WSC 13
Moscow, Russia

The 13th Winter Symposium on Chemometrics

Modern Methods of Data Analysis

Russia, Moscow,
February 28 – March 4, 2022
wsc.chemometrics.ru/wsc13

ABSTRACT BOOK

Organized by:
Russian Chemometrics Society
Moscow State University

ORGANIZED BY

RCS

Russian Chemometrics Society

Lomonosov Moscow State University
DEPARTMENT
OF CHEMISTRY
1929

Moscow State University

PARTNERS

HORIBA
Explore the future

TECHNICAL SUPPORT

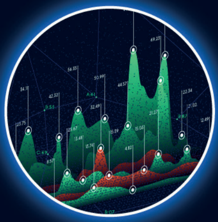
mesol

PROFESSIONAL CONGRESS ORGANISER
WWW.MESOL.RU

WSC 13 2022

Moscow, Russia

February
28 – March 2, 2022

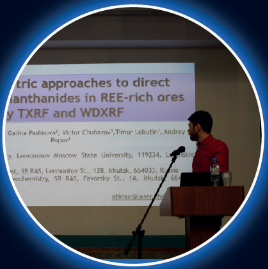


Russian Chemometrics Society presents the 13th Winter Symposium on Chemometrics, taking place February 28 – March 4, 2022. The Winter Symposium on Chemometrics (WSC) has been started as an international scientific event in a framework of Drushbametrics project by a small group of enthusiasts from Russia and Scandinavia to support the rapid development of multivariate data analysis and its application to analytical chemistry. The conference brings together both academic scientists and industrial experts, providing many opportunities for personal contacts and fruitful discussions.

The symposium schedule includes plenary lectures (40 min), oral presentations (20 min) and flash presentations (8 min). Selected contributions will be published in a special issue of the Journal of Chemometrics.

Pre-conference short courses on advanced topics in chemometrics (in English) will be held on Monday just before the symposium.

In view of the new wave of Covid-19, when many participants cannot travel abroad, the Organizing Committee has decided to hold the Symposium in online-only meeting. And we hope that we will have an opportunity to meet again at offline event in 2023.



LOCAL ORGANIZING COMMITTEE:

Timur Labutin, Chair (MSU)

Nikolai Sushkov, Secretary (MSU)

Andrey Samokhin (MSU)

PROGRAMME COMMITTEE:

Kim Esbensen (GEUS)

Dmitry Kirsanov (SPbSU)

Timur Labutin (MSU)

Paolo Oliveri (University of Genova)

Alexey Pomerantsev (ICP RAS)

Oxana Rodionova (ICP RAS)

THE 13TH WINTER SYMPOSIUM ON CHEMOMETRICS (ONLINE EVENT 2022)*

Monday, 28 February		Tuesday, 1 March		Wednesday, 2 March	
Time	Event	Time	Event	Time	Event
9:25-11:30	Online school (Krylov I.)	8:55-9:40	Invited talk (Parastar H.)	9:00-9:40	Invited talk (Madzhidov T.)
		9:40-10:05	Giebelhaus R.	9:40-10:00	Vladimirova N.
		10:05-10:25	Surkova A.	10:00-10:20	Viegas I.
		10:25-10:45	Zontov Yu.	10:20-10:40	Ferubko A.
		10:45-11:05	Kalambet Yu.	10:40-11:00	Guliev R.
		11:05-11:13	Bratchenko L.	11:00-11:08	Ikhalaynen Yu.
		11:13-11:21	Armstrong M.	11:08-11:16	Selivanovs Z.
		11:21-11:29	Silaev D.	11:16-11:24	Krylov I.
11:30-12:00	Coffee break	11:29-11:50	Coffee break	11:24-11:50	Coffee break
12:00-14:00	Online school (Monakhova Yu.)	11:50-12:30	Invited talk (Monakhova Yu.)	11:50-12:30	Invited talk (Oliveri P.)
		12:30-12:50	Anisimov D.	12:30-12:50	Krylov I.
		12:50-13:10	Marchesi C.	12:50-13:10	Lindner S.
		13:10-13:30	Sushkov N.	13:10-13:30	Skvortsov A.
		13:30-13:50	Trukhin I.	13:30-13:50	Zagitova L.
		13:50-14:10	Shuba A.	13:50-13:58	Yuskina Ye.
		14:10-14:18	Skorobogatov Ye.	13:58-14:06	Yarkaeva Yu.
		14:18-14:26	Khairullina V.		
14:00-15:00	Lunch	14:26-15:30	Lunch	14:06-15:30	Lunch
15:00-16:30	Online school (Kirsanov D.)	15:30-16:10	Invited talk (Bro R.)	15:30-16:10	Invited talk (Mishra P.)
		16:10-16:30	Bogomolov A.	16:10-16:30	Aidene S.
		16:30-16:50	Turova P.	16:30-16:50	Akhmetzhanov T.
		16:50-17:10	Cruz Sanchez J.	16:50-17:10	Malegori C.
		17:10-17:30	Burger R.	17:10-17:30	Kravich N.
		17:30-17:38	Shik A.	17:30-17:38	Yevseeva A.
		17:38-17:46	Plyushchenko I.	17:38-17:46	Nazyrov M.
				17:46-17:54	Matveeva I.

🟩 invited talk

🟦 oral talk

🟧 flash talk

*All times are in Moscow time (MSK = GMT+3)

The 13th Winter Symposium on Chemometrics (online event 2022)

Symposium schedule

All times are in Moscow time (MSK = GMT+3)

Monday, 28 February:
online school on chemometrics

Time		Speaker	Talk title
9:25	9:30	Timur Labutin	Welcome address
9:30	11:30	Ivan Krylov	Basics of PARAFAC decomposition in fluorescence excitation-emission spectroscopy
11:30	12:00	*Coffee break*	
12:00	14:00	Yulia Monakhova	Independent components analysis (ICA) in practice: tips and tricks
14:00	15:00	*Lunch*	
15:00	16:30	Dmitry Kirsanov	Multisensor systems' data processing

Tuesday, 1 March: online symposium, day 1

Time		Speaker	Talk title
8:55	9:00	Timur Labutin	Welcome address
9:00	9:40	Hadi Parastar	There is plenty of gloom in quantitative mass spectrometry imaging: can deep learning help?
9:40	10:05	Ryland Giebelhaus	Region of Interest Selection for GC-MS Data with a Pseudo Fisher Ratio Moving Window
10:05	10:25	Anastasia Surkova	Aquaphotomics approach for cancer diagnosis and monitoring
10:25	10:45	Yury Zontov	Study of the applicability of modern machine-learning techniques to the task of similarity search of objects represented by series of images and metadata
10:45	11:05	Yuri Kalambet	Peak modelling using exponentially modified functions
11:05	11:13	Lyudmila Bratchenko	Recognition and classification of skin Raman spectra using convolutional neural networks analysis
11:13	11:21	Michael Armstrong	PARAFAC2×2: a PARAFAC Model for Drift in Two Modes
11:21	11:29	Dmitry Silaev	Spectrophotometric determination of active substances in ophthalmic formulation using chemometrics
11:29	11:50	*Coffee break*	
11:50	12:30	Yulia Monakhova	Independent components analysis (ICA) at the “cocktail-party” in analytical chemistry
12:30	12:50	Jordi Cruz Sanchez	Quantitative determination of curcuminoids in turmeric powder in just 5 seconds
12:50	13:10	Claudio Marchesi	MicroNIR and chemometrics for microplastics pollution
13:10	13:30	Nikolai Sushkov	Compositional features of zooplankton species as studied by laser-based analytical techniques
13:30	13:50	Ivan Trukhin	Application of cluster analysis to identify the breakthrough of injected water in oil wells

Tuesday, 1 March:
online symposium, day 1

13:50	14:10	Anastasia Shuba	Primary and secondary data fusion from a sensor array for prediction of qualitative and quantitative indicators
14:10	14:18	Yevgeni Skorobogatov	Discrimination of plant samples using carbocyanine fluorophore additives
14:18	14:26	Veronika Khairullina	Prospects for the search for cyclooxygenase-2 inhibitors among diarylheterocycle derivatives
14:26	15:30	*Lunch*	
15:30	16:10	Rasmus Bro	Using chemical insight and AI to automate untargeted GC-MS profiling
16:10	16:30	Andrey Bogomolov	Development of a Multispectral Fiber Probe
16:30	16:50	Polina Turova	Various machine learning methods in HPLC-MS datasets treatment
16:50	17:10	Daniil Anisimov	Electronic nose based on Organic Field Effect Transistors for food quality monitoring
17:10	17:30	Rene Burger	Benchtop NMR versus High Field NMR: Comparison of Chemometric Molecular Weight Analysis of Lignin
17:30	17:38	Anna Shik	The use of artificial neural networks for the simultaneous determination of several analytes by the oxidation reactions of carbocyanine fluorophores
17:38	17:46	Ivan Plyushchenko	Application of gradient boosting machine for signal processing in LC-MS metabolomics

Wednesday, 2 March:
online symposium, day 2

Time		Speaker	Talk title
9:00	9:40	Timur Madzhidov	Chemoinformatics and machine learning in synthetic chemistry: from data to models, from models to robots
9:40	10:00	Nadezhda Vladimirova	Prediction of potentiometric selectivity for carbonate-selective plasticized membrane sensors using QSPR modelling
10:00	10:20	Isabelle Viegas	Advanced coupled factorization of multi-way fluorescence data to assess the conjugation of quantum dots to proteins
10:20	10:40	Andrey Ferubko	Fluorescent determination of the total content of humic substances in waters using multidimensional calibration models
10:40	11:00	Rustam Guliev	Generalization and geometrical interpretation of N-FINDR algorithms for unmixing hyperspectral data
11:00	11:08	Yuri Ikhalaynen	Untargeted metabolomics study of *Humulus lupulus* brewing cultivars, for genetic origin classification task
11:08	11:16	Zahars Selivanovs	Nonlinear dimensionality reduction methods for multisensor system data analysis
11:16	11:24	Ivan Krylov	Stochastic optimisation of spectra of laser-induced plasma for plasma composition calculation
11:24	11:50	*Coffee break*	
11:50	12:30	Paolo Oliveri	Is signal pre-processing a trivial step? A focus on pitfalls and challenges
12:30	12:50	Ivan Krylov	Modelling of scattering signal for direct PARAFAC decompositions of excitation-emission matrices
12:50	13:10	Simon Lindner	Is the Calibration Transfer of Multivariate Calibration Models Between High- and Low-Field NMR Instruments Possible?
13:10	13:30	Alexey Skvortsov	Grey modelling of absorption spectra of silver nanoisland films
13:30	13:50	Liana Zagitova	Voltammetric sensors in the analysis of enantiomeric mixture using PLS
13:50	13:58	Yekaterina Yuskina	Chemometric processing of the data from high-frequency inductor for chemical sensing

Wednesday, 2 March: online symposium, day 2			
13:58	14:06	Yulia Yarkaeva	Chemometric methods in the recognition of antibiotics using voltammetric sensory systems based on molecularly imprinted polymers
14:06	15:30	*Lunch*	
15:30	16:10	Puneet Mishra	Deep learning for spectral data modelling in Chemometrics: the benefits and the hypes
16:10	16:30	Soraya Aidene	Correcting sample matrix effects in XRF data using chemometrics
16:30	16:50	Timur Akhmetzhanov	Direct analysis of lanthanides in REE-rich ores by Laser-Induced Breakdown Spectrometry coupled with Chemometrics
16:50	17:10	Cristina Malegori	A moving-block-PCA based approach for real time monitoring of a powder blending process using a miniaturized near infrared sensor
17:10	17:30	Nadan Kravich	Improving precision of potentiometric multisensor analysis using nonlinear regression methods: spent nuclear fuel reprocessing case study
17:30	17:38	Alexandra Yevseeva	Handling of optical multisensor data
17:38	17:46	Marat Nazyrov	Sensor system based on Cu(II) and Zn(II) amino acid complexes for recognition of atenolol enantiomers in racemic mixture using the partial least squares
17:46	17:54	Irina Matveeva	Decomposition of in vivo skin Raman spectra using multivariate curve resolution method
17:54	18:00	Timur Labutin	Closing remarks

Short courses at WSC-13

Prof. Yulia Monakhova.
Independent components analysis (ICA) in practice: tips and tricks

Independent components analysis (ICA) is a probabilistic method, whose goal is to extract underlying component signals and their concentration ratios, that are maximally independent and non-Gaussian, from mixed observed signals. In this short course, theoretical background of ICA algorithms will first be explained. After that, several practical problems of using different ICA approaches (MILCA, SNICA, JADE, RADICAL, FastICA) will be solved in the Matlab environment. The methods will be applied to quantitative and qualitative analysis of non-trivial mixture types as well as to exploratory analysis using spectroscopic techniques (UV-VIS, NMR, IR). Special attention will be paid on the influence of spectra acquisition parameters, spectra preprocessing, decomposition parameters and the interpretation of ICA outcome. Approaches to absolute analyte quantification will be discussed. The differences to common chemometric approaches – PCA and MCR-ALS – will be highlighted.

Prof. Dmitry Kirsanov.
Multisensor systems' data processing

Multisensor systems for chemical analysis is a rapidly developing field of research in analytical chemistry. These systems comprise several (typically from 3 to 30) chemical sensors that do not show a sharp selectivity towards a particular substance, but rather have so called cross-sensitivity - a response towards a variety of analytes. In multicomponent samples these multisensor systems produce complex non-resolved analytical signals containing information about several analytes simultaneously. These signals can be effectively treated with chemometric tools to derive qualitative and quantitative information about the samples under study. The lecture will provide the basics of multisensor system for chemical analysis and will demonstrate several examples of multisensor data processing from real experiments using R.

Ivan Krylov.

Basics of PARAFAC decomposition in fluorescence excitation-emission spectroscopy

The seminar covers the following topics:

- Theoretical basis for the use of canonical tensor decomposition (PARAFAC) in excitation-emission fluorescence spectroscopy
- Deviations of excitation-emission data from the assumptions of PARAFAC and how to counter them with pre-processing: inner filter effects and the scattering signal
- PARAFAC uniqueness and validation of the model: split-half analysis and jack-knifing
- Performing the PARAFAC analysis of the data using the 'albatross' package in the R programming language

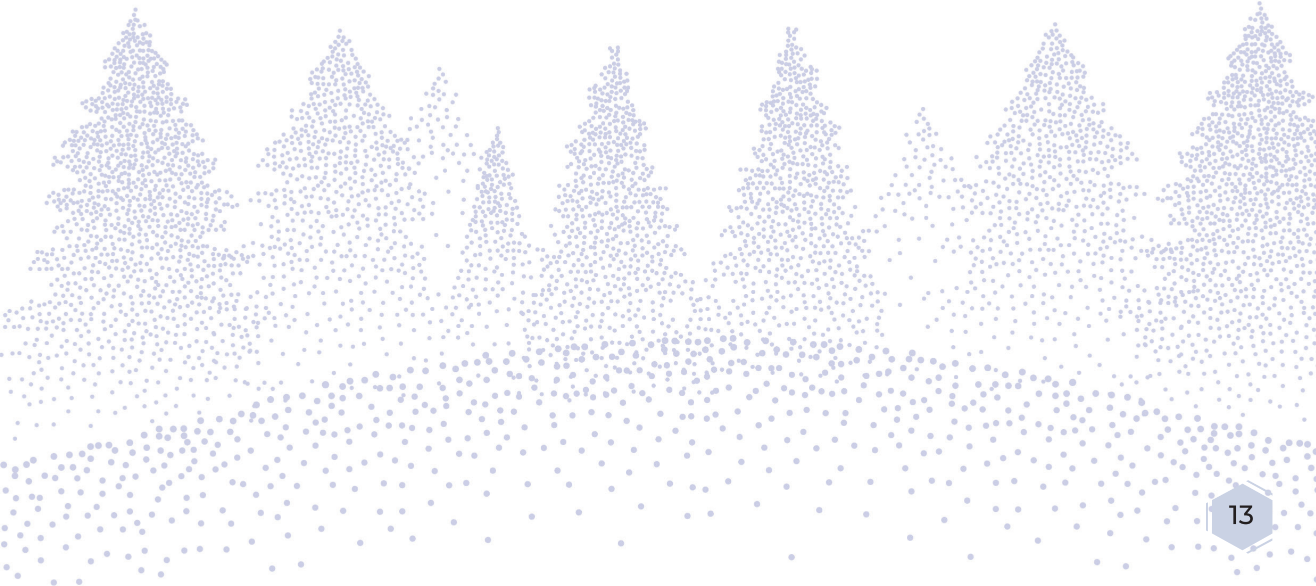
Recommended prerequisites: experience in the R programming language or related environments

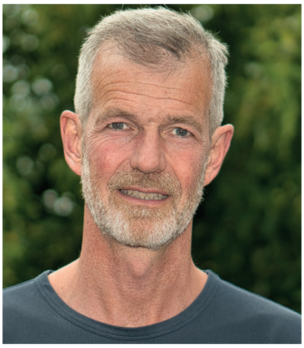
Software:

- the R programming language <https://r-project.org>
- the "albatross" package <https://CRAN.R-project.org/package=albatross>
- RStudio <https://www.rstudio.com/> or another graphical interface is not required but may be convenient

Please find details and updates at https://files.libs.chem.msu.ru/~ivan/WSC-13/ .

