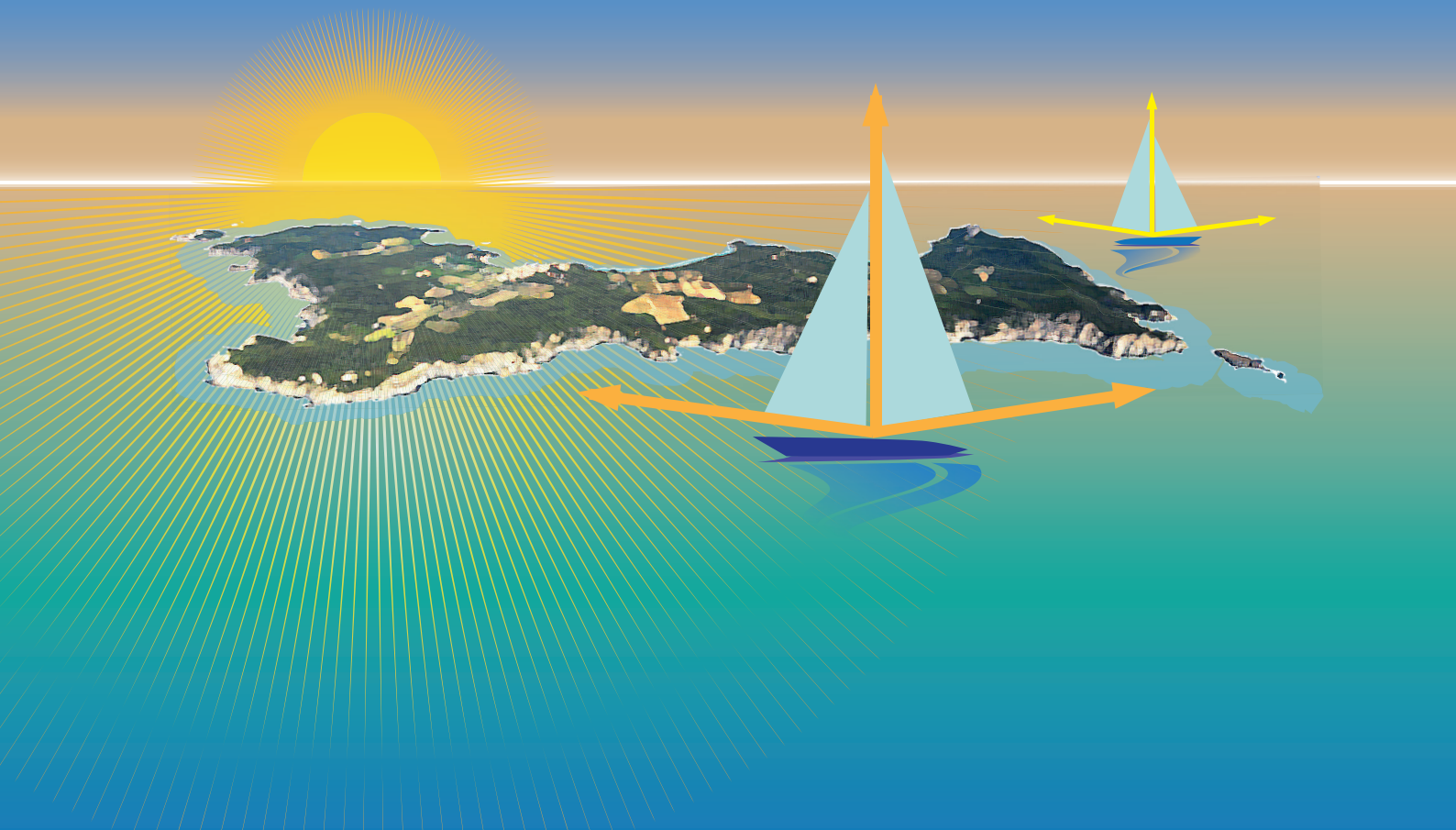
XII

COLLOQUIUM CHEMIOMETRICUM MEDITERRANEUM

ÎLE DE PORQUEROLLES
9 - 12 SEPTEMBER 2025
BOOK OF ABSTRACTS



The 12th edition of the Colloquium Chemiometricum Mediterraneum (CCM2025) will take place on the island of Porquerolles (France) between September 9th and 12th 2025.
The origin of the Colloquium Chemiometricum Mediterraneum (CCM) traces back to 1986 when, during an international conference, several French, Spanish and Italian researchers, each one speaking their own mother tongue, managed to share their own experience in the fields of chemometrics and multivariate data analysis. This was the seed from which the CCM sprang together with one of its main characteristics: the official languages of the conference are all Romance or Neo-Latin languages as well as English.
The past editions of CCM were held in Barcelona (Spain) in 1987, in San Miniato (Italy) in 1991, in Bastia (France) in 1994, in Burgos (Spain) in 1998, on the island of Ustica (Italy) in 2003, in Saint-Maximin-La-Sainte-Baume (France) in 2007, in Granada (Spain) in 2010, in Bevagna (Italy) in 2013, in Arles (France) in 2017, on the island of Menorca (Spain) in 2019 and in Padova (Italy) in 2023.
CCM constitutes a unique scientific forum where new advances in chemometrics are presented and discussed, and where French, Spanish, Portuguese and Italian chemometricians can express themselves in their own languages exchanging knowledge on cutting-edge developments. All over the last four decades, CCM has vastly promoted international collaborations and provided international exposure especially to students and young researchers.
CCM2025 will welcome contributions on - among others - topics like:

- algorithm development;
- calibration;
- Design of Experiments (DoE) and system optimisation;
- image analysis and hyperspectral imaging;
- multivariate analysis of variance;
- multivariate curve resolution;
- multiway and multiset data analysis;
- pattern recognition;
- Process Analytical Technology (PAT) and process data analysis;
- qualimetrics and chemical metrology;
- Quantitative Structure-Activity Relationships (QSAR) and Quantitative Structure-Property Relationships (QSPR).

PROGRAM

	September 9th 2025	September 10th 2025	September 11th 2025	September 12th 2025
09:00-09:20		Plenary 3 Astrid Jourdan	Plenary 4 Rosalba Calvini	Plenary 6 Rodrigo Rocha de Oliveira
09:20-09:40				
09:40-10:00		Oral 9 Charloto	Oral 19 Guerrini	Oral 36 Ortiz
10:00-10:20		Oral 10 Voccio	Oral 20 Olarini	Oral 37 Solarino
10:20-10:40		Oral 11 Sarabia	Oral 21 Caponigro	Oral 38 Cruz Muñoz
10:40-11:00		Coffee Break + Poster Walk	Coffee Break + Poster Walk	Coffee Break + Poster Walk
11:00:11:20		Oral 12 Camacho	Oral 22 Gariglio	Oral 39 Barron
11:20-11:40		Oral 13 Cariou	Oral 23 Ferrari	Oral 40 Lopez
11:40-12:00		Oral 14 Tauler	Oral 24 Cavallini	Oral 41 Roca-Nasser
12:00-13:20		Lunch	Lunch	Lunch
13:20-13:40	Opening Session			
13:40-14:00	Plenary 1 Elise Dumas	Oral 15 Gorla	Plenary 5 Ramin Nikzad-Langerodi	Oral 42 Kamp
14:00-14:20		Oral 16 Santarelli		Oral 43 Barbera
14:20-14:40	Oral 1 Gómez-Sánchez	Oral 17 Mazzoleni	Oral 25 Facco	Oral 44 Malinconico
14:40-15:00	Oral 2 Quintanilla-Casas	Oral 18 Strani	Oral 26 Clément	Oral 45 Pellacani
15:00-15:20	Oral 3 Sorochan Armstrong	Social activities	Oral 27 Kucheryavskiy	Award Presentation + Closing Session
15:20-15:40	Oral 4 Brandolini-Bunlon		Oral 28 Slanina	
15:40-16:00	Coffee Break + Poster Walk		Coffee Break + Poster Walk	
16:00-16:20	Plenary 2 Alberto Ferrer		Oral 29 Roussel	
16:20-16:40			Oral 30 Ginebreda	
16:40-17:00	Oral 5 Lacoue-Negre		Oral 31 Biancolillo	
17:00-17:20	Oral 6 Orecchio		Oral 32 Abhishek	
17:20-17:40	Oral 7 Dowek		Oral 33 Marini	
17:40-18:00	Oral 8 Giglio		Oral 34 Ezenarro	
18:00-18:20	Welcome Cocktail		Oral 35 Sozzi	
18:20-18:40				
From 19:15	Dinner	Dinner	Social Dinner	

SCIENTIFIC COMMITTEE

José Manuel Amigo Rubio, Universidad del País Vasco - Fundación Vasca para la Ciencia, Spain
Marina Cocchi, Università di Modena e Reggio Emilia, Italy
Riccardo Leardi, Università degli Studi di Genova, Italy
Federico Marini, Università degli Studi di Roma "La Sapienza", Italy
Jean-Michel Roger, Institut Agro Montpellier - Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement, France
Cyril Ruckebusch, Université de Lille, France
Luis Sarabia, Universidad de Burgos, Spain
Michelle Sergent, Aix Marseille Université, France
Romà Tauler, Instituto de Diagnóstico Ambiental y Estudios del Agua - Consejo Superior de Investigaciones Científicas, Spain
Roberto Todeschini, Università degli Studi di Milano-Bicocca, Italy

ORGANISING COMMITTEE

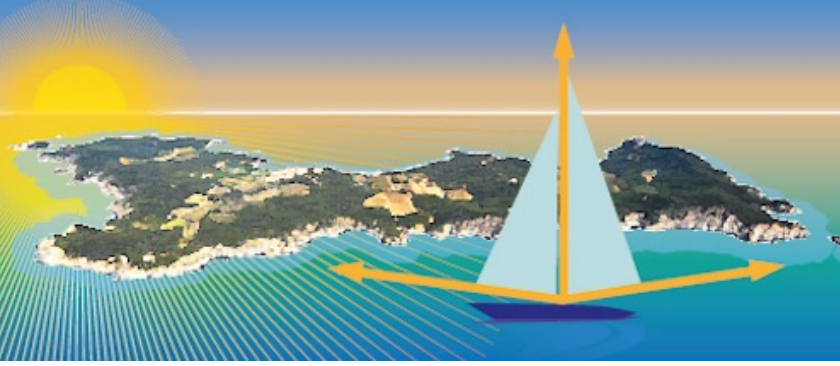
Philippe Bastien, L'Oréal Research and Innovation, France
Marion Brandolini-Bunlon, Université Clermont Auvergne - Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement, France
Magalie Claeys-Bruno, Aix Marseille Université, France
Ruggero Guerrini, Université de Lille, France
Benoît Jaillais, École Nationale Vétérinaire, Agroalimentaire et de l'Alimentation de Nantes-Atlantique (Oniris) - Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement, France
Silvia Mas-García, Institut Agro Montpellier - Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement, France
Jean-Michel Roger, Institut Agro Montpellier - Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement, France
Michelle Sergent, Aix Marseille Université, France
Raffaele Vitale (Chairman), Université de Lille, France

INVITED SPEAKERS

Rosalba Calvini obtained her Ph.D. in Agri-food Sciences, Technologies and Biotechnologies at the University of Modena and Reggio Emilia (Italy) in 2017, defending a thesis entitled "Chemometric tools for food characterisation through RGB and hyperspectral imaging". She has then been post-doctoral researcher at the same institution (2017–2021), working on national and international projects related to the development of analytical methods based on spectroscopy, imaging, and chemometrics for food analysis. Currently she is researcher in analytical chemistry at the Department of Life Sciences of the University of Modena and Reggio Emilia. Her research interests encompass the application and development of chemometric algorithms for hyperspectral and digital image analysis, with a focus on variable selection and data dimensionality reduction. In 2024, she received the Young Researcher Award from the Chemometrics Group of the Analytical Chemistry Division of the Italian Chemical Society.
At CCM2025, Rosalba will deliver a keynote talk entitled **"Overcoming the curse of dimensionality in hyperspectral data"**.

Elise Dumas is post-doctoral researcher in biostatistics at the École Polytechnique Fédérale de Lausanne (EPFL - Switzerland). Her research focuses on the use of large-scale observational datasets and causal inference methods for medical applications.
At CCM2025, Elise will deliver a keynote talk entitled **"Causal inference: a primer"**.

Alberto Ferrer is professor of statistics at the Universitat Politècnica de València (Spain), head of the Multivariate Statistical Engineering Group, chief scientific officer at Kenko Imalytics, S.L., and scientific advisor at Kensight Solutions, S.L. His research mainly focuses on (big) data analytics - especially in the domains of Industry 4.0 and Health 4.0 - industrial statistics, process chemometrics, machine learning, and multivariate statistical approaches for medical image analysis, systems and synthetic biology, omics sciences, fault detection and diagnosis, predictive maintenance, and optimisation.
At CCM2025, Alberto will deliver a keynote talk entitled **"PLS in the digitalized Industry 4.0: exploiting the causality in the latent space"**.



Astrid Jourdan is full professor at the Université de Pau et des Pays de l'Adour. She is an expert in statistics, machine learning, and experimental design. Her research mainly focuses on statistical analysis applied to numerical experiments and artificial intelligence. Her research interests include space-filling designs and sensitivity analysis, with applications in industry, environmental sciences, and artificial intelligence.
At CCM2025, Astrid will deliver a keynote talk entitled **"Space-filling designs for mixture experiments"**.

Ramin Nikzad-Langerodi is researcher team leader at the Software Competence Center Hagenberg (SCCH GmbH - Austria). Before joining SCCH, he worked as chemometrician at the Research Center for Non-Destructive Testing (RECENDT GmbH - Austria) and as machine learning researcher at the Department of Knowledge-Based Mathematical Systems of the Johannes Kepler University in Linz (Austria). Ramin Nikzad-Langerodi holds a degree in Biochemistry from the University of Zurich (Switzerland) and a Ph.D. in Pharmaceutical Sciences from the University of Vienna (Austria).
At CCM2025, Ramin will deliver a keynote talk entitled **"Two-norm penalties in multivariate calibration: beyond the usual suspects"**.

Rodrigo Rocha de Oliveira is lecturer at the Department of Chemical Engineering and Analytical Chemistry of the Universitat de Barcelona (Spain). He obtained his PhD from the same institution, focusing on process data fusion, modelling, and control strategies. His post-doctoral work, conducted in the Chemometrics Group of the University of Genova (Italy), advanced the analysis of near-infrared chemical images of biological samples and enhanced the understanding of near-infrared radiation interactions with solid matter. His research expertise spans the development and application of chemometric tools for Process Analytical Technology (PAT), with a particular emphasis on spectroscopic sensors and chemical imaging. Recently, he has begun exploring the potential of generative Artificial Intelligence (AI) models to promote developments in chemometrics. His contributions to the PAT field were recognized with the prestigious Siemens Process Analytics Prize in 2021 for an outstanding publication by a young scientist.
At CCM2025, Rodrigo will deliver a keynote talk entitled **"Harnessing the chemical, spatial, and temporal information: from spectral images to process monitoring"**.

PARTNERS

GFC

SFdS

amU
Aix Marseille Université

cnrs

L'INSTITUT agro Montpellier

INRAE

Université de Lille

SPONSORS

VILLE D'HYÈRES
LES PALMIERS

imbe
institut méditerranéen de biodiversité et d'écologie
marine et continentale

LASIRE

Département
MathNum

Lovelace's Square
An Open Source Hub for Chemometrics
www.lovelacesquare.org

Società
Chimica
Italiana
Divisione di Chimica Analitica
Gruppo di lavoro di Chemiometria

INDEX

PLENARY

PRESENTATIONS

The abstracts are organized in alphabetical order according to their titles.

Causal inference: a primer

Elise Dumas[1]
[1] *Institute of Mathematics, Chair of Biostatistics, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland*
elise.dumas@epfl.ch

As humans, we intuitively reason in terms of cause and effect. However, most statistical estimates, including those based on machine learning, are based on association, not causation. While measures based on mere association may be sufficient for certain problems, causal reasoning is required for many others, e.g., to estimate the effects of policies at the population level, to identify the units that would benefit most from an intervention, or to reason about the mechanisms of action of drugs. In recent years, causal reasoning has been mathematically formalized in the field of causal inference and has played an increasing role in disciplines that intersect with statistics, including econometrics, biostatistics, epidemiology, and machine learning. In this talk, I will introduce the basics of causal inference [1], including the potential outcomes framework, the definition of causal effect estimands, direct acyclic graphs, and core causal assumptions for identification. While most applications of causal inference have focused on economics and health, formal causal inference can provide a foundation for new intra- and inter-disciplinary developments in chemometrics.

References

[1] Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman&Hall/CRC; 2020

Harnessing the chemical, spatial and temporal information: From spectral images to process monitoring

Rodrigo Rocha de Oliveira[1]
[1] *Universitat de Barcelona, Gran Via de les Corts Catalanes, 585, 08007 Barcelona, Spain*
rodrigo.rocha@ub.edu

This communication presents advanced chemometric approaches for extracting and integrating multidimensional information from hyperspectral imaging (HSI) data [1] to enhance process monitoring capabilities. While HSI inherently provides both chemical (spectral) and spatial information, recent methodological developments allow for its acquisition across time, also known as HSI movies.

The application of different chemometrics methods such as PCA, PLS and MCR-ALS to handle such data will be demonstrated in case studies from pharmaceutical, food processing, and biological applications. Beyond conventional manufacturing, we showcase novel implementations for shelf-life testing of plant-based foods, where chemical changes can be monitored spatiotemporally to predict quality degradation. Additionally, we explore applications in environmental monitoring through the characterization of dehydration processes in epiphytic organisms such as lichens and moss, providing insights into stress responses and adaptation mechanisms.

Building upon previously developed heterogeneity assessment from chemical images, which quantifies both global (GHI) and distributional (DUI) heterogeneity indices, we extend this methodology to capture temporal dynamics in continuous manufacturing processes. This approach allows for real-time monitoring of chemical composition evolution and spatial distribution changes throughout blending process. The methodology leverages multivariate curve resolution techniques for assessing chemical component distribution maps and variographic analysis for spatial heterogeneity quantification, while implementing novel algorithms to track these parameters across time [2].

The results indicate that this comprehensive analytical framework significantly improves process understanding, facilitates root cause analysis of processing anomalies, and enables more responsive monitoring strategies across diverse fields from industrial manufacturing to ecological studies.

References

[1] Amigo, J.M., ed., Hyperspectral imaging, in: Data Handling in Science and Technology. Vol. 32., Elsevier, 2020

[2] Rocha de Oliveira, R., de Juan, A., Anal. Chem., 2020 (92)

Overcoming the curse of dimensionality in hyperspectral data

Rosalba Calvini[1]
[1] *Department of Life Sciences, University of Modena and Reggio Emilia, Italy*
rosalba.calvini@unimore.it

The curse of dimensionality is a major challenge in the analysis of hyperspectral imaging (HSI) data. In fact, hyperspectral images are extremely data-rich arrays, as each image can contain thousands of spectra, with each spectrum consisting of hundreds of wavelengths [1].
While such extensive data acquisition enables a detailed chemical mapping of the analysed-samples, the large size of hyperspectral images often leads to high computational demands and long processing times. This problem is even more relevant in practical applications of HSI, where it is often necessary to simultaneously analyse many images at the same time. To address these challenges, data reduction methods are mandatory to streamline data processing while preserving essential analytical information. Data reduction strategies can be broadly categorised into two main groups: pixel-level and image-level reduction.
Pixel-level approaches preserve the pixel-related information and are based on obtaining a reduced version of the original hyperspectral data spanning the same variance. Among these methods, random projection resulted a very efficient approach for the exploratory analysis of hyperspectral data [2, 3].
On the other hand, image-level data reduction methods consist of extracting the relevant spatial and/or spectral information contained in each image into a feature vector. In this way, a large dataset of hyperspectral images can be transformed into a two-dimensional data matrix in which each row represents the feature vector of the corresponding original hypercube. Then, this matrix of feature vectors can be analysed with chemometric methods for exploratory analysis or for the development of calibration or classification models [4].
While the simplest image-level data reduction method consists in calculating the average spectrum of each image, more advanced methods have been proposed, such as the hyperspectrograms approach, allowing to retain both spatial and spectral information [5]. This presentation will discuss the main advantages and drawbacks of pixel- and image- level data reduction methods, illustrating them in real-case scenarios.

References

[1] Burger, J., Gowen, A., Chemometr. Intell. Lab., 2011 (108), 13-22
[2] Cruz-Tirado, J. P., Amigo, J. M., Varbin, D. F., Kucheryavskiy, S., Anal. Chim. Acta, 2022 (1209), 339793
[3] Calvini, R., Amigo, J. M., Chemometr. Intell. Lab., 2024 (248), 105118

[4] Calvini, R., Foca., G., Ulrici., A., Anal. Bioanal. Chem., 2016 (408), 7351-7366
[5] Calvini, R., Michelini, S., Pizzamiglio, V., Foca, G., Ulrici, A., Food Control, 2020 (112), 107111

PLS in the digitalized Industry 4.0: exploiting the causality in the latent space

Alberto Ferrer[1,2]
[1] *Universitat Politècnica de València, Dpt. of Statistics, GIEM, Spain*
[2] *Kensight Solutions S.L.*
aferrer@eio.upv.es

In the modern industrial environment guided by Quality-by-Design, process optimization depends on understanding input-output relationships through causal models. Although deterministic models are the standard, they are often too costly or complex to develop. This has prompted a shift toward more feasible data-driven approaches. However, ensuring causality from these data-driven models demands independent variation in the inputs, commonly introduced via Design of Experiments (DOE). However, in Industry 4.0 environments, where the number of potential factors to consider as inputs can be really high, and due to the complex correlation structure among them there are a lot of restrictions that prevent moving some factors independently from others, DOE can be difficult to carry out, if not impossible [1].

Manufacturers now have access to vast amounts of production data, the problem is that these historical datasets often lack the input independence required for sound causal inference. This limitation undermines the effectiveness of conventional statistical and machine learning predictive models in driving optimization. In response, there is a growing movement toward extracting causal insights directly from historical data, and Partial Least Squares (PLS) regression can be a useful tool for this purpose. PLS models provide uniqueness and causality in the reduced latent space no matter if the data come either from a DOE or daily production process; therefore, optimization can be done in the latent space [2].

This talk examines different ways of exploiting the causality-in-the-latent-space PLS property by the PLS model inversion: i): obtaining the settings of the manipulable X variables that guarantee the desired value in the critical to quality attributes (CQA) of the manufactured products [3]; ii) running DOEs in the latent space [4]; iii) defining multivariate raw material specifications providing assurance of quality with a certain confidence level for the CQA [5,6]; and iv) developing a novel Latent Space-based Multivariate Capability Index (LSb-MCpk) that quantifies the capacity of each raw material supplier of providing assurance of quality with a certain confidence level for the CQAs of the manufactured product before manufacturing a single unit of the product [7].

References

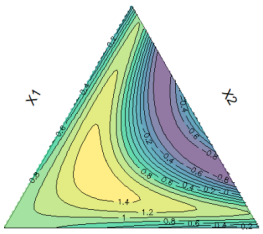
[1] Ferrer, A., Quality Engineering 2021 (33), 758-763
[2] Jaeckle, C. M., MacGregor, J.F., Chemometr. Intell. Lab., 2000 (50), 199–210

[3] Palací-López, D., Facco, P., Barolo, M., Ferrer, A., Chemometr. Intell. Lab., 2019 (194), 103848
[4] Wold, S., Sjöström, M, Carlson, R., Lundstedt, T., Hellberg, S., Skagerberg, B., Wikström, C., Öhman, J., Anal. Chim. Acta, 1986 (191), 17–32
[5] Borràs-Ferrís, J., Palací-López D., Duchesne, C., Ferrer, A., Chemometr. Intell. Lab., 2022 (225), 104563
[6] Borràs-Ferrís, J., Duchesne, C., Ferrer, A., Chemometr. Intell. Lab., 2023 (240), 104912
[7] Borràs-Ferrís, J., Duchesne, C., Ferrer, A., Chemometr. Intell. Lab., 2025 (258), 105339

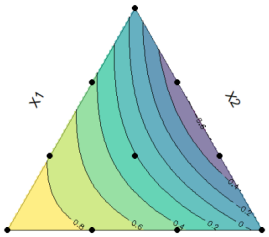
Space-filling designs for mixture experiments

Astrid Jourdan[1]
[1] *Laboratoire LMAP, Université de Pau et des Pays de l'Adour, France*
astrid.jourdan@univ-pau.fr

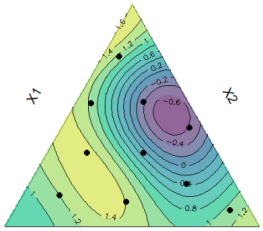
Space-filling designs are mainly used in experimentation when the response of interest is expected to be non-linear. In this case, the experiments must be uniformly spread in the entire experimental domain to ensure sufficient exploration and to facilitate the detection of nonlinearities. This kind of experimental design has been widely studied in the case where the experimental domain is a hypercube. In this presentation, we will focus on the case of uniform designs for mixture formulation, i.e. with the constraint that each variable varies between 0 and 1 and the sum of the coordinates is 1. We will compare the classical approach with a latticesimplex design and a second order polynomial model and the use of space-filling designs with a kriging model.



True response



A lattice-simplex design with a second-order polynomial model



A space-filling design with a kriging model

Two-norm penalties in multivariate calibration: Beyond the usual suspects

Ramin Nikzad-Langerodi[1]
[1] *Software Competence Center Hagenberg GmbH, Austria*
ramin.nikzad-langerodi@scch.at

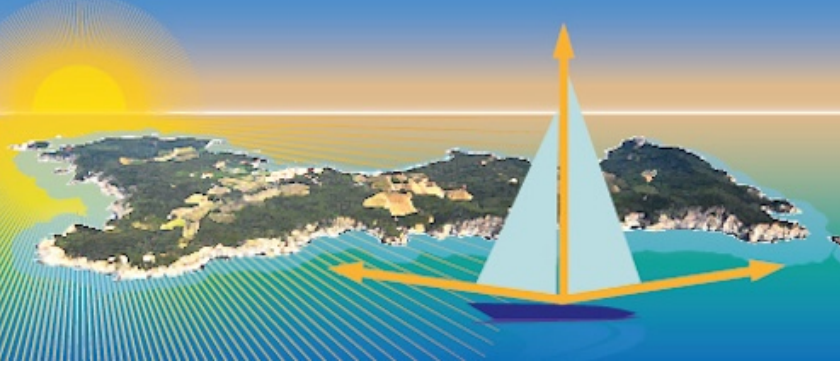
Two-norm or Tikhonov regularization is a fundamental technique to improve numerical stability and prevent overfitting in ill-posed, linear regression problems. With the adoption of latent variables-based techniques such as principal components and partial least squares (PLS) regression in chemometrics, that take a different approach to handle multicollinearity and controlling model complexity, the use of two-norm regularization has been somewhat overshadowed over the past decades. Recently, however, two-norm penalties have been revisited in the context of calibration model maintenance [1, 2], as well as for enforcing particular properties of the PLS model parameters based upon domain knowledge [3]. While the former approaches employ two-norm penalties to enforce a priori structure in the (PLS) scores, e.g., that the samples corresponding to matched calibration transfer standards roject to the same coordinates in the latent-variable space, the latter uses two-norm penalties on the PLS weights to enforce smoothness. In this contribution I will revisit the two-norm regularization framework in partial least squares modelling, underpinning its versatility, and introduce some more "exotic" penalty functions that allow for flexible control over scores, weights and loadings. Application of the framework will be illustrated using simulated and real-world data covering use-cases of semi-supervised calibration, calibration maintenance/transfer, implicit baseline removal and the incorporation of prior knowledge into multivariate calibrations.

References

[1] Kalivas, John H., et al. Appl. Spectrosc 63.7 (2009): 800-809
[2] Nikzad-Langerodi, R., Andries E., J. Chemometrics 35.11 (2021): e3373
[3] Goutis, C., Fearn T., J AM STAT ASSOC 91.434 (1996): 627-632

Acknowledgements

The research reported in this contribution has been funded by BMK, BMAW, and the State of Upper Austria in the frame of the SCCH competence center INTEGRATE (FFG grant no. 892418) part of the FFG COMET Competence Centers for Excellent Technologies Program.



ORAL PRESENTATIONS

The abstracts are organized in alphabetical order according to their titles.

A comparative study of anomaly detection methods in hyperspectral imaging. Targeted and untargeted approaches.

Giulia Gorla[1], Reaha Goyetche[1], Artzai Picon Ruiz[2], Aitor Alvarez Gila[2], Neal B. Gallagher[3], José Manuel Amigo[1,4]

[1] *Department of Analytical Chemistry, University of the Basque Country (UPV/EHU), Barrio Sarriena S/N, Leioa 48940, Spain*
[2] *TECNALIA, Basque Research and Technology Alliance (BRTA), Astondo Bidea, Edif. 700, 48160 Derio, Biscay, Spain*
[3] *Eigenvector Research, Inc., Manson, WA, USA*
[4] *Ikerbasque, Basque Foundation for Sciences, María Díaz de Haro, 3, Bilbao 48013, Spain*
giulia.gorla@ehu.eus

In chemometrics, anomaly detection methods are essential for identifying observations that deviate significantly from the expected data structure, often revealing rare, subtle information within highly variable chemical datasets [1]. Specifically, in chemical imaging, such as hyperspectral imaging, these anomalies may correspond to minor contaminations, defects, or unexpected chemical changes that can be easily masked by normal process variation or sample heterogeneity. Additional applications include identifying target compounds within complex matrices, where challenges may arise in both the spatial and spectral dimensions [2]. This study investigates and compares several strategies for anomaly detection in hyperspectral data, focusing on untargeted and targeted methods on real-case datasets of microplastic contaminants in sand. We begin by framing the concept of an anomaly, emphasising how such deviations can be subtle and context dependent. Unsupervised approaches examined include PCA-based detection using Hotelling's T^2 statistic and variational autoencoders, which aim to reconstruct expected patterns and highlight deviations. On the supervised side, we evaluate CLS-based methods such as Generalized Least Squares (GLS) and Extended Least Squares (ELS), as well as Multivariate Curve Resolution–Alternating Least Squares (MCR-ALS), which can disentangle overlapping chemical contributions. By comparing these methods, we identify their respective strengths in detecting anomalies of varying nature and intensity. Our results provide insight into when and how anomalies can be effectively captured in chemical imaging workflows, aiding in quality control, process monitoring, and exploratory analysis.

Acknowledgements
This research was supported by the Innovation Fund Denmark under the project "HyperSort - Hyper-spectral imaging for advanced textile sorting and food control." Additionally, funding was provided by the Elkartek Programme of the Basque Government (Spain) under the SMART-EYE project (KK-2023/00021)

References
[1] Geladi, P., Teugels, L.J., Extreme value theory and its potential role in chemometrics, J. Chemometr., 2022 (15), 1-12

[2] Racetin, I.; Krtalić, A., Systematic review of anomaly detection in hyperspectral remote sensing applications, Appl. Sci. 2021 (11), 4878

A quantitative analysis of hydrocortisone filament using a portable Raman spectrometer and chemometrics to ensure quality control in 3D printing

Antoine Dowek[1,2], Aruzhan Seidakhanova[1,2], Quentin Bourcy[3], Olivier Jennotte[3], Brigitte Evrard[3], Robin Crunenberg[3], Bernard Do[2], Maxime Annereau[2], André Rieutord[2], Anna Lechanteur[3]

[1] *Clinical Pharmacy Department, Gustave Roussy Cancer Campus, 94800 Villejuif, France*
[2] *Université Paris-Saclay, Lipides : systèmes analytiques et biologiques, 91400, Orsay, France*
[3] *Laboratory of Pharmaceutical Technology and Biopharmacy, Center for Interdisciplinary Research on Medicines (CIRM), Department of Pharmacy, University of Liege, Avenue Hippocrate 15, 4000 Liege, Belgium*
antoine.dowek@universite-paris-saclay.fr

Using 3D printing for drug development offers advantages, such as fulfilling specific patient needs and accommodating complex drug formulations. Ensuring the safety and efficacy of these printed medications requires rigorous quality control, for which Raman spectroscopy is a powerful tool. This technique can be integrated into the hot melt extrusion (HME) process or the printing process itself, analyzing the drug content in the filament after its production and before it is melted and formed into the final dosage form.

