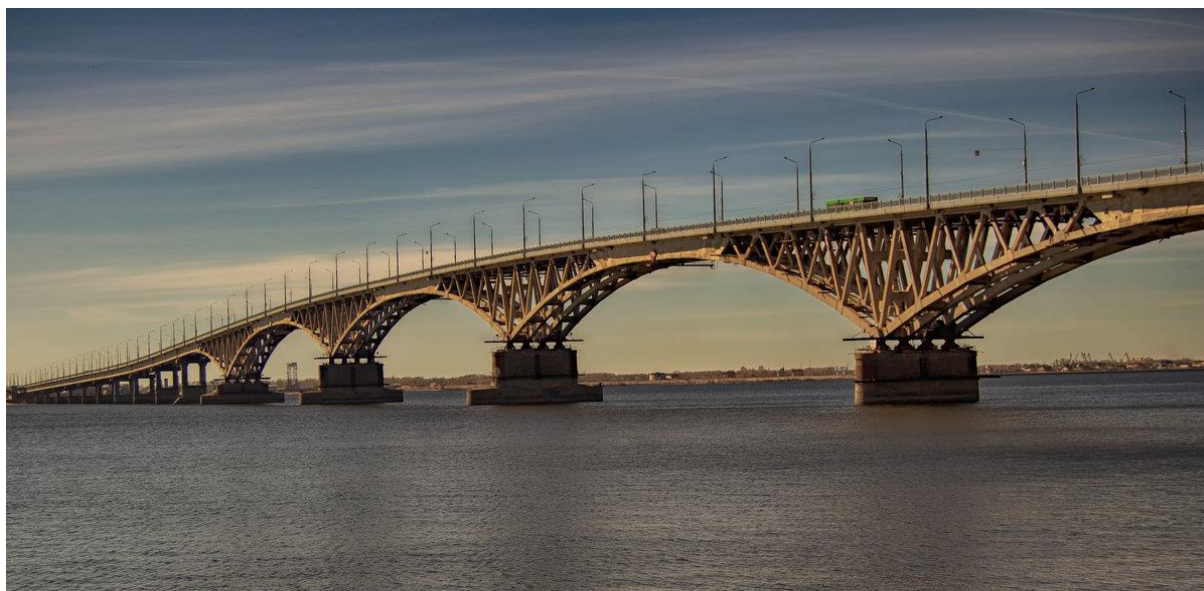**Drushbametrics Project**

Twelfth Winter Symposium on Chemometrics

Modern Methods of Data Analysis

Russia, Saratov, February 24–28, 2020

Russian Chemometrics Society

Saratov State University

Twelfth Winter Symposium on Chemometrics

Modern Methods of Data Analysis

The organizing committee

Co-chairmen

Natalia Burmistrova

Secretary

Yulia Monakhova

Members

Dmitry Kirsanov

Sergey Kucheryavskiy

Alexey Pomerantsev

Oxana Rodionova

Tatyana Rusanova

Address

Saratov State University, Astrakhanskaya Street 83, 410012 Saratov, Russia

https://wsc.chemometrics.ru/wsc12

e-mail: wsc12@chemometrics.ru

Thanks

The WSC-12 organizers and participants wish to express greatest appreciation to the following conference sponsors for their valuable economic and friendly help:

Saratov State University

Russian Chemometrics Society

Spectral Service

Chromos Engineering

American Elements

Finally, we are grateful to all the WSC-12 attendees, lecturers, accompanying persons, and visitors for their interest in the conference.

See you again at the next WSC-13 conference!

Useful information

Conference and activities

Conference sessions will be held in a conference hall (building 3, nr. 7 in the map below). In free time participants can try various winter activities: skiing, skating. In addition, swimming pool, sauna and billiard are also available.

Meals

All meals as well as the banquet will be served in the restaurant (5).

Scores & Loadings

Traditional "Scores and Loadings" meetings will be held in the bar (same building as the restaurant).

Communication

The main Russian cellular networks MTS, Beeline and Megafon have a proper coverage around the hotel. WiFi is available in all the buildings.

Money

You can exchange EUR and US dollars in Saratov banks. The organizing committee could also exchange a reasonable amount of currency.

Excursion

Friday morning all participants are invited to a Saratov city tour, where you will have a possibility to see the most impressive sites of the city.

Connection with Saratov

The recreation center Volzhskie Dali located on the bank of the Volga river in beautiful suburbs of city Saratov (15 km from the city center).

Besides the official conference transfer, you can reach the center by bus 236 from the stop "Radishchev street" (06:30, 09:00, 16:40, 18:20), the stop is "Volzhskie Dali".

The coordinates of the resort are:

Latitude 51°35'43.7"N (51.595477)

Longitude 46°11'09.4"E (46.185944)

Miscellaneous

The official conference language is English. Everyone is encouraged to have his/her badge attached, both during the symposium sessions and social activities.

Useful Phone Numbers

Natalia Burmistrova, local organizing committee +7 (927) 109-79-44

Tatyana Rusanova, +7 (927) 156-57-85

Volzhskie Dali reception +7 (8452) 41-17-11

Map of "Volzhskie Dali"

1. Bilding №1 (playroom for children - 10)
2. Bilding №2.
3. Bilding №3 (conference-hall – 7, hall for coffee-break – 8, pharmacy – 14, mud bath – 26, medical building – 28, admimstration - 29)
4. Parking.
5. Restaurant, karaoke-bar
6. Round conference-hall

9. Summer cafe
10. Playground for children
12. Skating ring
13. Shop
15. Fitness center (solarium – 16, aqua park - 17, sauna – 18, 19, sport hall – 20, 21, billiard - 22, SPA-27)
23. Promenade
24. Beach
25. Boat station

Monday, February 24, 2020

11:00-13:00	**Registration**
13:00-14:00	**Lunch**
14:00-15:00	**Opening**
Session 1	**Chair: Douglas Rutledge**
15:00-15:30	**T01** *Yurii Ishbulatov* Software adjustment for detection of circulating tumor cells using photoacoustic flow cytometry platform
15:30-16:00	**T02** *Andrey Samokhin* Evaluation of a pocket NIR-spectrometer: the case of oregano authentication
16:00-16:30	**Coffee break**
16:30-17:00	**T03** *Soraya Aidene* Expanding analytical capabilities of XRF spectrometry with chemometric tools
17:00-17:30	**T04** *János Elek* Use of near infrared spectroscopy in sports diagnostic examinations aiming personalized training methods
17:30-18:00	**T05** *Ivan Krylov* PARAFAC decompositions of Arctic seawater fluorescence data: numerical challenges and practical problems
18:00-18:30	**T06** *Polina Soboleva* Multi-spectroscopic techniques combined with multivariate analysis for heparin classification
18:30-19:00	**Free time**
19:00-20:00	**Dinner**
20:00-00:00	**Scores & Loadings**

Tuesday, February 25, 2020

08:00-09:00	**Breakfast**
Session 2	**Chair: Alejandro Olivieri**
09:00-10:00	**L01** *Douglas Rutledge* Novel extensions and applications of Common Components Analysis in chemometrics
10:00-10:30	**T07** *Alisa Rudnitskaya* Applications of the ASCA+ to the multivariate data with underlying unbalanced experimental design: the case study of the Portuguese grapes of Bairrada appellation
10:30-11:00	**Coffee break**
11:00-11:30	**T08** *Yulia Monakhova* Determination of the origin of acetaminophen tablets by spectroscopic methods and chemometrics
11:30-12:00	**T09** *Federico Marini* Advanced approaches for the analysis of multi-block data
12:00-12:30	**T10** *Yuri Kalambet* Peak sharpening by adding derivatives vs peak fitting: pro et contra
12:30-13:00	**Free time**
13:00-14:00	**Lunch**
Session 3	**Chair: Dmitry Kirsanov**
14:00-14:30	**L02** *Andrey Pakhomov* Development of chemometric database for chromatographic analysis. Practical application in industrial and laboratory analyzes
14:30-15:00	**T11** *Ekaterina Oleneva* Comparison of different multivariate model validation approaches on the spectroscopic data
15:00-15:30	**T12** *Victor Cardoso* Non-invasive method to identify the type of green tea inside teabag using NIR spectroscopy, support vector machines and Bayesian optimization
15:30-16:00	**T13** *Konstantin Brazovsky* Regression artificial neural network for modeling activation of NO synthase in peritoneal macrophages by humic acids
16:00-16:30	**Coffee break**

Session 4	Chair: Oxana Rodionova
16:30-17:00	**T14** *Victor Hugo Ferreira* Detection of adulterations in cachaça using GC×GC chromatographic images and DD-SIMCA
17:00-17:30	**T15** *Liana Zagitova* Multisensory voltammetric systems for recognition of medicines
17:30-18:00	**T16** *Anatoly Karavaev* Method for the analysis of synchronization between the contours of autonomic blood circulation control from long records
18:00-19:00	**Free time**
19:00-20:00	**Dinner**
20:00-00:00	**Scores & Loadings**

Wednesday, February 26, 2020

08:00-09:00	**Breakfast**
Session 5	**Chair: Federico Marini**
09:00-10:00	**L03** *Beata Walczak* Chemometrics in chromatography
10:00-10:30	**T17** *Sergey Kucheryavskiy* Procrustes cross-validation: reboot to strap validation again
10:30-11:00	**Coffee break**
11:00-11:30	**T18** *Dmitry Kirsanov* Analytical performance characterization of potentiometric multisensor systems
11:30-12:00	**T19** *Carlos Teixeira* Breaching the cross-contamination identification barrier in S. cerevisiae strains using SERS and DD-SIMCA
12:00-12:30	**T20** *Natalia Burmistrova* Chemometric tools for microstructured optical fibers spectral data processing
12:30-13:00	**Free time**
13:00-14:00	**Lunch**
Session 6	**Chair: Alexey Pomerantsev**
14:00-15:00	**L04** *Andrey Bogomolov* Optical multisensor systems and their analytical applications
15:00-15:30	**T21** *Xihui Bian* Ensemble modeling for the spectral analysis of complex samples
15:30-16:00	**T22** *Anastasiia Surkova* Direct standardization and slope and bias correction for calibration model transfer between 4-channel optical multisensor systems and full-scale spectrometer
16:00-16:30	**Coffee break**
Session 7	**Chair: Sergey Kucheryavskiy**
16:30-19:00	**Poster session**
19:00-20:00	**Dinner**
20:00-00:00	**Scores & Loadings**

Thursday, February 27, 2020

08:00-09:00	**Breakfast**
Session 8	**Chair: Beata Walczak**
09:00-10:00	**L05** *Alejandro Olivieri* A general view of multi-way data processing and its advantages
10:00-10:30	**T23** *Nikolai Sushkov* Use of non-negative matrix factorization (NMF) for an exploratory analysis of atomic and molecular spectra of zooplankton
10:30-11:00	**Coffee break**
11:00-11:30	**T24** *Zuzanna Małyjurek* Different strategies of SIMCA model optimization- a comparison study
11:30-12:00	**T25** *Ekaterina Martynko* Prediction of selectivity for potentiometric sensors with QSPR modeling
12:00-12:30	**T26** *Timur Akhmetzhanov* Chemometric approaches to direct analysis of lanthanides in REE-rich ores by TXRF and WDXRF
12:30-13:00	**Free time**
13:00-14:00	**Lunch**
Session 9	**Chair: János Elek**
14:00-14:30	**T27** *Oxana Rodionova* Calculation of total distance and outlier detection in the projection regression models
14:30-15:00	**T28** *Ivan Bratchenko* Multivariative analysis of Raman and autofluorescence spectra of human tissues
15:00-15:30	**T29** *Alexey Pomerantsev* SIMCA. Decision rules
15:30-16:00	**T30** *Ekaterina Borovkova* Development of the hardware-software complex for diagnosing the state of the human cardiovascular system based on non-linear connectivity indices
16:30-17:00	**Closing**
17:00-17:30	**Coffee break**
17:30-19:00	**Free time**
19:00-00:00	**Banquet**

Friday, February 28, 2020

08:00-09:00	**Breakfast**
09:00-10:00	**Check out**
10:00–15:00	**Bus to Saratov and excursion**

Abstracts

L01. Novel extensions and applications of Common Components Analysis in chemometrics

Douglas N. Rutledge[1,2]

[1]AgroParisTech, UMR 1145 Ingénierie Procédés Aliments, F-75005 Paris, France

[2]National Wine and Grape Industry Centre, Charles Sturt University, Wagga Wagga, Australia

When a set of samples is analysed by p (p>1) different analytical techniques, the result is a set of p matrices, often containing redundant and complementary information. Therefore, instead of analysing each matrix separately, a simultaneous analysis of all matrices can result in a single model reflecting all the information present in the p matrices. This is the objective of multi-block data analysis methods.

These methods facilitate the simultaneous analysis of p data tables, by extracting the structure which is common to all of them. The multiblock analysis method called Common Components and Specific Weight Analysis [1], abbreviated as ComDim, will be presented. The calculated latent variables, called Common Components (CC), contain the Global Scores; the contribution of each table to the construction of a CC is given by a measure called the salience; the Loadings and the Local Scores of each table can also be calculated [2].

Over the past few years, the ComDim method has been adapted in several different ways:

- A modification of ANOVA-PCA with ComDim: ANOVA-ComDim

In ANOVA-PCA [3], a set of 'Factor' and 'Interaction' matrices is created from a single matrix, by considering the different factors and interactions accounted for by the data. Each of the derived matrices is then analysed by PCA. In ANOVA-ComDim [4], these separate PCA analyses are replaced by a single multiple-block analysis.

