

INFRARED SPECTROSCOPY COUPLED TO CHEMOMETRICS: A FAST APPROACH FOR CHARACTERIZING AND AUTHENTICATING LEBANESE HONEY

Rita El Hajj¹, Wadih Skaff², Nathalie Estephan¹

¹*Faculty of Sciences, Holy Spirit University of Kaslik, BP 446 Jounieh, Lebanon*

²*ESIAM-Université Saint-Joseph, Taanayel, Rue de Damas, BP 159 Zahlé, Lebanon*

nathalieestephan@usek.edu.lb

Honey production in Lebanon is predominantly mountain poly-floral honey, and orange blossom honey.

The aim of this work is to authenticate and characterize Lebanese honey depending on its botanical origin. As several analytical methods are simultaneously necessary for a reliable authentication of honeys, such work is time-consuming and costly as it is based on sensory and pollen analysis as well as several physico-chemical methods containing at least measurements of the electrical conductivity and the sugar content. Thus, there is a need for methods that allow a rapid and reproducible authentication of the botanical origin of honey at low cost.

Infrared Spectroscopy is becoming one of the commonly used methods because of its reputation as a rapid and non-destructive powerful analytical tool requiring minimum sample preparation.

In this study, classical physico-chemical methods as well as near and mid infrared spectroscopy are used to authenticate and characterize Lebanese honey. In total, 100 honey samples were collected from different Lebanese regions, altitudes and floral characteristics.

Numerical data of the physico-chemical and spectroscopic analysis are treated by chemometric methods including: exploratory methods (PCA, ICA) and discriminant methods (FDA, PLS-DA).

Chemometric analysis shows a correlation between physico-chemical and spectral data. According to the results obtained, honey samples are classified with high discrimination between the 2 categories of honey: floral and honeydew. PCA and ICA applied on mid and near infrared show a distribution of honey samples in function of their botanical origin.

Acknowledgement: CNRS - Lebanon for the financial support

INSPEC: CHEMOMETRICS MEETS ADVANCED PROCESS CONTROL

Tim Offermans¹, Jeroen Jansen¹

¹*Radboud University, Nijmegen*
tim.offerma@science.ru.nl

'Integrating Sensor Based Process Monitoring and Advanced Process Control' (INSPEC) is an ISPT-managed project in which academic institutes and industrial chemical companies collaborate to provide dedicated process control and monitoring solutions for optimizing energy and raw material use of chemical plants. As industrial processes have to deal with many exogenous sources of variability, they are operated within overly conservative control regimes. INSPEC: Chemometrics meets Advanced Process Control A more advanced use of available knowledge, process data and quality data will allow industries to operate in control regimes that will lead to a more economic use of energy and raw material. The development of chemometrics routines to integrate Process Analytical Techniques (PAT), like on-line spectroscopic measurements, and process variable data, like in-line temperature measurements, with Advanced Process Control (APC) is a key effort in this project.

Acknowledgement: Corbion, DSM, FrieslandCampina, Huntsman, ISPT, Technical University Eindhoven

SINC: A NEW SOFTWARE FOR THE AUTOMATED AND SIMULTANEOUS PHASE AND BASELINE CORRECTION OF NMR DATA

Denise Meinhardt^{1,2}, Mathias Sawall¹, Henning Schröder¹, Klaus Neymeyr^{1,2}

¹University of Rostock, Ulmenstr. 69, 18057 Rostock, Germany

²LİKAT, Leibniz-Institute for Catalysis, Albert-Einstein-Str. 29A, 18059 Rostock, Germany

denise.meinhardt@uni-rostock.de

Data preprocessing of NMR data by means of phase and baseline corrections is often a necessary step for successful chemometric analyses [1,2,3]. Usually, these two corrections are performed in separate steps. Here, we present a new algorithm for the automated and simultaneous correction of the phase and baseline of NMR data called SINC (simultaneous NMR signal correction) [4]. Testing the algorithm on sample data results in improved outcomes and significant time savings, compared to separate corrections.

The numerical algorithm is based on the minimization of a target function taking chemical objectives into account. For the solution of the resulting nonlinear least squares problem an extended Gauss-Newton method (including quasi-Newton update formula) and a genetic algorithm are combined in order to find the global minimum. In addition, a modified Savitzky-Golay filter is used for baseline detection, and the baseline is subsequently corrected by a Whittaker smoother. The software is written in C; a MATLAB-GUI is provided for the data handling.

References

- [1] Q. Bao, J. Feng, L. Chen, F. Chen, Z. Liu, B. Jiang and C. Liu. A robust automatic phase correction method for signal dense spectra. *J. Magn. Reson.*, 234, pp 82-89, 2013.
 - [2] J.C. Cobas, M.A. Bernstein, M. Martin-Pastor and P.G. Tahoces. A new general-purpose fully automatic baseline-correction procedure for 1D and 2D NMR data. *J. Magn. Reson.*, 183, pp 145-151, 2006.
 - [3] H. de Brouwer. Evaluation of algorithms for automated phase correction of NMR spectra. *J. Magn. Reson.*, 201(2):230-238, 2009.
 - [4] M. Sawall, E. von Harbou, A. Moog, R. Behrens, H. Schröder, J. Simoneau, E. Steimers and K. Neymeyr. Multi-objective optimization for an automated and simultaneous phase and baseline correction of NMR spectral data. *Accepted for J. Magn. Reson.*, 2018.
-

DESIGN OF EXPERIMENTS FOR THE REDUCTION OF THE ROTATIONAL AMBIGUITY IN MCR

Denise Meinhardt^{1,2}, Mathias Sawall¹, Christoph Kubis², Detlef Selent², Henning Schröder¹, Armin Börner², Klaus Neymeyr^{1,2}

¹University of Rostock, Ulmenstr. 69, 18057 Rostock, Germany

²LIKAT, Leibniz-Institute for Catalysis, Albert-Einstein-Str. 29A, 18059 Rostock, German

denise.meinhardt@uni-rostock.de

The analysis of spectroscopic data by multivariate curve resolution methods suffers from the so-called rotational ambiguity. Unique solution cannot be expected but a continuum of possible solutions exists. The area of feasible solutions (AFS) covers the ambiguity of the factorization. Often additional information and chemically expertise is required to extract meaningful solutions.

In this contribution, we investigate and discuss ideas to design experiments with the goal of reducing the intrinsic ambiguity of the factorization. The focus for developing strategies is on the concentration factor. Additionally, we impose constraints to divide the frequencies and spectra into relevant and nonrelevant ones with respect to their impact on the rotational ambiguity. The investigations are done against the background of an AFS representation. Results are presented for several experimental data sets.

References

- [1] Manne: On the resolution problem in hyphenated chromatography. *Chemom. Intell. Lab. Syst.* 27(1), pp 89-94, 1995.
 - [2] Sawall, Kubis, Selent, Börner, Neymeyr: A fast polygon inflation algorithm to compute the area of feasible solutions for three-component systems. I: Concepts and applications. *J. Chemom.* 27(5), pp 106-116, 2013.
 - [3] Rajkó, Abdollahi, Beyramysoltan, Omidikia: Definition and detection of data-based uniqueness in evaluating bilinear (two-way) chemical measurements. *Anal. Chim. Acta*, 855, pp 21-33, 2015.
 - [4] Sawall, Jürß, Neymeyr: FACPACK: A software for the computation of multi-component factorizations and the area of feasible solutions, Revision 1.3, <http://www.math.uni-rostock.de/facpack/>, 2015.
-

CHARACTERIZATION OF FOOD-BORNE BACTERIA ON CULTURE MEDIA BY MULTIPLE MODALITIES OF HYPERSPECTRAL IMAGING

Ana Herrero-Langreo^{1,2}, Gauri Ravindra Chemburkar¹, Amalia G.M. Scanell^{2,3,4}, Aoife Gowen^{1,2}

¹*UCD School of Biosystems and Food Engineering, University College of Dublin (UCD), Belfield, Dublin 4 Ireland*