INVITED TALKS





Rasmus Bro

Professor, University of Copenhagen, Denmark
rb@food.ku.dk

Using chemical insight and AI to automate untargeted GC-MS profiling

Modern GC-MS systems combined with efficient sampling techniques produce chromatograms with a large number of peaks of which many are not well-resolved. Turning such data into chemical information (concentrations and chemical identities) is known to be time-consuming, often user dependent and with a number of problems e.g. in handling low signal to noise, severe coelution etc. We show that using the so-called PARAFAC2 tensor modelling we can provide a solution that is user-friendly, time-saving, and produces reliable results that are not user-dependent.

Information on the speaker

Prof. Rasmus Bro (b. 1965) is a world-recognized chemometrician, performing research on most aspects of chemometrics and in particular on multi-way analysis both from a theoretical and a practical point of view. He is heading an industrial research consortium, ODIN, focusing on Process Analytical Technology (PAT) and has also started a new master of science in the same area. He has been an editor of Journal of Chemometrics for many years and is the author of a number of MATLAB toolboxes that are made available at www.models.life.ku.dk. Primary fields of research: chemometrics, data mining, multivariate calibration, classification, multi-way analysis, exploratory analysis, experimental design, fluorescence spectroscopy, numerical analysis, spectroscopy metabolomics, process analytical technology.

Yulia Monakhova

Professor, University of Applied Sciences Aachen, Germany
Professor in Analytical chemistry, Institute of Chemistry, Saratov State University, Saratov, Russia
yul-monakhova@mail.ru

Independent components analysis (ICA) at the “cocktail-party” in analytical chemistry

Independent components analysis (ICA) is a probabilistic method, whose goal is to extract underlying component signals, that are maximally independent and non-Gaussian, from mixed observed signals. Since the data acquired in many applications in analytical chemistry are mixtures of component signals, such a method is of great interest. In this presentation recent ICA applications for quantitative and qualitative analysis in analytical chemistry will be reviewed. The following experimental techniques will be covered: fluorescence, UV-VIS, NMR, vibrational spectroscopies as well as chromatographic profiles. Furthermore, applications of ICA as a preprocessing tool as well as existing hybrid ICA-based multivariate approaches will be outlined. Finally, further research directions will be proposed. It will be concluded that ICA is starting to play an important role in analytical chemistry, and this will definitely increase in the future.

Information on the speaker

Dr. habil. Yulia Monakhova's (b. 1986) research interests are related to chemometrics and include: development of the methodology for chemometric modeling of one- and multi-dimensional spectroscopic data (NMR, UV, IR, fluorescence); theoretical improvement of machine learning methods to be used in analytical chemistry; establishment of analytical methods based on multivariate modeling for quantitative analysis of complex pharmaceutical, food and environmental matrices; usage of discrimination and data fusion methods to develop methods for verifying authenticity of complex matrices (provenience, potency, labeling validation). Her studies are currently oriented to the development of independent component analysis (ICA). Dr. Monakhova is a member of the chemometric committee at the Analytical Chemistry Council of Russian Academy of Sciences and a reviewer for the leading international journals in analytical chemistry and chemometrics.

Paolo Oliveri

Assistant Professor, University of Genova, Italy
oliveri@difar.unige.it

Is signal pre-processing a trivial step? A focus on pitfalls and challenges

Mathematical pre-processing includes a wide number of transformations generally aimed at minimising the unwanted variations that may affect analytical signals, with the result of improving data quality and, consequently, the conversion of data to valuable information. In addition to the desired corrections, application of mathematical transforms may produce undesired secondary effects. In particular, some transforms may introduce artefacts. Other may complicate interpretation of the final results of signal processing – a risk that is often underestimated.

The present study is aimed to analyse and critically discuss the consequences of row pre-processing (conversion of measurement units, derivatives, and standard normal variate transform) on the evaluation of final outcomes of chemometric data analysis. An in-depth focus on pre-processing effects both on the signal shape and on misinterpretation of results – a crucial and disregarded issue in the analytical field – will be presented. It will be shown how this preliminary step of data processing may lead, in many cases, to draw incongruous conclusions, not actually based on real information embodied within data, but on artefacts arising from the mathematical transforms. Key aspects related to the sequential order of pre-processing transforms – when they are applied in combination – will be also examined. The overview will not be limited to a description of the problem, it will also introduce strategies and tools for overcoming such unwanted effects, allowing a direct interpretation of the importance of original variables to be performed, explaining the chemical information that actually characterises samples. The dangerous implications of row pre-processing on instrumental signals will be demonstrated on real datasets coming from different analytical techniques applied to real case studies. Hence, the impact of this widespread problem in most of the branches of applied analytical chemistry will be illustrated.

Information on the speaker

Dr. Paolo Oliveri (b. 1982) works in the field of analytical spectroscopy and hyperspectral imaging as well as on their applications in food analysis, forensic science and cultural heritage studies. Chemometrics is what brings all this together. Dr. Oliveri's most cited works are related to classification topics, including the use of UNEQ, SIMCA, POTFUN and MRM. Dr. Oliveri is a member of editorial boards in several journals, including the Chemometrics and Intelligent Laboratory Systems, and a peer reviewer for 22 scientific journals.

Hadi Parastar

Associate Professor, Sharif University of Technology, Iran
h.parastar@sharif.edu

There is plenty of gloom in quantitative mass spectrometry imaging: can deep learning help?

Hyperspectral imaging (HSI) is an emerging technique that combines regular imaging with spectroscopic analysis methods. This analytical tool provides three-dimensional data, which represents a spectrum for each pixel on the sample surface. Information about chemical composition of samples can be obtained using spectral dimension and spatial dimension is utilized to map their distribution along sample. Mass spectrometry imaging (MSI) is a relevant subgroup of HSI techniques, including almost high-resolution mass spectrum beside the distribution map. MSI provides label-free detection and localization of various (bio)chemical compounds present in various types of tissues.

Despite the widespread application of MSI in various fields, the interpretation of the data obtained from this method faces significant challenges because of the high-resolution mass spectrometers that generate huge and complex data. Besides the challenges in digital processing and data storage in the big data era, the main concern is handling these large and complex data, especially in quantitative approaches where even very small signal changes affect the final measurements.

In recent years, deep learning (DL) methods as a sub-class of machine learning (ML) have received much attention due to their unique capabilities in solving various complex problems in different areas from voice recognition and image analysis to drug discovery and omics. Also, many efforts have been made to overcome the challenges in chemistry and especially in analytical chemistry.

Although DL is known as a powerful method for modeling complex data, no study has been yet conducted in using DL for MSI with a supervised quantitative approach because the main limitation for these studies is lack of sufficient labeled training data. In this presentation, for the first time, we aim to investigate the potential of DL methods based on DL algorithms (convolutional neural network, CNN, and autoencoder, AE) to interpret the big data received from mass spectrometry images for quantitative analysis of chlordecone known as a carcinogenic pesticide and map its accumulation pattern in mouse liver. To provide a big enough data we have used pixels of mass spectrometry images as calibration data and we have employed multivariate curve resolution- alternating least squares (MCR-ALS) method for the first time as the labeling approach to provide sufficient calibration labeled data for optimal training of a CNN model.

The results of this study demonstrate the power of DL methods for modeling MALDI-MSI data for quantitative analysis, minimizing the effects of tissue heterogeneity on calibration and unknown data, and predicting unknown values of chlordecone in mouse liver.

Information on the speaker

Dr. Hadi Parastar (b. 1983) works at the development of different branches of chemometrics. His latest works are related to convolutional neural networks (CNN), support vector machine (SVM) and independent component analysis (ICA). Studies on multivariate curve resolution (MCR) are also notable. Other interests include mathematical chromatography, multiway analysis methods such as PARAFAC and investigations in the field of food authenticity and adulteration. The material for the studies is usually obtained by techniques like GC-MS, NIR and UV-Vis spectroscopy. Dr. Parastar is involved in many national and international collaborations and is distinguished with a number of Iranian and international awards.

Timur Madzhidov

Associate Professor, Volga Region Federal University, Russia
tmadzhidov@gmail.com

Chemoinformatics and machine learning in synthetic chemistry: from data to models, from models to robots

The productivity of organic synthetic chemistry has greatly increased recently due to specific technologies of analysis, high-throughput experimentation, and parallel synthesis. The biggest database contains more than 100 mln data on chemical reactions. On the other side, it requires that a synthetic chemist has to be very productive and efficient, especially in industrial settings. He has to develop a robust and efficient pathway to the desired compound, select optimal conditions, foresee and avoid possible selectivity issues. At the same time, none of the chemists can follow all of the knowledge generated in the field. Thus, big data analytics and machine learning techniques (including artificial intelligence approaches) attract much attention [1].

We review some most important and interesting applications of machine learning in synthetic chemistry. We will show how text mining technologies can be applied to extract knowledge about chemical reactions from chemical experiments description. Also, the workflow of structure-activity modeling in the field of chemoinformatics will be described. We will show how chemical reactions can be represented for machine learning techniques application. We will give an overview of how the latter is applied for assessing reaction kinetics, thermodynamics, and optimal reaction condition.

We will give an overview of recent progress in the generation of a synthetic plan leading to the desired compound. Approaches for synthetic *de novo* design (e.g. design of compound with desired property) will be specially described. We will show how the new chemical reactions can be proposed using artificial intelligence approaches.

And finally, we will give a brief review of how the models can be coupled with computer-governed chemical tools. The first examples of such chemical robots will be described.

References

1. [1] Baskin I.I., Madzhidov T.I., Antipin I.S., Varnek A. *Russ. Chem. Rev.* 2017, *86*, 1127.
2. Acknowledgements: This work was supported by the Russian Science Foundation (project No 19-73-10137).

Information on the speaker

Dr. Timur Madzhidov (b. 1984) is a group leader in the Lab of Chemoinformatics and Molecular Modeling in Kazan Federal University, Russia. His team is working on the application of machine learning, data mining and AI for reaction informatics. They have developed tools for reaction data curation, manipulation, atom-to-atom mapping fixing, reaction property (rate, yield) prediction, reaction discovery, condition assessment, etc. The team also works at the development of novel QSAR and virtual *screening approaches based on multi-instance learning and pharmacophore representation of molecules, and also involved into collaboration with big industrial companies.*

Dr. Madzhidov is a founder, lecturer (as associate professor), and supervisor of the Master Program in Chemoinformatics and Molecular Modeling of Kazan Federal University, the first master program of this kind in Russia, which is also running as a double-diploma program with the University of Strasbourg, France.

Puneet Mishra

Scientist, Wageningen University and Research,
The Netherlands
puneet.mishra@wur.nl

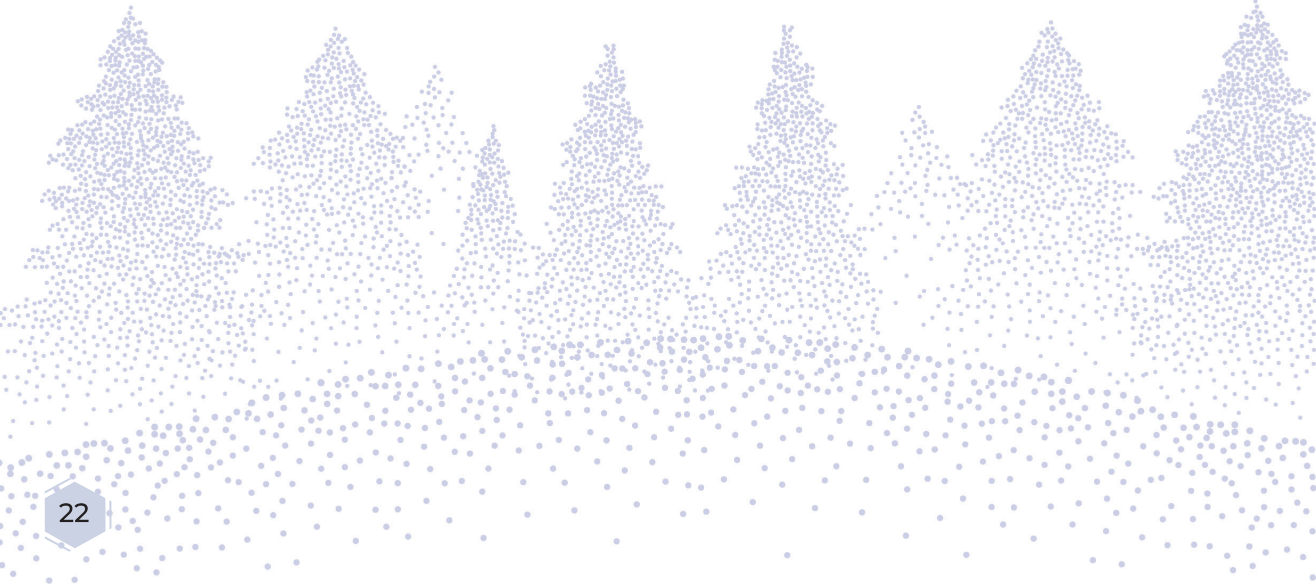
Deep learning for spectral data modelling in Chemometrics: the benefits and the hypes

Deep learning is emerging as a potential new direction for data modelling. Its applications can be found in almost all scientific fields ranging from agriculture to medicine to astrophysics. In the domain of chemometrics as well an increasing trend towards deep learning is emerging, where particularly for spectral data modelling, deep learning has shown to outperform several state-of-the-art chemometric approaches. Furthermore, for spectral image processing as well, the advanced concept of deep learning such as convolution modelling, semantic segmentation and generative adversarial networks have shown efficient modelling by combining both the spatial and spectral information. On one hand where the deep learning is emerging as a powerful tool, however, on the other hand, its applications are over-hyped in scientific chemometric literature, where usually the comparison of deep learning is performed with simple linear models and deep learning is demonstrated as a winner. This talk aims to bring the attention of the audience towards the potential and hypes of deep learning for spectral data modelling. The key idea is that once the audience is aware of the do and don’ts of deep learning, then they can use deep learning as a complementary tool to chemometrics and without competing.

Information on the speaker

Dr. Puneet Mishra (b. 1991) is a passionate expert in sensing technologies and chemometrics for non-destructive analysis of material properties. He has multidisciplinary educational background in international institutions of Spain, Belgium, UK and the Netherlands. Particularly contributed and published various research articles dealing with applications of spectroscopy and development of new chemometrics approaches to process spectral data. His studies deal with fluorescence, near-infrared and optical spectroscopy, spectral imaging, multisensory technologies, neural networks and artificial intelligence, machine vision, multi-block analysis techniques, food quality and related topics. Dr. Mishra is a reviewer for many high-ranked international journals. He is also an author of three software packages for data analysis (FRUITNIR-GUI, MBA-GUI, and CT-GUI).