- A ComDim- based regression method: PLS_ComDim

ComDim resembles Principal Component Analysis (PCA) in that it is an exploratory method based on the extraction of Components related to the dominant sources of variability in the tables. It can be modified to give a method we will call PLS_ComDim [5], which resembles Partial Least Squares regression (PLS) in that the extracted components are related both to the variance of the explanatory tables and their

covariance with another table containing values to be predicted. The application of PLS_ComDim to a metabolomic LC-MS data set will be presented.

- A ComDim- based MultiBlock Independent Components Analysis: ComDim-ICA

The classical ComDim algorithm is based on a Singular Value Decomposition (SVD) of a matrix which is the salience-weighted sum of all the sample "variance-covariance matrices" calculated from the data blocks.

A recent reformulation of the ComDim algorithm [6], based on the direct concatenation of the individual salience-weighted blocks, now makes it possible to replace the SVD by other multivariate decomposition methods such as ICA. This results in the Scores and Loadings of the classical SVD-based ComDim being replaced by ICA-based Proportions and Source Signals.

It will be shown that just as ICA is in many ways better than PCA for analyzing a single data table, ComDim -ICA is better than ComDim for analyzing several data tables.

References

1. E. Qannari, I. Wakeling, P. Courcoux, H.J.H MacFie, Defining the underlying sensory dimensions, Food Quality and Preference (2000) 11, 151–154.

2. V. Cariou, D. Jouan-Rimbaud Bouveresse E.M. Qannari, D.N. Rutledge, ComDim methods for the analysis of multiblock data in a data fusion perspective, in Data Fusion Methodology and Applications (Data Handling in Science and Technology), (ed. Marina Cocchi), Elsevier Science Publishers, Amsterdam, 2018.

3. P. de B. Harrington, N.E. Vieira, J. Espinoza, J.K. Nien, R. Romero, A.L. Yergey, Analysis of variance - Principal Component Analysis: a soft tool for proteomic discovery, Analytica Chimica Acta (2005) 544, 118-127.

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

5. A. El Ghaziri, V. Cariou, D. N. Rutledge, E. M. Qannari, Analysis of multiblock datasets using ComDim: Overview and extension to the analysis of (K + 1) datasets, Journal of Chemometrics, (2016) 30, 420-429.

6. M. Hanafi, E.M Qannari, Nouvelles propriétés de l'analyse en composantes communes et poids spécifiques, Journal de la société française de statistique, (2008) 149(2), 75-97.

L02. Development of chemometric database for chromatographic analysis. Practical application in industrial and laboratory analyzes

Andrey Pakhomov, Margarita Tyutlikova

Ltd. "Chromos Engineering", Nizhny Novgorod Region, Dzerzhinsk

Industrial companies technologists are responsible for the development of technological solutions, processes and the production of commercial products. Based on the technological requirements they take part in the choice and purchase of measuring instruments to provide the high-quality production.

The error of the measuring instruments used in the industry is determined by the normative documents and the department of metrology. However, due to technical problems (equipment wear, slow processes, the inability of process monitoring, etc.), technologists often neglect the accuracy of measurements and choose the measurement equipment with an error much higher than necessary. The result is a decrease in the quality of the final product.

Ltd "Chromos Engineering" is developing a software "Sonar-Analyst" that will provide an operator to control technological processes in real time and solve the above-mentioned quality problems.

"Sonar-Analyst" software implements the business process of unified information (cloud) digital control of key indicators of the industrial products quality, with elements of self-diagnostics of the reliability of indications, a hierarchy structure of informing line managers and top managers of the company, based on the measurement results obtained from streaming analytical and laboratory equipment. The "Sonar Analyst" software allows to:

− create a complex system for checking the analytical equipment correctness by comparing the data developed by a special department (competence center) with the data calculated by the device in the enterprise in online mode;
− inform managers and specialists of the company directly connected with the activity of analytical equipment about the results of testing devices;

- carry out the annual verification of equipment online without the specialist departure;
- store data and to carry out certification of products in the "cloud."

L03. Chemometrics in chromatography

Beata Walczak

Institute of Chemistry, University of Silesia 40-006 Katowice, 9 Szkolna Street, Poland

Chemometrics is considered as part of analytical chemistry. In its arsenal we can find methods that help analytical chemists to deal with all steps of analytical procedures, starting from the design of an experiment to extraction of information and to the final decision making. A majority of chemometrics methods are general, i.e. they can be applied to any type of analytical experiment and to any type of instrumental signal. However, there are problems associated with the specific types of instrumental signals or with the particular analytical techniques that need special treatment. Only with the knowledge of the system studied and of the principles of the performed measurements, the well suited methods can be chosen.

For instance, in separation science, many new possibilities are associated with the fact, that chromatograms can be treated as analytical signals and use for comparative analysis of different samples or systems in the same manner as spectra. In the arsenal of chemometrics methods we have all methods for signal enhancement, mixture analysis, warping, compression, visualization, calibration, classification, etc.

Different aspects of chromatographic data preprocessing and analysis will be discussed and illustrated for the simulated and real data sets.

Acknowledgement

L04. Optical multisensor systems and their analytical applications

Andrey Bogomolov

Endress+Hauser Conducta GmbH+Co. KG, Anton-Huber-Strasse 20, 73420 Aalen, Germany

Optical multisensor systems is a special class of low-resolution spectral analyzers. The main feature distinguishing them form traditional optical spectroscopy is intrinsic specialization for a particular practical application, which enables their miniaturization, portability and autonomic work at a significant price reduction. Due to these properties multisensor systems are increasingly used in the online-, field and personal analysis solving various analytical problems that could not be addressed by the traditional lab spectroscopy. Thus, optical multisensor systems introduce the chemical analysis in new areas and increase the level of analytical control in general.

Having much less (sometimes only a few) information channels compared to hundreds of wavelengths in conventional spectroscopy multisensor systems are capable of achieving the same or even better precision. This efficiency is reached due to the deep system optimization to the needs of analysis with the key role of data analysis. Chemometrics is applied both at the system development stage and to equip it with the working predictive model. Optical multisensor systems are often based on modern technical solutions, such as light-emitting diodes as the light sources and interferometric detectors.

This work presents a number of optical multisensor systems' development and application examples from different fields including food and pharmaceutical industry, biotechnology, ecological monitoring and medical diagnostics.

L05. A general view of multi-way data processing and its advantages

Alejandro C. Olivieri

Department of Analytical Chemistry, Faculty of Biochemical and Pharmaceutical Sciences, National University of Rosario, and Rosario Institute of Chemistry (IQUIR-CONICET), Suipacha 531 (2000) Rosario, Argentina

Multi-way data processing and calibration display outstanding properties from the analytical chemistry perspective. For example, by measuring chromatographic-spectral data matrices, it is possible to quantitate analytes in complex samples, without the need of a large training set. Only a few pure analyte standards are needed; after measuring their matrix data and joining them with those for the unknown sample, a multiway calibration model mathematically separates the analyte contribution from those of the interferences. This makes it unnecessary to apply sample pre-treatment or clean-up procedures, to achieve baseline resolution of every analyte peak or to correct for background signals. Chromatographic protocols become simpler, isocratic, cheaper, faster and, perhaps more importantly, greener. All these benefits stem from the second-order advantage, which is inherent to data second-order and beyond, i.e., matrices, three-dimensional arrays, etc.

In this lecture, the following topics will be covered: (1) data structures, (2) multi-way models and (3) experimental applications. Instrumental data will be classified according to the so-called "order" (number of modes for a single sample) and "ways" (number of modes for a sample set). More importantly, multi-way data need to be classified as multi-linear or non-multilinear, and in this case according to the cause for the multi-linearity loss. It will be shown that this latter classification scheme is highly helpful for selecting the correct data processing model.

A number of experimental reports will serve to illustrate the various data properties and the corresponding multi-way models employed for calibration. The examples include the use of matrix excitation-emission fluorescence spectroscopy and of liquid chromatography with multivariate spectral detection. Some cases involving four- and five-way calibrations will be shown, where a kinetic evolution or a pH gradient are implemented as additional instrumental modes.

T01. Software adjustment for detection of circulating tumor cells using photoacoustic flow cytometry platform

Yu.M. Ishbulatov [1,2], A.S .Karavaev [1,2,3], D.N. Bratashov [1], O.A. Inozemtseva [1],

V.P. Zharov [1,4]

[1]Saratov State University, Russia

[2]Institute of Cardiological Research, Saratov State Medical University, Russia [3]Saratov Branch of the Institute of RadioEngineering and Electronics of RAS, Russia

[4]University of Arkansas for Medical Sciences, USA

Oncology and melanoma in particular are among the most predominant causes of death in the whole world. Early diagnostics is a major problem since modern techniques for detection of circulating tumor cells (CTC) are not sensitive enough and undetected CTC are fast to cause metastasing, which dramatically decreases the probability of successful treatment.

Recently an innovative noninvasive technique of photoacoustic diagnostic was proposed to detect the CTCs [1], using laser pulses for generation of acoustic signals from CTCs and their detection with ultrasound transducers. Circulating melanoma cells has greater light absorption and therefore generate the acoustic signals of grater amplitude in comparison to blood cells. The proposed approach is capable to detect a single melanoma cells in the whole volume of blood and has much higher sensitivity [2].

The study aimed to adjust the parameters of the technique of melanoma cells detection. The technique was tested on the mice injected with melanoma cells culture (B16F10). We investigated how the filtration, threshold value and gating parameters change the sensitivity and specificity of the algorithm. After the parameters adjustment we managed to achieve the probability of singular melanoma cell detection of 0.53 and probability of false-positive detection lower than 0.001.

Acknowledgement

References

1. Galanzha E.I., Menyaev Y.A., Yadem A.C., Sarimollaoglu M,, Juratli M.A., Nedosekin D.A., Foster S.R., Jamshidi-Parsian A., Siegel E.R., Makhoul I., Hutchins L.F., Suen J.Y., Zharov V.P. In vivo liquid biopsy using Cytophone platform for photoacoustic detection of circulating tumor cells in patients with melanoma // Sci. Transl. Med. 11, eaat5857 (2019).

2. Galanzha E.I., Zharov V.P. Photoacoustic flow cytometry // Methods. 57(3) 280-296 (2012).

T02. Evaluation of a pocket NIR-spectrometer: the case of oregano authentication

Andrey Samokhin[1], Oxana Rodionova[2], Alexey Pomerantsev[2]

[1]Division of Analytical Chemistry, Chemistry Department, Lomonosov Moscow State University, Moscow, Russia

[2]N.N.Semenov Federal Research Center for Chemical Physics RAS, Moscow, Russia

In this work, we evaluated performance of a pocket NIR-spectrometer (SCiO, Consumer Physics) for differentiating oregano from the related plant species and some known adulterants. Previously the performance of this pocket device was assessed by various groups in application to such tasks as detection of counterfeit tablets, the freshness assessment of various foods, crop cultivar identification, fruit quality evaluation, determination of protein content and etc.

More than 20 oregano samples are investigated in our study. They were purchased in shops, marketplaces, and drugstores in Russia and in Europe. The samples are originated from different regions (Greece, Italy, Russia, Spain, Turkey). Alternative classes are the related plant species (basil, marjoram, mint, rosemary, thyme), known adulterants (olive and hazelnut leaves), and some other plants (celandine, nettle, parsley).

Two series of three replicate measurements are performed using SCiO instrument on different days. To reduce the amount of sample material required, a 3D-printed container of about 9 cm^3 in volume is used. Spectra are recorded in the range of 740–1070 nm with the step of 1 nm. Benchtop NIR-spectrometer Nicolet 6700 (Thermo

Fisher) is used as a reference instrument. In this case, spectra are recorded using a fiber probe in the range of 12000–4000 cm^{-1} with 8 cm^{-1} resolution.

A number of preprocessing methods are tested (including data reduction, smoothing, standard normal variate, derivative calculation and their combinations). DD-SIMCA approach is used for the target class modeling.

T03. Expanding analytical capabilities of XRF spectrometry with chemometric tools

S. Aidene[1], V.V. Panchuk[1, 2], V.G. Semenov[1, 2], D.O. Kirsanov[1]

[1]Institute of Chemistry, St. Petersburg State University, St. Petersburg, Russia

[2]Institute for Analytical Instrumentation RAS, St. Petersburg, Russia

Quantification of organic substances by Energy Dispersive X-ray Fluorescence (EDXRF) is considered to be a difficult task since EDX spectrometers cannot provide quantification of the elements with atomic numbers below 11 (sodium). In case of the light elements, the incident X-ray beam is mainly scattered and it is found in the form of coherent and incoherent peaks in addition to fluorescent lines in the EDXRF spectrum. The intensity of scattered radiation is influenced by average mass absorption coefficient of the sample, which in turn depends on numerous fcators: elemental composition of the sample; density of the sample; geometry of the experiment; on the scatter factors [1]. Thus, this region gives powerful properties for quantitative determination of light elements when it is combined with chemometric tools. The possibility of quantitative determination of light elements in organic compounds using scattered radiation was already demonstrated in [2, 3]. The present study deals with the comparison of performance of Rh X-ray tube induced and monochromatic induced energy dispersive X-ray fluorescence in generating scattering profiles when plastic samples are irradiated. The resulting spectra were processed with partial least square (PLS) and it was possible to construct predictive models relating EDX signals with the content of organic elements C% and H% and integral physical properties (density, mass per atom ,water absorption). The presentation will show the use of chemometric tools to expand analytical capabilities of XRF.

Acknowledgement

References

1. Espen, P.V., dack, L.V., Adams, F., Grieken, R.V.Effective Sample Weight from Scatter Peaks in Energy-Dispersive X-Ray Fluorescence(1979) Analytical Chemistry, 51 (7), pp. 961-967.