This project developed a quantitative analytical method using a portable Raman spectrometer to measure the concentration of hydrocortisone (HCT) in a filament. This filament, developed by the University of Liège's laboratory, is intended for 3D printing of low-dose solid oral forms for pediatric patients with adrenal insufficiency, using fused deposition modeling.

Following ICH Q2(R2) guidelines, a validated method for HCT quantification in solution was established. This model was subsequently used to predict the HCT content in a filament composed of 20% HCT, 54% Affinisol® 15LV (hydroxypropylmethylcellulose), 23% Kollidon® VA 64 (vinylpyrrolidone-vinyl acetate copolymer), and 3% tri(ethyl)citrate as a plasticizer. The initial step involved defining a specific spectral region unique to HCT, thereby avoiding interference from excipients.

Subsequently, preprocessing methods were optimized, including Savitzky-Golay (SG) smoothing, asymmetric least squares (ALS) baseline correction, first and second derivatives. Extended Multiplicative Signal Correction (EMSC), a technique used in Raman spectroscopy to mitigate unwanted spectral variations—particularly physical artifacts encountered when analyzing complex solid forms like filaments—was also investigated.

The optimized preprocessing sequence, consisting of shifting correction, SG, ALS, EMSC, and second derivative analysis, followed by partial least squares regression, achieved a mean absolute error of 4.11% across 187 filament measurements. Notably, the method demonstrated high accuracy for the target 20% w/w HCT concentration ($n = 81$) across three batches, yielding a mean absolute error of 3.23%. This successfully validated method

enables the direct and quantitative analysis of HCT within the 3D printing filament.

Adaptive soft sensors for highly time-varying processes through ProportionalIntegral-Derivative PLS

Andrea Botton[1], Francesco Bonacini[2], Gianni Marchetti[2], Angelo Ferrando[2], Pierantonio Facco[1]

[1] *CAPE-Lab – Computer-Aided Process Engineering Laboratory, Department of Industrial Engineering, University of Padova, Via Marzolo 9, 35131 Padova (PD), Italy*

[2] *Basic Chemicals and Plastics Research Centre, Versalis SpA, Via Taliercio 14, 46100 Mantova (MN), Italy*

pierantonio.facco@unipd.it

The global plastics market continues to expand, driven by the importance of materials such as ethylene propylene diene monomer (EPDM), which is valued for its elasticity and weather resistance in the automotive, construction and electrical industries [1]. In plastics manufacturing the real-time quality estimation is critical for process optimization and waste reduction. Laboratory-based quality measurements are typically collected, but they are infrequent and delayed, making them unsuitable for decision-making in a time-varying production environment [2]. Soft sensors based on Partial Least Squares (PLS) regression [3] offer an interpretable alternative to complex methodologies used for polymer quality estimation [4]. However, their performance degrades in processes with low cross-correlation between process and quality variables and strong temporal dependencies. To overcome these limitations, we propose a novel methodology for soft sensing: Proportional-Integral-Derivative Partial Least Squares (PID-PLS). Inspired by control engineering, PID-PLS integrates short-term deviations (Proportional), long-term drifts (Integral), and abrupt changes (Derivative) of the response variable residuals to the predictors of an adaptive PLS model together with the process data. This allows the model to adaptively capture both transient and persistent patterns in quality data, enhancing prediction accuracy in highly time-varying conditions.

The proposed approach is tested on an industrial case study (Versalis S.p.A., Mantova, Italy), in which process data from EPDM production are used to estimate the Mooney viscosity, a critical product quality parameter [2]. PID-PLS outperforms global [5], recursive [6] and moving-window PLS [7] models in terms of estimation accuracy (Table 1) and robustness, especially under fluctuating conditions due to subtle drifts, such as equipment aging and/or fouling, and/or abrupt changes in process conditions, such as mode changes.

The proposed modelling techniques is further tested on a simulated case study which confirms the model accuracy and versatility across datasets exhibiting varying temporal patterns, as shown by the results of Table 2. While the proportional component alone induces most of the predictive accuracy, the integral and derivative terms contribute meaningfully in scenarios where drifts and sharp transitions, respectively, are present in the process under study. The model is computationally efficient and remains interpretable, thus supporting its deployment in Industry 4.0 environments. In conclusion, this work

presents a scalable framework for adaptive soft sensing in polymer production which can be transferred to other manufacturing sectors facing time-variant challenges.

Table 1. Comparison of the performance of the proposed PID-PLS model with different modelling strategies for the estimation of the Mooney viscosity in an industrial production of EPDM

Model	Mean Relative Error (%)	Standard Deviation (%)
Global PLS	6.63	5.45
Recursive PLS	4.40	3.32
Moving Window PLS	2.78	2.24
Proportional-PLS	2.05	1.64
PID-PLS	2.25	1.78
PID regression[1]	3.15	2.60

Table 2. Comparison of the performance of the proposed PID-PLS model with different modelling strategies for the estimation of the response variable of a simulated case study

Model	Mean Relative Error (%)	Standard Deviation (%)
Moving Window PLS	5.60	3.28
Proportional-PLS	4.70	3.21
Integral PLS	2.88	2.02
Derivative-PLS	5.60	2.97
PID-PLS	2.52	1.86

References

[1] Mark J.E., Erman B., Eirich F.R., The Science and Technology of Rubber. 3rd ed. Elsevier Academic Press, 2005, San Diego, CA, USA, p. 1395

[2] Malač, J., Journal of Materials Science and Engineering with Advanced Technology, 2001 (3), 1, 67–87

[3] Wold S., Sjostorm M., Eriksson L., Chemometr. Intell. Lab., 2001 (58), 109–130

[4] Yang Y., Lu J., Entropy, 2022 (24), 528

[5] Geladi P., Kowalski B., Anal. Chim. Acta, 1986 (185), 1-17

[6] Helland, K., Berntsen, H.E., Borgen, OS. and Martens, H., Chemometr. Intell. Lab., 1992 (14), 129-137

[7] Qin S., Comput. Chem. Eng., 1998 (22), 503-514

[1] Multi-linear regression on the proportional, integral and derivative components of the response residuals at the previous time instants.

Can a simple visible image guide us in our chemometric explorations?

Ruggero Guerrini[1],Federico Marini[2], Herreyre Nicolas[3,4], Vincent Motto-Ros[3], Ludovic Duponchel[1]

[1] *Univ. Lille, CNRS, UMR 8516 – LASIRE – Laboratoire de Spectroscopie pour Les Interactions, La Réactivité et L'Environnement, Lille, 59000, France*

[2] *Department of Chemistry, Sapienza University of Rome, Piazzale Aldo Moro 5, Rome, 00185, Italy*

[3] *Institut Lumière Matière UMR 5306, Université Lyon 1. CNRS, Villeurbanne, 69622, France*

[4] *Archéologie et Archéométrie, UMR 5138, Univ. Lyon 2-CNRS-Univ. Lyon 1, Maison de l'Orient et de la Méditerranée, 7 rue Raulin, 69007 Lyon, France*

ruggero.guerrini.etu@univ-lille.fr

Spectroscopic imaging is now more than ever at the heart of analytical chemistry. It enables us to thoroughly investigate complex samples by combining, in a single acquisition, molecular or elemental spectral analysis with spatial information. Over time, each spectroscopic technique has developed its own imaging modality—typically through the combination of a spectrometer and a microscope or another focusing system. Without being exhaustive, we can cite infrared, Raman, MALDI, and LIBS imaging—each offering unique characteristics in terms of spatial resolution, sensitivity, speed, and chemical insight. The sheer volume of spectra generated by these instruments quickly led us to develop chemometric approaches to explore and extract the most unbiased chemical information possible. In fact, it is now quite rare to see published work in spectroscopic imaging without an accompanying chemometric component. One might assume, then, that everything is progressing optimally. However, the results we will present in this work stem from an observation that applies broadly across all spectroscopic imaging techniques. Most of these instruments are also capable of acquiring a visible image of the sample under investigation. This is often used to observe the precise area to be analyzed or to ensure the optical system is correctly aligned for spectral acquisition. Yet despite the rich information contained in these visible images, they are rarely used in subsequent data processing—only the spectroscopic data are typically exploited. The aim of this presentation is to demonstrate how data fusion between a visible image and the spectroscopic data cube can significantly enhance sample exploration. We will introduce aspects related to image registration as well as colorimetric space transformations. To illustrate our approach, we will focus on the characterization of a fragment of mortar taken the Saint-Irénée Church (Fourvière Hill in Lyon, France) using LIBS imaging.

Characterization of parotid gland tissue via data geometry analysis of MALDI MSI Images

Alessandra Olarini[1], Vicky Caponigro[2], Emanuela Salviati[2], Pietro Campiglia[2], Eugenio Alladio[1], Ludovic Duponchel[3], Cyril Ruckebusch[3], Marina Cocchi[4]

[1] *1Dipartimento di Chimica, Università di Torino, Italy*

[2] *Dipartimento di Farmacia, Università di Salerno, Italy*

[3] *LASIRE, Université de Lille, France*

[4] *Dipartimento di Scienze Chimiche e Geologiche, Università di Modena e Reggio Emilia, Italy*

alessandra.olarini@unito.it

Salivary gland tumors are relatively uncommon neoplasms, accounting for less than 5% of head and neck tumors, with approximately 90% occurring in the parotid gland [1]. The broad histological variability and tumor heterogeneity present significant challenges in both diagnosis and treatment [2].

This study presents preliminary results from the analysis of a matrix-assisted laser desorption/ionization mass spectrometry imaging (MALDI-MSI) dataset of parotid tissue. The dataset includes 44 images from 11 patients, comprising both healthy and pathological samples acquired in positive and negative ion modes [3]. A geometric interpretation based on the convexity of normalized scores and loadings, obtained through Singular Value Decomposition (SVD), is proposed to extract useful (essential) information, in the context of linear mixture analysis, from complex biological tissues [4, 5]. The analysis of the geometry of the normalized score (X) and loading (Y) spaces reveals the potential to discriminate between healthy and pathological tissues. In the X-space, healthy tissues display a uniform spatial distribution characterized by a single cluster, while pathological tissues exhibit two to three distinct, globularly shaped groupings, particularly evident in the negative ion mode. In the Y-space, m/z features radiate outward from a central region for both tissue types. Extreme points in both spaces were selected to identify representative pixels and m/z images, whose spectral profiles and m/z values were validated through comparison with histological data and previous studies. This approach aims to offer a computationally efficient framework for tissue characterization, with the potential to support future unmixing and classification tasks.

References

[1] Gatta, G., Guzzo, M., Locati, L.D., et al., Crit. Rev. Oncol. Hematol., 2020 (152), 102959

[2] Ali, N.S., Akhtar, S., Junaid, M., et al., ISRN Surg., 2011, 721625

[3] Sommella, E., Salviati, E., Caponigro, V., et al., Metabolites, 2022, (12), 530

[4] Grande, B.V., Manne, R., Chemometr. Intell. Lab. Syst., 2000, (50) 19–33

[5] Olarini, A., Cocchi, M., Motto-Ros, et al., Chemometr. Intell. Lab. Syst., 2024, (252), 105174

Chemometric approaches for volatilomic data: profiling of *in vitro* cell metabolism

Elisabetta Santarelli[1], Matteo Delli Carri[1], Maria Rosaria Miranda[1], Vincenzo Vesuto[1], Vicky Caponigro[1], Giacomo Pepe[1], Pietro Campiglia[1], Agnieszka Smolinska[2], Carlo Crescenzi[1]
[1]*1Department of Pharmacy, University of Salerno, Fisciano, Italy*
[2]*Department of Toxicology, Maastricht University, Maastricht, The Netherlands*
esantarelli@unisa.it

The term volatilome refers to the study of Volatile Organic Compounds (VOCs) produced and released by a biological system. These low-molecular-weight compounds originate from various metabolic processes and can, therefore, offer valuable insights into the metabolic pathways. Moreover, this fraction of metabolites is easily accessible, and it was poorly explored in last decades. For this reason, the volatilome has recently emerged as a promising analytical target for applications in biomarker discovery and disease diagnosis [1]. In this study, a chemometric workflow was applied to investigate the volatilome in the headspace of human melanoma cell cultures and to evaluate alterations induced by the pharmacological treatment with thapsigargin, a known pro-apoptotic agent.
The samples, divided into two batches, were analyzed using Solid Phase MicroExtraction coupled with Gas Chromatography–Mass Spectrometry (SPME-GC/MS). Data preprocessing included peak detection and integration, deconvolution and alignment, Probabilistic Quotient Normalization (PQN), and log10 transformation. Principal Component Analysis (PCA) on autoscaled data revealed clustering patterns, while External Parameter Orthogonalization (EPO) [2] was applied to remove batch-related variance. ANOVA Simultaneous Component Analysis (ASCA) [3] confirmed both effective batch correction and a significant treatment effect.
For supervised analysis, Soft Independent Modelling of Class Analogy (SIMCA) [4] and complementary classification methods were used to highlight relevant variables. Univariate analysis refined feature selection and supported biological interpretation. The results of this study highlight the potential of volatilomics to identify treatment-related biomarkers and provide insight into metabolic pathways of in vitro cancer models.

References

[1] Schmidt K., Podmore I., J Biomark, 2015 (2015), 1–16
[2] Roger J.M., Chauchard F., Bellon-Maurel V., Chemometr. Intell. Lab, 2003 (66), 191–204
[3] Smilde A.K., Jansen J. J., Hoefsloot H. C. J., Lamers R. J. A. N., van der Greef j., Timmerman M. E., Bioinformatics, 2005 (21), 3043–3048
[4] Vitale R., Cocchi M., Biancolillo A., Ruckebusch C., Marini F., Anal. Chim. Acta, 2023 (1272), 341304

Class-Modelling Strategies for Microplastic Detection in Liquid Food Products Using Stainless Steel Filtration and Micro-Raman Analysis

Vicky Caponigro[1], Antonio Malinconico[1], Fabrizio Bobba[2], Edoardo Medolla[3], Pietro Campiglia[1], Federico Marini[3], Carlo Crescenzi[1]

[1] *Department of Pharmacy, University of Salerno, Via Giovanni Paolo II, 132, 84084 Fisciano, Italy*
[2] *Department of Pharmacy, University of Salerno, Via Giovanni Paolo II, 132, 84084 Fisciano, Italy*
[3] *Department of Chemistry, University of Rome "La Sapienza", Piazzale Aldo Moro 5, 00185 Rome, Italy*
vcaponigro@unisa.it

Microplastics (MPs) are pervasive contaminants that infiltrate food chains and pose significant ecological and human health risks [1]. Despite growing concerns, standardised protocols for detecting MPs in food remain limited. This study introduces an approach for low-concentration MPs identification in liquid food matrices, combining stainless steel (StS) frit filtration, micro-Raman hyperspectral imaging (HSI), and multivariate data analysis.
StS substrates (4.6 mm and 2 mm diameters) were selected for their mechanical robustness and minimal Raman interference. A standard solution containing polyethylene terephthalate (PET) particles and real liquid food products were filtered and analysed with micro-Raman HSI. Two Raman data acquisition strategies were compared: a high-resolution sampling grid using the "Edge" laser configuration, and a faster continuous imaging mode using "LineFocus". The collected hypercube underwent extensive data pre-processing, including cosmic ray removal, baseline correction, and multiplicative scatter correction, to optimise spectral quality.
Two class-modelling strategies were built using StS substrate imaging: Soft Independent Modelling of Class Analogy and Support Vector Machine classification. These models were optimised and validated through leave-one-image-out cross-validation and then, applied to PET and real sample images to identify pixels not classified as StS. The spectra of these pixels were compared to PET reference standards and spectra collected from individual components of the plastic bottles (e.g., lids, bodies).
The workflow demonstrated high specificity in differentiating MPs from substrate interference. The analysis conclusively traced MPs to the plastic push-on bottle caps for real samples, attributed to mechanical stress during repeated opening and closing.

References

[1] Browne, M. A., Galloway, T., Thompson, R., Integrated Environmental Assessment and Management, 2007 (3), 559–561

Compressive imaging for biomedical applications

Jean-Emmanuel Clément[1], Shunsuke Ono[2], Riley Hargreaves[1], Cyril Ruckebusch[1], Peter Dedecker[1,3], Fujita Katsumasa[4], Tamiki Komatsuzaki[5]
[1] *CNRS, LASIRE (UMR 8516), University of Lille, France*
[2] *Department of Computer Science, Institute of Science Tokyo, Japan*
[3] *Department of Chemistry, KU Leuven, Leuven, Belgium*
[4] *Departement of Physics, Osaka University, Japan*
[5] *Department of Chemistry, Hokkaido University, Japan*
jean-emmanuel.clement@univ-lille.fr

3D fluorescence imaging, fluorescence lifetime imaging (FLIM), and Raman imaging are powerful techniques for revealing the structural and metabolic properties of biological samples. However, their inherently slow acquisition speeds limit their potential for imaging dynamics on timescales shorter than a few minutes [1]. To address this limitation, we developed a compressed sensing formulation that enables video-rate measurements. In this setting, we rapidly acquire undersampled spatio-spectral or spatio-temporal data using random patterns and shear transformations, as illustrated in Fig. 1. To reconstruct full images from these undersampled measurements, we formulate an inverse problem with a custom regularization term that captures both the low-rank structure and local smoothness typical of bioimages. Specifically, we define a constrained optimization problem that is efficiently solved using a primal-dual splitting method, making it suitable for large-scale data. We demonstrate that our algorithm reduces the oversmoothing of reconstructed images commonly observed in state-of-the-art algorithms.

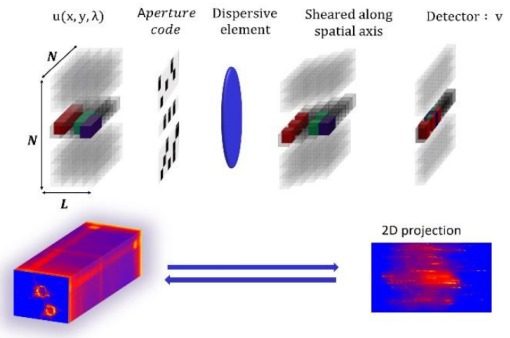


Figure 1. Compressive Raman imaging measurement

References

[1] K. Mochizuki, Y. Kumamoto, S. Maeda, M. Tanuma, A. Kasai, M. Takemura, Y. Harada, H. Hashimoto, H. Tanaka, N. Smith, and K. Fujita, BIOMED OPT EXPRESS 14, 1015-1026 (2023)

Contribution of multiblock methods for predicting the severity of COVID-19 from clinical, biochemical and metabolomic data

Marion Brandolini-Bunlon[1], Benoit Jaillais[2], Audrey Legouellec[3], Jean-Michel Roger[4], Estelle Pujos-Guillot[1]

[1] *Universitè Clermont Auvergne, INRAE, UNH, PFEM, MetaboHUB Clermont, Clermont-Ferrand, France*
[2] *INRAE, TRANSFORM, Nantes, France*
[3] *Univ. Grenoble Alpes, CNRS, UMR 5525, VetAgro Sup, INP, CHU Grenoble Alpes, TIMC, Grenoble, France*
[4] *ITAP, INRAE, Institute Agro, University Montpellier, Montpellier, France*
marion.brandolini-bunlon@inrae.fr

During the first wave of the COVID pandemic, elderly men with metabolic diseases were found to be at higher risk of severe forms. In order to improve the early detection of patients who will require intensive care, 62 patients were included in our study. Phenotypic data (age, sex, obesity) and blood biochemistry measurements were collected within the first 2 days of hospitalisation. In addition, a multiplatform approach, consisting in untargeted metabolomics (HILIC and C18) and complementary targeted lipidomics LC-MS/MS analyses, was carried out. The patients were classified at the end of their follow-up as "moderate" or "severe", according to the severity of the developed COVID. Exploratory (PCA, consensus-PCA) and supervised (PLS-DA, multiblock-PLS-DA (MB-PLS-DA), sequential and orthogonalised- PLS-DA (SO-PLS-DA) to predict the COVID severity) data analyses were then performed, using the R package rchemo, to identify early and predictive biomarkers of evolution towards severe forms of COVID.

PCA on separate blocks failed to visualise a variability in the datasets related to the COVID severity on the first axes. PLS-DA models, were valid (based on permutation tests) only for biochemistry and HILIC data, with cross-validated error rates (5-fold repeated 30 times) of 26.56±0.92% and 31.08±2.54%, respectively. Consensus-PCA on the 7 datasets, revealed only a very subtle effect in the data related to the COVID severity. In contrast, the MB-PLS-DA and SO-PLS-DA models were valid, with cross-validated error rates of 25.32±2.40% and 23.33±3.08%, respectively. The most important variables in the MB-PLS-DA model were age, sex and markers of inflammatory response from the different biochemistry and metabolomic datasets. The SO-PLS-DA model used only one latent variable from the biochemistry block and one from a lipidomics dataset. The significant variables were inflammatory markers and metabolites associated with altered metabolic status.

In conclusion, MB-PLS-DA may be an interesting first step to highlight and explain slight effects in the datasets related to the outcome. SO-PLS-DA may be used to deepen the analysis and bring out complementary information. It may also be used to select blocks of data, an advantage for biomarker validation purposes.

DDSIMCA with LOVE

Alexey Pomerantsev[1], Oxana Rodionova[1], Sergey Kucheryavskiy[2]
[1]*Semenov Federal Research Center for Chemical Physics RAS, Russia*
[2]*Aalborg University, Department of Chemistry and Bioscience, Denmark*
svk@bio.aau.dk

Variable selection (VS) is a critical step in improving the performance of chemometric models. While well-established VS methods exist for regression and discrimination, their applicability to one-class classifiers (OCC) remains limited. This study introduces LOVE (Leave One Variable Excluded), a novel VS method specifically designed for OCC, addressing the existing gap in chemometric modeling.

LOVE is a wrapper-based VS approach that integrates interactivity, allowing for iterative refinement of selected variables based on multiple performance indicators. Unlike automated methods, LOVE enables analysts to examine and adjust selection decisions at each step, fostering a more informed and flexible modeling process. The algorithm operates in two modes: discrete (individual variable selection) and interval (grouped variable selection for collinear data, e.g., spectra). The procedure iteratively removes variables or intervals while evaluating performance indicators such as sensitivity, specificity, selectivity and efficiency, ensuring robust classifier optimization.

LOVE was evaluated using several case studies across different analytical platforms, demonstrating its ability to:

1. Enhance classifier performance by selecting optimal subsets of variables.
2. Prevent overfitting and improve model stability.
3. Handle outliers without direct exclusion.
4. Provide deeper insight into influential variables affecting classification decisions.

The interactive nature of LOVE distinguishes it from existing VS methods, offering greater transparency and adaptability in OCC. The ability to adjust performance indicators and explore variable interactions in real-time makes it particularly suitable for chemometric applications. Furthermore, LOVE has been implemented as a freely available web-based tool for DD-SIMCA, facilitating its adoption in diverse analytical scenario

Dealing with non-linearity in the study of near infrared light penetration through matter

Sara Gariglio[1], Cristina Malegori[1], Carolina Scagliarini[3], Paolo Oliveri[1], Monica Casale[1], Eugenio Alladio[3], Giorgia Sciutto[4], Alberto Mazzoleni[3], Emilio Catelli[4], Zelan Li[4], Sergey Kucheryavskiy[5]

[1]*Department of Pharmacy, University of Genova, Viale Cembrano, 4, Genova, Italy*
[2]*Department of Chemistry Industrial Chemistry, University of Genova, Via Dodecaneso, 31, Genova, Italy*
3*Department of Chemistry, University of Torino, Via Pietro Giuria, 5, Torino, Italy*
[4]*Department of Chemistry "G. Ciamician", University of Bologna, Via Guaccimanni, 42, Ravenna, Italy*
[5]*Department of Chemistry Bioscience, Aalborg University, Niels Bohrs Vej, Esbjerg, Denmark*
sara.gariglio@edu.unige.it

The two-year research project INSIDE, funded by the Italian Ministry of Universities and Research – MUR, aims to understand and quantify the penetration of incident near-infrared (NIR) radiation coupled with hyperspectral imaging (HSI), depending on materials analysed and experimental settings, with the belief that a deeper knowledge on the topic would open up for new applications, with the help of advanced chemometrics.
To this goal, ad hoc multi-material samples were built employing a 3D printer, in the shape of 1×1×1 cm and 1×1×3 cm (base × height) parallelepipeds or cylinders. The resulting samples, respectively 37 and 8, were composed with strata of different thickness of polymers with distinct spectral signatures (PLA and PETG). The samples were then analysed via a NIR-HSI camera (Specim, Finland) operating in the reflectance mode within the 1000-2500 nm spectral range, and the obtained spectra were submitted to chemometric data analysis. Exploratory analysis through principal component analysis (PCA) was performed. Subsequently, the stratified components in terms of total height were modelled, and the quantification precision was investigated in terms of root mean square error in prediction. Specifically, partial least squares (PLS) regression was initially applied on the parallelepipedal samples and tested on the cylindric ones. The obtained results pointed out the presence of two consequent linear trends with different slopes, which needed to be modelled separately to optimize performances. Non-linearity of light penetration was then confirmed by ordinary least squares (OLS) and multivariate curve resolution-alternating least squares (MCR-ALS), since both estimated the pixel-based composition of samples with limited accuracy.
Subsequently, specific algorithms to deal with non-linearity were considered and applied. Then, regression by means of convolutional neural networks (CNN) was performed, with the aim of describing the whole range with a single model. Though CNN correctly dealt with the non-linear behaviour of the data, it confirmed that quantification becomes more challenging for deeper layers, as shown by a sudden increase of within-level standard deviation in the prediction results.

Decoding sensory quality by combining analytical chemistry, natural language processing and chemometrics: a multiplatform study on wines

Beatriz Quintanilla Casas[1], Frederikke Hjertholm[1], Julius Jessen Terp[1], Isabella Mei Ling Kledal[1], Sarah Winkler[1], Paul Albert Schneide[1], Mikael Agerlin Petersen[1], Qian Janice Wang[1], Rasmus Bro[1]

[1] *University of Copenhagen, Department of Food Science. Rolighedsvej 26, 1958 Frederiksberg, Denmark*
beatriz@food.ku.dk

Wine quality is a complex attribute shaped by a combination of a wide range of chemical and sensory factors. Traditional assessment methods primarily rely on sensory evaluation, which, despite its relevance, is time-consuming, resource-intensive and subject to variability [1]. This subjectivity is especially pronounced in consumer studies assessing hedonic preferences. Nevertheless, recent advances in analytical chemistry and data-driven approaches offer new opportunities to establish more objective, reproducible, and efficient methods for predicting wine sensory quality and consumer acceptance based on chemical composition.

In the initial part of this study, volatile and non-volatile metabolites were measured in 29 red wines from Margaux (Bordeaux, France) by multiple analytical platforms, including fast spectroscopic techniques e.g., Fourier transformed infrared spectroscopy [2], as well as hyphenated chromatographic techniques. Both targeted and untargeted approaches were conducted, the latter one required utilizing tensor methods - namely parallel factor analysis 2 (PARAFAC2) – to efficiently decompose the chromatographic signal [3].

Besides the above-mentioned analytical measurements, a panel of wine experts performed sensory evaluation of quality endpoints, following the Rate-All-That-Apply (RATA) method for aroma and flavour characteristics. Professional reviews on the wines were also collected from major sources (e.g., Decanter, Purple Pages, Wine Spectator). A natural language processing framework was then applied to extract most common descriptors used by critics, as well as for integrating text, sensory and chemical data into prediction models. The use of machine learning methods is essential for developing robust predicting models of wine quality, while elucidating relationships between wine's chemical makeup and its perceived qualities.

References

[1] Cayuela, J.A., et al. Eur Food Res Technol, 2017 (243), 941-953
[2] Basalekou, M., et al. Beverages, 2020 (6), 30
[3] Johnsen, L. G., et al. J Chromatogr A, 2017 (1503), 57-64

Detection of nutrients and contaminants in the agri-food industry evaluating the probabilities of false compliance and false non-compliance through PLS models and NIR spectroscopy

D. Castro-Reigia[1,2], I. Garcia[2], S. Sanllorente[1], M.C. Ortiz[1], L.A. Sarabia[3]
[1] *1Departamento de Química, Facultad de Ciencias, Universidad de Burgos, Plaza Misael Bañuelos s/n, Burgos 09001, Spain*
[2] *AOTECH (Advanced Optical Technologies S.L.), Edificio Ilgner-BIC Bizkaia Ezkerraldea, Módulo C-7, Avda. Altos Hornos de Vizcaya, 33, 48901, Barakaldo, Bizkaia, Spain*
[3] *Departamento de Matemáticas y Computación, Facultad de Ciencias, Universidad de Burgos, Plaza Misael Bañuelos s/n, Burgos 09001, Spain*
mcortiz@ubu.es

NIR spectroscopy has become one of the most prominent techniques in the food industry due to its easy and fast use. Coupled with PLS, it is a well stablished method for determining nutrients, contaminants or adulterants in foods. Nevertheless, it is not always common calculating the capability of detection or discrimination given a target or a permitted concentration value (in $x \neq 0$) providing probabilities of false non-compliance (α) or false compliance (β). That is the main purpose of this work, where these figures of merit will be evaluated generalizing the ISO-11843 (regulation that is only valid for $x = 0$) by means of the accuracy line as explained in [1], and applied when NIR combined with PLS is used [2]. In that sense, this work shows a versatile procedure that was applied to several analytical determinations in different food matrices (butter, flour, milk, yogurt, olive oil, and olives) and in different agri-food industries for the quantification of protein, fat, salt, and two agrochemicals.

When maximum limits were established at 5 % for the determination of fat in milk, and at 0.6 mg kg^{-1} and 0.5 mg kg^{-1} for the determination of diflufenican and piretrin in olives, detection capabilities of 5.20 %, 1.20 mg kg^{-1}, and 2.34 mg kg^{-1} were obtained, respectively. Also, when minimum limits were established at 1.2 % for salt in butter, at 12 % for protein in flour, at 4 % for protein in milk and at 3 % for protein in yoghurt, capabilities of detection of 1.02 %, 11.45 %, 3.78 % and 2.65 % were obtained, respectively. In every case, probabilities of false non-compliance (α) and false compliance (β) were established at 0.05.