²*UCD Institute of Food and Health*

³*UCD Center for Food Safety*

⁴*UCD School of Agriculture and Food Science*

ana.herrero-langreo@ucd.ie

The application of hyperspectral imaging for the identification of microbiological samples is a recently growing field with promising practical applications and specific challenges. While traditional methods require highly specialized training and long-time periods (up to 7 days), hyperspectral imaging can provide a quick, user-friendly and label-free tool for microbial assessment. The present work is a preliminary study for the identification of foodborne bacteria by combining different spectral modalities. Two strains of *Bacillus* and one of *Cronobacter* were grown in nutrient agar from six decimal dilutions using the spread-plate technique. Samples were imaged with a VIS-vNIR (400–1000nm) and a NIR (880-1720nm) push broom hyperspectral systems. Spectral profiles for each bacteria genus were compared through Principal Components Analysis. Partial Least Squares Discriminant Analysis for each spectral range was calibrated on the highest microbial concentration and applied on subsequent dilutions to predict microbial identification. Preliminary results show 99 % of correct pixel classification (CC) between *Cronobacter* and *Bacillus* on calibration images, while prediction on lower dilution test images presented lower CC (65% to 18%). Future work will assess the effect of spectral pre-treatments to correct the effect of variations in agar thickness. Further experimentation is planned to provide a validation set by replicating the experimental setup and including FT-IR microscopy imaging of specific colonies identified in the VIS-vNIR and NIR spectral images. This multimodal approach is being investigated to provide the knowledge base for a more robust and early-stage identification of food borne bacteria.

Acknowledgement: Authors would like to thank Multi-scale Hyperspectral Imaging for Enhanced Understanding and Control of Food Microbiology (HyperMicroMacro) project, funded by Science Foundation of Ireland.

THE USE OF MIR-FTIR COUPLED TO CHEMOMETRIC TOOLS TO DIFFERENTIATE BETWEEN DIESEL FUELS FROM DIFFERENT SUPPLIERS

A. Bouklouze¹, I. Barra¹, M. Alaoui¹, Y. Cherrah¹

¹Laboratory of Pharmacology and Toxicology, Faculty of Medicine and Pharmacy, University Mohammed V- Rabat- Morocco.

a.bouklouze@um5s.net.ma

In order to discriminate between products of four known diesel suppliers in Morocco, a chemometric model was built by the use of the MIR-FTIR known with its ease in use, reduced price of analyzes, non-destructive method and does not require sample preparation. Coupled with two chemometric tools, on the one hand, PCA was useful to visualize the data and to verify the existence of any outliers, with 4 PCs 70% of variability was explained, and it was noticed the formation of four groups perfectly separated, each Group represents samples belonging to each supplier. And on the other hand, the real model of discrimination was developed by PLS-DA, after fixing the number of optimal latent variables of 3 LVs, the discrimination of each group of samples compared to others has been achieved with a perfect selectivity and Sensitivity (almost 100%) and very low RMSEC and RMSEP errors, indicating the success of our model.

CLASSIFICATION OF FLUCONAZOLE API POLYMORPHIC FORMS IN PHARMACEUTICAL PRODUCTS BY ASSOCIATING FTIR, NIR AND RAMAN TO PLS-DA

Mohammed Alaoui Mansouri^{1,2}, Pierre-Yves Sacré¹, Eric Ziemons¹, Mourad Kharbach^{2,3}, Yahia Cherrah², Philippe Hubert¹, Abdelaziz Bouklouze², Roland Marini¹

¹ *University of Liege (ULiege), CIRM, VibraSanté Hub, Department of Pharmacy, Laboratory of Pharmaceutical Analytical Chemistry, Quartier Hôpital, Avenue Hippocrate 15, B36, B-4000 Liege, Belgium*

² *Laboratory of Pharmacology and Toxicology, Biopharmaceuticals and Toxicological Analysis Research Team, Faculty of Medicine and Pharmacy, Rabat, Morocco*

³ *Department of analytical chemistry and pharmaceutical technology, CePhaR, Vrije Universiteit Brussel (VUB), Laarbeeklaan 103, 1090 Brussels, Belgium*

med.alaoui@doct.uliege.be

The main goal of this work was to prove the ability of the combination between vibrational spectroscopy techniques and PLS-DA to differentiate between different polymorphic forms of fluconazole in pharmaceutical products. These are mostly manufactured based on fluconazole polymorphic form- II and form- III. These crystalline forms may undergo polymorphic transition during the manufacturing or storage conditions process. Therefore, it is important to confirm if the expected polymorphic form is still present or not.

FT-IR, FT-NIR and Raman spectroscopies were associated to PLS-DA and used to build robust classification models to distinguish between form- II, Form- III and monohydrate form. Based on the results, it is shown that PLS-DA models have a high efficiency to classify various fluconazole polymorphs, with a high sensitivity and specificity. Finally, the selectivity of the PLS-DA models is proven based on the analysis of two samples of itraconazole and miconazole that belong to the same antifungal class as fluconazole. These two samples mimic potential contaminants. Based on the plots of Hotelling T² vs Q residuals, miconazole and itraconazole are significantly considered outliers and rejected.

References:

[1] E. Ziémons, H. Bourichi, J. Mantanus, E. Rozet, P. Lebrun, E. Essassi, Y. Cherrah, A. Bouklouze, P. Hubert, *Journal of Pharmaceutical and Biomedical Analysis Determination of binary polymorphic mixtures of fluconazole using near infrared spectroscopy and X-ray powder diffraction : A comparative study based on the pre-validation stage results*, 55 (2011) 1208–1212. doi:10.1016/j.jpba.2011.02.019.

[2] N.L. Calvo, T.S. Kaufman, R.M. Maggio, *Mebendazole crystal forms in tablet formulations. An ATR-FTIR/chemometrics approach to polymorph assignment*, *J. Pharm. Biomed. Anal.* 122 (2016) 157–165. doi:10.1016/j.jpba.2016.01.035.

Rapid detection of smuggled non-compliant diesel in Moroccan market using mid infrared spectroscopy coupled to a PLS-DA model

Issam Barraa¹ , Mohammed Alaoui Mansouri^{1,2}, Yahia Cherrah¹, Abdelaziz Bouklouze¹

¹*Pharmaceutical and Toxicological Analysis Research Team. Laboratory of Pharmacology and Toxicology, Faculty of Medicine and Pharmacy, University Mohammed V- Rabat- Morocco.*

²*University of Liege (ULiege), CIRM, VibraSanté Hub, Department of Pharmacy, Laboratory of Pharmaceutical Analytical Chemistry, Quartier Hôpital, Avenue Hippocrate 15, B36, B-4000 Liege, Belgium*

issam.barra@gmail.com

In order to detect smuggled diesel samples and discriminate them from authentic products Mid Infrared Spectroscopy was chosen as a quick, non-destructive, less expensive and a very sensitive technique especially when coupled to Chemometric tools. After recording infrared spectra of all diesel samples, the data was projected to Principal Components Analysis (PCA) to visualize relations between samples and even detect if there is any outliers. The classification was carried out by a Partial Least Square Discriminate Analysis model. Results obtained by PCA shows that dissimilarities between the two diesel classes are easy to highlight thanks to the formation of two separable groups on PCA scores plot, using just two components already explained more than 68% of the total variance (PC1 explicated 51,09% and 16,3% for the second).

PLS-DA model allowed good discrimination between the two diesel classes with high specificity and sensitivity and reduced errors, it can also be said that the result is better than that obtained when applying PCA even with two latent variables. Then in this case 62,15% of the total variance was explained (52,32 by LV1 and 9,83 by LV2).

DUTCH SURFACE WATER QUALITY

André van den Doel^{1,2}, Geert van Kollenburg^{1,2}, Lutgarde Buydens¹ and Jeroen Jansen¹

¹Institute of Molecules and Materials, Department of Analytical Chemistry, Radboud University, PO Box 9010 6500 GL Nijmegen, The Netherlands

²TI-COAST, Science Park 904, 1098 XH, Amsterdam, The Netherlands

h.vandendoel@science.ru.nl

The European Water Framework Directive sets goals for the ecological and chemical condition of surface water and ground water [1]. Rijkswaterstaat is part of the Dutch ministry of Infrastructure and the Environment and is responsible for the management of the main waterway networks. At several measurement stations along the major rivers they measure the water that flows past with LC-MS or GC-MS. There is a limited number of target compounds that are screened daily, but it is not feasible to monitor all (possible) contaminants on a daily basis. That is why we use GC-MS fingerprints to analyse the water quality and to identify emerging contaminants. Compared to the univariate thresholds that are currently used our approach has the advantage of taking into account correlations and co-occurrences of compounds. This leads to a more holistic view of the water quality and better identification of processes that influence it.