2. Goraieb, K., Alexandre, T.L., Bueno, M.I.M.S. X-ray spectrometry and chemometrics in sugar classification, correlation with degree of sweetness and specific rotation of polarized light (2007) Analytica Chimica Acta, 595 (1-2 SPEC. ISS.), pp. 170-175.

3. Pereira, F.M.V., Bueno, M.I.M.S. Calibration of paint and varnish properties: Potentialities using X-ray Spectroscopy and Partial Least Squares (2008) Chemometrics and Intelligent Laboratory Systems, 92 (2), pp. 131-137.

T04. Use of near infrared spectroscopy in sports diagnostic examinations aiming personalized training methods

J. Elek [1], E. Markovics[2], Zs. Komka[3], M. Szász[4]

[1]Science Port Ltd

[2]University of Debrecen

[3]University of Physical Education

[4]Scitec Institute

The daily water intake is a very important generally, but has a pronounced importance in case of professional sportsmen. The hydration of the nervous system plays the more important role in the concentration efforts, thus has many times more impact on the sport performance under extreme conditions than the actual state of the muscles.

In a previous study near infrared spectra from different points of the body of male and female volunteers in different ages were recorded and evaluated. The volunteers were asked not to drink for a certain time, then they drunk water or

isotonic drinks, and the NIR spectra were recorded and evaluated: visualization of the hidden patterns was carried out by principal component analysis, while the conclusions of these visual observations were checked by ANOVA. Isotonic drinks were also tested and the hydration effect was compared to the use of pure water.

As mentioned, for the sportsmen the fatigue does not only appear due to the exhaustion of the muscles, but the dehydration of the nervous system which is essential for a focused concentration not only the absorption, but the retention of the water is particularly important. We have examined the effect of using isotonic drinks instead of water on the ability of the body to retain the water under physical/thermal stress.

The extreme stress tests were carried out in the framework of a cooperation with the Hungarian Olympic Canoe Team and Scitec Institute where the traditional sport diagnostic parameters (such as cardiovascular and respiratory parameters, blood chemistry, etc.) and the NIR infrared spectra were recorded during the experimental phases.

A very typical figure of PC1 extracted from the NIR data vs. time is shown below. The maximal strain was put on the volunteers at 25 min, where their vascular system was practically collapsed. The collapse and the restitution can be excellently followed by analysing the NIR spectra.
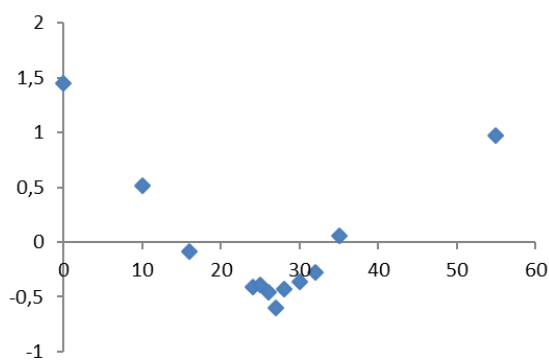

Figure 1. PC1 vs. sampling time.

Our recent efforts target the mapping of the relationship between the principal components extracted and the medical parameters recorded during the study. The presentation gives an insight of the recent state of these researches.

T05. PARAFAC decompositions of Arctic seawater fluorescence data: numerical challenges and practical problems

Ivan Krylov[1], Anastasia Drozdova[2], Timur Labutin[1]

[1]Department of Chemistry, Lomonosov Moscow State University, Moscow, Russia

[2]Shirshov Institute of Oceanology, Russian Academy of Sciences, Moscow, Russia

Fluorescence excitation-emission matrices (EEMs) combined with parallel factor analysis (PARAFAC) are a powerful tool for studying fluorescence properties of a mixture of fluorophores. One of the most common fields of the use of EEMs decomposition by PARAFAC is the study of distribution, dynamics and transformation of natural dissolved organic matter (DOM).

While detailed characterisation of DOM is usually achievable by means of high-resolution mass-spectroscopy, practical constraints (such as inability to ship a mass spectrometer on a polar research vessel or impracticality of taking 5-10 l of seawater samples for pre-concentration) result in the widespread use of less informative, but much more performant optical techniques combined with chemometrics.

In this work, 80 samples of DOM from shelf seas were collected during the cruises to the Kara, Laptev, White, and East Siberian seas in autumn (2015-2017) and spring (2016). A few samples were taken from freshwater ponds of Novaya Zemlya archipelago. Since the real world fluorescence data is subject to inner filter effect, scaling problems, and interference from scattering signal, these effects were accounted for by means of, respectively, absorbance-based correction, normalization to Raman peak belonging to water, and interpolation of regions affected by scattering signal.

Being an ill-posed unsupervised learning problem, canonical tensor decomposition is hard to perform properly: comparing two solutions is complicated by the fact that the values are defined up to a constant multiplier and can be arbitrarily shuffled. Stability of the PARAFAC solutions is investigated by means of the split-half method and by comparing results of multiple random restarts of the model. Effect of various preprocessing steps and algorithm parameters (such as number of degrees of freedom and termination criterion) on the stability and meaningfulness of decompositions is investigated.

Results from different software packages implementing the PARAFAC algorithm ("nwaytoolbox" MATLAB package and "multiway" R package) are compared. Potential to employ line search in addition to already implemented alternating least squares to speed up the decomposition computation in the "multiway" R package is also investigated.

Acknowledgement

T06. Multi-spectroscopic techniques combined with multivariate analysis for heparin authentication

Polina M. Soboleva [1], _Natalia A. Burmistrova [1]_, _Yulia B. Monakhova [1,2]_

[1]Institute of Chemistry, Saratov State University, Saratov, Russia

[2]Spectral Service AG, Cologne, Germany

Heparin is widespread medicine applied as an anticoagulant drug for the treatment and prevention of various thrombotic diseases. Contamination and overdose of pharmaceutical heparin have a significant impact on the results of its use in clinical practice, which requires proper quality control of pharmaceutical heparin. Heparin is a natural polysaccharide that belonged to glycosaminoglycans (GAG) and consists of alternating amino-sugars and uronic acids [1]. The structure of heparin is variable and depends both on the animal source and on the specific production process. Unstable composition of heparin makes difficult its quality certification.

Currently, numerous methods have been published for the quality control of heparin [2]. One way of heparin analysis is spectroscopy, which allows performing non-destructive, fast and sensitive analysis with minimal sample preparation. Moreover, spectroscopy provides direct structure-related information. Nowadays preference is given to the most sensitive and reliable method of analysis - NMR spectroscopy, however, other spectrometric techniques combined with chemometrics algorithms can be applied as screening tools and could be of great importance [3].

This study demonstrated the possibility of using various spectroscopic methods in combination with chemometric analysis without sample pre-treatment to classify samples according to the purity and origin. We tested 124 heparin samples from various sources and purification obtained in powdered form from different

manufacturers. Heparin samples were obtained from Spectral Service's routine quality control laboratory (Cologne, Germany) and were analyzed by 1H-NMR spectroscopy. Several spectroscopic methods (UV-, fluorescence and IR-spectroscopy) were compared to determine which method is the most suitable for analysis. Chemometric analysis of the spectral data was performed using Principal Component Analysis (PCA), SIMCA (Soft Independent Modeling of Class Analogy), PLS-DA (Partial Least Squares Discriminant Analysis) and k-means algorithm.

Therefore, a combination of spectroscopic and chemometric methods makes it possible to authentication heparin by purity and animal source. FTIR spectroscopy was more efficient in obtaining information from spectra, but it requires pre-processing of spectral data, which allows increasing the differences between spectra. The best result of FTIR spectral data analysis was obtained using PLS method and value of sensitivity amounted 97%.

Acknowledgement

References

1. Casu, B. Structure and Biological Activity of Heparin. Adv. Carbohydr. Chem. Biochem. 1985, 43, 51–134.

2. Tian, L., Hu, X., Liu, Z. Studies of the interaction of heparin with lysozyme by multi-spectroscopic techniques and atomic force microscopy // Spectrochim. Acta A. 2016, vol.154, p. 27-32.

3. Devlin, A., Mycroft-West, C., Procter, P. Tools for the quality control of pharmaceutical heparin// Medicina (ISSN 1010-660X). 2019, vol. 55.

T07. Applications of the ASCA+ to the multivariate data with underlying unbalanced experimental design: the case study of the Portuguese grapes of Bairrada appellation

Jéssica Marques[1], <u>Alisa Rudnitskaya[2]</u>, Ana Raquel Santos[3], José Pedro Soares[3], Sílvia M. Rocha[1]

[1]Department of Chemistry & QOPNA, University of Aveiro, 3810-193 Aveiro, Portugal
[2]Comissão Vitivinícola da Bairrada, 3780-203 Anadia, Portugal

[3]Department of Chemistry & CESAM, University of Aveiro, 3810-193 Aveiro, Portugal

Multivariate data sets with underlying experimental design are becoming more and more common with advent of omics technologies such as metabolomics, proteomics, etc. Metabolomics data sets typically comprise hundreds of metabolites measured in tenths of samples, thus, requiring use of multivariate techniques for the analysis of effects of design factors. Several approaches combining ANOVA with multivariate data analysis were suggested in the literature, of which the most popular are ANOVA - Simultaneous Component Analysis (ASCA) and ANOVA - Principle Component Analysis (APCA). Generalization of these methods using General Linear Models (GLM), named ASCA+ and APCA+, allowing to analyze data with unbalanced designs has been proposed [1].

Sustainable viticulture and winemaking continue to represent challenging task, for which knowledge about the interactions between autochthonous grape varieties and climate and vineyard conditions is indispensable. Taking advantages of chemometric tools, this research aims to provide insights into these interactions. Thus, four white grape varieties (Arinto, Cercial, Bical and Maria Gomes, *Vitis vinifera L.*), from the Portuguese Bairrada Appellation, were selected as case study. For each variety, grapes from two vineyards were collected during two consecutive harvests (2017 and 2018) and analyzed with respect to their physical-chemical parameters and volatile aroma compounds both free and glycosidically bound. Effects of the vineyards and harvest year on the grape composition were assessed using ASCA+. The results showed that each grape variety presents a specific pattern of interaction with the ecosystem under study. The approach can be used to hierarchize impact of the different factors on the grape quality and estimate

adaptability of the four grape varieties serving for development of tools for efficient management and rational use of vines.

Acknowledgement

References

1. M. Thiel, B. Féraud, B. Govaerts, J. Chemometr. 31 (2017) e2895.

T08. Determination of the origin of acetaminophen tablets by spectroscopic methods and chemometrics

Yulia B. Monakhova[1,2], Alexander Becht[3], Curd Schollmayer[3], Ulrike Holzgrabe[3], Douglas Rutledge[4,5]

[1]Spectral Service AG, Emil-Hoffmann-Straße 33, 50996, Cologne, Germany

[2]Institute of Chemistry, Saratov State University, Astrakhanskaya Street 83, 410012 Saratov, Russia

[3]University of Würzburg, Institute for Pharmacy and Food Chemistry, 97074 Würzburg, Germany

[4]UMR Ingénierie Procédés Aliments, AgroParisTech, INRA, Université Paris-Saclay, Massy, France

[5]National Wine and Grape Industry Centre, Charles Sturt University, Wagga Wagga, Australia

Nowadays it is very difficult to determine the origin of a drug, since many different companies can be involved in its manufacturing. The excipients, active pharmaceutical ingredient or even the finished product can come from different manufacturing plants and suppliers from all over the world without the need of declaration.

The aim of this research was to identify the origin of drugs and the country of their process plant, respectively, solely based on measured spectra with the help of chemometric methods. Since paracetamol tablets are widespread and easy to acquire, they have been used here as a model product. 56 paracetamol drug samples of 50 pharmaceutical companies from all parts of the globe were analyzed. Three data sets were created using NIR, IR, and NMR spectroscopy. With the help of principal component analysis (PCA), cluster analysis (CA), discriminant analysis (DA), independent component analysis (ICA), common components and specific weights analysis (CCSWA or ComDim), the data were screened for patterns which allow a tracing of the tablets back to their origin.

It was shown that the chemometric evaluation of IR, NIR and NMR spectra using PCA was able to determine the origin of many drugs and their manufacturers. In addition, it was possible to identify relationships between companies and suppliers and to identify the existence of major differences in the formulations.

Acknowledgement

T09. Advanced approaches for the analysis of multi-block data

Federico Marini

Dept. of Chemistry, University of Rome "La Sapienza", P.le Aldo Moro 5, Rome, Italy

With the growing complexity of analytical problems, especially in multidisciplinary fields such as the –omic sciences, food authentication and traceability and forensics, and the increasing availability of high throughput techniques which allow the investigation of many samples per day, it is not infrequent that a matrix be characterized by multiple platforms. In this context, it may be useful to use the possibility to acquire signals from more than one instrumental technique on each sample and, therefore, to use the combined information from the two (or more) techniques to build the final chemometric model, be it exploratory or predictive: such an approach is generally called multi-block data analysis or "data-fusion." In this way, especially if the instruments chosen possess complementary features, it is possible to benefit from the advantages and the specific characteristics of each, to create a final model that is more reliable and robust.

In this communication, together with some more conventional application of data-fusion methodologies to the study of chemical systems, some novel approaches for the analysis of multi-block data will be presented. In particular, the possibility of implementing non-linearity in predictive modeling (classification and regression) by extending the locally linear modeling paradigm to multiple blocks will be discussed. Moreover, a recently developed method for variable selection in the context of multi-block data, namely the Sequential and Orthogonalized Covariance Selection method (SO-COVSEL) [1] will also be introduced and discussed. Finally, the possibility of extending the SIMCA class-modeling approaches to the multi-block configuration will also be presented.

The advantages and limitations of these methods will be illustrated by means of simulated and real-world examples taken from food characterization and metabolomics.