References

[1] Ortiz, M.C., Sarabia, L.A., Sánchez, M.S., Chemometr. Anal. Chim. Acta, 2010 (674), 123-142

[2] Oca, M.L., Ortiz, M.C., Sarabia, L.A., Gredilla, A.E., Delgado, D., Talanta, 2012 (99), 558-565

Development and evaluation of a new algorithm to extract morphological features from FESEM images: a case study on rice kernels

Mattia Sozzi[1], Nicola Cavallini[1], Francesco Savorani[1]
[1] *Department of Applied Science and Technology, Polytechnic of Turin, Corso Duca Abruzzi 24, Turin, Italy*
mattia.sozzi@polito.it

The automatic extraction of morphological features from images is of fundamental importance for efficiently processing large datasets. In this research, a novel algorithm was developed to extract diverse morphological features from FESEM images, starting from a specific application on rice kernels. In this context, a dataset of 220 images from 54 distinct rice varieties was used to develop and test the algorithm [1]. The aim of this approach is to identify the round-shaped starch granules naturally present in rice kernels from FESEM greyscale images and extract a set of morphological features from them.
The algorithm follows a structured workflow that includes image segmentation, object identification, and feature extraction. Through these steps, it quantifies morphological features such as granule area, perimeter and eccentricity, as well as the number of detected objects, and the empty spaces between them, which correspond to kernel porosity (Figure 1). The use of adaptive thresholds and correction steps enhances robustness, enabling the analysis of diverse images while effectively filtering out defective objects and non-representative areas.

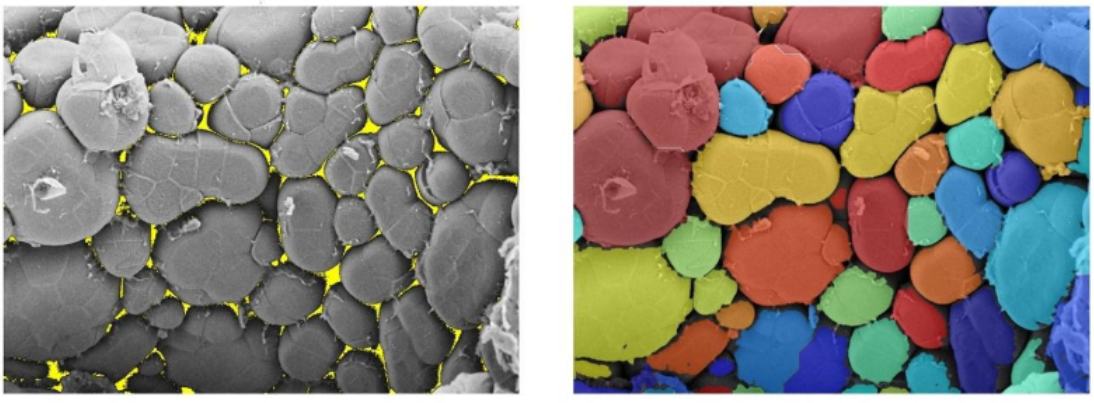


Figure 1. Output image generated by the algorithm. Left: representation of the empty spaces (in yellow) that correspond to the porosity. Right: representation of the starch particles identified by the algorithm and coloured differently to visualize the result obtained before the correction steps.

Furthermore, to evaluate and optimize the algorithm's performance, a Design of Exper-

iment (DoE) approach was employed to assess the influence of input parameters on the output results and on algorithm processing runtime. The experimental data were subsequently inspected using ANOVA Simultaneous Component Analysis (ASCA) and Multi-Linear Regression (MLR).

References

[1] F. Haxhari et al., "Endosperm structure and Glycemic Index of Japonica Italian rice varieties", Front Plant Sci, 2023 (14), doi: 10.3389/fpls.2023.1303771

Evaluation of Soil REE Perturbations on Plant Accumulation and Distribution Patterns by Multivariate Data Processing

Marcella Barbera[1], Riccardo Voccio[2], Cristina Malegori[2], Sara Gariglio[2], Paolo Oliveri[2], Filippo Saiano[3], Daniela Piazzese[1]

[1] *University of Palermo, Department of Earth and Marine Sciences, Via Archirafi, 22, 90123 Palermo, Italy*
[2] *University of Genova, Department of Pharmacy (DIFAR), Viale Cembrano, 4, 16148, Genova, Italy*
[3] *University of Palermo, Department Agricultural Food and Forestry Sciences, 90128 Palermo, Italy*
marcella.barbera@unipa.it

The demand for Rare Earth Elements (REE) is constantly increasing due to their use in advanced technologies, making them emerging contaminants. Although non-essential to plants, REE can accumulate in vegetal tissues and transfer through trophic levels. This raises questions about how plants respond to changes in REE composition in the soil.
The present work explores the influence of soil perturbation by REE on their distribution and accumulation in *Vitis Vinifera L.* plant using multivariate data processing. For this purpose, a full factorial (3^2) experimental plan was designed, varying two factors at three levels: treatment (control, low, high) and position (soil, roots, leaves). In more detail, for treatment two stress levels were applied by adding equimolar REE solutions to the soil: in the "low" condition, REE concentration [REE] was $\sim$3 times higher than in the "control" one; in the "high" condition [REE] was $\sim$50 times higher. At full leaf development, roots, leaves, and soil were collected and REE determined by ICP-MS. The high differences in REE concentration between treatments and positions, along with the effects of equimolar addition, prevented the detection of evident correlation patterns between factors and variables. Therefore, a two-step pre processing strategy was implemented: orthogonalization with respect to PC1, to minimize concentration effects while preserving the design structure, followed by sum-100 normalization, to minimize global intensity effects due to the equimolar addition. Subsequently, Principal Component Analysis (PCA) successfully differentiated REE profiles across treatments, enabling unsupervised differentiation of experimental conditions based on multi-elemental distribution patterns, rather than on absolute concentrations. To identify meaningful associations between factors and REE patterns, ANOVA–Simultaneous Component Analysis (ASCA) was applied. As expected from the experimental design, treatment had the strongest effect. Notably, treatment × position interactions were more significant than the position alone, indicating that REE distributions may change with treatment in a compartment-specific manner. In conclusion, the present study reveals that REE accumulation and distribution in plant tissues depend not only on total REE levels but also on their relative distribution in the soil, suggesting valuable implications for environmental monitoring, pollution assessment and geographical traceability.

Experimental and computational decomposition of complex biomedical samples

Teresa Slanina[1], Jana Hahn[1], Pooja Gune[1], Wolf Holtkamp[1], Sascha Hein[1], Christel Kamp[1]
[1] *Paul-Ehrlich-Institute, 63225 Langen, Germany*
teresa.slanina@pei.de

Biomedical products like vaccines, therapeutic allergen products or antibody preparations are complex formulations of proteins and excipients. Their quality is of great interest for patient safety. This why it is strictly controlled according to European and national guidelines and legislation often involving batch release testing and animal experiments. Raman spectroscopy is investigated as an alternative method to study the chemical composition of biomedicines and to ensure the high quality of the product. Acknowledging the dominance of excipients spectra we explore whether and how it will be possible to recover protein spectra from the spectrum of the full product. To reach this goal both experimental and computational techniques are used: experimental methods include dialysis, sample concentration or freeze drying. Computational methods aim to differentiate between protein and excipient spectra using Machine Learning algorithms to unmix spectra. We discuss how a combination of experimental and analytical methods can lead to a characterization of product components based on the product's Raman spectral fingerprint.

Explainable ML: Machine learning interpretability methods applied to spectroscopic data

Astrid Maléchaux[1], Jordane Poulain[1], Sylvie Roussel[1]
[1] *Ondalys , 4 rue Georges Besse, 34830 Clapiers, France*
sroussel@ondalys.fr

In the past decades, Machine Learning (ML) models have become more and more complex, leading to improvements in their predictive performance. However, these models can often be described as "black boxes". There is a growing need to improve their interpretability in order to have greater confidence in their results [1], providing the so-called Explainable AI (Artificial Intelligence).
Thus, this study is focused on the interpretability of Machine Learning models based on near- infrared spectroscopic data, a.k.a. the "post-hoc explanation" of ML models. Different regression models such as Support Vector Machines (SVM), Artificial Neural Networks (ANN) and Extreme Gradient Boosting (XGBoost), have been trained and compared to the classical PLS calibration models.
Various ML model interpretability algorithms currently applied to non-spectroscopic data have been tested and compared: Local Interpretable Model-agnostic Explanations [2], Shapley Additive Explanations [3] and pseudo-samples prediction [4]. They have also been compared to the regression coefficients of an intrinsically interpretable Partial Least Squares model. While the applicability to spectroscopic data of pseudo-samples prediction has been demonstrated [4], Local Interpretable Model-agnostic Explanations and Shapley Additive Explanations are theoretically better suited to uncorrelated variables [5].
The results indicate that the outputs of the interpretability methods tested appear to be in good agreement with the Partial Least Squares regression coefficients. Thus, they allow the identification of wavelengths ranges used by "black box" Machine Learning models. Moreover, they could be used as indicators of the risk of overfitting for complex models developed on spectroscopic data.
In conclusion, this study shows the potential for applying different Machine Learning model interpretability methods to Near Infrared spectroscopic data. Further work on different NIR datasets will be conducted to confirm these results

References

[1] P. Linardatos, V. Papastefanopoulos, S. Kotsiantis, Explainable AI: A Review of Machine Learning Interpretability Methods, Entropy, 2021 (23), 18
[2] M. T. Ribeiro, S. Singh, C. Guestrin, "Why should I trust you?" Explaining the predictions of any classifier, Proceedings of the 22nd ACM SIGKDD international conference on

knowledge discovery and data mining, 2016, 1135-1144
[3] S. M. Lundberg, S. Lee, A Unified Approach to Interpreting Model Predictions, Advances in Neural Information Processing Systems, 2017 (30), 4765-4774
[4] G. J. Postma, P. W. T. Krooshof, L. M. C. Buydens, Opening the kernel of kernel partial least squares and support vector machines, Analytica Chimica Acta, 2011 (705), 123-134
[5] A. M. Salih, Z. Raisi- Estabragh, I. B. Galazzo, P. Radeva, S. E. Petersen, K. Lekadir, G. Menegaz, A perspective on explainable artificial intelligence methods: SHAP and LIME, Advanced Intelligent Systems, 2024 (7), 2400304

Exploring the Limits of Chemometric Models in Raman Imaging of Biological Samples

Eneko Lopez[1], Giulia Gorla[2], Julene Aramendia[2], Kepa Castro[2], Jaione Etxebarria-Elezgarai[1,3], Nerea García Velasco[4], Manu Soto[4], Urtzi Izagirre[4], Andreas Seifert[1,5], Jose Manuel Amigo[2,5]

[1]*CIC nanoGUNE BRTA 20018 San Sebastián, Spain*

[2]*Department of Analytical Chemistry, University of the Basque Country UPV/EHU, 48940 Leioa, Spain*

[3]*OPTEC4LIFE 20018 San Sebastián, Spain*

[4]*Research Centre for Experimental Marine Biology & Biotechnology, University of the Basque Country (UPV/EHU) 48080 Plentzia, Spain*

[5]*Ikerbasque, Basque Foundation for Science 48009 Bilbao, Spain*

e.lopez@nanogune.eu

Raman imaging is a powerful tool for the chemical and surface analysis of complex biological samples. However, due to the complexity of the signals normally obtained, the limitations of chemometric modelling in this context have not yet been sufficiently explored, particularly when applied to heterogeneous matrices such as biological fluids and tissues. Here, we investigate the limits and potential of chemometric strategies such as principal component analysis (PCA), k-means clustering, and multivariate curve resolution-alternating least squares (MCR-ALS), to extract meaningful biochemical information [1]. Moreover, we propose an integrative framework that combines these approaches, exploiting their complementary strengths to improve data interpretation. Two different challenging scenarios for Raman imaging are presented: dried plasma droplets for Alzheimer's studies and cryosectioned mussel tissue with microplastics internalized. In the first case, hyperspectral Raman imaging revealed a complex spatial heterogeneity that goes beyond the classical "coffee ring" pattern and highlights how molecular biomarkers can concentrate in unexpected zones. In the second case, volumetric confocal Raman imaging [2] enabled the detection and 3D localisation of 1 μm polystyrene particles internalised within digestive epithelial tissues, while demonstrating the necessity of advanced image preprocessing and strict contamination control.

By evaluating the performance of these combined chemometric models in these two different cases, we illustrate both the power and the pitfalls of applying multivariate tools to Raman data. Ultimately, this study demonstrates the need for cautious interpretation, model validation, and contextual awareness when pushing the boundaries of chemometric applications in biological Raman imaging.

References

[1] Haessler, A., Jung, N., Windbergs, M., Anal. Chem., 2023 (95), 17646-17653

[2] Wang, N., Wang, L., Feng, G., et.al., Laser Photonics Rev., 2025 (19), 2401444

Extracting temporal patterns from longitudinal microbiome data through clustering of functional variables

Véronique Cariou[1], Jeremy Coindet[1], Camille Privard[1], Manon Salazar[1], Yoba Kandé[1], Olivier Chapleur[2]
[1] *Oniris, STATSC, 44300 Nantes, France*
[2] *Université Paris-Saclay, INRAE, PRocédés biOtechnologiques au Service de l'Environnement, 92160 Antony, France*
veronique.cariou@oniris-nantes.fr

In terms of sequencing, recent technological advances offer the possibility of rapid data acquisition at relatively low cost. These advances naturally pave the way for longitudinal studies to understand not only the composition of a community of microorganisms and their interactions, but also the kinetics of these microbial systems over time. Among the challenges associated with such longitudinal studies, the partitioning of microorganisms according to their evolutionary patterns is of crucial interest to microbiologists [1], in order to reduce the inherent high data dimensionality. To this end, we propose to adapt the clustering of variables CLV, traditionally performed on a matrix or three-way of quantitative variables [2, 3], to the case of functional variables in the scope of longitudinal omics. These data can be represented by an n x p matrix whose entry (i, j) is the function j corresponding to the individual i. Moreover, it is assumed that all functions are defined on the same compact real interval $[c, d]$ and take values in $\mathbb{R}$. Functional CLV aims at partitioning the set of the p functional variables into homogeneous groups, so that each variable is associated as closely as possible with the latent functional variable of the group to which it belongs. We show how to adapt the CLV criterion using the framework of multivariate functional PCA [4], and the associated partitioning algorithm is then detailed. Functional CLV is applied to longitudinal metataxonomic data in the context of the anaerobic digestion bioprocess [5, 6]. In particular, the aim is to identify groups of microorganisms based on common longitudinal profiles that also reflect potential differences in their dynamics towards salinity stress.

References

[1] Kodikara, S., Ellul, S., Lê Cao, K. A., Brief. Bioinform., 2022 (23.4), bbac273
[2] Vigneau, E., Qannari, E. M., Commun. Stat. Simulat., 2003 (32.4), 1131-1150
[3] Cariou, V., Alexandre- Gouabau, M. C., Wilderjans, T. F., J. Chemometr., 2021 (35.2), e3269
[4] Happ, C., Greven, S., J. Am. Stat. Assoc., 2018 (113.522), 649-659
[5] Bodein, A., Chapleur, O., Droit, A., Lê Cao, K. A., Front. Gen., 2019 (10.963)
[6] Chapleur, O., Poirier, S., Guenne, A., Lê Cao, K. A., Chemosphere, 2021 (283.131309)

Fusion by design: a new paradigm for the analysis of disparate datasets with independent samples and features in a common factor space

Michael Sorochan Armstrong[1], José Camacho[1]
[1] *Research Centre for Information and Communication Technologies (CITIC-UGR), University of Granada, Spain*
mdarmstr@go.ugr.es

Traditional omics analyses are expensive, which puts practical limits on the number of study participants. This has consequences for the statistical power of each study. Making the most of the fact that in most cases funding agencies and journals often mandate that the data be shared publicly in online repositories, an appealing possibility is to perform meta-analyses that combine several studies with increased statistical power. High level fusion techniques can test existing hypotheses across multiple datasets, only if the hypotheses (i.e. evidence of statistical significance) can be tested across a consistent group of features. Low level fusion techniques must contain some common modality across either the samples or features to properly concatenate multiple datasets [1]. For instances with no common sample or feature modality (herein: heteromodal data), intermediate-level data fusion - relating information across multiple datasets at the analysis stage, has been explored using kernel approaches and direct decompositions of sparse, block matrices [2]. Validating datasets without a common feature or sample modaility (herein: hetero-modal datasets) has relied on the ability of the model to reconstruct known data that is used as the test case. This does not establish the *significance* of the relationship, but rather highlights the model's ability to make predictions. This approach may also be inefficient at handling datasets variance associated different experimental controls.
The authors propose a new method for intermediate data fusion of heteromodal data, based on examining subspaces of similar dimensions that are dictated by common experimental information. A method for testing whether these data are sampled from a common distribution is proposed based on a newly-disclosed method for permutation testing that can handle the sign-ambiguity of decomposition by Principal Component Analysis (PCA).

References

[1] Smilde, A., Tormod, N., Hovde Liland, K., Multiblock data fusion in statistics and machine learning: Applications in the natural and life sciences, 2022
[2] Žitnik, M., and Zupan, B., 2014. Data fusion by matrix factorization. TPAMI, 2014 (37), 41-53.

G-CovSel: covariance oriented variable clustering

Alessandra Biancolillo[1], Bénédicte Favreau[2], Federico Marini[3], Jean-Michel Roger[4,5]
[1] *Department of Physical and Chemical Sciences, University of L'Aquila, Via Vetoio 67100, L'Aquila, Italy*
[2] *CIRAD, UMR AGAP CIRAD, Montpellier, France*
[3] *Department of Chemistry, University of Rome La Sapienza, Piazzale A. Moro 5, 00185, Rome, Italy*
[4] *ITAP-INRAE INRAE, Institut Agro, University Montpellier, 34196 Montpellier, France*
[5] *ChemHouse Research Group ChemHouse, Research Group, Montpellier, 34196, France*
alessandra.biancolillo@univaq.it

Analytical chemistry devices generate large quantities of variables with varying degrees of direct linkage to the phenomena of interest. Some measurement methods are highly indirect, resulting in strongly correlated variables, while others are more direct and yield less correlated variables.

Variable selection is essential for identifying the different phenomena linking predictors and responses. Covariance Selection (CovSel) [1] has been developed to address this need by selecting non-redundant variables parsimoniously. CovSel's deflation process ensures that the selected variables are linked to different sources of covariance between the measurements and the phenomena under study, making it a valuable tool for identifying key factors. However, in some cases, identifying one variable per factor is insufficient, as, for instance, in mass spectrometry.

G-CovSel [2], an extension of CovSel, is designed to select groups of variables, where each group is linked to a phenomenon explaining a set of responses. This approach has been tested on various real datasets and compared to CovSel. While CovSel focuses on selecting non redundant variables parsimoniously, g-CovSel enhances this by forming groups around the variables selected by CovSel. The CovSel algorithm has been modified to consider both the covariance of the selected variables with the predicted response and the inter-group correlation.

Two algorithms for creating groups will be proposed and illustrated. The variables selected by CovSel are found at the center of the groups formed, making g-CovSel a generalization of CovSel. Most groups formed can be interpreted functionally, making g-CovSel an excellent tool for biomarker identification in omics.

References

[1] Roger, J-M., Palagos, B., Bertrand, D., Fernandez-Ahumada, E., Chemometr. Intell. Lab., 2011 (106), 216-223

[2] Roger, J-M., Biancolillo, A., Favreau, B., Marini, F., Chemometr. Intell. Lab., 2024 (254), 105223

How to deal with large multispectral images comparison? PCA score distribution as a tool.

Marie-Françoise Devaux[1], Fabienne Guillon[1], Camille Alvarado[1], Oscar Main[2], Cécile Barron[2]
[1] *UR1268 BIA, INRAE, 44316 Nantes, France*
[2] *UMR IATE, Univ Montpellier, INRAE, Institut Agro, 34060 Montpellier, France*
cecile.barron@inrae.fr

Multispectral autofluorescence images can be acquired using automated microscopes with a spatial resolution of 1 μm and a large field of view (1 to 2 cm^2). Such acquisition systems are well suited to observe whole organ sections, while enabling tissue identification. They allowed the acquisition of an increased number of images, although the sample preparation is still the bottelneck. These large image series pave the way for statistical analysis, and could be used to explore the diversity of plant structure and cell wall spectral properties as a function of different factors (genetic, environmental and agricultural conditions), but also on technological fractions obtained from these plant ressources. To quantify the differences between images we proposed to take advantage of the pixel score distributions after Principal Component Analysis (PCA).
Series of 30-60 large images may contain millions to billions of pixels. Principal Component Analysis can be applied to image series by iteratively computing the variance-covariance matrix and the score images [1]. It is called *large* PCA. For each principal component, the pixel score distribution across the entire image series and its percentiles are determined. Local PCA score distributions are then calculated for each image using these percentiles. These local score distributions are regarded as quantitative characteristics enabling multispectral images to be compared.
Two examples based on maize stem section and ground maize fractions are presented. Autofluorescence multispectral images were acquired and specific attention was paid to UV and visible fluorescence variability. First of all, 4 maize inbred lines were compared based on 40 stem imaged sections. The overall comparison was carried out by looking at the principal component analysis of the score distributions. Similar approach was used to compare 6 grinding fractions obtained from 6 maize inbred lines.
Pixel score distributions are found to be a promising tool for statistical comparison of multispectral images. The method is easy to implement, can be used as a first descriptive analysis of the image series, and can be easily extended to other spectral imaging techniques.

References

[1] Devaux, M.F., Corcel, M., Guillon, F., Barron, C. Biomolecules., 2023 (7), 11

Funding: French ANR agency, PEPR B-BEST FillingGaps project (ANR-23-PEBB-0006)

Hyperspectral Imaging for Monitoring of Lanthanides in Ion Exchange Chromatography Columns

Cannon Giglio[1], Kristian Myhre[2], Luke Sadergaski[1]
[1] *Radioisotope Science and Technology Division, Oak Ridge National Laboratory, 1 Bethel Valley Road, Oak Ridge, TN 37830, USA*
gigliocj@ornl.gov

Ion exchange chromatography is ubiquitous within radioisotope production due to its relative ease of implementation and effectiveness for chemical separations. However, the performance of ion exchange chromatography columns remains challenging to optimize for certain chemical elements, such as the lanthanide series, which share especially similar chemistries. Enhanced performance could be achieved with a deeper understanding of how reactive flows behave throughout the column rather than only output monitoring, which is the current state of the art. This presentation will highlight the capability of visible–near-infrared hyperspectral imaging to allow for spatiotemporal mapping of key analytes during ion exchange chromatography and eventually more effective lanthanide separations with ion exchange chromatography.

The capabilities and usage of hyperspectral imaging has rapidly grown in recent years due to the ability to incorporate spatial information in addition to the chemical knowledge given by spectroscopic signals. Furthermore, the acquisition of hyperspectral imaging over time can also unlock a chronological dimension. Although hyperspectral imaging has seen massive growth in areas such as food analysis and pharmaceuticals [1], few studies have examined the potential for direct imaging of ion exchange chromatographic columns. The proposed approach will enable monitoring of the concentration of the mixture components along the length of the column, as well as an investigation of how each component moves down the column over time.

The development of this monitoring capability will enhance the radioisotope production process by providing real-time feedback on the status of a separation, which can be used to inform mid-run adjustments to key parameters, such as flow rate and solvent composition. This approach will also strengthen research and development capabilities by making it easier to quantitatively calculate key chromatographic parameters for testing new mobile phases or column packing materials. Furthermore, this methodology may be equally valuable in monitoring chromatographic columns in other fields.

References

[1] Amigo, J. M., Hyperspectral Imaging, Elsevier, 2019

Image-based modeling of three-way data for butter authentication

Ciro Orecchio[1], Sergey Kucheryavskiy[2]
[1] *University of Turin Via P. Giuria 7, Turin, Italy*
[2] *Aalborg University, department of Chemistry and Bioscience, N. Bohrs Vej 8, 6700, Esbjerg, Denmark*
ciro.orecchio@unito.it

In recent years, the issue of food authentication has received increasing attention due to the growing consumer demand for high quality and regionally distinctive food products [1]. As a result, food safety and authenticity detection play a crucial role in safeguarding consumers' interest, especially in cases involving high value-added ingredients [2]. Among these, butter, is practically susceptible to intentional adulteration with low-quality additives for economic gain.

Currently, some of the most effective analytical techniques for detecting food adulteration rely on three-way methods, such as fluorescence spectroscopy and GC-MS. Classical approaches for analysing such multiway data include well-established methods like PARAFAC and Tucker decomposition. However, these techniques are not always suitable for building robust discrimination or authentication models.

This study explores an alternative strategy: treating three-way datasets as images, so two modes (e.g. excitation and emission wavelength) serve as spatial dimensions and the spectral intensities represent pixels values. This transformation allows the application of powerful image analysis models, particularly Convolutional Neural Networks (CNN) – that are widely used for classification and anomaly detection. These models can be pre-trained on large image datasets and subsequently fine-tuned for specific cases, thereby reducing the need for extensive domain-specific training data.

To validate this approach, a butter authentication case study was conducted. Samples from six different commercial brands of butter were collected from various retail sources. A subset of these samples was adulterated with 5% and 10% w/w of coconut oil, resulting in a total of 91 samples. Excitation emission matrices (EEMs) were acquired for each sample using two acquisition setups.

As a preliminary step, a CNN was trained to perform a discriminant analysis for detecting the presence or absence of adulterants. Subsequently, since the focus is on identifying a single target class (pure butter), a one-class modelling approach was adopted [3]. Specifically, a Variational Autoencoder (VAE), a class of deep learning models designed for image representation - was combined with the SIMCA algorithm to construct a one-class classifier [4]. All experimental results and methodological specifics will be disclosed during the presentation.

References

[1] Rodionova, O. Y., Oliveri, P., Malegori, C., et al. Trends Food Sci. Tech., 2024 (147), 104429

[2] Wang, Y., Gu, H. -W., Yin, X. -L., et al. Trends Food Sci. Tech., 2024 (146), 104396

[3] Oliveri, P., Anal. Chim. Acta, 2017 (982), 9-19

[4] Petersen, A., Kucheryavskiy, S., Chemometr. Intell. Lab., 2025 (256), 105276

In-silico QSRR prediction of chromatographic retention time

Emmanuel Cruz Muñoz[1], Davide Ballabio[1], Viviana Consonni[1], Cristian Rojas[2], Veronica Termopoli[1], Roberto Todeschini[1]

[1] *Milano Chemometrics and QSAR Research Group, Department of Earth and Environmental Sciences, University of Milano-Bicocca P.za della Scienza 1, 20126, Italy*

[2] *Grupo de Investigación en Quimiometría y QSAR, Facultad de Ciencia y Tecnología, Universidad del Azuay Av. 24 de Mayo 7-77 y Hernán Malo, Cuenca 010107, Ecuador*

enmanuel.cruzmunoz@unimib.it

In-silico approaches providing information on chromatographic retention times are of great interest because they can support compound identification in untargeted analyses, thus reducing further experimental efforts during method development stages. Molecular properties influencing retention time can be encoded into chemical-structural descriptors, allowing the modeling with Quantitative Structure-Retention Relationships (QSRRs), which relies on the principle that retention time is influenced by a complex synergy of molecular properties, such as size, shape, polarity, and electronic distribution. By using appropriate chemometric techniques, QSRR endeavors to establish quantitative relationships between molecular descriptors and the chromatographic retention time. This can be used to predict the elution order and separation efficiency of compounds, which is fundamental in the design of chromatographic methods [1]. We will present two studies carried out in this framework.

In the first case study, we modelled 424 plant food bioactive compounds associated with experimental values of retention time measured in three different liquid chromatography mass spectrometry (LC-MS) systems. Feature selection with Genetic Algorithms (GAs) allowed the development of parsimonious regression models based on ordinary least squares, which were validated by means of a test set (Q^2 =0.86, 0.80, 0.83 for the three modelled chromatographic systems). In the second case study, supercritical fluid chromatography-high resolution mass spectrometry (SFC-HRMS) retention time data of 438 pesticides have been employed to develop a QSRR-based regression model. This model was adapted for the prediction of retention times of a different chromatographic system based on low-resolution mass spectrometry (SFC-LRMS). A kernel-based approach was used to account for prediction uncertainties, as well as to leverage model reliability by defining a structural domain in chemical space where lower uncertainty is expected [2]. Validation results (mean average error lower than 1.2 minutes for test compounds) demonstrated the possibility to predict retention times in chromatographic systems similar to those used for the calibration of QSRR models with enough accuracy. In analytical chemistry, this approach can therefore improve prediction transferability across chromatographic systems, enhancing QSRR model application for retention time prediction.

References

[1] Stanstrup, J., Neumann, S., Vrhovšek, U., Anal. Chem., 2015 (87)

[2] Consonni, V., Todeschini, R., Orlandi, M., Ballabio, D., Chemometr. Intell. Lab., 2024 (246)

Influence of a measurement procedure and evaluation of transflectance sensing system for quantifying sunflower oil adulterations in olive oil

D. Castro-Reigia[1,2], M. Sierra[2], I. García[2], S. Sanllorente[1], L.A. Sarabia[3], M.C. Ortiz[1]
[1] *Departamento de Química, Facultad de Ciencias, Universidad de Burgos, Plaza Misael Bañuelos s/n, Burgos 09001, Spain*
[2] *AOTECH (Advanced Optical Technologies S.L.), Edificio Ilgner-BIC Bizkaia Ezkerraldea, Módulo C-7, Avda. Altos Hornos de Vizcaya, 33, 48901, Barakaldo, Bizkaia, Spain*
[3] *Departamento de Matemáticas y computación, Facultad de Ciencias, Universidad de Burgos, Plaza Misael Bañuelos s/n, Burgos 09001, Spain*
lsarabia@ubu.es

The development of NIR instruments or their modification to adapt the measurements for each problem and improve its performance, are crucial steps for the optimal measurement procedures [1]. In this work it is presented the development of an accessory for cuvettes designed to have the possibility to collect NIR spectra in transflectance mode. In that sense, it is aimed to investigate how different factors in the measurement procedure using this accessory influence both the NIR spectra and the subsequent calibration models for detecting adulterations with sunflower oil in olive oil. Every measurement condition influencing the spectra was evaluated with ASCA [2], visualizing how the use of different NIR devices, the sensor arrangement regarding the cuvette, the activation of the internal compensation system of temperature of the sensor, or the concentration levels of the adulterant affected the resulting spectra. Afterwards every possible combinations of the factors were explored through eight different PLS calibration models and their validation, to examine if the factors also influenced the calibration models built for quantifying the sunflower oil present in different mixtures of olive oil. It was found that not only were all factors significant regarding NIR measurements, but also when quantifying adulterants. The best results were obtained by arranging the sensor in horizontal disposition regarding the cuvette and activating the internal compensation system of temperature. The capability of detection of the method was 1.4 % for probabilities of false positive and false negative of 0.05.

References

[1] Feng, X., Larson, R.A., Digman, M.F., Remote Sens., 2022 (14)
[2] Bertinetto, C., Engel, J., Jansen, J., Anal. Chim. Acta, 2020 (6), 100061

Integrating Multi-Analytical Data and Genetic Algorithm for Descriptor Identification: A Mid-Level Data Fusion Approach to Biomass Characterization

Ramtanon Ian[1], Lacoue-Negre Marion[1], Berlioz-Barbier Alexandra[1], Le Masle Agnès[1], Remy Simon[2], Martinez Agathe[2], Lameiras Pedro[2], Renault Jean-Hugues[2]
[1] *IFP Energies Nouvelles, Rond-point de l'Échangeur de Solaize BP 3, 69360 Solaize, France*
[2] *Université de Reims Champagne-Ardenne, CNRS, ICMR, UMR 7312, Reims, France*
marion.lacoue-negre@ifpen.fr

Considering the current environmental challenges, transformation of lignocellulosic biomass into bioethanol offers a sustainable alternative to petroleum fuels. In some bioconversion processes, a pretreatment step is applied to deconstruct the biomass and produce a cellulose rich solid soaked by an aqueous phase called hydrolysate. This complex mixture consists of various oxygenated compounds that may act as inhibitors during enzymatic hydrolysis (EH) of the cellulose-rich solid. Confident identification of these compounds is crucial for the optimization of the bioconversion process. Therefore, a complete analytical workflow was developed including nuclear magnetic resonance (NMR) and liquid chromatography hyphenated to tandem high-resolution mass spectrometry (LC-HRMS/MS) [1]. Then, a supervised chemometric approach involving variable selection through genetic algorithms (GA), fusion of the complementary analytical datasets and partial least squares discriminant analysis (PLS-DA) [2], was used to extract inhibitor candidates from the complex analytical datasets.