ORAL TALKS



Correcting sample matrix effects in XRF data using chemometrics

Soraya Aidene*[1], Maria Khaydukova[1,2], Valentin Semenov[1,3], Dmitry Kirsanov[1], Vitaly Panchuk[1,3]

[1] *Institute of Chemistry, St. Petersburg State University,*
St. Petersburg, Russian Federation
[2] *World-Class Research Centre "Centre for Personalized Medicine",*
FSBSI IEM, Saint-Petersburg, Russia
[3] *Institute for Analytical Instrumentation RAS, St. Petersburg, Russian Federation*
**e-mail: soraya.aidene@gmail.com*

Whatever the advantages of the energy-dispersive X-ray fluorescence (EDXRF) for determination of elemental composition in different types of samples, the matrix effects remain a major problem and hinder precise quantitative analysis. Various approaches were suggested to take these effects into account. The most widely applied methods are fundamental parameters (FP) and intensity correction (IC). FP is based on a set of physical parameters employed in the equations relating X-ray fluorescence intensities and element content. IC is based on linear equations employing XRF line intensities and their ratios. However, these methods have certain drawbacks: application of FP requires the knowledge on the exact chemical composition of sample; in IC, the number of calibration samples should be several times greater than the number of influence coefficients.

As an alternative to the traditional methods, various chemometric approaches have been extensively investigated in recent years. Chemometric tools were shown to be useful for quantitative treatment of X-ray spectra, however, no comprehensive comparison between traditional methods and chemometric approaches was done so far. In this work we perform such comparison for processing of EDXRF spectra obtained for two sets of samples with complex composition: steels and ores. The target elements were Cr, Ni, Mn and Si in steels and K, Ca, Fe, Ti, Mn and Si in ores. The classic multivariate calibration tools like partial least squares (PLS), K-nearest neighbor regression (KNNR) and artificial neural networks (ANN) were employed and compared with traditional XRF processing techniques (OLS, FP, IC). It was found that traditional IC demonstrates the best performance in most of the cases. After analyzing the results of comparison we have proposed a novel method for matrix effects correction. It is a hybrid technique that combines the best features of both PLS and IC. The input data is made according to IC algorithm; then, the model is built using PLS regression. The details of these studies and the results will be given in the presentation.

Direct analysis of lanthanides in REE-rich ores by Laser-Induced Breakdown Spectrometry coupled with Chemometrics

T. F. Akhmetzhanov*[1], A. M. Popov[1]

[1] *Department of Chemistry, Lomonosov Moscow State University,*
119234, Leninskie Gory, 1-3, Moscow, Russia
**e-mail: atimur@laser.chem.msu.ru*

Quantitative determination of Rare-Earth Elements (REE) in ore samples is an interesting and challenging task. Laser-Induced Breakdown Spectrometry (LIBS) provides a unique possibility of remote direct analysis (ideally on a conveyor line). However, there are two major problems. The first is line overlapping in the spectra of REE-rich Niobium and Uranium ores. The second problem is a strong correlation between lanthanide contents preventing the use of chemometric approach as a solution for the first problem. Thus, the aim of this work is to solve the abovementioned problems.

To solve the problem of lanthanide correlation in standard reference materials we used specialized design of experiment (DoE) based on Latin hypercube sampling (LHS) that already proved its effectiveness in case of XRF analysis [1]. To circumvent spectral interferences in LIBS emission spectra we used Partial Least Squares Regression (PLS-R). Calibration was performed by choosing the lowest RMSECV-LOO and at the same time simplest models in terms of PC number. Naturally the number of principal components (latent structures) strongly depends on the number of elements which lines are interfering in a spectrum. For this reason, we considered narrowing the spectral regions used to build a PLS model according to the positions of interfering lines leaving signals that are only related to the element of interest (and overlapped ones that we cannot separate). Variable selection procedure is an important step in our effort to give the model a reliable prediction power. Thereafter, our multivariate models were validated by direct analysis of lanthanides in 6 niobium (OREAS 460–465) and 4 uranium (OREAS 100a–102a) ore samples. Chemometrics provided quantitative determination (in range of ±10% relative error) of Ce (0.046–3.95 %), La (0.026–2.41 %), Nd (0.015–1.18 %) and semiquantitative results (in range of ±30%) for Pr (47–3800 ppm) and Sm (24–1500 ppm). In comparison a univariate calibration for elements other than Ce provided semiquantitative results at best. This level of accuracy shows that LIBS coupled with Chemometrics is a powerful tool for direct analysis of lanthanides in REE-rich ores.

The reported study was funded by RFBR according to the research project № 19-33-90242.

References

1. T.F. Akhmetzhanov, G.V. Pashkova, V.M. Chubarov, T.A. Labutin, A.M. Popov, Three calibration techniques combined with sample-effective design of experiment based on Latin hypercube sampling for direct detection of lanthanides in REE-rich ores using TXRF and WDXRF, J. Anal. At. Spectrom (2021) 36:224–232.

Electronic nose based on Organic Field-Effect Transistors for environment and food quality monitoring

<u>D.S. Anisimov*[1]</u>, V.P. Chekusova[1,2], A.A. Trul[1,2], A.A. Abramov[1], O.V. Borshchev[1], E.V. Agina[1], S.A. Ponomarenko[1].

[1]Enikolopov Institute of Synthetic Polymeric Materials of the Russian Academy of Sciences, Profsoyuznaya str. 70, 117393 Moscow, Russian Federation.
[2]Printeltech LLC, Profsoyuznaya str. 70, 117393 Moscow, Russian Federation.
**email: anisimov@ispm.ru*

Determination of food freshness, which is the most ancient role of the sense of smell, is still a challenge for compact electronic nose devices. Fast, sensitive and reusable sensors are long awaited in food industry to replace long and expensive bacteriological methods and would alert many poisons related diseases and drastically reduce unnecessary food waste, which is now up to 30% of all the food produced. Modern solid-state gas sensors are approaching ppb-level limit of detection, while cost much less compared to the benchmark optical or mass-spectroscopy equipment. Organic field effect transistors (OFETs) are especially promising due to their high sensitivity, cheap production, low power consumption and operation at room temperature. However, due to non-covalent sensing mechanism, single OFET sensors respond to a variety of analytes including water vapors, leading to their main drawbacks – poor selectivity and environmental instability.

Recently we have developed the first fully integrated OFET-based electronic nose with the whole sensor array located on a single substrate [1]. It features down to 30 ppb limit of detection provided by monolayer thick active layers and operates in air with up to 95% relative humidity. By means of principal component analysis, it is able to discriminate toxic air pollutants and monitor meat product freshness. The gases released during chicken breast spoilage are studied and the ability to see clearly its decay just after 10 h of storage at room temperature due to reducing gases emission was demonstrated. Further electronic nose pre-training leads to the composition unraveling through comparison with training dataset with simple gaseous substances such as NH_3 or H_2S (Fig. 1).

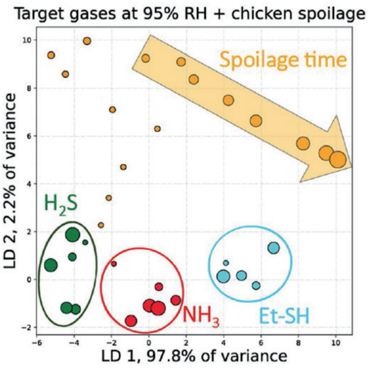


Fig. 1. LDA diagram of the signal growth during chicken spoilage projected onto training dataset.

Authors thank Russian Science Foundation (grant 19-73-30028) for funding.

References

1. Anisimov D.S., et al, Scientific Reports, 2021, 11, 10683

Development of a Multispectral Fiber Probe

A. Bogomolov*[1,2], T. Sakharova[2], I. Usenov[2], A. Artyushenko[2]

[1]Samara State Technical University, Molodogvardeyskaya 244, 443110 Samara, Russia
[2]art photonics GmbH, Rudower Chaussee 46, 12489 Berlin, Germany
**e-mail:* ab@globalmodelling.com

Multispectral measurements are used to enhance the data information content, and hence, to improve the accuracy and reliability of the chemical analysis. To enable simultaneous measurements, i.e. at the same time and sample point, the combined spectroscopic methods should be couped within the same measurement interface, such as a fiber probe. Qualitative analysis of the biological tissue is one of the most important applications of the developed multispectral probes [3].

We have developed a multispectral fiber optic probe that enables simultaneous analysis of various liquid and solid samples using attenuated total reflection (ATR) mid-infrared spectroscopy and fluorimetry [1,2]. Technical evaluation of the probe has confirmed its output signal quality that was comparable to that of respective probes for single methods. The capability of the probe to deliver complementary chemical information from the same measurement point has been illustrated using model samples of biological tissue.

One of the main disadvantages of the multispectral analysis limiting its wide dissemination is the need to use two separate spectrometers to register IR and fluorescence spectra. This can be avoided in a multisensor system, where specialization for a particular application enables a technological simplification of the spectroscopic methods so that they can be combined in one compact analytical device. The developed concept of combining ATR infrared and fluorescence spectroscopy provides a viable solution for the measurement interface in a multisensor system of this kind. The analysis can be limited to only two relevant information channels adjusted to a particular application by choosing appropriate excitation, emission and absorption wavelengths. Further development of the probe may include its modification to enable diffuse reflectance UV−vis, NIR and Raman measurements, thus opening new application areas.

References

1. A. Bogomolov, T. Sakharova, I. Usenov, C. Mizaikoff, V. Belikova, S. Perevoschikov, V. Artyushenko, O. Bibikova. Anal. Chem. 2021, 93, 6013−6018.
2. V. Artyushenko, T. Sakharova, I. Usenov, A. Bogomolov., D. Starodubov. Diagnostic System with Multispectral Optical Fiber Probe. Provisional U.S. Patent Application 63153426, 2021.
3. A. Bogomolov, V. Belikova, U. J. Zabarylo, O. Bibikova, I. Usenov, T. Sakharova, H. Krause, O. Minet, E. Feliksberger, V. Artyushenko. Sensors 2017, 17(11), art. no. 2548.

Benchtop NMR *versus* High Field NMR: Comparison of Chemometric Molecular Weight Analysis of Lignin

**R. Burger*[1], S. Lindner[1,2], J. Rumpf[1], X. T. Do[1], B. Diehl[3], M. Schulze[1],
Y. Monakhova[2,4]**

[1]*Department of Natural Sciences, Bonn-Rhein-Sieg University of Applied Sciences,
Rheinbach 53359, Germany*
[2]*Department of Chemistry and Biotechnology, FH Aachen University of Applied Sciences,
Jülich 52428, Germany*
[3] *Spectral Service AG, Köln 50996, Germany*
[4] *Institute of Chemistry, Saratov State University, Saratov 410012, Russia*
**e-mail:* rene.burger@h-brs.de

Lignin is a promising renewable biopolymer being investigated worldwide as an environmentally benign substitute of fossil-based aromatic compounds. For its successful implementation into process streams, a quick, easy, and reliable method is needed for the molecular weight determination of lignin. An approach using ^{1}H spectra of benchtop as well as conventional NMR systems in combination with simple PLS regression is presented to determine lignin's molecular weight (M_w and M_n) and polydispersity index (PDI). A set of 53 organosolv lignin samples was separated into calibration and external validation set. Validation errors between 5.6 % and 12.9 % were achieved for all parameters on four different NMR devices (43, 60, 500 and 600 MHz). Surprisingly, no significant differences in the performance of the benchtop and high-field devices were found. When signal sensitivity is not the main challenge of the analytical question, benchtop NMR systems appear to be a serious alternative to conventional high field NMR. This facilitates the application of this method for determining lignin's molecular weight in an industrial environment because of the low maintenance expenditure, small footprint, ruggedness, and low cost of permanent magnet benchtop NMR systems.

This study was supported by the BMBF (Grant 13FH102PX8), EFRE/NRW "Biobasierte Produkte" (Grant EFRE 0500035), and the FH Aachen University of Applied Sciences (internal project 1040343002).

References:

1. Burger, R.; Rumpf, J.; Do, X. T.; Monakhova, Y. B.; Diehl, B. W. K.; Rehahn, M.; Schulze, M. Is NMR Combined with Multivariate Regression Applicable for the Molecular Weight Determination of Randomly Cross-Linked Polymers Such as Lignin? *ACS omega* **2021**, 6 (44), 29516–29524. DOI: 10.1021/acsomega.1c03574.

Intantantly quantitative determination of curcuminoids in turmeric powder

J. Cruz [b], **P.Khongkaew** [a], **J. Puig** [c], **V. Cárdenas, M. Alcalà, Ch.Phechkrajang** [*,a]

[a] *Department of Pharmaceutical Chemistry, Faculty of Pharmacy, Mahidol University,*
447 Sri-Ayuthaya Road, Phayathai, Ratchathewi, Bangkok, Thailand
[b] *Escola Universitària Salesiana de Sarrià, Passeig Sant Joan Bosco,*
74, 08017 Barcelona), Spain
[c] *Universitat Autònoma de Barcelona, 08193 Bellaterra (Cerdanyola del Vallès),*
Barcelona, Spain
[d] *Instituto Nacional de Metrología Colombia,*
Av. Carrera 50 nº 26 - 55 Int. 2 Bogotá, D.C. Colombia
**e-mail:jcruz@euss.cat*

Near-infrared (NIR) spectroscopy and Raman Spectroscopy are significant non-invasive technologies, and with portable devices and using optic fiber probes, instruments can be taken to sample. These techniques can directly transmit through general plastic and glass bottles and possess a good capacity for qualification and quantification.

Turmeric, Curcuma Longa, is a plant with a long history of use in India and China's ancient traditional medicine[1]. Portable Raman and handheld NIR instruments have been used for the quantitative determination of total curcuminoids in turmeric samples by applying partial least squares regression, PLSR[2]. The results have been compared with different sample spectra registering methods: placing samples into a crystal vessel and direct measurement through the plastic bag which contains the sample.

When using the portable Raman with the standard sampling (crystal vessel), the optimal PLSR model has 3 latent variables and a mean square error of prediction (RMSEP) of 0.44 and in the case of handheld NIR with crystal vessel, it has 1 latent variable and a RMSEP of 0.41. For measures directly made through the plastic bag which contained the turmeric powder samples, the models made with the portable Raman spectra the optimal PLSR model has 1 latent variable and a mean square error of prediction (RMSEP) of 0.49. and for handheld NIR, it has 1 latent variable and a RMSEP of 0.41.

The results for Raman and NIR portable devices demonstrate that with the appropriate spectral pretreatments, we can create models that allow us to determine the total curcuminoids content without the need of placing the turmeric powder outside the plastic bag. This fact makes the determination almost instantaneous.

References

1. H.Hatcher, R.Planalp, J.Cho, F.M.Torti, S.V.Torti. *Cell. Mol. Life Sci.* (2008) *65*(11), 1631-1652.
2. M.Alcalà, M.Blanco, D.Moyano,N.W. Broad, N.O'Brien, D.Friedrich, H.W.Siesler. J. Near Infrared Spectrosc (2013)21(6), 445-457.

