitiveness).

References:

[1] M.R. Alcaráz, A. Schwaighofer, C. Kristament, G. Ramer, M. Brandstetter, H. Goicoechea, B. Lendl, Anal Chem, 87, 6980-6987 (2015).

P-124

RAMAN-FT AND NIR SPECTROSCOPY DATA FUSION FOR DETECTING FOOD FRAUD USING MULTIVARIATE QUALITATIVE ANALYSIS

M. Pilar Callao, Cristina Marquez, Itziar Ruisánchez

Department of Analytical and Organic Chemistry, Rovira i Virgili University, Marcel·lí Domingo s/n, Tarragona, 43007, Spain

Food fraud is becoming increasingly sophisticated due to the use of conventional and unconventional or synthetic adulterants. Because to the ever-increasing range of analytes that can be used in food fraud together with the impossibility of covering them all, there is increasing demand for the development of fast, easy-to-use and low-cost analytical methods to test for adulteration. Multivariate qualitative methodologies that use classification techniques allow to detect anomalous samples and they are an alternative of increasing application in detection food fraud, both adulteration and authentication [1].

The application of spectroscopic techniques has become a usual tool in food analysis and requires the use and development of chemometrics tools in order to display and interpret vast amounts of data. A trend that is expanding is the use of data fusion strategies that couple or merge data from two or more analytical techniques to improve the individual information of each one due to the exploitation of synergies between them.[2,3]

In this work a hazelnut adulteration problem is considered as a case study. Hazelnuts and their derivatives (oils and pastes) are widely used as ingredients in many desserts, ice creams and chocolates or can be eaten alone as a snack. The price of hazelnuts depends on the market and can be reduced by adding such other ingredients as almond, because of its similarity, but other unexpected products might be added (for example, chickpea).

NIR and FT-Raman data are used to determine whether the synergism between them can be exploited. The data have been processed both individually and fused by SIMCA model from a training set of unadulterated samples [4]. Two data fusion strategies (mid-level and high decision data fusion) are applied.

References:

[1] F. Marini (Ed.), Data Handling in Science and Technology: Chemometrics in Food Chemistry 28. Amsterdam: Elsevier B.V.

[2] E. Borras, J.Ferre , R. Boque, M. Mestres, L. Aceñ and, O. Busto (2015) Data fusion methodologies for food and beverage authentication and quality assessment. A review. Analytica Chimica Acta 891 1-14

[3] Carolina V. Di Anibal, M. Pilar Callao, Itziar Ruisánchez (2011), 1H NMR and UV-visible data fusion for determining Sudan dyes in culinary spices. Talanta 84 829–833

[4] M.I.López,; E.Trullols,; M.P. Callao, and I. Ruisánchez, (2014) Multivariate screening in food adulteration: untargeted versus targeted modelling. Food Chemistry, 147, 177-181.

P-125

QUALITATIVE MULTIVARIATE ANALYSIS. THE USE OF PROBABILITY OF DETECTION (POD) CURVES

Carina S. Gondim[1,2], Roberto Gonçalves[1]; Scheila V. C. Souza[1]; M. Pilar Callao[2]; Itziar Ruisánchez[2]

[1]Department of Food Science,Federal University of Minas Gerais (UFMG), , Av. Antônio Carlos, 6627, Campus da UFMG, Pampulha, 31270-010, Belo Horizonte, MG, Brazil
[2]Department of Analytical and Organic Chemistry ,Rovira i Virgili University, Marcel·lí Domingo s/n, Tarragona,43007, Spain

Most analytical problems require semi-quantifiable or non-quantifiable information: i.e. to authenticate a product or verify if a substance is present above or below a pre-established concentration level. In these cases, using qualitative methods that provide a binary response (positive/negative) might be suitable. They are an alternative to quantitative analysis, which generally gives more but often unnecessary sample information and requires a greater investment of money and/or time.

In this work, a qualitative method is developed to recognize four common adulterants in raw milk. The main performance parameters of the two-class SIMCA method, sensitivity and specificity, are established from the contingency table outputs which are [1]: True positive (TP), when the qualitative method gives a positive output for a positive sample; False positive (FP), when it gives a positive output for a negative sample. True negative (TN), when it gives a negative output for sample negative; False negative (FN) result, when it gives a negative output for a positive sample. In addition to these well-known parameters, we propose another parameter named as "Inconclusive ratio", which indicates the percentage of samples that cannot be undoubtedly assigned to one class. It considers the samples that are not assigned to any class and the samples assigned to the two class.

As we are dealing with a qualiative method addressed to detect if a substance (milk adulterand) is present in a sample, in the validation process it is also important to estimate quality performace parameters related to the concentration. For instace, the concentration at which an adulterant can be recognized by the qualitative multivariate technique with likely satisfactory success. A methodology based on nonlinear fitting of the experimental data to a probability of detection curve (POD) [2] is proposed to estimate the performance parameters related to the concentration.

Acknowledgement: Carina S. Gondim thanks to Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) - Programa de Doutorado Sanduíche no Exterior (PDSE) for providing a fellowship

References:

[1] M.I.; López, E. Trullols; M.P. Callao and I. Ruisánchez, (2014) Multivariate screening in food adulteration: untargeted versus targeted modelling. *Food Chemistry*, 147, 177-181.

[2] H. Schneeweiß, P.Th. Wilrich (2014) A simple method of constructing a confidence interval for the mean probability of detection in collaborative studies of binary measurement methods. *Accred Qual Assur* 19, 221–223.

P-126

DETECTION OF COMMON ADULTERANTS IN RAW MILK BY MID-INFRARED SPECTROSCOPY AND ONE-CLASS AND MULTICLASS STRATEGIES

Carina S. Gondim[1,2], Roberto Gonçalves[2]; Scheila V. C. Souza[2], Itziar Ruisánchez[1], M. Pilar Callao[1]

[1]Department of Analytical and Organic Chemistry ,Rovira i Virgili University, Marcel·lí Domingo s/n, Tarragona,43007, Spain

[2]Department of Food Science,Federal University of Minas Gerais (UFMG), , Av. Antônio Carlos, 6627, Campus da UFMG, Pampulha, 31270-010, Belo Horizonte, MG, Brazil

Milk fraud involves the addition of several types of compound from complex to simple substances. The last type is still today the most common because it allows to mask failures in good manufacturing practices or to increase the milk volume for greater profit. To reduce the microbial count, preservatives such as formaldehyde, hydrogen peroxide and sodium hypochlorite are usually added. To reduce the acidity, the most common neutralizers are bicarbonate, carbonate, hydroxide and citrate. To increase the milk volume, water is added and to occult this fraud, substances classified as density restoratives such as sodium chloride, starch and sucrose are also added [1].

Classical qualitative tests for detection of common milk adulteration includes independent determinations for different analytes, which results in a significant number of tests and large consumption of time and reagents, with the consequent generation of waste . To overcome these drawbacks in this work it is evaluated the ability to detect simultaneously common adulterants in milk, including water, density restoratives, preservatives and neutralizing agents, by classification techniques applied to mid-infrared (MIR) data.

The strategy proposed is sequential: in a first step, a one-class SIMCA model from a training set of unadulterated samples was established. The goal is to check if the combination of MIR data with multivariate classification strategy allows detecting unadulterated samples as belonging to the class modeled (fit the model) and detecting adulterated samples as not belonging to the modeled class (do not fit the model) [2]. As a result, we will be able to screen which are the adulterants that the model is able to differentiate.

In the second step, a screening process was implemented for determining if an unknown sample has been adulterated or not with one of the possible adulterants identified in the first step. Therefore, a multi-class SIMCA strategy was implemented in which, in addition to the unadulterated samples, multivariate classification models are established for each type of adulterant that has been added to the samples. At the end, 6 classes have been properly modeled, corresponding to unadulterated samples and adulterated samples with one of the 5 common milk adulterants (hydrogen peroxide, sodium citrate, sodium carbonate, formaldehyde and starch).

Acknowledgement: Carina S. Gondim thanks to Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) - Programa de Doutorado Sanduíche no Exterior (PDSE) for providing a fellowship

References:

[1] C.S. Gondim; R.C.S Souza; M.P.P Palhares; R.G. Junqueira and S.V.C. Souza, (2015) Performance improvement and single laboratory validation of classical qualitative methods for the detection of adulterants in milk: starch, chlorides and sucrose. *Analytical Methods*, 7, 9692-9701.

[2] M.I. López; E. Trullols; M.P. Callao and I. Ruisánchez, (2014) Multivariate screening in food adulteration: untargeted versus targeted modelling. *Food Chemistry*, 147, 177-181.

P-127

A CHEMOMETRIC PCR MODELLING STRATEGY FOR MONITORING THE GENERATION OF TOXIC ALDEHYDIC LIPID OXIDATION PRODUCTS IN FRYING MEDIA DURING SIMULATED SHALLOW-FRYING EPISODES

M. Grootveld[1], S. Moumtaz[1], K. Grootveld[1], D. Parmar[1], B. Percival[1], K. Desai[1], R. Ohurogo[1], P. Jansson[1], V. Ruiz-Rodado[1]

[1]*Leicester School of Pharmacy, De Montfort University,The Gateway, Leicester LE1 9BH, United Kingdom*
mgrootveld@dmu.ac.uk

Objectives/Hypotheses: Adverse health effects associated with the dietary consumption of lipid oxidation products (LOPs), such as cytotoxic/mutagenic aldehydes and their conjugated hydroperoxydiene precursors, has recently attracted a high level of clinical interest. Hence, in this investigation we have employed a ^{1}H NMR-linked Principal Component Regression (PCR) modelling approach to explore relationships between data matrices consisting of (1) aldehydic LOP concentrations produced in culinary oils/fats when thermally-stressed according to standard frying practices, and (2) the prior saturated (SFA), monounsaturated (MUFA) and polyunsaturated fatty acid (PUFA) contents of such frying media (FM), together with the times at which they were heated in laboratory-simulated shallow-frying episodes.

Methods: Corn, sunflower, extra virgin olive, rapeseed, linseed, canola and coconut oils, and butter and lard, were heated according to laboratory-simulated shallow-frying practices at 180^{o}C, and FM samples were collected at time-points of 0, 5, 10, 20, 30, 60 and 90 min. (n = 6 replicates per time-point). Aldehydes and FM fatty acid concentrations were determined by ^{1}H NMR analysis (Bruker AV 400 MHz spectrometer). The first (dependent) PCR data matrix comprised aldehyde concentration scores vectors (PC1* and PC2*), whilst the second (predictor) one involved those arising from the fatty acid content/heating time variables (PC1-PC4), together with their first-order PC1-3 x heating time-point (PC4) interactions.

Results: *Trans,trans*- and *cis,trans*-alka-2,4-dienals, 4-hydroxy-*trans*-2-alkenals and 4-hydroperoxy-*trans*-2-alkenals (phase I aldehydes predominantly arising from PUFA peroxidation) strongly/positively loaded on PC1*, whereas *n*-alkanals and *trans*-2-alkenals (phase II aldehydes derived from both MUFAs and PUFAs, those arising from the former appearing only after an extended time-lag period) strongly/positively loaded on PC2*. PCR analysis of these scores vectors (SVs) demonstrated that PCs 1 (positively-loaded linoleoylglycerols), 2 (positively-loaded oleoylglycerols), 3 (positively-loaded linolenoylglycerols) and 4 (exclusively positively-loaded sampling time-points) all powerfully contributed to aldehydic PC1* SVs ($p < 10^{-6}$), as did all PC1-3 x PC4 interaction ones. PC2* was also substantially dependent on two of the above PC SVs (specifically PC1 and PC2, with PC2's contribution being *ca.* twice that of PC1) and their interactions with heating time-point (PC4), but not those of PC3. A Q^2 value of 0.533 ($R^2Y = 0.810$) was obtained. These results were confirmed by Canonical Correlation Analysis of these two sets of PC SVs.

Conclusions: NMR-linked PCR analysis is a valuable strategy for (1) modelling the time-dependent generation of aldehydic LOPs in heated FM during standard shallow-frying practices, and (2) identification of their parent fatty acid sources therein.

P-128

APPLICATION OF INDEPENDENT COMPONENTS ANALYSIS ON GC-MS DATA TO STUDY THE AROMATIC PROFILE OF LEBANESE SYRAH WINE DURING FERMENTATION

S. Azzi [1], N. Estephan [1], N. Ouaini [1], D.N. Rutledge [2]

[1]Department of Chemistry and Biochemistry, Faculty of Sciences, Holy Spirit University of Kaslik (USEK), B.P.446, Jounieh, Lebanon
[2] UMR Ingénierie Procédés Aliments, AgroParisTech, Inra, Université Paris-Saclay, 91300 Massy, France
nathalieestephan@usek.edu.lb

In the competitive world market of today, wine industries constantly aim to improve the winemaking process, including the role of yeast. Hence, and to determine the contribution of certain yeast to the wine aroma, it is necessary to understand their role in the fermentation process, the chemical changes they induce and the influence of oenological factors on their growth kinetics.
The objective of this work is to study the evolution of the aromatic profile of Syrah wine, taking into consideration the effect of indigenous and commercial yeasts. Hence, inoculated and spontaneous alcoholic fermentations were conducted on the Syrah grape variety from two Lebanese regions: Kfifane and Mar Moussa.
Micro-vinifications were piloted with and without commercial yeast for the seasons 2011 and 2012. Samples were taken daily, from the beginning of the vinification until the end of the alcoholic fermentation. The extraction of flavors was conducted using headspace solid-phase microextraction (HS-SPME) and extracted compounds were identified and quantified by gas chromatography coupled to mass spectrometry (GC-MS).
The peak areas for the different compounds detected by GC-MS were exported in matrices and divided by the area of the internal standard peak. The resulting matrices were then processed by Independent Components Analysis (ICA).
In this study, the ICA was carried out using *JADE algorithm* (*Joint Approximate Diagonalization of Eigenmatrices*) which is a standard source separation algorithm [1,2]. Calculation was conducted using *Matlab R2011b* (*The MathWorks Inc., Natick, MA, USA*) and an extension of the ICA-by-Blocks function (*Random_ICA_2014*) was used to determine the optimal number of independent components (ICs) to extract.
Results led to the identification of critical factors for spontaneous fermentation depending on the geographical origin of the grape variety. Thus, low temperatures seem more appropriate for spontaneous fermentations for Kfifane grapes, favoring the synthesis of acetates, while the ordinary temperatures of fermentations are more suitable for spontaneous fermentations for Mar Moussa grapes, favoring the synthesis of higher alcohols floral notes.
In addition, spontaneous fermentations were distinguished from inoculated fermentations for both the Kfifane and Mar Moussa regions. The "terroir" effect is expressed in the variation in the volatiles profile. The evolution of volatile compounds in both regions was usually normal without notable features. Moreover, no influence of molecules of unpleasant characters was noted.

References:
[1] J.F. Cardoso, Neural Computation, 11(1), p. 157-192 (1999).
[2] D.N. Rutledge, D. Jouan-Rimbaud Bouveresse, TrAC Trends in Analytical Chemistry, 50, p. 22-32 (2013).

P-129

RANDOMIZED BLOCKED DESIGN APPLIED FOR THE MULTICRITERIA OPTIMIZATION OF HEADSPACE TRAP EXTRACTION OF FURAN AND FURFURAL FROM SPONGE CAKE

M. Cepeda-Vázquez, D. Blumenthal, V. Camel, B. Rega

UMR Ingénierie Procédés Aliments, AgroParisTech-Inra-Université Paris-Saclay,
1 avenue des Olympiades 91300, Massy, France.
mayela.cepedavazquez@agroparistech.fr

Furan, a possibly carcinogenic compound to humans, and furfural, a naturally occurring aroma compound, can be produced in heat-processed foods [1,2]. Therefore, their simultaneous determination is important for food safety and quality purposes. Headspace techniques coupled to gas chromatography-mass spectrometry (GC-MS) are often used to analyze process-induced compounds. Although headspace trap (HS Trap) extraction has been recently used for furan quantitation in bread crust [3], this technique has not been used for the simultaneous determination of furan and furfural before. This work presents an optimal design of experiments (O-DOE) approach using random blocks for the multicriteria optimization of HS Trap extraction of furan and furfural from sponge cake for their quantification by GC-MS. A bias related to sample preservation was observed during preliminary tests and eliminated in further assays by considering it as a blocking factor. A tailor-made experimental plan including four instrumental factors (i.e. thermostatting temperature and time, dry purge time, pressurization cycles) and two sample preparation variables (i.e. ratio of water and sample weight [dry basis] and total amount [dry basis]), considering the blocking factor was designed. Response surface models (P-value <0.0001; $R^2 \geq 0.974$ for all models) resulted in the identification of most linear and some quadratic terms and a few second order interactions as influencing the extraction of furan, furfural and their internal standards, *d4*-furan and *d4*-furfural. Interestingly, one of the sample preparation variables, namely the ratio of water and sample weight (dry basis), showed no effect on furfural extraction, while different profiles for furan and *d4*-furan were observed regarding this factor. A set of optimal conditions for a HS Trap extraction method was, for the first time, successfully defined by using the optimal design of experiments methodology and a desirability function.

Acknowledgements: The authors wish to thank Nicolas Descharles and Jessica Malfaire for technical support. In addition, M. Cepeda-Vázquez wishes to thank the Mexican National Council for Science and Technology (CONACYT) for her doctoral grant.

References:

[1] J. Pico, J. Bernal, M. Gómez, Food Research International, 75, 200 (2015).
[2] European Food Safety Authority, EFSA Journal, 9, 2347 (2011).
[3] L. Huault, N. Descharles, B. Rega, S. Bistac, V. Bosc, P. Giampaoli, Food Addit Contam A, 33, 236, (2016).

P-130

DETECTION OF COCOA COUNTERFEIT IN ROUTINE QUALITY CONTROL WITH A PORTABLE NIR ANALYZER (VISUM PALM)

L. Rodríguez-Turienzo [1], M. Hidalgo [2], M. Parra [1], S. Fluvià [1], Anna de Juan [2], A. Rosales [1]

[1] Science Department, IRIS-Innovació i Recerca Industrial i Sostenible. Parc Mediterrani de la Tecnología. Avda. Carl Friedrich Gauss nº11, 08860 Castelldefels, Spain.
[2] Grup de Quimiometria, Facultat de Química, Universitat de Barcelona, Diagonal 647, 08028 Barcelona, Spain.
lrodriguez@iris.cat

Detection of fraud and adulteration involves both economic and food safety issues, thus analysis of incoming raw material during routine quality control is crucial for food industry. Cocoa powder has been reported as one of the 25 food ingredients most adulterated [1]. Furthermore, recent cocoa powder adulterations in Europe have reported samples containing up to a 20% less cocoa compared to values claimed on labelling [2]. Cocoa products can be adulterated with a range of substances, including grain flours, peanut shell and other materials. Classical analytical techniques for testing cocoa powder, such as liquid chromatography or sensory evaluation, are time consuming and complicated. Near-Infrared (NIR) spectroscopy offers the advantage of rapid, non-destructive analysis and easier routine operation. In addition, a portable device results highly convenient to quality control analysis within an industrial environment.

In this communication, the suitability of a portable NIR analyser (Visum Palm) for detecting cocoa adulteration with wheat flour using NIR spectral fingerprinting coupled to multivariate calibration will be presented.

Samples consisted of a set of polyethylene bags (weight of approximately 60 g) with mixtures of cocoa powder and wheat flour at different proportions ranging from a 0% to a 100% of cocoa powder with a difference of 5% (w/w) between bags. Spectra of the samples were acquired directly through the plastic bags with the portable device. Afterwards, chemometric regression models were successfully developed and optimized. Best model obtained a calibration correlation (R^2) of 0.987, a *RMSEC* of 3.30 and a *RMSEV* of 3.36 applying Partial Least Squares (PLS). This model was implemented in the portable device and a new set of samples were measured, obtaining measurements of the cocoa powder content with an accuracy of 3.5%. These results showed the suitability of the Visum Palm device to detect and quantify adulteration of cocoa powder applying a PLS correlation model. Furthermore, spectra pre-treatments applied before PLS successfully solved variations due to geometrical irregularities of plastic bags and sample granularity.

As a conclusion, in the framework of food quality control, the possibility of using spectral fingerprinting coupled to chemometrics implemented in a portable device such as Visum Palm for detecting cocoa counterfeit was demonstrated.

References:

[1] J. C. Moore, J. Spink, M. Lipp, Development and application of a database of food ingredient fraud and economically motivated adulteration from 1980 to 2010. Journal of Food Science, 77(4), R118-R126 (2012).
[2] Czech Agriculture and Food Inspection Authority: Press Release Polish cocoa adulteration classes in Kaufland (2012).

P-131

PREDICTION OF WINE SENSORY DESCRIPTORS USING INSTRUMENTAL DATA FUSION

E. Borràs[1], A. Valls[2], R. Boqué[1], J. Ferré[1], M. Mestres[2], L. Aceña[2], O. Busto[2]

[1]*Chemometrics, Qualimetrics and Nanosensors Group, Universitat Rovira i Virgili, Tarragona (Spain)*
[2]*iSens Group, Universitat Rovira i Virgili, Tarragona (Spain)*
alba.valls@urv.cat

Wine is a highly valued product with excellent sensory characteristics and healthy benefits. But it is also vulnerable to different distortions and deceptions [1]. To guarantee the safety and excellence of wines, regulatory bodies are more and more interested in developing appropriate analytical procedures, based on chemical and sensory evaluations. Wine sensory quality is determined by the perception of sensory attributes that are directly related to several chemical compounds. Human taste panels are the most common methodology to assess wine sensory quality, but they present some limitations such as high costs, subjectivity and lack of reference standards due to the great variability of the products (regions and varieties) [2]. For this reason, research is being developed to find more objective and robust methodologies, using instrumental analytical techniques based on spectrometric or spectroscopic fingerprints and acting as electronic panels [1,2]. The complexity of the wine matrices does not allow to find suitable correlations between sensory descriptors and one single instrumental technique; therefore, the combination of several techniques is required.
In this study, an FT-MIR electronic tongue, a headspace-mass spectroscopy (HS-MS) electronic nose and UV-Vis electronic eye were combined to predict wine sensory descriptors of 78 red and aged wines from the same harvest (2011) and region (DO Priorat). Values for the sensory attributes were provided by an expert taste panel.

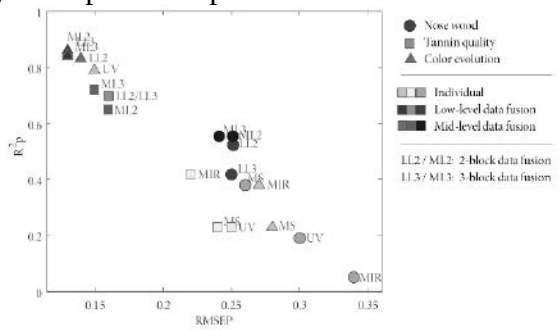


Figure 1. Determination coefficient (R^2) versus root mean square error of prediction (RMSEP) for all the models built.

Partial Least-Squares (PLS) regression models using individual instruments and different data fusion strategies at low- and mid-level (Figure 1) were built to predict sensory intensity values of nose wood, tannin quality and color evolution wine descriptors. The individual techniques were able to acceptably predict the sensory descriptors: HS-MS (0.26 RMSEP/0.38 R^2 for nose wood), FT-MIR (0.22 RMSEP/0.42 R^2 for tannin quality) and UV-Vis (0.15 RMSEP/0.79 R^2 for colour evolution). When fusing the data from the different instruments better results were obtained. Nose wood achieved higher correlations by combining MS and UV-Vis data (0.25 RMSEP/0.56 R^2), tannin quality (0.16/0.70) and colour evolution (0.13/0.86) by fusing both MIR and UV-Vis data, with slightly better results in all cases when applying mid-level fusion.

Acknowledgement: We thank the Spanish Ministry of Science and Technology (Project AGL2011-26456) for economic support.

References:

[1] S. Esslinger, J. Riedl, C. Fauhl-Hassek, Food Res. Int. 60 (2014) 189–204.
[2] A. Versari, V.F. Laurie, A. Ricci, L. Laghi, G.P. Parpinello, Food Res. Int. 60 (2014) 2–18.

P-132

GEOGRAPHICAL TRACEABILITY OF MOROCCAN ARGAN OILS ACCORDING TO ITS CHEMICAL FINGERPRINTING (FATTY ACID PROFILE, TOCOPHEROLS AND STEROLS COMPOSITION) AND CHEMOMETRIC ANALYSIS

M. Kharbach[1,2*], Y. Vander Heyden[2], A. Bouklouze[1]

1. *Pharmaceutical and Toxicological Analysis Research Team. Laboratory of Pharmacology and Toxicology, Faculty of Medicine and Pharmacy, University Mohammed V- Rabat- Morocco*
2. *Department of Analytical Chemistry and Pharmaceutical Technology, CePhaR, Vrije Universiteit Brussel (VUB), Laarbeeklaan 103, B-1090 Brussels, Belgium*

* *mokharba @vub.ac.be*

The Argan tree (*Argania spinosa L. Skeels*) is a tropical plant and represents the only endemic species of the genus *Argania* in Morocco. Currently, in Morocco the Argan forest covers an area of 840 000 ha including the fertile Souss valley region, the foothills of the Anti-Atlas mountains, and the coast region between Essaouira and Agadir [1-2]. Nowadays the origins of food are essential for import and export trading in order to ensure the traceability for consumers, traders or even food producers. Information about food's origin is necessary to verify its specifications and to guarantee its quality, because foods from different origin have distinct qualities [3-4].

Principal Component Analysis (PCA), and Partial Least Squares Discriminant Analysis (PLS-DA), were applied to data of content of the various chemical compound, fatty acid, tocopherols, sterols, or total data, to explore their capacity for classification and discrimination of Argan oil (AO), belonging to a Moroccan Protected Geographical Indication (PGI) according to their five geographical origins ('Ait-Baha', 'Agadir', 'Essaouira', 'Tiznit' and 'Taroudant').A total of 120 samples of PGI AO were analyzed between 2011 and 2014 harvests. Firstly, several physicochemical parameters were measured namely free acidity, peroxide value, spectrophotometric indices, fatty acid composition, tocopherolsand sterols content. The results obtained based on chemical composition indicate that the PCA was able to discriminate the samples in five classes. PLS-DA models gave a high prediction and an accurate discrimination of the samples from different region based on chemical composition.

Physicochemical data with chemometric tools can be recorded for the characterization and the quality traceability of the Moroccan Argan oils.

References

[1] Charrouf, Z., & Guillaume, D. Argan oil, the 35 years of research product. European Journal of Lipid Science and Technology, 116(10), 1316-1321, 2014.

[2] Matthäus, B., Guillaume, D., Gharby, S., Haddad, A., Harhar, H., & Charrouf, Z. Effect of processing on the quality of edible argan oil. Food chemistry, 120(2), 426-432, 2010.

[3] Gonzalvez, A., & de la Guardia, M. Basic Chemometric Tools. Food Protected Designation of Origin: Methodologies and Applications, 60, 299, 2013.

[4] Luykx, D. M., & Van Ruth, S. M. An overview of analytical methods for determining the geographical origin of food products. Food Chemistry, 107(2), 897-911, 2008.

P-133

FREE CHOICE PROFILE TO EVALUATE SENSORY ATTRIBUTES OF HONEY IN COMPETITION

S.F.L.S. Leal[1], M.B.S. Scholz[2], C.S.G. Kitzberger [2], A.F. Oliveira[1]

[1] Universidade Tecnológica do Paraná, Estrada dos Pioneiros Londrina-PR-Brasil
[2] Área de Ecofisiologia – Laboratório de Fisiologia Vegetal, IAPAR - Instituto Agronômico do Paraná, Rodovia Celso Garcia Cid, km 375 86047-902 - Londrina – PR, Brazil; mbscholz@iapar.br 1

Competitions that evaluate the sensory attributes of honey are ways for their valuation and acknowledgment of beekeepers [1]. Frequently hedonic evaluation by 3-5 panelists (consumers and beekeepers) is used in these events and this does not allow statistical test and the results are subjective [2]. Free choice profile (FCP) is a sensory technical statistically suitable to analyze data from the panelists without training after scaling, rotation and translation of scores applying the generalized Procrustes analysis (GPA). [3]. FCP was used to assess sensory attributes of 19 honey samples from different floral origins collected by Brazilian beekeepers. Eleven judges employed an average of 19 attributes to evaluate the attributes of appearance, aroma, flavor and texture. We observed low residues of the panelists and GAP-bidimensional solution is shown in Figure 1.

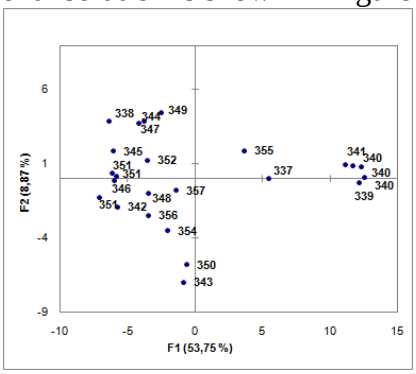


Figure 1. Honey positons in the plane definied by first two dimensions.

The attributes related to the positive dimensions (F1+ and F2+) and negative (F1- and F2-) and mentioned by several panelists promoted the separation of samples. Attributes of brightness and liquid appearance, aroma and flavor sweet and smooth texture were correlated with F1+. The samples creamy and fully crystallized textures are located in F1- dimension. Brightness, crystallized, transparent appearance and honey taste and aroma and smooth, creamy and crystallized textures were correlated with F2-. Samples 339, 340 and 341 are liquid, whereas the samples 355 and 337 have aspect pasty. Honey appearance and sweet, honey and wax aromas were used to describe the samples 338, 349, 352, 345 and 351. In honey competition the samples located of F2+ and F2- could be receiving the first and second award, respectively because they are associated with positive attributes. Samples in the F1+ have smaller award because had few positive attributes. So, the honey competition awards could be made on the basis of ratings analyzed mathematically without the subjective intervention of team leader.

References:

[1] CARI-Central Agriculture Research Institute. http://www.cari.be/medias/.../concours_mielvf2013.pdf. Accessed: January, 10 2013.
[2] M.M González, C., De Lorenzo, R. A. Pérez. Food Science and Technology International, 16, 19 (2010).
[3] G.B.Dijksterhuis, J.C Gower, Food Quality and Preference, 3, 67 (1992).

P-134

FUNCTIONAL PROPERTIES AND SENSORY ATTRIBUTES OF MODERN ARABICA COFFEE CULTIVARS

C.S.G. Kitzberger [1], M.B.S. Scholz [1], M.T. Benassi [2]

Área de Ecofisiologia – Laboratório de Fisiologia Vegetal, IAPAR - Instituto Agronômico do Paraná, Rodovia Celso Garcia Cid, km 375 86047-902 - Londrina – PR, Brazil; mbscholz@iapar.br 1
Departamento de Ciência e Tecnologia de Alimentos, UEL -Universidade Estadual de Londrina, Rod. Celso Garcia Cid (PR 445), Km 380 - Caixa Postal 10.011, 86057-970- Londrina- PR, Brazil 2

Positive sensory attributes and coffee contribution to health are decisive factor for commercialization and consumption of coffee. Hidrossoluble compounds such as phenolic compounds (PC), caffeine and melanoidins show antioxidant activity (AA) and trigonelline as flavor precursor have important role in functional and sensorial properties. The number of cultivars and growing conditions of Arabica coffees result in different chemical compositions and sensory attributes, which become challenging to combine functional and sensory coffee quality. This study aimed evaluation of diterpenes, AA and sensory attributes in different modern Arabica cultivars (Sarchimors: IPR 97, IPR104, IPR106, IPR107, IPR108; Catuaí SH_2SH_3: IPR100, IPR101, IPR105; Catuaí x Icatu: IPR102 and Icatu derived: IPR106) develop by IAPAR and growing in the same edaphoclimatic conditions. Hidrossoluble compounds were evaluated by HPLC while AA by ABTS, Folin Ciocalteau and conjugated dienes methods [1], melanoidins by spectroscopy method [2]. Free Choice Profile [3] analysis was used to describe sensorial attributes employed the Generalized Procrustes Analysis (GPA). Principal component analysis (PCA) was used to simultaneously assess the composition characteristics and AA, and was used to characterize different arabica coffee cultivars (Fig 1a). Conditions and processes postharvest were standardized, so functional and sensory properties are mainly attributed to genetic diversity. IPR97, IPR100 and IPR101 cultivars had higher AA by TEAC and dienes than the others cultivars. IPR 102, IPR104, PR105, IPR106, IPR107 and IPR108 cultivars showed high F-C. Sarchimor ana CatuaíxIcatu derived have higher values of trigonelline and 5-cafeoilquinic acid (5-CQA). Beverages of IPR100, IPR101, IPR105 and IPR106 cultivars showed positive attributes such as coffee color, turbidity, coffee, chocolate and sweet aromas, bitterness and fullbodied texture. IPR102, IPR104, IPR107 and IPR108 presented negative and positive attributes, while IPR97 showed attributes related to immature beans (Fig. 1b). Among modern cultivars, IPR100, IPR101, IPR105 and IPR106 showed high levels of nicotinic acid, caffeine and melanoidins and positive sensory attributes and higher AA, showing desirable functional properties and sensory attributes in the same cultivar. So, this study through multivariate analysis (PCA and GPA) its possible to demonstrate the existence of cultivars that dispense blends to obtain coffees which contain high quality sensory and many great health benefits.

Figure 1. a) Biplot of the first-two PCA, b) Consensus configuration of coffees from Mandaguari.

References:
[1] C.S.G. Kitzberger, M.B.S. Scholz, M.T. Benassi, Food Research International 61, 61 (2014).
[2] J.A. Vignoli, M.C. Viegas, D.G Bassoli, M.T. Benassi, Food Research International, 61, 279 (2014).
[3] C.S.G. Kitzberger, M.B.S. Scholz, J.B.G.D., Silva, M.T. Benassi, Brazilian Journal Food Technology, 38 (2011).

P-135

COMPREHENSIVE LCxLC-MS DATA RESOLUTION USING MCR-ALS

M. Navarro [1], J. Jaumot [1], T. van Beek [2], G. Vivó-Truyols [3], P.J. Schoenmakers[3], R. Tauler [1]

[1] Department of Environmental Chemistry, IDAEA-CSIC, C/Jordi Giroma 18-26, Barcelona, Spain
[2] Natural Products Chemistry Group, Wageningen University, Droevendaalsesteeg 4, Wageningen, Netherlands
[3]Van't Hoff Institute for Molecular Science, Universiteit van Amsterdam, Science Park 904, Amsterdam, Netherlands
meritxell.navarro@idaea.csic.es

Over the past years, the complexity of analytical samples to be analyzed is increasing in many applied research fields, such as environmental, health and food analysis. Liquid chromatography is often the analytical method selected for the analysis of these samples, due to its ability to resolve complex mixtures. However, one-dimensional liquid chromatography is often not capable of separating all the constituents present in these complex samples. Comprehensive two-dimensional liquid chromatography (LCxLC) appears to be a more powerful alternative to achieve a better separation. [1] Although LCxLC provides better peak capacity than one-dimensional LC, there is still the possibility that some of the constituents of the samples remain unresolved. In this work, an MCR-ALS based strategy is proposed to obtain a complete resolution of LCxLC-MS data. Furthermore, the possible use of trilinear models for the resolution of this type of data is studied. MCR-ALS results are compared with other chemometric methods suitable for three-way data sets, such as PARAFAC and PARAFAC2. [2] The proposed strategies were tested in the analysis of two experimental data sets.

The first data set corresponded to the analysis of triacylglycerols (TAGs) in corn oil samples. Although LCxLC-MS provided a rather good separation of TAGs, there were still some constituents showing unresolved co-elutions. In some cases, these compounds could be differentiated using the vendor software based on the differences in MS signals, but the distinction of TAGs positional isomers, such as PLO/POL, was unattainable. Application of the MCR-ALS based strategy to LCxLC-MS data allowed the successful resolution of all the components in corn oil samples, including TAGs positional isomers. The second data set corresponded to a non-targeted metabolomic study of rice growth. The complexity of the analysed rice samples and the fact that there was no prior knowledge of what metabolites they contained made very challenging the achievement of a complete resolution of them. Using a multisample column-wise augmented data matrix strategy, MCR-ALS resolved the elution profiles and mass spectra of the different metabolites and allowed their rapid identification using on-line libraries.

Results achieved in this work confirmed that the separation power of two-dimensional chromatography can be significantly improved when it is combined with chemometric methods. Moreover, among the applied methods, MCR-ALS showed the best data resolution, giving the best values of lack of fit (LOF) and percentage of explained variance (R^2).

Acknowledgement: The research leading to these results has received funding from the European Research Council under the EU Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n. 320737.

References:

[1] D.R. Stoll, X. Li, X. Wang, P.W. Carr, S.E.G. Porter, S.C. Rutan, Journal of Chromatography A, 1168, 3, (2007).
[2] H. Parastar, J.R. Radović, M. Jalali-Heravi, S. Diez, J.M. Bayona, R. Tauler, Analytical Chemistry, 9289, 83 (2011).

P-136

Effect of Adding Variables on Extents of Rotational Ambiguity Using Displacement method in Positive Matrix Factorization Solutions

F. Emami and P.K. Hopke

Center for Air Resources Engineering and Science
Clarkson University, Potsdam, NY USA
email phopke@clarkson.edu

One of the main problems that limit the use of the model-free analysis methods, like positive matrix factorization (PMF), for the resolution of multivariate data is that usually, there is rotational ambiguity in the result: different solutions (factors) provide equally good fits to the measured data. The PMF model imposes non-negativity of both source profiles and source contributions in order to reduce the rotational problem. However, such constraints are generally insufficient to ensure a unique solution [1]. The new version of EPA's positive matrix factorization (EPA PMF) software, 5.0, includes displacement (DISP) of factor elements as a error estimation method for analyzing factor analytic solutions. This method captures the uncertainty of PMF analyses due to rotational ambiguity. DISP diagnostics are consistently robust, indicating its use for understanding rotational uncertainty and as a first step in assessing a solution's viability [2]. To demonstrate the utility of the DISP method, results are presented for the submicron particles (12–470 nm), measured between January 2008 and December 2013 at New York State Department of Environmental Conservation (NYS DEC) site in Rochester, NY. The particle number size distribution (PNSD) of airborne particles not only provides us with information about sources and atmospheric processing of particles, but also plays an important role in determining regional lung depositon and dose [3].