T10. Peak sharpening by adding derivatives vs peak fitting: pro et contra

Yuri Kalambet[1], Yuri Kozmin[2]

[1]Ampersand Ltd, Moscow, Russia

[2]Shemyakin Institute of Bioorganic Chemistry, Moscow, Russia

Analysis of the technology of peak sharpening by summing up signal and its derivatives is presented. Adding first derivative can eliminate exponential trend introduced in the data by measuring apparatus/process. As an example, Exponentially Modified Gaussian (EMG) can be used to generate original Gaussian by the operation EMG+τ·EMG'=Gaussian. This sharpening technology is rather old and is re-invented recently already third time, period of re-invention being around 27 years.

We found additional proof of the not-very-widely-known statement, that any exponentially modified (smoothed) function can be reconstructed by summing up the smoothed image and its derivative. We made further study of EMG peak shape, including very fast EMG peak generation by exponentially weighted moving average (EWMA) smoothing of Gaussian peak. Peaks with multiple exponential modifications are studied.

The price of peak sharpening is noise amplification. We evaluated noise amplification coefficients for adding the derivatives and discuss a way to optimize the resulting signal-to noise ratio.

Peak sharpening technology is compared to peak fitting technology by peaks of EMG or any pre-defined shape. Multi-channel peak fitting technology is presented.

T11. Comparison of different multivariate model validation approaches on the spectroscopic data

Ekaterina Oleneva[1,2], Dmitry Kirsanov[1,2], Andrey Panchenko[3], Ekaterina Gubareva[3], Olga Bibikova[4], Viacheslav Artyushenko[4], Urszula Zabarylo[5,6], Andrey Bogomolov[7]

[1]Institute of Chemistry, St. Petersburg State University, St. Petersburg, Russia

[2]Laboratory of artificial sensory systems, ITMO University, St. Petersburg, Russia

[3]Laboratory of carcinogenesis and gerontology, FSBI «Petrov Research Institute of Oncology» of the Ministry of Healthcare of the Russian Federation, St. Petersburg, Russia

[4]Art photonics GmbH, Berlin, Germany

[5]Center for Diagnostic and Interventional Radiology and Nuclear Medicine, Charité – Universitätsmedizin Berlin, Berlin, Germany

[6]Berlin Brandenburg School of Regenerative Therapies, Charité – Universitätsmedizin Berlin, Berlin, Germany

[7]Laboratory of Multivariate Analysis and Global Modelling, Samara State Technical University, Samara, Russian Federation

In recent decades, numerous spectroscopic techniques have been applied for the distinction of tumor cells from normal ones for various clinical purposes, e.g., early diagnostics of cancer, determination of cancer subtype or intraoperative assessment of tumor margins. Usually, the structure of the obtained spectra of different tissues is complex that requires the application of machine learning methods for data processing. In the majority of the published studies, a certain multivariate classification algorithm combined with spectral data preprocessing was used for data treatment. This approach raises an important question: how the authors should validate their classification models? In the review of Sattlecker et al., published in

2014 [1], it is underlined that an external test set was used for validation only in 22% of publications concerned with the application of spectroscopy for cancer detection. Thus, which strategy of choosing an independent test set gives the most realistic results? In this work, two spectral data sets in near-infrared (NIR) and middle-infrared (MIR) regions were used for comparison of different model validation approaches. The former contains 238 spectra measured in several points (tumor sites and intact zones) of 56 paraffin blocks with rat brain tissues affected by glioma C6, a popular murine model of carcinogenesis. The latter consists of 100 spectra of fresh stomach tissues obtained from the patients with stomach cancer. Such well-known chemometric classification methods as linear discriminant analysis (LDA) and support vector machines (SVM) with different kernel functions were used for the data processing. Further, the optimized classification models were validated by several approaches: full cross-validation (FCV), a single independent test (randomly selected, chosen by Kennard-Stone and duplex algorithms) and by n-fold random splitting on train and test sets procedure. The sensitivity, specificity, and accuracy of each classification model were calculated. The final results and conclusions will be discussed during the presentation.

Acknowledgement

References

1. M. Sattlecker et al., Current trends in machine-learning methods applied to spectroscopic cancer diagnosis, Trends in Analytical Chemistry 59 (2014) 17–25.

T12. Non-invasive method to identify the type of green tea inside teabag using NIR spectroscopy, support vector machines and Bayesian optimization

Victor G. K. Cardoso, _Ronei J. Poppi_

Institute of Chemistry, University of Campinas, Campinas, Brazil

Green tea is amongst the most popular non-alcoholic beverages and it is largely consumed in Asiatic countries. Beyond its flavor, recent studies have shown that it bears veterinary, medicinal, thermogenic, antioxidant and antiaging proprieties due to caffeine and polyphenols [1,2]. Hence, ensuring an efficient quality control is necessary, preferably using clean, fast, non-destructive and non-invasive analysis methods [3]. Near-infrared (NIR) spectroscopy meets the requirements above and has been applied in several food analysis [4]. Therefore, our research group has developed a non-invasive method to identify four types of green tea inside the teabag.

Among several multivariate classification methods, machine learning techniques have excelled when applied to solve complex and non-linear problems. In this work, we have applied support vector machines discriminant analysis (SVM-DA) to identify tea classes that outperformed results from partial least squares discriminant analysis (PLS-DA), the most famous discriminant analysis technique. SVM-DA provided high values of accuracy in class discrimination, showing an excellent performance. Furthermore, we have applied Bayesian optimization to obtain the best parameters to train the SVM-DA model. This aproach has never been reported in analytical chemistry problems and presents a higher performance compared to other optimization techniques [5]. SVM-DA models were trained using spectra carried out in benchtop and handheld instruments, which were statistically compared. This comparison showed that although both instruments performed well with the tea-outside-teabag analysis, the teabag slightly affects the results for the handheld equipment.

We believe that this study has great potential in industrial application, which can be adapted as in-line or on-line process [6].

Acknowledgement

References

1. Zhuang, X. G., et al. J. Appl. Spectrosc. 86, 76–82 (2019).

2. Li, C. et al. Spectrochim. Acta - Part A Mol. Biomol. Spectrosc. 206, 254–262 (2019).

3. Li, X., et al. Food Chem. 270, 236–242 (2019).

4. Ye, Z.-H. et al. J. Anal. Methods Chem. 2019:1537568, 1–8 (2019).

5. Czarnecki, W. M., et al. J. Cheminform. 7:38, 1–15 (2015).

6. Porep, J. U., et al. Trends Food Sci. Technol. 46, 211–230 (2015).

T13. Regression artificial neural network for modeling activation of NO synthase in peritoneal macrophages by humic acids

Konstantin S. Brazovsky[1,4], Maria V. Zykova[1], Elena E. Veretennikova[2], Marina G. Danilets[3], Sergey V. Romanenko[1,4], Lyudmila A. Logvinova[1], Evgenia S. Trofimova[4], Anastasia A. Ligacheva[4], Kristina A. Bratishko[1], Mekhman S. Yusubov[1,4], Alexey A. Lyapkov[4], Michael V. Belousov[1]

[1]Siberian State Medical University, Tomsk, 634050, Russia

[2]Institute of Monitoring of Climatic and Ecological Systems RAS

[3]Goldberg Research Institute of Pharmacology and Regenerative Medicine, Tomsk NRMC, Tomsk, 634050, Russia

[4]National Research Tomsk Polytechnic University, Tomsk, 634050, Russia

Studying, modeling and predicting biological activity of very complex organic substances is a rather challenging problem. Although such substances are widely used as food additives with alleged positive biological effects, the intimate mechanisms of their activity have not been studied yet. Mainly this is due to the

extreme wide variety of possible metabolic pathways these molecules can interact with. For example, naturally formed humic substances reportedly have both pro- and anti-inflammatory activities, anti-oxidation, cardio-protective and immuno-modulating effects, but when it comes to pharmacological standardization, it is almost impossible to reproduce those effects with required precision and accuracy.

As a result, whole classes of naturally formed organic substances with very desirable pharmacological properties are out of consideration just because there is no obvious way to authenticate their structure and biological activity. Commonly agreed approach to standardize pharmacological substances is to describe their exact chemical composition and molecular structure. Unfortunately, this approach does not work for very complex naturally formed supra-molecules like humic substances.

The alternative option to estimate biological effects of such molecules is studying their activity with animal and cell culture models. Though reasonable it seems, this approach has not been accepted for many years due to lack of reproducibility and low precision.

Recently, a computational (*in silico*) approach to model and predict possible biological effects of very complex organic substances is taking gain. The main advantage of this approach is virtually limitless possibilities to set parameters of the models and initial conditions. On the other hand, the numerical models require strict validation against reference dataset to prove acceptable accuracy and precision.

In practice, however, creating comprehensive reference datasets is far beyond the capabilities of even largest laboratories.

The Quantitative Structure-Activty Relation (QSAR) approach could be considered as a good compromise between biological and numerical modeling because QSAR takes the best from two alternatives. Connecting chemical structure of the substances and their biological effects, QSAR exploits very useful way to design new drugs and predict bioactivity without strict requirements to the reference database.

The present research is aimed to propose a new approach to model and predict possible biological activity of complex supra-molecules. As an example, the immunotropic effects of humic acids on mices' peritoneal macrophages was successfully modeled.

The proposed approach makes use of an artificial neural network (ANN) modeling multiparametric QSAR, where "structure" (S) is described with infrared spectral parameters, while "activity" (A) is characterized through production of nitrogen oxide by activated peritoneal macrophages.

Among many parameters, which can be used to describe chemical compositions, infrared (IR) spectrum allows building the most suitable and easy-to-measure parametric set. The artificial neural network works here as an universal approximator connecting IR parameters with biological activity.

The ANN was built using teaching dataset comprised of 32 samples, while testing dataset consisted 10 samples. The overall performance of ANN, characterized by the coefficient of determination R^2, reaches 98 %.

The proposed approach to model and predict biological activity of very complex organic molecules, exploiting general QSAR paradimg, demonstrates quite good overall performance while being simple and reproducible.

T14. Detection of adulterations in cachaça using GC×GC chromatographic images and DD-SIMCA

Victor H. C. Ferreira, Leandro W. Hantao, Ronei J. Poppi

Chemistry Institute, University of Campinas, São Paulo, Brazil

Cachaça is the most renowned brazilian spirit. It is made from the distillation of sugarcane's fermented must and nowadays it is consumed over the world. Its production may involve a last step of aging in wooden barrels or not. Many different types of wood barrels are used for aging cachaças, from the conventional oak (*Quercus sp.*) barrels to native brazilian woods barrels, made from balsam (*Myroxylon peruiferum*) or amburana (*Amburana cearensis*). The diversity of production procedures results in distinct organoleptic sensations, which enhance the product's value. Adulteration is then a major problem and its identification is of great importance. As sensory properties are related to the products volatile organic compounds (VOC), gas chromatography is a suitable tool to monitor these adulterations. For better results in separating VOC, and thus enhancing adulterants discovery, a two-dimensional gas chromatography (GC×GC) may be an attractive alternative. Chromatograms from GC×GC are visualized in two dimensions as

colored contour plots, with one axis related to the retention time in the first column and the other one to the retention time in the second column. As the chromatograms are very reproducible, they may be regarded as pictures taken from samples' VOC and as such may be treated as a fingerprint image of the samples. These fingerprint RGB images should also be useful to identify adulterations in a more statistical approach. The discrimination tool chosen for this task was Data-Driven Soft Independent Modeling of Class Analogy (DD-SIMCA) [1]. To exemplify the proposed method, a brand of cachaça aged in balsam was used, 20 different batches of unadulterated samples were analyzed as well as 49 samples of lab-adulterated ones. Adulteration of 10% to 50% were performed with non-aged cachaças, cachaças aged in different wood barrels or with younger aged cachaças. No false positives or false negatives were found, which showcases that such approach is a promising tool to detect cachaça adulteration.

Acknowledgement

References

1. Zontov, Y.V., et al., Chemom. Intell. Lab. Syst., 167, 2017, 23-28.

T15. Multisensory voltammetric systems for recognition of medicines

L.R. Zagitova, Yu.A. Yarkaeva, D.I. Dubrovsky, R.A. Zilberg, V.N. Maistrenko

Bashkir State University, Ufa, Russia

The creation of sensors for the express recognition of enantiomers of drug compounds is one of the main tasks of analytical chemistry. It is known that the use of pure enantiomers in many cases can improve the quality of medicines due to the more specific effect and the reduction of side effects. The use of voltammetric sensors for quality control and recognition of optically active medicines is currently limited due to the lack of selectivity and a limited number of determined

components. The development of a multisensory approach and solutions for chemically modified electrodes and chemometrics makes it possible to consider voltammetric sensors as a promising direction in creating fairly simple and cheap express methods for recognizing medicines containing optically active compounds [1]. There are no universal electrochemical sensors because of the complexity of molecular recognition of individual enantiomers. However, manifestations of close electrochemical characteristics for electrodes with various chiral selectors differing in the value of the analytical signal for related optically active compounds can be expected. This provides the condition of cross sensitivity, necessary for the functioning of multisensory voltammetric systems of the "electronic tongue" type. To increase the sensitivity and selectivity of voltammetric sensors, chemometric methods (PCA, PLS-DA, SIMCA) are used.

This report addresses the electrochemical behavior of the enantiomers of medicines (atenolol, propranolol, methionine, tyrosine, warfarin) and their recognition using voltammetric sensors with various selectors (polymer composites based on polyarylenephthalides with cyclodextrins, a composite based on chitosan and cyclodextrins, melamine, cyanuric acid, 3,4,9,10-perylene tetracarboxylic acid). The possibilities of chemometric methods for processing multidimensional data in multisensory systems of the "electronic tongue" type were evaluated. Such devices can solve problems associated not only with the recognition of enantiomers, but also with the establishment of the authenticity and origin of the drug, and the recognition of the drug by shelf life.