Fluorescent determination of the total content of humic in waters using multidimensional calibration models

A.O. Ferubko*[1], **I.N. Krylov**[1], **T.A. Labutin**[1]

[1]*Department of Chemistry, Lomonosov Moscow State University,*
Moscow 119234, Russia
**e-mail:* ferubko1999@yandex.ru

The global biogeochemical carbon cycle largely determines the climate, the number, and the diversity of living organisms. In the World Ocean, the content of organic carbon is ~ 10^{12} tons. A significant part of the organic carbon in the water is in the humic substances (HS). Optical and extraction methods are used for the quantitative determination of HS. The content of HS in marine areas isn't large, and there is a need to take a lot of samples from the ocean. While a full characterisation is possible by means of high-resolution mass spectrometry, its sample volume requirements typically lead to the use of different methods which sacrifice the information value for sample throughput. For example, fluorescence presents a fairly selective signal, making it possible to perform linear regression on the concentration of humic substances. Additionally, the 3-way spectra of the excitation-emission matrices (EEM) of the HS satisfy the condition of a trilinear signal. One of the regression methods designed for multi-linear data is N-PLS1[1] using L2 regularization, which is also less sensitive to noise in the data than PLS on unfolded spectra[2]. Therefore, the use of the N-PLS algorithm is advisable.

We have implemented the N-PLS1 algorithm from scratch in the Python 3 programming language. Like conventional PLS, N-PLS1 seeks a transformation of the original data maximising the covariance of the resulting scores with the variable being predicted, except separate loadings are obtained for every mode of the predictor tensor. We used cross-validation to find the optimal number of components, calculating the root-mean-squared error and the coefficient of determination. Several data sets were used to check the quality of our model, including a purely trilinear synthetic dataset (with known ground truth loadings and an orthogonal score matrix), the dataset by Åsmund Rinnan [3] (405 EEMs with 136 emission wavelengths and 19 excitation wavelengths containing 6 fluorophores), and the dataset by Dorrit Baunsgaard [4] (27 samples of size 121 by 24 containing 4 fluorophores). In addition, we used a data set previously prepared in the laboratory. This dataset contains 35 samples consisting of a mixture of model humic substances from the International Humic Substances Society, tyrosine and tryptophan, simulating seawater, and corresponding EEM spectra with 206 emission wavelengths and 55 excitation wavelengths. We obtained reproducible results using the N-PLS1 algorithm, on all data sets except the laboratory data set. We were able to obtain physically significant loadings, by using this model. In addition, our implementation of the N-PLS1 algorithm showed better results than the standard PLS module. Therefore it can be recommended for the detection of fluorophores in the ocean.

The authors of this work thank all creators of the datasets which we used in this work.

The work is supported by the Russian Science Foundation (grant 21-77-10064).

References

1. [1] A. Smilde, R. Bro, P. Geladi. Multi-Way Analysis with Applications in the Chemical Sciences, WILEY, John Wiley & Sons, Ltd (2004).
2. [2] R. Bro, Å. Rinnan, N. (Klaas) M. Faber. Chemom. Intell. Lab. Syst. 75 (2005), 69–76.
3. [3] Å. Rinnan. Application of PARAFAC on Spectral Data, Royal Veterinary and Agricultural University, 2004.
4. [4] J. Riu, R. Bro. Chemom. Intell. Lab. Syst. 65 (2003), 35–49.

Peak modelling using exponentially modified functions

Y.Kalambet[*1], Y. Kozmin[2]

[1]*Ampersand Ltd., Moscow 123182, Russia*
[2]*Institute of Bioorganic Chemistry, Moscow 117437, Russia*
**e-mail:* kalambet@ampersand.ru

Exponential modification of functions [1,2] is discussed in context of experimental peak shape modelling. Super-fast algorithm of digital exponential modification based on exponentially weighted moving average (EWMA) smoother [3]”ISSN”:”15206882”,”abstract”:”The simple recursive generating function of eq 8, with a time-dependent function, E(t is discussed. Algorithm allows calculation of exponentially modified Gaussian, Cauchi or any other function, as well as double-, triple- and other multiply exponentially modified functions. Study includes computation accuracy evaluation and correspondence of “analog” exponential decay time τ and EWMA parameter α. Algorithm of pre-evaluation of model peak parameters based on experimental peak cumulants is constructed.

References

1. [1] J.W. Ashley, C.N. Reilley, De-Tailing and Sharpening of Response Peaks in Gas Chromatography, Anal. Chem. 37 (1965) 626–630. https://doi.org/10.1021/ac60225a001.
2. [2] J.C. Sternberg, Extracolumn contributions to chromatographic band broadening, in: J.C. Giddings, R.A. Keller (Eds.), Adv. Chromatogr., vol. 2, Marcel Dekker, New York, 1966: pp. 205–270.
3. [3] A. Berthod, Mathematical Series for Signal Modeling Using Exponentially Modified Functions, Anal. Chem. 63 (1991) 1879–1884. https://doi.org/10.1021/ac00017a038.

Improving precision of potentiometric multisensor analysis using nonlinear regression methods: spent nuclear fuel reprocessing case study

Nadan Kravić[1], Julia Savosina[2], Marina Agafonova-Moroz[2], Dmitry Kirsanov[1*]

[1]Institute of Chemistry, Saint Petersburg State University,
Saint Petersburg 198504, Russia
[2] Khlopin Radium Institute, Saint Petersburg 198504, Russia
**e-mail: d.kirsanov@gmail.com*

Chemical monitoring of spent nuclear fuel (SNF) reprocessing poses challenging technological task for the variety of reasons. To this day, PUREX (Plutonium-URanium Extraction) process remains the most common industrial way of SNF reprocessing. One of the biggest problems in current technology is the inability to determine analytes in real time. North-East Atlantic Environment Strategy and IAEA strongly accented on development of automated analytical techniques that are capable of reliable detecting discharges of radioactive substances. Finding a method that would allow on-line process monitoring would improve the process in terms of automation, ecological and working safety and reliability. Two different approaches were proposed so far to address this task. The first one is based on optical spectroscopy in various modes (UV-Vis, NIR, IR, Raman) combined with chemometrics.

While the applicability of this approach was already demonstrated for a variety of tasks in SNF, still the approach is not tolerant against light-scattering media containing colloidal particles and gas bubbles – a common feature of SNF solutions. The second approach is based on multisensor potentiometry and has no limitations with respect to inhomogeneous solutions. It also requires chemometric data processing and all the studies reported so far were employing classic PLS regression for quantitative modelling. This may induce a certain limitation of the performance of the potentiometric method as PLS is intrinsically linear method, while the response of potentiometric sensors in complex media is not.

In this work a thorough investigation was done on performance of multisensor system in combination with non-linear multivariate regression models for quantification of analytes in PUREX process. Multisensor system is composed of 17 cross-sensitive sensors with different membrane ligands. Regression models such as support vector machines (SVM), random forest (RF) and kernel-regularized least squares (KRLS) were tested and compared to the traditional partial least squares (PLS) method. It was shown that non-linear methods outperform PLS for most of the analytes with RMSEP values acceptable for technological monitoring needs.

Scattering signal as a bilinear component in PARAFAC decompositions of fluorescence excitation-emission matrices

I. N. Krylov*[1], T. A. Labutin[1], Å. Rinnan[2], R. Bro[2]

[1]Department of Chemistry, Lomonosov Moscow State University,
Moscow 119234, Russia
[2]Department of Food Science, University of Copenhagen,
1958 Frederiksberg, Denmark
**e-mail:* ikrylov@laser.chem.msu.ru

Fluorescence excitation-emission matrices (EEMs) are well matched to the PARAFAC decomposition due to the direct correspondence between the underlying physical process and the mathematical model being fitted. However, scattering signal typically present in EEMs is not well described by PARAFAC and must therefore be handled prior to calculating the decomposition.

Interpolation of the areas containing the scattering signal [1] is a simple and widely used preprocessing method in PARAFAC decomposition. Compared to the insertion of missing data, it typically gives better results by avoiding local minima and physically impossible solutions allowed by the zero-error weight in the scatter areas. However, if a fluorophore overlaps with a second diffraction order scattering band, removing and interpolating the band may hide the component. Moreover, some interpolation methods are sensitive to noise in the spectral data and may introduce artefacts in the shape of the estimated PARAFAC components.

In this work, an approach combining PARAFAC and a bilinear model is suggested for the purpose of modelling both fluorescence and scattering signal at the same time, similar to multivariate curve resolution (MCR) with trilinear constraints [2]. On every iteration of the algorithm, PARAFAC and MCR fit each other's residuals with appropriate constraints (nonnegativity for both methods; MCR loadings fixed to zero outside the scattering bands).

The approach has been tested on various EEM datasets, including synthetic data, amino acid mixtures, sugar process data, and seawater DOM. By eliminating a preprocessing step and taking more of the EEM information into account, we can conclude that the decomposition results obtained this way are more reliable than the classical methods.

The reported study was funded by the Russian Foundation for Basic Research according to the research project No. 20-33-90280.

References

1. M. Bahram, R. Bro, C. Stedmon, A. Afkhami. Journal of Chemometrics. 20 (2006), 99–105.
2. R. Tauler, I. Marqués, E. Casassas. Journal of Chemometrics. 12 (1998), 55–75.

Is the Calibration Transfer of Multivariate Calibration Models Between High- and Low-Field NMR Instruments Possible?

Simon Lindner[1,2,*], René Burger[2], Douglas N. Rutledge[3,4], Xuan Tung Do[2], Jessica Rumpf[2], Bernd W. K. Diehl[5], Margit Schulze[2], and Yulia B. Monakhova[1,5,6]

[1]*Department of Chemistry and Biotechnology,*
FH Aachen University of Applied Sciences, Jülich 52428, Germany
[2]*Department of Natural Sciences, Bonn-Rhein-Sieg University of Applied Sciences,*
Rheinbach 53359, Germany
[3]*Université Paris-Saclay, INRAE, AgroParisTech, UMR SayFood,*
75005, Paris, France
[4]*National Wine and Grape Industry Centre, Charles Sturt University,*
Wagga Wagga, Australia
[5]*Spectral Service AG, Emil-Hoffmann-Straße 33, Köln 50996, Germany*
[6]*Institute of Chemistry, Saratov State University, 410012 Saratov, Russia*
**e-mail:* simon.lindner@smail.bcw.h-brs.de

Although plenty of successful applications of benchtop *nuclear magnetic resonance* (NMR) spectroscopy in quantitative mixture analysis exist, the possibility of calibration transfer remains mostly unexplored, especially between high- and low-field NMR. For the first time, the calibration transfer of *partial least squares* (PLS) regressions (weight average molecular weight (M_w) of lignin) between high-field (600 MHz) NMR and benchtop NMR devices (43 and 60 MHz) is presented. For the transfer, *piecewise direct standardization* (PDS), calibration transfer based on *canonical correlation analysis* (CCA), and *transfer via extreme learning machine auto-encoder method* (TEAM) are employed. Despite the immense resolution difference between high-field and low-field NMR instruments, the results demonstrate that the calibration transfer from high- to low-field is feasible, achieving validation errors close to the original calibration (down to only 1.2 times higher root mean square errors (RMSE)). These results introduce new perspectives for applications of benchtop NMR, in which existing calibrations from expensive high-field instruments can be transferred to cheaper benchtop instruments to cut costs.

This study was supported by the BMBF (grant 13FH102PX8), EFRE/NRW "Biobasierte Produkte" (grant EFRE 0500035) and the FH Aachen University of Applied Sciences (internal project 1040343002).

A moving-block-PCA based approach for real time monitoring of a powder blending process using a miniaturized near infrared sensor

Cristina Malegori*[1], Paolo Oliveri[1], Vanina Borghi[2], Diana Carolina Angel[2], Paola Melli[2], Eleonora Mustorgi[1], Monica Casale[1].

[1] *DIFAR - Department of Pharmacy, University of Genova,*
Viale Cembrano 4, 16148 - Genova

[2] *Medi Nova s.a.s., Via Beethoven 2/A - 42122 Reggio Emilia*
**e-mail: malegori@difar.unige.it*

Semen extenders (SE) are zootechnical products added to animal semen as diluent and preservative before artificial insemination procedures. In the present work, the blending process of a commercial boar semen extender is monitored by means of a miniaturized near infrared (NIR) device coupled with multivariate process monitoring. A key-step of SE production is the blending phase, in which all the ingredients are mixed together until the endpoint of the process is reached, so that the product can be considered as homogeneous as possible. The definition of the process endpoint is, usually, done according to a specific blending duration, decided on experience bases – an approach that is totally blind to any modification occurring during the process itself.

Aim of the present work is the development of a strategy for the monitoring of the blending phase in a continuous and non-destructive way, determining a specific process endpoint for every single batch, in a quality by design (QbD) view.

In order to reach the aim, a MicroNIR PAT-U sensor (Viavi Solutions), working in the spectral range between 900 and 1700 nm, was directly clamped on the blender, enabling the acquisition of a spectrum every 4 s, through a sapphire window during the whole process. It is important to remark that, before the implementation of the NIR sensor, the endpoint of SE blending was set at 20 min. Spectra were acquired for 20 batches of SE, used as calibration set, and the strategy was validated on other 6 independent batches.

On the collected signals, a proper data pretreatment was performed by means of standard normal variate (SNV) transform combined with Savitzky–Golay first derivative (5 datapoint window, second polynomial order). Then, standard deviation spectra were calculated applying a moving-block strategy with a block size equal to 5 spectra. This step generates new profiles that summarize the informative spectral variation without being influenced by unwanted systematic variations between the batches, such as changes in particle size/humidity of the ingredients or variations of the environmental conditions. After a proper mean-centering, principal component analysis (PCA) was applied on the calibration batches, including only the spectra of the mixed products. Based upon knowledge and previous exploratory analyses, the product was considered mixed after 15 min of blending. The validation batches were, then, projected into the space defined by the two lowest-order principal components, accounting for more than 90% of the explained variance. The number of components was decided thanks to a dedicated cross-validation strategy. The influence plot (Hotelling's T2 vs. Q residuals) and its statistical boundaries at a 95% confidence level were implemented as a multivariate control chart, for the monitoring of the behavior of new batches in the orthogonal space defined by PCA. The endpoint criterion was defined after acceptance of 15 consecutive spectra accepted by the multivariate control chart. Thanks to the present approach, the average time of the process was reduced to 5 min.

The whole strategy was implemented in a dedicated software, called NIRNova, developed in house with Matlab App Designer. Thanks to NIRNova, the production of SE is now monitored continuously in a non-destructive way, with a definition of the process endpoint for every single batch. Uniformity tests demonstrated that, despite the significantly time reduction, the final product is properly mixed. Moreover, an improvement of the product quality was highlighted, in terms of reduction of undesirable chalking and of temperature fluctuation of the powder.

MicroNIR and chemometrics for microplastics pollution

C.Marchesi*[1,2], M.Rani[1,2], S. Federici[1,2], L.E. Depero[1,2]

[1]*Department of Mechanical and Industrial Engineering, University of Brescia,*
via Branze 38, 25123 Brescia, Italy
[2] *Consorzio Interuniversitario per la Scienza e Tecnologia dei Materiali (INSTM),*
via G. Giusti 9, 50121, Firenze, Italy
**e-mail: c.marchesi003@unibs.it*

The rate of plastic production has grown exponentially reaching an annual output of 368 million metric tons in 2020 [1]. Despite a lot of benefits, plastic showed several ecological and toxicological problems due to the indiscriminate abandonment in the environment. This leads to the formation of Microplastics (MPs), i.e., "any synthetic solid particle or polymeric matrix, with regular or irregular shape and with size ranging from 1 µm to 5 mm, of either primary or secondary manufacturing origin, which are insoluble in water" [2]. The evidence of seafood contamination by MPs was confirmed and the potential consequences of the presence of MPs in the marine environment for human food security, food safety and health have been demonstrated [3-4]. In this scenario, an efficient and reliable method to detect and quantify MPs is strongly required. This study aimed to investigate the capability and the feasibility of a Miniaturized Near-Infrared (MicroNIR) spectrometer coupled with chemometric tools for qualitative and quantitative analysis of MPs. First, 250 plastic samples were collected based on the Resin Identification Code from a recycling plant to set up a library of plastic waste. MicroNIR spectra of plastic macropieces were analysed through the application of multivariate modelling: Principal Component Analysis (PCA) and Partial Least Squares Discriminant Analysis (PLS-DA). The second step was the creation of MPs at laboratory scale by mechanical fragmentation of daily used plastics items and arranged in mixtures starting from 0 to 100%. Chemometrics was then apply for qualitative and quantitative analysis of ternary and quaternary microplastics mixtures by means of PCA and PLS-regression. At first, for ternary mixtures, polypropylene (PP), polyethylene (PE), and polystyrene (PS) were chosen, reflecting the most abundant part of environmental plastic waste; then, polyethylene terephthalate (PET) was also included in the quaternary mixtures. Finally, the models developed were applied to field MP samples obtained from the fragmentation of environmental abandoned plastic waste to account for weathered effects on the polymers. The results are poised to impact for a fast, reliable, and in-situ environmental microplastics quantification.