Besides the PNSD, some gaseous species and black carbon (2 wavelenghts) have been measured in this study. Thus, the effect of adding these species in size distribution data as new variables could be investigated on DISP results. Generally, informative variables can improve accuracy and precision of resolving process. In special cases when the new added variables introduce the selective sensors or windows to the data, unique resolution of one or more profiles and contributions is possible [4]. The results of the DISP-PMF showed the effect of adding new variables on extents of rotational ambiguities, which were decreased significantly. This is an evidence for “data based uniqueness”.

Clever choosing the new variables for achieving more accurate resolution of environmental data by moderately high numbers of factors (such as 7 in this case), can be investigated systematically. Displacement method can estimate rotational ambiguity during step by step adding the new variables. Regarding environmental data and PMF results, finding informative variables will be useful. In this case, it is possible to get more robust results, with lower rotational uncertainty, that are more easily interpreted and can be used in a subsequent study of the potential changes in myocardial infractions (heart attacks) and fetal growth restriction (low birth-weight babies) for different periods of large change in energy use and resultant emissions.

References:

[1] P. Paatero, P. K. Hopke, B. A. Begumb, S. K. Biswas, Atmospheric Environment, 39, 193 (2005).
[2] S. G. Brown, S. Eberly, P. Paatero, G. A. Norris, Science of the Total Environment, 518, 626 (2015).
[3] T. V. Vu, J. M. Delgado-Saborit, R. M. Harrison, Atmospheric Environment, 122, 114 (2015).
[4] R. Rajkó, H. Abdollahi, S. Beyramysoltan, N. Omidikia, Analytica Chimica Acta, 855, 21 (2015).

P-137

LIQUID CHROMATOGRAPHY WITH MULTIVARIATE DETECTION AND CHEMOMETRIC DATA PROCESSING: A BASIS FOR GREEN-ANALYTICAL METHODS

R.B. Pellegrino Vidal, M.D. Carabajal, G.A. Ibañez, J.A. Arancibia, G.M. Escandar

Instituto de Química Rosario (CONICET-UNR), Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Suipacha 531 (2000), Rosario, Argentina
email escandar@iquir-conicet.gov.ar

The general objective of the present work was the development of methods based on green-analytical chemistry principles for the quantification of organic pollutants in complex samples [1,2], by coupling liquid chromatography, multivariate detection and chemometrics data processing [3].

Liquid chromatography with dual UV/diode array (DAD) and fluorescence (FLD) detections was carried out in a single run, and the second-order DAD-elution time and FLD-elution time data obtained were treated with MCR-ALS (multivariate curve resolution/alternating least-squares) algorithm [4]. In this way, while analytes are measured through their more appropriate (absorbance and/or fluorescence) signals, chemometric treatment of the corresponding matrices allows the resolution of total or partial overlapped bands, and to overcome the presence of interferences in real samples.

The two investigated systems involved emerging contaminants belonging to plastics derived endocrine disruptors (bisphenol A, nonylphenol, octylphenol, diethyl phtalate, dibutyl phthalate and diethylhexyl phthalate), and ten agrochemicals of frequent use, namely fungicides, herbicides, insecticides and a plant growth regulator. The former were investigated in both soft and alcoholic beverages (mineral water, soda, juice, wine and beer) while agrochemicals were quantified in land cultivated vegetables, including mushroom, lettuce, alfalfa sprout, cucumber, and celery.

The particularities of each investigated system and the differences in the corresponding data treatment are presented and discussed.

The coupling of chromatographic approaches with second-order chemometric calibration allows us to considerably simplify the sample pretreatment and to significantly reduce the analysis time, resulting in analytical green methods that favorably compare with those usually employed in complex systems such as the ones here investigated. Low detection limits were achieved, even in the presence of partially overlapped chromatographic and spectral bands among analytes and potential interferents. The detection capabilities allow one to monitor the presence of the analytes in natural samples complying with international regulations.

Acknowledgement: The authors are grateful to the Universidad Nacional de Rosario (Project BIO 237), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET, Project PIP 0163), and Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT, Project PICT 2013-0136) for financial support.

References:

[1] S. Armenta, S. Garrigues, M. de la Guardia, Trends Anal. Chem. 27, 497 (2008).
[2] M. Farré, S. Pérez, C. Gonçalves, M.F. Alpendurada, D. Barceló, Trends Anal. Chem. 29, 1347 (2010).
[3] A.C. Olivieri, G.M. Escandar, Practical three-way calibration, Elsevier, Waltham, USA (2014).
[4] R. Tauler, M. Maeder, A. de Juan, Multiset data analysis: extended multivariate curve resolution, in: S. Brown, R. Tauler, B. Walczak (Eds.), Comprehensive Chemometrics, 2, Elsevier, Oxford, 473 (2009).

P-138

Variable Recognition from UV-Vis Chromatograms Obtained By Liquid Chromatography and Application to Trypsin Digests Classification

C. Burgos-Simón[1], E.J. Carrasco-Correa[2], M. Beneito-Cambra[2], G. Ramis-Ramos[2]

[1]Department of Applied Mathematics, University of Valencia, 46100 Burjassot, Spain.
[2]Department of Analytical Chemistry, University of Valencia, 46100 Burjassot, Spain

A recurrent issue in Chemometrics is the recognition of variables in complex chromatograms. This problem is normally addressed by aligning; however, the real problem is not alignment but recognizing the signals produced by the same sources of variance, i.e. the same analytes. Protein digestion with trypsin leads to rather complex mixtures of peptides and therefore also to very complex chromatograms. Further, the presence of many peaks over the detection limit in a number of chromatograms and down the limit in others makes alignment methods based on correlation to fail. In this work, a recognition method for peptide peaks was developed and applied to enzyme class prediction after trypsin digestion. The information was preprocessed not by alignment but using direct peak recognition of the peptides. The enzymes were digested with trypsin, followed by liquid chromatography of the peptides with UV-Vis detection at three wavelengths. The training set was constructed with enzymes of three classes, namely proteases, amylases and cellulases, belonging to industrial enzymes that are usual components of cleaners [1, 2].

The peaks were identified using the following six peak identifiers: peak reduced retention time (established according to the trypsin peaks), relative area (percentage of total peak area for the recording at 214 nm), peak base width, peak high ratio at different wavelengths (260/214 nm and 280/214 nm) and asymmetry. After peak collection, and baseline modelling and subtraction, multiple comparisons at the local level between the chromatograms of the training set led to the construction of 3rd order tensors containing the peaks recognized as common to two or more than two chromatograms, as well as those not recognized as belonging to at least two chromatograms. To accept a recognized peak, a requirement was that the peaks should be recognized as the same when target and sample were reversed, i.e. in the comparison of chromatogram A with B and vice versa. Another requirement was to reach a minimum score for the weighted sum of the identifiers. In each comparison, scores were given to the peak candidates according to the ranked distances between the identifiers of the target peak and those of the candidates. Further, the identifiers were weighted, thus, a high weight was given to the height ratios at 260/214 nm and 280/214 nm, which are related to the percentage of aromatic amino acids in the peptides. The rows of the third order tensors were the chromatograms, the columns were all the peaks, including those not recognized as common to at least two chromatograms, and the entries were the vectors of the identifiers. Using the final tensor representing the whole training set to feed the SPSS data editor, an excellent LDA classification of the enzyme classes of the training set was achieved. The model was refined by using the peaks identified as the same in several chromatograms, and/or with higher scores, as nodes to shift the chromatograms before repeating the peak identification procedure at local level. Another refinement was to rank the chromatograms according to the proximity to the class centroids to select the targets.

Acknowledgement: Project CTQ2014-52765-R.

References:

[1] M. Beneito-Cambra, J.M. Herrero-Martínez, E.F. Simó-Alfonso, G. Ramis-Ramos, Rapid classification of enzymes in cleaning products by hydrolysis, mass spectrometry and linear discriminant analysis. Rapid Commun. Mass Spectrom. 3667 (2008).

[2] M. Beneito-Cambra, V. Bernabé-Zafón, J.M. Herrero-Martínez, E.F. Simó-Alfonso, G. Ramis-Ramos, Enzyme class identification in cleaning products by hydrolysis followed by derivatization with o-phthaldialdehyde, HPLC and linear discriminant analysis, Talanta 275 (2009).

P-139

DEVELOPMENT OF LOCAL PARTIAL LEAST SQUARES MODELS FOR CHROMATOGRAPHIC RETENTION PREDICTION OF PEPTIDES

Jan P.M. Andries[1], Mohammad Goodarzi[2], Yvan Vander Heyden[2]

[1]*Research Group Analysis Techniques in the Life Sciences, Avans Hogeschool, University of Professional Education, P.O. Box 90116, 4800 RA Breda, The Netherlands*

[2]*Department of Analytical Chemistry and Pharmaceutical Technology, Center for Pharmaceutical Research, Vrije Universiteit Brussel-VUB, Laarbeeklaan 103, B-1090 Brussels, Belgium*

Email: yvanvdh@vub.ac.be

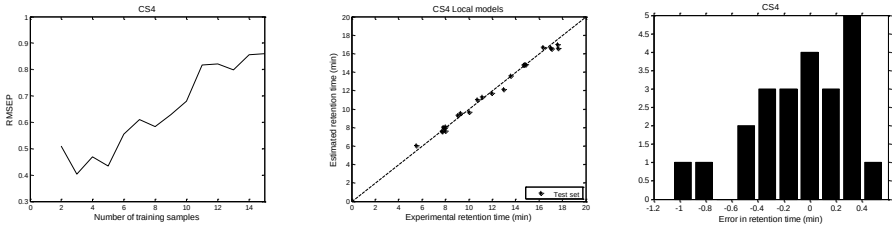
Keywords: Peptides / Prediction of retention / Local models

Aim: Development of local Quantitative Structure-Retention Relationships (QSRR) models for chromatographic retention prediction of peptides on 7 chromatographic systems, based on small calibration sets and calculated molecular descriptors.

Method: Retention data of 98 peptides for 7 Reversed Phase Liquid Chromatography (RPLC) systems are taken from Ref. [1]. Low retention times, which are situated nearby the dead time (*t*0), were eliminated. For the peptides, molecular descriptors are calculated with Dragon 6.0. All other calculations were made in Matlab 12.0. The data set was split in a training set and a test set, using the Kennard and Stone algorithm with software from the ChemoAC Toolbox.

FCAM variable selection [2], based on the absolute values of the PLS regression coefficients, REG, is carried out with autoscaled **X** data and mean-centered **y** data. The set of variables, corresponding to the minimum in the graph of RMSECV vs. the number of remaining variables, is used for the selection of the calibration samples to develop the local PLS models.

First, for each test sample individually, the *two* most similar training samples are determined, based on the Euclidean distances between the test sample and all training samples. A local PLS1-model is built with the two training samples, and the retention time of the test sample predicted. Then, the RMSEP is calculated, using the predicted retention times of all test samples. This procedure is repeated for the 3, 4, …, 15 most similar training samples. The optimal number of most similar training samples for the development of local models corresponds to the minimal RMSEP value, see Fig. Left for one RPLC system. For this optimal number of most similar training samples, predictions are made for all test samples (see Fig. Middle) and a histogram is made for the errors in the retention times, see Fig. (Right).



Conclusion

The results for the 7 chromatographic systems show that local PLS1 models with good predictive abilities can be built with two to four training samples.

References:

1. R. Put, Y. Vander Heyden, Proteomics 7 (2007) 1664.
2. J.P.M. Andries, Y. Vander Heyden, L.M.C. Buydens, Anal. Chim. Acta, 705 (2011) 292.

P-140

EVALUATION OF SEASONAL VARIATION IN UPLC FINGERPRINTS OF *CHRYSOBALANUS ICACO* L. (CHRYSOBALANACEAE) BY CHEMOMETRICS ASSOCIATED TO SOIL CHEMICAL ANALYSIS

N.E.N.P. Paracampo [1, 2], O.M.M. Teixeira [1], R.J. Poppi [2], J.A.F. da Silva [2]

[1]Embrapa Eastern Amazon, Tv. Dr. Enéas Pinheiro s/nº, 66095-100, Pará, Brazil.
[2]Chemistry Institute, State University of Campinas, R. J. de Castro nº 126, 13083-861, São Paulo, Brazil.
nadia.paracampo@embrapa.br

Chrysobalanus icaco leaves are traditionally used in Brazil as a hypoglycemic. This pharmacological activity has already been experimentally proved [1]. However, the biosynthesis of secondary metabolites are frequently affected by environmental conditions and variations in the total content and/or of the relative proportions of these compounds can take place [2] and alter the desired therapeutic effect. In this context, this work aims to study the effect of seasonality in UPLC fingerprints of *C. icaco* by chemometrics and to assess the soil fertility results. Twenty-two batches of wild *C. icaco* leaf and their soil samples were collected in the wet and dry season from 14 different cities in the northeast of the state of Pará (Brazil). The fingerprint analysis was performed on a Waters Acquity UPLC with photodiode array (PDA) detection acquiring at 273 nm. The hydroalcoholic extracts (70% v/v) (triplicate) were separated on an Agilent Zorbax Eclipse XDB-C18 (2.1 mm × 50 mm, 1.8 μm) column with a gradient mobile phase consisting of solvents A (1% aqueous formic acid, v/v) and B (acetonitrile). The flow rate was 0.3 mL/min and injection was 1 μL. Data files were converted to ARW files using Empower software (Waters), transferred to Excel® and input into Matlab® for peak alignment and chemometric analysis, by Principal Component Analysis (PCA). Correlated Optimized Warping (COW) [3] was used to accurate alignment of chromatograms. Then, the chromatograms were normalized and mean centered. Principal Component Analysis (PCA) was applied to evaluate the effect of seasonality in UPLC fingerprints of *C. icaco* samples. The first three PCs represent 80.01% of the total variance (PC1 = 42.56%, PC2 = 26.85% and PC3 = 10.60%). Examining the space defined by scores of first and second PCs, most samples of the same origin tends to cluster itself near. Indeed, some of *C. icaco* samples were located on the positive and negative side of PC2 axis separated like wet or dry season sample. Regarding soil chemical analysis, the macro elements (N, P, K, Ca and Mg) were analyzed quantitatively and it was observed that the higher potassium concentration determine the location of the *C. icaco* sample more negative in PC2 axis. The variables accounting for these separations were identified from the PC2 loadings plot. The results evidence that *C. icaco* samples from different geographical origin and harvested in the wet and dry season have varied the UPLC fingerprints. This approach should be refined to enable real and thorough understanding of plant interaction with the environment.

Acknowledgement: The authors are grateful to the Brazilian Agricultural Research Corporation for providing financial support.

References:

[1] P. A. S. Pollyanna. UFSE. 2011, 49p.
[2] L. Gobbo-Neto, N. P. Lopes. Química Nova. 2007, 30(2), 374-381.
[3] T. Skov, F. van den Berg, G. Tomasi, R. Bro, J. Chemometrics. 2006, 20, 484-497.

P-141

THE USE OF CHEMOMETRICS TO STUDY INDOLE ALKALOIDS FROM *Psychotria nemorosa*: EXTRACTION AND FRACTIONATION OPTIMIZATION BASED ON METABOLITE'S PROFILING

L.C. Klein-Júnior [1,2], J. Viaene [1], J. Salton [2], M. Koetz [2], A.L. Gasper [3], A.T. Henriques [2], Y. Vander Heyden [1]

[1]Department of Analytical Chemistry and Pharmaceutical Technology, Center for Pharmaceutical Research (CePhaR), Vrije Universiteit Brussel – VUB, Laarbeeklaan 103, B-1090 Brussels, Belgium
[2]Laboratory of Pharmacognosy and Quality Control of Phytomedicines, Faculty of Pharmacy, Universidade Federal do Rio Grande do Sul - UFRGS, Av. Ipiranga, 2752, 90610-000, Porto Alegre/RS, Brazil
[3]Herbarium Dr. Roberto Miguel Klein, Department of Natural Sciences, Universidade Regional de Blumenau - FURB, R. Antonio da Veiga, 140, 89012-900, Blumenau/SC, Brazil
e-mail: yvanvdh@vub.ac.be

The extraction is a difficulty to overcome in order to access the metabolome. Several approaches have been applied, mainly using the number of chemical features as optimization goal. However, these analyses may lack reliable data-handling methods. Thus, this study aims developing extraction and fractionation methods to access the alkaloids metabolite profile of *P. nemorosa*, a source of multifunctional indole alkaloids (IAs). Leaves of *P. nemorosa* were collected in Blumenau/SC, Brazil and dried, milled, and sieved. Based on earlier results, ultrasound assisted extraction was selected as extraction method. The alkaloid fraction was obtained by standardized liquid-liquid extraction (LLE), and analyzed by means of UPLC-DAD. In a first part, the extraction procedure was optimized, using a fractional factorial screening design (SD) to evaluate the significance of five factors: A) plant:solvent ratio (1:10 or 1:50, m/v), B) number of extractions (1 or 3 times), C) temperature (25 or 50 °C), D) time (10 or 60 min) and E) particle size (> 710 or ≤ 180 μm) on the Euclidean distance (*d*) between the profile and a blank chromatogram, and on the metabolite fingerprints. Distance *d* was used to quantify the dissimilarity between a blank and a fingerprint. Critical effects were estimated by Dong's algorithm. For optimization, temperature (X_1, 31-59 °C) and time (X_2, 24-66 min) were selected and examined via a central composite response surface design (RSD). Heights of important peaks, previously indicated by SD, were used as responses. In a second part, as an alternative to LLE, solid phase extraction (SPE) method was proposed, and the influence of X_I) sample concentration (50-250 mg/mL), X_{II}) % acetonitrile in dichloromethane (0-100%) and X_{III}) eluting volume (10-30 mL) were optimized in a second RSD (Box-Behnken) using sum of peak areas as response. SD analysis indicated that both E and A*D significantly affected the fingerprints dissimilarity, measured by d. To have detailed information how the factors affect the fingerprint, effects were calculated per time point and a critical effect (+ and -) was used as limit for graphical inspection. It was possible to detect thermo-labile peaks and to select these to be monitored in RSD optimization. X_1 and X_2 effects on the height of six peaks were modeled. Increasing X_2 resulted higher peaks, although at higher X_1 levels this effect was limited. For X_1, some peaks in the fingerprint were best extracted at low values (35-45 °C), while others were best at higher levels. To avoid compounds degradation, X_1 was set at 45 °C and X_2 at 65 min. For SPE fractionation, it was observed that X_I did not significantly affect the total area. However, X_{II}, X_{III}, the quadratic terms X_I^2 and X_{II}^2 and the interaction X_{II}*X_{III} had significant influence on the response. Based on contour plot analysis, X_I was set at 150 mg/mL, X_{II} at 30%, and X_{III} at 30 mL. Summarized, d and metabolite fingerprint analyses gave relevant responses for accessing IAs of *P. nemorosa*, allowing to develop reliable extraction and fractionation methods, avoiding degradation and decreasing time and amounts of solvent spent.

Acknowledgement: CNPq and FAPERGS/Brazil and FWO/Belgium for financial support.

P-142

SIMILARITY PARAMETERS TO DISCRIMINATE AND CLASSIFY HERBAL MEDICINES

J. Viaene [1], G. Alaerts [1], B. Dejaegher [1,2], Y. Vander Heyden [1]

[1] Department of Analytical Chemistry and Pharmaceutical Technology, Vrije Universiteit Brussel (VUB), Laarbeeklaan 103, B-1090 Jette (Brussel), Belgium.

[2] Laboratory of Instrumental Analysis and Bioelectrochemistry, Institute of Pharmacy, Université Libre de Bruxelles (ULB), Campus Plaine CP205/6, Boulevard du Triomphe, B-1050 Brussels, Belgium.

yvanvdh@vub.ac.be

Healthcare regulatory bodies (e.g. World Health Organization, European Medicines Agency, Food and Drug Administration, diverse Pharmacopoeias) allow a wide range of tools for identification and quality control of herbal medicines. Several of these tools include chromatographic analysis, which provides information on a multitude of sample compounds, and thus can be considered as fingerprinting. Nowadays fingerprints are accepted by the above mentioned institutions as tools to monitor the quality of herbal medicines.

Since fingerprints are composed of a multitude of measured points per sample, multivariate data analysis techniques are the tools of choice to extract the desired information. Linear or Quadratic Discriminant Analysis (LDA or QDA) and Soft Independent Modelling by Class Analogy (SIMCA), have been shown to provide satisfactory classification and discrimination using herbal chromatographic fingerprint data [1].

Similarity between fingerprints has been studied by Alaerts et al. [2] and warning and control limits for similarity parameters between pairs of fingerprints were defined. Among other similarity parameters, the Euclidean (d) was considered in the context of the identification of green tea samples and seemed to contain valuable information for identification purposes. The aim of our study was to explore the use of d to distinguish several groups of samples.

A data set, consisting of three types of *Curcuma* samples -i.e. 32 *C. longa* rhizoma, 29 *C.* rhizoma (different *C.* species possible) and 32 *C.* radix samples (different species possible, including *C. longa*)- was studied for this purpose. Duplicate analyses were performed per sample, applying the same chromatographic conditions. Chromatograms were aligned by means of Correlation Optimized Warping. Samples were split per type in a calibration and a test set applying a duplex algorithm on their averaged (n=2) fingerprints. Distribution of the calibration and test set fingerprints was evaluated by means of Principal Component Analysis (PCA). The d's between each pair of fingerprints were arranged in a matrix D. To estimate warning and control limits for a certain *Curcuma* type, the d's from its calibration set were extracted from D, resulting in submatrices $D_{C.\ longa}$, $D_{C.\ \text{rhizoma}}$ and $D_{C.\ \text{radix}}$. The off-diagonal elements of each submatrix were used to estimate upper and lower warning and control limit values for d of each *Curcuma* type. The calibration and test set fingerprints of all *Curcuma* types were compared to the warning and control limits of each *Curcuma* type.

Based on their behavior relative to the warning limits of the three classes, outliers could be identified and removed, and unknown samples could be assigned to a given class. The quality of the Euclidean distance based classification was quantified using the correct classification rate, and was found comparable to LDA and QDA and even better than SIMCA results.

References

[1] J. Viaene, M. Goodarzi, B. Dejaegher, C. Tistaert, A. Hoang Le Tuan, N. Nguyen Hoai, M. Chau Van, J. Queting-Leclercq, Y. Vander Heyden, Discrimination and classification techniques applied on Mallotus and Phyllanthus high performance liquid chromatography fingerprints. Anal. Chim. Acta, 877, 41-50 (2015).

[2] G. Alaerts, J. Van Erps, S. Pieters, M. Dumarey, A.M. Van Nederkassel, M. Goodarzi, J. Smeyers-Verbeke, Y. Vander Heyden, Similarity analyses of chromatographic fingerprints as tools for identification and quality control of green tea. J. Chromatogr. B, 910, 61-70 (2012).

P-143

Impact of UV radiation and urban atmospherics aerosols in paint dosimeters using Micro-Raman and Principal Component Analysis

A.Herrera[1], D. Ballabio [2], R. Todeschini[2], N. Navas[3] and C. Cardell[1]

[1] Dept. of Mineralogy and Petrology, Faculty of Science, University of Granada, Campus Fuentenueva s/n, 18071 Granada, Spain, Tel. 0034958242725, jaherrera@ugr.es.
[2] Dept. of Earth and Environmental Sciences University of Milano-Bicocca, P.za della Scienza 1, 20126 Milano, Italy
[3] Dept. Analytical Chemistry, and Biomedical Research Institute of Granada (IBIG), University of Granada, Faculty of Science, Campus Fuentenueva s/n, 18071 Granada, Spain .

Painted artworks undergo alteration under different scenarios, particularly in outdoor conditions, leading to changes in physico-chemical, mechanical and thus visual attributes. The main factors affecting paintings´ weathering are environmental parameters such as ultraviolet light (UV)[1], relative humidity (RH), temperature (T), or urban atmospheric pollutants (both particulate matter and gases) [2]. However regarding these latter determinants, there are relatively few studies that have examined the effects of urban air (atmospheric aerosols) on the pigment-binder interactions that take place in real and/or model paint samples (i.e. paint dosimeters)[3]. This work shows the results of micro-Raman data treated with Principal Component Analysis (PCA)[4] on paint dosimeters subjected to an accelerated UV ageing test and a natural ageing test, i.e. outdoor exposure to an urban atmosphere.

To this end, paint dosimeters were prepared according to medieval paint recipes mixing calcite (different crystal sizes) with egg yolk binder to mimic real paintings. Several paint layers of each binary paint mixture were applied on glass slides, which subsequently were exposed to the above-mentioned natural and accelerated ageing tests. Then, Raman spectra were collected from the blank paint dosimeters, the UV-aged samples at 1 month, and the naturally aged samples at 6 and 12 months of outdoor exposure. The Raman spectra were subjected to baseline subtraction, as well as to several types of pretreatments such as mean cantering, first derivative calculation by applying the Savitzky–Golay algorithm, and standard normal variate, as well as their combination. Afterwards PCA was applied to the Raman data. matrix by using spectral Raman regions containing relevant information, i.e. from 100 to 1800 cm–1 and from 2600 to 3800 cm–1. Multivariate analysis of spectra proved to be an excellent tool to assess changes in the naturally and artificially aged paint dosimeters. The loadings and score plots indicate that Raman spectra of blank, UV and naturally aged paint samples clearly separate in different clusters.

Acknowledgement Financial support was provided by Spanish Research Projects AERIMPACT (CGL2012-30729) and EXPOAIR (P12-FQM-1889), the European Regional Development Fund (ERDF), the Andalusian Research Groups RNM-179, FQM-118 and Milano Chemometrics and QSAR Research Group. Analyses were performed in the Scientific Instrumentation Centre (CIC) of the University of Granada (Spain).

References:

[1] E. Manzano, J. Romero-Pastor, N. Navas, L.R. Rodríguez-Simón, C. Cardell, Vib. Spectrosc, 53, 260 (2010).
[2] B. Horemans, C. Cardell, L. Bencs, V. Kontozova-Deutsch, K. De Wael, R. Van Grieken, Microchem, J. 99, 429 (2011).
[3] A. Herrera, N. Navas, C. Cardell, Talanta, (2016) [under review].
[4] D. Ballabio, V. Consonni, Anal. Methods, 5, 3790 (2013).

P-144

COMPARATIVE STUDY OF DIFFERENT THIRD-ORDER DATA GENERATION APPROACHES

M.R. Alcaráz, M. Montemurro, G.G. Siano, M.J Culzoni, H.C. Goicoechea

Laboratorio de Desarrollo Analítico y Quimiometría, Facultad de Bioquímica y Ciencias Biológicas, Universidad Nacional del Litoral, Ciudad Universitaria, 3000 Santa Fe, Argentina
email (alcarazmirtaraquel@gmail.com)

Over the last years, it has been demonstrated that the increase of multiway data dimensions has a positive impact on analytical figures of merit, e.g. higher sensitivity, lower limits of detection and quantitation, better selectivity, among others. First- and second-order data analyses have become excellent tools for the resolution of complex samples which would result experimentally challenging from the univariate calibration standpoint. On the other hand, even though no additional analytical advantages have been yet proved, third-order data analysis for analytical applications constitutes a field worth to be explored [1].
Although multidimensional instrumental signals are easy to be obtained with the available modern instrumentation, and several chemometric algorithms have been successfully developed to solve multi-way data problems, the way in which the multi-way data are generated may have a significant effect on the final results. In this work, a comparative study of different third-order data generation approaches was carried out. Three methods based on identical liquid chromatographic conditions but coupled to different emission and excitation fluorescence detection systems were developed for the quantitative analysis of antibiotics in aqueous matrices.
The first approach included the collection of several fractions at the end of the chromatographic procedure by means of a custom-built device that allows to collect fractions in a 96-wells ELISA plate. Then, emission and excitation spectra were registered for every fraction by using a spectrofluorometer equipped with a plate reader accessory coupled to an optical fiber and a gated photomultiplier. In this way, 25 emission-excitation matrices (EEM, size: 17×25) were obtained for each chromatographic run. [2] A second strategy was developed by using a 10 µL flow-cell connected at the end of the chromatographic instrument and placed in a fast-scanning spectrofluorometer. Here, it was possible to register sequential EEMs for a unique chromatographic run. Since the fast-scannig spectrofluorometer takes only few seconds to register each EEM, it was necessary neither to stop the chromatographic flow nor to collect fractions after the chromatographic procedure, and 25 sequential EEMs (size: 7×45) were obtained for each chromatographic run. Finally, a multi-chromatographic run method involving a liquid chromatographic instrument coupled to a fast-scanning fluorescence detector which allowed to register time-emission fluorescence data matrices in a specific spectral range at a fixed excitation wavelength was developed. In order to build excitation-emission data matrices, eight chromatographic runs at different excitation wavelengths (time-emission matrix size: 45×150) were required for the same sample.
The three methodologies aforementioned were evaluated using different algorithms, such as PARAFAC, *A*PARAFAC, PLS-RTL and MCR-ALS, and selectivity, sensitivity, robustness, and time processing were evaluated. Since the data generation was different, each methodology required a particular data pre-processing including smoothing, peak alignment, and baseline correction, among others. Furthermore, due to differences in sensitivity provided by the implementation of a variety of detection mode it was necessary to assess several sample preparation methods in order to reach good analytical figures of merit.

Acknowledgement: Authors are grateful to UNL, CONICET and ANPCyT. .

References:
[1] A. Olivieri, Anal Chem, 80, 5713 (2008).
[2] M. Alcaráz, G. Siano, M. Culzoni, A. Muñoz de la Peña, H. Goicoechea, Anal Chim Acta, 809, 37 (2014).

P-145

THIRD ORDER MULTIVARIATE CALIBRATION APPLIED TO THE QUANTITATION OF AZINPHOS-METHYL IN FRUITS BY EXCITATION-EMISSION-KINETIC MEASUREMENTS

M. Montemurro, G. G. Siano, M.J. Culzoni, H.C. Goicoechea

Laboratorio de Desarrollo Analítico y Quimiometría, Facultad de Bioquímica y Ciencias Biológicas, Universidad Nacional del Litoral, Ciudad Universitaria Paraje El Pozo, 3000 Santa Fe, Argentina
milagrosmontemurro@gmail.com

Azinphos-methyl (AZM) is a widely used organophosphate insecticide and acaricide, with demonstrated negative impacts on the environment [1]. The weak natural fluorescence of AZM in aqueous solution can be enhanced in a variety of ways, including UV photolysis, inclusion into cyclodextrins, and base hydrolysis. Upon absorption of UV-A radiation, this molecule undergoes photolysis to the highly fluorescent compound N-methylanthranilic acid, which undergoes subsequent photolysis to photochemically stable products [2].

Four-way data generated by excitation-emission fluorescence matrices (EEFMs) measured as a function of reaction time can be modelled with appropriate higher order algorithms to achieve the second order advantage and improve both the selectivity and sensibility of the method.

In this work, we developed a fluorescent kinetic method for the determination of AZM in fruit samples. It is based on third-order data, obtained by measuring the time evolution of the EEFMs of the photolysis of AZM in alkaline medium.

The experiments were carried out with an instrument consisting in an optic fiber connected to an UV radiation source, coupled to a fluorescence spectrophotometer. The matrices were recorded in a quartz cell of 1 cm of path length. The readings were made in the excitation range of 220 to 320 nm every 5 nm, and emission from 320 to 500 nm every 5 nm at a scan rate of 24,000 nm min^{-1}. EEFMs were measured every 30 seconds for six minutes, thus having 21 × 33 × 13 data points.

Calibration and validation sets, consisting in five concentration levels of AZM each, were analyzed. The validation samples were prepared by adding fuberidazole, bitertanol and thiabendazole as uncalibrated interferences. Data modelling was performed with the algorithms MCR-ALS, PARAFAC and U-PLS/RTL. Relative error prediction and analytical figures of merit were calculated. The results obtained were similar for the three algorithms.

The method was further used for the quantitation of AZM in apple, pear, peach and plum. For this purpose, the standard addition calibration method was used in order to overcome matrix effect, which involved spiking three levels of AZM standard to the real samples. The sample preparation procedure consisted in an extraction with acetonitrile followed by dispersive liquid-liquid microextraction. Recoveries and figures of merit were also calculated for real samples.

Acknowledgement: The authors are grateful to Universidad Nacional del Litoral, CONICET and ANPCyT.

References:

[1] L. Yeasmin, S.A. MacDougall, B.D. Wagner, J. Photochem. Photobiol. A: Chem., 204, 217 (2009)
[2] B.D. Wagner, A.C. Sherren, M.A. Rankin. J. Chem., 80, 1210 (2002)

P-146

THE USE OF CHEMOMETRICS TO STUDY INDOLE ALKALOIDS FROM *Psychotria nemorosa*: INDICATION OF PEAKS RELATED TO THE INHIBITION OF BUTYRYLCHOLINESTERASE AND MONOAMINE OXIDASE-A

L.C. Klein-Júnior [1,2], J. Viaene [1], J. Salton [2], A.L. Gasper [5], A.T. Henriques [2], Y. Vander Heyden [1]

[1]Department of Analytical Chemistry and Pharmaceutical Technology, Center for Pharmaceutical Research (CePhaR), Vrije Universiteit Brussel – VUB, Laarbeeklaan 103, B-1090 Brussels, Belgium
[2]Laboratory of Pharmacognosy and Quality Control of Phytomedicines, Faculty of Pharmacy, Universidade Federal do Rio Grande do Sul - UFRGS, Av. Ipiranga, 2752, 90610-000, Porto Alegre/RS, Brazil
[3]Herbarium Dr. Roberto Miguel Klein, Department of Natural Sciences, Universidade Regional de Blumenau - FURB, R. Antonio da Veiga, 140, 89012-900, Blumenau/SC, Brazil
e-mail: yvanvdh@vub.ac.be

Some *Psychotria* plants are used by Amazon Indians to prepare Ayahuasca, a hallucinogenic beverage. In addition, some tribes from Middle America use *Psychotria* species for the treatment of dementia. In fact, our research group has demonstrated the modulatory action of *Psychotria* alkaloid fractions and isolated compounds on enzymes related to neurodegenerative disorders. One of these species is *Psychotria nemorosa*, which displays prominent inhibitory activity on butyrylcholinesterase (BChE) and monoamine oxidase-A (MAO-A). As a strategy to indicate potential multifunctional indole alkaloids, a chemometric approach was used in this study. Forty three samples of *P. nemorosa* leaves, collected in different Brazilian regions, were submitted to ultrasound assisted extraction. In order to obtain an alkaloid enriched fraction, the solid-phase extraction technique was used. The fractions were analyzed by means of UPLC-DAD and assayed for their BChE and MAO-A inhibitory potencies. The IC_{50} ranged from 2.8 to 74 µg mL^{-1}, and from 1.0 to 18.3 µg mL^{-1}, respectively. The chromatographic fingerprint data was first aligned using correlation optimized warping and Principal Component Analysis was applied to explore the data structure. Linear multivariate calibration techniques, namely Partial Least Squares (PLS) and Orthogonal Projections to Latent Structure (O-PLS), were evaluated for modelling the activities as a function of the fingerprints. The regression coefficients from PLS modelling were very noisy, making the regression coefficients plot interpretation and indication of potentially active peaks difficult. On the other hand, the O-PLS model demonstrated lower error (RMSECV = 9.3 and 3.3 for BChE and MAO-A, respectively), and an improvement in the interpretability of the regression coefficients was seen. Plotting these regression coefficients relative to the original fingerprints, four peaks were indicated as multifunctional compounds, with the capacity to impair both BChE and MAO-A activities. In order to confirm these results, a semi-prepHPLC technique was used and a fraction containing the four peaks was purified and evaluated *in vitro*. It was observed that the fraction exhibited an IC_{50} of 2.12 µg mL^{-1} for BChE and 1.07 µg mL^{-1} for MAO-A. These results reinforce the prediction obtained by O-PLS modelling, confirming these four compounds as multifunctional indole alkaloids.

Acknowledgement: CNPq and FAPERGS/Brazil and FWO/Belgium for financial support.

P-147

Monitoring changes in organic matter during an ultrafiltration stage at a drinking water treatment plant using fluorescence spectroscopy and PARAFAC

Jose Luis Cortina [1], Oriol Gibert [2], Mercè Granados [3] , Jose Luis Beltran[4], Marc Vera[5]

[1]Department of Chemical Engineering, Universitat Politècnica de Catalunya (UPC), Diagonal 647, E-08028, Barcelona, Spain
[2]Department of Analytical Chemistry, Universitat de Barcelona, Martí i Franquès 1, E-08028, Barcelona, Spain
marc.vera.canudas@estudiant.upc.edu

Fluorescence excitation-emission matrix (EEM) combined with parallel factor analysis (PARAFAC) was used for charatcterising and assessing changes in dissolved organic matter (DOM). Fluorescence analysis were carried out in an ultrafiltration (UF) stage at a Drinking water treatment plant in the Barcelona area. The excitation-emission matrices were further decomposed into individual components using PARAFAC and a model of three components was obtained. According to literature, the regions were related to protein-like, humic-like and microbial by-products.

Water streams feeding the UF stage showed high variability among the identified organic components, being particularly significant when groundwater was filtered as compared to sand filtered water from the river. Permeate streams showed little rejection in organic matter due to ultrafiltration. Moreover, the backwash stream did not show a significant increase in organic matter either. A principal component analysis (PCA) was performed in the backwash stream and the scattering regions of the samples exhibited a high content in particulate/colloidal matter.

A lab-scale investigation was also conducted to correlate the water quality with membrane performance upon filtration. Different water sources were tested, including synthetic solutions of humic acid (HA) and bovine serum albumin (BSA). The results identified both humic-like and protein-like components as significant membrane foulants. Fluorescence spectroscopy was proved to be a succesful tool to monitor changes in organic matter and could be implemented in the future to anticipate fouling events.

References:

[1] H. Yu, F. Qu, H. Liang, Z. shuang Han, J. Ma, S. Shao, H. Chang, and G. Li, "Understanding ultrafiltration membrane fouling by extracellular organic matter of Microcystis aeruginosa using fluorescence excitation-emission matrix coupled with parallel factor analysis," *Desalination*, vol. 337, no. 1, pp. 67–75, 2014.

[2] H. Yu, F. Qu, H. Chang, S. Shao, X. Zou, G. Li, and H. Liang, "Understanding ultrafiltration membrane fouling by soluble microbial product and effluent organic matter using fluorescence excitation-emission matrix coupled with parallel factor analysis," *Int. Biodeterior. Biodegrad.*, vol. 102, pp. 56–63, 2015.

[3] S. Shao, H. Liang, F. Qu, H. Yu, K. Li, and G. Li, "Fluorescent natural organic matter fractions responsible for ultrafiltration membrane fouling: Identification by adsorption pretreatment coupled with parallel factor analysis of excitation-emission matrices," *J. Memb. Sci.*, vol. 464, pp. 33–42, 2014.