Acknowledgement

References

1. V.N. Maistrenko, A.V. Sidel'nikov, R.A. Zil'berg, Enantioselective voltammetric sensors: new solutions, J. Anal. Chem. 73 (2018) 1-9.

T16. Method for the analysis of synchronization between the contours of autonomic blood circulation control from long records

A.S. Karavaev [1,2,3], V.V. Skazkina [1,2], E.I. Borovkova [1,2,3], Y.M. Ishbulatov [1,2,3], A.R. Kiselev [3], Gridnev V.I. [3], V.I. Ponomarenko [1,2], M.D. Prokhorov [1,2], B.P. Bezruchko [1,2]

[1]Saratov Branch of the Institute of RadioEngineering and Electronics of RAS, Russia

[2]Saratov State University, Russia

[3]Institute of Cardiological Research, Saratov State Medical University, Russia

Our previous studies have shown the importance of analyzing the synchronization of the contours of blood circulation for understanding the fundamental features of cardiovascular system [1] and solving the problems of medical diagnostics [2] when analyzing short signals [3]. In this paper, we analyze this synchronization using long records of electrocardiograms and photoplethysmograms.

In this paper, we studied the statistical properties of the calculated synchronization index, including an estimationg of the level of statistical significance, correlation and spectral analysis, and an estimationg of the probability distribution density. The dynamics of the index over time was also studied, including in the low-frequency range, which is associated with chemical processes of humoral regulation [4].

Analysis of the synchronization of the contours of autonomic control in a sliding window showed that the dependence of the proposed index on time shows small low-frequency fluctuations, which are apparently associated with the modulation of the activity of the neural control loops from the slow processes of humoral regulation.

Acknowledgement

References

1. Karavaev A.S. et al. Autonomic control is a source of dynamical chaos in the cardiovascular system // CHAOS. 2019. V. 29. P. 121101.

2. Kiselev A.R. et al. A comprehensive assessment of cardiovascular autonomic control using photoplethysmograms recorded from earlobe and fingers // Physiological Measurement. 2016. V. 37. P. 580-595.

3. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, Heart rate variability: standards of measurement, physiological interpretation, and clinical use// Circulation. 1996. V. 93. P. 1043–1065.

4. Karavaev A.S. et al. Synchronization of low-frequency oscillations in the human cardiovascular system // Chaos. 2009. V.19. P. 033112.

T17. Procrustes cross-validation: reboot to strap validation again

Sergey Kucheryavskiy[1], Sergei Zhilin[2], Oxana E. Rodionova[3], Alexey Pomerantsev[3]

[1]Aalborg University, Department of Chemistry and Bioscience, Denmark

[2]Altai State University, Barnaul, Russia

[3]Semenov Federal Research Center for Chemical Physics, RAS, Moscow, Russia

Model validation is an important part of chemometric analysis. A large amount of research has been devoted to comparing of different techniques with a goal to find the best one. In general, there are two options — the test set and cross-validation. The former approach has been proven to be the most reliable for assessment the performance of a final model, while the latter is usually used at the optimization step.

One of the disadvantages of cross-validation is that it actually does not validate the original model. The cross-validation results are computed by using of local models — based on the subsets of the initial calibration set. This limits applicability of the validation results significantly, especially comparing to the possibilities provided by the test set validation. For example, in PCA, we neither can compute the explained variance or scores for the CV results, nor perform a direct comparison of the score and orthogonal distances obtained from the cross-validation segments.

To overcome these limitations, we propose a novel approach named Procrustes cross-validation (PCV). The method makes possible to assess the sample-to-sample variation by measuring the difference between the local models in CV (in form of angles between the principal component subspaces) and utilizing them to generate new set of data — pseudo-validation set. This set can then be employed similar to an independent test set. At the same time, pseudo-validation set can be used in the same way as the conventional cross-validation for evaluation of the model.

All details on the implementation of the proposed approach as well as the results obtained in validation of PCA models using simulated and real datasets will be shown.

T18. Analytical performance characterization of potentiometric multisensor systems

Hadi Parastar[1], Dmitry Kirsanov [2]

[1]Department of Chemistry, Sharif University of Technology, Tehran, Iran

[2]Institute of Chemistry, St. Petersburg State University, St. Petersburg, Russia

Potentiometric multisensor systems are now being actively developed for different analytical applications varying from quantitative analysis of particular chemical substances to evaluation of complex integral quality parameters of the samples, like e.g. taste descriptors, toxicity, and geographical provenance. Chemometric modeling of the response of the multisensor arrays is indispensable part of these instruments. In case of quantitative analysis a common way to report on multisensor system performance is to give the values of root mean square of prediction (RMSEP) for the studied set of samples. Although this indeed gives some idea about the quality of the regression model and applicability of the sensors to the particular analytical task, the RMSEP values can hardly be employed to compare the performance of two different multisensor systems unless absolutely the same samples were studied. Thus, objective numerical parameters with thoroughly established definitions are needed to evaluate and compare performance of multisensor systems. In the context of multivariate calibration these parameters (analytical figures of merit, AFOMs) were recently suggested by the group of Olivieri [1]. Sensitivity, analytical sensitivity, selectivity, detection limit and quantitation limit values can be calculated in the framework of this approach. The purpose of the present study was to explore the use of these metrics with several real cases of potentiometric multisensor systems. The presentation will give a summary of the obtained results.

Acknowledgement

References

1. A. O. Olivieri, "Analytical figures of merit: from univariate to multivariate calibration", Chem. Rev., 114 (2014) 5358-5378.

T19. Breaching the cross-contamination identification barrier in *S. cerevisiae* strains using SERS and DD-SIMCA

Carlos A. Teixeira, João R. Belinato, Fábio Augusto, Ronei J. Poppi

Institute of Chemistry, University of Campinas (UNICAMP), Campinas - SP, Brazil

Due to the impact that commercial fungicides can cause on human health, as well as in the environment, there is increasing interest for biopesticides such as some microorganisms-based products [1]. *Saccharomyces Cerevisiae* has been used as a bioproduct and it can act as antagonists to phytopathogens with varying degrees of efficiency. Therefore, it is important to develop fast and easy-to-use protocols to identify possible cross-contamination between different strains which play an important role on antagonistic efficiency. In this work Data Driven Soft Independent Modelling of Class Analogies (DD-SIMCA [2]) was used to distinguish SERS (Surface Enhanced Raman Scattering) spectra of cultures of pure lineages of *S. Cerevisiae* from cultures containing different strains, thus characterizing contamination situations which are at first visually identical even with the use of microscopes. This chosen approach considers just spectral information from a pure strain to determine whether it would have a contamination. Therefore, this situation characterizes an authentication problem and it is well suitable for a one-class classification method (OCC). All yeast strains were analyzed separately in pure colonies and as binary and ternary mixtures, simulating possible contaminations between them. All strains were grown on PDA culture medium and incubated for 3 days at $27\,^\circ$C in T24 cycles. To obtain SERS spectra 10 µL of the strain was added to 100 µL of concentrated gold colloid (synthesized by Lee and Meisel's method [3]). Then 5 µL of the mixture was deposited on an aluminum foil and allowed to dry at room temperature. With 100% sensitivity and selectivity, the proposed methodology is suitable for screening analysis on growth culture with suspicion of cross-contamination. Therefore, the proposed method avoids the indiscriminate use of analysis such as Polymerase Chain Reaction (PCR) and its use may be implemented in routine analysis in the microbiology field.

References

1. Fialho, M. B. et al, Afr. J. Microbiol. 2014, 8, 2077–2084.

2. Zontov, Y.V., et al., Chemom. Intell. Lab. Syst., 2017, 167, 23-28.

3. Lee, P. C. and Meisel, D., J. Phys. Chem. 1982, 86, 3391–339.

T20. Chemometric tools for microstructured optical fibers spectral data processing

Pavel P. Pidenko[1], Sergey A. Pidenko[1], Anastasia A Zanishevskaya[2], Yulia S. Skibina[2], Natalia A. Burmistrova[1]

[1]Institute of Chemistry, Saratov State University, 83 Astrakhanskaya Str., Saratov, Russia 410012

[2]LLC SPE Nanostructured Glass Technology Ltd, Saratov, Russia 410033

Nowadays optical sensors are widely used analytical tool. Microstructured optical fibers (MOF) are one of the types of optical fibers promising as new sensor elements basis. MOFs based sensors have several fundamental advantages, among which are high sensitivity to physico-chemical properties of medium filled the inner fiber cavity, and microvolume analysis of samples. MOFs with various architectural and optical properties were studied for creation optimal sensor elements for a wide range of analytes. However, the interpretation of the analytical response in the MOFs transmission spectrum is a complex task. This is due to several reasons, including the complex structure of the MOFs transmission spectra and the influence of various factors on the optical properties of the fibers. This makes one-dimensional analysis of MOFs transmission spectra difficult.

The main purpose of the present study was to estimation of effectivity the multidimensional processing MOFs transmission spectra data. The main mechanisms that influence on MOFs transmission spectra transformation are the

changing of the refractive index (RI) of the liquid sample filled the MOFs inner cavity and the change of the MOFs central defect diameter due to sorption of the target molecules. Solutions of inorganic salts of different nature and concentrations were used to study the effect of RI, ionic strength and absorption on the MOFs transmission spectra. The study of the effect of molecule sorption on MOFs inner surface was carried out using albumin from chicken egg white solutions of different concentrations. The results of one-dimension approach and partial least squares regression were compared.

It was shown that chemometric processing of the MOFs transmission spectra allow to effectively estimate the analytical characteristics and increase the correlation coefficients of calibration dependencies. The obtained results were confirmed in the analysis of complex matrixes.

Acknowledgement

T21. Ensemble modeling for the spectral analysis of complex samples

Xihui Bian[1,2], Xinyan Wu[1], Peng Liu[1]

[1]School of Chemistry and Chemical Engineering, Tiangong University, Tianjin, 300387, P. R. China

[2]Institute for Molecules and Materials, Radboud University, Heyendaalseweg 135, 6525 AJ Nijmegen, the Netherlands

Ensemble strategies, also referred to as data fusion, consensus, multimodel or population analysis, have gained increasing attention and combined with combined with multivariate calibration, chemical pattern recognition, calibration transfer, preprocessing, variable selection, outlier detection and parameter selection in chemometrics [1,2]. Among these methods, ensemble modeling for qualitative and quantitative analysis is the hottest research topic. The aim of ensemble modeling is to obtain a more accurate, stable and robust prediction by combining the predictions of multiple sub-models. The generation and modeling of the training subsets, as well

as the integration of the sub-models are three keys to the success of ensemble modeling [2].

This contribution firstly reviewed the above-mentioned strategies in relation to modeling of complex samples and then developed several ensemble modeling methods. Firstly, high and low frequency unfolded partial least squares [3,4], which integrates empirical mode decomposition (EMD) and unfolded strategy with PLS, was proposed for quantitative analysis of fuel oil samples and identification of milk samples. Secondly, subagging ELM and boosting ELM [5] were proposed for improving the stability and accuracy of ELM. Then, EMD weight multi-scale SVR was proposed for quantitative analysis of visible spectroscopy of edible blend oil. In addition, double ensemble PLS based on Lasso and Monte Carlo resampling for quantitative analysis of corn and edible blend samples were proposed. These methods all demonstrated the superiorities in accuracy, stability and robustness compared with single modeling.

Acknowledgement

References

1. R Wehrens, H Putter, LMC Buydens, Chemom. Intell. Lab. Syst. 2000, 54(1): 35-52.

2. X Bian, P Diwu, Y Liu, et al., J. Chemom. 2018, 32 (11): e2940.

3. Bian, X.H.; Li, S.J.; Lin, L.G.; et al. Anal. Chim. Acta 2016, 925:16-22.

4. X Bian, C Zhang, P Liu, et al., Chemom. Intell. Lab. Syst. 2017, 170, 96-101.

5. X Bian, P Diwu, C Zhang, et al., Anal. Chim. Acta 2018, 1009, 20-26.

T22. Direct standardization and slope and bias correction for calibration model transfer between 4-channel optical multisensor systems and full-scale spectrometer

Anastasiia Surkova[1,2], _Andrey Bogomolov[3], Andrey Legin[1], Dmitry Kirsanov[1]_

[1]Institute of Chemistry, St. Petersburg State University, St. Petersburg, Russia [2]Samara State Technical University, Samara, Russia

[3]Endress+Hauser Conducta GmbH+Co. KG, Aalen, Germany

Any type of spectroscopic analyzer requires a multivariate calibration model in order to relate its response with a measured property of the sample. To build the robust calibration model significant efforts are needed. A representative sample set has to be acquired and analyzed by a reference analytical method. However, it is not always possible to use the model derived for one spectroscopic instrument with the data obtained by another, even if they are of the same type. The application of various transfer techniques helps to solve the problem of the model adaptation to a new analytical device.

Nowadays, the development of various optical multisensor systems is experiencing intense growth. To reduce the price of such devices poorly standardizable materials (such as light fibers, 3D-printed parts, stamped optics) are used to construct them. Therefore, a response of each optical multisensor system is different and in this research domain calibration model transfer is urgently needed. Moreover, it seems very attractive to try the transfer of a good-quality calibration models derived in controlled laboratory conditions with traditional spectrometers to industrial multisensor systems operating in on-line mode.