References
1. PlascticEurope-Association of Plastics Manufactures Plastics – the Facts 2020. PlasticEurope 2020, 1–64.
2. Frias, J.P.G.L.; Nash, R..Mar. Pollut. Bull. 2019, 138
3. Lehtiniemi, M.; Hartikainen, S.; Näkki, P.; Engström-Öst, J.; Koistinen, A.; Setälä, O. Food Webs 2018, 17
4. Barboza, L.G.A.; Dick Vethaak, A.; Lavorante, B.R.B.O.; Lundebye, A.K.; Guilhermino, L. Mar. Pollut. Bull. 2018, 133, 336–348.

Primary and secondary data fusion from a sensor array for prediction of qualitative and quantitative indicators

A. A. Shuba*[1], T. A. Kuchmenko[1], E. V. Drozdova[1], A. E. Chernitskiy[2]

[1]Department of Ecology and Chemical Technology, Voronezh State University of Engineering Technologies, Voronezh 394006, Russia
[2] Research Center for Clinical Pharmacology and Therapy, Quality and Safety of Raw Materials and Products, All-Russian Scientific Research Veterinary Institute of Pathology, Pharmacology and Therapy, Voronezh 394000, Russia
**e-mail: shuba1nastya@gmail.com*

One of the urgent tasks of chemometrics is the prediction of various quantities, as a rule, which is difficult to determine (quantitative) or represent a conclusion based on a set of other parameters (qualitative). The prediction uses the data concerning the studied properties, obtained in a relatively simple way and method, for example, predicting the authenticity of pharmaceuticals, safety and quality of food products, diagnosis of the disease, based on the data from spectrometers, chromatographs, artificial tongues, noses. The most common data fusion includes information from several methods for predicting properties or classification. However, combining different types of signals obtained by the same method associated with the studied properties can also be helpful.

This work aims to analyze the primary and secondary data fusion from an array of eight piezoelectric sensors when analyzing the gas phase over samples of exhaled breath condensate to predict bacterial contamination of tracheal flush (quantitative indicator) and the state of the calves' respiratory organs (qualitative indicator) by the partial least squares regression.

Data fusion was investigated at a low- and mid-level. The raw data from an array of sensors (chronofrequencygrams) and maximum sensor signals during the sorption of the gas phase over biosamples - primary data, the calculated parameters of the sensor array - secondary data, reflecting a greater extent the qualitative composition of the gas mixture were the initial data blocks. Depending on the type of data, preprocessing was used - smoothing by Savitzky-Golay, Gaussian filters, Savitsky-Galey differentiation, scaling, centering. The features of primary and secondary data fusion from an array of sensors are considered for predicting qualitative and quantitative indicators. The choice of variables for each data block is scrutinized when constructing models according to statistical criteria, taking into account their information content in chemical analysis. The optimal composition of the initial matrix for data fusion with the best predictive properties is recommended for studied qualitative and qualitative indicators.

The data acquisition in this study was funded by the Russian Science Foundation according to research project No. 18-76-10015.

Region of Interest Selection for GC-MS Data with a Pseudo Fisher Ratio Moving Window

M.S. Armstrong[*1], **R.T. Giebelhaus**[1], **A. P. de la Mata**[1], **J.J. Harynuk**[1]

[1]*Department of Chemistry, University of Alberta, 11227 Saskatchewan Dr NW,*
Edmonton, Alberta, Canada T6G 2N4
**e-mail:* mdarmstr@ualberta.ca

Gas Chromatography – Mass Spectrometry (GC-MS) data present a number of challenges for analysis. This is largely because chemical factors drift along the chromatographic mode across different chromatographic runs, and there is a lack of reliable molecular ion measurements with which to align multiple samples. Matrix decomposition techniques such as Multivariate Curve Resolution (MCR) or Independent Component Analysis (ICA) operating on individual samples, and Parallel Factor Analysis (PARAFAC) or PARAFAC2 for multiple samples allow analysts to deconvolve closely eluting signals for quantitative and qualitative purposes[1]. These techniques make relatively few assumptions about chromatographic peak shapes or the relative abundance of the signal noise, and allow for highly accurate representations of the underlying chemical phenomena using well-characterised and scrutinised principles of chemometrics. However, expert intervention and supervision is required to select appropriate Regions of Interest (ROI) and numbers of components for each ROI.

Baccolo et al. have reported a fully automated approach to deploying PARAFAC2 for entire chromatograms[2]. In this presentation, the authors propose an alternate method for defining regions of interest specifically by observing the ratio of the first and second eigenvalues within a moving window across an entire GC-MS chromatogram. For auto-scaled data, the authors posit that the two eigenvalues can be interpreted as χ^2 values, whose ratio can be interpreted using the F-Distribution for values greater than 1. The algorithm outputs a continuous series of probabilities for each observation in the chromatogram, and the user can select a probability threshold to define the ROIs.

References

1. Johnsen, Lea G., et al. «Gas chromatography–mass spectrometry data processing made easy.» *Journal of Chromatography A* 1503 (2017): 57-64.
2. Baccolo, Giacomo, et al. «From untargeted chemical profiling to peak tables–a fully automated AI driven approach to untargeted GC-MS.» *TrAC Trends in Analytical Chemistry* (2021): 116451.

Aquaphotomics approach for cancer diagnosis and monitoring

Anastasiia Surkova [1, 2] *, **Andrey Bogomolov** [1], **Ekaterina Boichenko** [2],
Olga Bibikova [3], **Urszula Zabarylo** [3], **Viacheslav Artyushenko** [3],
Thaddäus Hocotz [4], **Jelena Munćan** [5], **and Roumiana Tsenkova** [5]

[1] Department of Analytical and Physical Chemistry,
Samara State Technical University,Samara 443100 Russia
[2] Institute of Chemistry, Saint Petersburg State University,
Saint Petersburg 198504, Russia
[3] art photonics GmbH, Berlin 12489, Germany;
[4] Charité – Universitätsmedizin Berlin 12203, Germany;
[5] Kobe University, Kobe 657-8501, Japan
**e-mail:* melenteva-anastasija@rambler.ru

Early diagnosis and monitoring of the patient’s condition in response to the treatment helps to reduce the cancer mortality. The standard procedures in cancer recognition such as biopsy and histopathology are expensive, often invasive, time-consuming, and injurious. Therefore, a development of a new rapid screening method for cancer diagnosis and further treatment is highly desired.

Cancer causes severe changes in the composition of the body fluids, such as urine and blood. The detection of cancer-related compositional changes in a patient’s biological fluid is a promising direction of the research in this field [1]. Near infrared (NIR) spectroscopy that provides fast and non-invasive measurements without difficult sample pretreatment is a great tool for the elucidation of water structure. However, there is no concerted strategy for the spectral data analysis that hinders wide application of NIR spectroscopy in the clinical practice.

Aquaphotomics is a young scientific discipline that studies water molecular conformations in the aqueous and biological systems [2]. The spectral region 680–2500 nm provides reach information about the water molecular structure. Water molecules tend to form hydrogen bonds between each other and with their surroundings, which makes water absorption bands in NIR region very sensitive to the sample composition.

The aim of this study is to investigate water molecular structure in the NIR spectra of biological fluids (urine, plasma and serum of blood). The samples were collected from patients before and after surgery of different cancer types. Raw and pre-processed spectra, as well as the loading vectors of the PCA analysis were studied to find and assign the characteristic water absorbance bands that change significantly in response to the surgical treatment. The reported results can be potentially used for the characterization of biological materials and cancer screening.

The reported study was funded by the President’s grant #MK-2192.2021.4.

References

1. P. Raja, P. Aruna, D. Koteeswaran , S. Ganesan. Vib. Spectrosc. 102 (2019), 1–7.
2. R. Tsenkova, J. Munćan, B. Pollner, Z. Kovacs. Front. Chem. 6 (2018), 363.

Compositional features of zooplankton species as studied by laser-based analytical techniques

N. I. Sushkov*[1], **N. V. Lobus**[2], **A. Kéri**[3], **Á. Bélteki**[3], **G. Galbács**[3], **T. A. Labutin**[1]

[1]Department of Chemistry, Lomonosov Moscow State University,
Moscow 119234, Russia
[2]Timiryazev Institute of Plant Physiology, Russian Academy of Sciences,
Moscow 127276, Russia
[3]Faculty of Science and Informatics, University of Szeged, Szeged 6720, Hungary
**e-mail:* nikolaisushkov@laser.chem.msu.ru

Laser-based techniques are advantageous for the analysis of vast variety of samples. Laser ablation sampling provides the basis for analytical techniques like laser-induced breakdown spectroscopy (LIBS) and inductively coupled plasma mass spectrometry with laser ablation (LA-ICP-MS). These are direct micro-destructive techniques, the former yielding information-rich atomic emission spectra, the latter making possible high sensitivity determination of trace elements. Raman spectroscopy, also based on laser interrogation, gives vibrational spectra characteristic of molecules comprising the sample. We used these techniques together to explore the untrivial compositional features of marine zooplankton, some species of which are known to accumulate Li, As and U. Our goal was to explore correlations between the elemental and molecular composition of these animals.

LIBS and LA-ICP-MS experiments were carried out using the Applied Spectra J200 Tandem LA-LIBS spectrometer (266 nm laser, detection within 186–1049 nm) and Agilent 7700 Series ICP-MS spectrometer. The signals of ^{7}Li, ^{11}B, ^{27}Al, ^{28}Si, ^{34}S, ^{35}Cl, ^{44}Ca, ^{56}Fe, ^{63}Cu, ^{66}Zn, ^{75}As, ^{88}Sr, ^{137}Ba, ^{238}U, ^{23}Na, ^{24}Mg, ^{31}P, and ^{39}K were recorded in the time-resolved mode. Raman measurements were done using the Thermo Scientific DXR Raman Microscope (780 nm laser) in the range of 45–3500 cm^{-1}. In all spectra, baseline was removed, prominent peaks selected and integrated. The resulting matrix of emission, mass, and scattering signals of 29 samples belonging to four biological taxa was analysed using PCA and non-negative matrix factorisation (NMF).

The PCA score plots allowed for a reasonably good discrimination of the animals, especially for two major taxa of *Calanoida* and *Euphasiaceae*. The loadings suggested that the classification was mainly driven by two latent vectors. One of them (more characteristic of *Calanoida*) contained the prominent signals of Li, Al, Cl, Fe, U and carotenoid pigments. The other included the signals of P, K, Ca, Cu, As, Sr, Ba and some Raman peaks. Thus, the accumulation patterns of Li and U are correlated, while As probably follows a different biochemical pathway (e.g. as a companion of P).

The reported study was funded by RFBR and BRFBR, project number 20-53-04036. The authors are grateful to D. Palásti and P. Janovszky (University of Szeged) for their assistance during the experiments.

Application of cluster analysis for identification of injection water breakthrough in oil production wells

I. S. Trukhin*[1], **P. A. Zadorozhny**[1], **S. V. Sukhoverkhov**[1], **A. N. Markin**[2]

[1]Institute of Chemistry, Far Easter Branch of Russian Academy of Science,
Vladivostok 690022, Russia
[2]Industrial University of Tyumen, Nizhnevartovsk 628616, Russia
**e-mail: truhin.ivan.91@gmail.com*

The breakthrough of injected water into the producing wells is a serious problem in oil production by waterflooding. The identification of the water inflow source is very important to taking correct measures [1]. Tracers are often used to control the movement of water in the oil fields. They are highly informative, but complex and expensive [2]. Regular monitoring of the chemical composition of the produced and injected waters is carried out at most of the fields. Therefore, the changing of a produced water chemical composition may be used to identify the water inflow source. In this work a method was developed for detecting and calculating the amount of injected water in the production wells based on the dynamics of changes in the statistical distances between the injected and produced water, obtained by the chemical composition of the water samples cluster analysis.

The calculations were based on the results of chemical composition studies of 240 samples of produced and injected water from two Sakhalin offshore oil and gas platforms using the “Statistica 12” software package. Injected seawater breakthrough in some wells has been established by k-means cluster analysis. The samples of produced and injected water from these wells were separated into three clusters with the intermediate by chemical composition parameters cluster, which confirmed the dynamics of water inflow. The ratio of injected / produced waters in the samples from the wells was calculated by hierarchical cluster analysis. The results correlated well with the sulphate content; no correlation with other components was observed. It was shown that this approach allows to identify latent oilfield processes and to minimize the influence of random factors or factors that do not affect the breakthrough of the injected water.

The work was carried out with the financial support of the Ministry of Education and Science, state assignment number 0265-2019-0002, using the equipment of the Center for Collective Use of the Far Eastern Center for Structural Research of the IC FEBRAS.

References

1. B. Bailey, M. Crabtree, J. Tyrie. The Challenge of Water Control. Oilfield Review. Schlumberger. 12 (2000), 30–51.
2. C. Serres-Piole, H. Preud’homme, N. Moradi-Tehrani etc. Water tracers in oilfield applications: Guidelines J. Pet. Sci. Eng. Elsevier. 98–99 (2012), 22–39.

Various machine learning methods in HPLC-MS datasets treatment

P. N. Turova*[1], A. N. Stavrianidi[1,2]

[1]*Department of Chemistry, Lomonosov Moscow State University,*
Moscow 119234, Russia
[2]*A.N. Frumkin Institute of Physical Chemistry and Electrochemistry Russian Academy of Sciences, 31 Leninsky Prospect, GSP-1, Moscow 119071, Russia*
**e-mail:* turova.polina@gmail.com

HPLC-MS method is most often used for the potential chemotaxonomic markers search in plants extracts due to its informativeness, sensitivity and lack of preliminary derivatization. In our work we obtained mass-chromatogram of samples from 19 species of the Apiaceae family. In the first part various unsupervised algorithms for multivariate statistical analysis were used to process the obtained data such as dimensionality reduction methods (PCA, ICA, NMF, PARAFAC) and unsupervised feature selection (UFS). Obtained clustering results (dendrograms) were compared with biological taxonomic tree. Also, 23 potential chemotaxonomic markers for plants from the Apiaceae family have been identified. These markers were mostly related to the coumarin group. Their preliminary identification was carried out using HPLC-HRMS (MS and MS^2 spectra)[1]. However, these biomarkers don't identify groups of plants by their genus.

In the second part of the work supervised machine learning methods were applied to the extended with other plants' parts dataset for classification and biomarkers' search. Basic approaches already widely described in literature[2] like PLS-DA and SVM were applied to unfolded dataset. After revealing class-specific features it was found out that they are not unique for each class. Further it was decided to try to apply neural networks because they are able to handle LC-MS data without vectorization and are known for huge flexibility and high level of accuracy. Because of the specific constraints in the dataset like class imbalance, low number of samples and expectation of specific features Siamese network with triples loss function was used. Moreover, dataset was extended using augmentation techniques with respect to LC-MS data specific. Results obtained by this method were compared with mentioned basic machine learning techniques.