[4] R. H. Peiris, M. Jaklewicz, H. Budman, R. L. Legge, and C. Moresoli, "Assessing the role of feed water constituents in irreversible membrane fouling of pilot-scale ultrafiltration drinking water treatment systems.," *Water Res.*, vol. 47, no. 10, pp. 3364–74, Jun. 2013.

P-148

ASSESSMENT OF BIOLOGICAL WASTEWATER TREATMENT SYSTEMS BY PROTOZOA AND METAZOA MONITORING AND CHEMOMETRICS ANALYSIS

A.L. Amaral [1,2], C.S. Leal [1], A.I. Vaz [1], J.C. Vieira [1], A.C. Quinteiro [1], M.L. Costa [3], L.M. Castro [1,4]

[1]Instituto Politécnico de Coimbra, ISEC, Rua Pedro Nunes, Quinta da Nora, 3030-199 Coimbra, Portugal
[2]CEB - Centre of Biological Engineering, Universidade do Minho, Campus de Gualtar, 4710-057 Braga, Portugal
[3]CERNAS - Center of Studies on Natural Resources, Environment and Society, Instituto Politécnico de Coimbra, ESAC, Bencanta, 3045-601 Coimbra, Portugal [4]GERST/ CIEPQPF – Faculty of Sciences and Technology, Universidade de Coimbra – Pólo II, 3030-790 Coimbra Portugal
lpamaral@isec.pt

Protozoa and metazoa communities in wastewater treatment plants (WWTP) are known to be dependent of the system itself (conventional activated system – CAS, oxidation ditch – OD, trickling filter – TF, etc.) and of working operational parameters (incoming effluent, dissolved oxygen, nitrification, hydraulic and sludge retention times, transient phenomena, etc.) [1,2]. Thus, for similar systems operating in comparable conditions it is expected to find the same protozoa and metazoa communities whilst differing from dissimilar WWTP. As such, the study of the protozoa and metazoa biota has already been employed for assessing the functioning of AS systems [3].

In the current study the protozoa and metazoa communities of three different types of WWTP, comprising one OD, four TF (TF1 to TF4) and three CAS (CAS1 to CAS3) reactors, were determined for each system characterization. Therefore, the metazoa contents, as well as the main protozoa groups (crawling, free-swimming and sessile ciliates, testate amoeba and flagellates) were determined in terms of contents and relative abundance. The collected data was further processed by principal components analysis (PCA) and the three first principal components (PC1, PC2 and PC3) were subsequently used for the overall characterization.

The obtained PC1 and PC2 results allowed to clearly individualize the related biota groups in different quadrants (testate amoeba and metazoa, linked to nitrification and high sludge retention times – SRT, in the up-left; swimming ciliates and flagellates, linked to transient phenomena, deficient aeration or low SRT, in the up-right; and sessile and crawling ciliates in the down-right quadrant). Furthermore, it could be found a clear division of the OD, CAS1 and CAS3 systems (in the left quadrant) regarding the four TF and CAS2 systems (in the right quadrant). This is in accordance to the fact that the OD is a extended aeration system, CAS1 presented high SRT (around 40 days) and CAS3 high nitrification abilities, thus leading to high contents (and correspondent relative abundance) of testate amoeba and metazoa. Given the high nitrification ability of CAS3, testate amoeba clearly predominated in that system leading to a clear distinction in the PC plot regarding CAS1 and OD. On the other hand, TF and CAS2 systems presented higher contents on flagellates and ciliates species, and were thus allocated towards the right quadrant. Furthermore a clear distinction could also be found between TF (with noticeable contents of flagellates and swimming ciliates) and CAS2 (high sessile and crawling ciliates contents). On the contrary, the four studied TF reactors appeared grouped together due to the fact that all belong to the same WWTP, and presented, to a given extent, similar operational conditions.

References:

[1] J.P. Canler, J.M. Peret, P. Duchéne, E. Cotteux. Aide au diagnostic des stations d'épuration par l'óbservation microscopique dês boues actives. Cemagref Editions, Bordeaux (1999).
[2] P. Madoni. Water Res. 28:67-75 (1994).
[3] L.A. Santos, V. Ferreira, M.O. Pereira, A. Nicolau. Eur. J. Protistol. 50, 319-328 (2014)

P-149

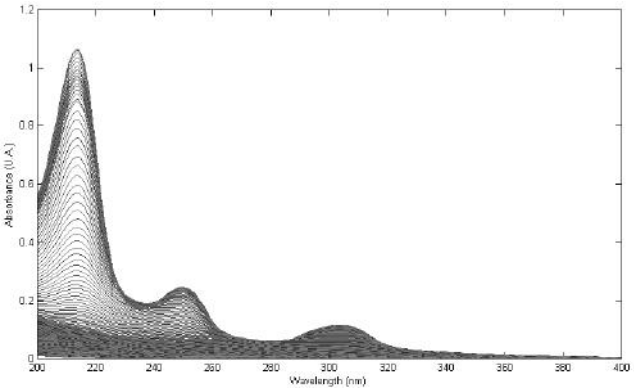
PHOTODEGRADATION PROFILE OF MYCOPHENOLIC ACID AT DIFFERENT STRESSING CONDITIONS BY MULTIVARIATE CURVE RESOLUTION OF SPECTROSCOPIC AND CHROMATOGRAPHIC DATA

Marc Marín-García[1], Giuseppina Ioele[2], Helena Franquet-Griell[1], Silvia Lacorte[1], Gaetano Ragno[2], Romà Tauler[1]

[1]Department of Environmental Chemistry, IDAEA-CSIC, Jordi Girona 18-26, 08034 Barcelona, Spain.
[2]Department of Pharmaceutical Sciences, University of Calabria, 87036 Rende (CS), Italy.
email: marcmaring@gmail.com

Multivariate curve resolution-alternating least squares (MCR-ALS) was extended to the simultaneous analysis of multiple data sets from different instrumental techniques from a detailed study of the kinetic photodegradation of mycophenolic acid (MPA) in environmental sample matrices. MPA is an immunosuppressant drug widely used to prevent a rejection in organ transplant treatments, normally at kidney. It is a non-competitive, selective and reversible inhibitor of inosine monophosphate dehydrogenase (IMPDH), an enzyme needed for the growth of lymphocytes, and responsible for the immune response of the body against external agents [1]. Nowadays many classes of drugs are being detected in wastewaters, after their excretion from human body (by urine and faeces), and due to their incomplete removal in wastewaters treatment plants, water quality is affected [2]. No previous evidences about the degradation of MPA in environmental sample matrices have been reported.

In this work, the effect of some parameters, like pH, on the stability of the MPA drug exposed to stressing irradiation conditions have been study. Photodegradation experiments (figure) were performed according to International Conference on Harmonization rules [3]. Concentration profiles of all the species involved in the photodegradation process as well as their pure spectra were resolved by MCR-ALS, and kinetic rate constants were estimated.



Experiments with different light sources (UV low-pressure mercury vapor lamp with a maximum of emission at 253.4 nm and a power of 15 W and sunlight simulator SUNTEST at several irradiation power) and pH values in the range 3 – 11 were investigated. Results of the MCR analysis explained the mechanism of the photodegradation process and provided the kinetic profiles of the photoproducts, together with their pure spectra and allowed the evaluation of their rate constants.

Photodegradation profiles in the applied conditions showed different degradation rates, with the formation of three different photoproducts. Degradation in the SUNTEST cabinet sunlight simulator was slower than the degradation in a reactor with a mercury lamp.

References:

[1] A.I. Sánchez-Fructuoso, Nefrología. 25, 234 (2005).
[2] C. Gómez-Canela, F. Ventura, J. Caixach, S. Lacorte, Anal. Bioanal. Chem. 406, 3801 (2014).
[3] ICH Q1A(R2), International Conference on Harmonization. IFPMA. Geneva, Switzerland, 2003.

P-150

SPECTROSCOPIC AND CHROMATOGRAPHIC MULTIVARIATE CURVE RESOLUTION INVESTIGATION OF THE STABILITY PROFILE OF TAMOXIFEN IN ENVIRONMENTAL SAMPLE MATRICES

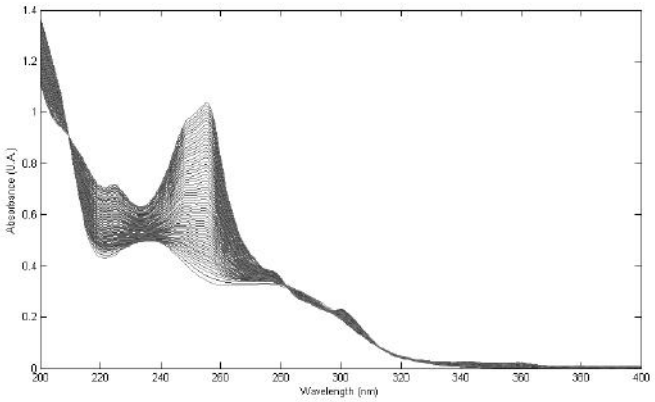
Giuseppina Ioele[1], Marc Marín-García[2], Helena Franquet-Griell[2], Silvia Lacorte[2], Gaetano Ragno[1], Romà Tauler[2]

[1]Department of Pharmaceutical Sciences, University of Calabria, 87036 Rende (CS), Italy.
[2]Department of Environmental Chemistry, IDAEA-CSIC, Jordi Girona 18-26, 08034 Barcelona, Spain.
email: marcmaring@gmail.com

An advanced and powerful chemometric approach is proposed for the analysis of multiset data obtained by fusion of coupled liquid chromatographic DAD/MS data with UV spectrophotometric data from the degradation experiment of tamoxifen (TAM) in solution and in environmental matrices.

The occurrence of this drug in environmental sample matrices has increased considerably in the least years because of its frequent use to treat some types of breast cancer in men and women [1]. Despite the physico-chemical properties of TAM, such as its low water solubility, this drug may enter easily into different environmental compartments [2]. The sensitivity to light of tamoxifen is well known, but its photodegradation profile and photoproducts have not been studied yet.

In this work, the tamoxifen drug was irradiated in a light cabinet SUNTEST, at 400 W/m^2, according to the International Conference on Harmonization rules [3]. The effects of temperature alone and of a combination of light and temperature were also investigated. Degradation profiles were monitored by UV spectrophotometric detection (figure) and aliquots at different reaction times were further analyzed by HPLC DAD/MS.



The Multivariate Curve Resolution-Alternative Least Squares (MCR-ALS) method was applied to describe the drug photodegradation process. This strategy involved the simultaneous analysis of multiple experimental data sets from different analytical platforms, including UV spectrophotometry and DAD/MS-chromatography. Resolution of the obtained multiset data allows the simultaneous description of the kinetic degradation process of this drug and the MS identification of the products formed after temperature and light exposures.

According to the literature [2] and to preliminary MCR-ALS results, the degradation process will include the formation of a photoproduct due to the isomerization reaction of the pure compound and subsequently, the cyclization of both, the pure compound and the isomer, respectively.

References:

[1] S. Iacobelli, M.E. Lippman, G. Robustelli della Cuna, The role of Tamoxifen in breast cancer, Raven Press, New York, 1982.
[2] M. Della Greca, M.R. Iesce, M. Isidori, A. Nardelli, L. Previtera, M. Rubino, Chemosphere, 67, 1933 (2007).
[3] ICH Q1A(R2), International Conference on Harmonization. IFPMA. Geneva, Switzerland, 2003.

P-151

B-HYDROXY FATTY ACIDS AS ENVIRONMENTAL MARKERS OF GRAM-NEGATIVE STRAINS

E.Paci [1], A.M. Marcelloni [1], E. Paba [1], G.Tranfo [1], F. Marini [2], A. Chiominto [1]

[1]Department of Occupational and Environmental Medicine, Epidemiology and Hygiene, 1INAIL, Research Area of Monte Porzio Catone (RM), Via di Fontana Candida 1, Italy
2 Department of Chemistry, Sapienza University, P.Le Aldo Moro 1, Rome Italy

The concentration of β-hydroxy-dodecanoic (C12) and β-hydroxy-tetradecanoic (C14) acids can be determined by HPLC coupled to triple quadrupole mass spectrometer for the quantitative determination of lipopolysaccharide in ATCC Gram-negative strains: the analysis of different bacterial strains showed that samples contain C12 and C14 acids in variable ratio, and that the sum of their concentrations showed a significant correlation of with results of LAL-test [1]. The objective of the present experiment was to measure the concentration of C12 and C14 in samples of airborne particulate matter and to verify if this information can be used for measuring the environmental contamination from endotoxins and/or Gram-negative bacteria. Stationary inhalable dust samples were collected in different occupational settings using airChek2000 pumps, at a flow rate of 2 l min^{-1}, with IOM sampler and fiberglass filters (pore size of 1,6 μm). In this study a total of 50 air samples were collected, 2 simultaneously at each sampling site 2. 25 filters were extracted and subjected to alkaline hydrolysis in order to release the β-hydroxy fatty acids from the bacterial membrane. These extracts were then analyzed by HPLC-MS/MS for their content of C12 and C14 and expressed as airborne contamination (ng/m^3). For endotoxin concentration (expressed as airborne pyrogenicity in EU/m^3) 25 filters were extracted and analyzed using a quantitative Kinetic Chromogenic LAL assay. All bacterial strains contained in each sample were identified.
Results show that at the concentration present in environmental samples the C12 is always below the analytical detection limit, and C14 is the only marker of lipopolysaccharide presence: however, as C12 is detected only at high levels of endotoxin (≈200 EU/m^3), it could be a useful chemical marker for a preliminary screening of the environmental contamination. The whole data set was then subjected to PCA in order to evidence the correlation between the experimental indices and the different bacterial strains. In particular, it was shown that the amount of endotoxins positively correlates with the measured fatty acids. Moreover, depending on the samples, different correlations with groups of bacterial strains was evidenced.These preliminary results demonstrated that the subject is complex but worth of further investigations and chemometrics will be used for analyzing the results.

References: [1] Validation of an high performance liquid chromatography-tandem mass spectrometry method for β-hydroxy fatty acids as environmental markers of lipopolysaccharide, A. Chiominto, A.M. Marcelloni, G. Tranfo, E. Paba, E. Paci. Journal of Chromatography A, 2014. 1353; 65 - 70

P-152

EVALUATION OF SECOND-ORDER MULTIVARIATE CALIBRATION AND FLUORESCENCE SPECTROSCOPY FOR MERCURY DETERMINATION IN ENVIRONMENTAL SAMPLES

M.A. Bravo, V. Escobar, S. Parra, J. Bergmann, W. Quiroz

Instituto de química, Pontificia Universidad Católica de Valparaíso, Valparaiso, Chile.
manuel.bravo@pucv.cl

Mercury has been declared a contaminant of global interest due to its high degree of persistence in the environment and its toxicity to different biological species. Exposure of humans to mercury and its compounds can lead to ataxia, loss of sensitivity in the hands and feet, weakness and, in extreme cases, paralysis and death. Several studies have showed that the toxicity of mercury depends on the concentration of its different chemical forms. In order to assess the exposition risk, it is extremely important have quantitative knowledge about mercury in environmental samples. For mercury determination, elemental techniques such as Atomic Absorption (AAS), Atomic Fluorescence (AFS) and inductively coupled plasma-mass spectrometry has been used.

In this work, a method based on fluorescent spectroscopy and second order multivariate calibration is evaluated for determination of inorganic mercury (Hg^{2+}) and methylmercury (CH_3Hg) in synthetic and real aqueous samples. A chemosensor based in a selenolactone [1] was used to quantify mercury species by fluorescence-induced. The experimental conditions are optimized in order to improve the analytical figures of merit of the proposed methodology. Two strategies for second order data generation were evaluated: A) Excitation-emission matrices (EEM); and B) Emission spectra-reaction time. The three-way data obtained were analyzed using Parallel Factor Analysis (PARAFAC), Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS) and Unfolding Partial Least Squares-Residual Bi-linearization (UPLS-RBL). Finally, the proposed methodology was applied to quantify Hg^{2+} and CH_3Hg in synthetic and real samples, in presence of uncalibrated interferences.

Acknowledgement: The authors thanks to Conicyt-Chile for financial support (Projects 1150950).

References:

[1] X. Q. Chen, K. H. Baek, Y. Kim, S. J. Kim, I. Shin, J. Yoon, Tetrahedron, 66, 4016 (2010)

P-153

MULTIVARIATE OPTIMIZATION OF PHOTOBIOREACTORS FOR WASTEWATER TREATMENT UTILIZING MICROALGAE AND LEDS

L.M.L. Silva[1], A.F.Santiago[1], G.A. da Silva[2], F. Vassoler[1], L.B. de Lima[1], L.P. do Amaral[1], M.C.R. Braga[1], R.S.L. Nascimento[1]

[1]Departamento de Engenharia Civil, Universidade Federal de Ouro Preto (UFOP), MG, Brazil
[2]Departamento de Química, Universidade Federal de Ouro Preto (UFOP), MG, Brazil
gilmare@gmail.com.br

The increase in population has resulted in an increase in the volume of domestic sewage, whose present in its composition water, organic matter, macro and micro nutrients, and pathogens. Among the nutrients found in wastewater, nitrogen and phosphorus are of critical environmental importance because they are essential for the growth of living things. But the release into water bodies can trigger the phenomenon of eutrophication. The microalgae can be used to remove and recover nitrogen and phosphorous from wastewater. These microorganisms can be grown in open systems, such as in high rate ponds (LAT) or in photobioreactors, which have the advantage of greater production of biomass [1]. To ensure a better energy utilization it must be selected a specific range of wavelength which promotes the best growth of algae, or removal of a particular nutrient [2]. The light emitting diodes (LEDS) are good options, as they have low power consumption, low heat generation and high durability. Moreover, they are free of toxic or polluting the environment. The use of photobioreactors with microalgae and illuminated by LEDS is presented as an alternative technology, since these systems can have low operating costs, simple operation and produce algal biomass that can be converted into new products such as bio-fertilizers and bio-diesel. From this, this study aimed to optimize the removal of nitrogen, phosphorus, organic matter and *Escherichia coli* (*E. coli*) in wastewater utilizing microalgae photobioreactors illuminated by LEDS. Other studies focusing this system access the univariate analysis; the use of multivariate optimization is new to the area. Full factorial designs 2^2 with quadruplicate in the central point [3] was applied, as the screening step, which was made for three ranges of wavelengths, corresponding to the colours: red, blue and red. The studied variables were time and light intensity. Best removal was obtained with the blue color, namely, 82.70% of organic matter, 30.00% of phosphorus, 80.00% of nitrogen and 99.00% of *E. coli*. The screening results showed that time and light intensities are significant for the system for nitrogen and organic matter parameters. Considering *E. coli* and phophorus only the time was the significant variable for the system. Based on these results, it was continued the optimization process by applying the methodology of the response surface based on the central composite design (CCD) [3]. This step was performed only for the blue color. All variables in screening were studied in the response surface methodology. The best removal results were: 85.80% of organic matter, 31.00% of phosphorus, 85.30% of nitrogen and 99.09% of *E.coli.* The results of the response surface methodology have shown that the best conditions for the removal of nitrogen, organic matter and *E. coli* are 700 µE m^{-2} s^{-1} and 15 days; the phosphorus removal was better with ~16 days and 450 µE m^{-2} s^{-1}. As the phosphorus removal of these two experiments was statistically the same, it was adopted the experiment with 16 days and 700 µE m^{-2} s^{-1} as the optimum for wastewater treatment.

Acknowledgement: To Brazilian agencies CAPES, CNPq, FAPEMIG, UFOP and Fundação Gorceix

References:

[1] Chisti, Y.Trends in Biotechnology, 26, 126 (2007).
[2] Yan, C. et al. Ecological Engineering, 51, 24 (2013).
[3] Teófilo, R. F., Ferreira, M. M. C., Química Nova, 29, 338 (2006).

P-154

EVALUATION OF WATER QUALITY INDEX USING MULTIVARIATE STATISTICAL ANALYSIS

K.P. Maia[1], G.A. da Silva[2], M. Libânio[3]

[1]Programa de Pós-graduação em Engenharia Ambiental (PROAMB), Universidade Federal de Ouro Preto (UFOP), MG, Brazil
[2]Departamento de Química, Universidade Federal de Ouro Preto (UFOP), MG, Brazil
[3]Departamento de Engenharia Civil, Universidade Federal de Minas Gerais (UFMG), Brazil
gilmare@gmail.com

This study proposes a new approach, considering the use of multivariate statistical analysis, for the reducion of the number of parameters of the Water Quality Index (WQI) [1]. In Minas Gerais, Brazil, the Instituto Mineiro de Gestão das Águas (IGAM), is responsible for controlling the quality of water of the rivers of this state. The monitoring stations by IGAM was conducted in the Velhas River basin and the respectives results of monitoring for each parameter of the WQI (dissolved oxygen, fecal coliform, pH, biochemical oxygen demand (BOD), nitrate, phosphate, temperature, turbidity and total solids) and the respect WQI scores obtained was choosen to this study, in the period between 2000 and 2010. The possibility of redundance information in the samples was firstly investigated by means of application of principal component analysis (PCA) [2], whose implementation was not suitable for evaluating the data set studied, considering its magnitude. Thus, the Kohonen neural network [2] was used, providing a friendly way for this evaluation, which indicated the possibility of work with 524 results of analyses from the 1834 monitoring results of the original data. The viability of this data reduction was assessed by comparing the correlation between the original and the reduced data set. The reduction of the number of data also showed that the monitoring frequency of all stations could be reduced, compared to the currently practiced by IGAM, culminating in important cost savings with monitoring. The new data set formed by 524 results were evaluated by the OPS (Ordered Predictors Selection) [3] to reduce the number of parameters of WQI. In this study there was the possibility of creating a new model for determining the WQI. This model was formed by five parameters: thermotolerant coliform, dissolved oxygen, BOD, turbidity and phosphate, in order of importance to the model, pointed out by OPS. It was used the RC vector, where R corresponds to the regression vector and C is the correlation vector, and it was obtained a root mean square error of cross validation (RMSECV) of 0.0309 and a coefficient of determination (R^2) and a correlation of 0.98, for both. The calibration of the new model was done by PLS (Partial Least Square), predicting WQI values generated by the new model for the prediction data set and comparing the results to the original WQI values, calculated by the traditional formula. The calibration indicated that the data fitted the model, whose R^2 was equal to 0.96, and the graph of residues showed a random distribution, with values close to zero. This model was further assessed for prediction of the results of WQI by applying the results of monitoring of water quality in the year 2011 carried out by IGAM. The results of the validation depicted very close to WQI results obtained by the traditional calculation, with a R^2 value of 0.92, confirming the applicability of the new constructed model.

Acknowledgement: CAPES, CNPq, FAPEMIG, PROAMB/ UFOP, UFOP.

References:

[1] Brown, R.M., et al., Water Sewage Works, 117, 339 (1970).
[2] E.S.J. Gontijo, et al. Journal of the Brazilian Chemical Society, 25, 208 (2014).
[3] R.F. Teófilo. J. P. A. Martins. M. M. C.Ferreira. Journal of Chemometrics, 23, 32 (2009).

P-155

Enhanced Representation of the Selectivity Profile of Solvents Using Molecular Descriptors and Chernoff Faces

N. Garrido-Sáez[1], E.J. Carrasco-Correa[2], G. Ramis-Ramos[2]

[1]Department of Applied Mathematics, University of Valencia, 46100 Burjassot, Spain.
[2]Department of Analytical Chemistry, University of Valencia, 46100 Burjassot, Spain

Proper graphical description of solvent selectivity is of interest in relation to explore the full range of possibilities during mobile-phase selectivity optimization in liquid chromatography and in other experimental situations including the design of reaction media and solubility of chemicals, to decide on selecting solvents maintaining mutual miscibility while having maximal differences in their properties. Other applications are the possibility of substituting a solvent by an equivalent one with improved non-chromatographic characteristics, such as price, availability, reactivity or better conformation to the principles of green chemistry. Owing to the complexity of the intermolecular forces than determine solvent selectivity, reactivity, miscibility and solubility, a multivariate set of descriptors is required. These are usually grouped and summarized into only 3 parameters to give rise to the classical triangular plots of Snyder and Rutan [1-3]. Spider diagrams, in which more than three descriptors can be handled, provide improved representations [4]; however, spider diagrams are multidimensional graphs projected on two-dimensions. The higher value of the spider diagrams over triangular plots rely on the selection of the axis by following a judicious order in which the correlations among them are taken into account. However, the interpretation of spider diagrams in order to choose the required solvent for a given application is not straightforward.

For these reasons, in this work, we propose the use of Chernoff faces, as a true multidimensional representation of the solvent properties. We have expressed the character of solvents by the judicious selection of the face traits, in such a way that with a minimum training any observer would be able of easily recognizing the features that are encoded in a given face. For this purpose, the face traits were assigned to the normalized solvent descriptors of Snyder, Kamlet-Taft, Hansen and Abraham [4]. The faces were programmed in R and then were plotted on the triangular and spider diagrams (instead of points or other symbols to represent the solvents), thus enhancing the information provided by the diagrams. Ambiguities due to the compensation of properties making two solvents with rather different properties to appear in the same location of the diagrams were easily detected using Chernoff faces. Using faces, the character of mixtures of solvents can be also represented. Representations by reduced groups of solvents, hierarchically selected, were of help in providing the best clues to select a solvent or solvent mixture for a given application.

Acknowledgement: Project CTQ2014-52765-R.

References:

[1] L.R. Snyder, Classification of the solvent properties of common liquids. J. Chromatogr. 92: 223–230 (1974).
[2] S.C. Rutan, P.W. Carr, W.J. Cheong, J.H. Park, L.R. Snyder, Re-evaluation of the solvent triangle and comparison to solvatochromic based scales of solvent selectivity. J Chromatogr. A 463: 21–37 (1989).
[3] A.R. Johnson, M.F. Vitha, Chromatographic selectivity triangles. J Chromatogr. A 1218: 556–86 (2011).
[4] E. Lessellier E. ∑pider diagram: A universal and versatile approach for system comparison and classification. Application to solvent properties. J Chromatogr. A 1389: 49–64 (2015).

P-156

CHARACTERIZATION OF SOLID WASTES FROM THE TAILING-PROCCESSING OF SILVER AMALGAMATION RESIDUES (JALES) FROM ZACATECAS (MEXICO) BY MULTIVARIATE TECHNIQUES

R. Pardo[1], I. Gavilán[2], B. Calles[1], A. Gavilán[3], M. Vega[1]

[1]Department of Analytical Chemistry, Faculty of Sciences, University of Valladolid, Paseo de Belén 7, 47011 Valladolid (Spain). [2]Environmental Management Unit, Faculty of Chemistry, Autonomous National University of Mexico (Mexico). [3]National Institute of Ecology and Climate Change (Mexico).
e-mail rpardo@qa.uva.es

The residues (jales) from the colonial silver production by the amalgamation technique, extensively used in Zacatecas (Mexico) until 1820, were deposited in low-lying areas used for crops and livestock farming. From 1920 onwards, there have been tailing-processing activities to recover precious metals from these soils, based upon their lixiviation with calcium thiosulfate and the subsequent recovery of silver and mercury in the lixiviate by reduction with metallic copper. The processed solids contain relatively high concentrations of mercury, silver, lead and other dangerous chemical elements and are stored in mounds within the treatment plant, posing a non-negligible environmental risk to the population of the nearby towns. In addition to the total metal concentrations, the environmental risk of a mining residue depends on the mobility or availability of the toxic elements, usually evaluated by chemical fractionation procedures such as the B.C. R. protocol, that allows distributing the total available concentrations into several fractions with decreasing mobility, corresponding to elements bonded to (i) exchangeable and carbonate phases, (ii) iron/manganese oxide and hydroxide phases, (iii) organic and sulphide phases and a fourth and almost non-movilizable fraction named residual bonded to the mineral components of the residue.

In this communication, we report the results of a survey carried out to assess the chemical stability and environmental safety of the residues from the storing mound of a treatment plant located in Tacoaleche (Zacatecas valley, Mexico). The samples were collected in two cuts of the mound at different depths, closely corresponding to the activity of the last ten years. The investigated elements were Ag, As, Cd, Cr, Cu, Ni, Pb and Zn. Their total mobilizable (pseudo-total) concentrations were determined after a single extraction with concentrated HNO_3/HCl (US-EPA 3051a method) followed by ICP-OES quantification. The B.C.R protocol was applied to the samples. Fractions (i), (ii) and (iii) were determined in the corresponding extract by ICP-OES, whereas the residual fraction was determined by difference.

The resulting dataset, a tridimensional matrix **X** with dimensions (samples=10, metals=8, fractions=4) was studied by multivariate statistical techniques: Hierarchical Clustering and normal and n-way Principal Component Analysis (PCA). Two different element associations were found: Cd-Cu-Pb-Zn were associated to iron/manganese oxide and hydroxide phases whereas Ag-As-Cr-Ni were bonded to the residual fraction. The applicaction of n-way PCA techniques confirmed these associations, and allowed to separate the information due to the samples from that corresponding to the fractions through a Tucker3 [2 2 1] model. The fractionation of the investigated elements in the samples showed a constant pattern irrespectively of their spatial location in the mound. The pseudo-total contents of the toxic elements allows to classify the samples as hazardous.

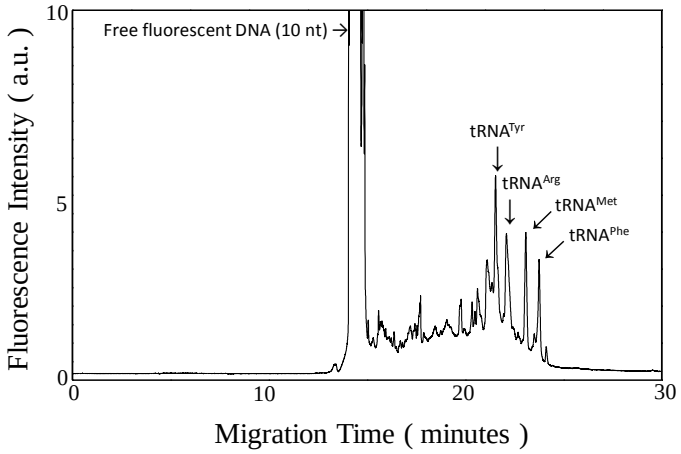
P-157

MULTIPLEXED DETERMINATION OF CANCER-ASSOCIATED TRANSFER RNA BY CAPILLARY ELECTROPHORESIS

Han-Yu Chen, Wei-Yu Lin, Po-Ling Chang*

Department of Chemistry, Tunghai University No.1727, Sec.4, Taiwan Boulevard, Taichung 40704, Taiwan
poling@thu.edu.tw

Transfer RNA (tRNA) is a well-known "house-keeping" molecule that carries amino acids to a synthesizing messenger RNA (mRNA) during translation based on codon recognition between tRNA and mRNA. Recently, several reports have indicated that some tRNA molecules may alter their expression levels in specific cancers [1]. In other words, tRNA may serve as a biomarker for cancer diagnosis in the early stage. To date, only a few available methods could be used for the determination of mature tRNA, for example, northern blot, microarray, and reverse transcription-quantitative PCR. In this work, we propose a method for multiplexed tRNA quantitation by a combination of splinted ligation and capillary electrophoresis with laser-induced fluorescence (CE-LIF) [2, 3]. We designed several novel probes that specifically recognize cancer-related tRNA as well as fluorescently labelled DNA reporter. By using well-designed probe in varies lengths, four tRNAs ($tRNA^{Tyr}$, $tRNA^{Arg}$, $tRNA^{Met}$, and $tRNA^{Phe}$) could be separated using 7% polyvinylpyrrolidone (M_r 13,000,000 Da), which dissolved in 10 mM HEPES buffer (pH 8.0). Electropherograms of different cancer cells showed varying expression levels of the four tRNAs, which correspond to those reported previously in the literature ($tRNA^{Tyr}$ for ovarian cancer; $tRNA^{Arg}$ for urothelial cancer; $tRNA^{Met}$ and $tRNA^{Phe}$ for breast cancer). Therefore, the multiplex and cost-effective characteristics of CE-LIF may allow it to become a powerful tool for high-throughput tRNA profiling for a rapid diagnosis in the early stage of cancers.



Acknowledgement: This work is supported by the Ministry of Science and Technology (Taiwan).

References:

[1] S. S. Grewal, "Why should cancer biologists care about tRNAs? tRNA synthesis, mRNA translation and the control of growth," *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, vol. 1849, 898-907, 2015.

[2] T.-H. Yang, D.-L. Ou, C. Hsu, S.-H. Huang, and P.-L. Chang, "Comparative microRNA detection from precursor-microRNA-transfected hepatocellular carcinoma cells by capillary electrophoresis with dual-color laser-induced fluorescence," *Electrophoresis*, vol. 33, 2769-2776, 2012.

[3] R.-M. Jiang, Y.-S. Chang, S.-J. Chen, J.-H. Chen, H.-C. Chen, and P.-L. Chang, "Multiplexed microRNA detection by capillary electrophoresis with laser-induced fluorescence," *Journal of Chromatography A*, vol. 1218, 2604-2610, 2011.

P-158

TECHNOLOGICAL DEVELOPMENT AND EVALUATION OF A PROTOTYPE FOR THE ANALYSIS OF WATER POLLUTION BY FLUORESCENCE EXCITATION-EMISSION MATRIX.

O.L. Lombana [1], J.C. Cely [1], R. Jimenez [1]

[1]Department of Chemical and Environmental Engineering, Universidad Nacional de Colombia - Bogota, Avenue 30 #. 45-03, Building 453 Bogotá, DC 111321, Office 316, Bogota , Colombia.
ollombanac@unal.edu.co

The contamination of surface and underground waters by oil is a major environmental problem, particularly for oil producing countries like Colombia. The statistics of accidental spills of oil and its product in Colombia are alarming (~10,000 barrels during 2008-2009). The oil boom in Colombia requires the application of operational and emergency procedures devised to minimize and mitigate the pollution of water bodies.

Our project aims at developing a real-time, *in situ*, low-cost analytical instrument and technique for the detection and quantitative analysis of water pollution by hydrocarbons, particularly crude oil and refined fuels [1].

We are developing a Xe lamp based instrument capable of producing fluorescence spectra suitable for pollutant identification and analysis through excitation-emission matrix spectroscopy (EEMS) [2]. The Xe lamp system uses band pass filters (320, 365, 377, 400, 436, and 568 nm) mounted on rotating wheel. The light source has a lens system for source light imaging and fluorescence light collection along with a fluorescence cell. Excitation and emission spectra are measured with a fiber coupled mini-spectrometer (AvaSpec-3648, Avantes). The emission spectra are normalized against the equivalent excitation intensity.

EEMS spectra for pollutant fingerprinting have been traditionally obtained using quasi-monochromatic excitation sources, such as monochromator filtered high pressure Xe lamps in bench-top, bulky instruments [3]. Our challenge is overcome the loss of analytical specificity and precision due to the larger spectral width and lower power of our light sources.

We have optimized the optical system and obtained EEMS fingerprints of fuels blends to detect fuels adulteration which can help control smuggling Colombian borders. We were able to detect these fuel blends from 100% to 0% Colombian gasoline, using Ecuadorian and Venezuelan gasoline.

References:

[1] Baker, A., Inverarity, R., Charlton, M., Richmond, S. Detecting river pollution using fluorescence spectrophotometry: case studies from the Ouseburn, NE England. Environ. Pollut. 124: 57-70, 2003
[2] Keshav, K. and Ashok K. M. Quantification of ethanol in ethanol-petrol and biodiesel in biodiesel-diesel blends using fluorescence spectroscopy and multivariate methods. J Fluoresc 22: 339–347, 2012
[3] Brown, C.E. and Fingas, M.F. Review of the development of laser fluorosensors for oil spill application. Marine Pollut. Bulletin 47: 477-484, 2003

P-160

IMPLEMENTATION OF IMAGE PROCESSING AS SPATIAL CONSTRAINTS IN MCR-ALS OF HYPERSPECTRAL IMAGES

S. Hugelier,[1] O. Devos,[1] C. Ruckebusch[1] and A. De Juan[2]

[1] *Université de Lille, LASIR CNRS UMR 8516, F-59000 Lille, France;*
[2] *Universitat de Barcelona, 08-007 Barcelona*

Hyperspectral imaging (HSI) is a technique to obtain spatial and spectral information associated with the distribution of the different compounds in a chemical or biological sample. Amongst the multivariate image analysis tools utilized to decompose the raw data into a bilinear chemically meaningful model, Multivariate Curve Resolution – Alternating Least Squares (MCR-ALS)[1] can be applied to obtain the distribution maps and pure spectra of the components of the imaged sample.

In the current MCR-ALS algorithm, the requirement to work with a two-way matrix involves unfolding the raw HSI data into a single pixel direction. Through this data manipulation, the information regarding pixel neighborhood is lost and the concentration profiles obtained are stretched arrays of pixel concentration values. This configuration of the concentration profiles allows incorporating general properties, such as non-negativity or presence/absence of constituents in pixels, but prevents the use of spatial information linked to the 2D variation of distribution patterns of compounds under the form of constraints.

Recently, we have proposed and discussed the benefit of adding, at each least-squares step, an additional refolding/unfolding of the pure concentration profiles into the related distribution maps for the components.[2,3] Once this is done, implementation of spatial constraints based on the use of 2D signal processing approaches directly on the distribution map is possible and improves the image modeling approach.

Dealing with simulated and real data, we will discuss the results obtained when implementing spatial constraints on distribution maps to allow image filtering and image fitting. Image filtering can be performed in different ways, including working in the frequency domain or by using 2D-splines to provide a continuous image representation. It results in noise removal, image sharpening, deblurring and smoothness/roughness properties that can be translated as a constraint. Image fitting with hard-modeling constraints can also be envisaged to allow further image processing such as sparse deconvolution.

References:

[1] De Juan, A.; Piqueras, S.; Maeder, M.; Hancewicz, T.; Duponchel, L.; Tauler, R. Chemometric tools for image analysis. Infrared and Raman Spectroscopic Imaging (Editors: R. Salzer, H.W. Siesler), 2nd. Edition. Wiley-VCH, (2014) 57-110.
[2] S. Hugelier, O. Devos, C. Ruckebusch, Journal of Chemometrics 29 (2015) 557-561.
[3] S. Hugelier, O. Devos, C. Ruckebusch, Journal of Chemometrics 29 (2015) 448-456.