References

[1] Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy. Official Journal L327, 22/12/2000 p0001-0073 (2000).

Acknowledgement: This research received funding from the Netherlands Organisation for Scientific Research (NWO) in the framework of the Programmatic Technology Area PTA-COAST3 of the Fund New Chemical Innovations. This publication reflects only the author's view and NWO is not liable for any use that may be made of the information contained herein.

LOCAL vs. SVM-BASED STACKING APPROACH FOR PREDICTING PROPORTIONS OF COMPLEX BLENDS IN FOOD PRODUCTS

Beatriz Carrasco¹, Alberto Gil¹, José Ramón Quevedo², Elena Montañés², Guanghui Shen^{3,4}, Vincent Baeten⁴, Juan Antonio Fernández Pierna⁴

¹*Blendhub, Murcia, Spain.*

²*Artificial Intelligence Center, Universidad de Oviedo, Gijón, Spain*

³*China Agricultural University (CAU), China*

⁴*Walloon Agricultural Research Centre (CRA-W), Belgium*

beatriz.carrasco@blendhub.com / j.fernandez@cra.wallonie.be

In this work, two different approaches are studied to characterize complex blends in food products: a Support Vector Machines (SVM)-based stacking approach and a local approach.

Stacking has become a quite common step in machine learning field to make more than one simple approach cooperate in order to obtain more accurate predictions. This work presents a stacking approach, where the output of one of the more simple approaches feeds the input of another forming a stack. Particularly, the proposal consists of using the output provided by a support vector machine approach as input of a support vector regression approach. Both support vector approaches have shown promising performance with high dimensional and noisy data.

A local approach is based 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 model 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. Here a simple local method based on PCA scores is applied.

Both approaches are applied for the characterization and prediction of complex blends used as additives in the food industry.

MULTIMODAL OPTICAL SPECTROSCOPY AND MULTIVARIATE DATA ANALYSIS FOR THE DEMARCATION OF HUMAN BRAIN TUMORS

Mona Stefanakis¹, Edwin Ostertag¹, Anita Lorenz¹, Marc Brecht¹, Karsten Rebner¹

¹*Process Analysis & Technology (PA&T), Reutlingen University, Alteburgstr. 150, 72762 Reutlingen, Germany*

mona.stefanakis@reutlingen-university.de

Glioblastoma multiforme is the most common and most malignant tumor of the human brain. The average long-term survival rate amount to 15 months. Due to the need for a good prognosis a nearly complete resection and a well-defined classification of the tumor is necessary.

We investigate different human brain tumors with various spectroscopic methods in combination with multivariate methods (e.g. principal component analysis). The objective is a label-free demarcation between malignant and normal brain tissue and a classification following the malignancy schema of the World Health Organisation (WHO). The results are compared to the pathologist's assessment.

A multimodal spectral imaging concept combining optical microscopy with multimodal molecular and elastic light scattering spectroscopy is presented. Laterally resolved spectroscopic information representing the chemical fingerprint is recorded together with the elastically scattered light linked to the morphology of the same area of the sample. Different spectroscopic techniques from the UV and VIS range to the middle infrared are applied for label-free identification, characterization, demarcation and classification of brain tumor cross-sections with multivariate methods. The combination of the spectroscopic methods and the fusion of their data can help to get rid of the need for staining and to obtain a higher diagnostic safety for the tumor classification.

Acknowledgement: The financial support by the "Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg" and the "Promotionskolleg IPMB" of the University of Tübingen and the Reutlingen University. The authors especially thank the team of the Neurosurgery Department of Tübingen University Hospitals, headed by Professor Dr. Marcos Tatagiba, for supplying the brain tumor tissue samples and preparing the cross-sections.

COMBINATION OF MCR-ALS AND K-MEANS INFORMATION FOR MASS SPECTROMETRY IMAGING DATA ANALYSIS

S. Mas¹, A. Torro², N. Bec², C. Larroque², P. Martineau², A. de Juan³, S. Marco¹,

¹*Signal and Information Processing for Sensing Systems, IBEC, Baldiri Reixac 4-8, 08028 Barcelona, Catalonia, Spain*

²*IRCM, Institut de Recherche en Cancérologie de Montpellier, INSERM U1194, UM, 208 Avenue des Apothicaires, Montpellier, F-34298, France.*

³*Chemometrics Group. Department of Chemical Engineering and Analytical Chemistry. Universitat de Barcelona. Av. Diagonal, 645. 08028 Barcelona, Catalonia, Spain*

smas@ibecbarcelona.eu

In this study, combination of multivariate curve resolution-alternating least squares (MCR-ALS) and K-means clustering is proposed as a general strategy for the characterization of cancer tissues in matrix-assisted laser desorption ionization-mass spectrometry images (MALDI-MSI).

In tissue-based cancer research, information on the spatial location of the cancer cell populations can be known from the traditional histology. This study proposes for the first time the introduction of this information as a constraint in the MCR resolution of MSI data. The use of this local rank constraint will help to decrease or remove the MCR inherent ambiguity. Moreover, this study aims at showing the advantages of combining a resolution method with histology-based local rank information to enhance the quality of information obtained from the combination of MCR-ALS and K-means methods [1-2].

The great potential of this strategy for the characterization of cancer tissues in MALDI-MSI data is shown on a set of fifteen images corresponding to different tissues of experimental colorectal cancer.

References:

[1] Piqueras, S., Duponchel, L., Tauler, R., & De Juan, A. (2011). *Analytica Chimica Acta*, 705(1-2), 182-192.

[2] Piqueras, S., Krafft, C., Beleites, C., Egodage, K., von Eggeling, F., Guntinas-Lichius, O., Popp, J., Tauler, R., de Juan, A. (2015). *Analytica Chimica Acta*, 881, 24-36.

Acknowledgements:

This work is part of the BEST Postdoctoral Programme, funded by the European Commission under Horizon 2020's Marie Skłodowska-Curie Actions COFUND scheme (Grant Agreement no. 712754) and by the Severo Ochoa programme of the Spanish Ministry of Science and Competitiveness (Grant SEV-2014-0425 (2015-2019)).

**MICROMETRIC MAPPING OF TOTAL ORGANIC CARBON IN LAKE SEDIMENT CORES
COMBINING FUSION OF MULTIREOLUTION HYPERSPECTRAL IMAGES AND PLSR ANALYSIS**

Kévin Jacq^{1,2}, Yves Perrette¹, Bernard Fanget¹, Didier Coquin², Pierre Sabatier¹, Fabien Arnaud¹, Ruth Martinez-Lamas^{3,4}, Maxime Debret³

¹ *University Grenoble Alpes, University Savoie Mont Blanc, CNRS, EDYTEM, 73000 Chambéry, France*

² *University Grenoble Alpes, University Savoie Mont Blanc, Polytech Annecy-Chambéry, LISTIC, 74000 Annecy, France*

³ *Normandie Univ, UNIROUEN, UNICAEN, CNRS, M2C, 76000 Rouen, France*

⁴ *IFREMER, Laboratory Géodynamique et Enregistrement Sédimentaire (LGS), France*

kevin.jacq@univ-smb.fr

Sedimentary cores are used, thanks to their physical, chemical and biological properties, to infer past climate and environment. Sampling methods (millimetre or centimetre) and routine analysis are destructive and non-spatially resolved methods that consume time and material. The use of hyperspectral imaging makes it possible to have micrometric area in each point of the core.

We use two hyperspectral cameras, the VNIR (spectral range: 400-1000nm, spatial resolution: 60µm) and the SWIR (spectral range: 1000-2500nm, spatial resolution: 189µm). Usually each camera is used separately. The goal of this work is to show the combination of sensor increase performance predictions. A pixel-level data fusion based on the ARSIS method [1] is applied to create a unique cube at the optimal resolution. This new cube can be used with a usual PLSR method to develop a model for the total organic carbon.