Two distinct approaches were compared to identify candidates for EH reactivity. The first approach used the GA-PLS-DA sample scores to pinpoint the latent variables that were the most relevant to EH reactivity-prediction. These latent variables were then used as EH reactivity-related directions on the individual ESI(+), ESI(-) and 13C NMR models for the identification of key descriptors. The second mid-level fusion approach explored the potential of applying PLS-DA to concatenated analytical datasets after GA subset selection. Ultimately, the variables highlighted by both approaches would be considered to be confident descriptors of EH reactivity. A selection workflow involving four criteria (VIP scores, SR, GA selection frequency, and cosine similarity score) was used to highlight 15 key descriptors of low EH reactivity trends. The ten ESI(+) and ESI(-) features were submitted to molecular annotation workflows, revealing the presence of levulinic acid and 4-hydroxybenzoic acid. EH tests were performed doping the samples with these compounds and inhibition effects were observed, validating the entire workflow developed in this study

References

[1] Ramtanon, I., Berlioz-Barbier, A., Remy, S., Renault, J-H., Le Masle, A., J. Chromatogr. A,

2023 (1706), 464277
[2] Ramtanon, I., Lacoue-Negre, M., Berlioz-Barbier, A., Le Masle, A., Renault, J-H., Chemometr. Intell. Lab., 2025 (261), 105381

Integration of Omics and Spectroscopic Approaches for Novel Potential Biomarker Discovery of Hepatocellular Carcinoma

Antonio Malincolinco[1,2], Vicky Caponigro[2], Fabrizio Merciai[2], Danila La Gioia[2], Eduardo Maria Sommella[2], Pietro Campiglia[2]

[1] *PhD program in RNA Therapeutics and Gene Therapy Department of Pharmacy*
[2] *Department of Pharmacy, University of Salerno Via Giovanni Paolo II, 132, 84084 Fisciano (SA), Italy*
antonio.malinconico@unina.it

Hepatocellular carcinoma (HCC) is one of the most common malignancies worldwide, with late diagnosis due to the initial absence of symptoms, significantly impacting survival rates [1]. New biomarker discoveries enable both earlier cancer detection and precisely targeted gene therapies, particularly through RNA-based approaches like siRNA and mRNA, poised to transform individualised patient care [2]. Plasma from 101 HCV-infected patients, including 69 with confirmed HCC, was analysed by mass spectrometry and infrared spectroscopy. For mass spectrometry, plasma was fractionated into polar and non-polar phases using the Matyash protocol [3], whereas ATR-FTIR spectroscopy was performed directly on the untreated plasma. The generated datasets underwent pre-processing and were evaluated independently and with multiblock approaches in Matlab R2024b (MathWorks Inc., Natick, MA, USA). Metabolomics dataset was normalised using TIC, whereas Lipidomics dataset was normalised by internal standards specific for each lipid class. The ATR-FTIR dataset was pre-processed using: Asymmetric Least Squares algorithm [4] (λ = 103, p = 10-4) for baseline correction, Multiplicative Scatter Correction (mean of the spectra as the reference) [5] and External Parameters Orthogonalization (removing 7 PCs) [6]. The dataset was split into training (80%) and test (20%) sets using the Kennard-Stone algorithm. Initial classification models based on PLS-DA [7] performed on each dataset with cross validation (leave one out) yielded suboptimal performance on the test set (Lipidomics = 85% accuracy, Metabolomics = 90% accuracy, ATR-FTIR = 80% accuracy). Subsequently, a data fusion strategy employing SO-CovSel-LDA [8,9] was implemented. By evaluating all combinations, two classification models were selected: one integrating Metabolomics and ATR-FTIR data, and the other combining Lipidomics and ATR-FTIR data, achieving accuracy $\geq$ 95% on the test set. Selected variables reveal hallmark metabolic reprogramming in tumorigenesis, marked by reduced lysophosphatidylcholines, acylcarnitines and ceramides, as well as an upregulation of the kynurenine pathway, a process implicated in immunosuppression. Additionally, ATR-FTIR data provided critical insights into modifications in protein content.

References

[1] Sayiner, M. et al., (2019). Digestive Diseases and Sciences, 64 (4), 910–917
[2] Younis, M. et al., (2020). Advanced Therapeutics, 3 (11)
[3] Matyash, V. et al., (2008). Journal of Lipid Research, 49 (5), 1137–1146

[4] Peng, J. et al., (2010). Analytica Chimica Acta, 683 (1), 63–68
[5] Geladi, P. et al., (1985). Applied Spectroscopy, 39 (3), 491–500
[6] Roger, J. et al., (2003). Chemometrics and Intelligent Laboratory Systems, 66(2), 191–204
[7] Ståhle, L., & Wold, S. (1987). Journal of Chemometrics, 1 (3), 185–196
[8] Biancolillo, A. et al., (2020). Journal of Chemometrics, 34 (2)
[9] Biancolillo, A. et al., (2015). Chemometrics and Intelligent Laboratory Systems, 141, 58–67

Linking Colour, Chemistry, and Craft: A Chemometric Study of 12th-Century Bronze Doors

Nicola Cavallini[1], Giorgia Ghiara[1], Marianne Mödlinger[2], Marina Cocchi[3], Francesco Savorani[1],

[1] *DISAT, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy*
[2] *DCCI, Università degli Studi di Genova, Via Dodecaneso 31, 16136 Genova, Italy*
[3] *DSCG, Università di Modena e Reggio Emilia, Via Campi 103, 41125 Modena, Italy*
nicola.cavallini@polito.it

The three leaded bronze doors by Barisanus of Trani, dating to the late 12th century, are among the largest surviving groups of medieval monumental bronzes in Europe. As part of the GAPAMET project [1], they were analysed using Energy Dispersive X-Ray Fluorescence (EDXRF, a fast, cost-effective method for in-situ metal analysis [2]) alongside high-definition RGB imaging [3]. Chemometric tools [4] were applied to validate and interpret the spectroscopic and imaging data, enabling insights into corrosion patterns and material composition. PCA, cluster analysis, and data-fusion techniques revealed spatial trends and relationships between chemical content and surface appearance. The study aims to connect EDXRF results with RGB-based assessments of corrosion and colour, relating the findings to historical knowledge of bronze production methods (i.e., casting and thermo-mechanical treatments) and supporting cultural heritage understanding and conservation.

Figure 1. The RGB images of the three analysed Barisanus doors: (a) Monreale, (b) Ravello, (c) Trani

References

[1] Gates to Paradise: Creating metal doors for 11th-12th century churches. Research funded

by the Austrian Science Fund (FWF) [10.55776/P-34477-G]

[2] Šatović D., Desnica V., Fazinić S., Spectrochim Acta B, 2013 (89), 7–13
[3] Prats-Montalbán J.M., de Juan A., Ferrer A., Chemometr. Intell. Lab., 2011 (107), 1–23
[4] Riu J., Giussani B., Chemometr. Intell. Lab., 2024 (247), 105095

Machine Learning by Hand: Localized Spectral Analysis using Chunk-Based Voting and Feature Rejection

Abhishek Abhishek[1], Eneko Lopez[1], Harun Hano[1], Andreas Seifert[1,2]
[1] *CIC nanoGUNE BRTA, 20018 San Sebastian, Spain*
[2] *Ikerbasque, Basque Foundation for Science 48009 Bilbao, Spain*
abhishek@nanogune.eu

Chemometric analysis of spectroscopic data plays an important role in biochemical research and will play a pivotal role in medical diagnostics in the future. Current classification methods commonly use either the entire spectral range or specific peak intensities, relying on techniques such as principal component analysis (PCA) and feature selection methods followed by a classifier [1]. This study presents a novel approach using chunk-based cosine similarity voting, which is enhanced by comprehensively analyzing all chunks to increase the reliability of classification. The spectral data is divided into chunks of fixed size, where each chunk acts as an independent sub-model for localized analysis across the spectrum. In each chunk, the spectral features are compared to the class averages using the cosine similarity metric, and the class with highest similarity to the reference spectrum is assigned a vote. Consequently, the class receiving the maximum votes after combining all votes from all chunks is designated as the most representative for the spectrum under test. Additionally, we conduct statistical analysis within each chunk by calculating cosine similarities with class averages and modeling the data through Gaussian distributions with 95% confidence intervals. Those chunks whose class distributions overlap above a threshold are excluded from voting (feature rejection), which ensures that only confident regions contribute to the final prediction. To improve model performance, we optimized the chunk size and threshold for the overlap of the voting chunks. The proposed technique was tested on two different datasets based on Raman spectroscopy measurements and compared with the classification performance based on PCA, feature selection and classifier. The first dataset consists of 36 dried cerebrospinal fluid samples, including 16 diagnosed with Alzheimer's disease and 20 healthy controls. The second dataset comprises 36 dried blood plasma samples, with 18 from healthy and 18 with lung cancer indication [2]. In all cases, the same data preprocessing steps were performed. In the first dataset, the standard method with a leave-one-out cross-validation strategy achieved an accuracy of 67%, while our approach reached a significantly higher accuracy of 81%. In the second dataset, the standard method with and without feature selection achieved an accuracy of 85% and 82%, respectively, while our approach attained an improved accuracy of 92%.

References

[1] Lopez, E., Etxebarria-Elezgarai, J., García-Sebastián, M., et al., Int. J. Mol. Sci., 2024 (25), 4737

[2] Hano, H., Lawrie, C.H., Suarez, B., et al., ACS Omega, 2024 (9), 14084-14091

Methods Validation App: Robust and Efficient Software for Analytical Method Validation

Giovanni Solarino[1,2], Eugenio Alladio[1,2], Marco Pazzi[1,2], Marco Vincenti[1]
[1] *Department of Chemistry, University of Turin, Turin, Italy*
[2] *DataBloom s.r.l., Turin, Italy*
giovanni.solarino@unito.it

Efficient and reliable analytical method validation is crucial for regulated laboratories, but traditional protocols are often time-consuming and error-prone. We developed Methods Validation App (MVA), a Python-based desktop application that integrates advanced statistical analysis with an intuitive, user-friendly interface to address this challenge. MVA streamlines validation by implementing a protocol that maximizes the number of validation parameters obtained from a minimal set of experiments [1, 2], automating key tasks such as calibration, limit of detection, precision, and accuracy. By reducing human error, ensuring consistency, and accelerating evaluations, MVA supports laboratories in meeting major regulatory requirements [3] while boosting efficiency. Its versatility makes it valuable across analytical fields where validated methods are essential, including pharmaceuticals, environmental testing, food safety, and toxicology. By modernizing method validation practices, MVA enhances productivity, improves result reliability, and broadens access to standardized validation tools.

References

[1] Alladio, E., Amante, E., Bozzolino, C., Seganti, F., Salomone, A., Vincenti, M., Desharnais, B., Talanta, 2020 (215), 120867
[2] Alladio, E., Amante, E., Bozzolino, C., Seganti, F., Salomone, A., Vincenti, M., Desharnais, B., MethodsX, 2020 (7), 100919
[3] Thompson, M., Ellison, S.L.R., Wood, R., Pure Appl. Chem., 2002 (74), 835-855

Multi-Way techniques and Ion Mobility Spectrometry for Honey Authentication

Samuele Pellacani[1], Daniele Tanzilli[1], Marina Cocchi[1], Raniero Marabottini[2], Angela Napolitano[2], Francesca Martone[1], Lorenzo Strani[1], Noemi Coppola[1], Caterina Durante[1]
[1] *Dipartimento di Scienze Chimiche e Geologiche, Università degli Studi di Modena e Reggio Emilia, Italy*
[2] *Ispettorato centrale della tutela della qualità e della repressione frodi dei prodotti agroalimentari, Modena, Italy*
samuele.pellacani@unimore.it

The honey industry is particularly vulnerable to fraudulent practices, such as mislabeling of botanical and geographical origins. However, melissopalinological analysis is time consuming and requires expert personnel. In this study, an analytical method based on Gas Chromatography coupled with Ion Mobility Spectrometry (GC-IMS) was used for assessing geographical origin of honey samples. This technique provides a very complex output with analytes separated into two dimensions: chromatographic dimension (retention time) and ion mobility dimension (drift time). In addition to the complexity arising from the presence of many chromatographic peaks, the signals from GC-IMS are subject to shifts in both retention and drift times. An example of output is reported in Figure 1. Two analytical strategies

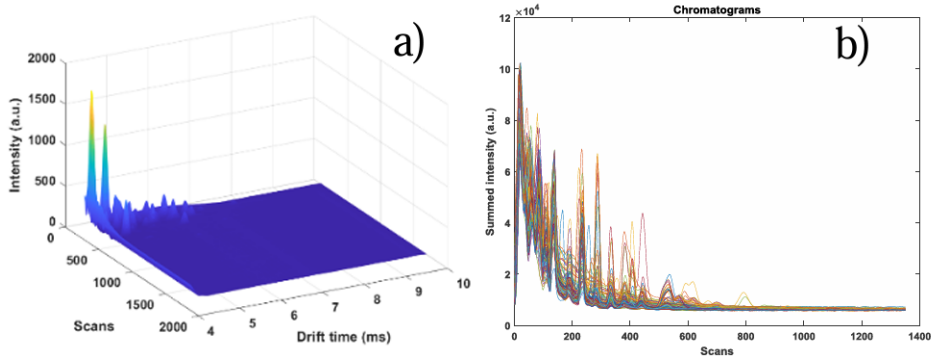


Figure 1. Surface plot of one sample (a) and overlaid sums of all samples on all the drift times (b)

were systematically evaluated. The first was a purely untargeted approach, based on the raw analytical signal (Figure 1b), aimed at capturing a comprehensive chemical fingerprint of the samples. The second involved a deconvolution-based method designed to resolve overlapping signals and enable semi-quantitative analysis. Each approach presented specific challenges: the untargeted fingerprinting required careful preprocessing steps such as scaling, while the deconvolution approach necessitated one-dimensional alignment and the selection of an appropriate number of components to accurately separate co-eluting compounds. For the first time, in this study, PARAFAC [1] and PARAFAC2 [2] were used

for the deconvolution of GC-IMS signals. The outcomes were used for the differentiation of honey samples based on their geographical origin and then compared with those of MCR-ALS [3].

Acknowledgements
This research was supported by the University of Modena and Reggio Emilia - Modena Foundation through FAR Mission Oriented 2024 funds.

References
[1] Bro, R., Chemometr. Intell. Lab., 1997 (38), 149-171
[2] Amigo, J. M., Skov, T., Coello, J., Maspoch, S., Bro, R., Trends in Anal. Chem. 2008 (8), 714-725
[3] Parastar, H., Yazdanpanah, H., Weller, P., Food Chem., 2025 (465), 14

Multivariate Curve Resolution for Environmental Source Apportionment and Pollution Monitoring

Romà Tauler[1]
[1] *IDAEA-CSIC, Barcelona, Spain*
roma.tauler@idaea.csic.es

Over 105 industrial, agricultural, and household chemicals are routinely used worldwide, many of which—along with their biotic/abiotic transformation products—enter environmental compartments (air, water, soils, biota). These substances pose significant toxic risks to ecosystems and human health, necessitating urgent mitigation strategies. Environmental systems are inherently complex, requiring advanced mathematical modeling, multivariate data analysis, and computational techniques to decipher contamination patterns. Modern environmental monitoring, source apportionment, and risk assessment generate *Big Data*, which often becomes a critical bottleneck in pollution research. Chemometrics, particularly Multivariate Curve Resolution (MCR), provides powerful tools for pollution source identification & apportionment, for tracking pollutant dispersion across time, geography, and environmental media, for pollution source characterization (e.g., industrial point source discharges, diffuse pollution from agricultural runoff), and for risk assessment and evaluation of ecological and human health risks of chemical mixtures. This presentation demonstrates practical applications [1-5] of MCR and allied chemometric methods in air and water quality studies, enabling data-driven insights, decoding complex pollution datasets, and supporting evidence-based environmental decision-making to develop actionable mitigation strategies.

References

[1] Terrado, M., Barceló, D., Tauler R. Environmental Science and Technology, 2009 (43), 5321–5326
[2] Gorrochategui, E., Hernandez, I., Tauler, R, Atmospheric Chemistry and Physics, 2022 (22), 9111-9127
[3] Perez-Lopez, C., Ginebreda, A., Jaumot, J., Yamamoto, F.Y., Barcelo, D., Tauler R., Chemometr. Intell. Lab., 2024, 105063
[4] Perez-Lopez, C., Santos, L.H.M.L.M., Ginebreda, A., Alonso, L.L., Buttiglieri, G., Statiris, E., Barcelo, D. Rodriguez-Mozaz, S., Tauler R., Green Analytical Chemistry, 2025 (12), 100203

Nesting the individual in Repeated Measures models with ASCA+

José Camacho[1], Torfinn Madssen[2], Carmen González[3], Caridad Díaz[4], Pedro Sanchez-Rovira[3]
[1] *Research Centre for Information and Communication Technologies, University of Granada, Spain*
[2] *Dept. of Circulation and Medical Imaging, Norwegian University of Science and Technology, Norway*
[3] *Medical Oncology Unit, University Hospital of Jaén, Spain*
[4] *Fundación MEDINA, Spain*
josecamacho@ugr.es

A repeated measures design with a multivariate response can be modelled as:

$$\mathbf{X} = \mathbf{1m^T} + \mathbf{A} + \mathbf{B} + \mathbf{C(A)} + \mathbf{AB} + \mathbf{E} \quad (1)$$

where **X** is the N × M data matrix of outcomes retrieved in the study, **1** is a vector of ones of suitable length, **m** represents the overall mean, **A**, **B** and **C(A)** represent the factor matrices, with **A** the treatment/control factor, **B** the time (repeated measures) factor, and **C(A)** the individual factor, which is nested in **A**, **AB** is the interaction between treatment and time, and **E** is the residual matrix. Both **C(A)** and **E** are random factors, while the rest are often interpreted as fixed.

Repeated measures models have recently gained a lot of attention in the chemometrics community. Martin and Govaerts [1] and Madssen et al. [2] propose alternative extensions of ASCA+ [3] to linear mixed models (LMMs). These solutions allow modeling the individual factor as a random factor, inheriting the good properties of LMMs [4]. A major shortcoming, though, is computation burden, which is not negligible in computational research. It should be noted that the traditional permutation testing approach most often employed in ASCA+ for significance testing can indeed handle random factors [5] in a much more computationally efficient way. So, it comes naturally to question when the use of LMMs means a competitive advantage.

ASCA+ is based on a general linear model (GLM) factorization, that is less parsimonious than LMMs, meaning more degrees of freedom are needed to capture the individual variation. A second advantage of LMMs is that they can provide measures of statistical uncertainty in **C(A)** estimates [4], which have not been exploited in the LMM versions of ASCA. On the other hand, a potential advantage of ASCA+ is a lower dependance on the assumption of random factors following a normal distribution. In this contribution, we introduce a straightforward approach to model nested design in the sum (or deviation) coding of ASCA+ and evaluate its performance using simulated power curves [5-7] and real data.

References

[1] Martin, M., Govaerts, B. J. Chemometr., 2020 (34), e3232

[2] Madssen, T.S., Giskeødegård, G.F., Smilde, A.K., Westerhuis, J.A. PLoS Comput. Bio., 2021 (17), e1009585
[3] Thiel, M., Feraud, B., Govaerts, B., J. Chemometr., 2017 (31), e2895
[4] Montgomery, D., Design and Analysis of Experiments. Wiley, 2020
[5] Anderson, M., Ter Braak, C. J. Stats. Comput. Simul., 2003 (73), 85–113
[6] Camacho, J., Armstrong, M.S. J. Chemometr., 2024 (38), e3596
[7] Camacho, J., Díaz, C., Sánchez-Rovira, P. J. Chemometr., 2022 (37), e3398

Pollution sources in water bodies and their spatiotemporal profiling disentangled using a chemometrics scalar field theory approach

Ginebreda Antoni[1], Romà Tauler[1], Carlos Pérez López[1],
[1] *IDAEA-CSIC, Barcelona, Spain*
agmqam@cid.csic.es

The present contribution proposes a new methodological approach to deal with discrete spatiotemporal environmental monitoring data applicable to waterbodies with known topology [1-3]. It encompasses two modeling steps: First, it uses the so-called MCR-ALS [4] method to factorize the raw three-way dataset into a trilinear model that yields the composition of the main components or sources of pollution and their respective spatial and temporal distributions (Fig. 1). Second, it accommodates the outcome of the previous

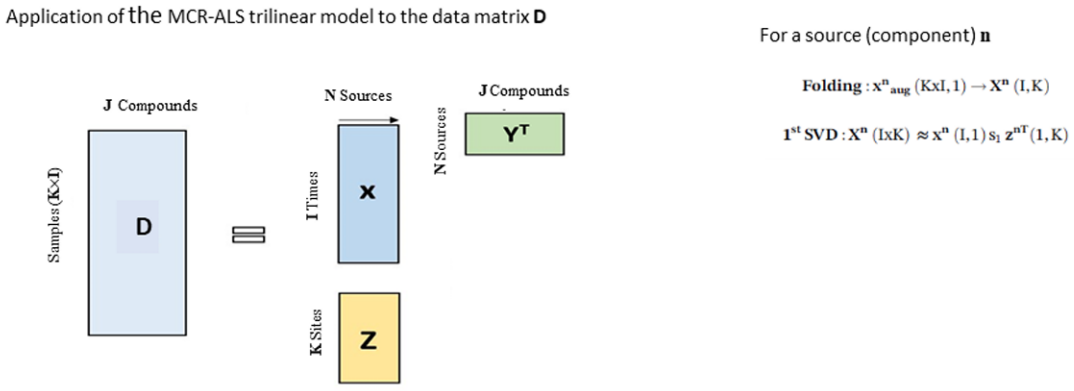


Figure 1. Trilinear decomposition of a dataset matrix D into its spatial, temporal, and composition components

step on a scalar-field framework, leading to a matrix equivalent to the Klein-Gordon scalar field equation, as outlined in Fig. 2. This involves the following steps that are applied to every source (MCR component): (a) matrix representation of the scalar field; (b) matrix representations for the differential square operators $\frac{\partial^2}{\partial r^2}$ and $\frac{\partial^2}{\partial t^2}$; (c) construction of a matrix analog of the Klein-Gordon equation and (d) Fourier spectral decomposition and construction of the dispersion relation (i.e., the matrix analog to that obtained from the Klein-Gordon scalar field equation). To do so, for every source **n**, a field $\mathbf{\Phi^n}(I,K)$ is defined as $\mathbf{\Phi^n} = \mathbf{x^n} s_1 \mathbf{z^{nT}}$. In turn, the corresponding spatial and temporal Laplacian matrices provide a matrix version of the required second-order spatial

Despite the simplifications assumed, the framework approach presented here is straightforward and physically meaningful. It provides useful quantitative information regarding the

Matrix analog of the Klein-Gordon scalar field equation for a MCR component (source)

$$\frac{1}{\bar{v}^2}\left(\frac{\partial^2\phi}{\partial t^2}\right) - \left(\frac{\partial^2\phi}{\partial r^2}\right) + m^2\phi = 0$$

$$\frac{1}{\bar{v}^2}\left(\mathbf{L_t x^n}\right)s_1\mathbf{z^{nT}} - \mathbf{x^n}\left(s_1\,\mathbf{z^{nT}L_s}\right) + m_n^2 \cdot \mathbf{x^n}\, s_1\, \mathbf{z^{nT}} = 0$$

$$\frac{1}{\bar{v}^2}\mathbf{x^{nT}L_t x^n} - \mathbf{z^{nT}L_s z^n} + m_n^2 = 0$$

$\mathbf{L_t}$: time Laplacian
$\mathbf{L_s}$: space Laplacian
$\mathbf{x^n}$: time profile for component (source) $\mathbf{n}$
$\mathbf{z^n}$: spatial profile for component (source) $\mathbf{n}$
$\mathbf{m_n^2}$: constant (to be determined)

- Dividing by scalar s_1
- Left multiplying by $\mathbf{x^{nT}}$
- Rigth multiplying by $\mathbf{z^n}$
- Considering that $\mathbf{x^{nT}x^n} = 1$ and $\mathbf{z^{nT}z^n} = 1$ by orthonormality

dynamic behavior of the variables monitored, such as the quantification of the field energy density into the kinetic, gradient, and internal potential (chemical conversion) terms, the transfer velocity, and the description of the energy contributions as a superposition of spatial and temporal plane waves through the corresponding Fourier Transform. This enables a quantitative alternative interpretation of the data in the frequential domain, such as the relative contribution of the different modes or states (expressed as *entropy*), and specifically, that of the fully *synchronized state* (i.e., the concentration of the measured variable equal for all sites or times), or classification of the different pollution sources. Our approach was tested on a polluted alluvial aquifer located in the Besos River Delta, near Barcelona, which was used as a suitable case study [3, 5]. However, its general applicability to datasets obtained from discrete monitoring of waterbodies characterized by a defined topology is emphasized [2].

References

[1] Ginebreda, A., Jurado, A., Pujades, E., Barceló, D., MethodsX, 2023 (10), 101948
[2] Ginebreda, A., Barceló, D., Water Emerg. Contam. Nanoplastics, 2022 (1), 12
[3] Labad, F., Ginebreda, A., Criollo, R., Vázquez-Suñé, E., Pérez, S., Jurado, A., Environ. Pollut., 2023 (316), 120504
[4] Tauler, R., J. Chemometr., 2021 (35), e3279
[5] Nikolenko, O., Labad, F., Pujades, E., Scheiber, L., Pérez, S., Ginebreda, A., Jurado, A., Environ. Pollut., 2024 (341), 123020

Quantifying Components in a Model Vaccine with Machine-Learning Augmented Raman Spectroscopy

Jana Hahn[1,2], Pooja Gune[1], Sascha Hein[1], Wolf Holtkamp[1], Marcel H. Schulz[2], Christel Kamp[1,2]

[1] *Paul-Ehrlich-Institut, Germany*
[2] *Goethe University Frankfurt, Germany*
jana.hahn@pei.de

The development and quality assurance of vaccines with aluminum-based adjuvants requires a thorough understanding of the interactions between protein antigens and the adjuvant material. Characterizing these antigens poses significant analytical challenges due to their low concentrations and the complexity of vaccine formulations. Raman spectroscopy, a non destructive and label-free analytical technique, shows considerable potential for overcoming these hurdles. In this study, Bovine Serum Albumin (BSA) and aluminum hydroxide ($Al(OH)_3$) were used as components of a model vaccine. Raman spectroscopy combined with advanced machine learning methods is used to deconvolute pure spectral components from the full vaccine spectrum and to quantify the component concentration ratios.

Raman spectra were acquired from mixtures of BSA and $Al(OH)_3$ across varying concentration ratios. Autoencoder-based spectral unmixing methods [1] were applied to extract pure spectral components and estimate their concentration ratios, demonstrating superior performance compared to Multivariate Curve Resolution (MCR) [2]. Additionally, synthetic Raman spectra were generated through Contextual Out-of-Distribution Integration (CODI) [3], incorporating experimental variability into the dataset to enhance model training and robustness. This additional synthetic data enabled autoencoders to accurately identify component spectra with improved signal-to-noise ratios. The findings revealed protein saturation phenomena at higher BSA:$Al(OH)_3$ ratios, consistent with adsorption principles. This study shows the capability of Raman spectroscopy, augmented by autoencoder-based machine learning techniques and synthetic data generation, to analyze protein-adjuvant interactions at concentrations relevant to vaccine development and quality control. By providing accurate identification of spectral fingerprints and superior quantitative estimates, this work advances analytical methodologies in the detailed characterization of biomedicines.

References

[1] Georgiev, D., Fernández-Galiana, Á., Pedersen, S., Papadopoulos, G., Xie, R., Stevens, M., Barahona, M., P. Natl. Acad. Sci. USA, 2024 (121), e2407439121
[2] Azzouz, T., Tauler, R., Talanta., 2008 (74), 1201-1210
[3] Eissa, T., Huber, M., Obermayer-Pietsch, B., Linkohr, B., Peters, A., Fleischmann, F., Žigman,

M., P. Natl. Acad. Sci. USA Nexus, 2024 (3), pgae449

Real-Time Monitoring of Milk Coagulation in Grana Padano Production Using MCR-ALS and NPLS on In-Line NIR Data

Lorenzo Strani[1], Dario Benedini[1,2], Marina Cocchi[1], Giovanni Cabassi[2]
[1] *Department of Chemical and Geological Sciences, University of Modena and Reggio Emilia, Italy*
[2] *Council for Agricultural Research and Economics, Research Centre for Animal Production and Aquaculture (CREA-ZA), Lodi, Italy*
lostrani@unimore.it

Milk coagulation is a critical phase in the production of Grana Padano DOP cheese, as it directly affects curd formation, syneresis, and, ultimately, cheese texture and quality. Traditionally, cheesemakers rely on empirical assessment to determine the optimal time for curd cutting, based on visual and tactile cues. However, this approach lacks reproducibility and makes it difficult to ensure process consistency across different batches. In this study, a combination of in-situ Near-Infrared (NIR) spectroscopy and chemometric methods was employed to investigate the dynamics of the coagulation process and to develop predictive tools for standardizing its monitoring.

Using a portable NIR probe, spectral data were collected every 13 seconds for 30 minutes directly in the coagulation vat under various processing conditions, including different types of rennet (traditional and microbial), pH levels, and enzyme concentrations. A Design of Experiments (DoE) strategy was employed to assess the effects of these factors on curd firmness (a30) and coagulation speed (k20), measured offline with a Formagraph.

The spectral data were analyzed using Multivariate Curve Resolution–Alternating Least Squares (MCR-ALS), which allowed the extraction of concentration and spectral profiles associated with the different phases of the coagulation process. The MCR-ALS algorithm, constrained by non-negativity, successfully resolved three major components interpretable as distinct coagulation stages: micellar destabilization, aggregation, and gel network formation. These profiles offered a time-resolved description of the physical transformations occurring during curd development and were found to vary across rennet types and experimental conditions, especially highlighting the sensitivity of microbial rennet to pH changes.

To translate the spectral information into actionable predictions, N-way Partial Least Squares (NPLS) regression models were developed using the full three-dimensional dataset (sample × time × wavelength). These models enabled robust prediction of a30 and k20 values with good accuracy, as validated through cross-validation and external test sets. Variable Importance in Projection (VIP) analysis further confirmed the relevance of scattering-driven spectral regions, consistent with the physical nature of the coagulation changes.

The integration of exploratory (MCR-ALS) and predictive (NPLS) chemometric tools with in-line NIR spectroscopy presents a powerful strategy to both understand and monitor milk coagulation in real time. This approach enables more objective curd-cutting decisions

and reduces variability, supporting a data-driven, standardized Grana Padano production while preserving its artisanal character.