P-161

MODEL TRANSFORM FOR RAMAN SPECTROSCOPY IN BIOLOGICAL APPLICATIONS

S.X. Guo [1], R. Heike [1], T.W. Bocklitz [1,2], J. Popp [1,2,3]

[1] *Institute of Physical Chemistry and Abbe Center of Photonics, Friedrich-Schiller-University Jena, Helmholtzweg 4, Jena, Germany.*
[2] *Leibniz Institute of Photonic Technology, Albert-Einstein-Straße 9, Jena, Germany.*
[3] *InfectoGnostics Research Campus Jena, Centre of Applied Research, Philosophenweg 7, Jena, Germany*
shuxia.guo@uni-jena.de

Efficient data treatment is one of the most important issues when applying Raman spectroscopy for biological diagnostics. This is achieved by chemometrics and a diagnostic result is obtained by predicting a spectral dataset or Raman spectrum with statistical models, which are trained with known samples beforehand. However, the accuracy of these models may dramatically decrease if the new dataset is measured under different condition compared with the training datasets. This can be solved by model transform, aiming to obtain a higher overall prediction accuracy for datasets measured under different conditions. Methods like standard calibration and multiple condition training are commonly used. Nevertheless, they are unable to provide satisfying results [1, 2]. Therefore, we attempted to further improve model transferability in two aspects:

(1) wavenumber alignment, achieved by genetic algorithm (GA) [3] after the standard calibration;
(2) model update, realized by Tikhonov regularization (TR).

Particularly, two TR methods were studied for model update, termed as TR_1 and TR_2 [4, 5]. The parameters of the both TR methods were optimized by genetic algorithm. We based our investigation on Raman spectra of three spore species (*B. mycoides*, *B. subtilis*, and *B. thuringiensis*) measured on four different devices (SPE raman.ID 0.5-Bio Particle Explorer). The model transferability was evaluated by classification accuracy of a three-class task by partial least square regression (PLSR). Each spectrometer was predicted by models trained by data measured with one or more other devices. As a comparison, we performed different combinations of model transform methods including standard calibration, GA based wavenumber adjustment, TR_1 and TR_2. An overview of the results is shown below, where label 'None' indicate that no model transform was applied. The abbreviation 'STD' represents the standard calibration. Overall, combinations of the standard calibration, GA wavenumber adjustment and both TR methods were tested and lead to an improvement of the model transferability. Furthermore, the model transferability is enhanced if mechanisms are combined.

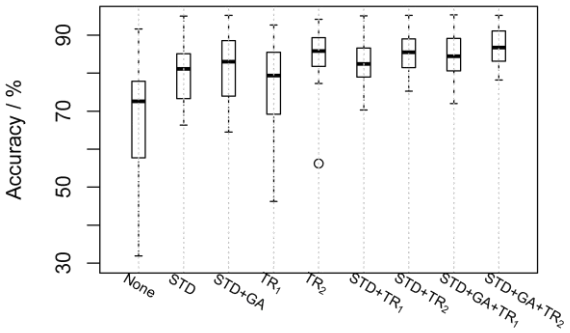


Fig 1. Accuracy of a classification model depending on different applied model transfer methods and combinations.

Acknowledgement: The financial support from China Scholarship Council (CSC) is greatly appreciated.

References:

[1] R.L. McCreery, John Wiley & Sons, 157(2000).
[2] T. Dörfer, T.W. Bocklitz, et al., Zeitschrift für Physikalische Chemie, 225(6-7), 753-764(2011).
[3] Z.M. Zhang, S. Chen, Y.Z. Liang, Talanta, 83(4), 1108-1117(2011).
[4] P. Shahbazikhah, J.H. Kalivas, Chemometrics and Intelligent Laboratory Systems, 120, 142-153(2013).
[5] J.H. Kalivas, G.G. Siano, et al., Applied spectroscopy, 63(7), 800-809(2009).

P-162

Detection of adulterants in roasted and ground coffee by NIR hyperspectral imaging and multivariate curve resolution

Debora A.P. Forchetti, Ronei J. Poppi

Institute of Chemistry, University of Campinas, P.O. BOX 6154, 13081-970, Campinas, Brazil.
debora.forchetti@iqm.unicamp.br

Coffee is an economic prominent product in the world due to its high consumption and acceptance. Seeking an illicit profit on the coffee market, several compounds have been used for its adulteration. Coffee quality control reported in the literature is based on microscopy techniques, as well as, chromatography and infrared spectroscopy [1]. The use of NIR hyperspectral imaging has gained importance in the analysis of food, helping in the combat of the adulteration [2]. Faced with the routine analysis of food quality control, NIR hyperspectral imaging provides results of the distribution of components in the surface of the sample in a short time, with little or no sample preparation, avoiding external contamination risks. In this work, the NIR hyperspectral imaging in combination with multivariate curve resolution method (MCR) was employed to detection of coffee adulterants.

The analysis of samples surfaces was conducted using a Spotlight 400N FT-NIR Imaging System from Perkin-Elmer. Tablets were made with coffee containing different concentrations of contaminants, such as: coffee husks, roasted and powdered corn grains, ground and twigs, which were placed on supports for microscopy and subsequently accommodated on the positioner. By using a joystick, the focus and the area to be analyzed was adjusted. The conditions for the analysis were 32 scans analyzing individual pixels of 25 μm^2 and 4 mm^2 area in the range of 4000-7800 cm^{-1}. The data obtained from the pure samples and mixtures were converted to log1/R and it was applied as a pre-treatment the MSC (multiplicative signal correction) with the aim of reducing the light scattering. The data set generated were converted in matrices and processed by MCR algorithm. From the generated images by the MCR, it was possible to distinguish each of the four contaminants and confirmation was accomplished comparing the pure spectra of each constituent with the spectra recovered. The MCR method was suitable for the detection of contaminants in the range of 5 to 30%.

Acknowledgement: Financial support CNPq and the members of LAQQA - Chemometrics Laboratory for Analytical Chemistry. Department of Analytical Chemistry – Chemical Institute – UNICAMP

References:

[1] E. D. Pauli, F. Barbiere, P. S. Garcia, T. B. Madeira, V. R. Acquaro junior, I. S. Scarmínio. Food Research International, 61, 112-119 (2014).
[2] X. Fu, M. S. Kim, K. Chao, J. Quin, J. Lim, H. Lee, A. Garrido-Varo, D. Pérez-Marín, Y. Ying. Journal of Food Engineering, 124, 97-104 (2014).

P-163

SPECTRAL AND CHEMOMETRICS ANALYSIS AS A COUPLED METHODS TO DETERMINATION OF CONTAMINATION DEGREE IN MATERIALS OBTAINED BY THERMAL CONVERSION OF BIOMASS

M. Sajdak

Institute for Chemical Processing of Coal
1 Zamkowa St. 41-803 Zabrze, Poland,
msajdak@ichpw.p

Biomass is the oldest and most widely used modern renewable energy source which are is the third largest in the world, natural source of energy. The growing demands for biomass and its processing products (pellets, briquettes, etc.) carry the risk of artificially raising the calorific value of the product known as a natural and renewable energy carrier. Substance such as fossil fuels, polymers, and wood industry which their waste not classified as biomass because of its composition containing substantial amounts of resins (polyester, alkyd, polyurethane) and adhesives (urea, polyvinyl acetate) can be added to biomass. Each chemical analysis excludes these contaminants would be time consuming and expensive, however there is the surest way to confirm the origin of biomass, which is 14C method. These methods are suitable only for the pure biomass analysis and solid recovered fuels, which are biodegradable in that they do not contain non-biodegradable substances to an extent deviating from the known natural properties of the biomass. Research conducted by the Institute for Chemical Processing of Coal showed that these methods do not provide sufficient proof that 100% of the material originates from biomass.

To meet the demands, it is being proposed to use of combining several methods of data exploration and chemometrics analysis to simple and fast verification of biomass in terms of its origin. It also identifies the relevant variables such as biomass parameters as a simple way to allow for confirmation or exclusion of contamination. For this purpose, the method used in classification and regression trees combined with principal component analysis and regression analysis to estimate quantitatively investigated capable of specifying the same depending on the degree of contamination. classification and regression trees method allowed the fast initial distribution of biomass of the study group for each subgroup for which the principal component analysis was performed to determine the origin of tested biomass. In the case a biomass of the different composition to natural samples is detected, the regression analysis allows the estimation of the degree of contamination of the test biomass.

As a raw data in our research was applied data from Fourier transform infrared spectroscopy and X-ray fluorescence analysis. In our study we proved that applied the chemometric methods was possible to show the differences between the pyrolysis solid products depending on the pyrolysis of polypropylene as an additive. Application of principal component was helped to identify these absorption bands that allow for a quantitative estimate of the additive polymer material and types for example in the case of styrene-butadiene rubber and polyethylene terephthalate.

Thanks to the applied methods we can prepared algorithms to predicted amount of polypropylene that was added to the biomass in the thermal conversion process and we showed that the polypropylene amount in char could only be predicted over narrow temperature ranges, for example 400-500°C, 500-600°C, and 600-700°C.

Acknowledgement: This research was conducted in the Institute for Chemical Processing of Coal (Zabrze, Poland) and funded by the Ministry of Science and Higher Education under the Iuventus Plus Program for 2015-2016 (project No. IP2014 041073)

P-164

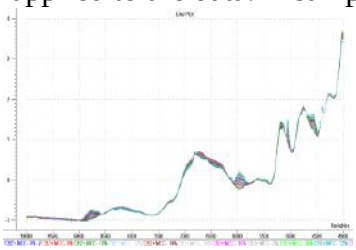
PXRD and NIR spectroscopic quantitative analysis of carbamazepine precipitated onto microcrystalline cellulose

N. Kamali [1], A. Erxleben [1], P. McArdle [1], K. Hodnett [2]

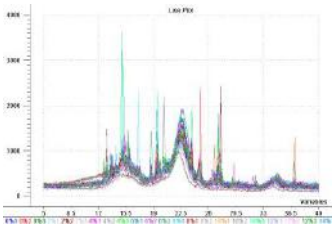
[1] *School of Chemistry, Nationa1University of Ireland, Galway, Ireland.*
[2] *Faculty of Science and Engineering, University of Limerick, Ireland*

naghmehkamali@gmail.com , n.kamali1@nuigalway.ie

Polymorphism investigations are particularly important in drug and product development in the pharmaceutical industry since the properties of a formulated product such as bioavailability and stability are often directly related with the physicochemical properties of the existing polymorphs in the formulation.
Carbamazapine (CBZ) is known as a leading antiepileptic drug which has been used for over 20 years. CBZ with four polymorphic forms and a hydrate have been reported.[1]
In this study we focuse on using PXRD and NIR spectroscopy for the quantitative analysis of CBZ form I and III precipitated onto microcrystalline cellulose (MCC) using rotary evaporation. Multivariate calibration models were developed to quantify CBZ/MCC samples containing 0-30% CBZ. Standard normal variate, SNV, pre-treatment as well as partial least squares, PLS1, regession analysis was applied to the data. A sample participated onto was found to be largely amorphous.



NIR Spectra-CBZ FIII



PXRD Pattern- CBZ FIII

	NIR Spectroscopy		PXRD	
Sample	Predicted%	Deviation	Predicted %	Deviation
CBZ-p4%	5.4	3.1	4.5	0.5
CBZ-p8%	9.7	4.9	7.7	0.5
CBZ-p12%	12.3	3.89	15.1	1.2
CBZ-p16%	15.4	4.7	17.1	0.9
CBZ-p20%	20.8	6.4	20.4	0.8

prediction samples- CBZ FIII

Keywords: Spectroscopic quantitative, Carbamazapine, Multivariate model.
Acknowledgement: This work was supported by Science Foundation Ireland under Grant No. [07/SRC/B1158] as part of the Synthesis and Solid State Pharmaceutical Centre (SSPC).

References:

[1] F.Tiana, J.Zeitler, C. J.Strachan, D.Saville, K.Gordonb, T.Rades. J. Pharm. Biomed. Anal. 40 (2006) 271-280.

P-165

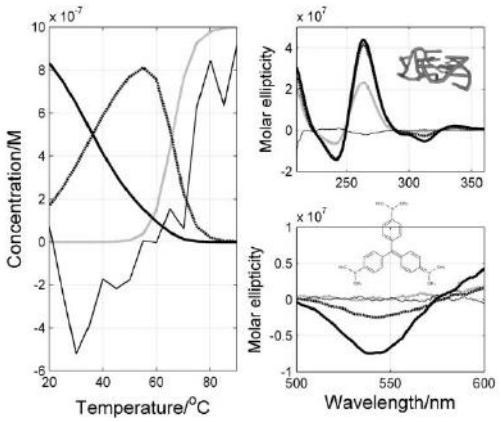
STUDY OF THE CONFORMATIONAL EQUILIBRIA OF G-QUADRUPLEX-FORMING DNA SEQUENCES BY MEANS OF SPECTROSCOPIC AND CHEMOMETRIC TECHNIQUES

A. Ferrer [1], S. Benabou [1], R. Gargallo [1]

[1]Department of Analytical Chemistry, University of Barcelona, Marti i Franques 1-11, Barcelona, Spain
raimon_gargallo@ub.edu

The best-known nucleic acid structure is the B-DNA proposed by Watson and Crick in 1953. However, nucleic acids may form other complex structures, such as triplexes, G-quadruplex or i-motif, among others. G-quadruplex structures, which are formed by guanine-rich sequences in appropriate conditions of pH, temperature and ionic strength, have been found *in vitro* at the end of telomeres and near the promoter regions of several oncogenes. Recently, the presence of these structures *in vivo* has been proposed [1]. Apart of its potential biological interest in the treatment of cancer, applications in nanotechnology have been proposed recently [2].

Guanine-rich sequences may fold into intramolecular (monomer) o intermolecular (multimer) G-quadruplex structures. In this work, the formation of monomeric and multimeric G-quadruplex structures is studied by means of spectroscopic and separation techniques. In addition, the interaction of the G-quadruplex structures with the crystal violet dye is investigated. With these goals in mind, multivariate data analysis methods, both based on soft- [3], hard- [4] and hybrid-modelling [5] have been used to extract information from the experimental data.



As example, this figure shows the results of the application of an hybrid-modelling approach [5] to analyze circular dichroism data recorded along the thermal denaturation of a dye:DNA mixture.

The contribution of base line drift has been modelled successfully, and reliable thermo-dynamic data may be obtained. On the other hand, UV and visible regions of pure spectra provide useful information about the G-quadruplex structure and dye, respectively.

Acknowledgement: We acknowledge funding from the Spanish government (BFU2014-52864-R, CTQ2012-38616-C02-02 and CTQ2014-52588-R).

References:

[1] G. Biffi, M. Di Antonio, D. Tannahill, S. Balasubramanian. Nat. Chem. 6, 75 (2014)
[2] J. Ren, T. Wang, E. Wang, J. Wang. Analyst, 140, 2556 (2015)
[3] R. Tauler. Chemom. Intel. Lab. Syst. 30, 133 (1995)
[4] R. Dyson, S. Kaderli, G.A. Lawrence, M. Maeder, A.D. Zuberbühler, Anal. Chim. Acta. 353, 381 (1997)
[5] R. Gargallo, Anal. Biochem. 466, 4 (2014)

P-166

LOW-CONTENT QUANTIFICATION IN THE SOLID-STATE BY RAMAN SPECTROSCOPY: AN EVALUATION OF METHODS.

B. Li, Y. Casamayou-Boucau, A. Calvet, A.G. Ryder

Nanoscale Biophotonics Laboratory, School of Chemistry, National University of Ireland, Galway, Galway, Ireland.
boyan.li@nuigalway.ie

Low-Content Quantification (LCQ) in the solid-state is feasible using Raman spectroscopy/mapping if the material has a sufficicent degree of heterogenity on the micrometre scale and the spectra are collected from small sample volumes [1]. Here we conducted a detailed comparison between five different chemometric methodologies for LCQ analysis (piracetam-proline mixtures) in order to determine the optimum methods for practical use. The methods were: partial least squares (PLS) regression [2], net analyte signal [3] coupled with classical least squares [4] (NAS-CLS), multivariate curve resolution (MCR) [5], principal component analysis [6] with CLS (PCA-CLS), and the ratio of characteristic bands at selective signal wavenumbers combined with shape-preserving piecewise cubic polynomial interpolation curve fitting [7] (BAND RATIO-PCHIP). The results showed that once appropriate data pretreatment had been applied and concentration-segmented strategy used for calibration modelling that PLS gave the best accuracy and lowest relative error of prediction (REP%=2.6%) for the 0.05−1.0% w/w analyte range [1]. However, the PLS method required a large sample set for building the calibration models and this is not always technically feasible particulary for complex materials.

For analytre content >1.0%, all methods except for PCA-CLS were able to accurately quantitate piracetam in powder mixtures, giving very low REP% values of: 0.68 (PLS), 0.74(MCR), 0.86(NAS-CLS), and 0.81 for BAND RATIO-PCHIP, respectively. PCA was found to not be suitable for quantitative analysis. For LCQ, the BAND RATIO-PCHIP method, although simple, and easy to implement, was inaccurate for LCQ (REP% = 36%) because the use of univariate selective signal was not robust. This unfortunatley was also the case for NAS-CLS and MCR, but these methods could be implemented without the need for large calibration sample sets as required by PLS. We are continuing with modifying both MCR and NAS based approaches to try and improve accuracy in order to provide a practical alternative to PLS for LCQ.

Acknowledgement: Research undertaken as part of the Synthesis and Solid State Pharmaceutical Centre was funded by Science Foundation Ireland and industry partners, and by Enterprise Ireland (Grant No: TC-2012-5106). Kaiser Optical Systems, Inc. (Ann Arbor, MI) and Mr. Harry Owen are thanked for the loan of the Raman instrumentation.

References:

[1] B.Y. Li, A. Calvet, Y. Casamayou-Boucau, C. Morris, A.G. Ryder, *Anal. Chem.*, 87, 3419 (2015).
[2] A.B. Smith, C.D. Brown, J. Good Results, Vol, Initial page (Year). S. Wold, M. Sjostrom, L. Eriksson, *Chemom. Intell. Lab. Syst.*, 58, 109 (2001).
[3] A. Lorber, K. Faber, B.R. Kowalski, *Anal. Chem.*, 69, 1620 (1997).
[4] H. Martens, T. Næs, *Multivariate Calibration*, Wiley: New York (1989).
[5] A. de Juan, R. Tauler, *Crit. Rev. Anal. Chem.*, 36, 163 (2006).
[6] I.T. Jolliffe, *Principal Component Analysis*, 2nd edn., Springer: New York (2002).
[7] S. Pruess, *IMA J. Numer. Anal.*, 13, 493 (1993).

P-167

MULTI-MODAL HYPERSPECTRAL DATA FUSION FOR CHARACTERISATION OF BIOMATERIALS

R.M. Dorrepaal[1], A.A. Gowen[1]

[1]Biosystems and Food Engineering, University College Dublin, Belfeld, Dublin 4, Ireland.
ronan.dorrepaal@ucdconnect.ie

In the fields of remote sensing and pharmaceutical quality control, hyperspectral imaging techniques have been used to great effect as such techniques can provide vast amounts of both spatial and spectral information [1,2]. However hyperspectral imaging is limited by its underlying analytical technique. For example, often infrared hyperspectral imaging can provide information which is not available to raman hyperspectral imaging, such as a changes to a dipole moment. Conversely, raman imaging can provide information on changes in polarizability within a molecule, which is not observed with infrared spectroscopy.

For this reason, there is an increasing interest in the fusion of complementary hyperspectral imaging techniques [3,4]. This is not simply the corroboration of more than one technique for characterisation, which can of course be useful in its own right [5]. Rather for fusion techniques, analysis is conducted on the combination of resultant data to arrive at a deeper understanding than the sum of its parts.

The objective of this study is to investigate multi-modal data fusion of complimentary hyperspectral imaging techniques for improved characterisation of biomaterials. Specifically, the research was conducted on magnesium oxychloride cement (for consideration in bone regeneration) and its derivatives. The study also seeks to resolve issues arising from the use of hyperspectral images with different spatial and spectral resolution. Near-infrared, mid-infrared and raman spectra and hyperspectral images were acquired and fusion techniques performed to analyse both the unique and correlated information provided by each hyperspectral technique and evaluate the synergy of combining these techniques using chemometric modelling.

Acknowledgement: The authors acknowledge funding from the EU FP7 under the European Research Council Starting Grant programme.

References:

[1] C. Rodarmel, J. Shan, Surveying and Land Information Systems, Vol. 62(2), 115(2002).
[2] F. Franch-Lage, J.M. Amigo, E. Skibsted, S. Maspoch, J. Coello, International Journal of Pharmaceutics, Vol. 411(1-2), 27(2011).
[3] F.C. Clarke, M.J. Jamieson, D.A. Clark, S.V. Hammond, R.D. Jee, A,C. Moffat, Analytical Chemistry, Vol. 73(10), 2213(2001).
[4] G.A. Licciardi, M.M. Khan, J. Chanussot, A. Montanvert, L. Condat, C. Jutten, EURASIP Journal on Advances in Signal Processing, Vol. 207, 1(2012).
[5] P.S. Wray, W.E. Sinclair, J.W. Jones, G.S. Clarke, D. Both, International Journal of Pharmaceutics, Vol. 493, 198(2015).

P-168

POWDER HOMOGENEITY CHARACTERIZATION USING HYPERSPECTRAL IMAGING AND CHEMOMETRICS

J.A. Fernández Pierna [1], D. Vincke [1], Z. Ouizem [2], P. Dardenne [1], V. Baeten [1]

[1]Valorisation of Agricultural Products Department, Food and Feed Quality Unit, Walloon Agricultural Research Centre, Henseval Building, 24 Chaussée de Namur, 5030 Gembloux, Belgium.
[2]Gembloux Agro-Bio Tech, University of Liège, 2 Passage des déportés, 5030 Gembloux, Belgium
j.fernandez@cra.wallonie.be

In many industrial fields such as pharmaceuticals, process engineering and food processing, handling and mixing of powders is an important unit step. During this step, many problems can happen, especially when the objective is to choose the right working parameters to perform a correct homogenization.

In the case of the food and feed industries, the characterization of the homogeneity of mixtures of powders by the use of a rapid and sensitive method is a priority. In this direction, several studies have been published, which are mainly linked to the procedure to obtain mixtures of powders and the problems of non-uniformity that can be faced at industrial level.

In this work, the aim is to propose a fast and easy methodology to define and characterize the homogeneity of feed mixture powders. This is an important requirement of the European Reference Laboratory for Animal Proteins (EURL-AP) when preparing inter-laboratory studies where homogeneity of samples should be assured. Here to characterize such homogeneity of samples, the proposed procedure includes the use of a hyperspectral NIR imaging system combined with an image treatment PCA-based protocol [1] as an alternative to well-known statistical methods, such as auto-correlation functions or variances.

Near infrared hyperspectral imaging combined with chemometrics has the advantages of its speed, robustness, and especially allows analyzing the products without major operations preparations. In one analytical step, it is possible to get spatial and spectral information of the mixture.

This study has allowed us to characterize the homogeneity of binary mixtures of feed ingredients as well as check the influence of homogenization duration. The results obtained indicate that degree of homogeneity of a feed powder mixture is influenced by several factors namely the homogenization time, the physical and chemical characteristics of the particles and the proportions of components. Moreover, the results have been modified in order to take into account the recommendations done by the working group for blend uniformity testing [2]. The results have been compared to other techniques to determine homogeneity as PCR, classical microscopy and NIR microscopy.

Acknowledgement: This work was performed within the framework of the European Reference Laboratory for Animal Proteins (EURL-AP).

References:

[1] H. Berthiaux, V. Mosorov, , L. Tomczak, C. Gatumel, J.F. Demeyre, Chemical Engineering and Processing, 45, 397 (2005).
[2] Trudy, J. of GXP Compliance, 12 (1), 46 (2007).

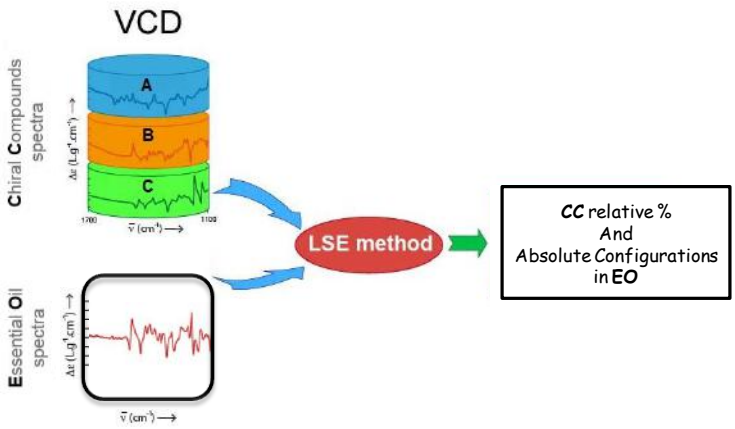
P-169

ESSENTIALS OIL VCD SPECTRA MODELISATION USING TERPENES VCD SPECTRAL DATABASE AND LSE.

M. E. A. Said[1,3], P. Vanloot[1], I. Bombarda[1], J. V. Naubron[2], N. Dupuy[1], C. Roussel[3].

[1]Aix-Marseille Université, EA4672 LISA Equipe METICA, Case 451, Av. Escadrille Normandie Niémen, 13397 Marseille Cedex 20, France
[2]Aix-Marseille Université, Spectropole, service 511, F-13397 Marseille, France
[3]Aix-Marseille Université, Ecole Centrale, CNRS, ISM2 UMR 7313, Marseille, France
email : saidmedamin@gmail.com

A method based on the VCD spectra of pure enantiomers database and VCD spectra of EOs was used to simultaneously assign the relative percentages of the major chiral compounds and their prevailing enantiomeric form in crude essential oils (EOs).[1] For this purpose, firstly, the EOs were analyzed by VCD and secondley they were modelized as a linear weighted combination of the individual spectra of pure enantiomers. The spectral modelizations were performed using a mathematical model (least square estimation) that gives us a weighting of each contributing compounds. The value of each weighting gives the relative percentage of the associate chiral compound in the EOs while the attached sign addressed the correctness of the enantiomer employed form used to build the model. In order to compare our results with the results of GC (gas chromatography), the chemical compositions of EOs of *Artemisia herba-alba*, grown in Algeria and Morocco, were determined by gas chromatography coupled to mass spectrometry (GC–MS).



References:

[1] M. E. A. Said, P. Vanloot, I. Bombarda, J. V. Naubron, E. M. Dahmane, A. Aamouche, M. Jean, N. Vanthuyne, N. Dupuy, et C. Roussel, *Anal. Chim. Acta*, 903, 121‑130, (2016).

P-170

FORENSIC SINGLE FIBER DISCRIMINATION BY COMBINING EXCITATION-EMISSION FLUORESCENCE MICROSCOPY WITH LINEAR DISCRIMINANT ANALYSIS

A. Muñoz de la Peña[1], A. D. Campiglia[2,3], N. Mujumdar[2], E. C. Heider[2], H. C. Goicoechea[4], D. Muñoz de la Peña[5]

[1]Department of Analytical Chemistry and IACYS, University of Extremadura, Badajoz, Avda. de Elvas, s/n, 06006, Spain, [2]Department of Chemistry, University of Central Florida, 4111 Libra Drive, P.O.Box 25000, Orlando, Florida 32816, United States, [3]National Center for Forensic Science, University of Central Florida, 12354 Research Parkway, Suite 225, Orlando, Florida 32826, United States,[4]Laboratorio de Desarrollo Analítico y Quimiometría,Universidad Nacional de Litoral, Santa Fe, S3000ZAA, Argentina,[5]Department of Automation and Systems Engineering, University of Seville, Seville, 41092, Spain, arsenio@unex.es

The potential of total excitation-emission fluorescence microscopy combined with multi-way linear discriminat analysis (LDA) was investigated for the forensic analysis of visually indistinguishable textile fibers. Forensic fiber comparison involves one questioned fiber (i.e the fiber found in the crime scene) versus one or more fibers (called ”known fibers”) collected from a textile of the same color belonging to a suspect. Our approach takes fluorescence microscopy to a higher level of selectivity with the collection of excitation emission matrices (EEMs) [1, 2]. For that, four pairs of visually indistinguishable fibers from four different textiles were selected. The fibers consisted of nylon 361 dyed with acid yellow 17 and acid yellow 23, acetate satin 105B dyed with disperse blue 3 and disperse blue 14, polyester 777 dyed with disperse red 1 and disperse red 19, and acrylic 864 dyed with basic green 1 and basic green 4, and were investigated using a non-destructive approach, which is of paramount importance to preserve the physical integrity of the fibers for further court examination. Excitation emission matrices were recorded with the aid of an inverted microscope coupled via a bifurcated fiber-optic probe to a commercial spectrofluorimeter [3]. The full information content of excitation-emission matrices was processed with the aid of PARAFAC supervised by linear discriminant analysis (LDA), and discriminant unfolded partial least squares (DU-PLS). The ability of the latter algorithm to classify the four pairs of fibers, and distinguish the questioned fiber and the known fiber to exclude the possibility that both could have originated from a common source, demonstrates the advantage of using the multidimensionality of fluorescence data formats for the non-destructive analysis of forensic fiber evidence.

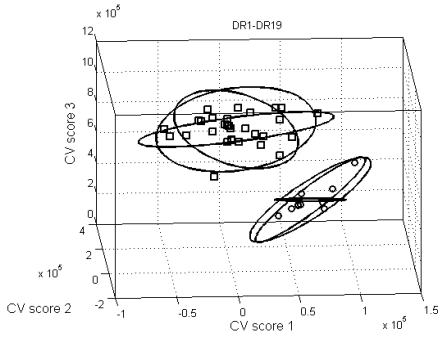


Figure 1.- LDA canonical variate (CV) scores (3 components model) for 40 samples of disperse red 1 (DR 1) (1 fiber, 10 replicates; circles) and disperse red 19 (DR19) (3 fibers, 10 replicates each; squares) polyester 777 fibers.

Acknowledgement: The authors are grateful to US National Institute of Justice (Grant # 2011-DN-BX-K553), UNL and CONICET, and Ministerio de Economia y Competitividad of Spain (Projects CTQ2014-52309-P and DPI2013-48243-C2-2-R) and Gobierno de Extremadura (GR15090-Research Group FQM003), both co-financed by European FEDER funds, for financially supporting this work. AMP and DMP acknowledge Awards (PRX14/00342 and PRX15/00138) under the Program Salvador de Madariaga-MECyD (Spain).

References:

[1] K. Appalaneni, E.C. Heider, A.F.T. Moore, A.D. Campiglia, Anal. Chem., 86, 6774 (2014).
[2] N.Mujumdar, E.C. Heider, A.C. Campiglia, Applied Spectroscopy, 69, 1390 (2015).
[3] N. Mujumdar, E.C. Heider, H.C. Goicoechea, D. Muñoz de la Peña, A. Muñoz de la Peña, A.D. Campiglia, 71st SWRM/67th SERM Meeting of the American Chemical Society, Memphis, Tennesse, United States, 2015.

P-171

CHEMOMETRICAL APPROACHES TO DETECT INFECTIONS AND FRAUDULENT ADDITION OF EXTRANEOUS SPECIES IN DRIED PORCINI MUSHROOMS

M. Casale[1], L. Bagnasco[1], M. Zotti[2], N. Sitta[3], P. Oliveri[1]

[1]Department of Pharmacy, University of Genoa, Via Brigata Salerno 13, Genoa, Italy
[2]Department of Earth, Environment and Life Sciences, University of Genoa, Corso Europa, 26, Genoa, Italy
[3]Professional Consulting Mycologist, Loc. Farné, 32, Lizzano in Belvedere, Italy
monica@difar.unige.it

Boletus edulis and allied species (known as "porcini" mushrooms) are among the most valued and prized edible wild mushrooms in the world [1].
In this study, two problems related to dried porcini mushroom were investigated:

1) infections by mycophilic fungi (*Sepedonium* spp.) that, at the final state, can lead to the total necrosis of the host [2];
2) fraudulent addition of extraneous species, especially *Tylopilus* spp. and *Boletus violaceofuscus*, within dried "porcini" mushrooms, mainly from those imported from China.

Both of them represent considerable problems from a commercial point of view.
Up to date, identification of *Sepedonium* infections and characterisation of mushroom species have been accomplished only through macroscopic and microscopic visual inspection performed by professional mycologists, while no instrumental analytical methods have been proposed to support safety and authentication of dried *Boletus edulis*.
In order to address these issues, two different chemometrical approaches have been proposed.

In particular, *Sepedonium* infection was evaluated investigating the joint use of hyperspectral imaging (HSI) and principal component analysis (PCA) [3].
Hyperspectral images were obtained using a pushbroom line-scanning HSI instrument, operating in the wavelength range between 400 and 1000 nm with 5 nm spectral resolution.
The results showed that the PCA-based hyperspectral approach is able to provide valuable information to the detection of contaminated parts in mushroom samples.

As far as the fraudulent addition of extraneous species is concerned, a set of dried mushrooms including *Boletus edulis*, *Tylopilus* spp. and *Boletus violaceofuscus* specimens were analysed by near-infrared spectroscopy (NIRS). NIR spectra were used to develop reliable and efficient class-models using a novel method, partial least squares density modelling (PLS-DM) [4], and the two most commonly used class-modeling techniques, UNEQ and SIMCA. In this case, the results showed that NIR spectroscopy, coupled with chemometric class-modelling techniques, can be suggested as an effective analytical strategy for the verification of authenticity of dried mushrooms of the *Boletus edulis* group.

Acknowledgement: DV-Optic *Tecnologie d'Avanguardia* (Padova, Italy) is gratefully acknowledged for having granted instrumentation (HSI) use.

References:

[1] N. Sitta and M. Floriani, Econ. Bot, 62, 307 (2008).
[2] K. Pöldmaa, D.F. Farr, E.B. McCray. ARS. USDA Agric. Res. Serv, United States Dep. Agric (2015).
[3 L. Bagnasco, M. Zotti, N. Sitta, P. Oliveri, Talanta, 144,1225 (2015).
[4] P. Oliveri, M. I. López, M. C. Casolino, I. Ruisánchez, M. P. Callao, L. Medini and S. Lanteri, Anal. Chim.Acta, 851, 30 (2014).

P-172

CHEMOMETRIC STRATEGIES FOR COMPRESSION AND RESOLUTION OF MASS SPECTROMETRY IMAGES

C. Bedia[1], C. Beltrand[1], R. Tauler[1], J. Jaumot[1]

[1]Department of Environmental Chemistry, IDAEA-CSIC, Jordi Girona 18-26, Barcelona, Spain
carmen.bedia@idaea.csic.es

Mass Spectrometry Imaging (MSI) technology is a useful tool for the analysis of complex mixtures in real biological samples, such as cells or tissues [1]. MSI couples the spatial information provided by the spectral imaging techniques with the chemical specificity based on the mass accuracy of the mass spectrometry techniques [2]. Up to now, the most common way to analyse MSI data is the a priori selection of mass values and the visual inspection of their distribution in the image, following a targeted-based approach. Only a few number of examples applying chemometric tools for MSI data analysis can be found in literature [3-5]. This is probably caused by the vast dimensionality and huge storage requirements of the generated data sets, which require a preliminary compression step, usually by binning mass spectra.

In this work, lipidomic changes occurring during germination of bean seeds were monitored using MSI technology. Evaluation of MS images obtained during the germination process allowed identifying changes in the seed composition along the process and detecting the spatial location of these lipidomic changes. In order to carry out these analyses, different strategies of data compression and of simultaneous analysis of multiple MS images were developed. An approach based on the detection of "region of interests" was utilized to reduce the size of the data set to be analysed, which together with the application of MCR-ALS using a column-wise matrix augmentation strategy allowed the full resolution of the different MS images describing the seed germination process. Thus, information related to the lipids present at each germination stage and their spatial distribution in the seed was finally attained.

Acknowledgement: The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n. 320737.

References:

[1] K. Chughtai and R.M.A. Heeren, Chemical Reviews, 110(5), 3273 (2010).
[2] D.S. Cornett, M.L. Reyzer, P. Chaurand, and R.M. Caprioli, Nature Methods, 4(10), 828 (2007).
[3] J.M. Fonville, C. Carter, O. Cloarec, J.K. Nicholson, J.C. Lindon, J. Bunch, and E. Holmes, Analytical Chemistry, 84(3), 1310 (2011).
[4] Y. Gut, M. Boiret, L. Bultel, T. Renaud, A. Chetouani, A. Hafiane, Y.M. Ginot, and R. Jennane, Journal of Pharmaceutical and Biomedical Analysis, 105, 91 (2015).
[5] J. Jaumot and R. Tauler, Analyst, 140(3), 837 (2015).

P-173

PHOTOCHEMICAL STUDY OF I-MOTIF STRUCTURES BY TIME-RESOLVED SPECTROSCOPY AND PROCESS MODELLING TECHNIQUES

S. Benabou [1], C. Ruckebusch [2], M. Sliwa[2], A. de Juan[1], R. Gargallo[1]

[1] Department of Analytical Chemistry, University of Barcelona, Martí i Franqués 1-11, 08028 Barcelona, Spain.
[2] Université de Lille Sciences et technologies, LASIR CNRS, Cité scientifique, 59655 Villeneuve d'Ascq, France.
sbenabou_13@ub.edu

The *i*-motif structure is the only known DNA structure formed by cytosine-rich sequences and consists of parallel-stranded duplexes held together by intercalated base pairs (Figure 1). The formation *in vitro* of these structures in DNA sequences corresponding to centromeres, to the end of telomeres and to the promoter regions of several oncogenes, such as *c-kit*, *c-myc* or *bcl-2*, has been demonstrated [1-3]. These structures are not only interesting from a biophysical and biomedical point of view [3], but also for their potential application as photoswitchers and in nanotechnology [4].

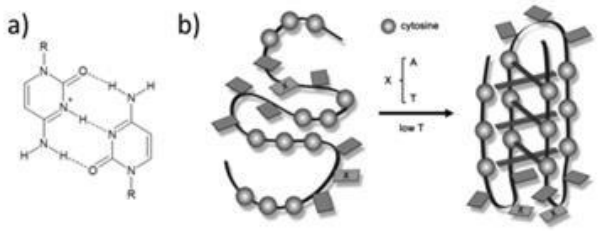


Figure 1. (a) Cytosine-protonated cytosine base pair; (b) Hypothetical scheme of the intramolecular *i*-motif structures adopted by the sequences proposed to study; X=A or T bases

To provide a different view on the structural changes and dynamic processes occurring after direct photoexcitation of cytosine-rich DNA sequences, the potential of time-resolved spectroscopic measurements to study the kinetics and mechanism of formation of *i*-motif structures has been tested. Different spectroscopic strategies have been used to monitor transitions between structural conformations and environmental changes occurring in and around short-lived transient species. Multiset analysis based on coupling spectroscopic data coming from experiments performed on structurally related DNA sequences allow a better distinction of concurrent events within the biochemical processes of interest. Multivariate data analysis approaches based on soft-modeling multivariate curve resolution methods, classical hard-modeling and hybrid hard- and soft-modeling methodologies enable the interpretation of processes involving *i*-motif structures.