Three cores from the lakes Le Bourget, Annecy and Geneva (Western Alps) are been tested (approximately 60cm long and 9cm width each). For both samples, the results show an increase prediction performance rather than data used separately. In the unique cube, the selected wavelength corresponds to those selected by each sensor. Although the analyzes were performed on bulk samples (5 mm x 90 mm x 45 mm slices), the prediction model provides access to the mapping of the surface with a micrometric resolution (the 60µm pixel can be interpreted as relevant information).

Reference

[1] Ranchin, T., & Wald, L. (2000). Fusion of high spatial and spectral resolution images: the ARSIS concept and its implementation. *Photogrammetric Engineering and Remote Sensing*, 66(1), 49–61.

FLEXIBLE VARIABLE SELECTION USING ELASTIC NET REGRESSION

Cannon Giglio¹, Steven D. Brown²

¹*Department of Chemistry, Dalhousie University, P.O. Box 15000, Halifax, NS, B3H 4R2, Canada*

²*Department of Chemistry and Biochemistry, University of Delaware, 163 The Green, Newark, DE 19716, USA*

cannon.giglio@dal.ca

Variable selection is frequently used to remove uninformative variables in multivariate data such as spectra. The most commonly used variable selection methods, such as variable importance in projection (VIP), are based on the partial least squares (PLS) regression. These methods assume a valid PLS model, which is often not the case. A potential alternative is to use the elastic net (EN) regression, which can select variables automatically by using a mixed-norm penalization function. An EN regression can select groups of correlated variables, and can select either sparse or non-sparse sets of variables. However, when using the EN regression on spectra, there is often a tradeoff between variable selection quality and prediction performance. In the present work, the use of the elastic net to select variables, followed by conventional PLS regression on the selected variables (EN-PLS), has been investigated. Variable selection using EN-PLS was compared with that from elastic net regression, sparse partial least squares regression (SPLS), variable importance in projection (VIP), and from selectivity ratio (SR) selection on two datasets of visible/near-infrared spectra. The EN-PLS method was found to give similar prediction performance and interpretability when compared with SR, and improved performance when compared with the other variable selection methods.

Acknowledgement: This work was supported by the United States National Science Foundation, grant number 1506853. CG gratefully acknowledges partial summer support from the University of Delaware Summer Scholars Program.

EVALUATION OF ACIDITY STRENGTHS ON ELECTRONIC GROUND AND EXCITED STATES OF 2-(2'-HYDROPHENYL)BENZOTHAZOLE USING EXCITATION-EMISSION FLUORESCENCE MATRIX AND MULTIVARIATE CURVE RESOLUTION WITH ALTERNATING LEAST SQUARES (MCR-ALS)

Willian Francisco Cordeiro Dantas¹, Luis Gustavo Teixeira Alves Duarte¹, Fabiano Severo Rodembusch², Teresa Dib Zambon Atvars¹ and Ronei Jesus Poppi¹

¹*Institute of Chemistry, University of Campinas, P.O. Box 6154, 13083-970, Campinas, SP, Brazil*

²*Grupo de Pesquisa em Fotoquímica Orgânica Aplicada, Institute of Chemistry, Federal University of Rio Grande do Sul, Bento Gonçalves Avenue 9500, Porto Alegre, Brazil*

willianfcd@gmail.com

In this work, we present a comprehensive procedure using excitation-emission fluorescence matrix (EEM) and multivariate curve resolution with alternating least squares (MCR-ALS) [1] for the analysis of the deprotonation balances from 2-(2'-hydroxyphenyl)benzothiazole (HBT) [2] in different concentrations of hydrogen ions. For acidity variation the Hammett acidity and pH scale were applied using several buffer solutions, allowing the identification of HBT structural variation on the electronic ground and excited states.

It was demonstrated that the presented methodology is also suitable for kinetic analysis of deprotonation balance on the excited state, since it can indicate the relative concentrations profiles of HBT prior and after ionization on the S_1 . MCR-ALS was performed with augmented matrix and using the constraints non-negativity, unimodality, closure and multiway. Moreover, it was possible the construction of concentrations estimation of all species on both electronic ground and excited states.

The combination of EEM analysis and MCR-ALS provided promising results for acidity strengths prediction on the S_1 and S_0 , which can result on a simple strategy to analyze the enhancement of acidity strength due to light absorption on organic aromatic dyes.

References

[1] J. Jaumot, A. de Juan and R. Tauler, *Chemom. Intell. Lab. Syst.*, 2015, 140, 1–12.

[2] J. K. Dey, S. K. Dograt, *J. Photochem. Photobiol. A: Chem.*, 1992, 66, 15-31.

Acknowledgement: The authors would like to thank the CAPES, CNPq and FAPESP for financial support.

EXPLORATIVE STUDY OF THE RED FRUIT JUICES FLUORESCENCE USING EXCITATION-EMISSION SPECTROSCOPY AND CHEMOMETRICS

Ewa Sikorska, Katarzyna Włodarska

Faculty of Commodity Science, Poznań University of Economics and Business, al. Niepodległości 10, 61-875 Poznań Poland

ewa.sikorska@ue.poznan.pl

In the present study we investigate the fluorescence of juices obtained from different red fruits using excitation-emission spectroscopy and multivariate data analysis. The red fruits are rich in phytochemicals, including phenolic antioxidants [1]. The intense colour of these fruits is due to the presence of anthocyanins. Phenolic compounds present in these fruits play a vital role in the prevention of diseases and in the health promoting properties. Anthocyanins present a wide range of biological activities including antioxidant, antimicrobial, anti-inflammatory, and anti-carcinogenic activities. Some of phenolic compounds are fluorescent. Thusly, the fluorescence spectroscopy with its multidimensional character, high sensitivity and selectivity, may be well suited for study properties of juices related to their antioxidants content and antioxidant capacity.

Juices obtained from red fruits, including: strawberry, red raspberry, black currant and chokeberry were the subject of this study. Commercial juices selected for study covered various categories of products available on the market. The total fluorescence spectra (excitation-emission matrices, EEM) were recorded for each juice. The tentative assignment of the specific emission bands based on the comparison to fluorescence of respective standards. The exploratory analysis was performed using Principal Component Analysis (PCA) and Parallel Factor Analysis (PARAFAC). Observed differences in fluorescence characteristics of juices were ascribed to different compositions and concentrations of phenolic compounds. This preliminary study shows the potential of using fluorescence for quality and authenticity assessment of red fruit juices.

References

[1] G. A. Manganaris, V.Goulas, A. R. Vicente, L. A. Terry, (2014). *J. Sci. Food Agric.*, 94, 825-833 (2014)

Acknowledgement: Grant 2016/23/B/NZ9/03591 from the National Science Centre, Poland, is gratefully acknowledged.

EVALUATION OF CHEMICAL PROPERTIES OF APPLE JUICES RELATED TO TASTE USING NEAR INFRARED SPECTROSCOPY AND CHEMOMETRICS

Ewa Sikorska, Katarzyna Włodarska

Faculty of Commodity Science, Poznań University of Economics and Business, al. Niepodległości 10, 61-875 Poznań Poland

ewa.sikorska@ue.poznan.pl

The aim of this study was to test feasibility of the NIR spectroscopy in developing the calibration models for predicting the chemical parameters related to the sweet and sour flavour of apple juices. The apple juice is one of the most consumed juices due to its pleasant flavour and beneficial health effects. The taste of juice is an important driver for consumer acceptance. For apples, it was demonstrated that sweet and sour flavours are related to the soluble solids content (SSC), titratable acidity (TA), and even more closely to the ratio of SSC to TA (SSC/TA) [1]. This ratio is used as an index of sensory acceptability of the fruit taste. The discrimination between juices with different SSC/TA ratio may be of interest from both, manufacturers and consumers perspectives, replacing the tedious sensory analysis with instrumental methods.