HPLC-MS-TOF analysis for biomarker identification was performed using the equipment of the demo laboratory of Bruker Ltd., Moscow, Russia

References

1. P. Turova, I. Styles, V. Timashev, K. Kravets, A. Grechnikov, D. Lyskov, T. Samigullin, I. Podolskiy, O. Shpigun, A. Stavrianidi, J Pharmaceut Biomed, 206 (2021).
2. P. Kharyuk, D. Nazarenko, I. Oseledets, I. Rodin, O. Shpigun, A. Tsitsilin, M. Lavrentyev, Sci Rep 8 (2018), 1-12.

Advanced coupled factorization of multi-way fluorescence data to assess the conjugation of quantum dots to proteins

I. M. A. Viegas*[1,2], R. Bro[3], J. M. Amigo[3,4,5], G. A. L. Pereira[1], C. F. Pereira[1]

[1]Department of Fundamental Chemistry, Federal University of Pernambuco,
Recife 50740-560, Brazil
[2]Centre for Oil and Gas, Technical University of Denmark,
Kgs. Lyngby 2800, Denmark
[3]Department of Food Science, University of Copenhagen,
Frederiksberg 1958, Denmark
[4]IKERBASQUE, Basque Foundation for Science, Bilbao 48009, Spain
[5]Department of Analytical Chemistry, University of the Basque Country UPV/EHU,
Bilbao 48080, Spain
**e-mail: iviegas@dtu.dk*

Quantum dots (QDs) are semiconductor nanocrystals with an active surface that enables their conjugation with a variety of biomolecules, forming inorganic-biological hybrid nanoparticles with exceptional optical characteristics from the QDs and the biochemical functions from the conjugated molecules [1]. The conjugation process is usually characterized by laborious and/or costly techniques that requires the sample preparation procedures, which can interfere with the properties of the bioconjugates and lead to misinterpretation of the results [2]. In this context, we proposed the coupled factorization of fluorescence excitation-emission matrices (EEMs) measured in two spectral regions, over nine weeks, to extract underlying characteristics of the QDs-protein conjugation and monitor it over time. That originated two tensors: $\underline{\mathbf{A}}$ for the spectral range of proteins and $\underline{\mathbf{B}}$ for the region of QDs, which were decomposed by PARAFAC into three and two components, respectively for $\underline{\mathbf{A}}$ and $\underline{\mathbf{B}}$. Subsequently, Advanced Coupled Matrix and Tensor Factorization (ACMTF) jointly decomposed the two tensors into three components, which allowed to improve the recovered profiles in the shared modes (relative concentrations and time). Therefore, we concluded that fluorescence spectroscopy associated with multi-way analysis has a great potential as a non-destructive, quick technique to extract implicit information about the conjugation of QDs to molecules of biological interest.

This work was funded by CNPq, CAPES, INCTAA and Instituto Serrapilheira.

References

1. L. Jing, S. V. Kershaw, Y. Li, X. Huang, Y. Li, A. L. Rogach, M. Gao. Chem. Rev. 116 (2016), 10623-10730.
2. K. E. Sapsford, K. M. Tyner, B. J. Dair, J. R. Deschamps, I. L. Medintz. Anal. Chem. 83 (2011), 4453-4488.

Prediction of potentiometric selectivity for carbonate-selective plasticized membrane sensors using QSPR modelling

Nadezhda I. Vladimirova*, **Julia S. Ashina, Dmitry O. Kirsanov**

Institute of Chemistry, Saint Petersburg State University,
Saint Petersburg 198504, Russia
**e-mail:* hada96@mail.ru

Despite the fact that PVC-plasticized membrane ion selective sensors are well-known and successfully applied analytical devices, the search for new, more sensitive and selective materials for these membranes, ionophores, is still in progress. The development of ion selective sensors for detecting hydrophilic anions such as e.g. hydrocarbonate or dihydrophosphate anions is one of the challenging areas in modern sensor research. This is due to the number of issues: 1) anions have a broad variety of geometric conformations in contrast to typically spherical metal cations; 2) anions also have denser solvation shell in comparison with cations of a similar charge; 3) some anions change their ionic forms depending on pH value of a sample solution.

An investigation of each new potential ligand for ion sensing consists of several stages such as a synthesis of the organic compound, sensor membrane preparation, and electrochemical evaluation of the resulted sensors with respect to its' sensitivity, selectivity, etc. That is why the overall process of developing novel ion sensors with new ionophores is expensive and time and labour consuming process.

Modern computational chemistry has a tool for *in silico* evaluation of the properties of organic compounds. QSAR/QSPR modelling allows relating the chemical structure of the compound to its' properties through mathematical equation. Calculation of the regression coefficients for such equation in QSPR typically requires thousands of compounds in a training set. Unfortunately, due to the economic reasons mentioned above such huge datasets are not available in potentiometric sensing. Nevertheless, the recent studies from our group have shown that QSPR modelling of potentiometric selectivity is possible at certain level even with extremely small datasets [1]

In this work we have extended this approach to anionic sensing and proposed a QSPR model for prediction of logK (HCO_3^-/Cl^-) selectivity based on the chemical structures of the ionophores described in literature. The details on the calculation and results will be given in the presentation.

This study was supported by RFBR project #20-33-70272.

Reference

1. E. Martynko, V. Solov‘ev, A. Varnek, A. Legin, D. Kirsanov. Electroanalysis. 2 (2020), 792–798.

Voltammetric sensors in the analysis of enantiomeric mixture using PLS

L.R. Zagitova*, E.V. Shestakova, Yu.A. Yarkaeva, R.A. Zilberg, V.N. Maistrenko

Bashkir State University, Ufa 450076, Russia
**e-mail:* kabirova.lian@ya.ru

The development of a chiral selective analytical technique and quantification of one enantiomeric isomer in a mixture is very important in the food and pharmaceutical industries as well as in vivo/vitro clinical analysis. Enantioselective voltammetric sensors (EVS) have proven to be highly sensitive, express, relatively inexpensive techniques for chiral analysis. It is known that most of the works in this area are mainly based on the differences in the peaks of the oxidation current of enantiomers. In fact, enantiomers of biologically active substances usually coexist in a mixture, and their detection without differences in oxidation potential is difficult to apply in the analysis of real objects. However, insufficient sensor selectivity can be compensated by chemometric techniques.

For the recognition of tryptophan, warfarin, and atenolol enantiomers voltammetric sensors based on different chiral selectors were developed [1-3]. The maximum enantioselectivity coefficient was 1.34 and the maximum potential difference of the oxidation peaks was 40 mV. The observed differences are small, which can lead to errors. Therefore, for reliable recognition enantiomers in a mixture PLS was used. Solutions were prepared with different contents of enantiomers (0, 20, 40, 60, 80, 100%). The differential pulse voltammograms (DPVs) for each of these mixtures were recorded between +500 and +1100 mV, producing a data matrix with 30 rows (5 parallel measurements for 6 solutions of mixtures) and 159 columns (instantaneous currents for each DPV). On the score plot DPVs formed clusters arranged in the order of change in the ratios of enantiomers in the mixture along PC1. To confirm the operability of the obtained PLS-model, a crossvalidation was carried out. The resulting model can be considered adequate and suitable for prediction, since the slope and the correlation coefficient are close to one, the offset is close to zero, and the RMSEP was 0.09. The composition of the test set was within the ranges of calibration set, but randomly designed. All samples were correctly identified using this model, the relative error was less than 8% in all cases.

Thus, it has been shown that chemometric data processing makes it possible to recognize the ratios of enantiomers in a mixture with a higher accuracy and to solve the problem of the proximity of analytical signals.

The study was supported by Russian Science Foundation (project number 21-13-00169), https://rscf.ru/project/21-13-00169/»

References

1. L.R. Zagitova, V.N. Maistrenko, Y.A. Yarkaeva, *et al.* // *J. Electroanal. Chem.*, **2021**, 880, 114939.
2. R.A. Zilberg, V.N. Maistrenko, L.R. Zagitova, *et al.* // *J. Electroanal. Chem.*, **2020**, 861, 113986.
3. Y.A. Yarkaeva, V.N. Maistrenko, L.R. Zagitova, *et al.* // *J. Electroanal. Chem.*, **2021**, 903, 115839.

Study of the applicability of modern machine-learning techniques to the task of similarity search of objects represented by series of images and metadata

Y. V. Zontov*[1,2]

[1]*HSE University,*
[2]*Federal Institute of Industrial Property, 121059, Moscow, Berezhkovskaya nab. 20*
**e-mail:* yzontov@hse.ru

In this study we are considering the applicability of modern machine-learning techniques to the task of similarity search of objects represented by series of images and metadata, containing textual descriptions. Such objects may be encountered in many research areas, but we are regarding the industrial design database of Federal Institute of Industrial Property as our dataset.

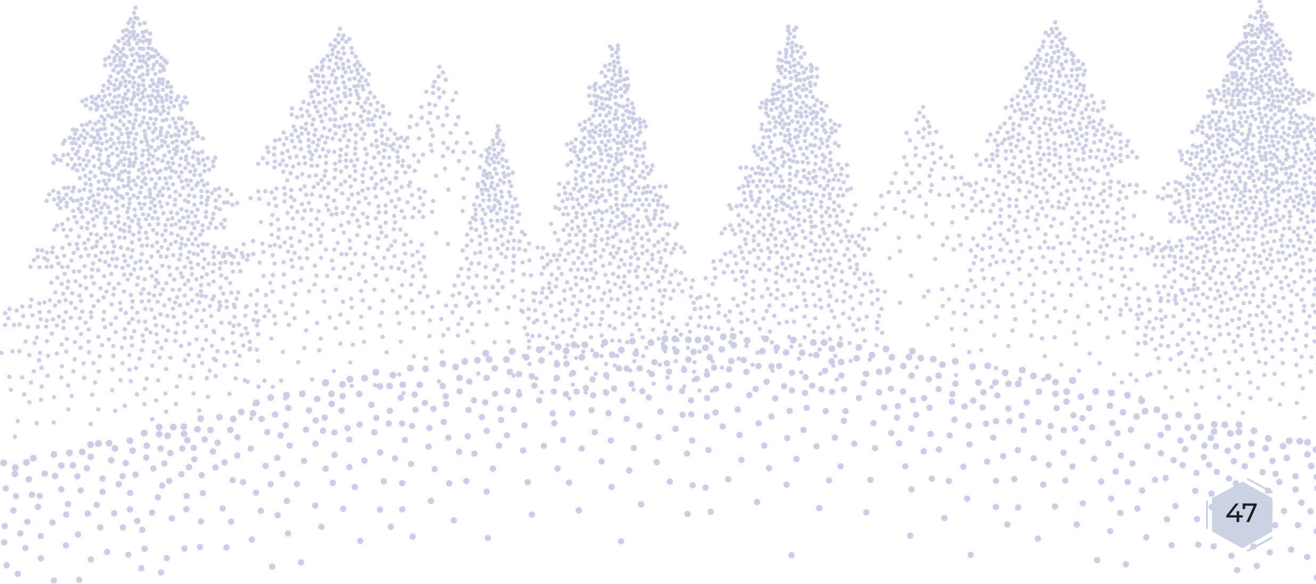
A significant aspect of industrial designs is the representation by multiple images, which in many cases are views of the same object from different angles, and presence of textual descriptions and classifications. This feature suggests the effectiveness of architectures that take into account the fact that each object in the training set is represented by a set of entities - a situation common in text analysis problems. In this regard, the architecture of Visual Transformers [1] or approaches based on CNN (Multi-view classification) [2] can be applied, as well as ensembles of a networks designed for image similarity search, and networks designed for text meaning similarity search, in the presence of a consolidated quality function (Multi-Modal Classification) [3].

In this study we use several existing neural network architectures to obtain vector representations for visual and textual properties of an industrial designs (embeddings) and then apply popular data analysis techniques such as DD-SIMCA [4] and PLS-DA [5] to the resulting vector representations.

References

1. M. Seeland, P. Mäder, Multi-view classification with convolutional neural networks // PLoS ONE 16(1): e0245230, 2021
2. Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, Baining Guo. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows, 2021, arXiv, URL:https://arxiv.org/abs/2103.14030v2 (Visited on 30.08.2021)
3. Stuart J. Miller et al. Multi-Modal Classification Using Images and Text // SMU Data Science Review, 2020, vol.3
4. Y.V. Zontov, O.Ye. Rodionova, S.V. Kucheryavskiy, A.L. Pomerantsev. DD-SIMCA – A MATLAB GUI tool for data driven SIMCA approach, Chemometrics and Intelligent Laboratory Systems, Volume 167, 2017, Pages 23-28
5. Y.V. Zontov, O.Ye. Rodionova, S.V. Kucheryavskiy, A.L. Pomerantsev. PLS-DA – A MATLAB GUI tool for hard and soft approaches to partial least squares discriminant analysis // Chemometrics and Intelligent Laboratory Systems. 2020. Vol. 203. P. 1-9

FLASH TALKS



PARAFAC2×2: a PARAFAC Model for Drift in Two Modes

M.S. Armstrong*[1], A. P. de la Mata[1], J.J. Harynuk[1]

[1]Department of Chemistry, University of Alberta, 11227 Saskatchewan Dr NW,
Edmonton, Alberta, Canada T6G 2N4
**e-mail:* mdarmstr@ualberta.ca

Comprehensive Two-Dimensional Gas Chromatography – Time-of-Flight Mass Spectrometry (GC×GC-TOFMS) is an advanced variant of Gas Chromatography – Mass Spectrometry (GC-MS) that utilises an additional dimension to separate complex mixtures. The additional column features a different phase than the first, and can be used to separate co-elutions that would otherwise be difficult to resolve using a single separation. Because of the resultant decrease in signal-to-noise ratio and improved peak resolution, GC×GC-TOFMS outperforms GC-MS on complex samples. Despite these advantages, and the increasing accessibility of GC×GC-TOFMS platforms for researchers, uptake in both industry and academia has been slow. One possible reason for this lack of enthusiasm is the absence of transparent and reliable tools to handle GC×GC-TOFMS data. For alignment of chemical features across multiple samples, these tools must be able to account for drift along both chromatographic modes, and researchers relying on commercial offerings must optimise several parameters that can have a dramatic effect on the representation of the underlying data.

PARAFAC2 offers a parsimonious solution for deconvolution and quantification of GC-MS data with drift along one mode, which exploits the advantages of multi-way decomposition. It has been successfully deployed as a semi-supervised tool for GC-MS as the freely available package: PARADISe [1], and a fully automated implementation has also been reported [2]. Equivalent GC×GC-TOFMS data is an I×J×K×L tensor of I spectral acquisitions (second dimension retention mode), J mass-to-charge ratios, K modulations (first dimension retention mode), and L samples, application of PARAFAC2 is not straightforward for these data. A possible work-around is to unfold the modulations to create an IK×J×L tensor. Doing so introduces a rotational ambiguity to the model, and decreases the number of replicates relative to the degrees of freedom along the unified retention mode.

We present an approach to processing GC×GC-TOFMS data sets, with PARAFAC2×2. PARAFAC2×2 calculates two regularised PARAFAC2 models to account for drift in both the first- and second-dimension retention modes. At convergence, the two models are averaged to estimate for the unique score tensors for each sample and their resultant mass spectra and relative quantities.

References

1. Johnsen, Lea G., et al. «Gas chromatography–mass spectrometry data processing made easy.» *Journal of Chromatography A* 1503 (2017): 57-64.
2. Baccolo, Giacomo, et al. «From untargeted chemical profiling to peak tables–a fully automated AI driven approach to untargeted GC-MS.» *TrAC Trends in Analytical Chemistry* (2021): 116451.