Space filling design for calibration sample selection

Carl Charloto[1,2], Maxime Metz[1], Michelle Sergent[2], Magalie Claeys-Bruno[2]
[1] *Pellenc Selective Technologies,125 Rue Francois Gernelle, BP124, 84120 Pertuis, France*
[2] *Institut Méditerranéen de Biodiversité et d'Ecologie marine et continentale, Aix-Marseille Université, UMR CNRS IRD Avignon Université, Site de l'Etoile, Marseille, France*
c.charloto@pellencst.com

Chemometrics associated with IR spectroscopy have been widely used in chemical analysis to establish predictive models [1]. However, one of the challenges faced by the user when building and maintaining models is to select calibration samples for reference analyses. To address this issue, we propose to use Space Filling Designs [2] (SFD), which allow an unsupervised exploration of the spectral space as evenly as possible. In this study, SFD was constructed using the Wootton, Sergent, Phan-Tan-Luu's algorithm [3]. This approach offers advantages in case of outliers and with nonlinear relationships, by not favouring extreme samples. This is not the case with classical methods such as Kennard-Stone [4] (KS). An additional consideration highlighted in literature is that the performance of a selection methods is influenced by parameters such as preprocessing and dimensionality reduction [5]. In this context, the WSP selection method was compared to KS and random selection to assess its performance and the impact of these parameters on the selection process. Experiments conducted on simulated and real datasets have shown that WSP selection displayed the same performance as KS selection. Preprocessing affected the selection process without enhancing performance, while dimensionality reduction via principal component analysis (PCA) made some slight improvements for both methods. Even though the improvements were modest, variable reduction offers significant improvements for calculation duration.

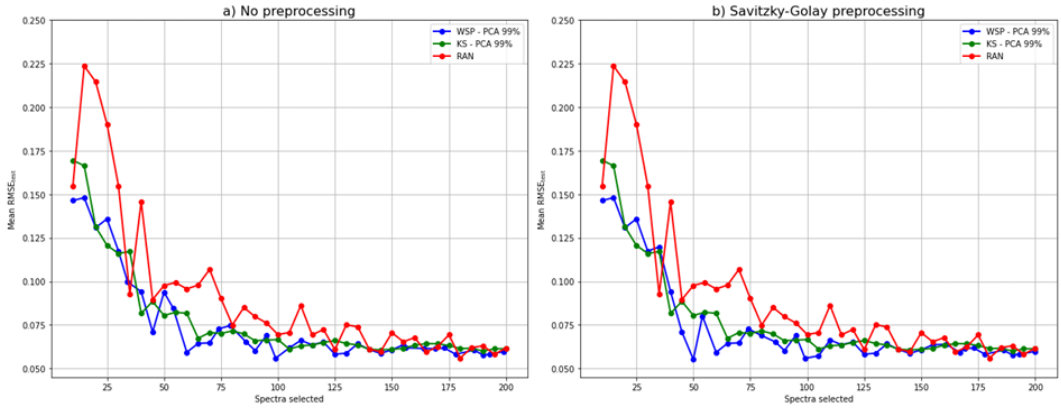


Figure 1. WSP (blue), KS (green) and random (red) performance selections a) without and b) with preprocessing

References

[1] M.S. Larrechi, M.P. Callao, TrAC Trends Anal. Chem., 2003 (22), 634–640

[2] T.J. Santner, B.J. Williams, W.I. Notz, Space-Filling Designs for Computer Experiments, in: T.J. Santner, B.J. Williams, W.I. Notz (Eds.), The Design and Analysis of Computer Experiments, Springer, New York, NY, 2003: pp. 121–161

[3] J. Santiago, M. Claeys-Bruno, M. Sergent, Chemom. Intell. Lab., 2012 (113), 26–31

[4] R.W. Kennard, L.A. Stone, Technometrics, 1969 (11), 137–148

[5] V. Fonseca Diaz, B. De Ketelaere, B. Aernouts, W. Saeys, Chemom. Intell. Lab., 2021 (215), 10435

Sparsity in the spatial direction as an exploratory tool for Hyperspectral imaging

Veronica Ferrari[1,2], Rosalba Calvini[1,2], Giulia Gorla[3], Alessandro Ulrici[1,2], Jose Manuel Amigo[3,4]

[1] *Department of Life Sciences, University of Modena and Reggio Emilia, Italy*
[2] *Interdepartmental Research Centre BIOGEST-SITEIA, University of Modena and Reggio Emilia, Italy*
[3] *Department of Analytical Chemistry, University of the Basque Country, Spain*
[4] *Ikerbasque, Basque Foundation for Sciences, Spain*
veronica.ferrari@unimore.it

Hyperspectral imaging (HSI) combines the benefits of spectroscopy and spatial imaging, enabling the visualisation of the distribution of chemical properties across a sample's surface. However, the high dimensionality of HSI data complicates analysis, storage, and interpretation [1]. Multivariate Image Analysis is essential in this context, but it often involves pixel-level data reduction to retain only those pixels most representative of the sample's variability. This work proposes sparse Principal Component Analysis (sPCA) with sparsity applied in the spatial direction as a promising exploratory tool for HSI. sPCA is an extension of classical PCA where a sparsity constraint is applied to the model coefficients [2, 3]. To extract informative spatial features (i.e., pixels), sparsity was applied to the score vectors using the Least Absolute Shrinkage and Selector Operator (LASSO). Only a reduced subset of pixels contributes to each principal component, while the score values of noisy, uninformative, or redundant pixels are forced to be zero [2]. The effectiveness of sPCA was evaluated on well-known datasets of increasing complexity: a Raman HSI of oil-in-water emulsion, Near Infrared (NIR) HSI of plastics pellets of different polymers and a dataset of NIR-HSI of stinkbugs on mixed vegetal backgrounds [3, 4]. The influence of varying sparsity levels was also evaluated, since the identification of the proper sparsity level represents a key aspect to refine the selection of the Regions of Interest (RoIs). Spatial feature extraction performed with sPCA reduces redundant information, enhancing interpretability and mitigating noise. In addition, it favours data compression by reducing storage and computational efforts. As a further development, sPCA can be used as an alternative approach to select representative pixels from classes of interest to develop supervised classification models.

References

[1] Gowen, A.A., O'Donnell, C., Cullen, P., Downey, G., Frias, J.M., Trends Food Sci. Tech., 2007 (18), 590-598

[2] Rasmussen, M. A., Bro, R., Chemometr. Intell. Lab., 2012 (119), 21–31

[3] Calvini, R., Ulrici, A., Amigo, J. M., Data Handl. Sci. Tech., 2016 (30), 613–634

[4] Ferrari, V., Calvini, R., Boom, B., Menozzi, C., Rangarajan, A. K., Maistrello, L., Offermans,

P., Ulrici, A., Chemometr. Intell. Lab., 2023 (234), 104751

SpectrApp: An Open-Source Dashboard for Chemometric Data Analysis in Education and Research

Alberto Mazzoleni[1,2], Giovanni Solarino[1,2], Marco Vincenti[1], Marco Pazzi[1,2], Rosario Casamassima[3], Francesco Pellegrino[1], Lorenzo Castellino[1], Eugenio Alladio[1,2]
[1] *University of Turin, Department of Chemistry, 10125 Turin, Italy*
[2] *DataBloom s.r.l., 10125 Turin, Italy*
[3] *Carabinieri, Department of Scientific Investigation (RIS), 00191 Rome, Italy*
alberto.mazzoleni@unito.it

Chemometric methods are now integral to modern analytical chemistry, providing robust strategies for extracting relevant information from high-dimensional and complex datasets. Despite their importance, the effective application of these techniques often relies on programming skills that are rarely addressed in traditional chemistry education. To bridge this gap, we developed *SpectrApp*, an open-source application built in R Shiny, aimed at facilitating chemometric analysis for students, educators, and researchers without the need for coding expertise.

SpectrApp offers a user-friendly and interactive environment that guides users through the entire data analysis workflow, from data visualization - ranging from univariate and bivariate plots to multivariate representations - to preprocessing, exploratory analysis, model development, and validation.

It includes tools for unsupervised learning (e.g., PCA, clustering, t-SNE, UMAP), classification (e.g., k-Nearest Neighbors, PLS-DA, LDA, SVM, Random Forest, Logistic Regression, Naive Bayes), class modeling (SIMCA), and regression (e.g., MLR, PCR, PLS-R). The platform leverages well-established R packages such as *mdatools* [1] for multivariate analysis, *prospectr* for spectral data preprocessing, and *caret* [2] for supervised machine learning workflows. By integrating a wide range of chemometric tools within an intuitive interface, SpectrApp lowers the barrier to data analysis, enabling users to focus on interpreting results rather than coding. It is particularly valuable in educational contexts, where it facilitates the teaching of chemometric concepts through hands-on learning, and promotes reproducible, transparent workflows in research.

References

[1] Kucheryavskiy, Sergey. "mdatools–R package for chemometrics". Chemometrics and Intelligent Laboratory Systems 198 (2020): 103937

[2] Kuhn, Max. "Building predictive models in R using the caret package". Journal of Statistical Software 28 (2008): 1-26

The successful combination of hyperspectral imaging, information theory, and chemometrics for the multigrain bread conformity assessment

Esteban A. Roca-Nasser[1], Miriam Medina-García[1], Miguel A. Martínez-Domingo[2], Eva M. Valero[2], Luis Cuadros-Rodríguez[1], Ana M. Jiménez-Carvelo[1]

[1]*Department of Analytical Chemistry, University of Granada, Av. Fuentenueva S/N, 18071 Granada, Spain*

[2]*Department of Optics, University of Granada, Av. Fuentenueva S/N, 18071, Granada, Spain*

erocan@ugr.es

Deceptive product labelling is still a present byproduct of consumers' increasing awareness of the impact of food on health. Therefore, ensuring the authenticity of foods like commercial bread becomes imperative, as fraudulent cereal blends and substitution of cheaper flour in premium bread can mislead consumers and compromise product integrity. Bread is a type of food that can be made from a variety of flours (e.g. wheat, oat, spelt, rye) and there is currently no way to tackle its authentication. To combat this problem, the present study proposes a novel methodology for the evaluation of flour content in bread using hyperspectral imaging (HSI), information theory, and chemometrics. By evaluating spectral data captured from HSI in short-wave infrared (SWIR, 900-1700 nm) and visible-NIR (VNIR, 400-1000 nm) ranges, Shannon's entropy and information index [1] were employed as objective criteria for selecting the most informative spectral range, ensuring optimal classification performance. After spectral range selection, exploratory and screening methods such as PCA and PLS regression were employed, and, after observing a natural grouping of samples, supervised learning methods such as PLS-DA and SVM were introduced to successfully distinguish between pure cereal breads (wheat, oat, spelt, and rye) with an accuracy of up to 0.96. Furthermore, a quantification approach based on QPC (quantification based on pixel counting by classification) [2] was developed to estimate the proportion of oat flour in binary bread blends. By merging conventional regression with HSI´s capability of counting individual pixels, a model achieving an R^2 of 0.94 was generated. All in all, results show that integrating information theory with HSI is a viable option, providing an objective metric to evaluate and objectively select signals before commencing a study, thus improving classification and quantification outcomes. Moreover, the proposed methodology offers a new alternative for bread conformity assessment, addressing the lack of standardized post-production methods.

References

[1] Cuadros-Rodríguez, L., Arroyo-Cerezo, A., Jiménez-Carvelo, A.M., Anal. Chim. Acta, 2025 (1344), 343693

[2] Medina-García, M., Roca-Nasser, E.A., Martínez-Domingo, M.A., Valero, E.M., Arroyo-Cerezo, A., Cuadros-Rodríguez, L., Jiménez-Carvelo, A.M., Food Control, 2024 (166), 110715

Understanding the main factors affecting TD-NMR data on tyre products: MDOE and data compression index.

Riccardo Voccio[1,2], Cristina Malegori[1], Paolo Oliveri[1], Oliver Albohr[3], Thorsten Zeltner[3], Marco Arimondi[4], Giorgio Luciano[5], Mattia Cettolin[4]

[1] *1University of Genova, Department of Pharmacy (DIFAR), Viale Cembrano, 4. Genova, 16148, Italy*
[2] *University of Genova, Department of Chemistry and Industrial Chemistry (DCCI), Via Dodecaneso 31. Genova, 16146, Italy*
[3] *Pirelli Deutschland GmbH, Höchster Str. 48 – 60, Breuberg, 64747, Germany*
[4] *Pirelli Tyre SpA Viale Piero e Alberto Pirelli 25, 20126 Milano, Italy*
[5] *Istituto di Scienze e Tecnologie Chimiche "Giulio Natta", Consiglio Nazionale delle Ricerche, SCITEC-CNR Via De Marini 6, 16149 Genova, Italy*
riccardo.voccio@edu.unige.it

Time Domain – Nuclear Magnetic Resonance (TD-NMR) is a non-destructive analytical technique increasingly used in the analysis of tyre and rubber products. Its greater use is due to the nature of the signal obtained by this technique. Indeed, a decay over time is obtained which is strongly correlated with the polymeric mobility of the materials analyzed. Therefore, any factor that affects polymeric mobility will also affect the signal obtained, such as the temperature of the sample analyzed, the compound formulation and the curing or not of the compound.
It is therefore essential for the daily use of this technique to know the impact of these factors. To the aim, a Design of Experiment (DoE) strategy was applied in which the effect of the following three factors was evaluated:

1. *Temperature of the sample* → Three levels (coded as: -1, 0, +1)
2. *Ratio between the amount of component A and component B* → Two levels (coded as: -1, +1)
3. *Curing or not curing* → Two levels (coded as: -1, +1)

For each level of temperature, all compounds obtained by combining the second and third factor levels were analysed at TD-NMR (^{1}H-20 MHz applying Carr Purcell Meiboom Gill sequence). On the obtained dataset, a strategy of data compression was applied in order to obtain a parameter which holds all the information contained in the data.
In this regards, several parameters were studied. The most promising one is the time at which a specific intensity is reached by the signal. The strategy DoE + data compression index allows to obtain response surfaces that are an indication of how the parameter, and subsequently the signal, vary with the variation of the factors analysed. Through this approach, interesting information can be obtained which may allow a more conscious use of the analytical technique.

Unmixing photochromic fluorescent protein kinetics for fast bioimaging

Adrián Gómez-Sánchez[1], Hana Valenta[2], Raffaele Vitale[1], Wim Vandenberg[2], Cyril Ruckebusch[1], Peter Dedecker[2]

[1] *Univ. Lille, CNRS, LASIRE (UMR 8516), Laboratoire Avancé de Spectroscopie pour les Interactions, la Réactivité et l'Environnement, F-59000 Lille, France*

[2] *Department of Chemistry, KU Leuven, Leuven, Belgium*

adrian.gomez-sanchez@univ-lille.fr

Smart imaging labels like photochromic fluorescent proteins (FPs) are being widely used in fluorescence microscopy. When illuminated with specific light patterns in rapid capture sequences, these FPs switch reversibly between bright and dark states according to their chemical kinetics. The resulting intensity variations across images produce characteristic fluorescence patterns for the fluorophores. Most existing unmixing methods, though, depend on prior knowledge of these fluorescence patterns, which hinders their use in complex specimens [1]. An alternative approach is to measure the full kinetic decay curves at every pixel and to apply chemometric analysis to extract pure decay profiles and spatial maps, but this is often too slow for large areas or volumes, which can engender phototoxicity and jeopardize sample integrity. Here, we propose a novel approach to leverage the advantages of these two measurement procedures while circumventing their respective limitations. In a nutshell, the data collected via both methodologies are fused and simultaneously subjected to Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS) [2-4]. The combination of both datasets allows one to perform accurate blind unmixing without prior knowledge about the fluorophore kinetic properties. Once the kinetic fingerprints are recovered, they can be used to unmix new image stacks acquired solely through the fast pattern protocol, eliminating the need for repeated slow measurements. We tested our method on fixed HeLa cell samples and achieved a good resolution of pure fluorophore kinetic fingerprints as well as the identification of fluorophore spatial locations, which led to a significant improvement over existing state-of-the-art approaches.

References

[1] Valenta, H., Bierbuesse, F., Vitale, R., Ruckebusch, C., Vandenberg, W., Dedecker, P., Talanta, 2024 (269), 125397

[2] de Juan, A., Tauler, R., Anal. Chim. Acta, 2021 (1145), 59-78

[3] Gómez-Sánchez, A., Vitale, R., Devos, O., de Juan, A., Ruckebusch, C., Anal. Chim. Acta, 2023 (1273), 341545

[4] Gómez-Sánchez, A., Marro, M., Marsal, M., Loza-Alvarez, P., de Juan, A., Anal. Chem., 2020 (92), 9591 9602

w-CovSel: an efficient algorithm for selecting intervals of predictors

Miriam Aledda[1], Alessandra Biancolillo[2], Achim Kohler[1], Jean-Michel Roger[3,4], Federico Marini[5]

[1] *Faculty of Science and Technology, Norwegian University of Life Sciences Drøbakveien 31, 1432 Ås, Norway*
[2] *Department of Physical and Chemical Sciences, University of L'Aquila, 67100, Coppito, L'Aquila, Italy*
[3] *ITAP-INRAE, Institut Agro, University Montpellier, 34196 Montpellier, France*
[4] *ChemHouse Research Group, 34196 Montpellier, France*
[5] *Department of Chemistry, University of Rome "La Sapienza", 00185 Rome, Italy*
federico.marini@uniroma1.it

Selecting intervals of predictors in multivariate calibration and classification is motivated by the need to isolate and exploit contiguous spectral regions that carry relevant chemical information while filtering out noise and redundant data, thereby improving both model interpretability and predictive accuracy [1]. By segmenting spectra into interpretable intervals, chemometric models can more effectively address issues like multicollinearity and overfitting by focusing on chemically meaningful predictor groups rather than individual wavelengths, resulting in more robust and efficient calibration and classification performance [2]. Moreover, recent advancement in the development of miniaturized tunable lasers in the mid-infrared region, requires selecting of few contiguous variables to develop tailored portable and handheld spectroscopic sensors with high predictive accuracy.
Based on these considerations a new algorithm, named window-Covariance Selection (w-CovSel), was developed, which extends the principles of Covariance Selection [3] to groups of a fixed number of adjacent predictors. Building on the main concepts behind CovSel, w-CovSel iteratively selects the signal window having the highest (residual) squared covariance with the response(s), until a stopping criterion is met. In the communication the effect of tunable (meta-)parameters, such as the window size or the way the deflation is conducted (rank-1 or full rank), on the selection outcome will be discussed, by means of examples, mainly based on spectroscopic data.

References

[1] Nørgaard, L., Saudland, A., Wagner, J., Nielsen, J., Munck, L., Engelsen, S., Appl. Spectrosc., 2000 (54), 413-419.
[2] Máquina, A., Souza, L., Buiatte, J., Santos, D., Neto, W., Energy Fuels, 2016 (31), 571-577.
[3] Roger, J., Palagos, B., Bertrand, D., Fernández-Ahumada, E., Chemometr. Intell. Lab., 2011 (106), 216-223.

XYOnion: a layer-based method for dataset splitting

Jokin Ezenarro[1], Daniel Schorn-García[2]
[1] *Department of Food Science, Faculty of Science, University of Copenhagen, Denmark*
[2] *South African Grape and Wine Research Institute, Stellenbosch University, South Africa*
jokin@food.ku.dk

Validation is an essential step when proposing new analytical methods. And in chemometric modelling, when an independent test set for validation is unavailable, using dataset splitting algorithms becomes a usual alternative [1]. However, common approaches to do so (random splitting, Kennard-Stone, SPXY, or Onion) often struggle with representativity, over-optimism, or extrapolation issues. To address these limitations, we introduce the XYOnion method, a novel algorithm that integrates two key principles: the consideration of both predictor (X) and response (y) variability, as done in SPXY [2]; and the structured, layer-based allocation of samples, inspired by the Onion algorithm [3].
XYOnion constructs concentric sample layers using a joint distance metric that incorporates variability in both the X and y spaces; distances that can be Euclidean, Mahalanobis, or even after a dimensionality reduction (by principal component analysis or other means). These layers are then alternately assigned to calibration and validation subsets, maintaining the desired ratio and ensuring full coverage of the data domain, while preventing extrapolation in the validation subset. To enhance scalability for bigger datasets, the method is combined with the DISTSLCT algorithm [4], which reduces computational needs by avoiding full pairwise distance matrix calculations. The method was validated using both simulated and real spectroscopy datasets and compared against standard splitting algorithms. XYOnion consistently generated more representative splits, avoided extrapolation, and produced more conservative and reliable figures of merit, suggesting a more realistic assessment of model performance. This makes XYOnion a robust and practical choice for external validation in chemometric workflows, particularly in data-limited scenarios.

References

[1] Lopez, E., Etxebarria-Elezgarai, J., Amigo, J.M., Seifert, A., Anal. Chim. Acta, 2023 (1275)
[2] Saptoro, A., Tadé, M.O., Vuthaluru, H., Chem. Prod. Process Model., 2012 (7)
[3] Gallagher, N.B., O'Sullivan, D., Eigenvector Res. Inc., 2020
[4] Gallagher, N.B., Shaver, J.M., Martin, E.B., Morris, J., Wise, B.M., Windig, W., Chemometr. Intell. Lab., 2004 (73), 105-117.

POSTER PRESENTATIONS

The abstracts are organized in alphabetical order according to their titles.

Analytical characterization of grapes at different stages of ripening for the valorisation of by-products

Núria Vera[1], Jokin Ezenarro[2], Ángel García-Pizarro[1,3], Montserrat Mestres[1], Ricard Boqué[1], Itziar Ruisánchez[1]

[1] *Universitat Rovira i Virgili, Department of Analytical Chemistry and Organic Chemistry C/ Marcel·lí Domingo 1, Tarragona, 43007, Spain*

[2] *Department of Food Science, University of Copenhagen, Rolighedsvej 26, Frederiksberg C 1958, Denmark*

[3] *Fruit Production Program, IRTA Mas Bové, Ctra. Reus-El Morell Km 3.8, Constantí, 43120 Tarragona, Spain*

nuria.verai@urv.cat

Optimizing the valorisation of grape by-products is highly depending on understanding the chemical changes during ripening, as they influence composition, functionality, and extraction efficiency. Identifying the optimal ripening stage improves the presence of key compounds and supports the sustainable reuse of these by-products in various industries. To address this challenge, we combined Mid-Infrared (MIR) spectroscopy, to predict the polyphenol content in grape skins and seeds, with Gas Chromatography-Mass Spectrometry (GC-MS), for the detailed characterization of volatile compounds in the must. Due to the complexity of the data sets generated and the large variability associated with grape ripening, a set of chemometric tools were applied to process and interpret the data, including Principal Component Analysis (PCA) to explore the overall variability of the data, ANOVA Simultaneous Component Analysis (ASCA) to assess the effect of ripening stages on chemical composition, and Partial Least Squares Regression (PLSR) to correlate MIR spectra with polyphenol content. In addition, the ProSpecTool software was used to select the optimal preprocessing for the spectral data and improve the performance of the predictive models [1]. These multivariate techniques were fundamental for the identification of key chemical markers that differentiate ripening stages and for reducing data complexity while maintaining interpretability.

The use of classical and instrumental techniques, together with chemometric modelling enabled the development of predictive models to accurately classify ripening stages based on chemical composition. This approach optimizes the timing of harvest and promotes the sustainable reuse of grape by-products, supporting quality control and circular economy strategies in the wine industry.

References

[1] J. Ezenarro, D. Schorn-García, O. Busto, R. Boqué. ProSpecTool: A MATLAB toolbox for spectral preprocessing selection, Chemometr. Intell. Lab., 2024 (247), 105096

Application of chemometric methods to the traceability and quality control of "Canestrato di Castel del Monte" cheese

Alessandra Biancolillo[1], Mattia Montanaro[1], Angelo D'Archivio[1]
[1] *Department of Physical and Chemical Sciences, University of L'Aquila, Via Vetoio 67100, L'Aquila, Italy*
alessandra.biancolillo@univaq.it

Italy is renowned for its vast and diverse cheese production, with hundreds of varieties crafted across its regions. A significant portion of this production comes from sheep milk, especially in central regions like Abruzzo. Known for its strong pastoral tradition and mountainous landscapes, Abruzzo is home to a wide array of sheep milk cheeses that are deeply tied to local identity and tradition.

Sheep milk cheeses are a cornerstone of Abruzzo's agro-food heritage; nevertheless, globalization of food markets has increased the risk of fraud and loss of regional authenticity, threatening consumer trust and the economic sustainability of local producers. This project aims to develop a cost-effective method for authenticating the geographical origin of sheep milk cheeses using portable Mid-Infrared (MIR) spectroscopy, complemented by advanced data analysis techniques.

Samples of sheep milk cheese will be collected from Abruzzo and other Italian regions and analyzed through chemometric classifiers. The use of traditional chemometric methods will allow the detection of markers tied to regional authenticity, enhancing classification accuracy. This approach not only addresses a critical gap in research on cheese authenticity but also provides a scalable tool for food traceability and fraud prevention.

The authentication of geographical origin using MIR often requires a robust chemometric approach. In this regard, discriminant and class-modelling classification techniques, such as Partial Least Squares Discriminant Analysis (PLS-DA) and Soft Independent Modeling of Class Analogy (SIMCA), are commonly employed to link MIR spectral data with specific chemical markers tied to geographical origin.

Despite the availability of such analytical tools, there is limited research specifically focusing on sheep milk cheeses from Abruzzo, highlighting the need to explore the potential of this technology in supporting and safeguarding regional authenticity.

Assessment and Modeling of the Stability of Complex Dispersed Systems through Physico-chemical Methods

A. Viton[1,2], B. Anthouard[1], Pr. Claeys-Bruno[2], Pr. Piccerelle[2]
[1] *Laboratoire SICAF cosmétiques 280 All. Bellecour, 84200 Carpentras, France*
[2] *Institut Méditerranéen de Biodiversité et d'Ecologie marine et continentale, Aix-Marseille Université, UMR CNRS IRD Avignon Université, Site de l'Etoile, Marseille, France*
axel.viton@etu.univ-amu.fr

The physico-chemical stability of cosmetic emulsions is crucial for their marketability [1, 2]. This study aims to anticipate emulsion stability by combining innovative experimental techniques such as rheology, turbidimetry, granulometry, microscopy, and texture analysis with in-depth data analysis. Stable and unstable formulas were compared, revealing that granulometry, rheology, and turbidimetry effectively distinguish between them [3, 4]. The results were then studied using experimental design [5].

References

[1] ISO/TR 18811:2018 [Internet]. Afnor Editions. 2023. https://www.boutique.afnor.org/fr-fr/norme/iso-tr-188112018/cosmetiques-lignes-directrices-relatives-aux-essais de-stabilite-des-produi/xs132002/129753
[2] Angardi V, Ettehadi A, Yücel Ö. Critical Review of Emulsion Stability and Characterization Techniques in Oil Processing. J. Energy Resour. Technol. 2021; 144
[3] Tadros T. Application of rheology for assessment and prediction of the long-term physical stability of emulsions. Adv. Colloid. Interface Sci. 2004; 108-109:22-58
[4] Santos J, Calero N, Trujillo-Cayado LA, Garcia MC, Muñoz J. Assessing differences between Ostwald ripening and coalescence by rheology, laser diffraction and multiple light scattering. Colloids Surf. B Biointerfaces. 2017; 159:405-411
[5] Navarro-Pérez YM, Cedeño-Linares E, Norman-Montenegro O, Ruz-Sanjuan V, Mondeja-Rivera Y, Hernández-Monzón AM, et al. Prediction of the physical stability and quality of O/W cosmetic emulsions using full factorial design. J. Pharm. Pharmacogn. Res. 2021; 9:98-112

Beyond Curve Fitting: Chemometric Methods to Estimate the Metamorphic Temperature in Metasedimentary Rocks via Raman Spectroscopy

Simone Cerruti[1], Elisa Robotti[1], Alessandro Croce[1,2], Giorgio Gatti[3], Emilio Marengo[1]
[1] *Department of Sciences and Technological Innovation, University of Piemonte Orientale, Viale T. Michel 11, 15121 Alessandria, Italy*
[2] *SSD Research Laboratories, Research and Innovation Department (DAIRI), Azienda Ospedaliero-Universitaria SS. Antonio e Biagio e Cesare Arrigo, Via Venezia 16, 15121 Alessandria, Italy*
[3] *Department for Sustainable Development and Ecological Transition, University of Piemonte Orientale, Piazza Sant' Eusebio 5, 13100 Vercelli, Italy*
simone.cerruti@uniupo.it

The chemical structure of carbonaceous material in metasedimentary rocks progressively transforms from highly disordered to crystalline graphite as a function of metamorphic temperature (typically between 160 and 650°C) [1]. To estimate this property, Raman spectra acquired on carbonaceous particles contained in these samples are usually processed using curve fitting procedures. The resulting signal characteristics (positions, widths, integrated areas, and relative intensities) are then related to temperature via appropriate computational methods [2]. However, curve fitting is often performed manually, introducing subjectivity bias [3]. Previous attempts to overcome this issue include automatic procedures such as the IFORS program [4], which ensures good reproducibility but remains slow and time-consuming. Moreover, the IFORS method estimates metamorphic temperature using polynomial regression based solely on the standard total area of the spectrum [5], thereby not fully exploiting the information available from curve fitting process.
In this study, three chemometric methods are proposed as effective alternatives to improve the estimation of peak metamorphic temperature from Raman spectra of carbonaceous material in 23 metasedimentary rock samples (14 in the training set and 9 in the test set), the same used to develop the IFORS model [4, 5]. For each sample, 40 spectra were recorded and pre-processed. The methods include: a MLR model using disorder-related descriptors obtained through curve fitting or chemometric procedures, if available [6]; MLR applied to the contributions of seven MCR components, previously computed from a larger dataset [6]; PLS regression, optionally combined with sequential feature selection, applied directly to the spectra themselves. For each method, predictive performance was assessed using both Monte Carlo cross-validation and the external test set. All models achieved excellent results, with $R^2 > 0.97$ and RMSE < 15°C in prediction.

References

[1] Beyssac, O., et al. Journal of Metamorphic Geology, 2002 (20), 859–871
[2] Kouketsu, Y., et al. Island Arc, 2014 (23) 33–50
[3] Lupoi, J.S., et al. Front. Energy Res., 2017 (5), 24
[4] Lünsdorf, N.K., Lünsdorf, J.O. Int. J. Coal Geol., 2016 (160–161), 51–62

[5] Lünsdorf, N.K., et al. Geostand. and Geoanal. Res., 2017 (41), 593–612
[6] Cerruti, S., et al. Estimation of structural disorder degree from carbon materials' Raman spectra by Multivariate Curve Resolution and Partial Least Squares Regression. XXVIII National Congress SCI2024, 26-30 august 2024

Contribution of SO-PLS to the exploratory multiblock approach

Marion Brandolini-Bunlon[1], Benoit Jaillais[2]
[1] *Universit Clermont Auvergne, INRAE, UNH, PFEM, MetaboHUB Clermont, Clermont-Ferrand, France*
[2] *INRAE, TRANSFORM, Nantes, France*
benoit.jaillais@inrae.fr

Many studies consist of characterizing the same samples using different procedures - chemical, physical, structural or combining several spectroscopic techniques - to take the best possible account of the complexity of the samples and the complementarity of the analytical techniques. These multivariate data are generally in different tables matched by individuals with different number and type of variables. To process these data, it is generally necessary to analyze the blocks separately and then together, to identify whether certain blocks are specific to a source of variability.

In our work, 3 blocks of spectroscopic data (visible, near infrared - NIR - and mid-infrared - MIR) acquired on fat samples are considered. We present the separate PCA of the 3 blocks and the contribution of the consensus PCA. To identify the proportion of variability inherent in each block, linked to the fat content, we used the SO-PLS, which enables us to estimate the variability explained by a block without including the variability already explained by the block(s) already present in the model.