Near infrared (NIR) spectra were recorded for commercial apple juices. The partial least squares regression (PLS-R) was used to develop the model for determination of SSC/TA ratio, and partial least squares discriminant analysis (PLS-DA) was used to discriminate between juices with low and high SSC/TA ratio. Various spectral pre-processing methods were used for models optimization. The optimal spectral variables were chosen using interval PLS (iPLS) and jack-knife based method. For limited number of samples the PLS-R analysis was performed using NIR spectral data and scores for sweet and sour flavour obtained from sensory analysis. The present results show the potential of the NIR spectroscopy for screening for the important quality parameters of apple juices.

References

[1] N. Abu-Khalaf, B. S. Bennedsen, *Int. Agrophys.*, 18, 203-211 (2004)

Acknowledgement: Grant 2016/23/B/NZ9/03591 from the National Science Centre, Poland, is gratefully acknowledged.

CHEMOMETRIC PROFILING OF CANNABIS STRAINS TO EVALUATE THEIR PHYTOCHEMICAL DIVERSITY

Elizabeth M. Mudge^{1,2}, Susan J. Murch¹, Paula N. Brown^{1,2}

¹*University of British Columbia, Kelowna, BC*

²*AC Institute of Technology, Burnaby, BC*

Paula_Brown@bcit.ca

Flowers of *Cannabis sativa* L. are used medicinally for pain management, nausea, appetite regulation, anxiety, depression, and spasticity. The cannabinoids, delta-9-tetrahydrocannabinol (THC) and cannabidiol (CBD) are the most studied phytochemicals in Cannabis, while strains with similar THC/CBD profiles can exhibit significantly different pharmacological effects. There may be up to 30,000 unique phytochemical metabolites present in a single plant leaf, so chemometric approaches are necessary to fully elucidate the phytochemical diversity and variation of Cannabis strains. Targeted and untargeted metabolomics were there employed for strain characterization. Over thirty strains of Cannabis were obtained from licensed producers in Canada and subjected to cannabinoid analysis by HPLC-UV and untargeted metabolomics using ¹H NMR. The strains were categorized into five different groups based on varying THC and CBD contents. Chemometric analysis showed significant overlap across the classes of strains with separation in the first principal component (PC) primarily impacted by the THCA and CBDA contents while the second and third PCs were associated with untargeted metabolites. Unsupervised clustering identified several groups that were not dependent on the cannabinoid classes, where regions of highest variance indicated the presence of unique metabolite clusters. This demonstrates that metabolomic approaches to evaluate Cannabis strains may be better suited than other approaches to improve product quality and explore the pharmacological variation of medicinal Cannabis.

USE OF MOVING WINDOW PCA OF CHEMICAL PROFILES FOR AUTHENTICATION OF BOTANICAL PRODUCTS

Jamie Finley¹, Paula Brown¹

¹*BC Institute of Technology, Burnaby, BC*
Paula_brown@bcit.ca

The authenticity of botanical ingredients in marketed products has been the focus of widespread concern in the natural product and dietary supplement industry. Classically, plants are identified by physical examination of minimally processed biomass for diagnostic macroscopic and microscopic features. However, botanical ingredients in modern supply chains are so highly processed that diagnostic anatomical features have been destroyed or removed. The use of chemical profiling coupled with chemometric analyses shows promise as an approach for establishing the authenticity of botanical products. A multivariate moving window PCA approach based on the PC score residuals has been used to evaluate adulteration of American ginseng roots with leaves and goldenseal from common adulterants. The model computes standard residuals within windows defined in clean spectra and applies PCA to those residuals to determine if they are within the PCA model limits. MWPCA detected leaf adulteration as low as 2% in ginseng roots, and 15% *Rumex crispus* in goldenseal roots. This represents a significant improvement compared with traditional regression models for detection of low level of contamination.

PIXEL-BASED CHEMOMETRIC ANALYSIS OF LC-MS ANALYZED SAMPLES FROM AN INDUSTRIAL ENZYME PRODUCTION: CHARACTERIZATION OF DIFFERENT PRODUCTS AND INVESTIGATING IF THERE IS A LINK BETWEEN THE SMALL MOLECULE FINGERPRINT AND PRODUCT QUALITY?

Sofie B. Knudsen^{1,2}, Nikoline J. Nielsen¹, Johan Qvist², Kim B. Andersen², Jan H. Christensen¹

¹*Analytical Chemistry Group, Section for Environmental Chemistry and Physics, Department of Plant and Environmental Sciences, University of Copenhagen, Thorvaldsensvej 40, 1871 Frederiksberg C, Denmark.*

²*Department of Liquid Formulation, Novozymes A/S, Krogshøjvej 36, 2880 Bagsvaerd, Denmark.*

sofiebk@plen.ku.dk

Physical stability of liquid enzyme products is one of the major quality parameters in enzyme production and unfortunately also one that may be influenced by even small variation in the dry matter background. A measure of steady-state physical stability can take weeks to obtain, and hence it can be a slow process to detect poor quality by e.g. visual evaluation of precipitate in the sample. This study aims to identify small molecule endpoints (<1200 Da) that could be indicators of the physical stability of a product, and in that way indicate poor quality products in an early stage. 10 different products have been analyzed by LC-TOFMS in positive and negative mode to distinguish the difference between samples. Furthermore, a set of one product were exposed to forced degradation to distinguish between non-degraded and degraded products (good and poor quality). Both datasets, including facilitator samples, were processed by pixel-based chemometric analysis[1]. Different baseline removal, alignment and normalization procedures were tested and finally variable selection was performed on a subset of relevant features. Different enzyme product could be discriminated by use of a pixel-based chemometric approach, as could non-degraded and degraded products.

[1] Furbo, S., et al. (2014). "Pixel-Based Analysis of Comprehensive Two-Dimensional Gas Chromatograms (Color Plots) of Petroleum: A Tutorial." *Analytical Chemistry* 86(15): 7160-7170.

Acknowledgement: Thanks to the Innovation Fund Denmark for financial support.

CHEMICAL IMAGE PROCESSING BY DISCRETE GABOR TRANSFORM COMBINED WITH GAUSSIAN APODIZATION FACTOR ANALYSIS FOR RESOLUTION OF THE EMBEDDED PEAKS IN COMPLICATED HYPER SPECTRAL DATA

Amir Hossein Alinoori^{1,2}

¹*Institute of Materials and Energy, Iranian Space Research Center, I.R. Iran 1*

²*Department of Analytical Chemistry, Faculty of Chemistry, University of Kashan, Kashan, I.R. Iran 2*
ahalinoori@yahoo.co.uk

Discrete Gabor transform combined with Gaussian apodization factor analysis (GAFA) has been developed as an enhanced algorithm to assess the purity of hyper spectral data in decomposed data in time-frequency domain by short time Fourier transform with a Gaussian window. In GAFA method, submatrices in time-frequency domain are extracted by Gaussian apodization moving window by weighting the fixed-size moving window via Gaussian formula. Therefore, each submatrix mainly characterizes an spectrum and by performing factor analysis on this Gaussian weighted submatrix, the number of principal components for each evaluated spectrum, is determined by singular value decomposition (SVD). This precise and quick determination of a rank map is successfully used for extract pure components from hyper spectral data in time-frequency domain. An algorithm based on GAFA was applied to resolve different types of overlapped simulated and real complex hyper spectral data. This algorithm find spectra of pure component with GAFA one by one and eliminate obtained components in time-frequency domain and search for next pure component spectra until all of the components are determined. On the next steps reconstruction of data from in time-frequency domain to time domain was applied.