Recognition and classification of skin Raman spectra using convolutional neural networks analysis

L. A. Bratchenko*[1], Y. A. Khristoforova[1], A. A. Moryatov [2,3], S. V. Kozlov[2,3], V. P. Zakharov[1], I. A. Bratchenko[1]

[1]Department of Laser and Biotechnical Systems, Samara University,
Samara, 443086, Russia
[2]Department of Oncology, Samara State Medical University, Samara, 443095, Russia
[3]Department of Visual Localization Tumors, Samara Regional Clinical Oncology
Dispensary, Samara, 443095, Russia
**e-mail: shamina94@inbox.ru*

Portable Raman spectroscopy setups for skin "optical biopsy" are utilized to detect human body pathologies based on skin spectral features caused by the comparative presence of different chemical components. However, low signal-to-noise ratio in such systems may prevent accurate skin spectra classification. Thus, there is a challenge to develop methods for efficient recognition and classification of skin Raman spectra.

Methods: We compare the performance of convolutional neural networks and the projection on latent structures with discriminant analysis for discriminating skin tissues according to pathologies using the analysis of noisy Raman spectra with a high autofluorescence background stimulated by a 785 nm laser. To check the classification models stability, a 10-fold cross-validation was performed for all created models. To avoid models overfitting, the data was divided into a training set (80% of spectral dataset) and a test set (20% of spectral dataset).

The results for different classification tasks demonstrate that the convolutional neural networks significantly ($p<0.01$) outperforms the projection on latent structures. The performance of the convolutional neural networks classification of skin according to pathologies based on Raman spectra analysis is higher or comparable to the accuracy provided by trained dermatologists. The increased accuracy with the convolutional neural networks implementation is due to a more precise accounting of low intensity Raman bands in the intense autofluorescence background.

This study was partly supported by grant from Russian Foundation of Basic Research № 20-32-90040 (L.A. Bratchenko) and grant from the Russian Science Foundation No. 21-75-10097, https://rscf.ru/project/21-75-10097/.

Handling of optical multisensor data

Evseeva Alexandra, Ignatiev Egor ,Bogomolov Andrey

Samara State Technical University, Molodogvardeyskaya 244, 443110 Samara, Russia
**e-mail:* evseevasaha1111@gmail.com

Development of simplified miniature devices specialized for a particular application is an important trend in the modern analytical chemistry. Such devices called optical multisensor systems (OMS) have a number of distinguishing features that make them different from the traditional spectroscopy. One of the main differences of the new multisensor technique is the nature of data increasingly generated by OMS.

OMS is an analytical device including a few (at least two) optical sensor channels selected and optimized to solve a wide range of analytical problems in different areas, such as quality control in the food and pharmaceutical industry, ecological monitoring and medical diagnostics [1].

The emergence and rapid dissemination of the optical sensor analysis is starting to change the entire situation in the practical chemical analytics worldwide. The purpose of the present paper is to attract attention to the problem of incomplete appropriateness of the current well-established toolkit that is widely used for the data handling in spectroscopy and to propose alternative approaches to the data acquisition, visualization and analysis, which could be more efficient when dealing with OMS.

References

1. A. Bogomolov, V. Belikova, U. J. Zabarylo, O. Bibikova, I. Usenov, T. Sakharova, H. Krause, O. Minet, E. Feliksberger, V. Artyushenko. Sensors 2017, 17(11), art. no. 2548.

Generalization and geometrical interpretation of N-FINDR algorithms for unmixing hyperspectral data

R. R. Guliev*[1], U. Neugebauer[1,2,3], C. Beleites[4]

[1]*Leibniz Institute of Photonic Technology, 07745 Jena, Germany*
[2]*Institute of Physical Chemistry and Abbe Center of Photonics, 07743 Jena, Germany*
[3]*Jena Center for Soft Matter (JCSM),*
Friedrich Schiller University Jena, 07743 Jena, Germany
[4]*Chemometrix GmbH, 61200 Wölfersheim, Germany*
**e-mail:* rustam.guliev@leibniz-ipht.de

Unmixing, the process of determination of certain basis spectra and representation of each pixel with respect to that basis, is an important tool for the analysis of hyperspectral data. N-FINDR is one of the most commonly used algorithms for unmixing. Although it is widely used, the implementation of the algorithm is not well established. Since the first publication of the algorithm in 1999 [1], multiple works suggested different implementations and optimizations of the algorithm under many different abbreviations: SM N-FINDR, SQ N-FINDR, SC N-FINDR, IN-FINDR, LDU-N-FINDR, LDU-S-N-FINDR, S-N-FINDR, MN-FINDR, MN-FINDR2, and so on. With this and the lack of publicly available implementations, the definition of the algorithm became somewhat blurred. Although there were attempts to classify implementations of the N-FINDR algorithm[2], it happened that the same abbreviation is used for different implementations of the algorithm[3].

In this work, we generalize the N-FINDR algorithms giving a geometrical interpretation. The algorithm split into main steps and the use of abbreviations is replaced by explicitly naming the chosen option for each step. In addition to covering most of the previous implementations, this also allows us to introduce some previously unknown implementations. The generalized version is implemented in R and publicly available on GitHub (r-hyperspec/unmixR). In addition, each modification of the algorithm was vectorized by replacing loop steps by matrix operations. Finally, benchmarking of each algorithm implementation was performed and recommendation for default algorithm usage is given.

We gratefully acknowledge the contributions to the R package by Conor McManus (2013) and Anton Belov (2016) as well as their funding by Google Summer of Code, and Bryan Hanson's contributions. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 861122 (ITN IMAGE-IN).

References

1. M. Winter, *Proc. SPIE 3753, Imaging Spectrometry V*, 3753 (1999), 2689–2692
2. W. Xiong *et al.*, *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, 4 (2011), 545–564
3. S. W. Dowler, R. Takashima, and M. Andrews, *IEEE Trans. Image Process.*, 22 (2013), 2835–2848

Untargeted metabolomics study of *Humulus lupulus* brewing cultivars, for genetic origin classification task

Y. A. Ikhalaynen*, **I.A. Rodin**, **I.V.Plyushchenko**

Department of Chemistry, Lomonosov Moscow State University,
Moscow 119991, Russia
**e-mail:* ikh.ya@yandex.ru

In plant analysis omics approaches help to solve a wide range of different problems – from botanic classification to environment pollution studies. Genomic and metabolomic approaches were main instruments in plant studies for many years. In spite of the fact that DNA based methods seem to be most suitable for plant classification tasks, the resulting phenotype of a species could vary because of different factors, such as different gene expression influence on phenotype and specifics of plant genome. In this situation, an untargeted metabolomics approach seems convenient for confirmation of other types of classification and cross-omics studies, since it does not require difficult conditions of sample preparation and interpretation of raw data. In our study, we provide workflow of brewing *Humulus lupulus* cultivars classification according to previously obtained EST-SSR based results [1] by LC-HRMS profiling, including feature selection and identification of significant metabolites.

18 cultivars of hops (in pellets) were analyzed, each in 3 biological repeats (54 samples in total). Randomization of the experimental sequence and quality control (QC) samples were used to reduce influence of batch effect and systematic error. Sample preparation consisted of ultrasonic extraction of homogenised pellets using methanol. Mixture of all samples were used for QC samples preparation. LC-HRMS profiling was provided via Shimadzu LCMS-IT-TOF system, equipped with electrospray ionization ion source (ESI) and reversed-phase column Thermo Acclaim RSLC C18. Peak picking and chromatogram alignment were performed using xcms R package with parameters were optimized by IPO package. Random forest based missing value imputation and XGboost based signal correction were performed for the peaks table. abels, obtained in reference genomic study, were fine-tuned by hierarchical cluster analysis based on principal components coordinates.

Marker compounds were selected using rational cut off according to three tests: moderated t-test, fold change and PLS-DA algorithm. Collision induced dissociation fragmentation mass-spectra were than obtained for selected biomarkers and processed by various annotation tools, including MetFrag, CFM-ID and McSearch.

References

1. Josef Patzak, Alena Henychová. *Czech Journal of Genetics and Plant Breeding* 54.2 (2018): 86-91.
2. (2018), 86-91.

Stochastic optimisation of spectra of laser-induced plasma for plasma composition calculation

Ivan Krylov, **Sergey Zaytsev, Timur Labutin**

Department of Chemistry, Lomonosov Moscow State University. Moscow. Russia.
e-mail: ikrylov@laser.chem.msu.ru

Laser-induced breakdown spectroscopy (LIBS) is an atomic emission technique which contactless quantifies elemental composition of solids, liquids or gases using the emission of plasma produced by a high-power Q-switched laser on the surface or in a volume of a sample. Thus LIBS can be used in unique applications such as analysis of Mars surface, ocean floor, or process control of molten metal and dross. In most of those cases it is extremely difficult or even impossible to obtain certified reference materials for calibration. Common approaches to calibration-free (CF) LIBS are based on Boltzmann plots and, therefore, tend to impose constraints on the experiment, namely homogeneity of the source, local thermodynamic equilibrium in plasma and absence of self-absorption of involved atomic lines. Fulfilling these requirements thorough experimental optimisation and preliminary studies may prove impossible in space missions or in deep sea research.

We have earlier developed an algorithm based on a stationary model of laser-induced plasma [1] to simulate emission spectra for a given sample composition. In this work, we applied optimisation techniques to fit experimental and model spectra, in which relative elemental concentrations served as variables. For relative simplicity of an implementation, a gradient-free method has been chosen to minimise the sum of squared differences between scaled model and experimental spectra.

Various shapes of the loss function are also considered, making it possible to account for analytical lines of very different intensity. We have tested the accuracy by 100 consequent runs of the algorithm with random initialisation. It has been shown that the homogeneous plasma model is suitable for narrow spectral regions, while a multi-zone model with different temperatures and electron densities provides the best results for the full spectral region, although in the latter case the accuracy is lower.

The reported study was funded by Grant of the President of the Russian Federation (No. MK-5513.2021.6).

References

1. [1] Zaytsev S.M.; Popov A.M.; Labutin T.A.; Stationary model of laser-induced plasma: Critical evaluation and applications, Spectrochimica Acta Part B: Atomic Spectroscopy, **2019**, *158*, 105632.

Decomposition of in vivo skin Raman spectra using multivariate curve resolution method

I. A. Matveeva*[1], **I. A. Bratchenko**[1], **O. O. Myakinin**[1], **Y. A. Khristoforova**[1],
L. A. Bratchenko[1], **A. A. Moryatov**[2,3], **S. V. Kozlov**[2,3], **V. P. Zakharov**[1]

[1]*Department of Laser and Biotechnical Systems, Samara National Research University,*
Samara 443086, Russia
[2]*Department of Oncology, Samara State Medical University, Samara 443099, Russia*
[3]*Department of Visual Localization Tumors, Samara Regional Clinical Oncology*
Dispensary 443031, Samara, Russia
**e-mail:* m-irene-a@yandex.ru

It is known that during the development of the disease, metabolic changes in the patient's body occur, which leads to changes in the biochemical composition of biological tissues and fluids. In recent years, Raman spectroscopy has been increasingly used to investigate these changes. However, Raman spectra of most biological tissue components overlap significantly, and it is difficult to separate individual components from the analysed Raman spectrum. Besides, most methods for analysing Raman spectra have an impossibility of physical interpretation of the components. The aim of our study is to investigate the possibilities of the multivariate curve resolution method [1] for the analysis of in vivo skin Raman spectra using a portable low-cost spectroscopy setup [2]. The multivariate curve resolution analysis of Raman spectra of normal skin, keratosis, basal cell carcinoma, malignant melanoma and pigmented nevus is performed.

As a result, we obtained Raman spectra corresponding to the contribution of the optical system and skin components: melanin, proteins, lipids, water, etc., that demonstrate the possibility of unmixing several spectrally similar components using MCR analysis even under noisy conditions of the recorded Raman spectra. The classification of the Raman spectra of various diseases (malignant vs. benign neoplasms, malignant melanoma vs pigmented neoplasms) by the contribution of the spectra of the components shows the classification accuracy about 70%. The obtained results show that the multivariate curve resolution analysis can provide new information about the biochemical profiles of the skin tissues. Such information may be used in medical diagnostics and screening of the population for the analysis of Raman spectra with a low signal-to-noise ratio.

This study was supported by Russian Science Foundation grant No. 21-75-10097

References

1. J. Jaumot, A. de Juan, R. Tauler. Chemometr. Intell. Lab. Syst. 140 (2015), 1–12.
2. I.A. Bratchenko, L.A. Bratchenko, A.A. Moryatov, Y.A. Khristoforova, D.N. Artemyev, O.O. Myakinin, A.E. Orlov, S.V. Kozlov, V.P. Zakharov. Exp. Derm. 30(5) (2021), 652–663.

Sensor system based on Cu(II) and Zn(II) amino acid complexes for recognition of atenolol enantiomers in racemic mixture using the partial least squares

M.I. Nazyrov*, Y.A. Yarkaeva, L.R. Zagitova, V.N. Maystrenko

Department of Chemistry, Bashkir State University, Ufa 450076, Russia
**e-mail:* mnazyrov@list.ru

Recognition and determination of enantiomers of biologically active compounds plays a huge role in modern medicine, because alive organisms react differently to their presence in medicines. In this work, composites of polyarylenephthalide and amino acid complexes of Cu(II) and Zn(II) – glycine-L-alaninate and glycine-L-argininate Cu(II), bis-L-phenylalaninate Zn(II), were used as modifiers of a glassy carbon electrode in enantioselective voltammetric sensor systems for the recognition and determination of atenolol (Atn) enantiomers [1].

For reliable recognition of enantiomers, we used a sensor system, such as an electronic tongue, which involves chemometric processing of voltammetric data. To do this, we combined individual sensors into a sensor system (i.e., the data obtained from each sensor was combined into a single matrix) and carried out their joint chemometric processing, in particular, using principal component analysis (PCA). In this case, each voltammogram was transformed into a point and located on the plane of the principal components (PC). The location of the points on the PC plane indicates the similarity and difference of the analyzed samples. The next step is to build a model based on calibration set containing a known amount of Atn enantiomers, and to recognize test set prepared independently using the partial least squares discriminant analysis (PLS-DA). The use of a sensor system allows to eliminate errors of the first kind and minimize errors of the second kind. The percentage of correctly recognized samples reaches 99%.

And finally, the proposed sensor system was used to determine the ratio of the enantiomers Atn (mass fraction) in the mixture, which is of great practical importance. The plots of PLS-modeling show that the points corresponding to different S-Atn contents in the mixture form clusters, and the clusters corresponding to 0.0 to 1.0 mass fractions of S-Atn are located on opposite sides on the coordinate plane, and the rest clusters are located between them. The results show that in all cases the content of S-Atn in the mixture was established correctly with RSD not exceeding 8.9% and recovery of 95–102%.

The reported study was funded by RSF, project number 21-13-00169.

References

1. Yu.A. Yarkaeva, V.N. Maistrenko, L.R. Zagitova, M.I. Nazyrov, T.V. Berestova. J. Electroanal. Chem. 903 (2021), 115839.

Application of gradient boosting machine for signal processing in LC-MS metabolomics

I. V. Plyushchenko*, **I. A. Rodin**

Chemistry Department, Lomonosov Moscow State University, Moscow 119991, Russia
**e-mail:* plyushchenko.ivan@gmail.com

Signal drift correction with batch effect removal in LC-MS based metabolomics is aimed to remove unwanted variation components, such as within- and between-batch variations, both sample and instrument sensitivity changes over time in a long injection sequence and dependency of measured values by injection order and/or batch number. "QC-LOESS" [1] regression for signal correction has become the "gold" standard over the last decade. Some improvements include the implementation of other nonlinear algorithms, i.e. random forest (RF) [2].

A complex examination of other regression models was performed [3] by various numeric criteria [4]. As a result, the new signal correction method, which outperformed other QC regression algorithms, based on gradient boosting machine (XGboost, QC-XGB) was proposed. A more precise comparison based on graphical estimation was carried out on Amide [5] metabolomics dataset for QC-LOESS, QC-RF and QC-XGB methods.