In the present work, direct standardization and slope and bias correction methods were tested for transfer of calibration models from the full-scale spectrometer to optical multisensor system consisted of four light-emitting diodes (LEDs) and between two optical systems of the same kind. To study the possibility of calibration transfer, the metal nitrates aqueous solutions prepared according to the diagonal design were measured with UV/Vis/NIR spectrometer and two LED-based optical multisensor systems. The prediction performance of different transfer techniques will be discussed.

The results of this study are important for further development of optical multisensor systems especially in the context of their industrial application.

Acknowledgement

T23. Use of non-negative matrix factorization (NMF) for an exploratory analysis of atomic and molecular spectra of zooplankton

N. Sushkov[1], T. Labutin[1], N. Lobus[2], G. Galbács[3], R. Rajkó[4]

[1]Lomonosov Moscow State University, Moscow, Russia

[2]Shirshov Institute of Oceanology of the Russian Academy of Sciences, Moscow, Russia

[3]University of Szeged, Szeged, Hungary

[4]University of Pécs, Pécs, Hungary

Some of zooplankton organisms (viz. several taxa of sea crustaceans) are found to accumulate lithium. Causes and mechanisms underlying this phaenomenon are unclear and interesting for marine biologists and oceanographers. The goal of this study was to establish correlations between analytical signals of elements and molecules in animal tissue using laser-induced breakdown spectroscopy (LIBS) and Raman scattering spectroscopy. These techniques are rapid and micro- or non-destructive. Bulk composition data, obtained by ICP-AES and ICP-MS after digestion, were also available. In total, 29 samples were analysed.

The NMF is well suited for our purposes as it provides principal components (PCs) that have physical sense and are therefore relatively easy to interpret. Prior to processing Raman spectra, all remains of fluorescence background were carefully removed. We found that Raman PCs obtained by NMF correlate well with the available taxa (copepods, malacostracans, arrow worms, molluscs, and mixed samples). The number of PCs was found using the scree plot provided by the PCA. As copepods contain the highest amounts of Li, they were then considered as a separate sample subset. This subset could further be decomposed to give 3 PCs, the 1st containing carotenoid peaks, the 2nd exhibiting quite intensive amino acid signals, and the 3rd featuring bands of hydroxyl-containing compounds.

Decomposition of LIBS data yields 4 PCs (lines of Mg, Sr and Ca; lines of Li, Na and K; molecular emission; Ca and Cu lines), while the copepod subset suggests that Li and K (but not Na) share the same component.

NMF of fused datasets (composed from LIBS and Raman spectra) has suggested that Li accumulation is related to the content of amino acids. Indeed, Li content linearly depends on the area of tryptophan peak (R = 0.91, n = 8) for copepods containing less than 70 mg/kg of Li. It has also been found that high copper and lithium contents are not found simultaneously, Cu being more abundant in malacostracans than in copepods.

Acknowledgement

T24. Different strategies of SIMCA model optimization- a comparison study

Zuzanna Małyjurek, *Beata Walczak*

Institute of Chemistry, University of Silesia, Katowice, Poland

Classification approach aims to define the belongingness of new objects to predefined studied classes, where for each of studied classes (target classes) an individual model is constructed. Unlike the discrimination where new sample will always be assigned to one of the studied classes, in classification new sample may be assigned to none of the studied classes. Classification approach is widely exploited in food and drug quality testing, authenticity verification, origin confirming and monitoring of technological processes [1].

One of the most popular classification methods is SIMCA (Soft Independent Model Class Analogy). Since SIMCA model is based on PCA (Principal Component Analysis), crucial stage is estimation of PCA model complexity. Strategies for the selection of SIMCA model complexity can be divided into two types: "compliant" when target and nontarget class samples are required in the model complexity optimization

process, and "rigorous" when only the target class samples are used [2]. Another model optimization step is the selection of the final decision threshold. Two types of decision threshold establishment can be considered: fixed and optimized threshold [3].

The aim of this study is comprehensive comparison of different strategies of SIMCA model optimization considering the compliant and rigorous approaches combined with four different resampling methods implemented in validation process, i.e. leave-one-out, 10-fold Cross-Validation, Monte Carlo Cross-Validation and bootstrapping. Moreover, every combination is tested in reference to fixed and optimized decision threshold. For evaluation of different strategies of SIMCA model optimization, seven real-world datasets are employed.

The study revealed that compliant approach with optimized decision threshold and rigorous approach with fixed threshold can lead to the most satisfactory classification results, thus they are recommended for optimization of SIMCA model.

Acknowledgement

References

1. P. Oliveri, Anal. Chim. Acta 982 (2017) 9-19.

2. O.Y. Rodionowa, P. Oliveri, A.L. Pomerantsev, Chemometr. Intell. Lab. 159 (2016) 89-96.

3. R. Vitale, F. Marini, C. Ruckebusch, Anal. Chem. 90 (2018) 10738-10747.

T25. Prediction of selectivity for potentiometric sensors with QSPR modeling

Ekaterina A. Martynko, Dmitry O. Kirsanov

Saint Petersburg State University, Saint Petersburg, Russia

Ion-selective electrodes (ISE), or potentiometric sensors, are electrochemical devices, which can determine target ion activity in solution in the presence of other

ions. They are widely applied for various analytical tasks, both individually and as a part of sensor arrays. However, the development of new ISEs is a tedious process: it requires the synthesis of the ligands, the production of the polymeric membranes and the electrochemical measurements in order to characterize the potential ISE, but in the end, there is no guarantee that a new sensor with the required parameters will be successfully created.

Quantitative structure-property relationship (QSPR) is an approach that connects the structure of a molecule formalized in a set of descriptors with its physicochemical properties via a mathematical model [1]. QSPR approach was successfully employed for the prediction of metal-ligand interactions in liquid extraction processes [2] and studies of metal toxicity [3]. The response of polymeric membrane potentiometric sensors is also based on metal-ligand interactions. Therefore, the application of QSPR modeling to connect the ligand structure and electrochemical properties of the corresponding polymeric sensor membrane could significantly facilitate the development of sensors with desired properties.

In this work, the feasibility of QSPR modeling was studied in the prediction of the potentiometric Mg^{2+}/Ca^{2+} selectivity of Mg-selective electrodes based on the ionophore structure. The dataset was created using literature data on various lipophilic amide magnesium ionophores with selectivity constants determination conditions as similar as possible.

The ISIDA/QSPR program package was used for modeling and prediction with multiple linear regression as a machine learning method and substructural molecular fragments descriptors as independent variables. The model predictive performance was evaluated with 10-fold external cross-validation. The coefficients of determination were 0.95 and 0.66 for calibration and test sets, respectively, showing a perspective for this model to become a screening tool for the promising ionophores with tailored selectivity.

Acknowledgement

References

1. Golbraikh A., Tropsha A. Chemoinformatics: Basic Concepts and Methods, 2018.

2. V. Solov'ev, A. Tsivadze, G. Marcou, A. Varnek, Mol. Inform., 2019, 38.6.

3. X. Li, L. Wang, Y. Li, Adv. Mater. Res., 2012, 155–158.

T26. Chemometric approaches to direct analysis of lanthanides in REE-rich ores by TXRF and WDXRF

Timur Akhmetzhanov[1], Galina Pashkova[2], Victor Chubarov[3], Timur Labutin[1], Andrey Popov[1]

[1]Department of Chemistry, Lomonosov Moscow State University, 119234, Leninskie Gory, 1-3, Moscow, Russia

[2]Institute of the Earth's Crust, SB RAS, Lermontov St., 128, Irkutsk, 664033, Russia

[3]Vinogradov Institute of Geochemistry, SB RAS, Favorsky St., 1A, Irkutsk, 664033, Russia

X-Ray fluorescence (XRF) is a useful tool for rapid multielemental analysis of geological samples that does not require acid digestion. One of the important tasks given to XRF is the simultaneous quantification of lanthanides in ores, which is a problem due to strong line overlapping. Thus, high resolution wavelength-dispersive XRF (WDXRF) is used. However, the question still stands: can cheaper low-resolution energy-dispersive spectrometers be used for this task? To answer this question we provided an assessment of total reflection XRF (TXRF) as a tool for direct analysis of REE-rich ores.

In this work, we used partial least squares (PLS) and principal component regression (PCR) to circumvent spectral interferences in both TXRF (S2 Picofox, Bruker GmbH) and WDXRF (S8 Tiger, Bruker AXS) spectra. To provide a uniform distribution of concentrations of lanthanides in a test set we implemented specialized design of experiment (DoE) based on Latin hypercube sampling (LHS) [1]. We obtained the matrix of 5 factors (elements) and 20 levels (concentrations as well as samples). The maximum correlation is 0.03, which is several times better than in [2]. In addition, this design has an advantage over DoE in [3] where the number of samples equals a square of the levels. Thereafter, our model was applied for the determination of Ce (460–39500 ppm), La (260–24100 ppm), Nd (150–11800 ppm) by both TXRF and WDXRF in standard reference materials of niobium and uranium ores used for validation. In addition, WDXRF provided quantitative

determination of Pr (47–3800 ppm), Sm (24–1500 ppm) in these samples. We also provide a comparison of PLS, PCR and integrated software results. Chemometrics provided twice the best average root mean square error of prediction (RMSEp) for Ce, La, and Nd for TXRF and thrice for WDXRF (average RMSEp for all elements) compared to the integrated software.

References

1. M. D. McKay, R. J. Beckman and W. J. Conover., Technometrics. 1979. 21. 239-245.

2. D. Kirsanov et al. Spectrochimica Acta Part B. 2015. 113. 126–131.

3. A. Maltsev, A. Shulyumova and N. Umarova. X-Ray Spectrometry. 2018. 5. 396–404.

T27. Calculation of total distance and outlier detection in the projection regression models

Oxana Ye. Rodionova, *Alexey L. Pomerantsev*

N.N.Semenov Federal Research Center for Chemical Physics, RAS, Kosygin 4, 119991 Moscow, Russia

We present a step-by-step procedure for outlier detection. This procedure utilizes the total distance which is calculated for each sample in the projection regression model. The approach is based on the previously introduced technique developed for full distance calculation in the PCA modeling. Employing the total distance, we introduce the outlier cut-off level, which is obtained by means of the data-driven estimation of DoFs and scaling parameters for the partial distances in the projection models. The procedure can be referred to as a sequential focused trimming. Starting from a model that employs the whole training set, we estimate various partial distances for X and Y data using classical and robust estimation methods. These estimates are then used for calculation of the total distance that is employed for detection of outliers. If outliers are found and removed, the model is rebuilt, and then the data are checked again for new possible outliers. The iterative nature of the procedure helps to overcome the "masking" effect. The backward step allows to

identify samples that are erroneously attributed to outliers and return these samples to the calibration set, thus, to overcome the 'swamping' effect.

The procedure is simple as it utilizes the conventional projection methods such as PCR and PLS, which are much faster than their robust counterparts. At the same time, the procedure is effective in application to the data contaminated with outliers due to employment of the robust estimates for unknown parameters in the distance distributions.

T28. Multivariative analysis of Raman and autofluorescence spectra of human tissues

Ivan A. Bratchenko[1], Lyudmila A. Bratchenko[1], Yulia A. Khristoforova[1], Dmitry N. Artemyev[1], Oleg O. Myakinin[1], Alexander A. Moryatov[2], Sergey V. Kozlov[2], Valery P. Zakharov[1]

[1]Samara National Research University

[2]Samara State Medical University

The pathological state of a human body leads to altered biochemical composition of body fluids. Conventional biochemical analysis of body fluids is notable for its low-informative value in localizing a particular pathology. As an alternative, Raman spectroscopy provides detailed evaluation of blood characteristics at the molecular level. Raman blood spectra are characterized by multicollinearity as well as by the presence of autofluorescence background and noises of different nature. Choice of a proper method for experimental data processing of blood spectra is crucial for obtaining statistically reliable information regarding a pathological process in the body. This study examines different approaches to multidimensional analysis of the various-size Raman spectral dataset of human blood samples by a cost-effective Raman setup in a clinical setting. To discriminate blood samples by the pathology type, statistical processing of experimental data is performed by factor analysis, logistic regression, discriminant analysis, classification tree, projection to latent structures discriminant analysis (PLS-DA), and soft independent modeling of class analogies. Comparative analysis of the discussed multivariate methods demonstrates that the PLS-DA method (sensitivity 0.75, specificity 0.81, and accuracy 0.76) proved to be the most effective for the classification of blood samples by cancer localization. In terms of classification for the presence of

hyperproteinemia, the most efficient are the logistic regression method (sensitivity 0.89, specificity 0.99, and accuracy 0.97) and the discriminant analysis method (sensitivity 0.83, specificity 1.0, and accuracy 0.97). In general, the selected multivariate methods could serve as a reliable tool for analyzing spectral characteristics of body fluids.

T29. SIMCA. Decision rules

Alexey Pomerantsev, Oxana Rodionova

N.N. Semenov Federal Research Center for Chemical Physics, RAS, Moscow, Russia

SIMCA was first proposed in its original version by S. Wold more than 40 years ago. Afterwards it underwent several modifications. Nowadays, SIMCA is designed as an approach that consists of three stages. The first is PCA followed by the calculation of the orthogonal, Q, and score, H, distances. The second step is a subsequent estimation of the distributions parameters. At the last stage, Q and H statistics are combined and a decision-making rule is developed.

The main goal of SIMCA is to establish a decision rule that determines whether a new sample belongs to the target class that has been modeled using the training set **X**. Typically, this decision is developed by means of cut-off values, which delineate an acceptance area in the space of the PCA statistics, sometimes called 'distance-distance' plot.

In making a decision, two typical errors could occur. The type I error happens in case the sample is rejected, while it actually belongs to the class. The type II error is the acceptance of the sample, which in fact does not belong to the class. The type I error is directly connected with the acceptance area. The smaller the type I error is, the larger the acceptance area will be.