GAUSSIAN APODIZATION FACTOR ANALYSIS FOR RESOLUTION OF THE EMBEDDED PEAKS IN ION TRAP TIME OF FLIGHT ION MOBILITY SPECTROMETRY

Amir Hossein Alinoori¹, Saeed Masoum¹

¹*Department of Analytical Chemistry, Faculty of Chemistry, University of Kashan, Kashan, I.R. Iran 1*
ah.alinoori@grad.kashanu.ac.ir

A unique portable ion trap ion mobility spectrometry (IMS) was designed and fabricated. This design consists of fast sampling and injecting system (FSI) coupled with ion trap IMS. FSI design provides fast and easily applied method for real time injection of air sample accompanied by particles with no additional sample preparation. Required analysis cycle time for each run is less than 15 sec. The high sample load and sharp injection with the fast separation by programmed thermal desorption decrease the peak widths, and improve detection limits. In addition to coupling of FSI to IMS detector and other consideration in hardware design, miniaturization of IMS cell was done to increase sensitivity and selectivity and warm up time. Data analysis was enhanced with adapted Gaussian apodization factor analysis (GAFA) as a multivariate curve resolution algorithm. This homemade customized instrument is an alternative to other time consuming technologies for monitoring of organic particles in air samples without sample preparation. GAFA has been developed as an enhanced algorithm to assess the purity of ion trap-IMS data. In GAFA method, submatrices are extracted by Gaussian apodization moving window by weighting the fixed-size moving window via Gaussian formula. Therefore, each submatrix mainly characterizes a spectrum and by performing factor analysis on this Gaussian weighted submatrix, the number of principal components for each evaluated spectrum, is determined by singular value decomposition (SVD). This precise and quick determination of a rank map is successfully used for extract pure components from ion trap-IMS data.

FOUR-DIMENSIONAL QUANTITATIVE STRUCTURE-ACTIVITY ANALYSIS OF 1,4-NAPHTHOQUINONES DERIVATIVES TESTED AGAINST HL-60 HUMAN PROMYELOCYTIC

Márcia Miguel Castro Ferreira, Maria Cristina Andreazza Costa, Pedro Oliveira Mariz Carvalho

Laboratory for Theoretical and Applied Chemometrics, Institute of Chemistry, University of Campinas
Campinas-SP, Brazil, 13083-970
marcia@iqm.unicamp.br

A series of 45 1,4-naphthoquinones derivatives, tested against human HL-60 leukemic cells and previously analyzed by two-dimensional structure-activity relationships [1], was submitted to a 4D-QSAR study, with the main objective to investigate the possible interactions with the receptor. Dynamic molecular (MD) calculations [2] were applied to the 3D optimized structures (Gaussian 3.0, DFT-B3LYP, 6-31G**), in order to obtain a conformational ensemble profile (CEP) for each compound. The CEP were aligned and the field descriptors which are the electrostatic and van der Waals interaction energies (Coulomb and Lennard-Jonnes potentials) were calculated through LQTAgrid module from LQTA-QSAR program [3]. The selected descriptors were used to build the regression model using PLS regression method.

The regression model (eq.1) with 2 factors presented $R^2 = 0.90$ and $SEC = 0.23$; $Q^2 = 0.88$ and $SEV = 0.26$. For external validation (eight molecules on test set), $R^2 = 0.75$ and $SEP = 0.25$.

$$pIC50 = 0.27 C140 - 0.28 L143 + 0.38 C225 - 0.21 L133 - 0.40 C160 \quad (\text{eq. 1})$$

The positive Coulomb coefficients (C) indicate that the biological activity is favored by polar substituents in these regions, while the negative Lennard-Jones coefficients (L) are indicating that the biological activity increase with bulky groups.

[1] Costa, M.C.A.; Ferreira, M.M.C; SAR QSAR Environ. Res. 2017, 28, 325-339.

[2] Lindahl, E.; Hess, B.; van der Spoel, D.; Journal of Molecular Modeling 2001, 7, 306-317.

[3] Martins, J. P. A.; Barbosa, E. G.; Pasqualato, K. F. M.; Ferreira, M. M. C.; J.Chem. Inf. Model. 2009, 49, 1428.

Acknowledgements: CNPq, FAPESP.

MULTIVARIATE ACCELERATED SHELF-LIFE TESTING FOR DETERMINING THE SHELF-LIFE OF PAINTS

Márcia Miguel Castro Ferreira¹, Nelson Mituo Matsumoto²

¹Laboratory for Theoretical and Applied Chemometrics, Institute of Chemistry, University of Campinas Campinas-SP 13084-971, P. O. Box 6154, Brazil

²PPPG Industrial do Brasil Tintas e Vernizes Ltda, Sumaré - SP - Brasil

marcia@iqm.unicamp.br

In this study we propose a multivariate accelerated shelf-life approach for determining the shelf-life of tinting paste and consequently for industrial coatings. The pigments used were red, yellow, blue, white and black. The tinting pastes were prepared and properly stored at 0 °C, 25 °C and 60 °C. The colorimetric parameters ΔL , Δa and Δb were measured in a spectrophotometer (X-Rite) every 15 days, for 460 days. According to the univariate approach, the kinetic charts would be built and by evaluating the reaction velocity profile it would be possible to determine the reaction order and then convert the results from accelerated tests to actual market conditions. For this, a proportionality constant between different storage temperatures should be determined. On the other hand, the multivariate method is based on a Principal Component Analysis (PCA), in which the scores of the time-related components, instead of individual responses, are taken for estimating the multivariate kinetic rate constants (k_m), the multivariate acceleration factor (α_m) and the multivariate activation energy (AE_m). Finally, the multivariate parameters from accelerated tests are converted to actual market conditions.

Shelf-life monitoring in room temperature was also performed to validate the method. Shelf-life obtained for all colours by the PCA accelerated method were quite consistent with the shelf lives obtained under standard conditions. The multivariate accelerated shelf-life test has proven its value by reducing the kinetic study to a single variable, whilst giving information on what are the main parameters affecting product degradation in a direct fashion.

Acknowledgements: CNPq, FAPESP.

A PIGMENT-BASED TAXONOMIC DISCRIMINATION ANALYSIS FOR THE IN-SITU CLASSIFICATION OF (HARMFUL) ALGAL BLOOMS

Zieger S. E.¹, Mistlberger G.¹, Troi L.¹ and Klimant I.¹

¹*Graz University of Technology, Institute of Analytical Chemistry and Food Chemistry (TU Graz, ACFC), Stremayrgasse 9, 8010 Graz, Austria*

zieger@tugraz.at

Early stage identification of (harmful) algal blooms (HABs) has gained in significance for monitoring systems over the years. HAB pose a serious threat to marine and human life, having an adverse effect on their surrounding ecosystem. Various biotoxins produced by algae and accumulated across the food web, can lead to massive fish kill and human disorders.

Facing these problems, various approaches for in-situ classification and early-stage identification have been developed. Among them, pigment-based taxonomic classification of HABs is one promising principle for in-situ characterization of bloom compositions, although it is yet underestimated in marine monitoring programs. However, to demonstrate the applicability and importance of this powerful approach for monitoring programs, we developed a miniaturized and low-cost multi-wavelength fluorometer for in situ detection of relevant algal groups.^[1]

Based on chemotaxonomic principles, we characterized algae of eight different phytoplankton classes with a spectro-fluorometer and by the means of our fluorometer. Relevant marker pigments were investigated and significant spectral differences were extracted as key features. These key features were used within Fisher's linear discriminant analysis for reliable differentiation at order level. A comprehensive investigation of the system performance of axenic algal cultures was conducted in terms of standard statistical measures and independent figures of merits. The focus of attention throughout our studies lies on the reliable discrimination of cyanobacteria and dinoflagellates from co-occurring algae. Besides this, robustness evaluations of the algorithm were assessed during growth and under abnormal light conditions. The separation capability of the linear discriminant analysis was further successfully examined in mixed algal suspensions.

References

[1] Zieger, S. E. et al. Compact and low-cost fluorescence based flow-through analyzer for in situ quantification and early-stage classification of toxic algal blooms. *Environ. Sci. Technol.* (submitted), (2018).