The PCA applied to each block shows a slight effect linked to fat content in blocks NIR and MIR. The consensus-PCA provides an explanation of the variability common to the blocks. For the SO-PLS, the evolution of the prediction error as a function of the combination of latent variables from each block is presented in the Mageplot (Fig. 1a). Fig. 1b shows the individuals, colored according to the percentage of fat, in the first factorial plane. The SO-PLS therefore highlights that the NIR block is the one that best predicts the amount of fat in the samples.

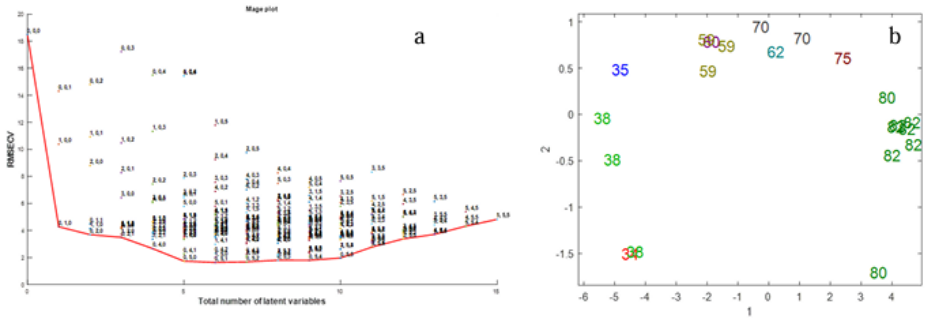


Figure 1. Mageplot SOPLS (a) and factorial design (b) for the first two dimensions of the MIR&NIR&VIS model

D-optimal design to optimize the extraction/desorption step of a HS-SPME chromatographic procedure in the determination of two adulterants in origanum vulgare

V. Vállez-Gomis[1], B. Company[1], M.C. Ortiz[2], M.S. Sánchez[3], L.A. Sarabia[3]
[1]*AINIA, Laboratorio de Cromatografía, Av. Benjamín Franklin 5-11, Paterna, Valencia, 46980, Spain*
[2]*Departamento de Química* [3]*Departamento de Matemáticas y Computación, Facultad de Ciencias, Universidad de Burgos, Plaza Misael Bañuelos s/n, Burgos 09001, Spain*
mcortiz@ubu.es

In the optimization of the extraction/desorption step of a HS-SPME (Head Space-Solid Phase Microextraction) chromatographic procedure, four experimental factors need to be found to achieve the maximum area for five different ratios m/z (responses), related to two different adulterants (olive tree and marjoram) when trying to determine oregano (origanum vulgare). The factors are the temperature of incubation (U1) and three times, namely the incubation (U2), desorption (U3), and extraction/exposition of the fiber to HS (U4) time. The last one is expected to have a possible nonlinear effect on the responses being studied, so it is defined at three level. The remaining three factors are at two levels. Also, only interactions between the temperature U1 and U2 and U1 and U4 are expected. temperature U1 and U2 and U1 and U4 are expected. A full factorial design will have 2^3×3^1 = 24 experiments per adulterant, a total of 48 experimental chromatographic runs. Instead, a D-optimal design is found [1] with 12 experimental runs to fit the model in Equation (1) with nine coefficients.

$$y = b_0 + b_{1A}x_{1A} + b_{2A}x_{2A} + b_{3A}x_{3A} + b_{4A}x_{4A} + b_{4B}x_{4B}$$
$$+ b_{1A2A}x_{1A}x_{2A} + b_{1A4A}x_{1A}x_{4A} + b_{1A4B}x_{1A}x_{4B}$$

The analysis of the experimental results and the visualization via a Parallel Coordinates Plot make it possible to find experimental conditions for the four factors that achieve a compromise among the five responses, reducing the experimental effort to a single design with 12 experiments, saving time and cost. The repeatibility (RSD%) obtained for n=5 replicates in the found conditions varies from 3 and 7% for the five responses.

Acknowledgements
This project (IMDEEA/2024/60) is supported by the Conselleria d'Innovació, Indústria, Comerç i Turisme de la Generalitat Valenciana, through IVACE, and is financed by the European Union, through the FEDER Comunitat Valenciana 2021-2027 Program.

References
[1] Mathieu D, Nony J, Phan-Than-Luu R, Beal A, NemrodW®, 2017, Marseille, France

Data Fusion for the Food Industry: Leveraging Machine Learning and Deep Learning Algorithms

Giorgio Felizzato[1], Eloisa Bagnulo[1], Alessandro Guglielmetti[2], Cristian Bortolini[2], Erica Liberto[1]

[1] *Department of Science and Drug Technology, University of Turin, Italy*
[2] *Soremartec Italia, Alba, Italy*
giorgio.felizzato@unito.it

Food quality perception is significantly influenced by off-flavours and taints, which encompass various sensory attributes, including grassy, bitter, and astringent notes. Traditional quality control relies on human inspectors, whose evaluations are often subjective and inconsistent. Molecular analysis techniques provide a fast and objective method for quality assessment. However, while modern analytical tools can quantify food quality markers with high precision, they generate large, complex datasets that are difficult to interpret. Machine Learning (ML) and Deep Learning (DL) algorithms have been proven effective in extracting meaningful insights from these datasets, enabling fast and objective quality assessments. The accuracy of exploratory and predictive models can be further improved by integrating analytical signals from different instruments, creating a synergistic effect. This approach, known as Data Fusion (DF), enhances analytical performance and reliability [1-3].

This study analysed off-flavour cocoa liquor samples identified by an internal company sensory panel and compared them with 'industrially accepted' samples from two Ecuadorian cultivars: CCN51 and Arriba. The samples were examined using Liquid Chromatography and Gas Chromatography coupled with Mass Spectrometry. Initially, data from both instruments were analysed separately, and the results were compared. Different ML and DL models were applied, including K-Nearest Neighbors, Linear Discriminant Analysis, Logistic Regression, Partial Least Squares Discriminant Analysis, Random Forest, Support Vector Machines, and Multi Layer Perceptron. Finally, DF strategies, Low-Level (LLDF), Mid-Level (MLDF), and High Level (HLDF), were explored. Among these, HLDF based on Bayesian Consensus proved most effective, achieving 93% accuracy in identifying off-flavor cocoa samples.

References

[1] Jakhar, Deepack, and Ishmeet Kaur. "Artificial intelligence, machine learning and deep learning: definitions and differences". Clinical and Experimental Dermatology 45.1 (2020): 131-132

[2] Caratti, Andrea, et al. "Boosting comprehensive two-dimensional chromatography with artificial intelligence: Application to food-omics". TrAC Trends in Analytical Chemistry (2024): 117669

[3] Ballabio, D., R. Todeschini, and V. Consonni. "Recent advances in high-level fusion methods to classify multiple analytical chemical data". Data Handling in Science and Technology 31 (2019): 129-155

Deep-Chemometrics: integrating pre-processing with aggregation strategies

Daniele Tanzilli[1,2], Florent Abdelghafour[1,2], Jean-Michel Roger[1,2], Maxime Metz[1,3,4]
[1] *1 LabCom Aioly, Artificial Intelligence and Optics Laboratory, 34196, Montpellier, France*
[2] *ITAP, Univ.Montpellier, INRAE, Institut Agro, 34196, Montpellier, France*
[3] *Institut Méditerranéen de Biodiversité et d'Ecologie Marine et Continentale Aix-Marseille Université, UMR CNRS IRD Avignon Université, Site de l'Etoile Marseille, France*
[4] *Pellenc ST, Applied Research Group, 84120, Pertuis, France*
daniele.tanzilli@inrae.fr

Raw spectral measurements frequently suffer from contamination by noise, instrumental interference, and extraneous fluctuations. When these unwanted variations remain unaddressed,they can substantially alter data and negatively influence model efficiency. Therefore, the preprocessing of spectral measurements represents an essential procedure for improving data integrity and isolating valuable chemical insights from intricate spectral patterns [1]. The significance of data preprocessing has been firmly established within chemometrics, where extensive research spanning multiple decades has generated thorough knowledge regarding diverse preprocessing methodologies and their ideal implementation scenarios. Nevertheless, this knowledge has yet to be comprehensively incorporated into deep learning architectures designed for spectral data.

Recent studies have proposed using multiple pre-processing techniques by concatenating their outputs and feeding them as inputs to neural networks [2]. While the results are promising, a key limitation of this strategy is that each pre-processing transformation generates a new version of the input data, and concatenating several of them leads to a substantial increase in model parameters. In the context of spectroscopy, where large annotated datasets are often limited, this increase in model complexity can hinder the practical use of diverse pre-processing operations. To address this issue, we propose an alternative data aggregation strategy that limits the impact of the number of pre-processing transformations on the overall parameter count of the model.

The performance of this approach is demonstrated using real spectral data relating to dry matter (DM) prediction in mangoes. The model's hyperparameters were optimised using Design of Experiment, and the results are then compared with those of existing deep learning strategies.

References

[1] Rinnan, Å., Anal. Methods, 2014 (6), 7124-7129
[2] Mishra, P., Passos, D., Chemometr. Intell. Lab., 2021 (212), 1042

Degradation analysis of aromatic disulfide-based vitrimers: defining optimal reprocessing conditions using ^{1}H NMR and chemometrics

Gorka Albizu[1], Paula Fanlo[2,3], Miren Ostra[1]
[1] *Department of Applied Chemistry, Faculty of Chemistry, UPV/EHU, 20018 San Sebastian, Spain*
[2] *CIDETEC, Basque Research and Technology Alliance (BRTA), 20014 San Sebastian, Spain.*
[3] *POLYMAT, Department of Polymers and Advanced Materials: Physics, Chemistry and Technology, Faculty of Chemistry UPV/EHU, 20018 San Sebastian, Spain.*
gorka.albizu@ehu.eus

Aromatic disulfide-based vitrimers have emerged as promising dynamic polymeric materials, combining the excellent performance of thermoset materials with the processability of thermoplastic materials. However, ensuring efficient reprocessing minimizing degradation remains a key challenge [1]. In this work, we present an approach coupling ^{1}H Nuclear Magnetic Resonance (NMR) spectroscopy with chemometric tools to evaluate the thermal stability and degradation behavior of aromatic disulfide networks using model compounds instead of the vitrimeric networks, where the procedure is more laborious.

A model system based on glycidyl phenyl ether (GPE), and 4-aminophenyl disulfide (4-AFD) was synthesized, varying the amine-to-epoxy stoichiometry. These model compounds were exposed to different thermal treatments, modifying both temperature and processing time. Their ^{1}H NMR spectra were then analyzed using Principal Component Analysis (PCA) to identify the factors influencing degradation. The results revealed that processing time had the greatest impact, followed by temperature, while stoichiometry had a negligible effect. In addition, a Multivariate Curve Resolution–Alternating Least Squares (MCR-ALS) model was applied to quantify degradation levels and generate a predictive degradation map.

These results were validated in epoxy vitrimer networks, where Dynamic Mechanical Analysis (DMA) measurements were performed before and after reprocessing. The degradation trends observed in model compounds were confirmed in the vitrimeric materials, demonstrating that ^{1}H NMR combined with chemometrics is a valuable strategy for rapidly identifying optimal reprocessing conditions for guiding the repair processes of vitrimers, ultimately advancing their sustainable design and industrial application [2].

References

[1] Schenk, V., D'Elia, R., Olivier, P., Labastie, K., Destarac, M., Guerre, M., ACS Appl. Mater. Interfaces, 2023, 15 (39), 46357–46367

[2] Fanlo, P., Ruiz de Luzuriaga, A., Albizu, G., Ximenis, M., Rekondo, A., Grande, H.J., Sardon, H., RSC Appl. Polym., 2024, 2 (5), 826–837

Effect of pedoclimatic conditions on the chemical composition of essential oils of *Thymus vulgaris* L.

Elodie Mezzatesta[1], Anne-Marie Farnet[1], Isabelle Bombarda[1], Nathalie Dupuy[1]
[1] *MBE, Aix Marseille Univ, CNRS, IRD, Marseille, France*
elodie.mezzatesta@imbe.fr

Common thyme (*Thymus vulgaris* L.) is a perennial herbaceous plant native to the western Mediterranean region, belonging to the *Lamiaceae* family. Its great adaptability to different environmental conditions allows this species to be cultivated throughout Europe, also in temperate regions. Essential oils from the leaves of thyme are sought after for their varied properties (antiseptic, antioxidant, insecticide and food preservative). In France, *T. vulgaris* presents different chemotypes [1]. Several factors influence the chemotypes, such as the soil type, temperature, humidity and the phenologic stage of the plant. For example, the so-called phenolic thymes, mainly containing thymol and carvacrol, grows on rocky soils such as regosols, while the carvacrol chemotype grows on drier soils [2].
In this study, thyme was chosen as a marker of bioclimatic change by studying the impact of the chemical composition of thyme under two sub-climates of the Mediterranean region (Subhumid and Humid) and also as an indicator of past land use since, forests with different ages coexist after agriculture abandonment during the last century [3]. For the collection of samples, the chosen study site was the Parc Naturel Régional of Luberon. This park was chosen for its extensive documentation on the climate and its forest history. The study focused on statistical processing of Thymus data by global analyses by IRTF (Fourier Transform Infrared) and targeted analyses on the essential oil part by GC (Gas Chromatography). The spectral data can be explored as a first approach using principal component analysis (PCA), it is also necessary to pre-processing of the raw spectra to reduce the effect of interference and artefacts on the subsequent development of a predictive model. First or second derivatives can solve overlap problems. Other algorithms, such as Standard Normal Variate (SNV) and Multiplicative Scatter Correction (MSC), are useful when additive and multiplicative effects caused by light scattering are present. For the GC analyses, multivariate analyses will be carried out to obtain a graphical representation of the data using supervised methods (PCA; PLS-DA) to obtain information on the classes and/or groups to guide the construction of the components in order to maximise discrimination between the classes and/or groups.

References

[1] Satyal, P., Murray, B., McFeeters, R., Setzer, W., Foods, 2016 (5), 70
[2] Gouyon, P.H., Vernet, P., Guillerm, J.L., Valdeyron, G., Heredity, 1986 (57), 59–66
[3] Delcourt, N., Farnet-Da Silva, AM, Rebufa, C., Foli, L., Dupuy, N., Geoderma, 2023, (430), 116299

Effect of the calculation method of the critical values for the Q and T^2 statistics on calibration and class modelling using PLS and NIR spectroscopy

L.A. Sarabia[1], O. Valencia[1], D. Castro-Reigia[1], M.C. Ortiz[1]
[1] *Departamento de Matemáticas y Computación, Facultad de Ciencias, Universidad de Burgos, Plaza Misael Bañuelos s/n, Burgos 09001, Spain*
[2] *Departamento de Química, Facultad de Ciencias, Universidad de Burgos, Plaza Misael Bañuelos s/n, Burgos 09001, Spain*
lsarabia@ubu.es

When applying PLS, statistics T^2 and Q define the boundaries where the model can be applied (the so-called PLSbox). For each sample, T^2 is defined as the squared Mahalanobis distance from the model centre to the sample within the scores subspace, while Q is the squared Euclidian distance from the sample to the scores subspace. The calculation of the critical values of these two statistics has been extensively investigated in principal component analysis [1], and it has received renewed attention [2] in relation to multivariate modelling techniques (SIMCA), where two procedures to estimate the degrees of freedom of the chi-square distribution were proposed. The first one uses the mean and the standard deviation of the Q and T^2 values, and the second one the robust version of both parameters. In this work, the performance of these three procedures is comparatively studied with a new fully data-driven procedure based on fitting a Gaussian Kernel to the values of Q and T^2 obtained with PLS and NIR spectroscopy data.
When using PLS for calibration (PLSR), the impact of the four computational procedures is evaluated by means of some figures of merit such as the accuracy line (concentration values predicted by PLS versus actual values). However, when modelling two classes A and B with PLS (PLS-CM), the procedures are evaluated by considering the assignment of the samples to one, to both, or to none of the classes, analysing the sensitivities and specificities as well as the DMCEN value achieved. DMCEN [3] is an entropy-based criterion for class-modelling that is more sensitive than the usual accuracy or the total efficiency.

References

[1] Jackson, J. (1991). A User's Guide to Principal Components. John Wiley
[2] Pomerantsev, A.L., Rodionova, O.Y., J. Chemom. 2014 (28), 429-438
[3] Valencia, O., Ortiz, M.C., Sánchez, M.S., Sarabia, L.A., Chemometr. Intell. Lab., 2021 (217), 104423

Hyperspectral imaging and partial least squares for early-stage product development of hybrid sausages

Victor Gustavo Kelis Cardoso[1], Giulia Gorla[2], Daniel Breiner[3], Ivan Nielsen[3], Daniella Jensen[3], Frederik Nielsen[4], Angelo Manetta[4], Dung-Han Yeh[4], Marchen Hviid[3], Patrick Montague[4], Jose Manuel Amigo[2,5], Rasmus Bro[1]

[1] *Department of Food Science, University of Copenhagen, Denmark*
[2] *Faculty of Science and Technology, University of Basque Country, Spain*
[3] *Danish Meat Research Institute, Denmark*
[4] *NKT Photonics A/S, Denmark*
[5] *Ikerbasque, Basque Foundation for Sciences, Spain*
victor.cardoso@food.ku.dk

Hybrid sausages are products that combine plant-based ingredients and meat, aiming to create more sustainable foods [1]. This is also an opportunity for upcycling food by-products, which through fermentation can enhance nutrient availability as well as functional and gastronomic properties [2]. Fermenting brewing spent grains (FBSG), a major by-product of beer production,represents an attractive option to support the circular economy while increasing dietary fiber in hybrid products [2]. In early-stage product development, however, it is often challenging to understand nutrient and ingredient distribution, which directly affects firmness, chewiness, and mouthfeel. In this context, hyperspectral imaging (HSI) in the near infrared (NIR) region was applied to assist product characterization and design [3]. Six sausage formulations with varying levels of meat and FBSG were analyzed using a Specim FX17 line-scanner (900-1700 nm). Partial least squares (PLS) regression was used to predict ingredient proportions and macronutrient contents from median spectra of sausage slices, achieving errors as low as 8%. The PLS models were further applied pixel-wise, which generated spatial distribution maps of ingredients and macronutrients. These maps revealed regions of formulation heterogeneity, correlations between ingredients with proteins and carbohydrates, and uneven distribution of fat, water, and ash. Overall, the results demonstrate the potential of NIR-HSI as a rapid, non-destructive tool to guide product design of new hybrid meat products.

References

[1] Neville, M., Tarrega, L., Hewson, L., Foster, T., Food Sci. Nutr., 2017 (5), 852-864
[2] Weberskov, C., Foley, J., Yang, L., et al., Food Bioproc. Tech., 2025 (10), 1-20
[3] Squeo, G., Angelis, D., Summo, C., et al., J. Food Compos. Anal., 2022 (108), 104459

Impact of data subset selection on the robustness of MVDA models for Critical Quality Attributes prediction of mAb using in-line Raman spectroscopy

Morandise Rubini[1], Martin Soucé[1], Igor Chourpa[1]
[1] *Centre de Biophysique Moléculaire (UPR CNRS 4301), Département Nanomédicaments et Nanosondes, UFR de Pharmacie Philippe Maupas, Université de Tours, 31 Avenue Monge, 37000 Tours, France*
morandise.rubini@univ-tours.fr

The identification and control of Critical Quality Attributes (CQAs) are important during monoclonal antibody (mAb) production due to their impact on product stability, pharmacokinetics, and efficacy. Process Analytical Technology (PAT), particularly Raman spectroscopy combined with multivariate data analysis (MVDA), enables real-time, in-line monitoring and prediction of cellular metabolism and CQAs. However, developing robust predictive models remains challenging due to the complexity of cell culture matrices and the limited availability of representative reference datasets, often leading to subjective and potentially biased data subset selection.
This study investigates the impact of two data subset selection strategies (batch-wise vs. DUPLEX) on the performance of PLS2 models.
Data were generated from five fed-batch CHO cell cultures expressing an IgG1 mAb in 10 L bioreactors, under standard controlled conditions. In-line Raman spectra were collected using a Viserion spectrometer (Indatech, Clapiers), acquiring one spectrum (250 3200 cm^{-1}) of 12–15 min., every 20–30 min. Reference CQAs, including fragmentation, aggregation, isoforms, and glycoforms, were quantified using conventional analytical methods (e.g., LC-MS). The fingerprint region of the spectra (800–1800 cm^{-1}) was selected and preprocessed using Savitzky–Golay smoothing, first derivative, Standard Normal Variate (SNV), and mean centering.
As a result, the batch-wise method was shown to produce over- or under-optimistic model performance due to subjective and clustered data partitioning. In contrast, the DUPLEX method enabled the creation of objective and representative calibration and validation subsets, leading to improved model generalizability. Moreover, PLS2 models based on DUPLEX consistently exhibited high predictive performance, with R^2 values exceeding 0.8 and low, comparable RMSECV and RMSEP values.
This highlights the importance of objective data selection for robust MVDA model development in bioprocess monitoring.

References

[1] CMC Biotech Working Group, "A–Mab: A Case Study in Bioprocess Development"
[2] Rubini et al., Pharmaceutics, 2025 (17), 473

Implementation of a PAT framework for in-line monitoring in a miniaturized pneumatic powder blender

Fabien Millan[1], Florian Seeling[1], Greta Fragnière[2], Matthias Rüdt[1]
[1] *Institute of Life Sciences, School of Engineering, HES-SO University of Applied Sciences and Arts Western Switzerland, Rue de l'Industrie 19, 1950 Sion, Switzerland*
[2] *Dietrich Engineering Consultants sa, Chemin du Dévent 3, 1024 Ecublens, Switzerland*
fabien.millan@hevs.ch

The pharmaceutical industry is undergoing a paradigm shift driven by personalized medicine, continuous manufacturing, and the demand for smaller batch sizes. These trends challenge traditional powder blending technologies and motivate the development of compact, flexible, and automated equipment. In this context, we present the first implementation of a Process Analytical Technology (PAT) framework tailored for a novel miniaturized pneumatic batch blender, developed in collaboration with industrial and academic partners.
A PAT framework was developed to support in-line monitoring of powder mixing performance using Near-Infrared Spectroscopy (NIR). The applied methodology comprises high-frequency in-line spectral acquisition and multivariate data analysis. A grid of ternary powder mixtures consisting of ascorbic acid and salicylic acid (as API mimics) with starch (as excipient) was used to evaluate homogeneity detection capabilities. Spectral pre processing involved Standard Normal Variate (SNV) correction and second derivative transformation, followed by chemometric modeling using Partial Least Squares Regression (PLSR) and Principal Component Analysis (PCA). The results demonstrate that in-line NIR combined with off-line calibration enables quantification of active pharmaceutical ingredient (API) concentrations during blending. Despite limitations at low API concentrations, the approach provides valuable insight into powder mixing dynamics. PCA revealed consistent progression of blend homogeneity across cycles and PLSR achieved good predictive power when models were trained on representative concentration gradients.
This work presents the implementation of a PAT framework in a miniaturized pneumatic batch blender, for which no prior PAT integration had been reported. The in-line spectroscopic monitoring combined with chemometric analysis provided valuable insights into the evolution of blend homogeneity over repeated mixing cycles. These initial results demonstrate how PAT can be used to characterize the dynamics of powder mixing in this pneumatic blending process and pave the way for future integration with dynamic process models to enable real-time control and, ultimately, real-time release testing.

References

[1] Bellon, C., Truffer, C., Steiner, A., Moreillon, A., Nicolay, L., Adv. Powder Technol., 2013 (24), 43–50

Leveraging "Golden Batch" Analysis for Enhanced Monitoring of mAb Production in Sanofi R&D CMC

Paul David[1], Louis Joos[2], Hajer Souaifi-Amara[2]
[1] *Ividata Life Sciences for Sanofi / CMC Data Sciences, 3 rue de Liège, 75009 Paris, France*
[2] *Sanofi RD / Global CMC Development / CMC Data Sciences, 13 Quai Jules Guesde, 94400 Vitry sur Seine, France*
davidolivier.paul-ext@sanofi.com

Biotherapeutics production using cell lines involves complex processes that traditionally require restrictive daily sampling and monitoring of key variables. This approach can be costly, generate significant waste, and demand constant operator presence. "Golden Batch" modeling emerges as an innovative solution, allowing continuous in situ control of bioreactors thanks to Orthogonal Partial Least Squares (OPLS) models which define an ideal trajectory as a reference for future batches.

While Statistical Process Control (SPC) proves insufficient for multi-variable batch processes, Multivariate Statistical Process Control (MSPC) offers a simplified yet comprehensive view of complex processes, effectively evaluating interactions between diverse variables.

The Golden Batch model aims to:

1. Model process time as a response variable
2. Assess batch comparability, identifying atypical batches or out-of-trend behaviors
3. Rapidly evaluate conformity of new batches against the model
4. Implement multivariate Continued Process Verification (CPV) utilizing all process variables

This approach leverages historical data to construct a statistical "Golden Batch" model under nominal conditions, serving as a reference for ongoing production batches. The model assumes consistent process conditions across batches. By implementing Golden

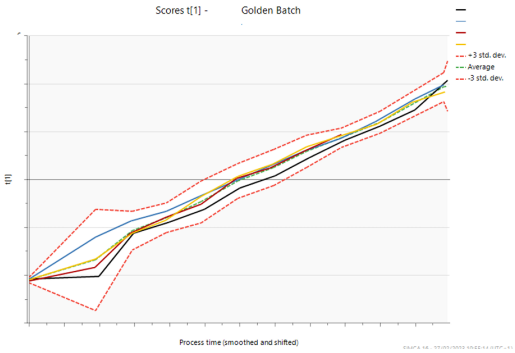


Figure 1. Example of an OPLS (Orthogonal Partial Least Squares) Golden Batch score plot

Batch analysis, we enhance our ability to monitor mAbs production efficiently, reducing costs and improving process understanding in Sanofi R&D CMC.

Made by PCA, inspired by The Renormalization Group: Complementing PCA beyond variance

Sarin Nhek[1], Ana Millán[2], Nicoló Defenu[1]
[1] *CNR-INO, Area Science Park, Basovizza, I-34149 Trieste, Italy*
[2] *Department of Electromagnetism and Matter Physics and Institute "Carlos I" of Theoretical and Computational Physics, University of Granada, Spain*
sarin.nhek@ino.cnr.it

PCA provides a valuable visual perspective for exploring structure in high-dimensional data [1]. By projecting data onto orthogonal directions of maximal variance, it highlights dominant modes of variability and offers insight into large-scale patterns. However, in multivariate time series or structured systems such as large graphs—particularly those involving latent dynamics, coupling, or multi-scale organization—data can appear nearly Gaussian across many low-variance principal components, accompanied by a deceptively smooth eigenvalue spectrum. In such cases, relying solely on explained variance as a selection criterion may obscure informative but subtle structure.

This work is a simple, renormalization group–inspired strategy to enhance principal component investigation. By systematically reconstructing the data using progressively larger subspaces and monitoring higher-order indicators—specifically, average columnwise kurtosis—this approach reveals deviations from Gaussianity that variance-based methods fail to detect. The method introduces a single structural criterion, grounded in statistical physics [2], that complements standard PCA output and draws attention to components that may carry relevant information despite low variance. This perspective aligns with recent efforts to link dimensionality reduction with coarse-graining strategies [3], and is particularly suited to high-dimensional, structured time series where second-order statistics may be insufficient to uncover latent features.

References

[1] Martens, H., Næs, T., Multivariate Calibration, Wiley, 1989
[2] Goldenfeld, N., Lectures on Phase Transitions and the Renormalization Group, Addison-Wesley, 1992
[3] Bradde, S., Bialek, W., PCA Meets RG, J. Stat. Phys., 2017 (167), 462–475

Mapping Emerging Organic Contaminants in wastewater, superficial and sub-superficial water by untargeted uHPLC-HRMS

Samuele Pellacani[1], Lorenzo Strani[1], Mattia Pietropaolo[1], Guido Perra[1], Daniele Tanzilli[1], Marina Cocchi[1], Barbara Benedetti[2], Francesco Ronchetti[1], Caterina Durante[1]
[1] *University of Modena and Reggio Emilia, Department of Chemical and Geological Sciences, Via Campi 103 41125 Modena, Italy*
[2] *University of Genova, Department of Chemistry and Industrial Chemistry, via Dodecaneso 31, 16146 Genova, Italy*
samuele.pellacani@unimore.it
marina.cocchi@unimore.it

Emerging organic contaminants (EOCs) can be of both natural and synthetic origin, and include a wide range of industrial, agricultural, pharmaceutical and consumer products, such as plasticisers, pesticides, fertilisers, antibiotics, hormones, parabens and PFAS (perfluoroalkyl substances) and their metabolites [1]. Their poor biodegradability and tendency to bioaccumulate in environmental matrices makes them potentially hazardous for causing health damage in the long term, see [2] for a first list of concern. This study aims at obtaining a first mapping of EOCs in different kind of water samples, such as surface water (three sampling sites) and groundwater (six sampling sites), since these are vital resources for water supply for drinking, agricultural and industrial purposes, as well as wastewater (six sites sampled at entrance and exit of a wastewater treatment plants, WWTP). In fact, numerous studies have shown a correlation between the presence of these contaminants and treated water from conventional municipal, WWTPs [3], considering that WWTPs were not originally designed to remove these types of pollutants. An untargeted approach based on uHPLC (using a ZIC-cHILIC stationary phase) coupled to high resolution mass spectrometry (by using Orbitrap Q-Exactive) was carried out to explore the presence of EOCs and for some of them estimate the content by using reference standards. EOCs were sampled and preconcentrated by solid phase extraction (by testing different SPE cartidreges). The data arising from uHPLC-HRMS are very complex and we resort to Multivariate Curve Resolution – Region of Interest (ROI-MCR) method [4], taking care of properly optimizing the hyperparameters, followed by further multivariate data analysis to compare the different samples sites.

Acknowledgements
This study is funded from the Ministry of University and Research under the Call for Proposals related to the scrolling of the final rankings of the PRIN 2022 call for proposals. Project Title: EXploitation of targeted and untargeted analytical strategies for WASTEwateR monitoring: toward a sustainable water management according to the principles of circular economy- EXWASTER. Prot. n° 2022L3AH34. Principal Investigator: C. Durante.

References

[1] Pérez-Fernández, V., Mainero Rocca, L., Tomai, P., Fanali, S., Gentili, A., Anal. Chim. Acta, 2017 (983), 9-41

[2] NORMAN list of emerging substances, https://www.norman-network.net/?q=node/81

[3] Peña-Guzmán, C., Ulloa-Sánchez, S., Mora, K., Helena-Bustos, R., Lopez-Barrera, E., Alvarez, J., Rodriguez-Pinzón, M., J. Env. Management, 2019 (237), 408-423

[4] Gorrochategui, E., Jaumot, J., Tauler, R., BMC Bioinformatics, 2019 (20), 256

Micro-Scale Spectral Imaging and Chemometrics for Rapid Detection and Monitoring of Microplastics

Emilio Catelli[1], Ilenia Saraceno[1], Cristina Malegori[2], Barbara Bravo[3], Jacopo La Nasa[4], Greta Biale[4], Angelo Traina[3], Silvia Prati[1], Francesca Modugno[4], Paolo Oliveri[2], Giorgia Sciutto[1],

[1] *Department of Chemistry, University of Bologna, via Guaccimanni, 42, 48121, Ravenna, Italy*
[2] *Department of Pharmacy, University of Genova, Viale Cembrano, 4, 16148 Genova, Italy*
[3] *Thermo Fisher Scientific, Via San Bovio, 3, 20054 Segrate, Italy*
[4] *Department of Chemistry and Industrial Chemistry, University of Pisa, via G. Moruzzi 13, 56124 Pisa, Italy*
emilio.catelli2@unibo.it

Microplastics (MPs), defined as plastic particles smaller than 5 mm, originate from a wide range of sources, including everyday plastic use and industrial processes. Their small size, chemical inertness, and high durability allow them to persist in the environment for decades, raising significant concerns regarding their ecological and human health impacts. A major challenge in analytical chemistry is developing non-destructive, rapid, and reliable methods for identifying MPs, essential both for environmental monitoring and the implementation of mitigation strategies.