Using two statistics, H and Q, a number of the decision rules can be established. In this presentation we compare and discuss the most popular rules.

Acknowledgement

T30. Development of the hardware-software complex for diagnosing the state of the human cardiovascular system based on non-linear connectivity indices

E.I. Borovkova[1,2,3], _O.M. Posnenkova[1,3]_, _B.P. Bezruchko[1,2]_, _Yu.M. Ishbulatov[1,2,3]_, _V.V. Skazkina[1,2]_, _A.S. Karavaev[1,2,3]_

[1]_Saratov Branch of the Institute of RadioEngineering and Electronics of RAS_

[2]_Saratov State University_

[3]_Saratov State Medical University_

Earlier, we proposed a method of determining a condition of cardiovascular of a person, based on the study of phase locking between autonomic control of heart rate variability and baroreflex control of arterial pressure. The developed method is based on the calculation using the methods proposed by us of the analysis of complex nonlinear non-stationary signals of a quantitative indicator - the total percentage of phase synchronization [1]. Comprehensive statistics (more than 1000 experimental records of different groups of patients and conditionally healthy subjects) showed that the method provides important additional diagnostic information, allowing one to statistically reliably predict the occurrence of cardiovascular complications, assess the likelihood of survival of patients after acute myocardial infarction, and help in the selection of medication treatment of such patients [2]. In this case, the diagnostics were carried out using of simultaneously recorded ten minutes records of the electrocardiogram and photoplethysmogram in the post-processing mode.

In this work, we expanded the capabilities of the proposed method, developed an autonomous small-sized portable device with non-volatile data storage with microprocessor control. Device allows to registrate of a digital finger photoplethysmogram for many hours and do long-term monitoring of phase locking between autonomic control of heart rate variability and baroreflex control of arterial pressure in a real time. For this device uses only an univariate signal photoplethysmogram. We compared the results of monitoring the phase locking by the developed device and the earlear developed method using the photoplethysmogram and electrocardiogram signals. The operability and prospects of the device were shown.

References

1. Karavaev A.S. et al. Synchronization of low-frequency oscillations in the human cardiovascular system // Chaos. 2009. V. 19. P. 033112.

2. Kiselev A.R. et al. Selection of optimal dose of beta-blocker treatment in myocardial infarction patients basing on changes in synchronization between 0.1 Hz oscillations in heart rate and peripheral microcirculation // Journal of Cardiovascular Medicine. 2012. V. 13. № 8. P. 491-498.

P01. Express technique for nondestructive testing of medicines quality using NIR spectroscopy and chemometrics. Application for mobile laboratories in Russia

Anna V. Titova[1], K.S. Balyklova[1,2], Oxana Ye. Rodionova[1,3]

[1]Information and Methodological Center for Expertise, Stocktaking and Analysis of Circulation of Medical Products, Roszdravnadzor, Moscow, Russia

[2]I.M. Sechenov First Moscow State Medical University, Moscow, Russia

[3]N.N.Semenov Federal research center for Chemical Physics RAS, Moscow, Russia

The system of mobile laboratories for revealing of substandard and falsified medicines has been organized in Russia. The task for development of non-destructive express procedures for the medicines testing has been set. The NIR-spectroscopy is used as an instrumental method for testing of solid dosage forms (tablets, dragee, and capsules). One-class classification method based on Euclidean distances ("Ident" by OPUS, Bruker) is used for data processing.

The general procedure for model development and validation is presented [1]. It consists of the following main steps: (1) sampling (2) NIR-spectra acquisition; (3) visual analysis of the spectra for excluding gross errors; (4) division of NIR-spectra into training (for model building) and test (for assessing e model sensitivity) sets; (5) preprocessing of the spectra (spectral alignment with the conditionally standard

spectrum; selection of the informative region; unit spectral normalization, depending on the model, smoothing and/or first or second odder derivation; (6) model building; (7) assessment of model sensitivity I – the degree of acceptance of samples from the training set; (8) assessment of model sensitivity II- the degree of acceptance of the samples from the test set; (9) model validation for specificity I - the degree of deviation of the blister spectra; (10) model validation for specificity II - the degree of deviation samples of generic medicines in the same packaging; (11) model validation for specificity III - the degree of deviation of medicines sample in the same packaging, but containing similar in structure active substance; (12) model validation for specificity IV - the degree of deviation of counterfeit or substandard samples (if available).Optimization of the one-class model area boundary is carried out on the basis of validation parameters. A year later, the model is tested using samples of new medicines batches; it additionally allows to control the stability of the medicines technology process.

Currently, the database used in mobile laboratories in Russia contains about 2500 NIR-models.

References

1. O.Ye. Rodionova, A.V. Titova, K.S. Balyklova, A.L. Pomerantsev, Detection of counterfeit and substandard tablets using non-invasive NIR and chemometrics - A conceptual framework for a big screening system, Talanta, 205, 120150 (2019) . DOI: 10.1016/j.talanta.2019.120150.

P02. Multivariate curve resolution of data with heteroscedasticity: from PCA to WLS and back

Alexej N. Skvortsov

Peter the Great Saint Petersburg Polytechnic University. Polytekhnicheskaja, 29. 195251 St-Petersburg, Russia

Multivariate curve resolution (MCR) aims to resolve data matrices into spectra and concentrations of individual chemical components, given a set of constraints, ranging from nonnegativity of concentrations to complex tailored physicochemical models. In general, additional work is also required to estimate the ambiguity of the resolution. The core assumption of ordinary MCR is the bilinearity of the measured

response: $\mathbf{X}=\mathbf{CS}^T+\mathbf{E}$ (Eq. 1), where $\mathbf{X}$ – data matrix, $\mathbf{C}$ and $\mathbf{S}$ – concentration and spectra matrices (chemical factors), $\mathbf{E}$ – homoscedastic uncertainty matrix. However, in many practical cases, nonlinear effects in $\mathbf{X}$ or heteroscedasticity of $\mathbf{E}$ may ruin MCR accuracy and hinder rank analysis. Typical examples are moderate nonlinearity (e.g. saturation or 'peak flattening' in absorption spectra) or noise spikes. We have previously shown (Skvortsov 2017) that the use of weighted least-squares (WLS) instead of ordinary least squares (OLS) in Eq. 1 with carefully selected weight matrix may produce unbiased estimates for $\mathbf{C}$ and $\mathbf{S}$ and their ambiguity in many practical cases. The approaches are related to MCR-WALS (Stanimirova et al., 2011) and PMF (Paatero, Tapper, 1994). Still, generation of model spaces in WLS was much slower and less stable than finding principal components (PCs, actually, OLS model spaces). Principal component analysis is done anyway for assessment of data matrices, but, in WALS, PCs are actually wasted.

It is well known in chemometrics that, in case of moderate nonlinearity, the number of significant principal components (PCs) is larger than the number of factors in the system, and additional PCs somehow account for nonlinearity. Methods exist, e.g. principal polynomial analysis (Laparra et al., 2016), which try to grasp it explicitly. In the present study we tried an alternative empirical approach: we constructed *A* chemical factors ($\mathbf{C}$, $\mathbf{S}$) as linear combination of *B* PCs ($\mathbf{T}$, $\mathbf{P}$) of matrix $\mathbf{X}$: $\mathbf{C}=\mathbf{TU}$, $\mathbf{S}=\mathbf{PV}$, so that *B>A*. Then the elements of transformation matrices $\mathbf{U}$ and $\mathbf{V}$ were found by alternating least squares or non-linear optimization to minimize the residues in Eq. 1 in WLS sense. This approach resembles resolving factor analysis (RFA), but the transformation matrices $\mathbf{U}$ and $\mathbf{V}$ are rectangular and generally unrelated to each other. Once a single ($\mathbf{C}$, $\mathbf{S}$) was found, the whole set of feasible WLS solutions could be generated by typical RFA means. It has been shown that this approach is operable and gives estimates, which are practically similar to the results of proper WALS, but it is more stable and requires few operations on large matrices (data matrix $\mathbf{X}$ and weight matrix). The generalization of the calculation procedure is discussed, which allows to read the large matrices once and estimate the success of the subsequent nonlinear search. The ability of the proposed approach to resolve simulated and real flattened absorption spectra, matrices with missing data, and matrices with point outliers (spikes) is demonstrated.

P03. Spectral index of oil in soil based on ultraviolet and near-infrared

Danil N. Sherstobitov, V.V. Ermakov

Petroleum Engineering Faculty, Samara State Technical University, Samara, Russia

In this work, we studied the spectral characteristics of previously dehydrated soil under an artificial light source. The main task of the work is to develop a spectral index for determining the content of oil and oil products in the soil based on the ultraviolet and near infrared spectral regions.

The weight of each test sample of contaminated soil was 30 grams. The concentration of oil and oil products in each of the samples did not exceed 21% to more accurately determine the spectral characteristics of the studied soil samples.

The experiment was conducted under an artificial light source - an incandescent lamp (power - 95 watts, luminous flux -1250 lumens), located at a distance of 15 cm from the surface of the studied soil.

The spectral characteristics of the soil were determined using 2 spectrometers - Ocean Optics (STS-NIR) and (STS-UV). The distance of the spectrometer from the studied object was 5 cm.

In the work, to determine the characteristic features of contaminated soil, multidimensional classification and calibration methods were used. The most acceptable way to determine such implicit dependencies are chemometric methods. The simulation was carried out using the principal component method (SAR), with the help of which areas characteristic for this type of pollution were identified in multidimensional space. The spectral characteristics of contaminated soil were divided into 2 parts of the data array - ultraviolet and near infrared spectral regions. The selected patterns were differentially analyzed to highlight subgroups and search for additional information.

For each individual selected cluster, calibration was performed using regression methods for the main components (PCA) or latent structures (PLS). As a result of processing high-quality data, calibration dependences were found inside the data arrays, which in turn made it possible to identify the most valuable characteristics for creating a spectral index.

P04. Chemometric approaches in the quality control of multicomponent drugs

Natalia B. Shestopalova[1], Yulia A.Fomina [1], Natalia A.Burmistrova [2], Tatiana Yu. Rusanova[2]

[1]Saratov State Medical University named after V. I. Razumovsky, Saratov, Russia

[2]Saratov State University, Saratov, Russia

Modern drugs are complex therapeutic systems, which, along with active drug substances (DSs), include auxiliary components. Combined drugs containing two or more DSs are popular with consumers, as they are convenient to use, economically beneficial, allow you to strengthen and expand the pharmacological effect, level out side effects. One indicator of the quality of tablet forms is an indicator of dosage uniformity, which is most often determined by spectrophotometric methods in the UV region or HPLC. The presence in substances of combined drugs of similar physicochemical properties requires, in most cases, a laborious procedure for preliminary separation of the components and complicates the quality control of dosage forms. The development of alternative rapid approaches for the analysis of multicomponent drugs is an urgent task. One of the promising areas is the use of chemometric algorithms.

In this regard, the aim of the work was to assess the possibility of using spectrophotometry methods in combination with the principal component analysis (PCA) for controlling the quality of "Papazol" tablets in terms of the dosage uniformity. "Papazol" tablets of two Russian manufacturers were selected as model drugs. To quantify the content of active substances - papaverine hydrochloride (30 mg) and dibazole (30 mg) - the spectrophotometry method in the UV region was used with the results being processed by the Vierordt's method. According to the requirements of the State Pharmacopoeia XIV (OFS 1.4.2.0008.18) "Dosing Uniformity", the test results in all test samples of tablets were found to be satisfactory.

Processing of the spectral data of aqueous solutions of tablets by the PCA showed that the samples form a single cluster on the plot of scores, which also confirms the

uniformity of the dosage of the main active substances in the drugs of the two manufacturers.

An express spectrophotometric technique with a chemometric algorithm for determining the uniformity of dosage of active substances in the "Papazol" combination preparation has been developed. The correctness and reproducibility of the developed methodology were statistically evaluated and the possibility of using chemometric algorithms to control the quality of drugs by the indicator of "uniformity of dosage" was shown.

P05. Simultaneous determination of food colorants by solid phase spectroscopy coupled to chemometric analysis

E.V. Larionova, <u>S.V. Romanenko</u>, Yu. M. Vodova

Tomsk Polytechnic University, Lenin av., 30, Tomsk, Russia, 634050

Food colorants are used to preserve or provide color to manufactured products that can lose their natural color and to attract attention to edible products in order to make them more appealing for consumers. Nevertheless, their use should be limited because of toxicity effects.

We have early proposed a new solid phase absorbance spectroscopy technique applied for determination of food colorants that are low cost, mobile, possess expressivity, low detection limits and do not require preliminary sample preparation [1]. Nevertheless, the limitation of using the technique is the lack of selectivity in case of relatively equal of maximum absorbance of colorants. The use of chemometrics is a very common approach to overcome these difficulties.

The aim of this work is to develop the technique of simultaneous determination of food colorants for example Ponceau 4R (E124) Cocheneal red A and Sunset Yellow (E110) Orange Yellow S by solid phase spectroscopy coupled to chemometric resolution of overlapped peaks. We use several chemometric procedures such as curve fitting, stepwise resolution method and MCR-ALS in order to choose simple, quick and precise resolution of overlapped peaks. The resolution method should be appropriate for low cost portable devises for on-site analysis. For the successful application of the resolution procedures, we use the following specific features: the

prior estimation of signal shape and appropriate model functions for the analytical signal.

Accuracy and precision of resolution are verified by using a set of overlapping signals of colorant standards with different ratios of heights. The proposed technique is validated by analysis of food samples compared with HPLS with satisfactory results.