HIGH-THROUGHPUT AUTHENTICATION OF CANNABIS EXTRACTS BY ULTRAVIOLET MICROPLATE READER AND MULTIVARIATE CLASSIFIERS

Peter Harrington, Zewei Chen

Center for Intelligent Chemical Instrumentation, Department of Chemistry and Biochemistry, Ohio University, Athens, OH 45701, USA
peter.harrington@ohio.edu

Authentication of *Cannabis* samples from different cultivations is increasingly important since *Cannabis* has been used for medical purposes in the recent decade. Two samples sets of botanical extracts were studied. The first set is referred to as *Cannabis* that contained plant material from the *sativa*, *indica*, and hybrids of the two species. The second set, contained extracts from the variety of *Cannabis sativa* with low tetrahydrocannabinol (THC) concentrations (<0.2%) and is referred to as hemp. An ultraviolet microplate reader provides a cost-effective and high-throughput method for classifying 15 *Cannabis* and 20 hemp extracts in this study. By coupling with five multivariate classifiers, i.e. fuzzy rule-building expert system (FuRES), super partial least squares-discriminant analysis (sPLS-DA), support vector machine (SVM) and two tree-based support vector machines (SVMtreeG and SVMtreeH), good classification rates were achieved from the evaluation of bootstrapped Latin partitions. For the *Cannabis* extracts, SVMtreeG yielded the best performance and the classification rate is high as 99.1% for the extract/solvent ratio of 1:10 and 97.1% for the ratio of 1:20. The results from a matched sample t-test indicates that the classification for the 15 *Cannabis* extracts can be affected by different dilution ratios. For the hemp extracts, SVM classifier performed the best with 97.4% of classification rate. These results demonstrate that ultraviolet microplate reader coupled with multivariate classifiers can be used as a high-throughput and cost-effective approach to authenticate *Cannabis* and hemp extracts from different cultivations.

NIR-HYPERSPECTRAL IMAGING TO VISUALIZE THE WETTING AND DRYING CYCLE IN CYANOLICHENS

Cristina Malegori, Paola Malaspina, Paolo Oliveri, Paolo Giordani, Monica Casale

Department of Pharmacy, University of Genoa, Viale Cembrano, 4 -14148, Genoa, Italy
malegori@difar.unige.it

Lichens are symbiotic associations between a fungal partner and photosynthetic partners which is most often either a green alga or a cyanobacterium. Lichens, as poikilohydric organisms, cannot actively regulate their water content, so that they are completely dependent on the atmosphere for their water uptake. Thus, lichens are desiccation-tolerant organisms that can tolerate deep desiccation periods but, as water becomes available again, they do recover their metabolic process. Cyanolichens can achieve positive net photosynthesis when directly exposed to liquid water. The time necessary to fill the internal water storage depends on the thickness of the lichen thallus: thick lichens generally need a longer time, but they can also retain water for extended periods after hydration events. Until now, numerous research works have dealt with the effects of thallus hydration on the lichen metabolism, but only a limited number of them have investigated on the intrathalline water localization. In this work, we aimed to visualize the pattern of distribution of water during wet and drying cycle in cyanolichens. Authors analyzed thalli of four cyanolichen species characteristic of environments with a high relative humidity but with different thallus structures and different thallus thickness, using hyperspectral imaging (NIR-HSI) system operating in the 1000-2500 nm spectral region. Lichens were hydrated with deionized water and the desiccation process was followed by NIR-HSI until the complete dehydration of the thallus. The promising results show that NIR-HSI can be regarded as an effective analytical technique for mapping and chemically understanding the wetting and drying cycle in cyanolichens.

HOW SIGNAL PRE-PROCESSING AFFECTS INTERPRETATION OF CHEMOMETRIC PROCEDURES – REMARKS AND SUGGESTIONS

Paolo Oliveri¹, Cristina Malegori¹, Monica Casale¹

¹*University of Genova – Department of Pharmacy (DIFAR) – Viale Cembrano, 4 – 16147 Genova – Italy*
oliveri@difar.unige.it

Tools for signal pre-processing include a wide number of mathematical transformations generally aimed at minimising unwanted effects (both random and systematic), with the result of improving data quality and, consequently, converting raw data to exploitable useful information.

Examples of common pre-processing corrections are the standard normal variate transform (SNV) – or row autoscaling – and derivatives of different orders, usually applied in combination with smoothing, to overcome the enhancement of random noise, which is usually characterised by high-frequency slope variations.

Along with the desired corrections, application of mathematical transforms may produce undesired secondary effects. In particular, some transforms may introduce artefacts; other may confound interpretation of the final outcomes of the whole signal processing process – a risk that is often underestimated.

In the present work, desired and undesired effects of the most commonly applied signal transforms are reviewed. Moreover, their effects on the interpretation of the outcomes of common chemometric methods (such as principal component analysis – PCA) are critically described, and efficient strategies to overcome these hurdles are presented.

Acknowledgement: Financial support by the Italian Ministry of Education, Universities and Research (MIUR) is gratefully acknowledged – Research Project SIR 2014 “Advanced strategies in near infrared spectroscopy and multivariate data analysis for food safety and authentication”, RBS14CJHJ (CUP: D32115000150008).

CONTRAST ENHANCEMENT AND TOTAL VARIATION: A COMPARISON BETWEEN TWO SPATIAL CONSTRAINTS IN MCR-ALS OF HYPERSPECTRAL IMAGES

Dario Cevoli¹, Raffaele Vitale^{1,2}, Siewert Hugelier¹, Cyril Ruckebusch¹

¹Laboratoire de Spectrochimie Infrarouge et Raman - LASIR CNRS - UMR 8516, Université de Lille, 59000, Lille, France

²Molecular Imaging and Photonics Unit, Department of Chemistry, Katholieke Universiteit Leuven, Celestijnenlaan 200F, B-3001, Leuven, Belgium
rvitale86@gmail.com

MCR-ALS [1] has demonstrated to be a valuable tool to resolve hyperspectral images (HSI). However, traditional MCR-ALS analysis of an HSI requires the unfolding of the hyperspectral data cube into a two-way array, which causes the loss of information on adjacency between pixels. In such cases, exploiting the spatial structure of HSI to improve the corresponding MCR-ALS solutions is unfeasible [2]. Several approaches have been developed to overcome this issue. The angle constraint (also known as contrast enhancement constraint) proposed by Winding et al. [3, 4] takes advantage from the duality of the spectral and spatial domain to enhance the selectivity of one of the two. Alternatively, a refolding step can be added in the least squares loop of the MCR-ALS decomposition [2]. This allows image processing constraints like sparseness [5], segmentation [2] or total variation, [6, 7] to be employed for improving the visualisation of regions of interest with sharp edges.

The aim of this work is to compare the outcomes resulting from the MCR-ALS analysis of hyperspectral images when contrast enhancement and total variation constraints are applied. Three simulated and a real dataset are here investigated, with the aim of exploring different experimental scenarios in which their effect is tested. The resolution of each dataset is discussed and the pros and cons of the two approaches are highlighted. Particular attention is given to how overlap among components (both in the spectral and spatial domain) can affect the possibility of achieving a physico-chemical meaningful MCR-ALS solution.

References

- [1] R. Tauler, A. Smilde, B. Kowalski, *J. Chemometr.*, 9, 31-58 (1995)
 - [2] S. Hugelier, O. Devos, C. Ruckebusch, *J. Chemometr.*, 29, 557-561 (2015)
 - [3] W. Windig, M. R. Keenan, *Appl. Spectrosc.*, 65, 349-357 (2011)
 - [4] W. Windig, J. M. Shaver, M. R. Keenan, B. M. Wise, *Chemometr. Intell. Lab.*, 117, 159-168 (2012)
 - [5] S. Hugelier, S. Piqueras, C. Bedia, A. de Juan, C. Ruckebusch, *Anal. Chim. Acta*, 1000, 100-108 (2018)
 - [6] L. Rudin, S. Osher, E. Fatemi, *Physica D*, 60, 259-268 (1992)
 - [7] S. Hugelier, R. Vitale, C. Ruckebusch, *Appl. Spectrosc.*, 73, 420-431 (2018)
-

PENALTY OPTIMIZATION IN SPARSE DECONVOLUTION OF SUPER-RESOLUTION FLUORESCENCE IMAGES

Hugelier Siewert¹, Ahmad Mohamad¹, Vitale R.^{1,2}, Ruckebusch Cyril¹

¹Université de Lille, LASIR, F-59655 Villeneuve d'Ascq, Cedex, France.