References:

[1] M. Garavís, N. Escaja, V. Gabelica, A. Villasante and C. González, *Chemistry - A European Journal*, **21**, 9816 (2015).
[2] S. Benabou, A. Avino, R. Eritja, C. Gonzalez and R. Gargallo, *RSC Advances*, **4**, 26956(2014).
[3] J. Amato, N. Iaccarino, A. Randazzo, E. Novellino and B. Pagano, *ChemMedChem*, **9**, 2026(2014).
[4] Y. Dong, Z. Yang and D. Liu, *Acc. Chem. Res.*, **47**, 1853 (2014).

P-174

LOCAL REGRESSION APPROACHES FOR LARGE NEAR INFRARED SPECTROSCOPY (NIRS) DATASETS BASED ON PARTIAL LEAST-SQUARES (PLS) SCORES

F. Allegrini[1], J.A. Fernández Pierna[2], W.D. Fragoso[3], A.C. Olivieri[1], V. Baeten[2], P. Dardenne[2]

[1] Univ. Nacional de Rosario, Facultad de Ciencias Bioquímicas y Farmacéuticas, IQUIR, CONICET, Argentina
[2] Valorisation of Agricultural Products Dpt, Walloon Agricultural Research Centre, Gembloux, Belgium.
[3] Departamento de Química, Universidade Federal da Paraíba, Campus I, João Pessoa, Brazil.
j.fernandez@cra.wallonie.be

Extensive spectral databases compiling thousands of NIR reflectance spectra have been created during the last 30 years. Among these libraries, the ones related to typical parameters used in grain monitoring (as moisture, proteins, fat, starch and ash) in the agricultural sector are the most extended. In order to deal with these large databases, many chemometric tools have been developed, with Partial Least Squares (PLS) regression [1] being the most popular because of its robustness, simplicity and efficiency.

The approaches considered here are based on the idea of local regression [2,3]. This consists on a group of methods based on selecting from a large database, a set of samples spectrally similar to an unknown sample whose properties are to be predicted. Following this strategy, a specific local calibration is then developed for that sample using the previously selected "neighborhood" samples as calibration set. This means that each sample is predicted with a different calibration equation. Up to now, the main types of local regression methods described in the literature are Locally Weighted Regression (LWR) [4], the LOCAL algorithm [5], Comparison Analysis using Reconstructed Near Infrared and Constituent Data (CARNAC) [6], and the formulation of complex indexes to measure distances between samples after the application of data reduction (Principal Component Analysis or Fast Fourier Transform) [7].

In this work, we present two new approaches to perform sample selection in local regression methods based on PLS scores: Local Calibration by Percentile Selection (LCPS) and the Local Calibration by Customized Radii Selection (LCCRS). These approaches differ in the way they automatically fix and select a certain number of samples to optimally predict the unknown sample. Preliminary results obtained after applying these strategies over a large corn dataset show improved predictions if compared with a standard PLS model using the complete set of samples. The proposed methodologies could be extended to the prediction of more than one product from unique and large data sets, with the consequent savings in time and effort required for the development of individual calibration models.

References:

[1] - S. Wold, M. Sjostrom, L. Eriksson, PLS-regression: a basic tool of chemometrics, Chemometrics Intell, Lab. Syst. 58 (2001) 109–130.
[2] - T. Fearn, A.M.C. Davies, Locally-biased regression, J. Near Infrared Spectrosc. 11 (2003) 467–478.
[3] J.S. Shenk, P. Berzaghi, M.O. Westerhaus, Investigation of a LOCAL calibration procedure for near infrared instruments, J. Near Infrared Spectrosc. 5 (1997) 223–232.
[4] F.E. Barton, J.S. Shenk, M.O. Westerhaus, D.B. Funk, The development of near infrared wheat quality models by locally weighted regressions, J. Near Infrared Spectrosc. 8 (2000) 201–208.
[5] P. Berzaghi, J.S. Shenk, M.O. Westerhaus, LOCAL prediction with near infrared multi-product databases, J. Near Infrared Spectrosc. 8 (2000) 1–9.
[6] A.M.C. Davies, H.V. Britcher, J.G. Franklin, S.M. Ring, A. Grant, W.F. McClure, The application of Fourier-transformed near-infrared spectra to quantitative analysis by comparison of similarity indexes (CARNAC), Mikrochim. Acta 1 (1988) 61–64.
[7] F. Gogé, R. Joffre, C. Jolivet, I. Ross, L. Ranjard, Optimization criteria in sample selection step of local regression for quantitative analysis of large soil NIRS database, Chemometrics Intell, Lab. Syst. 110 (2012) 168–176.

P-175

SPECTROSCOPY AND CHEMOMETRICS FOR QUALITY EVALUATION AND CHARACTERIZATION OF AÇAI (*Euterpe oleracea*) FRUIT

G. Bartolomeo[1], A.L.S. Dias[2], H. Rogez[2], O. Abbas[3], F. Marini[1], V. Baeten[3], J.A. Fernández Pierna[3]

[1] Department of Chemistry, University of Rome "La Sapienza", P.le Aldo Moro 5, I-00185 Rome, Italy
[2] Faculty of Food Engineering and Center for Agro-food Valorisation of Amazonian Bioactive Compounds, Universidade Federal do Pará, Av. Perimetral s/n, 66.095-780, Belém-PA, Brazil
[3] Valorisation of Agricultural Products Department, Food and Feed Quality Unit, Walloon Agricultural Research Centre, Henseval Building, 24 Chaussée de Namur, 5030 Gembloux, Belgium.

The açai drupe is the fruit of a native Amazonian palm (*Euterpe oleracea*) known all over the world for its high polyphenol content supposed to have several health benefits. A large portion of these polyphenols is represented by anthocyanins, typical blue-reddish compounds that have high anti-oxidant and radical scavenging effects. Several approaches (mainly involving the use of chromatography) have been proposed in the literature to characterize and quantify such compounds in different food matrices. In the present study the possibility of obtaining the same kind of information through the use of spectroscopic techniques (with their advantageous characteristics of rapidity and non-destructivity, together with their low operating costs) is addressed. Moreover it was also verified whether it was possible to sort samples according to their breeding ground and if some nutritional differences were present.
A total of 97 powdered samples of freeze-dried açai juice and pulp from Brazil were analyzed using MIR, NIR, Raman and Fluorescence spectroscopy. Methanolic extract and fat fraction were also analyzed, looking for an improvement of the method. All acquired data were processed with MatLab, performing basic tools such as PCA, PLS regression against chromatographic reference data and PLS-DA. The results indicated that samples were able to be distinguished according to their provenience using MIR on powder or Raman spectroscopy on fat fraction with an accuracy of over 90%. Regarding the predictive capacity of the models in terms of phenolic compounds and anthocyanin content, the best results were achieved using Fluorescence on methanolic extract ($R^2 > 0.80$), but also Infrared techniques gave good results, even on powdered samples ($0.7 < R^2 < 0.8$) proving the efficacy of MIR and NIR as screening techniques. Nonetheless chemometric improvements could be of great commercial interest, because most of the spectroscopic techniques can work on-line in factories, allowing real-time analysis of processing fruit. These results indicate that this approach could be extended to similar polyphenol-rich fruits, to perform an easy, fast and reliable analysis.

Acknowledgement: This work was performed within the framework of the collaboration project PhotonFruit between CAPES (Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) and WBI (Wallonie-Bruxelles International).

P-176

DETERMINATION OF URIC ACID USING SURFACE-ENHANCED RAMAN SPECTROSCOPY AND MULTIVARIATE CURVE RESOLUTION

J.E.L. Villa, R.J. Poppi

[1]*Institute of Chemistry, University of Campinas, P.O. Box 6154, 13081-970, Campinas, SP, Brazil*
villa.javier03@gmail.com

Uric acid is the end product of the catabolization of purine nucleotides and is considered to be an important biomarker in urine and serum [1]. Previous studies showed that high uric acid concentrations in these biological fluids can be associated with renal diseases and preeclampsia, a hypertensive disorder that occurs during pregnancy and is the primary cause of maternal morbidity and mortality worldwide (more than 50,000 deaths per year), mainly in developing countries [1,2]. Raman spectroscopy is a powerful technique that has grown in application because it can provide chemical and structural information with minimal sample preparation. Moreover, in recent years, the development of portable Raman spectrometers introduced the possibility of rapid on-site detection in a wide variety of sample types [3,4]. The low efficiency of inelastic scattering restricted the use of Raman spectroscopy to relatively high concentration analyses; nevertheless, surface-enhanced Raman spectroscopy (SERS) is an interesting alternative to overcome this limitation. SERS can dramatically increase the efficiency of Raman scattering through a combination of electromagnetic and chemical contributions when the target molecule (analyte) is attached to the surface of metallic nanostructures (typically made of gold or silver) [5]. In this work, a portable quantitative method for on-site determination of uric acid in urine using surface-enhanced Raman spectroscopy (SERS) and gold nanoparticle-coated paper as a substrate. Gold nanoparticles were synthesized by Turkevich method and a procedure was developed for the rapid preparation of cost-effective SERS substrates that enabled the adequate control of a homogeneous active area and the use of small quantities of gold nanoparticles per substrate. The standard addition method and multivariate curve resolution-alternating least squares (MCR-ALS) were applied to compensate for the matrix effect and to address overlapping bands between uric acid and interferences SERS spectra. The proposed methodology enabling the analysis of complex matrix without the necessity to identify and include the interferences in the model (second order advantage), demonstrated better performance than conventional univariate methods (in terms of linearity, accuracy and precision), a wide linear range (0-3.5 mmol L^{-1}) and an adequate limit of detection (0.11 mmol L^{-1}). For the first time, a portable SERS method coupled with chemometrics was developed for the routine analysis of uric acid at clinically relevant concentrations with minimal sample preparation. This method could be implemented for the rapid preliminary testing of preeclampsia and extended for the on-site determination of other biomarkers in complex samples matrices.

Acknowledgement: The authors gratefully acknowledge the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), São Paulo, and Metrohm for providing the portable Raman spectrometer.

References:

[1] D.M. Carty, C. Delles and A.F. Dominiczak, Trends Cardiovasc. Med., 18, 186–194 (2008).
[2] L. Duley, Semin. Perinatol., 33, 130–137 (2009).
[3] P. R. Carey, J. Biol. Chem., 274, 26625–26628 (1999).
[4] M. Pérez-Alonso, K. Castro and J. M. Madariaga, Anal. Chim. Acta, 571, 121–128 (2006).
[5] G. McNay, D. Eustace, W. E. Smith, K. Faulds and D. Graham, Appl. Spectrosc., 65, 825–837 (2011).

P-177

DETECTION OF ADULTERATION IN MEAT PRODUCT USING OF HYPERSPECTRAL IMAGING

A. Karrer[1], A. Stuart[2], C. Craigie[2], K. Taukiri[2], M. M. Reis[2]

[1]University of Applied Science, Am Hofgarten 4, D-85354 Freising, Germany. [2]Food Assurance and Meat Quality, AgResearch, 10 Bisley Road, Private Bag 3123, Hamilton, New Zealand.

a.karrer@gmx.net

Food fraud is an increasing concern worldwide as is estimated to cost the global food industry US$30 to US$40 billion every year[1]. In 2013 it was found that meat of several species were mixed with lamb meat on the production of lamb rolls, a traditional item in Chinese cuisine[2]. New Zealand is China's largest supplier of imported lamb, with just over 50 percent market share and there is a growing interest of applying fast and non-invasive techniques for authentication of meat products to prevent food fraud [3]. This study investigates the application of hyperspectral imaging to detect whether meat in lamb-roll-type product is only from lamb or if meat of other species is present.

Meat and fat samples from different species (i.e. lamb, beef and pork) were used to produce templates mimicking a meat product. The meat was purchased in a local market from different butcheries and used to prepare 5 different frames. Partial Least Square Discriminant Analysis (PLS-DA) [4] and Soft Independent Modelling of Class Analogies (SIMCA) [5] were applied as classification models. Models were fitted using cross validation with data obtained in one frame and subsequently applied to data from the other three frames as validation. The hypersepectral data (550-1700 nm) was collected from frames as fresh unpacked (just after preparation), fresh vacuum packed, frozen vacuum packed, and frozen unpacked. The best performance was observed when using PLS-DA was applied to vacuum packed fresh samples. The best PLS-DA model was then applied to a 5th frame with a design different from the original frames used to fit and validate this model. This resulted in very good classification, with lamb meat correctly identified and no misclassification. The model based on SIMCA was tested as it would mainly require a representative spectral database with lamb meat. However the best approach was based on PLS-DA, which indicates the need for a comprehensive database including meat of several species.

Acknowledgement: AgResearch core fund.

References:

[1] https://www.pwc.nl/nl/assets/documents/pwc-food-fraud-vulnerability-assessment.pdf (access 02/02/2016).
[2] http://www.interest.co.nz/rural-news/64330/chinese-authorities-investigating-fake-nz-lamb-products-sold-major-chain-hot-pot (access 02/02/2016).
[3] http://www.nzherald.co.nz/business/news/article.cfm?c_id=3&objectid=11246243 (access 02/02/2016).
[4] Bharati, M.H., J.J. Liu, and J.F. MacGregor, *Image texture analysis: methods and comparisons.* Chemometrics and Intelligent Laboratory Systems, 2004. **72**(1): p. 57-71.
[5] Liu, D., et al., *Lychee Variety Discrimination by Hyperspectral Imaging Coupled with Multivariate Classification.* Food Analytical Methods, 2014. **7**(9): p. 1848-1857.

P-178

Calculation of Molecular Mixture Ratios Using Single-Pixel Time-of-Flight Secondary Ion Mass Spectrometry Analysis

Rainer Kassenböhmer, Felix Draude, Martin Körsgen, Andreas Pelster, Heinrich F. Arlinghaus

Physikalisches Institut, Westfälische Wilhelms-Universität Münster, Wilhelm-Klemm-Straße 10, 48149 Münster, Germany
kassenboehmer@uni-muenster.de

We report an algorithm which for the first time offers the capability to calculate the percentages of the ingredients from time-of-flight secondary ion mass spectrometry (ToF-SIMS) spectra in model systems when all possible constituents are known. The algorithm is based on a linear mixing model. In a first step the dimension of the data space is reduced using discriminant analysis. Discriminant scores are then used to calculate the ratios of the mixture compounds. As an example we calculated the percentages of phospholipid mixtures separately for positive and negative ion single-pixel ToF-SIMS spectra, which revealed good correlations to real mixing ratios [1]. It could be shown that even phosphatidylcholine and sphingomyeline which have identical headgroups could be distinguished from another in mixtures. ToF-SIMS imaging combined with discriminant analysis can be a powerful label-free technique to identify domain structures in model membranes with submicrometer-scale resolution. The algorithm outlined may also be useful to calculate the percentages of ingredients of mixtures not only in ToF-SIMS imaging but in other mass spectrometric techniques such as electrospray ionization or matrix-assisted laser desorption/ionization mass spectrometry. Since fragmentation of molecules is significantly lower with these techniques, it should be possible to examine mixtures of large molecules, for example proteins, and perhaps even more complex mixtures such as pharmaceutical drugs and their metabolites. Moreover, analytical techniques such as Raman spectroscopy, Fourier transform infrared spectroscopy, or hyphenated chromatographic methods, which provide linear combinations of chemical spectra, could also be potential application areas for this algorithm.

References:

[1] R. Kassenböhmer, F. Draude, M. Körsgen, A. Pelster, H. F. Arlinghaus, Anal. Chem. 87, 7795 (2015)

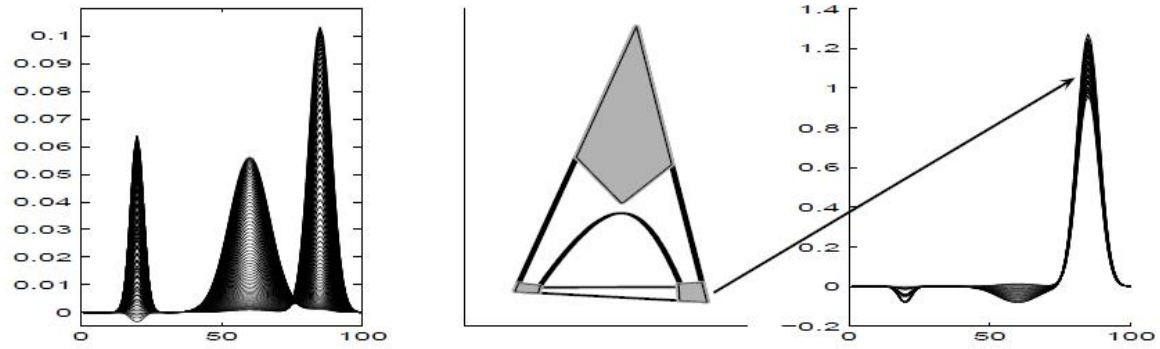
P-179

GENERALIZED BORGEN PLOTS - A NEW TOOL TO ANALYZE PERTURBED TWO-WAY DATA

A. Jürß [1,2], M. Sawall [1], K. Neymeyr [1,2]

[1]*Department of Mathematics, Universität Rostock, Ulmenstraße 69,18057 Rostock, Germany*
[2]*Leibniz Institute for Catalysis, Albert-Einstein-Straße 29a, 18059 Rostock,Germany*
annekathrin.juerss@uni-rostock.de

Borgen plots are known as a geometric constructive tool to analyze bilinear spectroscopic data $D = CA$ [1]. They are generalization of Lawton-Sylvestre plots [2] which can be considered as one of the first SMCR-methods. Lawton-Sylvestre plots and Borgen plots serve to explore the spectral profiles and concentration lines which reconstruct the given spectroscopic data sample D. Whereas Lawton-Sylvestre plots can be applied only to spectroscopic data with two absorbing species, Borgen plots allow analyzing three component systems. However, these approaches are limited to nonnegative factorizations and can hardly be applied in the presence of noise and perturbations.
We present a generalization of the classical Borgen method, which serves to analyze perturbed data and allows small negative entries in the factors C and A [3]. The algorithm is implemented in the *FAC-PACK* toolbox. Further we investigate whether the geometric concepts behind the Lawton-Sylvestre plots and Borgen plots can also be applied to data samples with four absorbing species. We show that the problem for four-component systems is NP-hard. Therefore constructive methods seem to be inappropriate to solve the reconstruction problem.



Left: Model data of a three-component system D. Middle: Generalized Borgen plot of D. Right: Range of feasible spectra for one component.

References:
[1] O.S. Borgen, B.R. Kowalski, Analytica Chimica Acta, 174, 1 (1985)
[2] W.H. Lawton, E.A. Sylvestre, Technometrics, 13, 617 (1971)
[3] A. Jürß, M. Sawall, K. Neymeyr, Journal of Chemometrics, 29, 420 (2015)

P-180

Multivariate Approaches to Mass Spectrometry Image Fusion

B.J. Tyler[1] and R. Kassenböhmer[1]

[1] *Physikalisches Institut, Universität Münster, Wilhelm-Klemm Straße 10, Múnster, Germany, 48149*
tyler@uni-muenster.de

Mass Spectrometry Imaging (MSI) encompasses a range of techniques including time-of -flight secondary-ion mass spectrometry (ToF-SIMS) imaging, matrix assisted laser desoprtion ionization (MALDI) MSI, and Ambient MSI. Because the different techniques contain complementary strengths, there is growing interest in combining them to obtain higher quality results. In general, obtaining the highest spatial resolution comes at the expense of lower quality spectral information (i.e. reduced mass range, lower mass resolution and accuracy, and decreased signal-to-noise). Image fusion has been proposed as a solution to this problem. In this paper we present a modified Principal Component Analysis (PCA) image fusion approach for combining high mass resolution MS images with high spatial resolution MS images. This dual-hyperspectral PCA fusion algorithm makes optimal use of the spectral and spatial information in both image modes to reduce fusion artifacts and improve sharpness of image features without loss of the high mass resolution information. Performance of the algorithm has been evaluated on both synthetic images and mass spectral images of mouse diaphragm muscle. By fusing the two imaging modes, an order of magnitude improvement in spatial resolution and signal-to-noise ratio is achieved while retaining the high mass resolution information. Figure 1 illustrates the improvements in the mouse diaphragm muscle images. An overlay of three components corresponding to cell nuclei (green), adipose tissue (blue) and lipid inclusion (red) are shown for the fused image (left) and the original high mass resolution image (right). In the fused image, individual nuclei, muscle fiber bundles and substructures in the adipose tissue are clearly resolved. These improvements are achieved by combining the complementary information from the two imaging modes rather than through interpolation or filling of the data. Dual-spectral PCA image fusion shows promise for improving the quality of mass spectrometry images of biological specimens and has potential applications with other complementary spectral image techniques.

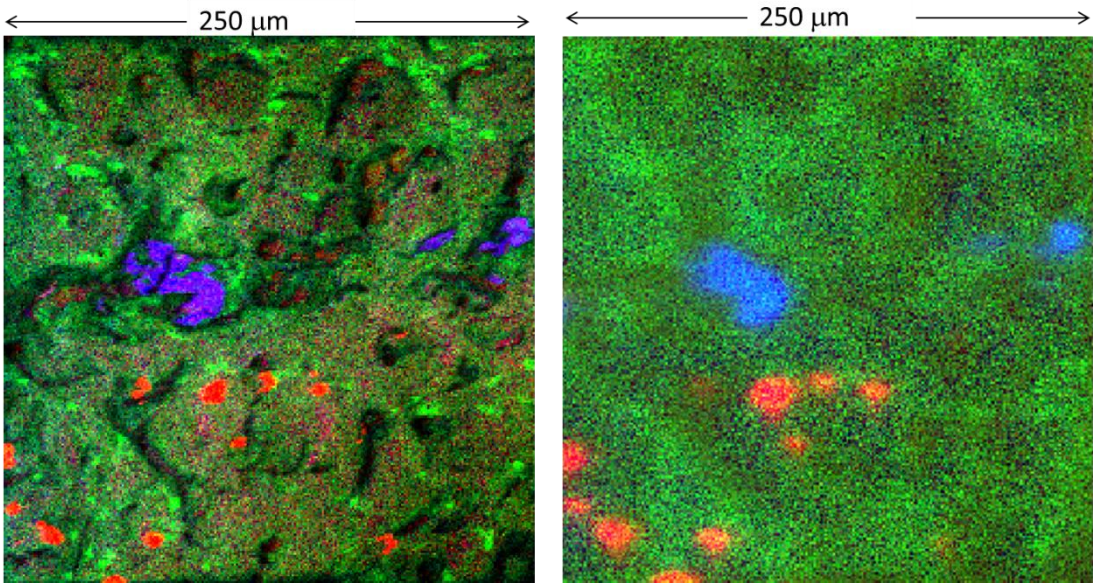


Figure 1. Overlay of three components for the fused image (left) and original high mass resolution image (right). After image fusion the cell bodies and sub-cellular structures are clearly resolved.

P-181

Quantification of Biodiesel and HEFA in petroleum diesel fuel blends using MCR-ALS with correlation constraint

Willian F.C. Dantas[1], Julio C. L. Alves[1] and Ronei J. Poppi[1]

[1]Institute of Chemiatry, University of Campinas, P.O. BOX 6154, 13081-970, Campinas, Brazil.
willianfcd@gmail.com

Due to the increase use of blends of diesel with conventional and advanced biofuels, there is a need in the development of quality control methodologies, enabling the quantification and fraud identification.There are a lot of techniques available that can help on these developments, such as spectroscopic, mass spectrometry, chromatography and others [1].
In addition to this diversity of techniques the analyst should be aware that mostly of the samples to be analyzed has limited quantity. Furthermore, many samples can be contaminated, and to identify and quantify an analyte can be a challenge. Therefore, techniques with minimal casualties or without sample destruction becomes interesting [2].
In this work Raman Spectroscopy was employed as analytcal technique, since it encompasses all desired characteristics. The samples studied were blends of diesel with biodiesel (fatty acid methyl esters) and HEFA (hydrotreated esters and fatty acids) fuels. Firstly, for a quantification study, it was built a calibration curves for each fuel. The calibration set had 22 samples prepared acording to the following experimental design: the first 10 samples had the HEFA content in each sample increases by 1% in the range of 1 – 10%, the biodiesel content in each sample decreased by 0.5% in the range of 25.5 – 21% and the petroleum diesel was used to complete the volume of each sample; the next six samples had the biodiesel content increased by 10% in the range of 0 – 50% and the petroleum diesel content decreased by 10% in the range of 100 – 50%; the last six samples had the same concetration profile only that it was replaced biodiesel by HEFA. The validation set had 14 samples and it was prepared with HEFA, biodiesel and petroleum diesel contents differents of calibration sets, but included in the calibration set analytical range.
Next, for a fraud study, samples of diesel B-10 (90% of petroleum diesel asd 10% of HEFA or biodiesel or 5% of each biodiesel and HEFA) were contamined with ethanol, cyclohexane, hexane, gasoline, kerosene and blend of gasoline/ethanol. The validation set had five samples with diesel B-10 and the adulterant content in each sample increased by 2,44% in the range of 0 – 9,09%.
Multivariate curve resolution with alternating least squares (MCR-ALS) employing the toolbox MCR-ALS GUI 2.0 [3] and the correlation constraint was used to construct the calibration curves. The chosed data preprocessing was smoothing and the first derivative using Savitzky–Golay filter with normalization.
The results showed that it was possible to obtain the pure spectra, allowing its identification, of each compound and to quantify accurately the amount of biodiesel and HEFA in diesel samples. The RMSEP in the validation set for biodiesel and HEFA were 0.53% and 0.98%, respectively, and the spectra recover was 99%. The fuels were also satisfactory quantified in the presence of contaminants, in the range of 1.4 to 5.8%. In this case, it was possible to obtain the spectra of the contaminat, allowing its identification.

Acknowledgements: Authors would like to thank Neste Oil, Porvoo, Finland, for providing the HEFA sample and Petrobras, Brazil, for providing the Diesel and Biodiesel samples and for financial support.

References:
[1] G. McLaughlin, I.K. Lednev, J. Forensic Sci. 60, 595 (2015).
[2] C.K. Muro, K.C. Doty, J. Bueno, L. Hala, I.K. Lednev, Anal. Chem. 87, 306 (2015).
[3] J. Jaumot, A. de Juan, R. Tauler, Chemom. Intell. Lab. Syst. 140, 1 (2015).

P-182

MECHANISM ELUCIDATION OF COMPLEX REACTIONS USING SENSITIVITY ANALYSIS WITH HARD AND SOFT MODELLING

Jesus A. Ágreda[1], Iván F. Robayo[1]
[1]*Department of Chemistry, Universidad Nacional de ColombiaAv. Cra. 30 # 45-03, Bogotá, Colombia.*
ifrobayom@unal.edu.co

A computational package was developed to support the elucidation, understanding and improvement of reaction mechanisms. The package has four programs: *ProgPrincipal* resolves the differential equations of a reaction mechanism using the Gear algorithm, *Raskolnikov* and *Dostoyevsky* make concentration and reaction rate sensitivity analysis[1], respectively, and *Karamazov* performs non-linear fitting using the Levenberg-Marquardt algorithm. This package was used to improve the previously proposed mechanism of the Uncatalyzed Bromate Oscillator (UBO)[2] a well known chemical oscillator, whose reaction mechanism is still unclear because it has plenty of intermediates[3]. With the use of the package and the results obtained by Multivariate Curve Resolution Alternanting Least Squares of a kinetic UV/Vis spectrophotometric study of the reaction, a mechanism of this chemical oscillator was proposed taking as starting point the György, Varga, Körös, Field and Ruoff model[4]. The agreement between the experimental and modeled data is fairly good. In the next future, it is expected that this protocol will help to elucidate the dynamics of other oscillators that has species which absorb in the UV-VIS region of the electromagnetic spectrum.
Acknowledgement: The authors thank the support of the Universidad Nacional de Colombia through the proyects DIB-14234, DIB-19076, and for providing a fellowship to Iván Robayo.

References:
[1] T. Turanyi. *J. Math. Chem.*, 5, 203 (1990).
[2] E. Körös, M. Orban, I. Habon. *J. Phys. Chem.*, 84, 559 (1980).
[3] M. Orban, G. Szókán. *ACH ¨C Models Chem.*, 132, 179 (1995).
[4] L. Györgyi, M. Varga, E. Körös, R. Field, P. Ruoff. *J. Phys. Chem.* 93, 2836 (1989).

P-183

ANALYSIS OF ACTIVE PRINCIPLES IN FISH PELLETS USING FOURIER TRANSFORM INFRARED IMAGING AND CHEMOMETRIC TECHNIQUES.

C.Y. Bastidas [1], C. von Plessing [2], J. Troncoso [3], R. del P. Castillo [1].

[1]Departamento de Análisis Instrumental, Facultad de Farmacia, Universidad de Concepción, Concepción, Chile. [2] Departamento de Farmacia, Facultad de Farmacia, Universidad de Concepción, Concepción, Chile.
[3] Ewos Innovation, Colaco, Chile.
***e-mail**: cbastidas@udec.cl*

Fourier Transform Infrared spectroscopic imaging (FT-IR imaging) in ATR (Attenuated Total Reflection) mode is an important and very useful tool to analyze chemical composition in pharmaceutical industry. For example, FT-IR imaging has been used to study drug release [1] and compaction of pharmaceutical tablets [2]. In this work, we analyze and compare the distribution of the active principle FF and the inclusion complex FF-C at microscopic level in fish pellets at two low concentrations, using mid infrared spectra (4000-740cm^{-1}) of different zones of pellets. Principal Component Analysis (PCA) and Multivariate Curve Resolution - Alternating Least Squares (MCR-ALS), were evaluated and compared to analyse top, side stand, cross and lenghtwise surfaces from both pellets. Results show ability of both techniques to detect the active principles at low concentrations in a very complex matrix. Obtained MCR-ALS pure spectra is in agreement with the spectra of isolated active principles. We realize that active principles are located just in top surface in the pellets of least concentration while principles are in all surfaces in the other pellets, meaning that analysis of FT-IR ATR hyperspectral images combined with chemometrics techniques is an effective analytical tool to study FF and FF-C distribution in fish pellets with ability to detect the active principles until pixel sizes of 1.56 μm x 1.56 μm in the images.

Acknowledgement: Authors thanks Fondecyt 11130388 project by the chemometrics tools availability.

References:
[1] S.G. Kazarian, A.V. Ewing. Expert Opin Drug Deliv, 2013. Vol. **10**(9): p. 1207-21.
[2] P. Wray, et al., Journal of Pharmaceutical Sciences, 2008. Vol. **97**(10): p. 4269-4277.

P-184

QUANTIFICATION OF LIGNOCELLULOSIC COMPONENTS IN PRETREATED WOOD USING FT-IR IMAGING AND MULTIVARIATE IMAGE REGRESSION (MIR) FOR EVALUATION OF WOOD RECALCITRANCE AT MICROSCOPIC SCALE

Juan Araya [1,2], Juanita Freer[2,3], Cristian Arévalo[2,3], Rosario del P. Castillo[1,2]

[1]Department of Instrumental Analysis, Faculty of Pharmacy, University of Concepcion, Concepcion, Chile.
[2] Biotechnology Center, University of Concepcion, Concepcion, Chile
3 Faculty of Chemical Sciences, University of Concepcion, Concepcion, Chile

E-mail: rosariocastillo@udec.cl

The complex arrangement of lignocellulosic componentes in wood is responsible for the difficulties of most of processes in production of biofuels from this bioresource. Wood recalcitrance is attributed to multiple factors, among them, accesible area of cellulose (AAC) to enzymatic hydrolysis can be one of the most important in production of bioethanol [1]. AAC can be evaluated at microscopic level by different techniques suchs as infrared microspectroscopy, confocal microscopy and scannig electronic microscopy (SEM). Infrared microspectroscopy acompanied by multivariate techniques has been useful to describe areas in pulp rich in cellulose, lignin or complex cellulose-lignin [2]. In this work, mid infrared spectra of microscopic areas with pixel sizes of 1.56 μm x 1.56 μm of wood pretreated by different processes were evaluated by multivariate image regression (MIR) with the aim to quantify glucans and lignin present in fibers. Partial least squares was applied as MIR method [2] on a group of hyperspectral images in order to build a predictive model to predict the bulk and pixel concentration of lignocellulosic components in the fibers. Lignin and glucans were quantifed using HPLC as bulk reference method. MIR models were validated by cross and external validation. Histograms of pixel concentrations versus pixel points were used to compare the homogeneity of samples. Bulk concentration predictions obtained by MIR show errors of prediction similar to the conventional macroscopic infrared analysis. Both, concentrations and homogeneity show in a more detailed way the presence of cellulose or lignin in samples and allow us to make a comparison between AAC of samples to understand the recalcitrance of pretreated wood.

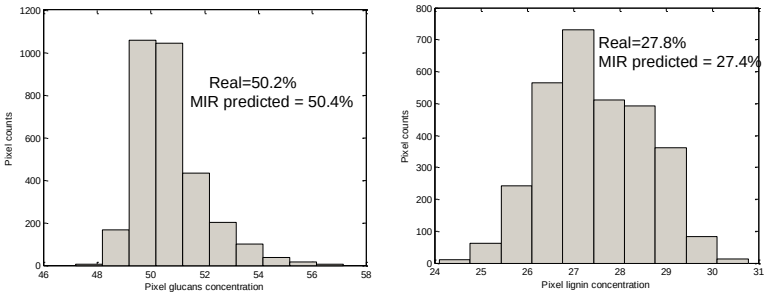


Figure 1. Glucans (left) and lignin (right) histograms for evaluation of homogeneity in one image of validation set.

Acknowledgement: Authors thank the support of projects Fondecyt 11130388, FONDECYT, CONICYT, Chile. Juan Araya thanks CONICYT-PCHA/Doctorado Nacional/2013-21130641 scholarship.

References:

[1] X. Zhao, L. Zhang, D. Liu. *Biofuels, Bioprod. Bioref*. 6, 465 (2012)

[2] R. P. Castillo, J. Araya, E. Troncoso, S. Vinet, J. Freer. *Anal. Chim. Acta*, 866,10 (2015)

[3] J. Burger & P.Geladi. *J. of Chemom*, 20, 106 (2006).

P-185

Multivariate Analysis of Large X-ray Fluorescence Spectroscopic Images

P. Van Espen [1], A. Cabal [1,2]

[1]AXES, Dept. of Chemistry, University of Antwerp, Groenenborgerlaan 171, Antwerp, Belgium,
piet.vanespen@uantwerpen.be
[2]Dept. of Physics, CEADEN, Calle 30 #502, Havana, Cuba

Scanning macro X-ray fluorescence (MA-XRF) is used frequently for the chemical characterisation of large planar objects. A focussed x-ray beam is scanned over the sample and x-ray spectra are acquired at each spot. Due to the non-invasive nature, the method is especially useful for the study of cultural heritage items such as paintings [1]. It provides information about the creative process of the artist, help in the restoration process and can often visualise hidden under-paintings.

The amount of data that is generated by MA-XRF is enormous. Even a small part of a painting of 50 × 50 cm^2 scanned in steps of 0.5 mm results in a data cube of 10^6 x-ray spectra.

We have recently developed a method to analyse this large number of spectra in a reasonable amount of type (minutes) using a "hybrid" least squares fitting. The method alternates between the very accurate but slow non-linear least squares fitting and a very fast variant of linear least squares. The results of such analysis are maps (images) with the net, i.e. interference free and background corrected, intensity of the characteristic x-ray lines of the element present in the sample.

However "element images" are not the most useful way to study paintings as they are made with pigments rather than with elements.

In this study we have investigated how various multivariate methods such as PCA, PLS en SVM can help in the interpretation of the data.

Further we have studied the possibility to use the spectral data directly, i.e. apply multivariate methods directly to the measured spectra rather that analysing them first to obtain the elemental data. It turns out that very similar and interesting results can be obtained providing a background correction is applied to the spectral data.

References:

[1] A. Cabal, O. Schalm, P. Eyskens, P. Willems, A. Harth and P. Van Espen, *X-Ray Spectrom*, **44**, 141 (2015).

P-186

COMPARISON BETWEEN UV-MCR-ALS AND HPLC-DAD TO MONITORING THE REACTION OF FORMATION OF 5-HMF FROM GLUCOSE USING NIOBIC ACID AS CATALYST

M.N. Catrinck[1], L.R. Terra[1], J.V. Roque[1], M.H.P. Barbosa[2], R.M. Ribas[3], R.S. Monteiro[3,4] and R.F. Teófilo[1]

[1]Departamento de Química, Universidade Federal de Viçosa, Viçosa, MG, Brazil
[2]Departamento de Fitotecnia, Universidade Federal de Viçosa, Viçosa, MG, Brazil
[3]Companhia Brasileira de Metalurgia e Mineração, CBMM, Araxá, MG, Brazil
[4]Catalysis Consultoria Ltda., Rio de Janeiro, RJ, Brazil
rteofilo@gmail.com

The 5-hydroxymethylfurfural (5-HMF), produced from dehydration acid catalyzed of six carbon sugars, such as glucose, is considered to be one of the top value-added chemicals [1]. Traditionally, the 5-HMF is obtained by homogeneous acids catalyst. However, several problems are found, such as corrosion, difficulty of separation and recover of products and catalysts. Therefore, their replacement with solid acids catalyst is highly desirable in the chemical industry. In this way, niobic acids ($Nb_2O_5 \cdot nH_2O$) have attracted great interest as catalysts, due to its strong acid properties which can be preserved in polar liquids [2]. The conversion of biomass into chemicals is traditionally monitored by chromatographic techniques, but they are expensive and often time consuming. Otherwise chemometric methods, such as multivariate curve resolution - alternating least squares (MCR-ALS), can be used to solve and identify overlapping analytical signals, *e.g.*, UV/Vis spectra [3]. In this way, the aims of this study were monitoring the glucose conversion into 5-HMF over niobic acid catalyst using UV spectroscopy combined with MCR-ALS and compare these results with the ones obtained by high performance liquid chromatography with diode array detector (HPLC-DAD). Calibration curve using UV-MCR-ALS was built to HMF in the presence of levulinic acid in the range of 2.10 to 16.10 mg L^{-1}. For HPLC-DAD, 5-HMF concentration range used for regression was 10.0 to 800.0 mg L^{-1}. The calibration models were obtained and the comparison between them was performed by analyzing statistical parameters of quality such as root mean square error calibration (*RMSEC)* and determination coefficient (R^2). The parameters obtained for UV-MCR-ALS were: *RMSEC* 0.76 mg L^{-1} and R^2 0.968. For HPLC-DAD were obtained: *RMSEC* 4.92 mg L^{-1} and R^2 0.999. Glucose conversion into 5-HMF was performed under the following conditions: glucose concentration 2 %(w/w), 0.6 g of catalyst mass, reaction temperature and time 120 ºC and 100 min, respectively. The samples were analyzed by UV-MCR-ALS and HPLC-DAD. In order to evaluate the performance of both methods was plotted the predicted concentration of 5-HMF obtained by UV-MCR-ALS *versus* HPLC-DAD. It was observed a linear fit between the values. The highest selectivity for HMF was 26.3%. The results suggest that both methods are able to predict the concentration of HMF in mixture reactions. The niobium acid is a promising catalyst for glucose conversion into HMF. Regarding the UV-MCR-ALS method, it is possible to obtain the relative concentrations and pure spectra of the compounds in a mixture without chromatographic separation. Furthermore, it showed to be simple, fast, fairly selective and efficient for monitoring this system.