References
1. Dunn W.B., Broadhurst D., Begley P., Zelena E., Francis-McIntyre S., Anderson N., Brown M., Knowles J.D., Halsall A., Haselden J.N., Nicholls A.W., Wilson I.D., Kell D.B., Goodacre R. Nat. Protoc. 6 (2011), 1060–1083.
2. Luan H., Ji F., Chen Y., Cai Z. Anal. Chim. Acta. 1036 (2018), 66–72.
3. Plyushchenko I. V., Fedorova E. S., Potoldykova N. V., Polyakovskiy K. A., Glukhov A. I., Rodin I. A. J. Proteome Res. (2021), 1c00392.
4. Sánchez-Illana Á., Piñeiro-Ramos J.D., Sanjuan-Herráez J.D., Vento M., Quintás G., Kuligowski J. Anal. Chim. Acta. 1019 (2018), 38–48.
5. Deng K., Zhang F., Tan Q., Huang Y., Song W., Rong Z., Zhu Z.-J., Li K., Li Z. Anal. Chim. Acta. 1061 (2019), 60–69.

Nonlinear dimensionality reduction methods for multisensor system data analysis

Zahars Selivanovs[1]*, Regina Belugina[2], Julia Savosina[3], Marina Agafonova-Moroz[3], Dmitry Kirsanov[1]

1Institute of Chemistry, Saint Petersburg State University,
Saint Petersburg 198504, Russia
[2] ITMO University, Saint Petersburg Russia
[3] Khlopin Radium Institute, Saint-Petersburg, Russia
**e-mail: st073551@student.spbu.ru*

Processing the data from multisensor systems for chemical analysis is mostly based on traditional chemometric tools, like PCA. In case of complex data structures it has been shown that nonlinear dimensionality reduction methods outperform linear ones [1]. Potentiometric multisensor systems in real complex multicomponent samples provide the responses containing information on numerous chemical species and these responses are not linear in nature. We have hypothesized that application of non-linear dimensionality reduction methods may yield valuable outputs for analysis.

In order to test this hypothesis, we have compared linear (PCA) and nonlinear tools (isomap, self-organizing maps (SOM) and autoencoder) as applied to two different multisensor datasets. The first one contained the results of potentiometric measurements in urine samples obtained from 65 individuals (30 with confirmed bladder cancer diagnosis and 35 volunteers with no complaints). The purpose of modeling was in classification of the samples into normal and pathological. The second dataset was obtained during online multisensor potentiometric measurements in spent nuclear fuel reprocessing media. It contained 8469 observations with 18 sensors acquired within 17.5 hours. The purpose of the modelling was to explore the relation between sensor responses and the process status.

The results of the comparison between the performance of linear and non-linear dimensionality reduction methods will be given in the presentation.

References

1. A Nonlinear Dimensionality Reduction Framework Using Smooth Geodesics. Gajamannage Kelum, Paffenroth Randy, Bollt Erik M. Pattern Recognition (2017), Vol. 87, 226-236

The use of artificial neural networks for the simultaneous determination of several analytes by the oxidation reactions of carbocyanine fluorophores

A. V. Shik*[1], M. K. Beklemishev[1]

[1]Department of Chemistry, Lomonosov Moscow State University,
Moscow 119991 GSP, Russia
**e-mail:* shik.1966@mail.ru

The aim of this work was to demonstrate the possibility of simultaneous determination of two drugs by their effect in the reaction of catalytic oxidation of a NIR dye. The determination is possible for the compounds that give different shapes of kinetic curves of dye oxidation. Consideration of a number of dyes and oxidizers made it possible to choose the oxidation of a carbocyanine dye with hydrogen peroxide catalyzed by copper(2+) at pH 7.3 as giving the most reproducible NIR fluorescence intensities.

As model analytes, the components of biseptol drug were introduced into the reaction: sulfamethoxazole (5 µM – 0.25 mM) and trimethoprim (30 µM – 0.3 mM, final concentrations are indicated) in various molar ratios (from 5:1 to 1:50). A total of 11 different ratios were used, each was studied in three parallel experiments (three wells of the plate). The progress of the indicator reaction was monitored by taking photographs of the plate at different times (0, 6, 9, 17, 21, 29 and 39 min). The calibration model was built using artificial neural networks in the Statistica 7 and 12 programs. The indicated concentrations of analytes, as well as the total intensities obtained from the photographs at all indicated reaction times, were introduced into the program as variables. Of the 33 experiments, 18 were used as a training set, 8 as a validation set, and 7 as a test set. An automatic enumeration of ANN architectures was set.

The most efficient network turned out to be the MLP 7:7-11-2:2 multilayer perceptron (7 input neurons, 11 hidden and 2 output neurons, 200 epochs, learning rate 0.1). The training error was 0.056, the validation error was 0.147, and the test error was 0.31. When using this neural network, the error in predicting the concentration of sulfamethoxazole and trimethoprim was less than 10% for 5 out of 7 samples. Thus, using the proposed system, it is possible to determine semiquantitatively two analytes in an aqueous solution.

The work was supported by the Russian Science Foundation (grant No 20-13-00330). The authors thank Dr. T.A.Podrugina and Dr. I.A.Doroshenko for the samples of dyes.

Spectrophotometric determination of active substances in ophthalmic formulation using chemometrics

D. V. Silaev*, T. Yu. Rusanova

Institute of Chemistry, Saratov State University, Saratov410012, Russia
**e-mail:* sil_diman@mail.ru

The combination of two or more active pharmaceutical ingredients (APIs) that form a complex therapeutic system in a medicinal product is popular with consumers due to its convenience, economy, wide pharmacological effect, and reduced side effects. For the treatment of allergic reactions and inflammatory eye irritations in the spring-autumn period, ophthalmic dosage forms are used, in which a combination of APIs that have anti-allergic and vasoconstrictive effects is widely used.

The labor-intensive and complex preliminary separation of APIs that have similar physicochemical properties makes it difficult to carry out express analysis for quality control of medicines in the laboratory. Chemometric algorithms, which are currently promising, express and do not have large labor costs, are an alternative to labor-intensive procedures associated with the preliminary separation of components during analysis.

The aim of the work was to evaluate the possibility of quantitative determination of the API of diphenhydramine hydrochloride and naphazoline nitrate in eye drops “Sigida Duo” by using a combination of spectrophotometry with chemometric algorithms (projections on latent structures (PLS1 and PLS2) and principal components regression, PCR).

The drug “Sigida Duo”, containing diphenhydramine hydrochloride 1.0 mg/ml and naphazoline nitrate 0.33 mg/ml, refers to combined antiallergic drugs for the treatment of allergic conjunctivitis, inflammatory conditions of irritated conjunctiva of the eyes.

19 mixtures were prepared with various API concentrations in the range of 0.3-3.25·10-4 mol/l. Electronic absorption spectra were recorded using a Shimadzu UV 1800 spectrophotometer in the range of 200–400 nm, the recording step was 0.2 nm. The absorption maxima for naphazoline hydrochloride and naphazoline nitrate were 258 and 280 nm, respectively. The number of mixtures in the calibration and testing sets was selected by the cross-validation method (14 and 6 mixtures, respectively). Based on the values of root mean square error of prediction (RMSEP), the optimal numbers of latent variables or principal components are selected. Errors in the determination of API in the model solutions and in the drug “Sigida Duo” do not exceed 10%.

Discrimination of plant samples using carbocyanine fluorophore additives

E. V. Skorobogatov*[1], M. K. Beklemishev[1]

[1]Department of Chemistry, Lomonosov Moscow State University,
Moscow 119234, Russia
**e-mail:* skorobogatoveg@gmail.com

The tasks of recognition and classification of samples are solved by obtaining their "fingerprints" (spectra, chromatograms, voltammograms, etc.) and their analysis by chemometric methods. Fluorimetric methods are successfully used to discriminate and classify samples of various nature.

In this work, we studied the feasibility of using fluorescent fingerprinting to establish the fact of irradiation of food samples with high energy electrons with sterilization purposes. We studied potato samples irradiated on an UELR-1-25-T-001 industrial electron accelerator with an energy of 1 MeV**. The extracts of potatoes were prepared by cutting the central part of the tuber into cubes (about 5 mm) and shaking with water or alcohol for 24 h. Extracts of potatoes with irradiation doses 1, 5 and 10 kGy were used for the analysis along with non-irradiated samples as blank. The experiment was carried out in 96-well plates in 6 parallel runs. The cell was filled with the potato sample extract, the carbocyanine as fluorophore, surfactant, necessary for the implementation of the aggregation mechanism of dye fluorescence, and buffer with pH 7,4. The images of the plates were digitized using ImageJ software.

The obtained digital data were processed by the PCA method, followed by the construction of confidence ellipses for a confidence level of 95% (by using XLStat software). For the analysis, PC-1 and PC-2 components were used as the most informative. Based on fluorimetry data and using the PCA method, it was possible to divide samples of potatoes into classes according to the level of irradiation. Complete separation was not achieved; however, in the case of aqueous extracts, only classes with close levels of irradiation overlapped; the groups of samples with irradiation of 1 and 10 kGy do not overlap. In the case of alcoholic extracts, all groups strongly overlapped with no separation. In addition, extracts of non-irradiated potatoes of a different variety were studied, which formed a separate class, different from non-irradiated samples. This is a serious problem of the described approach, since the difference in the type of non-irradiated matrix (in this case, the potato variety) has an effect on thediscrimination along with the level of irradiation.

***Irradiation was carried out in Skobeltsyn Institute of Nuclear Physics, Moscow State University by A. P. Chernyaev and V. V. Rozanov.*

The work was supported by the Russian Science Foundation (grant No 20-13-00330). The authors thank Dr. U.A. Bliznyuk for the samples of potatoes and Dr. T.A.Podrugina and Dr. I.A.Doroshenko for the samples of dyes.

Grey modelling of absorption spectra of silver nanoisland films in aqueous solutions

A. N. Skvortsov*[1,2], **E. S. Babich**[1]

[1] Peter the Great Saint Petersburg Polytechnic University, Saint Petersburg 195251, Russia
[2]Laboratory of the molecular biology of stem cells, Institute of Cytology of
the Russian Academy of Science, Saint Petersburg 194064, Russia
**e-mail:* colbug@mail.ru

Outstanding photophysical properties of silver nanoparticles (SNP) make them a very powerful tool for biosensing applications. However SNP are known to corrode in natural waters and biological media. Recently we characterized the properties of silver nanoisland films formed on glass surface by out-diffusion and exposed to phosphate buffers with various ion compositions [1]. During the exposure, the absorption spectra of the film in the visible range showed complex and reproducible dynamics, caused by changes in silver species concentration and size in glass volume and on its surface, and related shift of the plasmon resonance of SNP. The aim of the present study was to build a multivariate curve resolution (MCR) model, which would allow comparing the spectral changes in different experiments semi-quantitatively and estimating kinetic parameters.

The applications of conventional self-modelling MCR methods or simple kinetic model fitting were unsuccessful. The resolutions had an improbably high rank and were unstable, which disagreed with similar spectral changes and final chemical results in the experimental replicas. Augmentation of the data matrix failed to improve the results. The main reason was the fact that chemical modification of SNP affected both the amplitude and the position/width of the plasmon peak(s). Thus the bilinearity assumption of MCR ($\mathbf{X}=\mathbf{C}\mathbf{S}^t+\mathbf{E}$, **S** does not depend on time) was broken and the effective rank of the data matrix increased. So, we attempted to use semi-empirical MCR by alternating least squares (grey MCR-ALS) allowing certain smooth changes of selected columns of the spectral matrix **S** with time. Different sets of constraints were tested (monotony, smoothness, peak shape fitting, peak alignment). The best results were obtained using scaled/shifted empirical peak shape to constrain plasmon peak(s). The approaches for estimation of the empirical peak shape and implementations of constraints are discussed. The ambiguity, stability and convergence of the models are also considered.

The work was supported by Ministry of Education and Science of the Russian Federation as a part of the State program for the fundamental research (theme code FSEG-2020-0024).

References

1. A. Skvortsov, E. Babich, A. Redkov, A. Lipovskii, V. Zhurikhina. Biosensors.11 (2021), 448. https://doi.org/10.3390/bios11110448

Chemometric methods in the recognition of antibiotics using voltammetric sensory systems based on molecularly imprinted polymers

Yu.A. Yarkaeva*, **D.A. Vorontsova, D.Kh.Farahova, V.N. Maistrenko**

Department of Chemistry, Bashkir State University, Ufa 450076, Russia
**e-mail:* Julijajarkaeva05@gmail.com

Antibiotics are one of the most important therapeutic classes that have a huge impact on human health. However, their unwise use has led to the spread of antibiotics and related metabolites in ecosystems and food. Consequently, the analysis of antibiotics in environmental samples and in biological samples becomes an urgent task. In the field of pharmaceutical and biological analysis, electrochemical methods are widely used due to their high sensitivity, low cost of equipment with relatively fast response times. In particular, voltammetric methods have recently found wide application in the field of drug analysis, in particular, antibiotics. However, it should be noted that a little-studied area in this case is the determination of antibiotics in the joint presence. An approach based on the use of molecularly imprinted polymers (MIPs) [1] has great prospects in the development of voltammetric sensors. It consists in the polymerization of the monomer around the template molecule, as a result of which the formed polymer has specific cavities for the selected analyte after its removal. In this work, we propose the use of polyaniline and its derivatives o-toluidine and o-anisidine as MIPs, as well as the creation of sensor systems based on them for the determination of antibiotics in the joint presence. In this case, sensor systems can be assembled from sensors designed to detect different antibiotics and, accordingly, made from the same polymer, but with molecular imprints of different templates. In this case, it is possible to simultaneously determine several antibiotics belonging to different classes. In addition, for a more accurate and selective determination of one antibiotic in the presence of others, for example, those belonging to the same class and having a similar structure, sensor systems can be assembled from sensors made from different polymers, but in the presence of the same template. To ensure the functioning of such sensory systems, obtaining and interpreting the results, chemometric data processing methods are used, in particular, the principal component analysis as a data visualization method, projection on latent structures and soft independent modeling of class analogies as classification methods.

The reported study was funded by the Russian Science Foundation according to the research project No. 21-73-00295, https://rscf.ru/project/21-73-00295/.

References

1. M. Bougrini, A. Florea, C. Cristea, R. Sandulescu, F. Vocanson, A. Errachid, B. Bouchikhi, N. El Bari, N. Jaffrezic-Renault. Food Control. 59 (2016), p. 424-429.

Chemometric processing of the data from high-frequency inductor for chemical sensing

Ekaterina A. Yuskina, Nikodim A. Makarov, Valentin.G. Semenov, Vitaly.V. Panchuk, Dmitry.O. Kirsanov

Saint Petersburg State University, Saint Petersburg, Russia
**e-mail:* lazyvelikaya@yandex.ru

In this study we propose a very simple and low-cost detector for contactless chemical sensing which may have many analytical applications. The sensor device is based on an inductance coil connected to a high-frequency electric field generator (4-114MHz). When a sample solution is introduced inside the coil as an inductor core, the receiver detects variations in the registered “spectrum” depending on the dielectric permittivity and conductivity of the sample. Chemometric processing of these spectral signals allows distinguishing between various media, various substances and it is concentration sensitive.

The proposed concept was tested in several applications to confirm it’s potential. The applicability of the device to distinguish between different inorganic salts solutions ($Cu(NO_3)_2$, $Pb(NO_3)_2$, Na_2SO_4, KBr, KJ) was demonstrated. Moreover, it was shown that the detector is able to distinguish between various concentrations of the electrolytes in the concentration range 10^{-4} – 10^{-1} M. Principal component analysis applied to the results of measurements in different organic solvents revealed the clear dependence of the registered electrical signal at certain frequencies on solvent’s dipole moment.

The applicability of the detector to real samples was demonstrated in the experiment with milk and cream samples with various fat contents. Partial least squares regression modeling demonstrated that the device can be employed for contactless evaluation of fat content in dairy products. The sensor also can detect inorganic salt additions in milk and ethanol content in water-ethanol mixtures.

The details on the experimental set-up and the data processing will be provided in the presentation.