²KU Leuven, Laboratory of Photochemistry and Spectroscopy, B-3001 Heverlee, Belgium.

siewert.hugelier@univ-lille.fr

SPIDER [1] is a super-resolution image deconvolution approach for high density fluorescence microscopy images. It applies penalized least squares regression with an L0-norm penalty to reach a spatial resolution beyond the diffraction limit of light. In this framework, a parameter λ has to be tuned to balance between the goodness of fit (LS term) and the number of emitters found (penalty term) in the final super-resolved image. As of writing, no automatic procedure to select this parameter is available due to the non-convex nature of the optimization, which results in it being tuned manually via visual inspection.

In this work, we propose a heuristic approach for the optimization of the trade-off parameter λ . We look for λ values that would correspond to the minimum of the Sum of the Normalized Terms (SNT), in the penalized least squares loss function. This heuristic approach has been tested on simulations and turns out to provide exceptional results. We validated it by investigating the robustness of the parameter selection in situations with low to high noise, at different emitter density (ranging from 0.5 – 15 μm^{-2}). The obtained value gives the best relative balance between a low reproduction error and the smallest number of emitters; one tries to find the sparsest solution that gives the best reproduction of the original signal. Furthermore, the approach has also been applied to experimental data sets (i.e. a HEK293-T cell, of which the mitochondria are labelled with the fluorescent protein DAKAP-Dronpa), offering an easy and quick decision criterion for the penalty parameter. Additionally, it reveals an interesting evolution of the penalty parameter with time – which can be related to the fluorophore bleaching – that provides an additional validation of the method. In all, the proposed method is a valuable criterion for the optimization of the penalty parameter and can be used for automatic selection in real experimental situations.

[1] Hugelier et al., Sci. Reports, 2016, **6**. 21413.

THE CRAMER-RAO LOWER BOUND ON THE RESOLUTION OF PEAK SHAPED SIGNALS

L. Fernandez^{1,2}, S. Mas^{1,3}, Maria Padilla², A. Pardo², S. Marco^{1,2}

¹*Signal and Information Processing for Sensing Systems, Institute for Bioengineering of Catalonia (IBEC), The Barcelona Institute of Science and Technology, Baldiri Reixac 10-12, 08028 Barcelona Spain.*

²*Department of Electronics and Biomedical Engineering, University de Barcelona, Martí i Franqués 1, 08028, Barcelona, Spain.*

³*Department of Chemical Engineering and Analytical Chemistry, University of Barcelona, Martí I Franqués 1, 08028, Barcelona, Spain.*

lfernandez@ibecbarcelona.eu

The resolution of overlapping peak shaped signals is a basic problem in analytical chemistry, since most of the instrumentation techniques provide peak-shaped signals [1]. In estimation theory, the Cramer-Rao inequality[2] provides a lower bound on the variance of unbiased estimators of parametric deterministic signals in the presence of noise. We consider that the empirical signal can be modeled by the two Gaussian functions with unknown parameters $A_{1,2}$ (the peak height), $x_{1,2}$ (the position of the center of the peak), and $\sigma_{1,2}$ (a measure of the peak width), plus noise assumed to be Additive White Gaussian Noise (AWGN). Other peak shapes are possible in the same formalism. Using this framework, we propose a method to determine the minimum distance between two peak shaped signals so that their positions can be considered statistically different at a certain risk level. We compute the Cramér-Rao bound (CRB) for all the unknown parameters but for the present work we are interested in the variance-covariance of the position of the peaks. Assuming these estimators are gaussian distributed and using as null-hypothesis that both peaks have the same position we can calculate the minimum distance to reject the null hypothesis at a certain risk level. In other words, this minimum distance is the minimum peak separation to resolve the presence of two distinct peaks in the empirical signal. We will show how this minimum separation depends on the relative peak intensities, the peak widths and the noise variance.

References

- [1] Danzer K (2007) Analytical chemistry : theoretical and metrological fundamentals. Springer.
- [2] Kay SM (1993) Fundamentals of statistical signal processing. Prentice-Hall PTR.

Acknowledgement: This work was partially funded by the Spanish MINECO program, under grant TEC2014-59229-R. IBEC is a member of the CERCA Programme/Generalitat de Catalunya. CIBER-BBN is an initiative of the Spanish ISCIII. Additional financial support has been provided by the Institut de Bioenginyeria de Catalunya (IBEC).

MACHINE LEARNING METHODS FOR FT-IR SPECTRA TO PREDICT SOLUBILITIES OF ROCK SAMPLES

Pál Péter Hanzelik¹, Szilveszter Gergely², Csaba Gáspár^{3,4}, László Győry¹

¹*Exploration & Production Applied Materials Technology & Labs, MOL Group, Október huszonharmadika utca 18., H-1117 Budapest, Hungary*

²*Department of Applied Biotechnology and Food Science, Budapest University of Technology and Economics, Szent Gellért tér 4., H-1111 Budapest, Hungary*

³*Department of Telecommunication and Media Informatics, Budapest University of Technology and Economics, Magyar tudósok körútja 2., H-1111 Budapest, Hungary*

⁴*Dmlab Ltd, Aradi utca 8-10, H-1062 Budapest, Hungary*

pphanzelik@mol.hu

Interests in use of chemometric and data science methods for laboratory techniques have grown rapidly over the last 10 years, for the reason that they are cheaper and faster than traditional analytical methods of material testing.

This study uses 888 rock samples collected from the exploration and production sector of oil industry. Based on the Fourier-transform infrared (FT-IR) spectra of these rock samples their solubility predictions have been developed and investigated with nine methods including both linear and non-linear ones. These programs that were written by the authors are based either on commercial applications or open source libraries. The investigation starts with spectral data pre-processing carried out by standard normal variate (SNV), baseline correction and feature selection methods creating the feature set for all machine learning (ML) applications. The accuracy of predictions have been evaluated with mean squared error as a performance metric for each investigated method.

The comparisons of predicted values to real data of test samples have shown that mineral solubility in acids can be well predicted in the range of the uncertainties of real laboratory measurements, therefore it can be used to improve the response time of these investigations and reduce the risk in industrial applications. In those cases, where the unknown samples have got some out of the range features, the limitations in the accuracy of predictions have become clear. This finding further emphasizes the need for data base building efforts, so that the real potential in big data and machine learning can be realized.

Acknowledgement: The foundation of this work has been laid down by our colleague, Imre Drávucz, whose contribution to the data base building and introducing chemometric methods into our laboratory practice has been instrumental.

Fully automated PARAFAC2 based analysis of GC-MS data

Anne Bech Risum¹, Rasmus Bro¹

¹*Chemometrics and Analytical Technology, Univ. Copenhagen, Rolighedsvej 26, Frederiksberg, Denmark*

anne.risum@food.ku.dk

Untargeted GC-MS often gives complex data. No specific chemical analytes are in focus a priori and hence it is not possible to optimize the procedure as in classical analytical chemistry. It has been shown that the PARADISE software^[1], based on PARAFAC2, is capable of separating co-eluting peaks and baselines, outperforming state of the art analytical chemical software both in the number of quantified compounds, and in the reproducibility of the analysis.

However, analysing GC-MS datasets using the PARADISE software still involves several manual steps, which are labour intensive and require a skilled tensor data analyst. We want to develop a fully automated “one button” solution, allowing non-chemometricians to easily utilize this advanced data analysis technique.

To achieve this goal, a number of subproblems have been identified and solved: 1) Division of the chromatogram into intervals suitable for PARAFAC2 analysis has been automated using a brute force approach involving many representations of each peak to find the optimal, 2) Automatic choice of the number of PARAFAC2 components by an expert system, developed and extended from earlier work^[2], and 3) Classification of the PARAFAC2 elution profile components as either attractive chemical peaks for quantification or not (e.g. baselines or half-peaks) by a deep convolutional neural network.

Here we show the combined result of these efforts, in the first ever fully automated PARAFAC2 based GC-MS analysis.

References

[1] Johnsen, L. G., Skou, P. B., Khakimov, B., & Bro, R. Gas chromatography–mass spectrometry data processing made easy. *Journal of Chromatography A*, 1503, 57-64 (2017)

[2] Johnsen, L. G., Amigo, J. M., Skov, T., & Bro, R. Automated resolution of overlapping peaks in chromatographic data. *Journal of Chemometrics*, 28(2), 71-82 (2014)