Acknowledgement: The authors thankful to the financial support from Companhia Brasileira de Metalurgia e Mineração, CBMM, CAPES, CNPq, FAPEMIG and RQ-MG.

References:

6. J.M.R. Gallo, D.M. Alonso, M.A. Mellmer and J.A. Dumesic, Green Chem., 15, 85 (2013).
7. Y. Zhang, J. Wang, X. Li, X. Liu, Y. Xia, B. Hu, G. Lu, Y. Wang, Fuel, 139, 301 (2015).
8. J.J.A. de Juan, R. Tauler, Chemometr. Intell. Lab., 140, 1 (2015).

P-187

Fluorescence soft-sensor for the classification of bacteriophage antigen – antibody complexes

M. Zabadaj [1], E. Roźniecka[2], J. Niedziółka-Jönsson[2], P. Ciosek[1],

[1] *Chemistry Department, Warsaw University of Technology, Noakowskiego 3, 00-660 Warszawa, Poland*
[2] *Institute of Physical Chemistry, Polish Academy of Sciences, Kasprzaka 44/52, 01-224 Warszawa, Poland*
email: mzabadaj@ch.pw.edu.pl

Herein, we report a fluorescent soft-sensor for bacteriophage with covalent bonded anti-bodies classification. The new soft-sensor allow for a fast, simple and selective detection of T1, T4 and T7 viruses. The staining of viral nucleic acids with acridine orange and their subsequent fluorescence examination permits for the detection of the bacteriophages.

Phages are viruses whose hosts are bacterial cells. They identify their hosts by a specific receptor molecules on the outside of the host cell. Once the phages find their specific receptors, they bind to the bacterial cell and inject their nucleic acid inside the cell. The binding between phage and host can be so specific that only certain strains of a single species can be infected [1]. The antibodies are bare protein molecules. Since antibodies are available for a large variety of pathogenic viruses, the described concept is very flexible and can be adapted to detect many different viruses, not only bacteriophages [2].

In this work a fluorescence soft-sensor was used to the study of bacteriophage antigen-antibody interaction. The bacteriophages were labelled using fluorescence dye and then the changes in 2D-fluorescence spectrum were observed. The viruses complexed with various antibodies were examinated. To show the changes before and after adding specific or non-specific antibodies to the virus samples, 2D-fluorescent spectra were analyzed with the use of PCA or PLS-DA. The various samples of bacteriophag-antibody complexes formed clusters on PLS-DA plot, which were usually easily differentiated (Fig. 1). This means that 2D-fluorescence soft-sensor is useful for classification of bacteriophages with covalent bonded antibodies. Moreover, the developed soft-sensor allows for qualitative analysis of interaction between bacteriophage and antibody. Results can be obtained in a one-step procedure in a short time.

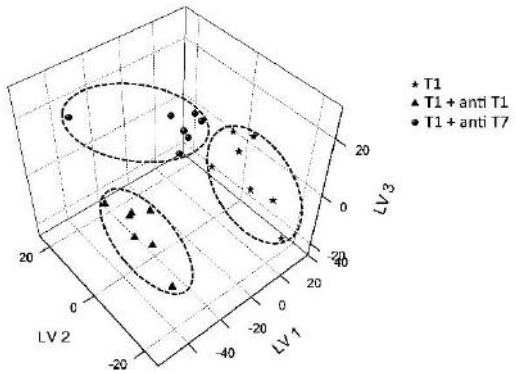


Figure 3. PLS-DA plot of chemical images of various bacteriophage antigen-antibody complexes.

References:

[1] P. A. Mosier-Boss, S. H. Liberman, J. M. Andrews, F. L. Rohwer, L. E. Wegley, M. Breitbart, *Use of Fluorescently Labeled Phage in the Detection and Identification of Bacterial Species*, Society for Applied Spectroscopy, Volume 57, Number 9, 1138 – 1144 (2003)

[2] A. Lesniewski, M. Los, M. Jonsson-Niedziółka, A. Krajewska, K. Szot, J. M. Los, J. Niedziolka-Jonsson, *Antibody Modified Gold Nanoparticles for Fast and Selective, Colorimetric T7 Bacteriophage Detection*, Bioconjugate Chem., 25, 644−648 (2014)

Acknowledgement:
This work has been supported by National Science Centre within a framework of OPUS project "(Bio)electrochemical imaging for cellular microphysiometry" UMO-2014/15/B/ST4/04807.

P-188

Multivariate calibration using pure component spectra measurements

C. Srinesh[1], S. Narasimhan[1], G. Jayaraman[2]

[1]Department of Chemical Engineering, Indian Institute of Technology Madras, Chennai, India
[2]Department of Biotechnology,, Indian Institute of Technology Madras, Chennai,, India
email: ch14s300@smail.iitm.ac.in

(Bio) chemical processes require online concentration measurements for kinetic modeling or for the purpose of process control. Near-Infrared Spectroscopy (NIRS) is a popular spectroscopic technique that is widely used for estimating concentrations directly without hassles of offline analysis enabling its use for online process correction/control. Multivariate calibration (MVC) or curve resolution (MCR) methods can be used to estimate concentrations from the spectral data. However due to high intrinsic absorption of water along with presence of broad overlapping peaks that characterizes the region, the spectral signature of most of the compounds of interest is obscured, making it difficult to analyze these data[1]. Commonly used methods to monitor the analytes are PLS, MLR or PCR, these methods perform a regression between offline concentration measurements and spectral data based on the different assumptions [2,3]. The model developed is used to predict the new concentrations.

The process of building a calibration model by this method is laborious and time consuming requiring re-calibration when the compound to be monitored changes or in case of drifts which occurs due to change in process or instrumental conditions. As an alternative to traditional PCR or PLS calibration, we propose a calibration method that uses pure component spectrum of all the major compounds that are present in the mixture. The calibration procedure here involves augmenting the pure component spectra with the spectral measurements and using a modified Non-negative Matrix factorization algorithm (NMF)[4] coined as g-NMF (guided NMF) that factorizes the augmented spectral matrix into two non negative matrices keeping certain elements of the factored matrices fixed. These correspond to elements whose prior knowledge is available. This factorization is physically realizable as the factors estimated has a meaning, since both concentrations and pure component spectra are positive. Additionally a wavelength band selection is carried out to enhance the prediction quality.

The methodology is tested on NIR data obtained from an NIR spectrometer using an aqueous mixture of ethanol, glucose and biomass. The performance of the method is quantified using prediction RMSE of the individual compounds. The RMSE values are found to be satisfactory for all the three analytes. This technique can be easily extended to quantify more compounds and could be potentially employed for calibration of bio fermentor experiments that predominantly produce compounds whose spectra can be catalogued.

References:

[1] F. Ham, I Kostanic, G.Cohen, B.Gooch, IEEE transactions on Biomedical engineering, 44 (1997), 475 – 485
[2] S. Tosi, M. Rossi, E. Tamburini, G. Vaccari, A. Amaretti, and D. Matteuzzi (2003). Biotechnology Progress, 19(6), 1816–1821
[3] S. Gary, Trends in Analytical Chemistry, 25 (2006), 1057-1066
[4] D. Lee and S. Seung, Nature,401 (1999), 788-791

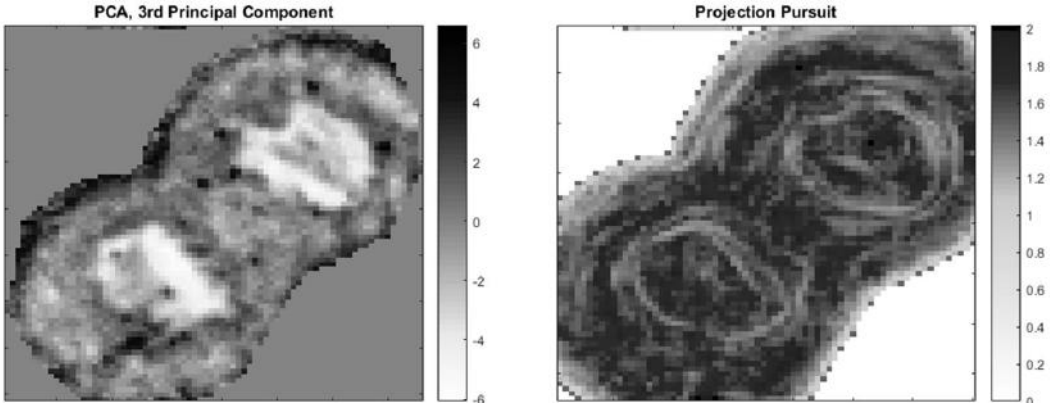
P-189

PROJECTION PURSUIT USING KURTOSIS AS A HYPERSPECTRAL IMAGE FEATURE ENHANCEMENT TOOL

C. Wicks[1], P.D. Wentzell [1], T.Karakach[2]

[1]Department of Chemistry, Dalhousie University, Halifax, NS, B3H 4R2, Canada, ch565235@dal.ca
[2]National Research Council of Canada, Halifax, NS, B3H 3Y8, Canada

Hyperspectral imaging techniques collect massive amounts of data that potentially carry important chemical information, but the extraction of this information requires specialized analysis tools. Many exploratory techniques have been applied to data of this type with the goal being to visualize characteristics of the image, both quantitatively and qualitatively. Some of these methods include principal component analysis (PCA) and classical least-squares (CLS) regression, as well as many clustering methods such as K-means [1]. However, the use of variance or simple distance-based algorithms as a means of clustering pixels is not always an effective means for distinguishing among different topological features in complex images with subtle spectral differences. As an alternative to these traditional approaches, projection pursuit (PP) analysis [2] can be applied using a projection index to search for more "interesting" projections of the data. The PP algorithm employed in this work was developed by Hou and Wentzell [3] and uses kurtosis as the projection index. The strategy used here analyzes smaller frames of the overall image. This method searches for differences within each frame, giving less distinctive features the opportunity to stand out from their surrounding pixels. Each frame is then represented by the kurtosis value calculated by the algorithm and a map can be use to create a visual representation of the image as a whole with potentially higher contrast. The figure below gives an example of using PCA and PP on an image of a splitting cell. In the left-hand frame the features extracted by the third principal component can be seen, namely the nuclei and a few other small structures. The right-hand frame shows results from PP where much detail surrounding the nuclei, as well the cell wall, can be seen as well as some other small structures. One drawback of the PP algorithm alone is that the data must be compressed if the sample-to-variable ratio is not large enough. This is typically done using PCA. Though it may sound computationally laborious, the algorithms actually run very quickly and efficiently. Results will be presented to illustrate the characteristics of this approach to image analysis compared to traditional methods.



References:

[1] J.M. Amigo, J. Cruz, M. Bautista, S. Maspoch, J. Coello, M.Blanco, Trends Anal. Chem, 27,696 (2008)
[2] J.H. Friedman, J.W. Tukey, IEEE Trans. Comput., 23, 881 (1974)
[3] S. Hou, P.D. Wentzell, Anal. Chim. Acta, 704, 1 (2011)

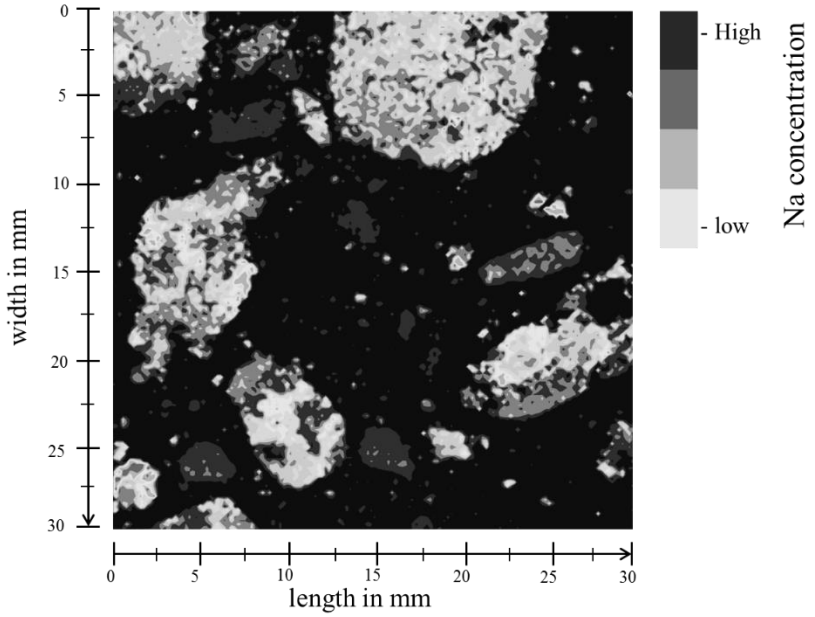
P-190

CLUSTER ALGORITHM FOR THE EVALUATION OF HETEROGENEOUS MATERIALS BY LIBS

C. Gottlieb [1], S. Grothe, G. Wilsch

[1]BAM Federal Institute for Materials Research and Testing, Unter den Eichen 87, 12205 Berlin, Germany
cassian.gottlieb@bam.de

The laser-induced breakdown spectroscopy (LIBS) is a fast method to provide multi-elemental analysis of any sample [1, 2]. At the Federal Institute for Materials Research and Testing (BAM) the LIBS technique is applied on building materials to measure ingress profiles of harmful species like sodium and chloride. These elements are triggering different damage processes in civil engineering like the alkali-silica reaction or chloride-induced corrosion. Due to a scanning procedure a two dimensional element distribution of a sample surface can be measured. Therefore a laser with a pulse energy of 3 mJ, a pulse length of 1.5 ns, a wavelength of 1064 ns and a repetition rate of 80 Hz has been used. In order to have an automated seperation method to evaluate heterogeneous materials, different cluster algorithm have been tested. Best results have been achieved with the Expectation-Maximization-Algorithm (EM-Algorithm). After clustering, different multivariate methods like PLS [3] and ICR [4] have been applied to provide quantitative element analysis.



References:

9. R. Noll, Laser-Induced Breakdown Spectroscopy – Fundamentals and Application, Springer-Verlag Berlin Heidelberg, 2012
10. S. Millar, G. Wilsch, T. Eichler, C. Gottlieb, H. Wiggenhauser, Laser Induced Breakdown Spectroscopy (LIBS) im Bauwesen - automatisierte Baustoffanalyse, Beton- und Stahlbetonbau, WILEY-VCH Verlag, 110, 501-510, 2015
11. W. Kessler, Multivariate Datenanalyse: für die Pharma, Bio- und Prozessanalytik, Wiley-VCH, 2007
12. A. Hyvärinen, E. Oja, E. Independent Component Analysis: Algorithms and Applications Neural Networks, 411-430, 2000

P-191

MULTIVARIATE ANALYSIS OF SPECTROSCOPIC SIGNALS FOR AFLATOXIN M1 DETERMINATION IN BOVINE MILK

M. Li Vigni[1,2], C. Durante[1], J.M. Amigo[3], M. Cocchi[2]

[1]ChemSTAMP srl, Via Giuseppe Campi, 183, 41125 Modena, Italy.
[2]Dept of Chemical and Geological Sciences, University of Modena and Reggio Emilia, Via Campi, 103, 41125 Modena, Italy
[3]Dept of Food Science, University of Copenhagen, Rolighedsvej, 30, DK-1958 Frederiksberg C, Denmark

email (only for the presenting author)

The contamination of food by mycotoxins produced by moulds (aflatoxins and ochratoxin) is of great importance to health and science, and has important effects on the population when news of requisitions of contaminated foodstuff are reported. For the health and safety of consumers there is a clear direct risk associated with these molecules (declared as probable carcinogens), due to the ubiquity of food contamination, which can be found in vegetables for direct use as food, or is distributed through the processes of transformation in fruit juices or other products, and, through animal feeding, can arrive to milk and related products. European guidelines provide very stringent minimum values; e.g. for aflatoxin M1 in milk the limit is 50 ng kg^{-1}. Therefore, the analytical techniques required for such determinations must ensure high sensitivity to the molecule, which is obtained at the expense of aspects such as the speed of sample preparation and analysis (determination by chromatography), which has been only partially tackled by the introduction of immunological methods such as ELISA.
In this work, a feasibility study for the development of an immediate and efficient analytical method for the determination of mycotoxins in milk (Aflatoxins, in particular M1) is proposed, by coupling spectroscopic and chemometric techniques. The target of the activity is to obtain a fast and simple screening method of the positivity of samples to be easily implemented by manufacturers, associations, consortia of milk collection and production for quicker and well-timed internal controls. A total of eighty milk samples, collected between September 2015 and January 2016, have been considered, all presenting different values of contamination by Aflatoxin M1: true negatives (the toxin is absent), below the limit (20-50 ng kg^{-1}, slightly above the limit (50-80 ng kg^{-1}) and showing severe contamination (up to 300 ng kg^{-1}). Two different spectroscopic approaches, each with a particular target, have been taken into account. Infrared Spectroscopy, namely Near Infrared Spectroscopy, is a technique which is already in use for product properties determination and quality control, and has been evaluated as a preliminary method to classify the sample according to the level of contamination (below the limit / close to the limit / above the limit / far above the limit), so that further analysis is performed only on milk samples which appear to be contaminated. Excitation-Emission Spectroscopy Fluorescence (EEM-F), due to the chemical nature of the toxin, appears suitable for quantification, even by analysing the sample “as-is” due to the possibility of resolving spectral contribution from different sources thanks to the bilinear advantage (PARAFAC) and by means of curve resolution tools (MCR).

Acknowledgement: the Authors would like to thank Dr. Nicoletta Rizzi, head of the Laboratory of ARAL (Associazione Regionale Allevatori della Lombardia, Via Kennedy, 30, 26013 Crema, Italy) for providing the samples and performing the reference analysis by ELISA and HPLC.

P-192

STANDARDIZATION PROCEDURE USING MATHEMATICALLY MIXED NIR SPECTRA AS TRANSFER SAMPLES FOR DETERMINATION OF FUEL QUALITY PARAMETERS

N.C. da Silva[1], C.J. Cavalcanti[2], F.A. Honorato[2], M.F. Pimentel[2], J.M. Amigo[3]

[1]Department of Fundamental Chemistry, Federal University of Pernambuco, Recife, Pernambuco, Brazil.
[2]Department of Chemical Engineering, Federal University of Pernambuco, Recife, Pernambuco, Brazil.
[3]Department of Food Science, University of Copenhagen, Denmark.
email (neirivaldocavalcante@gmail.com)

The interest in performing in field measures using portable instruments is growing increasingly [1]. Calibration transfer techniques enable the use of multivariate calibration models developed from benchtop to portable instruments, saving money and time. To perform standardization techniques such as Direct Standardization (DS), it is necessary to measure representative samples (transfer samples - TS) in the instruments involved. Nevertheless, volatilization of the compounds present in fuel samples may cause changes in their composition during transportation or over time. To avoid the storage of the TS, virtual standards (VS) can be created by mathematically mixing spectra from the pure solvents present in gasoline [2] or diesel/biodiesel (D/B) blends. These VS can then be used as transfer sets in the DS procedure. In this work, a reverse standardization (RS) technique was employed to transfer the calibration set obtained with a Frontier FT-NIR (PerkinElmer) master instrument (MI) to a MicroNIR™ 1700 (JDSU) slave instrument (SI). RS transforms the spectra from the MI to resemble those from the SI. Then, a new model is built with the corrected spectra which can be directly used in the SI. To perform RS, gasoline VS were created using ten and five pure solvents for gasoline and D/B blends, respectively. Partial least squares regression (PLS) models were built for five quality parameters of gasoline (distillation temperatures at 10%, 50%, 90% and final boiling point (FBP) volume recovered – ASTM D86 and density – ASTM D1298) and one of D/B blends (biodiesel content – EN14078). 103 and 130 gasoline and D/B blends samples, respectively, were collected in Pernambuco State (Brazil) fuel stations. Calibration and external validation sets were composed by 70% and 30% of the samples, respectively. The spectra were acquired with a 20mm path length cuvette. A labmade transmitance acessory was employed to acquire the spectra in the SI. From the pre-processing techniques tested, the best results were obtained with Standard Normal Variate . Gasoline models for the MI showed root mean square error of prediction (RMSEP) values of 0.404ºC (T10%), 0.412ºC (T50%), 1.82ºC (T90%), 2.60ºC (FBP), 0.887 $Kg.m^{-3}$ (density). RMSEP values for the SI were 0.563ºC (T10%), 0.815ºC (T50%), 4.15ºC (T90%), 3.91ºC (FBP), 2.39 $Kg.m^{-3}$ (density). Biodiesel content models showed RMSEP values of 0.45% for the MI and 1.27% for the SI. The new models, built after RS, were directly applied on the spectra of the external validation samples acquired in the SI. The RMSEP values obtained were 1.12ºC (T10%), 1.49ºC (T50%), 4.02ºC (T90%), 3.93ºC (FBP) and 3.73 $Kg.m^{-3}$ (density) for the gasoline models and 2.08% for the biodiesel content model. After standardization, RMSEP values were equivalent in reproducibility of the reference methods, except for density and biodiesel content. RS procedure provided promising results showing that it is possible to transfer gasoline or D/B blend spectra acquired with a high-resolution benchtop instrument to the portable MicroNIR using VS as transfer samples. A improved number-type of solvents can probably enhance the performance of the method for the biodiesel content model.

Acknowledgement: Capes for financial support, CNPQ, NUQAAPE and INCTAA.

References:

[1] E. M. Paiva, J. J. R. Rohwedder, C. Pasquini, M. F. Pimentel, C. F. Pereira, Fuel. v. 160, p. 57 (2015).
[2] J. B. Cooper; C. M. Larkin; M. F. Abdelkader, J. Chemometrics, v. 25, p. 496 (2011).

P-193

DATA FUSION OF HYPERSPECTRAL IMAGES FOR A BETTER COMPREHENSION OF BIOLOGICAL TISSUE ANALIZED BY MCR-ALS METHOD

Sara Piqueras[1], Marcel Maeder[3], C. Beleites[4], C.Krafft[4], J.Popp[4], C. Bedia[2], Romà Tauler[2], Anna de Juan[1]

[1]Chemometrics group. Universitat de Barcelona. Diagonal, 645. 08028 Barcelona, Spain. piqueras.sara@gmail.com
[2]IDAEA-CSIC. Jordi Girona, 18. Barcelona, Spain.
[3]Dept. Chemistry. The University of Newcastle. Newcastle (Australia)
[4]Leibniz Institute of Photonic Technology, Jena, Germany

Hyperspectral images can be acquired by different techniques, such as Raman, IR, fluorescence or mass spectrometry, among others. The combined information from different spectroscopic techniques is an excellent option to improve the capacity to differentiate among sample constituents. This aspect is highly beneficial for the understanding of biological tissue composition as well as for diagnostic purposes, since the spectral signatures associated with the biological constituents can be very similar [1].

Data fusion of different imaging techniques allows a comprehensive description of the composition of biological tissue, but joining images collected with different spectroscopic platforms is complex due their different sample orientation and spatial resolution. A matching procedure is necessary to ensure coherence among the fused images and, as a consequence, a correct interpretation of the system studied. Chemometric tools are required for reliable image matching and subsequent multitechnique image analysis [2].

The main goal of this work is the description of a full chemometric workflow starting by matching image spatial properties and finishing by performing multitechnique image analysis on a data structure that can cope with the different spatial resolution (pixel size) of the coupled techniques. The first step is carried out with a new preprocessing to perform the necessary translation/rotation transformations that use all pixels in the images to be matched. Multitechnique image analysis is afterwards performed with a new variant of MCR-ALS for incomplete multisets that allows analyzing simultaneously images collected with different spectroscopic techniques without losing spatial resolution and ensuring coherence among the images treated [3]. Different biological systems and coupled techniques will be presented as examples of data fusion for multitechnique analysis. The potential of the methodology proposed will allow a better comprehensive understanding of the composition of biological tissue as compared with the information obtained from the separate analysis of images coming from different hyperspectral platforms.

References:

[1] T.W.Bocklitz, A.C. Crecelius, C.Mattius, N.Tarcea, F.von Eggleling, M.Schmitt, U.S. Schubert and J.Popp. *Anal. Chem*. 85,10829 (2013).
[2] A. de Juan, S.Piqueras, M. Maeder, T. Hancewicz, L. Duponchel, R. Tauler. Chemometric tools for image analysis, in Infrared and Raman Spectroscopic Imaging (Editors: R. Salzer, H.W. Siesler), 2nd. Edition, Wiley-VCH, 57-110 (2014).
[3] M. Alier and R. Tauler. *Chemom. Intell. Lab. Sys.* 127 (2013) 17-28.

P-194

TIME OF FLIGHT SECONDARY ION MASS SPECTROMETRY AS A SCREENING TECHNIQUE - A MULTIVARIATE DATA ANALYSIS APPROACH

Danica Heller[1,2], Lothar Veith[1,2], Michael Fartmann[1], Rik ter Veen[1], Birgit Hagenhoff[1], Carsten Engelhard[2]

[1] Tascon GmbH, Mendelstr. 17, 48149 Münster, Germany
[2] Department of Chemistry and Biology, University of Siegen, Adolf-Reichwein-Str. 2, 57076 Siegen, Germany
danica.heller@tascon-gmbh.de

Time of Flight Secondary Ion Mass Spectrometry (TOF-SIMS) is a "workhorse" screening technique for samples of completely unknown composition. This is due to the fact that inorganic and organic molecules can be measured simultaneously and with high sensitivity. Modern TOF-SIMS instruments are highly automated and routinely offer the possibility of three-dimensional (3D) analyses. This promises quick and comprehensive analysis of complex samples. However, the interpretation of such data is often time consuming and complex. For example, a 3D analysis with 256 x 256 pixels running for 1000 scans (data points) results in 6.4 x 10^7 individual mass spectra.

To handle such complex data sets different Multivariate Data Analysis Techniques (MVA) were applied to various sample systems in the past.[1,2] MVA methods help to classify the data and thereby reduce the complexity of the data set and the time of data analysis. The application of MVA requires a pre-treatment of the data. This usually comprises a manual selection of peaks, which is time consuming and includes the possibility to overlook essential peaks with low intensity. Also, appropriate scaling, normalization, and centering are mandatory. Depending on the selected scaling approach, it is possible to enhance peaks of low or high intensity.

The aim of this study is to develop a suitable pre-treatment approach, which enables the application of MVA to sample systems with a completely unknown surface composition. Aged Li-ion battery electrodes containing unknown degradation products are analyzed as a test system. It was shown that Principal Component Analysis (PCA) facilitates the detection of small variations in the data set, which might be missed by manual data analysis. By this approach the *a priori* unknown degradation products of could be classified according to different additives and the chemical identification of the signals was simplified. The next step, an efficient way to resolve the layer structure of these sample types was put into praxis by the utilisation of Multivariate Curve Resolution (MCR). An additional PCA applied on the MCR results enabled the comparison of the different layer structures for all samples, simultaneously.

Acknowledgement: The samples were kindly provided by Björn Hoffmann, Timo Schwieters, and Sascha Nowak (MEET, Münster, Germany). This work is part of the project OptiLIB (grant number 310103702). It was supported through funding instrument "CheK.NRW" by the federal Government of NRW and the European Union (EFRE program).

References:

[1] M. S. Wagner, D. J. Graham, B. D. Ratner, and D. G. Castner, *Surf. Sci.*, 570, 78 (2004).
[2] D. J. Graham and D. G. Castner, *Biointerphases*, 7, 49 (2012).
[3] D. Heller, B. Hagenhoff, and C. Engelhard, *J. Vac. Sci. Technol. B*, submitted for publication.

P-195

MULTIVARIATE DATA ANALYSIS IN MASS SPECTROMETRY: A CASE STUDY ON THE IMPACT OF DATA PRE-TREATMENT

D. Heller [1,5], A. Albert[2], F. Ebert[3,4], L. Leffers[3], T. Schwerdtle[3,4], B. Hagenhoff[1], C. Engelhard[5]

[1]Tascon GmbH, Mendelstraße 17,48149 Münster, Germany. [2]University Münster, Institute of Inorganic and Analytical Chemistry, Corrensstraße 30, 48149 Münster, Germany [3]University Münster, Institute of Food Chemistry, Corrensstraße 45, 48149 Münster, Germany [4]University of Potsdam, Institute of Nutritional Science, Athur-Scheunert-Allee 114-116, 14558 Nuthetal, Germany [5]University of Siegen, Department of Chemistry and Biology, Adolf-Reichwein-Straße 2, 57076 Siegen, Germany

danica.heller@tascon-gmbh.de

In various applications in mass spectrometry multivariate data analysis (MVA) is typically used to simplify the analysis and reduce analysis time. In this study, we focus on data analysis for two different types of mass spectrometry, namely Secondary Ion Mass Spectrometry (TOF-SIMS) and Low Temperature Plasma Ambient Desorption Ionization Mass Spectrometry (LTP-ADI-MS). Data pre-treatment is a crucial step because it significantly influences the MVA results.[1–3] However, in many applications standard pre-treatments are often applied on mass spectrometric data sets. In this work, we demonstrate how a carefully optimised data pre-treatment can provide results of better quality with respect to the analytical problem. Specifically, three selected examples from different types of mass spectrometry are discussed.

In a first example, TOF-SIMS data of aged battery anodes with different additives are used. Data analysis focused on the best centering and normalization of this data to identify the influence of different additives on the surface of cycled battery samples. Here, a pre-treatment adapted to the analytical question and data set, respectively, yields better results than standard methods.

In order to further investigate ageing processes in batteries, all anode samples were subsequently measured in three different ageing states. It was found that the largest variation in the data set was induced by these ageing states. In order to identify the influence of different additives on battery anodes, a centering different than in the previous example was required. Here, data centered around the mean of each ageing state enabled to identifiy the individual influence of each additive.

In a last example, a LTP-MS data set from the volatile headspace of different cancer cells *in vitro* was analysed. The aim of this study was to classify different cell types and their media, respectively. Here, it was shown that although the magnitude of each signal in LTP-MS is in general not proportional to its impact, autoscaling was not the best scaling technique. This scaling method enhanced variations among different measurement days. Only after optimised scaling, a clear separation and classification of different cell types and their media was successfully achieved.

Acknowledgement: Battery samples were kindly provided by Björn Hoffmann, Timo Schwieters, and Sascha Nowak (MEET, Münster, Germany). Parts of this work were part of the project OptiLIB (grant number 310103702). It was supported through funding instrument "CheK.NRW" by the federal Government of NRW and the European Union (EFRE program).

References:

[1] R. Bro, A.K. Smilde, *J. Chemom.*, 17, 16 (2003).
[2] S.N. Deming, J.A. Palasota, J.M. Nocerino, *J. Chemom.*, 7, 393 (1993).
[3] M.S. Wagner, D.J. Graham, B.D. Ratner, D.G. Castner, *Surf. Sci.*, 570, 78 (2004).

P-196

DETERMINATION OF THE LIQUID PHASE COMPOSITION IN AMMONIA/WATER MIXTURE BY NIR SPECTROSCOPY AND SIMCA ANALYSIS

M. Isabel Barba[1], Daniel Salavera[1], M. Soledad Larrechi[2], Alberto Coronas[1]

[1]CREVER. Mechanical Engineering Dept.
[2]Analytical and Organic Chemistry Dept.
Universitat Rovira I Virgili, Tarragona, Spain
misabel.barba@urv.cat

Ammonia/water mixture is a conventional working fluid pair in absorption refrigeration. The composition of the mixture is an input data to study the absorption process and in the design the refrigeration systems. In a previous paper [1], the near infrared spectroscopy (NIR) was used to monitor quantitatively "in situ" the liquid phase composition in the vapor-liquid equilibrium of an ammonia/water mixture.

This study proposes a strategy based on Soft Independent Modeling of Class Analogy (SIMCA) [2] to model the representative spectra of the liquid phase at different temperatures and compositions, in order to classify them. Thus, obtaining the spectrum at a specific temperature, using this methodology the liquid composition can be estimated. For that, the NIR spectra of twelve mixtures, with ammonia mass fractions between 5 % and 60 %, were measured between 900 nm and 1150 nm from 298.15 K to 373.15 K, and SIMCA classification rules were established.

For each mixture, the percentages of ability classification and of prediction were higher than 98 % and 95 %, respectively. The results allow establish a strategy based on near infrared spectroscopy and the chemometrics analysis to determine the ammonia liquid composition of the ammonia/water mixtures in vapor-liquid equilibrium.

Acknowledgement: The authors are grateful to FEDER and the Spanish Ministry of Economy and Competitiveness (Project DPI2012-38841-C02-01) for financial support

References:

[1] M. I. Barba, D. Salavera, M.S. Larrechi, A. Coronas, *Quantitative "in-situ" monitoring of the ammonia concentration in the liquid phase at vapor-liquid equilibrium using near-infrared spectrosocopy. (*In preparation.

[2] R.D. Maesschalck, A. Candolfi, D.L. Massart, S. Heuerding, *Decision criteria for soft independent modelling of class analogy applied to near infrared data*, Chemometr. Intell. Lab. Syst. 47 (1999) 65-77

P-197

ANALYSIS OF BEER USING MID INFRARED SPECTROSCOPY AND CHEMOMETRICS

K. Wójcicki, F. Wawrzyniak, Sz. Drygas, E. Sikorska

Department of Technology and Instrumental Analysis, Faculty of Commodity Science, The Poznań University of Economics, al. Niepodległości 10, 61-875 Poznań, Poland
krzysztof.wojcicki@ue.poznan.pl

Quality control and monitoring of chemical composition of food products are very important tasks in food industry.Spectroscopic techniques coupled with multivariate data analysis provide an alternative to the traditional chemical methods, in food quality control. This approach extracts quantitative or qualitative analytical information from non-selective spectroscopic signals. In comparison to conventional methods, these measurements are rapid and non-destructive, performed directly on intact samples without extensive pre-treatment.

In our work mid infrared (MIR) spectroscopy was used for determination of the quality parameters of beer.

Beer is an alcoholic beverage obtained by yeast fermentation of cereals germinated in water. The main beer constituents are water and ethanol, however it also contains a complex mixture of compounds of varied chemical nature and concentration range, originating from both the raw materials and products of alcoholic fermentation. Due to the complex nature of beer a variety of analytical methods is used to study its properties. Recently, several authors reported the usage of vibrational spectroscopy for direct analysis of beer [1, 2].

In this study we used MIR spectroscopy for determination of the selected physicochemical properties of beer.

Beer spectra were measured over the entire MIR region using the attenuated total reflection (ATR) technique. The spectra of beers comprise a complex system of overlapping bands attributable to various vibrational transitions. That includes the intense bands originating from water, and bands characteristic of ethanol. Bands from the other beer components have much lower intensity.

Partial least squares regression (PLS) was used to study the relationship between selected physicochemical properties of beer (polyphenols, flavonoids, bitterness, and color) and MIR spectra. The set of independent variables *X* were the MIR spectra and the set of the dependent variables *Y* were values of the parameters studied. The various calibration models were studied using the entire spectra or their selected regions, and different pretreatment methods. The performance of the regression models was evaluated using the adjusted R^2 and the root mean-square error (RMSE), as the term indicating the prediction error of the model.

The results indicate that MIR spectroscopy combined with chemometrics is an interesting alternative to the traditional methods, for assessment the quality of beer.

References:

[1] J.A. Fernandez Pierna, L. Duponchel, C. Ruckebusch, D. Bertrand, V. Baeten, P. Dardenne, Chemometrics and Intelligent Laboratory Systems, 113, 2-9 (2012).
[2] D.W. Lachenmeier, Food Chemistry, 101(2), 825-832 (2007).

P-198

A robust soft classification approach for the detection of cancers in the lymph nodes of the head and neck

G.R. Lloyd [1*], R.G. Brereton [2], N. Stone [3], C. Kendall [1]

[1]Biophotonics Research Unit, Gloucestershire Hospitals NHS Foundation Trust, Gloucester GL1 3NN, UK
[2]School of Chemistry, University of Bristol, Cantocks Close, Bristol BS8 1TS, UK
[3] Biomedical Physics, School of Physics, University of Exeter, Exeter EX4 4QL, UK
**g.lloyd@medical-research-centre.com*

Accurate detection and identification of cancer of the lymph nodes is vitally important for improving patient outcomes. However, identification of lymph node cancers is currently very difficult in vivo, resulting in the unnecessary removal of benign nodes and unpleasant side effects. Raman spectroscopy has the potential to revolutionise the diagnosis of lymph node cancers by allowing in vivo assessment through the use of a Raman needle probe [1]. Furthermore, by measuring the biochemical fingerprints within tissue and in combination with chemometric classification algorithms a non-subjective assessment of the lymph node can be made. This can then be used to enhance the diagnostic decision making process and reduce the unnecessary removal of benign lymph nodes and associated side effects. However, training of a chemometric model using Raman spectra has to be carefully considered. Raman spectra can be very noisy and contain cosmic ray artefacts that need to be removed. Furthermore, highly fluorescent backgrounds e.g. from blood can mask useful biochemical signals. The nature of the samples also means that it is not possible to include all cell types in the training of the classifier. Any classification algorithm will therefore need to be able to account for this during prediction of unknowns. In this work we investigate the use of robust covariance matrix estimators to make the training of classifiers such as LDA less sensitive to low quality and/or outlying spectra. Furthermore we develop our algorithm using a soft-classifier [2] to allow for previously unseen tissue types. We demonstrate our approach using an existing study in which we used Raman spectroscopy to detect primary and secondary cancers in the lymph nodes of the head neck ex vivo [3].