In this study, we present a novel approach that integrates micro-spectral imaging (micro-SI) techniques and chemometrics analysis to map MPs deposited on filters. SI data were collected using Thermo Scientific RaptIR+ and iN10 MX imaging microscopes, operating in the near-infrared (10000-4000 cm^{-1}) and medium infrared (4000-675 cm^{-1}) ranges, respectively. Aqueous suspensions of polypropylene (PP) and polystyrene (PS) MPs, both aged and not aged, were prepared and filtered through Whatman filters via vacuum filtration. The filters were then analyzed in mapping and imaging mode, generating hyperspectral datasets with pixel size of 50 μm × 50 μm (RaptIR+) and 25 μm x 25 μm (iN10 MX). To efficiently identify MPs, Principal Component Analysis (PCA) and Multivariate Curve Resolution – Alternating Least Squares (MCR-ALS) were applied. PCA was used to explore the dataset and highlight the major patterns related to MPs. Successively, MCR-ALS was employed to produced detailed distribution maps of the MPs, including the smallest particles, by decomposing the dataset into pure component spectra and corresponding concentration maps. Additionally, an automatic background segmentation was applied on concentration maps to automatically extract information on MP number and size. This chemometric approach significantly enhances the identification of MPs, allowing the distinction of the two compounds through the use of the concentration maps and characteristic overtone and combination bands in the reconstructed pure spectra profiles. Our findings demonstrate that integrating NIR and MIR micro-SI with MCR offers a powerful, non-destructive, and highly sensitive methodology for advancing microplastic analysis in environmental monitoring.

Minor component detection in imaging. Purest variable detection and variance-based approaches

Albert Sicre-Conesa[1], Adrián Gómez-Sánchez[1,2], Anna De Juan[1]
[1] *Chemometrics Group, Universitat de Barcelona, Martí i Franquès, 1, 08028 Barcelona, Spain*
[2] *LASIRE (UMR 8516), Univ. Lille, CNRS, Laboratoire Avancé de Spectroscopie pour les Interactions, la Réactivité et l'Environnement, Lille, France*
asicrec@ub.edu

Detecting minor components in hyperspectral images is crucial to determine the right number of components and derive initial estimates for unmixing analysis (e.g., MCR-ALS) [1]. Some factors like the relative contribution of the minor components to the overall signal and the noise level of the measurement can significantly affect the detection process [2]. This work is oriented to understand advantages and limitations of existing approaches when doing this detection task.

The methodologies envisioned to detect minor compounds can be roughly classified into those based on variance criteria, such as PCA and related methods, or those based on purest variable detection, where dissimilarity among image spectra is the driving force to detect different compounds in the system [3-5]. This work will test different methodologies from these two main families. To do so, simulated data will be generated with different configurations to mimic possible challenging scenarios in image analysis. The factors studied in these simulations will be the noise level of the measurement, the contribution of the minor component to the overall signal and the spatial and spectral overlap among components. Different qualitative and quantitative indicators will be used to assess the performance of the methods and the effect of the different factors studied. The ultimate goal is proposing some guidelines to choose the best approach for minor compound detection according to the characteristics of the imaging measurement.

References

[1] de Juan, A., Tauler, R., Anal. Chim. Acta, 2021 (1145), 59-78
[2] Vitale, R., Ruckebusch, C., J. Chemometr., 2023 (37), e3442
[3] Jolliffe, I.T., Principal Component Analysis, 2002
[4] Windig, W., Bogomolov, A., Kucheryavskiy, S., In Comprehensive Chemometrics: Chemical and Biochemical Data Analysis, Second Edition, 2020, 107-136
[5] Ruckebusch, C., Vitale, R., Ghaffari, M., Hugelier, S., Omidikia, N., TrAC Trends Anal. Chem., 2020 (132), 116044

Multiblock analysis of restored Raffaello painting by XRF-VNIR-SWIR hyperspectral system

Emmanuel Cruz Muñoz[1], Emilio Catelli[2], Zelan Li[2], Silvia Prati[2], Davide Ballabio[1], Michele Occhipinti[3], Giorgia Sciutto[2]

[1] *Milano Chemometrics and QSAR Research Group, Department of Earth and Environmental Sciences,University of Milano-Bicocca, P.za della Scienza 1, 20126, Milano, Italy*

[2] *Department of Chemistry, University of Bologna, via Guaccimanni, 42, 48121, Ravenna, Italy*

[3] *Bruker Nano Analytics, via Conte Rosso 23, 20134, Milano, Italy*

enmanuel.cruzmunoz@unimib.it

In recent years, chemometrics has emerged as a valuable complement in the field of cultural heritage [1], particularly in support of non-destructive analysis techniques such as spectral imaging. However, several challenges still need to be addressed, chief among them is the ability to extract stratigraphic information about the spatial distribution of organic and inorganic compounds within multilayered historical samples [2]. This information is crucial in restoration campaigns. Due to their historical and artistic value, understanding the precise pigments and binders originally used by the artist is essential before initiating any restoration work. Additionally, gaining insight into how the artist layered pigments to create the final composition provides valuable knowledge about historical painting techniques and the creative process itself.

In this context, *Il San Sebastiano* (1502) by the renowned Italian painter Raffaello has been analyzed using the IRIS hyperspectral device, which enables simultaneous acquisition of XRF (0.5–40 keV), VNIR (400–1087 nm), and SWIR (1080–2500 nm) data with a shared pixel resolution. This setup allows for the extraction of both elemental and molecular information from the surface down to subsurface layers, with each technique offering a different penetration depth. The painting is housed at the Accademia Carrara in Bergamo, and the analyses were conducted in collaboration with the museum. Each data block was initially pre-processed independently using dedicated techniques tailored to its nature. Following pre-processing, low-level data fusion was performed. The analysis focused primarily on two specific areas of the painting that had undergone recent restoration: the sky and the forest background. The multivariate analysis strategy involved a combination of Principal Component Analysis (PCA), spectral unmixing through Multivariate Curve Resolution–Alternating Least Squares (MCR-ALS), and correlation mapping (within and between blocks). This approach revealed notable pigment distributions across the paint stratigraphy and highlighted spectral differences that correspond to the recently restored features.

References

[1] G. Sciutto, P. Oliveri, S. Prati, M. Quaranta, S. Bersani, R. Mazzeo, Anal. Chim. Acta, 2012, 752

[2] E. Catelli, Z. Li, G. Sciutto, P. Oliveri, S. Prati, M. Occhipinti, A. Tocchio, R. Alberti, T. Frizzi, C. Malegori, R. Mazzeo, Anal. Chim. Acta, 2023, 1239

NIR and MIR-Hyperspectral Imaging for the detection of salt weathering in *buon fresco* and *secco* wall paintings

Iñaki Vázquez-De La Fuente[1], Giulia Gorla[1], Sara Puente-Muñoz[1], Nagore Prieto-Taboada[1], Gorka Arana[1], Federico Marini[2], Juan Madariaga[1]

[1] *IBeA Research Group, Department of Analytical Chemistry, Faculty of Science and Technology, University of the Basque Country UPV/EHU, Barrio Sarriena s/n, 48940, Leioa, Spain*

[2] *Department of Chemistry, Sapienza University of Rome, Piazzale Aldo Moro 5, 00185 Rome, Italy*

inaki.vazquez@ehu.eus

Buon fresco and *secco* are two wall painting techniques that differ in their preparation, leading to variations in material chemistry and durability. Despite these differences, both are vulnerable to salt crystallization, which can cause surface efflorescence or, more critically, subefflorescence, risking detachment of the outer layers [1]. Detecting these alterations at an early stage is crucial for conservation efforts. However, traditional analytical methods are often limited in scale and require sampling, which may not be feasible for large or fragile artworks.

Infrared hyperspectral imaging provides a rapid and non-invasive alternative for analyzing large surfaces. To assess its potential for detecting surface efflorescence and subefflorescence, mock-ups of *buon fresco*, *secco*, and their preparatory layers—both salt-weathered and unweathered—were analyzed using two Specim FX series cameras: FX17 (950–1700 nm) and FX50 (2710–5250 nm). Principal component analysis (PCA), performed using the HYPER-Tools toolbox [2], showed that each camera provided complementary information: FX17 was effective for mapping pigments, while FX50 highlighted differences in preparatory layers. Both cameras successfully differentiated samples affected by salt weathering in the *intonaco* layer (Figure 1). These findings highlight hyperspectral imaging's potential for detecting early degradation in wall paintings and its broader applications in cultural heritage preservation.

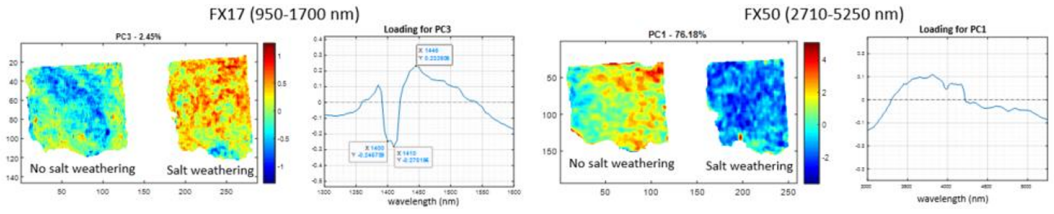


Figure 1. Score and loading plots showing separation of areas with and without salt weathering. Images were captured using FX17 (PC3 – 2.45%) and PX50 (PC1 – 76.18%) cameras

Acknowledgements

This work was supported by SIDIRE project (PID2023-146230OB-I00) funded by the Spanish

Agency for Research AEI (MICIU/AEI/10.13039/501100011033/FEDER/UE).

References

[1] Comite, V.; Bergomi, A.; Lombardi, C.A.; Borelli, M.; Fermo, P., Appl. Sci., 2023 13, 6623
[2] Mobaraki, N.; Amigo, J.M., Chemometr. Intell. Lab., 2018 (172), 174-187

NMR-Based Metabolomics for Occupational and Environmental Biomonitoring in Foundry Settings

Federico Marini[1,2], Michele De Rosa[1,2], Mariangela Spagnoli[3], Fabio Sciubba[2,4], Giovanna Tranfo[3], Silvia Canepari[4]

[1] *Department of Chemistry, Sapienza University of Rome, Rome, Italy*
[2] *NMR-Based Metabolomics Laboratory (NMLab), Sapienza University of Rome, Rome, Italy*
[3] *Department of Medicine, Epidemiology, Environmental and Occupational Hygiene, INAIL, Monteporzio Catone, Italy*
[4] *Department of Environmental Biology, Sapienza University of Rome, Rome, Italy*
federico.marini@uniroma1.it

The application of omics sciences to occupational health research has opened novel perspectives for early biomonitoring of workers exposed to complex toxic environments. Among these, metabolomics provides a powerful tool to detect subtle biological alterations associated with hazardous exposures. In this study, an NMR-based metabolomics approach was employed to investigate urinary metabolic profiles from three groups: foundry workers directly exposed to emissions (n = 64), residents living in the surrounding areas (n = 77), and non-exposed healthy volunteers serving as controls (n = 62). Spot urine samples were analyzed by ^{1}H NMR spectroscopy, and the resulting metabolic fingerprints were subjected to both unsupervised (PCA, URF) and supervised (PLS-DA) chemometric analyses.

The results revealed that both workers and residents exhibited analogous metabolic trends when compared with the control group. Perturbations were observed in amino acid metabolism, the tricarboxylic acid (TCA) cycle, and short-chain fatty acid metabolism. Moreover, common alterations were identified in key metabolites, including hypoxanthine, furoylglycine, hippuric acid, 4-hydroxybenzoic acid, and 1-methylnicotinamide. These metabolites are linked to pathways associated with oxidative stress, inflammation, and disruptions of energy metabolism, indicating that occupational exposure and environmental emissions from foundry activities have systemic biological effects that extend beyond the workplace.

The findings highlight the potential of metabolomics to complement conventional biological monitoring, enabling the detection of early metabolic responses even at exposure levels below established threshold limit values or in the presence of complex mixtures. Furthermore, the inclusion of residents in the study underscores the broader public health implications of industrial activities, pointing to possible long-term effects on surrounding populations. This integrative approach provides a valuable framework for refining risk assessment strategiesand for the design of preventive and protective measures to safeguard workers' and communities' health.

Optimization and classification of grasslands spots based on their management using remote sensing and chemometrics

Carlos Martin Santiago[1,2], Jose Manuel Amigo[2,3], Enmanuel Cruz Muñoz[1], Davide Ballabio[1], Arantza Aldezabal Roteta[2], Kepa Castro Ortiz De Pinedo[2]

[1] *University of Milano-Bicocca, Department of Earth and Environmental Sciences, Milan, Italy*
[2] *University of Basque Country, Faculty of Science and Technology, Leioa, Spain*
[3] *Ikerbasque, Basque Foundation for Sciences, Bilbao, Spain*
c.martinsantiago@campus.unimib.it

Mountain grasslands and meadows are open herbaceous habitats that make up one-third of the agricultural area of the European Union. These habitats provide a variety of services such as food, fodder production, freshwater retention and filtration. They are also important for biodiversity conservation of native species. Given the importance of these habitats, the European Commission committed to biodiversity protection and the conservation of natural habitats and species through the Habitats Directive 92/43/EEC (EUR-Lex - 31992L0043) [1]. Knowing the importance of this kind of habitats and considering the obligation to create a national report every six years on the state of these ecosystems, remote sensing combined with chemometrics could be a very useful tool.

Although this technology has been used in different countries, such as Italy for the green payments of the Common Agricultural Policy [2] or in the timber industry [3], it has not been so well-developed for the ecosystems of our interest. In this study, remote sensing data was obtained from the European Space Agency (ESA) through its Copernicus programme for the Basque Country area (Spain). We selected the satellite Sentinel-2, which measures in 12 different bands, from 425 nm to 2202 nm [4]. However, bands are measured with different resolutions (10 m, 20 m and 60 m). To face these differences, two different methods were used to align the data. On one hand, we used a bicubic interpolation where the pixel value is a weighted average of the pixels in the nearest 4-by-4 neighbourhood. On the other hand, we used a convolutional neural network, which allows to improve the resolution of the 20 m and 60 m bands [5]. Moreover, according to the authors, this CNN preserves better spectral characteristics. The upsampled data was analysed by Principal Component Analysis (PCA) finding the best separation between cut and no-cut grassland based on the spots that were measured in 2022 in Basque Country. These spots were selected by the botanist of the University of Basque Country from July to November and labelled as cut or uncut accordingly.

It is important to take into account the size of the different patches used for the analysis, as the homogeneity could be very different from one to the other, as the way in which a grassland can be managed is diverse, so it can be used solely for livestock fodder production, or it can be used directly for livestock grazing or the grassland can be abandoned.

This is why the selection of a good interpolation technique can be of significant importance

in the subsequent use of the different grading models. One reason is that techniques such as neural networks can be sensitive to the shapes found in the images. Interpolation could cause the lost of these details, which could diminish the quality of the predictions of these models. This interpolation is not only important in classification; it has been seen in this work how the use of the two used interpolations creates variations in the PCA results, where the same samples clustered differently depending on the technique used. It is for all these reasons that this work seeks to find the most effective way to make a correct interpolation, not only taking into account the quality of the data obtained, but also the time and computational resources required for each one.

References

[1] EUR-Lex - 31992L0043 - ES. D. Of. N° 206 22071992 P 0007 - 0050 Ed. Espec. En Finés Capítulo 15 Tomo 11 P 0114 Ed. Espec. Sueca Capítulo 15 Tomo 11 P 0114

[2] Filippo, S.; Samuele, D.P.; Enrico, B.M. Detection and Counting of Meadow Cuts by Copernicus Sentinel-2 Imagery in the Framework of the Common Agricultural Policy (CAP). Eur. J. Remote Sens., 2022

[3] Forstmaier, A.; Shekhar, A.; Chen, J. Mapping of Eucalyptus in Natura 2000 Areas Using Sentinel 2 Imagery and Artificial Neural Networks. Remote Sens., 2020 (12), 2176

[4] Sentinel-2 - Missions - Sentinel Online. https://copernicus.eu/missions/sentinel-2 (accessed 2023-06-21)

[5] Lanaras, C.; Bioucas-Dias, J.; Galliani, S.; Baltsavias, E.; Schindler, K. Super-Resolution of Sentinel-2 Images: Learning a Globally Applicable Deep Neural Network. ISPRS J. Photogramm. Remote Sens., 2018 (146), 305–319

Orthogonalization methods to cope with interferences from moisture and scattered solids content of virgin olive oil on NIRS methods. An example for acidity prediction

José Antonio Cayuela-Sánchez[1], Jean-Michel Roger[2,3]
[1] *Instituto de la Grasa (IG), CSIC. Campus University Pablo de Olavide, Ed. 46, Ctra. Utrera km 1, 41013 Sevilla, Spain*
[2] *ITAP, Univ. Montpellier, INRAE, Domaine de Lavalette, Montpellier, France*
[3] *ChemHouse Research Group, Montpellier, France*
jacayuela@ig.csic.es

Near infra-red spectroscopy (NIRS) or visible and NIR (VISNIRS) methods is suitable for measuring main quality markers of olive oil, such as acidity and peroxide value, according to several reports [1]. Nevertheless, freshly produced virgin olive oils (VOO) is a complex colloid emulsion-sol. The cause is it contain variable amounts of microscopic water droplets of vegetation water and solid particles from olive fruit, homogeneously scattered in the triglyceride matrix [2]. This physical state provides its cloudy appearance. Preparing olive oil samples by filtration for NIRS analysis is not referred frequently in the literature. However, the colloid emulsion-sol features of fresh VOO imply important interferences for NIRS measures. Therefore, for use correctly NIRS on VOO, some suitable method is required to avoid such interferences. The most simple and easy one can be filtering the samples, by standard procedures. Notwithstanding, doing so for on-line or in-situ NIRS uses is not easy. Characterizing fresh virgin olive oils in the mills on unaltered samples is of great interest, for simplifying decisions on batch management. VOO free acidity is one most relevant feature, which on-line determination in the olives mill is of great interest. Thus, alternative solutions are required. Correcting NIR spectra by signal orthogonalization methods has been reported as suitable solution to remove from the spectra the influence of various sources of interference, such as temperature or humidity, among others [3]. This study evaluated the external parameter orthogonalization (EPO) method for NIRS measuring free VOO acidity using PLS models. The method consists on treating NIR spectra to avoid the influence from the scattered phase of VOO, mainly consisting on moisture. The results indicated the EPO method allowed to extract the useful information from NIR spectra, by removing the useless wavelength intervals, for VOO free acidity measurement. The same method can apply to other VOO quality markers.

References

[1] Armenta, S., Moros, J., Garrigues, S., De La Guardia, M., Crit. Rev. Food Sci. Nutr., 2010 (50), 567-582
[2] Cayuela-Sánchez, J.A., Caballero-Guerrero, B., Trends Food Sci. Technol., 2019 (83), 78-85
[3] Roger, J.M., Boulet, J.C., Journal of Chemometrics, 2018, e3045

PARAFAC combined with DD-SIMCA on EEM fluorescence spectra: an analytical method for the authentication of Slovakian Tokaj Wine

Cristina Malegori[1], Elisa Giacalone[1], Adriano De Araujo Gomes[2], Mateus Pires Schneider[2], Paolo Oliveri[1], Ivan Spanik[3]

[1] *Department of Pharmacy (DIFAR), University of Genova, Viale Cembrano 4, 16148, Genova, Italy*
[2] *Institute of Chemistry, Federal University of Rio Grande do Sul (UFRGS), Av. Bento Gonçalves 9500, 90650-001, Porto Alegre, Brazil*
[3] *Institute of Analytical Chemistry, Slovak University of Technology in Bratislava, Radlinskeho 9, 81237 Bratislava, Slovakia*
cristina.malegori@unige.it

The Protected Designation of Origin certification is an essential instrument to ensure food safety, quality, and authenticity of products. For this reason, Slovak Tokaj wine, known for its unique production processes, represents a product of great interest for obtaining this certification. This wine is part of the botrytized wine group, in which the grapes infected by the fungus *Botrytis cinerea* produce a sweet and aromatic wine. In the present work, an analytical method for the authentication of Slovak Tokaj wine was developed and validated, using fluorescence spectroscopy, a selective and non-destructive analytical technique, combined with multivariate data processing techniques such as PARAFAC (Parallel Factor Analysis) and DD-SIMCA (Data-Driven Soft Independent Modelling of Class Analogy). The focus was to develop a method that allows the classification of Slovak Tokaj wine compared to other botrytized wines from different countries. 75 samples of botrytized wine from Slovakia, Hungary, France, Austria, and Ukraine were analyzed. EEM (Emission-Excitation Matrix) spectra were acquired covering all wavelengths of the fluorescence spectrum, with excitation wavelengths ranging from 280 nm to 675 nm, and emission wavelengths from 300 nm to 700 nm. The collected data, after eliminating the Rayleigh effect, were explored with PARAFAC, identifying four components. These four components were associated with four chemical compounds present in the wine, such as kaempferol, quercetin, riboflavin, and catechin/epicatechin. The use of DD-SIMCA allowed the development of a predictive classification model, with a correct prediction percentage of 90.67%. These results confirm that Slovak Tokaj wines can be distinguished from other botrytized wines due to differences in the concentration of fluorophores, which are a consequence of the specific production processes and climatic conditions of the different geographical regions. Thanks to the results of this work, further useful data are provided for the request of PDO recognition for Slovak Tokaj wine.

Acknowledgements
I.S. would like to thank the Slovak Research and Development Agency for Contract No. APVV-21-0211; M.P.S. would like to thank the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

Pattern recognition techniques for the study of saffron provenance. Third-order data generation by in-flow photodegradation setup

Arsenio Muñoz De La Peña[1], Mirta Alcaraz[2], Isabel Durán-Merás[1], Héctor C. Goicoechea[2]
[1] *Department of Analytical Chemistry, University of Extremadura, Badajoz, Spain*
[2] *LADAQ, QAI, FBCB, Universidad Nacional del Litoral-CONICET, Santa Fe, Argentina*
arsenio@unex.es

The management of higher-order data influences the advancement of analytical strategies for identifying characteristic patterns of analytes in complex matrices. Consequently, there is an increasing focus on developing second-order and third-order data modelling applications for pattern recognition (un- and supervised) tasks, although third-order data have been less commonly applied for these purposes. Two recent publications describe data characteristics, algorithms, processing procedures and applications of second-order and third-order data for classification issues [1, 2].

In this work, an in-flow photodegradation setup was developed to generate third-order data by monitoring the photodegradation reaction of a saffron solution by acquiring excitation emission matrices (EEM). Sixteen saffron brands from different Spanish regions and countries were analysed. For data generation, 10 EEMs were acquired, covering a total of 30 minutes of continuous UV irradiation, obtaining a (10×61×35) UV-EEM third-order data for each sample. First, PARAFAC was used to analyse the compounds present in the samples and their behaviour over time. After that, the data were modelled, using the well-known unsupervised pattern recognition method principal component analysis (PCA), to study the geographic origins of saffron grown in different countries and conditions. This study analysed the advantages of the third-order data (21350 data points) over the second-order data represented by the EEM without irradiation (2135 data points). Results showed that the different groups were better separated when using third-order data. More studies are being developed to exploit not only the amount of information but the data structure, for example, using algorithms exploiting the three-dimensional capabilities.

Acknowledgements
Ministerio de Ciencia, Innovación y Universidades of Spain (Project PID2024-156397NB-I00)

References

[1] Azcarate, S.M., de Araújo Gomes, A., Muñoz de la Peña, A., Goicoechea, H.C., TrAC - Trends in Analytical Chemistry, 2018 (107), 151-169
[2] Cabrera, M., Valdés-Sánchez, E., Muñoz de la Peña, A., Durán-Merás, I., Talanta, 2019 (199), 652-661

PoliBrush for processing spectral images – an educational software

Cristina Malegori[1], Paolo Oliveri[1], Rodrigo Rocha De Oliveira[2], Giorgia Sciutto[3]
[1]*Department of Pharmacy (DIFAR), University of Genova, Viale Cembrano 4, 16148, Genova, Italy*
[2]*Department of Chemical Engineering and Analytical Chemistry, University of Barcelona, Martí i Franquès 1, 08028 Barcelona, Spain*
[3]*Department of Chemistry, University of Bologna, Ravenna Campus, Via Guaccimanni, 42, 48121 Ravenna, Italy*
cristina.malegori@unige.it

Since the first introduction of hyperspectral imaging in chemistry, the need for dedicated software was recognized by the chemometric community. Nonetheless, considering that such an introduction has been relatively recent, availability of software for multivariate analysis of hyperspectral images is still limited, if compared with general chemometric applications. PoliBrush is a freely distributed, stand-alone software designed for teaching exploratory multivariate analysis in the frame of color RGB and spectral imaging.
Several approaches for spectral and spatial preprocessing are available: from a spectral perspective, the most used row transformation are implemented (scale transformation, SNV, Savitzky-Golay derivatives, normalization) while, from a spatial extent, several masking strategy are proposed (single channel, ratio wavelengths, NDI).
PoliBrush implements principal component analysis (PCA) as its core method. It features a single main window that provides users with essential tools for spectral image preprocessing and exploration. In more detail, the main window presents six separated graphical output panels: total intensity image, average/maximum spectral profile, PCA scree plot, loading line plot, score map, and score scatter plot. The software emphasizes an interactive brushing approach, enabling users to gain a comprehensive understanding of the relationships between PCA score space and image pixel space.
The software has already been successfully used in various educational settings and workshops on multivariate approaches for spectral imaging, proving its effectiveness in teaching key concepts. A dedicated tutorial paper was recently published [1], where PoliBrush usage and its operational procedure are demonstrated through two case studies, whose data are made available for download.

Acknowledgement
Financial support provided by the Italian Ministry of Universities and Research - MUR (Research Project PRIN 2022 No. 20223WBTH8, CUP: D53D23008950006. Research Project PRIN 2022 PNRR No. P20227N5YZ, CUP: D53D23016770001) is gratefully acknowledged.

References
[1] Rocha de Oliveira, R., Malegori, C., Sciutto, G., Oliveri, P., Chemometr. Intell. Lab., 2023 (240), 104918

Rapid viral detection in milk using near infrared spectroscopy combined with chemometrics and neural networks

Samy De Oliveira[1], Stéphane Guyot[1], Laurence Dujourdy[2,3], Emmanuel Denimal[2], Caera O'Neill[1], Alexis De Rougemont[1,4], Elias Bou-Maroun[1]

[1] *UMR PAM, Institut Agro, Université Bourgogne Europe, INRAE, F-21000 Dijon, France*
[2] *Institut Agro Dijon, Direction Scientifique, Cellule d'Appui à la Recherche en sciences des données, F-21000 Dijon, France*
[3] *Laboratoire d'Informatique de Bourgogne (LIB), Data Science, Université Bourgogne Europe, F-21000 Dijon, France*
[4] *National Reference Centre for Gastroenteritis viruses, Laboratory of Virology, University Hospital of Dijon, F21000 Dijon, France*
samy.de-oliveira@agrosupdijon.fr

Human norovirus (HuNoV) is the leading cause of foodborne illnesses, necessitating rapid, reliable, and cost-effective detection methods. Near InfraRed Spectroscopy (NIRS) offers real-time analysis with minimal preparation. Two series of quadruplicates of UHT semi skimmed milk samples spiked with MS2 bacteriophage concentrations (a HuNoV surrogate) performed at 8 months interval by two independent operators were analyzed [1]. Partial least squares regression based on 3-way multiblock models predicted virus concentrations and correlated NIRS spectra with the PFU reference method [2]. A 5-fold cross-validation with four pretreatments was applied to 88 samples. Calibration and prediction models using multiplicative scatter correction followed by standard normal variate yielded high R^2 values (0.98 and 0.95) with 14 latent variables. A two-layer perceptron achieved a R^2 of 0.947 showing a good prediction of viral concentrations [3] (Figure 1). NIRS quantified 80 PFU/mL in complex matrices, competing with RT-PCR (10^2-10^3 viral particles per mL), demonstrating its viability for viral detection [4].

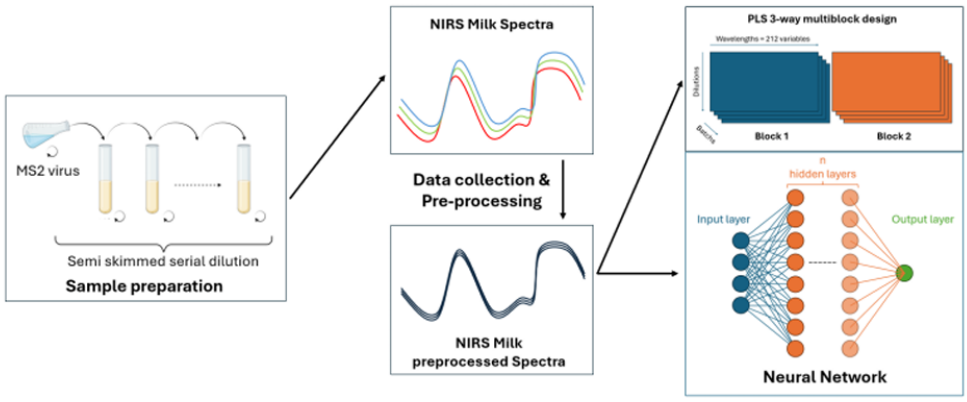


Figure 1. Sample preparation and chemometric methods applied to NIRS data

References

[1] Yavarmanesh, M., Alum, A., Abbaszadegan M., Food Environ. Virol., 2015 (7), 232-238

[2] Mishra, P., Liland, K. H., Anal. Chim. Acta, 2022 (1206), 339786
[3] Mishra, P., Passos, D., Marini, F., Xu, J., Amigo, J.M., Gowen, A., Jansen, J., Biancolillo, A., Roger, J.M., Rutledge, D., Nordon, A., TRAC-Trend. Anal. Chem., 2022 (157), 116804
[4] De Oliveira, S., Guyot, S., Dujourdy, L., Denimal, E., Gal, L., O'Neill, C., Cayot, P., de Rougemont, A., Bou-Maroun, E., Food Res. Int., 2025, In Press

XII COLLOQUIUM CHEMIOMETRICUM MEDITERRANEUM
ÎLE DE PORQUEROLLES
9 - 12 SEPTEMBER 2025

SERS quantitative analysis of uracil with multi-batches nanoparticles suspensions: how to deal with the sample variability?

Marion Berge[1,2], Eric Caudron[1,2], Charlotte De Bleye[3], Eric Ziemons[3], Laetitia Lê[1,2]
[1] *Pharmacy, European George Pompidou Hospital, APHP, Paris, France*
[2] *Lip(sys)², University of Pharmacy Paris-Saclay, Orsay, France*
[3] *University of Liege, CIRM, Vibra-Santé HUB, Laboratory of Pharmaceutical Analytical Chemistry, Belgium*
marion.berge@universite-paris-saclay.fr

Surface Enhanced Raman Scattering (SERS) spectroscopy combines Raman spectroscopy with metallic substrates as silver nanoparticles (AgNPs) to enhance Raman scattering of analyte. In order to determine DPD phenotyping, prerequisite for 5FU treatment of cancer patients, this technique was applied to quantify uracil (U). Despite proven ability for low concentration analysis, a limited reproducibility of NPs aggregation and interaction with analyte was observed. The aim of this study was to assess the optimal calibration strategy to deal with sample variability due to multi-batches nanoparticles suspension.