References:

[1] J. Day, N. Stone, Appl. Spectrosc. 67 (2003) 349–354
[2] R.G. Brereton, G.R. Lloyd, J. Chemom. In Press (2016)
[3] G.R. Lloyd, et al., Analyst. 138 (2013) 3900–3908.

P-199

ACOMDIM study of iodine-paint interactions after gamma irradiation by Raman spectroscopy

E. Chauvet[2], N. Dupuy[1], S. Amat[1], C. Duffieux[2], J. Colombani[2]

[1] *Aix Marseille Université, LISA EA 4672, 13397, Marseille, France*
[2] *Institut de Radioprotection et de Sûreté Nucléaire, L2EC, CEN Cadarache, 13115 St-Paul-Lez-Durance, France*

nathalie.dupuy@univ-amu.fr

The effects of radiation on polymeric materials are a topic of concern in a wide range of industries including the sterilization [1], and the nuclear power industry [2]. While much work has concentrated on systems like polyolefins that are radiation sterilized [1], some work has been done on epoxy systems [3]. The epoxy system studied is an epoxy/amine paint which is representative of the paint that recovers the inner surfaces of the French reactor containment buildings. In case of a severe accident in a nuclear power plant, a significant amount of fission products can be released from the nuclear fuel to the reactor containment building. Among these species, volatile iodine (I_2) can be produced and can interact with the epoxy/amine paint. This paint is also subjected to gamma radiation damages (due to the radionuclides released from the fuel). The epoxy/amine paint studied was exposed to gamma radiations under air after being loaded with I_2 or not. Aging of material submitted to γ-irradiation can be monitored by spectroscopic methods such as Raman spectroscopy.
The aim of this study is to characterize the iodine-paint interactions, then to identify the radiation damages on the epoxy-amine paint, and to check their effects on the iodine-paint interactions. Some recent works show the potential of multiblock methods as ANOVA-PCA or COMDIM for such studies [4].
Factors studied are thus the integrated radiation dose and the chemical loading by I_2. Spectral differences caused by several aging processes are studied. AComDim conduces to the extraction of Common Components (CC) to different tables and highlights factors of influence and their interactions.

Keywords: gamma irradiation, iodine, paint, ACOMDIM, Raman

[1]. H. J. Nusbaum, R. M. Rose. 13.4 557-76, (1979)
[2]. Clough, R. (1988). Radiation-resistant polymers. Wiley-Interscience, Encyclopedia of Polymer Science and Engineering., 13, 667-708
[3]. Clough, Roger L., Shalaby W. Shalaby. Radiation Effects on Polymers: Developed from a Symposium Sponsored by the Division of Polymer Chemistry, Inc. at the 200th National Meeting of the American Chemical Society, Washington, D.C., August 26-31, 1990. Washington, D.C.: American Chemical Society, 1991.
[4]. R. Korifi, S. Amat, C. Rébufa, V. Labed, D.N. Rutledge, N. Dupuy, Chemom. Intel Lab Sys 141, 12–23 (**2015)**

P-200

EFFECT OF CHLORPYRIFOS-OXON ON ZEBRAFISH BY HYPERSPECTRAL IMAGING AND CHEMOMETRIC METHODS

V. Olmos [1], M. Marro [3], B. Piña [2], P. Loza-Alvarez [3], R. Tauler [2], A. de Juan [1]

[1]Department of Analytical Chemistry, University of Barcelona, Diagonal 645, Barcelona, Spain
[2]Department of Environmental Chemistry, Institute of Environmental Assessment and Water Diagnostic (IDAEA-CSIC), Jordi Girona 18, Barcelona, Spain
[3]Institute of Fotonic Sciences (ICFO), Carl Friedrich Gauss 3, Castelldefels (Barcelona), Spain
Email: victor_olmos@ub.edu

The characterization of the changes produced on an organism due to the exposure to an environmental stress are usually performed by destructive techniques (e.g. HPLC-MS, immunoassays...) [1]. Hyperspectral imaging (HSI) techniques provide spatial and chemical information and preserve the natural morphology of the samples. These properties may allow HSI to be a potentially useful methodology in this kind of studies. HSI methods have been proposed in several studies to differentiate tissues, e.g. healthy and tumor or inflamed tissues [2]. HSI analysis requires the use of multivariate chemometric techniques in order to handle and interpret the large quantities of data obtained from the raw spectral signals. Multivariate curve resolution-alternating least squares (MCR-ALS) is often used in hyperspectral image analysis because it can provide the pure spectral signatures and distribution maps of compounds in a sample from the raw image measurement and it does not need previous information [3].
In this work, several Raman HSI have been acquired on zebrafish embryo tissue cryosections in order to assess the effect of chlorpyrifos-oxon (CPO) (a metabolite of the pesticide chlorpyrifos) on zebrafish eye tissues. Zebrafish embryos are genetically identical and have been obtained by natural breeding. Some of the embryos have been exposed to CPO during 24h and the rest have been used as control samples.
HSI from either control or contaminated samples have been analyzed together in two separate column-wise augmented multiset structures related to each sample population by MCR-ALS. Pure spectral signatures and distribution maps obtained for each population multiset (control and contaminated) have been visually compared for a qualitative interpretation of the effect of the contaminant.
On the other hand, a new and sounder statistical approach to perform partial least squares-discriminant analysis (PLS-DA) on HSI data is performed. This new approach consists of a pixel resampling of the images that form each class multiset, control or contaminated, to obtain many small multisets that contain representative pixels from all the images. MCR-ALS is performed on all these small multisets. The matrix submitted to PLS-DA is formed by the sets of pure spectral signatures resolved for each component tissue in the resampling step for each one of the classes. In this way, there are as many PLS-DA models as analogous components resolved in both the control and exposed sample populations. The classification parameters achieved can be used to assess the significance of CPO exposure on the different zebrafish eye tissues and the variable importance in projection (VIPs) is the chosen criterion to indicate the most important Raman bands related to this effect.

References:

[1] C. Bedia, N. Dalmau, J. Jaumot, R. Tauler, Environ. Res. , vol 140, 18 (2015)
[2] S.V. Panasyuk, *et al*., Cancer Biol. Ther., vol 6(3), 439 (2007).
[3] A. de Juan, S.Piqueras, M. Maeder, T. Hancewicz, L. Duponchel, R. Tauler. Chemometric tools for image analysis, in Infrared and Raman Spectroscopic Imaging (Editors: R. Salzer, H.W. Siesler), 2nd. Edition, Wiley-VCH, 57-110 (2014).

P-201

MCR-ALS and BTEM: comparative evaluation and potential integration points

C.G. Bertinetto [1], A. De Juan [2]

[1]Department of Forest Product Technology, School of Chemical Technology, Aalto University
Vuorimiehentie 1, 02150 Espoo, Finland.
[2]Chemometrics Group, Department of Analytical Chemistry, Universitat de Barcelona
Diagonal, 647, 08028 Barcelona, Spain
email: carlo.bertinetto@aalto.fi

This study aims at assessing the relative performance and range of applicability of two linear methods for spectral unmixing and curve resolution: Multivariate Curve Resolution – Alternating Least Squares (MCR-ALS) and Band-Target Entropy Minimization (BTEM). MCR-ALS [1,2] performs a factorization of the data matrix into spectral and concentration profiles that satisfy constraints expressing the physico-chemical knowledge on the analyzed system. BTEM [3] reconstructs the pure components' spectral profiles as linear combinations of loadings that minimize the spectral entropy and contain specific peaks. Both methods have been applied to several different analytical problems, but to date there has been no systematic investigation on understanding the suitability of both methods for the different data typologies and problems of interest.

In the present work, this question is addressed by applying the two mentioned methods to data sets that include diverse instrumental signals (IR, Raman, UV-Vis, MS, etc.) and experimental problems (images, chromatograms, reaction monitoring, etc.). In addition to comparing results obtained from existing algorithms, this study explores the possibility of incorporating aspects of BTEM into the MCR algorithm, in the scope of developing an approach that combines advantages of both methods.

Acknowledgement: The Walter Ahlström foundation is gratefully acknowledged for the financial support.

References:

[1] A. de Juan, J. Jaumot, R. Tauler, Anal. Methods, 6, 4964 (2014).
[2] J. Jaumot, A. de Juan, R. Tauler, Chemom. Intell. Lab. Sys., 140, 1 (2015).
[3] E. Widjaja, C. Li, W. Chew, M. Garland, Anal. Chem., 75, 4499 (2003).

P-202

RAMAN IMAGING AND MCR-ALS PROBING OF POSSIBLE INTRACELLULAR TARGETS FOR ORGANOMETALLIC ANTITUMORALS

M.B. Mamián-López, R. Bernardi M, M.F.G. Huila, K. Araki, A.M. da Costa Ferreira, M.L. A. Temperini

Instituto de Quimica, Universidade de São Paulo, PO Box 26.077, 05513-970, São Paulo/SP, Brazil
monikml@gmail.com

Multivariate curve resolution with alternating least-squares (MCR-ALS) [1] aiming to decompose Raman spectra in order to identify possible intracellular targets for [Cu(ISA-EPY)(H_2O)], a complex with antitumoral activity. Confocal Raman hyperspectral images were acquired from living HeLa cells, following treatment with the complex and without, as control. Spectra were obtained in a WITec alpha-300R equipment using frequency-doubled Nd:YAG (λ= 532 nm) and an oil immersion objective (100x, N.A= 1.25). The area was scanned in the *xy* direction with a piezo-driven *xyz* feedback controlled scan stage.

Data were pre-processed with Weighted Least-Squares (WLS) and smoothing (5 pt window) procedures. MCR-ALS method was applied using the MCR-ALS toolbox V. 2.0 on Matlab R 2013a. The number of constituents was estimated using singular value decomposition (SVD) and its respective profiles were calculated by SIMPLe-to-use Interactive Self-modeling Mixture Analysis (SIMPLISMA). Three factors (F) were found for control cells in all cases (Fig1, left), with remarkable differences around 3000 cm^{-1} where C-H stretching bands are expected [2]. Images showed F1 mainly describing the nucleus, F2 corresponding to lipid- or phospholipid-rich regions in the cytoplasm and F3 showing the aqueous medium distribution, while for treated cells (Fig 1, right), only two factors were calculated. An interesting feature observed is the clear absence of the nucleus and the fact that corpuscles in the cytoplasm are apparently not affected by the drug. Further calculations using an initial estimate that included the nucleus profile, allowed to detect its residuals indicating drug-induced cellular death (Image not shown). Raman imaging along with MCR-ALS have given insight to possible targets where the antitumoral is causing cell damage and allows to complement studies focused on elucidating its mechanism of action inside cell systems.

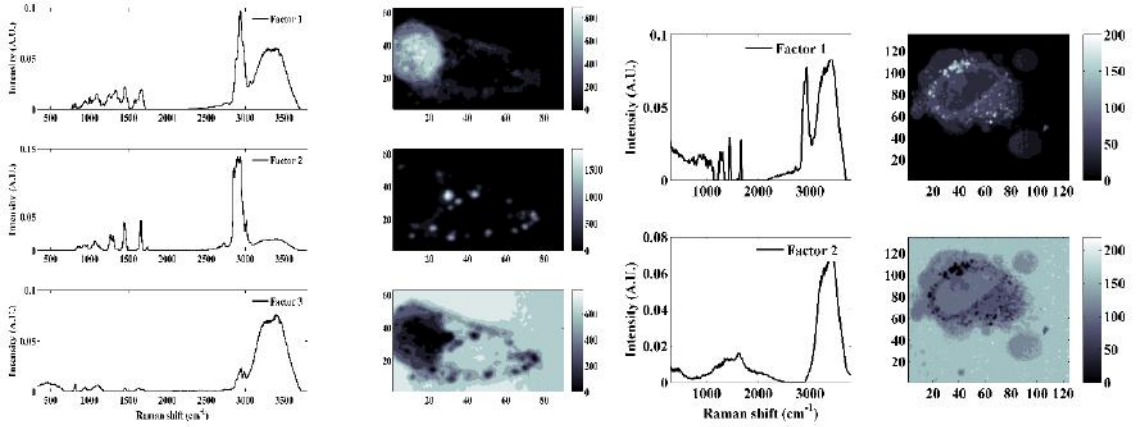


Figure 1. Raman images and calculated factors from MCR-ALS of HeLa cells without treatment (left) and after treatment with [Cu(ISA-EPY)(H_2O)] (right).

Acknowledgement: The authors acknowledge FAPESP (Process number 2014/07841-3 and 2012/13119-3), CNPq and CAPES for fellowships and financial support.

References:

[1] J. Jaumot, A. De Juan, R. Tauler. Chemometrics and Intelligent Laboratory Systems, 140, 1 (2015).
[2] T. Dieing, O. Hollricher, J. Toporski. Confocal Raman Microscopy; Springer: Heidelberg, Germany, 2011.

P-203

COMBINATION OF MALDI IMAGING MASS SPECTROMETRY WITH CHEMOMETRIC TOOLS FOR INVESTIGATION OF TUMOR HETEROGENEITY

S. Mas[1,2], A. Torro [2], N. Bec[2], G. Erschov[2], C. Larroque[2], A. de Juan[3], P. Martineau[2]

[1]Equipe Sciences Analytiques & Modélisation Moléculaire of Institut des Biomolécules Max Mousseron (IBMM). UMR5247, Université Montpellier et CNRS, 15 avenue Charles Flahault, 34093 Montpellier, France.

[2]IRCM, Institut de Recherche en Cancérologie de Montpellier, INSERM U1194, Université Montpellier, 208 Avenue des Apothicaires, Montpellier, F-34298, France.

[3]Chemometrics Group. Department of Analytical Chemistry. Universitat de Barcelona. Av. Diagonal, 647. 08028 Barcelona, Catalonia, Spain

silviamasgarc@hotmail.com

Colorectal cancer is a major worldwide cause of morbidity and mortality. It is the third most common cancer and the fourth most common cause of cancer death throughout the world. It affects men and women almost equally. Neoadjuvant treatment is the standard of care for colorectal cancer. The use of preoperative chemotherapy combined with surgery of the primary tumor and/or liver metastasis has been shown to improve the 5-year survival rates of up to 50% and even complete recovery in some patients. In the absence of the surgery, 100 % of patients will relapse and become resistant to chemotherapy.

Nowadays, MALDI Imaging mass spectrometry is frequently used in tissue-based research [1]. The great advantage of MALDI Imaging is the correlation of molecular information with traditional histology by keeping the spatial localization information of the analytes during mass spectrometric measurement. The combination of mass spectrometry profiles with spatial information helps to investigate and understand molecular processes happening in specific cells and tissues. MALDI Imaging mass spectrometry can generate MS profiles containing hundreds of distinct biomolecular ions. In order to achieve the distribution of these biomecular ions and demonstrate the significant spatial heterogeneity that can be present in the biomecular ion distribution the use of chemometrics tools is necessary.

The aim of this study is to demonstrate that the combination of MALDI Imaging mass spectrometry with chemometric tools allows investigation of tumor heterogeneity, identifying molecular phenotypes that can drive tumor progression, and correlate these data with each patient's progression. With this purpose and as preliminary study, human colon adenocarcinoma cell lines sensitive HCT116 (S) and resistant HCT116-SN50 (R) to iriontecan, were used as a model of experimental tumors. Clonogenic tumor xenographs were generated by subcutaneous injection of both unique cell line on athymic mice whereas a model of heterogeneity was created injecting various mixture of (R) and (S) HCT116 clones. Tumors were then collected, sliced, and analyzed by MALDI Imaging mass spectrometry. Factor analysis methods such as principal component analysis (PCA) and multivariate curve resolution-alternating least squares (MCR-ALS) [2,3] were used in order to recover the molecular distribution patterns within these experimental tumors. Potential identification of R and S populations without any molecular knowledge in advance will be tested.

References:

[1] J. Bianga, A. Bouslimani, N. Bec, F. Quenet, S. Mounicou, J. Szpunar, C. Larroque, C. Complementarity of MALDI and LA ICP mass spectrometry for platinum anticancer imaging in human tumor. Metallomics, 6(2014), 1382-1386.

[2] A. de Juan, S. Piqueras, M. Maeder, T. Hancewicz, L. Duponchel and R. Tauler. Chemometric tools for image analysis in Infrared and Raman spectroscopic imaging (2nd ed. R. Salzer and H.W. siesler, eds.), pp. 57-110.

[3] J. Jaumot, R. Tauler. Potential use of multivariate curve resolution for the analysis of mass spectrometry images. Analyst. 140 (2015) 837-46.

P-204

Using PLS-DA to evalutate the quality of spore inoculum to optimize biotechnological process control

K. Wieland [1], C. Koch [1], J. Kuligowski[2], J. Ofner[1], D. Ehgartner[3], C. Herwig[3], B. Lendl[1]

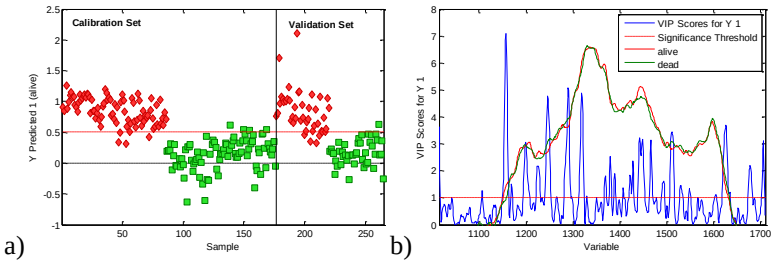
[1]Institute of Chemial Technologies and Analtyics, TU Wien, Getreidemarkt 9/164 UPA, A-1060 Vienna, Austria
[2]Neonatal Research Unit, Health Research Institute Hospital La Fe, Avda Fernando Abril Martorell 106, 46026 Valencia, Spain
[3]Institute of Chemical Engineering, TU Wien, Christian Doppler Laboratory on mechanistic and physiological methods for improved bioprocesses, Gumpendorferstraße 1a, A-1060 Vienna, Austria

Improved understanding of the biochemical processes that occur in a bioreactor is of great importance for process optimization. The chemical composition of microorganisms can be linked to their morphological and physiological state. This information is crucial for operators to optimize product yield and reduce production costs escpecially in downstream processing which usually is the most expensive part in a pharmaceutical production line. An improved understanding of the biochemical processes in a bioreactor may lead to prevention or reduction of undesired by-products. Also, bioreactors can be operated more efficiently at an optimal work load.

The filamentous fungus *Penicillium chrysogenum* is famous for producing the β-lactam antibiotic penicillin [1]. In order to improve process control at a very early stage the quality of spore inoculum is of great importance. Knowledge about the viability and germination ability of spores is crucial for adapting C-source rate, stirrer velocity, oxygen rate,.. and other process parameters needed for optimal growth and production condition and regulation.

Using Raman spectroscopy as label-free, non-invasive analytical tool for accessing the necessary information to differentiate living from dead spores, chemometric methods such as clustering and PCA were applied to see if the wanted information could be extracted from the vibrational fingerprint region of the investigated spore sample. In a next step, a PLS-DA model was established based on a calibration set of approx. 170 single spectra (each spectrum representing one single spore). For the validation set consisting of 90 single spectra 5 misclassifications (4 living and 1 dead spore) were detected (Fig. 1). Applying the model to another data set shows 98% correct class assignment.

Based on this model, an automated sceening method can be developed paving the way towards label-free, non-invasive in-situ investigation of the spore inoculum's quality.



a) b)

Fig. 1 a) Calibration and validation set of PLS-DA with 3 latent variables showing 5 misclassified spores in the validation set (red: living spores, green: dead spores); b) VIP scores indicating importance of spectral features

Acknowledgement: Financial support was provided by the Christian Doppler Gesellschaft and the Austrian research funding association (FFG) under the scope of the COMET program within the research project „Industrial Methods for Process Analytical Chemistry – From Measurement Technologies to Information Systems (imPACts)" (contract # 843546).

References:

[1] T. R. van der Lende; P. Breeuwer, T. Abee, W. N. Konings, A. J. M. Driessen, *Biochimica et Biophysica Acta*, 2002, **1598**, 104-111

P-205

CHEMOMETRIC METHODS TO ENHANCE SPECTRA QUALITY AND EVALUATE DATA OBTAINED BY A NOVEL LASER-BASED IR TRANSMISSION SETUP FOR PROTEIN ANALYSIS

M.R. Alcaráz[1,2], A. Schwaighofer[1], H.C. Goicoechea[2], B. Lendl[1]

[1] *Institute of Chemical Technologies and Analytics, Vienna University of Technology, Getreidemarkt 9/164-UPA, 1060 Vienna, Austria.*

[2] *Laboratorio de Desarrollo Analítico y Quimiometría, Facultad de Bioquímica y Ciencias Biológicas, Universidad Nacional del Litoral, Ciudad Universitaria, 3000 Santa Fe, Argentina*

alcarazmirtaraquel@gmail.com

Chemometrics has been proved to be an excellent complementary tool in several research fields. In combination with advanced instrumentation technology, it could significantly improve the performance of analytical methods. In the analysis of protein secondary structure by infrared spectroscopy, the limits of the method are defined by the high absorption of the solvent (i.e. water) as well as the low emission powers of the thermal light sources commonly employed in conventional FTIR spectrometers. Recently, a novel IR transmission setup has been developed, employing an external cavity-quantum cascade laser (EC-QCL) as light source that provides spectral power densities several orders of magnitude higher than thermal sources. In spite of being commercially available, these new light sources still suffer from imperfections in the tuning mechanism and shifts in the mode-hope fine structure of the emission curve within consecutive scans. To overcome this constant and variable offset between scans, correlation optimized warping (COW) was applied. This well-known algorithm, previously employed for spectra and chromatographic peak alignment, is here used to eliminate high noise levels in absorbance spectra obtained by QCL-IR spectroscopy [1].

Furthermore, to showcase the potential and quality of the IR absorbance spectra obtained by QCL-IR spectroscopy, dynamic changes of proteins secondary structure in aqueous solution were studied at varying pH values and across a wide concentration range. To this end, extended-MCR-ALS was employed. Pure spectral and concentration profiles of the temporal transition between different protein conformational structures were obtained [2].

We successfully demonstrate the application of COW to significantly reduce the noise level in IR spectra acquired with EC-QCL, thus getting this novel type of light source ready for routine measurements. The high quality of the obtained spectra has allowed to build chemometric models for describing dynamic protein secondary structure changes at varying experimental conditions.

Acknowledgement: The authors acknowledge the Austrian Research Promotion Agency (FFG) within the K-project imPACts (contract no. 843546) for funding. M.R.A. gratefully acknowledges the financial support provided by CONICET.

References:

[1] M.R. Alcaráz, A. Schwaighofer, C. Kristament, G. Ramer, M. Brandstetter, H. Goicoechea, B. Lendl, Anal Chem, 87, 6980-6987 (2015).

[2] M.R. Alcaráz, A. Schwaighofer, H. Goicoechea, B. Lendl, Anal Bioanal Chem. doi:10.1007/s00216-016-9464-5 (2016).

P-206

CHEMICAL SPECIATION DETERMINED BY MCR-ALS FITTING OF QUICK-X-RAY ABSORPTION SPECTROSCOPY DATA

O. Roudenko [1], C. La Fontaine [1], S. Belin [1], V. Briois [1]

[1]Synchrotron SOLEIL, ROCK beamline, L'Orme des Merisiers, BP48, 91912 Gif sur Yvette, France.
roudenko@synchrotron-soleil.fr

With the advent of 3rd generation synchrotron radiation facilities allowing a sub-second time resolution in the monitoring of kinetics by X-ray Absorption Spectroscopy (XAS), experimentalists can now access a deeper and more accurate temporal description of the chemical species involved in such processes provided that the data analysis bottleneck has been overcome. Time-resolved XAS beamlines users indeed have to face a huge amount of data in a couple of minutes, making the common strategy of XAS data evaluation, reduction and analysis quite inefficient. New tools must be proposed to users in order to interactively optimize the outcome of an experiment carried out at the beamline. In this context, chemometric tools combining Principal Component Analysis (PCA) and Multivariate Curve Resolution optimized with Alternating Least Square (MCR-ALS) fitting are emerging as a powerful method to quantify and extract spectra of intermediate species from XAS spectra of evolving mixtures upon reaction [1-6]. By accessing pure spectra, the gain of such chemometric tools can be sometimes unique insofar as those spectra are related to chemical species which are often different from the known bulk phases, for which XAS spectra are reported in the literature. This is particularly true in the field of heterogeneous catalysis involving nanometric phases well dispersed on a support. Thanks to the local order structure extracted from further analysis of the XAS data of the MCR-ALS components, the identification of these particular nanometric species is highly facilitated. At the ROCK beamline of SOLEIL synchrotron, a user-friendly interface has been recently developed to handle large amounts of data collected upon reaction monitoring. Normalized matrices of several hundreds of absorption spectra are quickly generated after data recording, and are then directly suitable as input files for MCR-ALS analysis available on the Matlab platform using free toolboxes developed by Tauler's group [7]. A preliminary analysis can thus be performed during the beamtime, allowing an optimization of the following experiments in light of the first results.

In this poster, a description of the tools for Quick-XAS pre-data normalization treatment used at the ROCK beamline will be presented. The importance of robust energy data alignment procedures for the output of PCA analysis will be discussed. The critical choice of methods for guessing the input of minimization loops will be examined on selected Quick-XAS data sets acquired upon monitoring the activation of heterogeneous catalysts.

Acknowledgement: The authors are grateful to Ludovic Duponchel (LASIR, Lille) for fruitful discussions on the use of MCR-ALS methodology

References:

[1] W. Cassinelli, L. Martins, A. R. Passos, S. H. Pulcinelli, C. V. Santilli, A. Rochet, V. Briois, *Cat. Today* **229**, 114 (2014).
[2] J. Hong, E. Marceau, A. Khodakov, L. Gaberova, A. Griboval-Constant, Jean-Sebastien Girardon, C. La Fontaine, V. Briois, *ACS Catal.* **5**, 1273 (2014).
[3] H.W. P. Carvalho, S. H. Pulcinelli, C. V. Santilli, F. Leroux, F. Meneau, V. Briois, *Chem. Mater.* **25**, 2855 (2013).
[4] A. Rochet, B. Baubet, V. Moizan, C. Pichon, V. Briois, *Comptes Rendus Chimie* (2016) http://dx.doi.org/10.1016/j.crci.2016.01.009.
[5] A. Voronov, A. Urakawa, W. van Beek, N. E. Tsakoumis, H. Emercih, M. Ronning, *Analytica Chimica Acta*, **840**, 20, (2014).
[6] M. Staniuk, O. Hirsch, N. Kränzlin, R. Bölhen, W. van Beek, P. M. Abdala, D. Koziej, *Chem. Mater.* **26**, 2086 (20134).
[7] J. Jaumot, A. de Juan, R ; Tauler, *Chemom. Intell. Lab. Syst.***140**, 1, (2015).

P-207

FEASIBILITY OF NEAR INFRARED HYPERSPECTRAL IMAGING AND RAMAN HYPERSPECTRAL IMAGING FOR ADULTERATION DETECTION IN FOOD INGREDIENTS

E.M. Achata, C. Esquerre, A.A. Gowen, C.P. O'Donnell

UCD School of Biosystems and Food Engineering, University College Dublin, Dublin 4, Ireland
eva.m.achatagonzales@ucdconnect.ie

Adulteration and mislabelling, whether economically-motivated or accidental, are reported to be widespread practices in the food industry. Manufacturers rely on certificates of analysis provided by suppliers and may also perform additional analytical tests, which are expensive, time consuming and require sample preparation. Near Infrared – Hyperspectral Imaging (NIR – HSI) and Raman Hyperspectral Imaging (Raman HSI), emerging analytical techniques for non-destructive fast analysis, were investigated in combination with chemometric techniques, (Principal Component Analysis (PCA), Partial Least Squares Regression (PLS) and Multivariate Curve Resolution (MCR)), to assess their potential to detect adulteration of food ingredients [1,2]. Binary mixtures of Corn Flour – Icing Sugar (CF – IS), Rice Flour (RF) – Wheat Flour (WF) (RF-WF) were obtained by mixing samples at different concentrations (0, 10, 20, 30, 40, 50, 60, 70, 80, 90 and 100 % (w/w)). NIR – HSI images were acquired in the wavelength range of 880 to 1720 nm at 7 nm intervals using a line scanning NIR HSI system. Raman HSI images were acquired at the Raman shift range of 106.94 to 2369.11 cm^{-1} using a Raman imaging system with 785 nm line laser as excitation source. PCA was employed to analyse the variability in spectral data, while PLS and MCR models were developed to predict concentrations of the components present. This work demonstrates the potential of both NIR-HSI and Raman HSI as suitable techniques for adulteration detection within the studied contaminant and concentration range.

Acknowledgement: The authors acknowledge funding for this project from FIRM as administered by the Irish Department of Agriculture, Food and the Marine.

References:

[1] A.A. Gowen, C.P. O'Donnell, P.J. Cullen, G. Downey, J.M. Frias, Hyperspectral imaging—an emerging process analytical tool for food quality and safety control. Trends Food Sci. Technol. 2007, 18, 590–598.

[2] J. Jaumot, J. C. Menezes, R. Tauler, Quality assessment of the results obtained by multivariate curve resolution analysis of multiple runs of gasoline blending processes, J. Chemometrics 2006; 20: 54–67.

P-208

Comparison of processing techniques for FT-IR images of tissue sections

M. Hermes[1], J. Nallala[1], N. Stone[1]

[1]College of Engineering, Mathematics and Physical Sciences, University of Exeter, EX4 4QL, United Kingdom
m.hermes@exeter.ca.uk

Fourier transform infra-red (FT-IR) imaging is supposed to become a versatile tool for the investigation of biological materials [1]. However handling huge data sets provided by state of the art FT-IR imaging systems is still an area of ongoing research.

The aim of this poster is to provide a hands-on practical comparison of different explorative approaches currently used by spectroscopists to access information from IR data. In order to do so a high quality hyperspectral image of a tissue section is processed with different approaches and the outcome of all of those techniques is compared. Furthermore the influence of the assumptions made in different algorithms will be evaluated.

A major challenge for handling IR data sets is their mere size, though cloud and out-of memory computing techniques exist, as well as the use of super computers is possible, this is not feasible to use for everyday applications. Therefore, the memory requirements of a workflow are an issue which needs to be considered as well as computation time in order to be applicable for biomedical imaging.

Techniques which will be evaluated for these requirements are principal component analysis (PCA), the k-means algorithm, the N-FINDR algorithm [2] as well as multivariate curve resolution alternating least squares (MCR-ALS) based techniques [3].

Acknowledgement: Funding within the scope of Horizon 2020 by the European Union is highly appreciated. This work was conducted as part of the Mid-TECH Marie Curie innovative training network [H2020-MSCA-ITN-2014- 642661].

References:
[1] Baker *et al.*, Nature Protocols, 9, 1771 (2014).
[2] M. A. B. Hedegaard, M. S. Bergholt and M. M. Stevens, J. Biophoton, n/a, 2016
[3] J. Felten, H. Hall, J. Jaumot, R. Tauler, A. de Juan and A. Gorzsas, Nature Protocols, 10, 217 (2015)

P-209

Exploratory Data Analysis Combined with Entropy-based Template Selection for the Identification of Mixed Substances

V. Mendiola-Lau [1], F.J. Silva-Mata [1], Y. Martínez-Díaz [1], I. Talavera Bustamante [1] and M. De Marsico [2]

[1]Advanced Technologies Application Center (CENATAV), 7th A Avenue #21406 % 214 and 216, Siboney, Playa, P.C. 12200, Havana, Cuba.
[2]Sapienza University of Rome, Piazzale Aldo Moro, 5, 00185, Rome, Italy.
vmendiola@cenatav.co.cu

This paper proposes a new methodology for the identification of mixed substances, which combines the results of a computational analysis of chemical and biochemical data obtained from three analytical techniques: Gas Chromatography (GC), Ultraviolet Spectrometry (UV) and Thin Layer Chromatography (TLC). With this methodology, the identity of mixed substances is determined with greater efficiency and effectiveness due to a template selection procedure. First, the output data extracted from these analytical techniques is properly normalized, aligned and transformed with the goal of obtaining comparable spectra for each technique. Then, an exploratory data analysis procedure is applied for discovering groups or classes of mixed substances that provide evidence concerning their families; whether it is according to their common industrial or craft processing, areas of crops in cases of plant extracts, date of production, etc. Finally, a refinement step is proposed based on the entropy of each group of samples aiming to remove the less significant samples from each group, which allows to improve the quality of the training set for further classification. The experiments conducted show the validity of the proposal, showing an improvement of the classification results for each analytical technique once the entropy-based template selection is applied.

KeyWords: Exploratory data analysis, entropy-based template selection, mixed substance identification.

P-210

Using a neural network and concentrations of calcium and oxalates to predict crystalluria type

A. Ait Ider[1], C.Tcheka[2], A. Ben Ali[3], M. Maaouni [1]A. Merbouha[1] and M. Mbarki[1]

[1] University of Sultan Moulay Slimane, Faculty of Science and Technology, P.B 523, BeniMellal, Morocco
[2] University of Yaoundé 1, Faculty of Science, P.B 812, Yaounde, Cameroun.
[3] National Control Laboratory Medicines,Ministry of Health, Madinat Al Irfane, P.B 6206, Rabat, Morocco.
Email: a.aitider@usms.ma

The majority of the analyzed calculi from patients are composed of calcium oxalate (CaOx) monohydrate whewellite (Wh) and CaOx dihydrate wedellite (Wd). The urinary calculi were identified by chemical and morphological analysis based on106 urine samples from human voluntary. The Crystalluria made by an optical polarized light microscopy. The oxaluria and urinary calcium were determined by conventional volumetric assays. The aim of this paper was to develop a simple system to predict and classify the type of crystalluria using Artificial Neural Networks (ANNs) algorithm.

Key words: calcium oxalate; urinary calculi; Artificial Neural Networks.

P-211

Exploiting X-ray diffractometry and optical polarized light microscopy for the diagnosis of urolithiasis

M. Maâouni[1], A. Ait ider[1], S. Rabi[1], M. Mbarki[1]

[1] Transdisciplinary team of Analytical Sciences for Sustainable Development, Department of Chemistry and Environment, Faculty of Science and Technology, University Sultan Moulay Slimane, 23000 Beni Mellal, Morocco; Email: maaounifstbm@gmail.com

Urolithiasis is a common disease that affects from 4 to 20% of the population in different countries. This pathology requires a lot of multidisciplinary research to understand its etiology and its possible etiopathogenic mechanisms responsible for the formation of urinary stones.
This work is on the research of the etiological parameters that contributes to the urinary stones development by chemical analysis, physical technics and using chemometric methods.
30 urinary stones collected just after their surgical removal, from both sexes. At first, the stones have been cleaned, dried, ground and were the subject of a morpho-constitutional analysis. A powder X-ray diffraction, Fourier Transform Infrared Spectrophotometry and the Stereomicroscopy-Optical Polarizing Light Microscopy have been used.
Physical and statistical methods are not very used to diagnose urolithiasis while they can give, with only a small investment and minimal time, spectacularly reliable results when used together and rationally.

Keywords : X-ray diffraction, Optical Polarizing Light Microscopy, chemometrics.

P-212

Repetition Rate Priority Combination Partial Least Squares with Application to Near-infrared Spectroscopic Analysis

Lijun Yao, Jiemei Chen, Tao Pan*

Department of Optoelectronic Engineering, Jinan University, Guangzhou 510632, China.
(**Presenting Author: tpan@jnu.edu.cn*; *466945939@qq.com*)

For the rapid measurement of a complex analyte using near-infrared (NIR) spectroscopy, appropriate wavelength selection is an important and albeit difficult aspect, which is essential for improving prediction performance, reducing model complexity and designing small dedicated spectrometers with a high signal-to-noise ratio.

Equidistant combination partial least squares (EC-PLS) method [1,2] focused on the choice of the combination of equidistant wavelengths by using initial wavelength, number of wavelengths and number of wavelength gaps as the parameters, which covers moving-window PLS (MW-PLS) in term of the wavelength screening algorithm. Due to the low degree of freedom of the parameters, EC-PLS can achieve the ergodic choice of wavelength combination in large range. Under statistical consideration, the equivalence model set was proposed based on EC-PLS. Those wavelengths appear in the equivalent model set were selected and sorted according to their repetition rate. A further wavelength selection method, called repetition rate priority combination PLS (RRPC-PLS), was thus proposed and successfully applied to NIR analysis of soil organic matter. The successive projections algorithm PLS (SPA-PLS) and EC-PLS, which are well-performed wavelength selection methods, were also conducted for comparison. A rigorous process based on the various divisions of calibration and prediction sets was performed to achieve modeling optimization with stability.

The results of the prediction performance showed that the optimal SPA-PLS model (28 wavelengths) was significantly superior to the full PLS model (860 wavelengths); the optimal EC-PLS model (48 wavelengths) was superior to the optimal SPA-PLS model. While the optimal RRPC-PLS model (39 wavelengths) was further superior to the optimal EC-PLS model in two aspects of prediction performance and model complexity. Thus, the RRPC-PLS is the good improvement on EC-PLS, which can be more effective to extract the information wavelengths and remove the redundant wavelengths.

It is very interesting to note that the RRPC-PLS method also found a simpler and high performance model with nine wavelengths for soil organic matter, which provided valuable reference for development of small and dedicated spectrometer. We believe that proposed method has such applicability and can be applied to other fields.

Acknowledgement: This work was supported by National Key Technology Support Program of China (No. 2012BAD18B01), the Science and Technology Project of Guangdong Province of China (No. 2014A020213016, No. 2014A020212445).

References:

[1] T. Pan, M.M. Li, and J.M. Chen, *Appl. Spectrosc.*, 68, 263-271 (2014).
[2] Y. Han, J.M. Chen, T. Pan, and G.S. Liu, *Chemometr. Intell. Lab. Syst.*, 145, 84-92, (2015).

P-213

Visualizing model dependencies by projections - milk phenotyping using vibrational spectroscopy and chemometrics

C.E. Eskildsen

Department of Food Science, University of Copenhagen, Rolighedsvej 26, DK-1858 Frederiksberg, Denmark.
carle@food.ku.dk

Livestock breeders and the dairy industry are showing increased interest in high-throughput methods for milk phenotyping. Vibrational spectroscopic techniques are widely used throughout all stages of food production and have also been proposed as a method for milk phenotyping. A number of previous studies reported successful prediction of individual fatty acids and protein fractions from Fourier transform infrared measurements of milk. However, the regression vectors for such phenotype regression models are estimated in the direction of the signal of interferents such as total fat and protein content. Hence, the predictions are trapped in a *cage of covariance* with the major milk constituents. The prediction models for detailed milk composition are not thus based on causal relationships and this may seriously compromise calibration robustness.