Thus, U was analyzed a various concentrations in aqueous matrix from 5 to 50 ng mL^{-1} with AgNPs and $NaNO_3$ at 0.8 M as the aggregating agent. Six independent colloidal suspensions of AgNPs (S1 to S6), synthesized according to Lee and Meisel procedure, were tested. Data acquisition was carried out using an Agilent TRS-100 Raman spectrometer equipped with a 830 nm laser during 3 seconds. Using Matlab® software, partial least squares regression models were developed to predict U concentration. To develop predictive quantitative model, two strategies based on independent mono batch (including samples only obtained with S1 or S2 or S3) or multi batch (including samples obtained with S1, S2, S3) NPs constituting calibration set were tested to predict samples prepared with different NPs batches (S4, S5, S6). To compare the predictions of each developed model with each other, a paired data Student's t-test was used.

A total of 990 spectra were acquired. For both strategies, the best models were obtained using partial least squares regression after baseline correction and semi logarithmic transformation of U concentrations with R^2 higher than 0.990 and root mean square error of cross validation between 1.05 and 1.96 ng mL^{-1} for mono-batch calibration models and 1.36 ng mL^{-1} for multi-batch calibration model. For each mono-batch model, the predictions were significantly different (*p-value* < 0.05): there was an influence of the suspensions batch used for the calibration on the prediction. However, the predictions obtained with the multi-batch calibration model were significantly higher than predictions obtained by mono-batch calibration S1 and S3 (*p-value* = 0.0035 and 7.29×10^{-13} < 0.05). Despite high performance calibration models with low RMSECV, multi-batch calibration allows the construction of more powerful and robust prediction models. Therefore, for prediction of unknown samples from suspensions unknown to the model, this calibration strategy

appears to be optimal.

Shelf-Life Analysis of Plant-Based Foods Using RGB and NIR Hyperspectral Imaging Combined with MCR-ALS

Carlotta Sapienza[1], Anna de Juan[2], Rodrigo Rocha De Oliveira[2]
[1] *Università di Bologna, Bologna, Italy*
[2] *Chemometrics Group, Dept. of Chemical Engineering and Analytical Chemistry, Universitat de Barcelona, Barcelona, Spain*
anna.dejuan@ub.edu

Plant-based products have seen an increasing demand in recent years. Despite this trend, there is a lack of studies investigating the chemical and color evolution of these products over their shelf-life. Only scarce research on the capabilities of hyperspectral imaging to detect contaminants in plant based food has been recently stated [1]. This work is oriented to provide a methodology for plant based food degradation processes by combining RGB imaging and near-infrared (NIR) hyperspectral imaging with chemical information, such as pH measurements and weight loss. The approach is tested on some commercial plant-based foods, such as seitan-based burgers and pea protein salame. Samples were stored at three temperatures (5°C, 10°C, 20°C) and monitored for up to 14 days, or until visible spoilage occurred.

Processed plant-based foods are inherently heterogeneous, implying a spatial variation of their physicochemical properties. NIR hyperspectral imaging enabled the study of the spatial distribution of product components and their evolution over time. In this context, Multivariate Curve Resolution (MCR) was applied to image time series associated with the degradation process to extract qualitative (spectral), quantitative (spatial) information and temporal degradation trends of the food sample constituents [2].

As a complementary measurement, color RGB image time series were also acquired. RGB images were corrected for white-dark variation and lighting fluctuations using a checkerboard. From these images, colorgrams were created to study color composition, distribution and its changes through time. MCR-ALS modeling was applied to the RGB colorgrams [3]. Concentration maps were reconstructed to visualize the spatial distribution of the chemical components detected by NIR. Color components and their change during the shelf-life analysis were also identified.

The results demonstrate that RGB imaging can serve as a fast, low-cost indicator of product degradation, while hyperspectral imaging coupled with MCR provides deeper insights into the spatial and chemical changes occurring during storage.

References

[1] T. Yu, S. Yao, Z. Zhang, A. F. Victorio Ramos, L. Rodriguez-Saona, and J. Wang, Trends Food Sci. Technol., 2025 (160), 105017

[2] A. De Juan, R. Rocha De Oliveira, and A. Gómez-Sánchez, in Data Handling in Science and Technology, Elsevier, 2024 (33), 111–132

[3] R. Calvini, G. Orlandi, G. Foca, and A. Ulrici, Chemom. Intell. Lab., 2020 (196), 103915

Spectroscopic monitoring of table olive fermentation process

Jokin Ezenarro[1], Daniel Schorn-García[2], Elisavet Vakouli[3], Maria Kazou[3], Efstathios Panagou[3]

[1] *Department of Food Science, Faculty of Science, University of Copenhagen, Denmark*
[2] *South African Grape and Wine Research Institute, Stellenbosch University, South Africa*
[3] *Laboratory of Microbiology and Biotechnology of Foods, Department of Food Science and Human Nutrition, School of Food and Nutritional Sciences, Agricultural University of Athens, Greece*
jokin@food.ku.dk

Table olives are a basic Mediterranean food, valued for their cultural, nutritional, and economic importance. However, table olives are traditionally processed in brines rich in sodium chloride (NaCl), and excessive sodium intake is a major public health concern, as it contributes to hypertension and cardiovascular diseases. Therefore, strategies to reduce the sodium content, such as partial substitution with potassium chloride (KCl), have gained increasing attention [1]. However, it is essential to assess the impact of such changes during fermentation and on final product quality, and to develop reliable analytical tools to do so [2].

This study proposes the use of middle-infrared (MIR) and Raman spectroscopy to monitor the fermentation process of 'Kalamata' olives over 145 days. And also, the use of spectroscopy to evaluate the effect of sodium replacement on the spectroscopic profiles; for this, batches were produced with varying NaCl/KCl ratios (0%, 25%, 50%, 75% KCl). Spectral data were analysed using multivariate methods, including error covariance matrices, Principal Component Analysis (PCA), ANOVA – Simultaneous Component Analysis (ASCA), and Partial Least Squares (PLS) regression to identify variability sources, assess the influence of time and salt substitution on spectral signatures and physicochemical properties and to monitor the fermentation process.

Error covariance matrix analysis showed that intra-sample variability was greater than the variability between salt substitution levels, highlighting the dominant role of natural biological variation. ASCA confirmed that time was the main driver of spectral changes, while potassium substitution had a smaller effect, but with clear differences between the 0% KCl batches and those with higher substitution levels. PLS regression models accurately predicted physicochemical parameters such as pH and titratable acidity. And PCA based models proved useful for monitoring and potentially control of the fermentation process. This underscores the value of Raman and MIR spectroscopy, together with chemometrics, for both monitoring and predictive purposes, offering a useful analytical toolbox for the production of table olives.

References

[1] López-López, A., Moreno-Baquero, J.M., Garrido-Fernández, A., Foods, 2024 (13), 4061
[2] Tzamourani, A. et al., Int. J. Food Sci. Technol., 2024 (59), 4093-4108

Study of the environmental variability of coffee beans using NIR spectroscopy

Francesco Savorani[1], Nicola Cavallini[1], Silvia Milgietta[2], Carlotta Caruso[2]
[1] *Politecnico di Torino, Torino, Italy*
[2] *Luigi Lavazza S.p.A., Torino, Italy*
francesco.savorani@polito.it

Coffee is a worldwide significant commodity, with its quality and characteristics influenced by various factors, including environmental conditions [1, 2]. This study investigated the chemical variations in coffee beans resulting from different environmental conditions, using Near-Infrared (NIR) spectroscopy [3]. Samples were collected across different production stages in a single year: green coffee beans (June-October), roasted coffee (April-October) and final product, i.e. capsules, (February-October). Both 100% Arabica blends and blends containing Robusta were analyzed. NIR spectra were acquired using two different spectrometers: one operating from 4000 to 10000 cm^{-1} for both the green coffee and the final product and the other one operating from 5500 to 12500 cm^{-1} for the roasted coffee. Data analysis was performed using chemometric tools [4] such as preprocessing methods and Principal Component Analysis (PCA) on the SNV-corrected and mean-centered data to evaluate spectral variations ascribable to changes in chemical properties. After a global range investigation, specific spectral ranges were analyzed, including common wavelength ranges between the two spectrometers and focused spectral regions containing peaks related to important chemical components, to highlight differences among samples. The results indicate that environmental variability has no detectable effect on green coffee beans, but it becomes evident in roasted coffee and final product. Further analysis of NIR spectral intervals suggests that variations in lipid, protein and moisture contents are correlated with the industrial process. Overall, these findings demonstrate the potential of NIR spectroscopy for monitoring and understanding the impact of environmental factors and process parameters on coffee quality during processing. This approach can enable targeted adjustments to maintain consistent product characteristics.

References

[1] Galanakis, C. M. (2017). Handbook of Coffee Processing By-Products: Sustainable Applications. Elsevier, London, UK
[2] Adrià, F. (2018). Coffee Sapiens: comprendere per innovare. Giunti, Firenze, Italy
[3] Griffiths, P.R. (2010). Theory and instrumentation for vibrational spectroscopy. In: Applications of Vibrational Spectroscopy to Food Science. Wiley, Chichester
[4] Bro, R., Van den Berg, F., Thybo, A., Andersen, C.M., Jørgensen, B.M., Andersen, H. (2002). Multivariate data analysis as a tool in advanced quality monitoring in the food production chain, Trends in Food Science & Technology, 13, 235-244

The Role of Mindfulness in Modulating Stress and Well-Being Among Healthcare Students: A Clinical and Metabolomic Analysis

Célie Da Silva[1,2], Mathieu Galmiche[1,2], Oriane Strassel[1,2], Sergey Girel[1,2], Mathis Fiault[1,2], Ben Meuleman[3,4], Isabel Meister[1,2], Camille Piguet[3,5], Françoise Jermann[6], Julien Boccard[1,2], Serge Rudaz[1,2]

[1] *School of Pharmaceutical Sciences, University of Geneva, Geneva, Switzerland*
[2] *Institute of Pharmaceutical Sciences of Western Switzerland, University of Geneva, Geneva, Switzerland*
[3] *Psychiatry Department, Faculty of Medicine, University of Geneva, Switzerland*
[4] *Swiss Center for Affective Sciences, University of Geneva, Geneva, Switzerland*
[5] *General Pediatric Division, Geneva University Hospital, Switzerland*
[6] *Department of Psychiatry, Geneva University Hospitals, Geneva, Switzerland*
celie.dasilva@unige.ch

Today, there is growing evidence that healthcare workers are experiencing high levels of stress and psychological distress. This can alter their prosocial abilities, such as empathy and the capacity to support others, ultimately affecting their daily performance and the quality of care provided to patients. Although this issue is beginning to gain recognition, little has been done to formally address it – particularly during their training, when students already face considerable stress and mental strain [1]. In this context, Mindfulness-Based Interventions (MBIs) have shown effectiveness in various clinical settings [2]. This work presents an extensive metabolomic study involving multiple data blocks to explore the potential of Mindfulness-Based Cognitive Therapy for Life (MBCT-L) for reducing anxiety and stress while enhancing the overall well-being of healthcare students. First, clinical scores obtained from questionnaires were investigated using dimensionality reduction and hierarchical clustering (PCA-HCPC) to stratify the cohort into three distinct groups, namely "High Well-Being Improvement", "Moderate Well-Being Improvement", and "Low to No Well-Being Improvement". Then these groups were further used to drive chemometric analyses on several metabolomic data blocks (steroidomics and lipidomics) to deconvolute the biological modulations underlying the effects of MBCT-L and link these signatures to the observed clinical outcomes. The preliminary results obtained make it possible to discuss biological hypotheses to describe the beneficial effects of the intervention on participants' well-being. This work opens new perspectives for a better understanding of the benefits associated with mindfulness practice in the studied population.

References

[1] Franzen, J., Jermann, F., Ghisletta, P., Rudaz, S., Bondolfi, G., Tran, N.T., Int. J. Environ. Res. Public Health, 2021, 18, 2151
[2] Gu, J., Strauss, C., Bond, R., Cavanagh, K., Clin. Psychol. Rev., 2015 (37), 1–12

Unmasking synthetic caramel: A fluorescent investigation into the sweet mystery of balsamic vinegar of Modena PGI

Monica Casale[1], Samuele Pellacani[2], Rasmus Bro[3], Beatriz Quintanilla Casas[3], Francesca Durante[1], Xhesmina Bregu[1], Franca Ladogana[4], Lorenzo Strani[2], Marina Cocchi[2], Caterina Durante[2]

[1] *Department of Pharmacy, University of Genoa, Viale Cembrano 4, Genova, 16149, Italy*
[2] *Department of Chemical and Geological Sciences, University of Modena and Reggio Emilia, via Campi 103, Modena, 41125, Italy*
[3] *Department of Food Science, University of Copenhagen, Department of Food Science, University of Copenhagen, Rolighedsvej 30, Frederiksberg C, DK-1958, Denmark*
[4] *Consorzio Tutela Aceto Balsamico di Modena, Via Ganaceto 113, Modena, 41121, Italy.*

Synthetic caramel is one of the most commonly used colorants in the food industry. In the case of Balsamic Vinegar of Modena PGI, its use is regulated by the Consortium of Balsamic Vinegar of Modena. The addition of caramel may be permitted for color stabilization but must be declared on the product label. Among the various coloring agents, the most frequently used is Class IV additive, E-150d, also known as sulphite ammonia caramel. However, caramel can also occur "naturally" in balsamic vinegar due to the cooked must used as a raw material in vinegar production. Currently, no official analytical method exists to differentiate between sulphite ammonia caramel and naturally occurring caramel in Balsamic Vinegar of Modena PGI.

To address this issue, an analytical approach based on Fluorescence Spectroscopy was employed. Vinegar samples were analysed both in their original state and after the addition of synthetic caramel. Fluorescence spectroscopy produced three-dimensional Excitation Emission Matrices (EEMs), which required appropriate data analysis. Specifically, Parallel Factor Analysis (PARAFAC) was applied to identify and quantify fluorophores associated with caramel.

Finally, the EEM data were used to quantify the presence of synthetic caramel in Balsamic Vinegar of Modena samples using PLS regression on the PARAFAC loadings. The results were satisfactory, confirming Fluorescence Spectroscopy as an effective analytical method for distinguishing sulphite ammonia caramel from naturally occurring caramel in Balsamic Vinegar of Modena PGI and for quantifying its presence.

References

[1] https://www.balsamicotradizionale.it/il-prodotto/disciplinare-di-produzione/
[2] Bro, R., PARAFAC. Tutorial and applications. Chemometrics and Intelligent Laboratory Systems, 1997, https://doi.org/10.1016/S0169-7439(97)00032-4

Unravelling the DoE in a multiway approach: GEMANOVA

Jokin Ezenarro[1], Daniel Schorn-García[2], Morten Rasmussen[3], Rasmus Bro[1], Olga Busto[4], Ricard Boqué[4]

[1] *Department of Food Science, Faculty of Science, University of Copenhagen, Denmark*
[2] *South African Grape and Wine Research Institute, Stellenbosch University, South Africa*
[3] *COPSAC, Copenhagen Prospective Studies on Asthma in Childhood, Herlev and Gentofte Hospital, University of Copenhagen, Denmark*
[4] *Universitat Rovira i Virgili. ChemoSens group. Department of Analytical Chemistry and Organic Chemistry, Spain.*
jokin@food.ku.dk

Design of Experiments (DoE)-based datasets often contain complex factor interactions that are not easy to capture or interpret using traditional additive models. While effective for simpler designs, these models fall short in representing the intricate, non-linear relationships that can be found in multivariate datasets. Generalised Multiplicative ANOVA (GEMANOVA) overcomes these issues by using a multiplicative framework to directly analyse high-order interactions, unlike additive methods such as ANOVA-Simultaneous Component Analysis (ASCA) and Parallel Factor Analysis with ASCA (PARAFASCA) [1, 2].
Despite its advantages, the use of GEMANOVA in chemometrics has been limited by a lack of standardised validation tools. Traditional techniques like cross-validation and permutation testing, which work well with additive models, do not directly apply due to the distinct mathematical structure of GEMANOVA [2]. This gap has made researchers hesitant to fully trust its results.
To address this, we propose a new validation method based on permutation testing. This approach systematically permutes individual factors, while maintaining the overall experimental structure, to generate a distribution of permuted loadings. Comparing these to the original model states the statistical significance of observed effects, thereby improving reliability and interpretability.
Furthermore, by quantifying the pseudovariances for each mode, we introduce a metric to assess the effect of experimental factors on the data; a critical aspect in DoE. GEMANOVA also inherits the ability of PARAFAC to efficiently handle missing data, making it adaptable to incomplete designs encountered in real-world experiments. Overall, our work enhances the usability of GEMANOVA and aligns its reliability with established methods like ASCA.

References

[1] Guisset, S., Martin, M., Govaerts, B., Chemom. Intell. Lab., 2019 (184), 44-63
[2] Bro, R., Jakobsen, M., J. Chemometr., 2002 (16), 294-304

Untargeted metabolomics and chemometric strategies for biomarkers discovery in bipolar disorder: insights from the BORDER project

Sara Tanilli[1], Alessandra Olarini[1], Giovanni Solarino[1], Marta Massano[2], Gabriele Lando[3], Marco Vincenti[1], Eugenio Alladio[1,2]
[1]*Dipartimento di Chimica, Università di Torino, Turin, Italy*
[2]*Centro Regionale Antidoping e di Tossicologia "A. Bertinaria", Orbassano (Turin), Italy*
[3]*Dipartimento di Scienze Chimiche, Biologiche, Farmaceutiche ed Ambientali, Università degli Studi di Messina, Italy*
alessandra.olarini@unito.it

Bipolar disorder (BD) is a complex and multifactorial psychiatric condition characterized by alternating episodes of mania and depression, with a severe impact on daily functioning [1]. Diagnosis currently relies on clinical assessment of symptoms and family history, but overlaps with other psychiatric disorders often cause misdiagnosis, impacting treatment strategies [2].

The BORDER project (*Bipolar disORDER: beyond psychology and towards multi-omics analysis*) aims to develop a minimally invasive diagnostic protocol for objective diagnosis and treatment selection. Blood and urine samples from 50 BD patients and 50 healthy controls (HC) were analyzed using Ultra-High-Performance Liquid Chromatography High-Resolution Mass Spectrometry (UHPLC-HRMS), generating high dimensional datasets. An untargeted metabolomics workflow was implemented to identify reliable BD biomarkers. Data preprocessing, independent statistical analysis, and chemometric techniques, including Principal Component Analysis (PCA) and Partial Least Squares-Discriminant Analysis (PLS-DA), revealed distinctive metabolic fingerprints associated with BD and confirmed their diagnostic potential. Candidate biomarkers were identified using Compound Discoverer® (v3.3, Thermo Fisher Scientific), integrated with multiple databases and libraries. Significant metabolites included Ala-Tyr dipeptide, γ-glutamyl cysteine, phenylacetyl glycine, xanthine, 1,3,7-trimethyluric acid, 3-methylxanthine, 8-hydroxy-methyl guanine, galactonic acid, offering promising targets for further validation.

The untargeted metabolomics pipeline, combined with unsupervised analysis and validated PLS-DA models, enabled the discovery of novel BD biomarkers, laying the groundwork for future clinical validation studies and advancing personalized medicine.

References

[1] Vieta, E., Berk, M., Schulze, T., et al., Nat. Rev. Dis. Prim., 2018 (4), 18008
[2] Grande, I., Berk, M., Birmaher, B., et al., Lancet, 2016 (387), 1561–1572

UV–Vis Diffuse Reflectance Spectroscopy as a Complementary Tool for Paper Based Analytical Devices (PAD)

Ainara Cirión[1], Ane Bordagaray[1], Miren Ostra[1],Rosa García-Arrona[1], Maider Vidal[1]
[1] *Department of Applied Chemistry, UPV/EHU, Paseo Manuel Lardizabal 3, 20018 San Sebastián, Spain*
ainara.cirion@ehu.eus

Ultraviolet–Visible Diffuse Reflectance Spectroscopy (UV–Vis DRS) is an analytical technique used to investigate the surface dispersion of a fraction of solid materials, particularly powders and surfaces. This technique relies on the Kubelka–Munk function to relate reflectance data to absorbance, enabling the qualitative and quantitative analysis [1].
One of the primary advantages of UV–Vis DRS is that it is non-destructive, requires minimal sample preparation, and allows rapid spectral acquisition. These features make UV–Vis DRS especially attractive for applications involving pigments, sensors, and functionalized surfaces.
In the present work, UV–Vis DRS is proposed as a complementary technique to traditional liquid-phase UV–Vis spectroscopy and paper-based analytical devices (PADs). By combining DRS with PADs, it becomes possible to perform in situ measurements using minimal amounts of reagents, since the reactions occur directly on the paper substrate. Furthermore, acquiring full spectra rather than single-point measurements (e.g. the signal of the RGB channels) enables multivariate analysis, offering a more comprehensive understanding of the chemical system under investigation.
The combination of UV–Vis DRS with paper-based platforms opens new ways for low-cost, portable, and efficient analytical tools. The multivariate data derived from these measurements made it a powerful tool for complex sample matrices and improving the reliability and interpretability of results.

Acknowledgements
Authors acknowledge financial support from Basque Government (Research Group IT1662/22 and Provincial Council of Gipuzkoa 2022-CIEN-000062-01).

References
[1] Torrent, J., Barron, V., Diffuse Reflectance Spectroscopy. *Mineralogical Methods*, SSSA Book Series, USA, 2008 (5), 367–385

LIST OF PARTICIPANTS

Surname	Name	Email	Institution
Abdelghafour	Florent	florent.abdelghafour@inrae.fr	INRAE
Abhishek	Abhishek	abhishek@nanogune.eu	CIC nanoGUNE
Albizu	Gorka	gorka.albizu@ehu.eus	University of the Basque Country
Barbera	Marcella	marcella.barbera@unipa.it	Università degli Studi di Palermo
Barron	Cécile	cecile.barron@inrae.fr	INRAE
Bastien	Philippe	pbastien@rd.loreal.com	L'Oréal
Berge	Marion	marion.berge@aphp.fr	Hôpitaux Universitaires Henri Mondor
Bertrand	Noémie	noemie.bertrand@loreal.com	L'Oréal
Biancolillo	Alessandra	alessandra.biancolillo@univaq.it	Università degli Studi dell'Aquila
Boccard	Julien	julien.boccard@unige.ch	Université de Genève
Bombarda	Isabelle	isabelle.bombarda@univ-amu.fr	IMBE
Bonnefoy	Christelle	christelle.bonnefoy@isa-lyon.fr	CNRS
Boqué	Ricard	ricard.boque@urv.cat	Universitat Rovira i Virgili
Borg	Antoine	antoine.borg@euroapi.com	EUROAPI
Brandolini-Bunlon	Marion	marion.brandolini-bunlon@inrae.fr	INRAE
Calvini	Rosalba	rosalba.calvini@unimore.it	Università degli Studi di Modena e Reggio Emilia
Camacho	José	josecamacho@ugr.es	Universidad de Granada
Caponigro	Vicky	vcaponigro@unisa.it	Università degli Studi di Salerno
Cardoso	Victor	victor.cardoso@food.ku.dk	University of Copenhagen
Cariou	Véronique	veronique.cariou@oniris-nantes.fr	ONIRIS
Casale	Monica	monica.casale@unige.it	Università degli Studi di Genova
Catelli	Emilio	emilio.catelli2@unibo.it	Università di Bologna
Caufourier	Cédric	noemie.retout@servier.com	Oril Industrie
Cavallini	Nicola	nicola.cavallini@polito.it	Politecnico di Torino
Cayuela-Sánchez	José Antonio	jacayuela@ig.csic.es	IG-CSIC
Cerruti	Simone	simone.cerruti@uniupo.it	Università degli Studi del Piemonte Orientale
Charloto	Carl	carl.charloto@orange.fr	Pellenc ST
Claeys-Bruno	Magalie	m.claeys-bruno@univ-amu.fr	Aix Marseille Université
Clément	Yohann	yohann.clement@univ-lyon1.fr	Université Claude Bernard Lyon 1
Clément	Jean-Emmanuel	jean-emmanuel.clement@univ-lille.fr	Université de Lille
Cocchi	Marina	marina.cocchi@unimore.it	Università degli Studi di Modena e Reggio Emilia
Cruz Muñoz	Emmanuel	enmanuel.cruzmunoz@unimib.it	Università degli Studi di Milano-Bicocca
D'Alessandro	Alessandro	alessandro.dalessandro@barilla.com	Barilla G. e R. F.lli
Da Silva	Célie	celie.dasilva@unige.ch	Université de Genève
de Juan	Anna	anna.dejuan@ub.edu	Universitat de Barcelona
Denimal	Emmanuel	emmanuel.denimal@agrosupdijon.fr	Institut Agro Dijon
De Oliveira	Samy	samy.de-oliveira@agrosupdijon.fr	Institut Agro Dijon
Dermitzakis	Nicolás	ndermitzakis@fhrt.com	Universitat Autònoma de Barcelona
Dowek	Antoine	antoine.dowek@gmail.com	Hôpitaux Universitaires Henri Mondor
Dufour	Olivia	olivia.dufour@loreal.com	L'Oréal
Dujourdy	Laurence	laurence.dujourdy@agrosupdijon.fr	Institut Agro Dijon
Dumas	Elise	elise.dumas@epfl.ch	École Polytechnique Fédérale de Lausanne
Duponchel	Ludovic	ludovic.duponchel@univ-lille.fr	Université de Lille
Ezenarro	Jokin	jokin@food.ku.dk	University of Copenhagen
Facco	Pierantonio	pierantonio.facco@unipd.it	Università degli Studi di Padova
Felizzato	Giorgio	giorgio.felizzato@unito.it	Università degli Studi di Torino
Ferrari	Veronica	veronica.ferrari@unimore.it	Università degli Studi di Modena e Reggio Emilia

Ferrer	Alberto	aferrer@eio.upv.es	Universitat Politècnica de València
Gariglio	Sara	sara.gariglio@edu.unige.it	Università degli Studi di Genova
Giglio	Cannon	gigliocj@ornl.gov	Oak Ridge National Laboratory
Ginebreda	Antoni	agmqam@cid.csic.es	IDAEA-CSIC
Gomez-Sanchez	Adrian	gomez.sanchez.adr@gmail.com	Université de Lille
Gorla	Giulia	giulia.gorla@ehu.eus	University of the Basque Country
Guerrini	Ruggero	ruggero.guerrini.etu@univ-lille.fr	Université de Lille
Jaillais	Benoît	benoit.jaillais@inrae.fr	INRAE
Jourdan	Astrid	astrid.jourdan@univ-pau.fr	Université de Pau et des Pays de l'Adour
Kamp	Christel	christel.kamp@pei.de	Paul-Ehrlich-Institut
Kucheryavskiy	Sergey	svk@bio.aau.dk	Aalborg University
Lacoue-Negre	Marion	marion.lacoue-negre@ifpen.fr	IFPEN
Lanteri	Pierre	pierre.lanteri@univ-lyon1.fr	Université Claude Bernard Lyon 1
Lopez	Eneko	e.lopez@nanogune.eu	CIC nanoGUNE
Malegori	Cristina	malegori@difar.unige.it	Università degli Studi di Genova
Malinconico	Antonio	antonio.malinconico@unina.it	Università degli Studi di Napoli "Federico II"
Marini	Federico	federico.marini@uniroma1.it	Università degli Studi di Roma "La Sapienza"
Martin Santiago	Carlos	c.martinsantiago@campus.unimib.it	Università degli Studi di Milano-Bicocca
Mas-García	Silvia	silvia.mas-garcia@inrae.fr	INRAE
Mazzoleni	Alberto	alberto.mazzoleni@unito.it	Università degli Studi di Torino
Mezzatesta	Elodie	elodie.mezzatesta@univ-amu.fr	Aix Marseille Université
Millan	Fabien	fabien.millan@hevs.ch	HES-SO Valais-Wallis
Muñoz de la Peña	Arsenio	amunozdelapena@gmail.com	Universidad de Extremadura
Nhek	Sarin	sarin.nhek@gmail.com	National Institute of Optics (CNR-INO)
Nikzad-Langerodi	Ramin	ramin.nikzad-langerodi@scch.at	Software Competence Center Hagenberg
Olarini	Alessandra	alessandra.olarini@unito.it	Università degli Studi di Torino
Orecchio	Ciro	ciro.orecchio@unito.it	Università degli Studi di Torino
Ortiz	María Cruz	mcortiz@ubu.es	Universidad de Burgos
Ostra-Beldarrain	Miren	miren.ostra@ehu.eus	University of the Basque Country
Paul	David	davidolivier.paul@ividata.com	Ividata Life Sciences
Pellacani	Samuele	samuele.pellacani@unimore.it	Università degli Studi di Modena e Reggio Emilia
Portaluri	Vincent	vincent.portaluri@gmail.com	Servier
Quintanilla-Casas	Beatriz	beatriz@food.ku.dk	University of Copenhagen
Roca Nasser	Esteban	erocan@ugr.es	Universidad de Granada
Rocha de Oliveira	Rodrigo	rodrigo.rocha@ub.edu	Universitat de Barcelona
Roger	Jean-Michel	jean-michel.roger@irstea.fr	INRAE
Roussel	Sylvie	sroussel@ondalys.fr	Ondalys
Rubini	Morandise	morandise.rubini@univ-tours.fr	CNRS
Ruckebusch	Cyril	cyril.ruckebusch@univ-lille1.fr	Université de Lille
Rudaz	Serge	serge.rudaz@unige.ch	Université de Genève
Ruisánchez Capelastegui	Itziar	itziar.ruisanchez@urv.cat	Universitat Rovira i Virgili
Saccenti	Edoardo	edoardo.saccenti@wur.nl	Wageningen University & Research
Santarelli	Elisabetta	esantarelli@unisa.it	Università degli Studi di Salerno
Sarabia	Luis Antonio	lsarabia@ubu.es	Universidad de Burgos
Savorani	Francesco	francesco.savorani@polito.it	Politecnico di Torino
Scotté	Camille	camille.scotte@inrae.fr	INRAE
Seifert	Andreas	a.seifert@nanogune.eu	CIC nanoGUNE
Sergent	Michelle	michelle.sergent@univ-amu.fr	Aix Marseille Université
Sicre-Conesa	Albert	asicrec@ub.edu	Universitat de Barcelona
Slanina	Teresa	teresa.slanina@pei.de	Paul-Ehrlich-Institut

Solarino	Giovanni	giovanni.solarino@unito.it	Università degli Studi di Torino
Sorochan Armstrong	Michael	mdarmstr@ualberta.ca	Universidad de Granada
Sozzi	Mattia	mattia.sozzi@polito.it	Politecnico di Torino
Strani	Lorenzo	lostrani@unimore.it	Università degli Studi di Modena e Reggio Emilia
Tanzilli	Daniele	daniele.tanzilli@inrae.fr	INRAE
Tauler	Romà	roma.tauler@idaea.csic.es	IDAEA-CSIC
Vera	Núria	nuria.verai@urv.cat	Universitat Rovira i Virgili
Vitale	Raffaele	raffaele.vitale@univ-lille.fr	Université de Lille
Viton	Axel	axel.viton@etu.univ-amu.fr	Aix Marseille Université
Voccio	Riccardo	riccardo.voccio@edu.unige.it	Università degli Studi di Genova