Sanchez and Kowalski [1] identified the regression vector for a given analyte as a contravariant vector. Hence, the regression vector is the part of the analyte signal orthogonal to the signals of the interferent species. If signals of the interferents vary, then based on the orthogonal constraint the regression vector for the analyte contravaries [1,2]. This is essential in order to obtain predictions of the analyte independent of the interferents.

Figure 1 sketches the basic idea on how to investigate whether predictions of a given analyte, $\boldsymbol{a}$, are independent of an interferent component, $\boldsymbol{k}_1$. If the estimated regression vector, $\boldsymbol{b}^{hat}$, has predicting power for the signal of $\boldsymbol{a}$ and is orthogonal to the signal of $\boldsymbol{k}_1$, $\boldsymbol{b}^{hat}$ will also have predicting power for the signal of $\boldsymbol{a}_{\text{-k1}}$ (that is, the signal of $\boldsymbol{a}$ orthogonalized with the signal of $\boldsymbol{k}_1$) as shown in **Figure 1A**. **Figure 1B** illustrates an example where $\boldsymbol{b}^{hat}$ is estimated (partly) in the direction of $\boldsymbol{k}_1$. Hence, the predictions of $\boldsymbol{a}$ are dependent on $\boldsymbol{k}_1$. Then it is found that $\boldsymbol{b}^{hat}$ has limited predicting power for $\boldsymbol{a}_{\text{-k1}}$ and this reveals that the predictions of $\boldsymbol{a}$ are dependent of $\boldsymbol{k}_1$.

P-214

DIRECT DETECTION OF ENDOGENOUS MICRORNAS AND THEIR POST-TRANSCRIPTIONAL MODIFICATIONS IN CANCER SERUM BY CAPILLARY ELECTROPHORESIS-MASS SPECTROMETRY

N. Khan[1], G.G. Mironov[1], M. V. Berezovski[1]

[1]Department of Chemistry and Biomolecular Sciences, University of Ottawa, 10 Marie Curie, Ottawa, ON, Canada K1N 6N5

MicroRNA molecules (miRNAs) are a class of small, single-stranded, non-coding RNA molecules that regulate cellular messenger RNA and their corresponding proteins. Extracellular miRNAs circulate in the bloodstream inside exosomes or in complexes with proteins and lipoproteins. The miRNA sequences and they quantitative levels are used as unique signatures associated with cancer diagnosis and prognosis after anticancer treatment. MicroRNAs are modified through a series of processing events after transcription like 5´-end phosphorylation, 3´- end adenylation or uridylation, terminal nucleotide deletion. The problem is that existing bioanalytical methods such as microarrays and a quantitative polymerase chain reaction are sensitive, but not capable of identifying the posttranscriptional modifications of miRNA. Here we report a Capillary Electrophoresis-Mass Spectrometry (CE-MS) method, which performs a multiplex, direct analysis of miRNAs from biological samples. Using the CE-MS method, we detected two endogenous human circulating miRNAs, a 23-nucleotide long 5´-phosporylated miRNA with 3´-uridylation (iso-miR-16-5p) and a 22-nucleotide long 5´-phosporylated miRNA (miR-21-5p) isolated from B-cell chronic lymphocytic leukemia serum. The CE separation and following MS analysis provides label-free quantitation and reveals modifications of miRNAs. MicroRNA profiling of serum samples with CE-MS has the potential to be a versatile and minimally invasive bioassay that could lead to better clinical diagnostics and disease treatment.

AUTHOR INDEX

SURNAME	NAME	PRESENTATION CODE
Abbas	Ouissam	P-175
Abel	Marie-Laure	O-29
Aceña	Laura	P-131
Achata	Eva Maria	P-207
Afanador	Nelson Lee	O-10, P-025, P-026
Ágreda Bastidas	Jesús Alberto	P-182
Ait Ider	Abdelouahed	P-210, P-211
Alaerts	Goedele	P-142
Alberch	Jordi	P-085
Alberts	Anastasia	P-195
Alcalà Bernárdez	Manel	P-040, P-044, P-096, P-101
Alcaráz	Mirta Raquel	P-123, P-144, P-205
Aldamiz-Echevarria	Luis	P-080
Alfonso	Ignacio	P-066
Aliakbarian	Bahar	P-094
Alladio	Eugenio	P-074, P-075
Allegrini	Franco	P-174
Alvarez-Segura	Tamara	P-001, P-002, P-003
Alving	Kjell	P-082
Amante	Eleonora	P-075
Amaral	Lucas	P-153
Amat	Sandrine	P-199
Amigo Rubio	José Manuel	P-072, P-191, P-192
Amrhein	Michael	O-02
Andrade	Fernando	P-080
Andries	Jan	P-139
Anggrahini	Sri	P-106
Antalick	Guillaume	P-097
Araki	Koiti	P-202
Arancibia	Juan	P-137
Araújo Honorato	Fernanda	P-192
Araya	Juan	P-184
Arce	Lourdes	O-40
Arévalo	Cristian	P-184
Ariño	Cristina	P-010
Arlinghaus	Heinrich F.	P-178
Arteaga	Francisco	P-030
Augusto	Fabio	P-099
Azcarate	Silvana Mariela	P-121
Azzi	Samar	P-128
Baeten	Vincent	P-168, P-174, P-175
Bagnasco	Lucia	P-094, P-171
Ballabio	Davide	O-47, P-143
Baquero	David	P-057
Barata	Carlos	P-069
Barba	M Isabel	P-196
Barbosa	Jose	P-085
Barbosa	Marcio	P-186
Barbosa	Sergio	P-109
Barbosa Torralbo	José	P-084
Barolo	Massimiliano	P-046
Barreiro Elorza	Pilar	P-111
Barrio	Ramon J.	P-080
Barroso Ramos	Albert	P-084
Bart	Jacob	P-024
Bartolomeo	Gianluca	P-175
Bastidas	Camila	P-183

SURNAME	NAME	PRESENTATION CODE
Batenburg	Max	O-55
Batista	André Mariano	P-072
Baum	Andreas	O-56
Bec	Nicole	P-203
Bedia	Carmen	P-078, P-172
Beleites	Claudia	O-13
Belin	Stephanie	P-206
Belis	Claudio A.	O-36
Beltran	Jose Luis	P-147
Beltrand	C.	P-172
Ben Ali	Abdelkader	P-210
Benabou	Sanae	P-165, P-173
Benassi	Marta De Toledo	P-134
Benavente Moreno	Fernando	P-084, P-085
Bendoula	Ryad	O-32
Beneito-Cambra	Miriam	P-138
Benito	Sandra	P-080
Berezovski	Maxim	P-214
Bergmann	Jan	P-152
Bernadó	Pau	O-38
Bernardi Miguel	Rodrigo	P-202
Bertinetto	Carlo G.	P-201
Bertrand	Samuel	O-48
Billeter	Julien	O-02
Blackman	John	P-097
Blanchet	Lionel	O-10, O-11, P-021, P-025, P-026
Blanco Romia	Marcelo	P-040
Blanco Romía	Marcel	P-044
Blumenthal	David	P-129
Bocklitz	Thomas	P-161
Bogomolov	Andrey	O-53
Bonvin	Dominique	O-02
Boqué	Ricard	P-131
Borràs	Eva	P-131
Botelho	Bruno	P-114
Botelho	Élvio	P-071
Bouklouze	Aziz	P-132
Braga	Jez	O-20, P-071
Braga	Michelle	P-153
Braga	Soraia Cristina Gonzaga Neves	P-099
Brandstetter	Markus	O-41, P-050, P-051
Brassier Font	Judit	P-040
Bravo	Manuel	P-152
Bravo	Sandra Daniela	P-117
Brereton	Richard G	P-198
Brezmes	Jesus	P-090
Briois	Valerie	P-206
Bro	Rasmus	P-060, P-113
Brouwer	Eric	P-112
Brown	Phil	O-55
Bruun	Sander	O-07
Burgos-Simón	Clara	P-138
Bustamante	Leonardo	P-083
Busto	Olga	P-131
Buydens	Lutgarde M. C.	O-11, O-48, O-55, P-021, P-024, P-025, P-026, P-027, P-028, P-031, P-062, P-080, P-093, P-112
Cabal	Ana	P-185

SURNAME	NAME	PRESENTATION CODE
Cabello	Zaida	P-104
Caixach	Josep	P-115
Callao	M. Pilar	P-124, P-125, P-126
Calles	Beatriz	P-156
Calvet	Amandine	P-166
Camacho	José	O-14, P-009, P-041
Camel	Valérie	P-129
Camiña	Jose Manuel	P-121
Campiglia	Andres D.	P-170
Campos	Bruno	P-069
Carabajal	Maira	P-137
Cardell	Carolina	P-143
Cárdenas Espita	Vanessa	P-040
Cariou	Véronique	O-08
Carrasco	Beatriz	P-032
Carrasco-Correa	Enrique Javier	P-138, P-155
Caruso	Roberto	P-075
Carvalho Vieira	Joana	P-148
Casale	Monica	P-094, P-171
Casals Missio	Joan	P-096, P-101
Casamayou-Boucau	Yannick	O-31, P-166
Castillo	Elianna	P-047
Castillo	Rosario Del P.	P-183, P-184
Castro	Luís Miguel	P-148
Catrinck	Mariana	P-186
Cavalcante Da Silva	Neirivaldo	P-192
Cavalcanti	Cláudia Jéssica Da Silva	P-192
Cavallini	Nicola	P-113
Cavanillas	Santiago	P-010
Cedeño	Miguel	P-117
Cédric	Rvier	P-064
Cely	Juan Camilo	P-158
Cepeda Vázquez	Mayela	P-129
Cernuda	Carlos	O-41
Chagas	Junior Olair	P-055, P-056
Chang	Po-Ling	P-157
Chauvet	Elodie	P-199
Checa	Antonio	P-082
Chen	Han-Yu	P-157
Chen	Jiemei	P-212
Chiominto	Alessandra	P-151
Ciosek	Patrycja	P-095, P-187
Clarke	Barrie	P-087
Claude	Roche	P-064
Clavaud	Matthieu	P-007
Clavel	Thomas	P-089
Cocchi	Marina	O-42, P-113, P-191
Codony	Rafael	P-115
Colombani	Juliette	P-199
Cordeiro Dantas	Willian Francisco	P-181
Coronas	Alberto	P-196
Correig	Xavier	P-090
Cortina	Jose Luis	P-057, P-147
Costa	Jean	P-118
Costa	Maria De Lourdes	P-148
Craigie	Cameron	P-177
Cristescu	Simona	O-48

SURNAME	NAME	PRESENTATION CODE
Crozet	Marielle	P-064
Cruz Sanchez	Jordi	P-096
Cuadrado	Teresa	P-111
Cuadros-Rodríguez	Luis	P-070, P-120
Cuevas	Francisco Julián	P-091
Culzoni	Maria Julia	P-144, P-145
Cunha	Isabela Cristina Matos	P-055, P-056
Da Costa Ferreira	Ana Maria	P-202
Da Silva	Alcides	P-042
Da Silva Friis	Helena	P-113
Da Silva Leal	Cristiano	P-148
Da Silva Lopes De Oliveira	Lise Mirela	P-092
Dalmau Solà	Núria	P-078
Danièle	Roudil	P-064
Dankowska	Anna	P-100
Dantas	Kele	P-114
Dardenne	Pierre	P-168, P-174
Daszykowski	Michal	P-076
Davies	Tony	P-024
De Gasper	André Luis	P-146
De Juan	Anna	P-042, P-160, P-173, P-200, P-201, P-203
De Juan Capdevila	Anna	P-130
De Lima	Kássio	P-042
De Marsico	Maria	P-209
De Noord	Onno E.	O-05
De Oliveira	Rodrigo Rocha	P-042
De Rooi	Johan	O-17, O-25
Dean	Nema	P-065
Dedecker	Peter	O-25
Degardin	Klara	O-28, P-007
Dejaegher	Bieke	P-142
Delgado Sánchez	Pablo	P-111
Deloire	Alain	P-097
Deng	Baichuan	P-029
Desai	Kavita	P-127
Descours	Emilie	P-034
Devos	Olivier	O-25, P-160
Dias	Aécio L.S.	P-175
Díaz	Marta	P-090
Díaz-Cruz	José Manuel	P-010
Diehl	Bernd	O-50
Domingo-Almenara	Xavier	P-090
Donat-Beneito	Rosa María	P-006
Dorn	Mary Frances	O-04
Dorrepaal	Ronan	P-167
Draude	Felix	P-178
Drygas	Szymon	P-197
Du	Congmin	O-51
Dubbelman	Anne-Charlotte	P-038
Dubuc	Perrine	P-061
Duffieux	Cyrille	P-199
Duponchel	Ludovic	O-37
Dupuy	Nathalie	P-118, P-199
Durante	Caterina	P-191
Ebert	Franziska	P-195
Egorov	Alexander	P-023
Ehgartner	Daniela	P-204

SURNAME	NAME	PRESENTATION CODE
Eilers	Paul	O-25
Emami	Fereshteh	P-136
Engel	Jasper	O-11, O-21, P-021
Engelhard	Carsten	P-194, P-195
Engelke	Udo	O-11
Engelsen	Søren B.	O-07
Erschov	Gennady	P-203
Erxleben	Andrea	P-164
Escandar	Graciela	P-098, P-137
Escobar	Veronica	P-152
Escobedo-Reyes	Antonio	P-117
Eskildsen	Carl Emil	P-213
Esquerre	Carlos	O-30, P-207
Esteban	Miquel	P-010
Estephan	Nathalie	P-128
F Gouveia	Francisca	P-060
Facco	Pierantonio	P-046
Fartmann	Michael	P-194
Fedorak	Richard	P-089
Fernández Pierna	Juan Antonio	P-168, P-174, P-175
Fernandez-Ahumada	Elvira	O-06
Ferraz Trindade	Gustavo	O-29
Ferre	Joan	P-017, P-131
Ferrer	Alba	P-165
Ferrer	Alberto	O-05, O-18, P-030, P-045, P-046, P-062
Filatova	Daria	P-110
Filgueiras	Paulo Roberto	P-092
Filzmoser	Peter	O-23
Fluvià Sabio	Sergi	P-130
Folcarelli	Rita	P-027
Folch-Fortuny	Abel	O-18, P-030
Fonollosa	Jordi	O-40
Fontes De Oliveira	Luciana	P-099
Forchetti	Débora	P-162
Fracassi Da Silva	José Alberto	P-140
Fragoso	Wallace Duarte	P-174
Franquet	Helena	P-149, P-150
Freer	Juanita	P-184
Freitas	Ana Isabel	P-054
Fu	Xiangxue	P-122
Fuentes	Edwar	P-083
Fuentes-García	Marta	O-14
Funsueb	Sujittra	P-015
G Felizardo	Pedro	P-060
G Huila	Manuel F	P-202
Galeano Díaz	Teresa	P-098
Galindo-Prieto	Beatriz	O-43, P-037
Gallart-Ayala	Héctor	P-082
Gárate	José	P-083
García Barroso	Carmelo	P-105
García Muñoz	Salvador	P-043
García-Alvarez-Coque	María Celia	P-001, P-002, P-003, P-006, P-014
Garcia-Munoz	Salvador	O-15
Gargallo	Raimundo	P-165, P-173
Garreta	Elba	P-011, P-069
Garrido Delgado	Rocío	O-40
Garrido-Sáez	Neus	P-155

SURNAME	NAME	PRESENTATION CODE
Gasper	A.l.	P-141
Gavilán	Arturo	P-156
Gavilán	Irma	P-156
Geladi	Paul	O-43, P-037
Georgouli	Konstantia	P-036
Gerace	Enrico	P-075
Gergely	Szilveszter	P-086
Gerretzen	Jan	P-024, P-025, P-026
Geurts	Brigitte	O-48
Ghasemi	Jahan B.	O-36 P-011
Ghassemi	Farbord	P-034
Gibert	Oriol	P-147
Gierlinger	Notburga	O-26
Giménez López	Estela	P-084
Gines	Silvia	P-085
Goicoechea	Hector C.	P-121, P-144, P-145, P-170, P-205
Goicolea	M. Aranzazu	P-080
Gomes	José Messias	P-055
Gomez	Cécile	O-06
Gonçalves	Roberto	P-125, P-126
Gondim	Carina S.	P-126
González	Susana	P-057
González Bosch	Josep Maria	P-040
González-Casado	Antonio	P-120
González-Sáiz	J.m.	P-119
Goodarzi	Mohammad	P-139
Gorretta	Nathalie	O-32
Gorrochategui	Eva	P-067
Gottlieb	Cassian	P-190
Govaerts	Bernadette	O-19
Gowen	Aoife	P-167, P-207
Gowen	Aoife A.	O-30
Grabowski	Ireneusz	P-076
Granados	Mercè	P-147
Grobério	Tatiane	P-071
Groenning	Minna	O-38
Grootveld	Kerry	P-127
Grootveld	Martin	P-127
Grothe	Sven	P-190
Guamán Balcázar	Maria Del Cisne	P-108
Guardiola	Francesc	P-115
Guhan	Jayaraman	P-188
Guilment	Jean	P-061
Guo	Shuxia	P-161
Hagenhoff	Birgit	P-194, P-195
Hakemeier	Thomas	P-038
Haller	Dirk	P-089
Hansen	Per Waaben	O-56
Hanson	Joshua	O-15
Harms	Amy	P-038
Harren	Frans	O-55
Harrington	Peter	O-46
Heider	Emily	P-170
Heinke	Ralf	P-161
Heller	Danica	P-194, P-195
Henriques	Amélia Teresinha	P-141, P-146
Hermes	Michael	P-208

SURNAME	NAME	PRESENTATION CODE
Hernández-Cassou	Santiago	P-109
Herranz	Fatima	O-38
Herrera	Agustín	P-143
Herrero	Ana	P-049
Herrero-Langreo	Ana	O-32
Herwig	Christoph	P-204
Hidalgo	Míriam	P-109
Hidalgo Serrano	Miriam	P-130
Hodnett	Kieran	P-164
Hoefsloot	Huub C.J.	O-18
Hopke	Philip	O-33, P-136
Hou	Ta	P-059
Hron	Karel	O-23
Hubert	Philippe	P-007
Ibañez	Gabriela	P-137
Ibáñez	Lourdes	P-090
Ioele	Giuseppina	P-149, P-150
Izadmanesh	Yahya	O-36, P-011
Jacobs	Doris	O-52
Jakmunee	Jaroon	P-015
Jančić-Stojanović	Biljana	P-004, P-005
Jansen	Jeroen	O-48, P-024, P-027, P-028, P-031, P-093
Jansen	Jeroen J.	P-080
Jansson	Pim	P-127
Jaumot	Joaquim	P-068, P-082, P-085, P-135, P-172
Jean	Guilment	O-12
Jimenez	Rodrigo	P-158
Jiménez Soto	Juan-Manuel	O-40
Jiménez-Carvelo	Ana María	P-120
Johnson	Kevin	P-012
Juerss	Annekathrin	P-179
Kalivas	John	P-017
Kamali	Naghmeh	P-164
Kantner	Wolfgang	O-41, P-051
Kaowphong	Sulawan	P-018
Karakach	Tobias	P-189
Karrer	Andrea	P-177
Kassenböhmer	Rainer	P-178, P-180
Kendall	Catherine	P-198
Kerkenaar	Harmen H.M.	P-062
Khan	Nasrin	P-214
Kharbach	Mourad	P-132
Khaydukova	Maria	P-035
Kiers	Henk A.I.	O-18
Kim	Inyoung	O-34
Kirsanov	Dmitry	P-035
Kittiwatchana	Sila	P-015, P-018
Kitzberger	Cíntia Sorane Good	P-133, P-134
Klein-Júnior	Luiz Carlos	P-141, P-146
Knapp	Adam	P-012
Koch	Cosima	P-204
Koeman	Mike	P-031
Koetz	M.M.	P-141
Kohler	Achim	O-13, O-39
Koidis	Anastasios	P-036
Körsgen	Martin	P-178
Kotwica	Andzelika	P-100

SURNAME	NAME	PRESENTATION CODE
Koussissi	Elisabeth	P-112
Koval	Slavik	P-038
Krafft	Christoph	O-13
Krakowska	Barbara	P-076
Krongchai	Chanida	P-015, P-018
Kucheryavskiy	Sergey	P-023
Kuligowski	Julia	P-079, P-123, P-204
La Fontaine	Camille	P-206
Lacorte	Sílvia	P-067, P-069, P-149, P-150
Lagkouvardos	Ilias	P-089
Lallemand	Jordane	O-12, P-061
Lam	Chloie	O-52
Larrechi	M Soledad	P-196
Larroque	Christian	P-203
Laurentino Alves	Julio Cesar	P-181
Le Dréau	Yveline	P-118
Leal	Ana	P-052
Leal	Silane Flor De Liz Da Silva	P-133
Leardi	Riccardo	P-094
Leça	João Micael	P-052, P-053, P-054
Lee	Thomas	P-089
Leffers	Larissa	P-195
Legin	Andrey	P-035
Leiva	Karla	P-083
Lendl	Bernhard	P-123, P-204, P-205
Li	Boyan	P-166
Li	Jian-Feng	P-019
Li	Peiwu	O-24
Li	Zhongfeng	P-073
Li Vigni	Mario	O-42, P-191
Liang	Yizeng	P-029
Libânio	Marcelo	P-154
Lima	Lívia	P-153
Lima	Suzana	P-045
Lin	Wei-Yu	P-157
Lindedam	Jane	O-07
Lissorgues	Gaelle	P-034
Liu	Linqi	P-122
Liu	Ya_juan	P-028
Lleó García	Lourdes	P-111
Lloyd	Gavin Rhys	P-198
Lobaton Villa	Javier Erick	P-176
Lobo	Fabiana Aparecida	P-055, P-056
Lombana	Oscar Libardo	P-158
Lopes	João A.	P-119
López-Ureña	Sergio	P-006, P-014
Loza-Alvarez	Pablo	O-27, P-200
Lu	Shaohua	O-51
Lucio	Marianna	P-081, P-088, P-089
Ludeña	Mishell	P-108
Lughofer	Edwin	O-41
Lugo-Melchor	Yadira	P-117
Luoma	Pekka	P-050
Maaouni	Mohamed	P-210, P-211
Maciá-Fernández	Gabriel	O-14
Magalhães Dias	Júlio César	P-092
Maia	Kelly	P-154

SURNAME	NAME	PRESENTATION CODE
Maia	Mariza	P-102
Maldonado	Jimmy	P-108
Malli	Birgit	P-050, P-051
Mallon	Timothy	O-33
Mamián López	Mónica B	P-202
Marcelloni	Anna Maria	P-151
Marco	Santiago	O-40
Marín	Marc	P-149, P-150
Marini	Federico	O-44, P-103, P-151, P-175
Marques	José C.	P-053
Marques	José Carlos	P-052, P-054
Marques Teixeira	Orivan Maria	P-140
Marquez	Cristina	P-124
Marro	Mónica	O-27, P-200
Martin	Francis L.	P-067
Martin	Jordi	P-057
Martineau	Pierre	P-203
Martinez	Ekaitz	O-54
Martinez Del Rincon	Jesus	P-036
Martínez-Díaz	Yoanna	P-209
Martins	Sara	O-55
Martyna	Agnieszka	O-22, P-065, P-074, P-075
Mas	Silvia	P-203
Mathias	Sawall	O-49
Mbarki	Mohamed	P-210, P-211
Mc Ardle	Patrick	P-164
Medenica	Mirjana	P-004
Medina-Rodríguez	Santiago	P-070
Meeks	Campbell	P-097
Meléndez	María Elena	P-022
Melenteva	Anastasiia	O-53
Melo	Nathalie Ferreira Silva De	P-055
Mendiola-Lau	Victor M.	P-209
Merbouha	Abdelkarim	P-210
Mestres	Montserrat	P-131
Michalke	Bernhard	P-081, P-088, P-089
Michalska	Aleksandra	O-22, P-065
Mikkelsen	Jørn Dalgaard	O-56
Mironov	Gleb	P-214
Mišković	Stefan	P-004
Monago Maraña	Olga	P-098
Monakhova	Yulia	O-50
Montagnier	Safia	P-061
Montañés	Elena	P-032
Monteiro	Robson	P-186
Montemurro	Milagros	P-121, P-144, P-145
Moreno-Rojas	Jose Manuel	P-091
Moumtaz	Sarah	P-127
Moyano	Encarnación	P-110
Mujumdar	Nirvani	P-170
Muñoz De La Peña	David	P-170
Muñoz De La Peña Castrillo	Arsenio	P-098, P-170
Næs	Tormod	O-05
Nallala	Jayakrupakar	P-208
Nascimento	Rafael	P-153
Natschläger	Thomas	P-050, P-051
Navarro-Huerta	José Antonio	P-002, P-003, P-014

SURNAME	NAME	PRESENTATION CODE
Navarro-Reig	Meritxell	P-135
Navas	Natalia	P-070, P-143
Naz	Shama	P-082
Neerincx	Anne	O-48
Neocleous	Tereza	P-065
Neth	Katharina	P-081, P-088
Neymeyr	Klaus	O-01, O-49, P-013, P-179
Niedziółka-Jönsson2	Joanna	P-187
Ningrum	Andriati	P-107
Nørgaard	Lars	O-56
Nouri	Maroua	O-32
Nunes Pinto Paracampo	Nádia Elígia	P-140
Núñez	Oscar	P-109, P-110
Oca	María Leticia	O-35
O'Donnell	Colm P.	O-30, P-207
Ofner	Johannes	P-204
Ohurogo	Richard	P-127
Olabarrieta	Idoia	O-54
Oleneva	Ekaterina	P-035
Oliveira	Ana Flávia	P-133
Oliveira	Isadora	P-102
Oliveri	Paolo	P-171
Olivieri	Alejandro C.	P-174
Oller Moreno	Sergio	O-40
Olmos	Victor	P-200
Ortiz	María Cruz	O-35, P-022, P-048, P-049, P-103, P-104
Ortiz-Villanueva	Elena	P-068
Orzeł	Joanna	P-076
Ouaini	Naim	P-128
Ouizem	Zina	P-168
Paba	Emilia	P-151
Paci	Enrico	P-151
Palací-López	Daniel	P-046, P-062
Palma	Miguel	P-105
Pan	Tao	P-212
Panchuk	Vitaly	P-035
Paolini	Julien	P-118
Pardo	Antonio	O-40
Pardo Almudí	Rafael	P-156
Park	Eun Sug	O-34
Parmar	Devki	P-127
Parra	Sonnia	P-152
Parra Borderías	María	P-130
Párta	László	P-086
Pasamontes	Alberto	P-038
Páscoa	Ricardo N. M. J.	P-119
Pasquini	Celio	P-072
Pawlak-Lemańska	Katarzyna	P-116
Pawliczek	Marcin	O-41, P-050, P-051
Pedroza	Ricardo	P-042
Pellegrino Vidal	Rocío	P-137
Pelster	Andreas	P-178
Percival	Benita	P-127
Perego	Patrizia	P-094
Pereira	Ana Cristina	P-052, P-053, P-054
Pereira	Claudete	P-058
Pereira	Vanda	P-054

SURNAME	NAME	PRESENTATION CODE
Pereira Do Amaral	António Luís	P-148
Pereira-Caro	Gema	P-091
Perera	Alexandre	P-090
Perez	Izaskun	O-54
Pérez	Irene	P-048
Pérez	Rocío Laura	P-098
Pérez	Rodrigo	P-047
Pérez-Castaño	Estefanía	P-120
Pérez-Del-Notario	Nuria	P-119
Pérez-Robles	Raquel	P-070
Perrine	Dubuc	O-12
Pimentel	Maria Fernanda	P-045, P-072, P-192
Piña	Benjamín	P-066, P-200
Piqueras Solsona	Sara	P-193
Pirro	Valentina	P-074
Pizarro	Consuelo	P-119
Platikanov	Stefan	P-057
Polyakov	Victor	P-023
Pomerantsev	Alexey	P-016
Pont	Laura	P-085
Popović	Igor	P-005
Popp	Jürgen	O-13, P-161
Poppi	Ronei Jesus	P-092, P-099, P-140, P-162, P-176, P-181,
Possas Abreu	Maira	P-034
Postma	Geert	P-028, P-080
Postma	Geert J.	P-062
Prats Mateu	Batirtze	O-26
Puig-Castellví	Francesc	P-066
Puignou	Lluís	P-109
Qannari	El Mostafa	O-08
Quevedo	José Ramón	P-032
Quintás	Guillermo	P-079
Quinteiro	Andreia	P-148
Quiroz	Waldo	P-152
R Mortensen	Asmus	P-060
Rabi	Souad	P-211
Ragno	Gaetano	P-149, P-150
Rahbek	Jesper P	P-060
Ramirez-Lopez	Leonardo	P-033
Ramis-Ramos	Guillermo	P-138, P-155
Rega	Barbara	P-129
Reguera	Celia	P-049
Reinke	Stacey N.	P-082
Reis	Marco	P-052, P-053
Reis	Marlon M.	O-54, P-177
Reischer	Thomas	O-41, P-051
Ren	Lei	P-122
Ribas	Rogério	P-186
Ribeiro De Castro	Eustáquio Vinicius	P-092
Richter	Pablo	P-083
Rinnan	Åsmund	O-07
Robayo Molina	Iván	P-182
Rodionova	Oxana	P-016
Rodrigues	Pedro	P-053
Rodriguez	Raquel	O-54
Rodriguez Turienzo	Laura	P-130
Rodriguez-Campos	Jacobo	P-117

SURNAME	NAME	PRESENTATION CODE
Roger	Jean-Michel	O-06, O-32
Rogez	Hervé	P-175
Roggo	Yves	O-28, P-007
Romão	Wanderson	P-092
Romero	Agustí	P-115
Roque	Jussara	P-102, P-186
Rosales	Alejandro	P-130
Roudenko	Olga	P-206
Rousseau	Lionel	P-034
Roussel	Sylvie	P-061
Roźniecka	Ewa	P-187
Ruckebusch	Cyril	O-25, P-160, P-173
Ruisanchez	Itziar	P-124, P-125, P-126
Ruiz	Facundo	P-111
Ruiz-Moreno	María José	P-091
Ruiz-Rodado	Victor	P-127
Ruíz-Samblás	Cristina	P-120
Rutledge	Douglas N.	O-08, P-128
Ryder	Alan G.	O-31, P-166
Sabate Reboll	Josep	P-096
Saccenti	Edoardo	O-09, P-009
Saccenti	Eduardo	O-52
Sáez	Pedro	P-083
Said	Mohammed El Amin	P-008, P-169
Saitua	Eduardo	O-54
Sajdak	Marcin	P-163
Salavera	Daniel	P-196
Saldanha Honorato	Ricardo	P-072
Sales	Rafaella	P-045
Salgó	András	P-086
Salomone	Alberto	P-074, P-075
Salton	Juliana	P-141, P-146
Samino	Sara	P-090
Sanchez	Alicia	P-080
Sánchez	M. Sagrario	P-048
Sánchez-Illana	Ángel	P-079
Sanjuan-Herráez	Daniel	P-079
Sanllorente	Silvia	P-049
Santiago	Anibal	P-153
Sanz Nebot	Victoria	P-084
Sanz-Nebot	Victoria	P-085
Sarabia	Luis Antonio	O-35, P-022, P-048, P-049, P-103, P-104
Saurina	Javier	P-109, P-110
Savorani	Francesco	P-113
Sawall	Mathias	O-01, P-013, P-179
Schmidt	Annemarie	P-089
Schmidtke	Leigh	P-097
Schmitt Kopplin	Philippe	P-081, P-088
Schmitt-Kopplin	Philippe	P-089
Schoenmakers	Peter J.	P-135
Scholz	Maria Brígida Dos Santos	P-133, P-134
Schoot	Mark	P-112
Schröder	Henning	P-013
Schwaighofer	Andreas	P-123, P-205
Schwerdtle	Tanja	P-195
Sena	Marcelo	P-114
Serrano	Núria	P-010

SURNAME	NAME	PRESENTATION CODE
Setiyoko	Agus	P-106
Setyaningsih	Widiastuti	P-105, P-108
Shankar	Narasimhan	P-188
Shapaval	Volha	O-39
Shi	Zhenqi	O-15, P-043
Siano	Gabriel	P-121, P-144, P-145
Siewert	Hugelier	O-25, P-160
Sikorska	Ewa	P-116, P-197
Silva	Carolina	P-072
Silva	Gilmare Antônia Da	P-055, P-056, P-153, P-154
Silva	Jailson	P-058
Silva	Ludymyla	P-153
Silva	Verônica	O-20
Silva	Vitor	P-058
Silva-Mata	Francisco J.	P-209
Simó	Joan	P-096
Simó Cruanyes	Joan	P-101
Simón Nicó	Estela	P-101
Sitta	Nicola	P-171
Sliwa	Michel	O-25, P-173
Smilde	Age K.	O-05, O-18, O-52
Smirnov	Kirill	P-089
Smolinska	Agnieszka	O-10
Soltero	Francisco	P-117
Sommer	Ulf	O-21
Sørensen	John	O-56
Sousa	Adriano	P-042
Souza	Scheila V. C.	P-125, P-126
Spagnoulo	María Laura	P-103
Spiegelman	Clifford	O-04, O-34
Srinesh	Chandrakesa	P-188
Stajić	Ana	P-005
Stanimirova	Ivana	P-020, P-076
Stanimirović	Zorica	P-004
Stevens	Antoine	P-033
Stone	Nick	P-198, P-208
Stragevitch	Luiz	P-045
Stringheta	Paulo	P-102
Stuart	Adam	P-177
Šuklje	Katje	P-097
Sun	Dong	P-044, P-096
Suppers	Anouk	P-112
Sylvie	Roussel	O-12
Szabó	Éva	P-086
Sznajder	Miroslaw	P-076
Szymanska	Ewa	O-55, P-024, P-025, P-026, P-112
T Pedersen	Mette	P-060
Tafintseva	Valeria	O-13, O-39
Talavera Bustamante	Isneri	P-209
Talhavini	Marcio	O-20, P-071
Tan	Shuman	O-34
Tan	Xiangyu	P-122
Tareilus	Erwin	O-52
Taukiri	Kevin	P-177
Tauler	Roma	O-36, O-38, P-011, P-057, P-066, P-067, P-068, P-069, P-078, P-085, P-135, P-149, P-150, P-172, P-200
Tcheka	Constant	P-210
Teglia	Carla	P-121

SURNAME	NAME	PRESENTATION CODE
Teixeira Dos Santos	C.A.	P-119
Temperini	Marcia L A	P-202
Tencate	Alister	P-017
Teófilo	Reinaldo	P-102, P-186
Ter Veen	Rik	P-194
Terra	Larissa	P-186
Terra	Luciana Assis	P-092
Terzić	Jelena	P-005
Teusink	Bas	O-18
Théodore	Cozzika	P-064
Theron	Roberto	P-041
Thiel	Michel	O-19
Timmerman	Marieke	O-09
Tinnevelt	Gerjen	P-093
Tisseyre	B.	O-32
Todeschini	Roberto	O-47, P-143
Torres	Francisco	P-083
Torres-Lapasió	José Ramón	P-001, P-002, P-003, P-006, P-014
Torro	Adeline	P-203
Tose	Lílian Valadares	P-092
Tran	Thanh	O-10, P-025, P-026, P-028
Tranfo	Giovanna	P-151
Tres	Alba	P-115
Trindade	Bruno	O-20
Troncoso	Jose	P-183
Trygg	Johan	O-43, P-037
Tumpa	Anja	P-004, P-005
Tyler	Bonnie	P-180
Unceta	Nora	P-080
Utell	Mark	O-33
Vallet	Nadine	P-034
Valls	Alba	P-131
Valverde-Som	Lucía	P-120
Van Beek	Teris	P-135
Van Den Heuvel	Edwin	P-024
Van Duynhoven	John	O-52
Van Espen	Piet	P-185
Van Maarschalkerweerd	Andreas	O-38
Van Manen	Henk-Jan	P-024
Van Velzen	Ewoud	O-52
Van Vliet	Michael	P-038
Vander Heyden	Yvan	P-132, P-139, P-141, P-142, P-146
Vassoler	Fábio	P-153
Vaz	Ana Isabel	P-148
Vaz	Boniekgontijo	P-092
Vega Alegre	Marisol	P-156
Vega-Montoto	Lorenzo	O-16
Vento	Máximo	P-079, P-123
Vera	Marc	P-147
Vestergaard	Bente	O-38
Viaene	Johan	P-141, P-142, P-146
Viant	Mark	O-21
Vichi	Stefania	P-115
Vigneau	Evelyne	O-08
Vinaixa	Maria	P-090
Vincenti	Marco	P-074, P-075
Vincke	Damien	P-168

SURNAME	NAME	PRESENTATION CODE
Vitale	Raffaele	O-05, P-045, P-062
Vivó-Truyols	Gabriel	P-077, P-135
Von Plessing	Carlos	P-183
Wahyuningtyas	Amalia	P-106
Walach	Jan	O-23
Walczak	Beata	O-23, P-021
Walker	Alesia	P-081, P-088, P-089
Wang	Ya-Hao	P-019
Watts	John F.	O-29
Wawrzyniak	Filip	P-197
Weinstock	B. Andre	P-059
Wentzell	Peter	P-087, P-189
Wesoły	Małgorzata	P-095
Westerhuis	Johan A.	O-05
Wevers	Ron	O-11
Wheelock	Craig E.	P-082
Wicks	Chelsi	P-087, P-189
Wieland	Karin	P-204
Wilsch	Gerd	P-190
Wiseso Marseno	Djagal	P-106
Wlasiuk	Patryk	O-22, P-075
Włodarska	Katarzyna	P-116
Wójcicki	Krzysztof	P-197
Woldegebriel	Michael	P-077
Wolfender	Jean-Luc	O-48
Wu	Hai-Long	O-45, P-028
Xu	Jian	O-03
Xu	Qingsong	O-03
Yanes	Oscar	P-090
Yang	Ying	P-118
Yao	Lijun	P-212
Yun	Yonghuan	P-029
Zabadaj	Marcin	P-187
Zacca	Jorge	O-20, P-071
Zadora	Grzegorz	O-22, P-065, P-074, P-075
Zalai	Dénes	P-086
Zaleszczyk	Grzegorz	P-076
Zhang	Liangxiao	O-24
Zhang	Xin	O-51 , P-073
Zhang	Zhuoyong	O-51 , P-073
Zhao	Chenxi	P-122
Zhou	Xiao-Shun	P-019
Ziemons	Eric	P-007
Ziere	Aldo	O-55
Zimmermann	Boris	O-39
Zotti	Mirca	P-171