References

1. Dudkina, A.A., Volgina, T.N., Saranchina, N.V., Gavrilenko, N.A., Gavrilenko, M.A. Colorimetric determination of food colourants using solid phase extraction into polymethacrylate matrix // (2019) Talanta, 202, pp. 186-189.

P06. A stepwise resolution method of overlapped peaks

S.V. Romanenko, N.S. Shekhovtsova, D.M. Karachakov, E.V. Larionova

Tomsk Polytechnic University, Lenin av., 30, Tomsk, Russia, 634050

The problem of overlapping peaks resolution is a topical issue in a number of analytical techniques. A lot of chemometric techniques based on hard and soft modelling are proposed for resolution of overlapped analytical signals. Nevertheless they have various disadvantages and limitations, for example selectivity to noise level, peak shape and signals ratio. Searching the new approaches to resolution of strongly overlapped signals (with more than two components) is still actual problem.

We propose a new stepwise resolution method (SRM).

The basic idea of the SRM approach is a one-by-one mathematical removal of a signal of a particular analyzed component from a complex (mixed) signal.

For the successful application of the proposed resolution procedure we suggest to use the following specific issues: (i) optimization criterion having the best defined minimum responsible for the optimal resolving signal parameters and the absence of local minimum; (ii) appropriate model functions for the subtracted signal, the choice of that based on the prior estimation of signal shape; (iii) reduction the number of optimized parameters of subtracted signals, based on the reliable prior estimation of peak shape parameters; (iv) effective resolution criterion that used to estimate robustness and efficiency of the proposed resolution procedure.

Effectiveness of SRM is verified by resolution of simulated overlapping signals and signals obtained by stripping voltammetry and X-ray diffraction method.

P07. Microstructured optical fibers optimization for proteins analysis supported by chemometric tools

Pavel P. Pidenko[1], _Kirill Yu. Presnyakov[1], Andrey A. Shuvalov[2], Yulia S. Skibina[2], Natalia A. Burmistrova[1]_

[1]Institute of Chemistry, Saratov State University, 83 Astrakhanskaya Str., Saratov, Russia 410012

[2]LLC SPE Nanostructured Glass Technology Ltd, Saratov, Russia 410033

Currently the analysis of individual protein and their mixtures is a powerful tool for solving different tasks in biochemistry, proteomic, biomedical research, clinical practice and food control. Optical sensors are one of the modern trends in protein analysis. They are successively used for protein screening and detection. Hollow core microstructured optical fibers (HC MOFs) are promising candidate for such systems design. The HC MOFs properties and their sensitivity to protein determination depends from their architecture and glass composition. In this regard, the choice of the most effective HC MOFs type is an important task. It should be mention that the complex dependence of HC MOFs spectral properties from several factors makes it difficult to interpret the obtained results using a one-dimensional spectral-data analysis.

The purpose of this work was the searching of HC MOFs with high sensitivity for proteins presence using chemometric tools for spectral data processing. Bovine serum albumin and albumin from chicken egg white were used as model proteins. The transformation of HC MOFs transmission spectra was investigated for albumins aqua individual and binary solutions with different concentrations. Four types of HC MOFs with different architecture and glass composition were used and have shown a different sensitivity to target albumins. Analysis of spectral data by principal component analysis allowed to obtain dependence of samples position on the score plots from proteins concentration in individual solutions and their mixture. Partial least squares regression (PLS) method was utilized for quantitative analysis of proteins. PLS model quality for different HC MOFs types was compared.

P08. Application of the MCR-ALS for processing of Mössbauer spectra

Vitaly V. Panchuk[1,2] , B. Debus[3], V.G. Semenov[1,2], D.O. Kirsanov[1]

[1]Institute of Chemistry, St. Petersburg State University, St. Petersburg, Russia

[2]Institute for Analytical Instrumentation RAS, St. Petersburg, Russia

[3]Air Quality Research Center, University of California, Davis, Davis, California, USA

Mössbauer spectroscopy has found broad applications in material science, corrosion studies, biology and various other fields as a useful tool to explore the chemical state of the atoms in condensed matter. Traditional way for processing the Mössbauer spectra is to represent them as a superposition of Lorentz multiplets using the algorithms based on least-squares, like e.g. Levenberg-Marquardt. In complex samples the chemical states of the probe atom are numerous and the derived Mössbauer spectra may contain a lot of overlapping signals, thus, a lot of uncertainties, subjectivity, and errors in fitting possible hypotheses to the experimental data may occur. The processing of such data requires very high expert level and some automated Mössbauer spectra processing tools would be highly attractive.

Multivariate Curve Resolution – Alternating Least Squares (MCR-ALS) is widely applied in analytical chemistry to resolve multi-component mixtures into individual concentration profiles and pure component spectra. It seems to be logical to try to apply MCR-ALS for processing of Mössbauer data and first attempts in this matter with simple systems were already made [1].

The purpose of this report was to explore the potential and the limitations of MCR-ALS as a data processing tool in Mössbauer studies in complex samples, where numerous phases can exist simultaneously. As an example, we address both simulated (Mössbauer spectra of a three-component kinetic system) and real-world datasets. The real-world data sets contained the Mossbauer spectra obtained in the study of formation and transformation processes in the system based on

nanocrystalline yttrium orthoferrite and in the study of corrosion products in steam generator pipes. The advantages and limitations of the MCR-ALS compared to the traditional processing of Mossbauer spectra will be discussed.

Acknowledgement

References

1. M. Fehse, D. Bessas, A. Darwiche, A. Mahmoud, G. Rahamim, C. La Fontaine, R.P. Hermann, D. Zitoun, L. Monconduit, L. Stievano, M.T. Sougrati, The electrochemical sodiation of FeSb2: New insights from operando 57Fe synchrotron Mössbauer and X-Ray absorption spectroscopy, Batteries & Supercaps, 2 (2019) 66-73.

P09. PCA-analysis of time series of biologically active substances under condition of continuous sensors functioning

E.I. Maksyutova[1], A.V. Sidelnikov[2], V.B. Boguslavskiy[1]

[1]Bashkir State University, Ufa, Russian Federation

[2]Ufa State Petroleum Technological University, Ufa, Russian Federation

In recent decades, drugs of inadequate quality are increasingly appearing. These include vitamins due to the great demand for them, especially in the autumn-winter period. This also includes the problem with biologically active additives, since they often absence the necessary biologically active substance. For reliable determination of falsification, it is necessary to develop methods that will allow rapid determination of biologically active substances in the analyzed drugs.

Vitamins are usually analyzed by two groups of methods: microbiological and physicochemical (for example, HPLC, fluorescence, colorimetry). These methods are economically costly and time-consuming. Voltammetry has several advantages over other instrumental methods. These include: high sensitivity, rapidity, the ability to analyze mixtures of substances without prior separation or determination of trace amounts of substances, fairly high reproducibility.

In analytical chemistry in recent years, chemometric methods have become actively used, which allow using mathematical and statistical methods to obtain chemical information, describe chemical systems and changes in them. These methods also

solve the problems of classification, identification and recognition of the studied objects.

In this work new information about the voltammetric behavior of biologically active substances solutions (B vitamins) under the conditions of data array accumulating was obtained. The data array for one sample included 1600 * 100 values under reproducibility conditions with 3 parallel measurements. The score plot of the PCA-modeling show that the samples of biologically active substances form clear clusters on the plane of the principle components PC1-PC2. In some cases, information was contained up to PC8. Useful information is provided throughout the potential range. So the information in the capacitive and Faraday regions is different, which is confirmed by the difference in the relative positions of the clusters. The background electrolyte solution is characterized by a relatively large scatter of points in the clusters.

Acknowledgement

P10. Virtual screening of ligand-based indole database

Pavel Majek[1], Jana Ballekova[2], Barbora Chrastinova[3], Magdalena Majekova[2]

[1]Institute of Analytical Chemistry, Faculty of Chemical and Food Technology, Slovak University of Technology, Radlinskeho 9, 81237 Bratislava, Slovakia

[2]Centre of Experimental Medicine, Institute of Experimental Pharmacology and Toxicology, Slovak Academy of Sciences, Dubravska cesta 9, 841 04 Bratislava, Slovakia

[3]Faculty of Pharmacy, Comenius University, Odbojarov 10, 832 32 Bratislava, Slovakia

Indole-acetic acid moiety is known as a useful starting fragment for design of efficient aldose reductase inhibitors. Starting from the group of indole-1-acetic acid derivatives, we have already identified three compounds with prospective inhibition properties, high selectivity towards aldehyde reductase and acceptable distribution properties: 5-carboxy-3-mercapto-1,2,4-triazino-[5,6-b]indole, [5-(benzyloxy)-1H-indol-1-yl]acetic acid and 2-(2-(ethoxycarbonyl)-8-methoxy-1,2,3,4-tetrahydro-pyrido [4,3-b] indol-5-yl) acetic acid. In order to assess distribution and toxicity potential of indole-acetic acid derivatives in designing

process, we performed ADMET and pharmacology prediction study of more than 500 structures constructed with indole skeleton substituted with (H, CH_2COOH, CH_3, C_2H_5, OCH_3, NH_2, $N(CH_3)_2$, OC_6H_6, $OCH_2C_6H_6$, SH) at R_1 – R_7 indole positions.

Our calculations revealed several important potential features for the compounds studied. In most cases, our compounds have low affinity towards human serum albumin. According to the values of MDCK parameter, derivatives with SH substituent at R_2 are able to penetrate readily to brain. Significant cardiotoxicity was predicted for 32 derivatives, with SH group at R_2 and OC_6H_6 or $OCH_2C_6H_6$ substituent at R_4, R_5 or R_6. The number of probable metabolic reactions was 4 or less for 439 compounds, which indicates low possibility of drug interactions. Most of the compounds have good oral absorption and 53 compounds achieved 100% absorption. The pharmacophore indole-based model was used for inhibitory prediction. Finally, the pharmacokinetic and pharmacodynamics prediction data have been assessed by means of cluster analysis and several compounds have been proposed for synthesis and inhibition activity experiments.

Acknowledgement

P11. Fluorometric determination of the total content of humic substances in Arctic seawaters by PLS regression

Krylov I.N., Puiman M.S., Labutin T.A.

Lomonosov Moscow State University, Moscow, Russia

Natural dissolved organic matter (DOM) is the main (~ 95%) form of organic carbon in the oceans, and, therefore, it is the important component of marine ecosystems and the key chain in the carbon cycle. The production of plankton and algae, and organic matter from land (terrigenous) are the main sources of DOM in the ocean. The basis of terrigenous organic matter is humic substances (HS). They have a

significant effect on the color of water, and, on the one hand, they can limit photosynthesis by absorbing light in the blue region, and on the other, HS limit the negative effects of UV radiation on plankton absorbing the UVB (280-320 nm) and the UVA (320-400 nm) bands of sunlight. Simultaneously the quantitative assessment of HS is very important for remote sensing (for example estimating chlorophyll concentrations using satellite imagery) in the visible range due to the effect on the optical properties of water. Rapid optical techniques are best suited for the HS quantitative determination. It is known that the DOM content has a good with absorption coefficient at 350 nm and the UV portable lidars (fluorescence) are used for the evaluation of DOC content in the surface layer. However, the fraction of the HS in DOM is not commonly estimated. It seems appropriate to use multivariate calibration models for the determination of the total HS content due to their diversity resulting in complex shape of fluorescence spectra.

The Latin hypercube provides a low correlation of factors in design even for a small number of experiments. This allowed us to use a compact training set of 35 samples for 5 factors. Although, we measured EEM fluorescence spectra, initially we constructed PLS calibration models based on fluorescence spectra at typical excitation wavelengths of lidars and the most common excitation wavelength in the OpenFluor database of natural fluorophores. Thereafter we have applied sparse PLS for the full EEM spectra. We selected the best models (1-4 principle components) in terms of prediction error by leave-one-out cross-validation. The models also provide physically significant loadings. The good predictive ability of the obtained calibration models was proved by a good correlation of the predicted HS contents with the dissolved organic carbon content in an independent set of Arctic determined by the combustion method.

P12. Discrimination of Tea by the Electrochemical Determination of its Antioxidant Properties by a Polyaniline–DNA–Polyphenazine Dye Modified Glassy Carbon Electrode

T. N. Kulikova[1], A. V. Porfireva[1], V. V. Vorobev[2], A. A. Saveliev[3], G. K. Ziyatdinova[1], G. A. Evtugyn[1]

[1]Chemistry Institute named after A. M. Butlerov of Kazan Federal University, Kazan, Russian Federation

[2]Interdisciplinary Center of Analytical Microscopy of Kazan Federal University, Kazan, Russian Federation

[3]Institute of Environmental Sciences of Kazan Federal University, Kazan, Russian Federation

Sensors based on electroconductive and electroactive polymers offer new opportunities for detection of biologically actives species based on their abilities to differently affect the redox properties of the matrices. We have proposed a family of polyaniline based sensors with DNA applied as a template / counter ion and phenazine dyes as selectors.

The interaction of the DNA molecules with polyaniline was confirmed by fitting a model of empirical representation of the oxidation of the polyaniline containing DNA caused by electrode polarization. The model describes interactions between the redox centers of polyaniline and takes into account both the faradic and the capacitance currents. The model analysis showed that the energy of interaction between the redox centers of the polymer depended on the concentration of the DNA added to the reaction media on the aniline electropolymerization stage.

The DNA-polyaniline sensors were utilized for tea brands classification based on intrinsic redox activity measured by direct current voltammetry. To separate the tea brands, the polyaniline-DNA layer was additionally covered with electropolymerized phenothiazine dyes (poly(methylene blue), poly(methylene green), poly(neutral red). The tea classification was first confirmed by principal component analysis utilizing redox peak currents (about ninety one % of explained variance for first two principal components). In parallel, the antioxidant capacity of the tea fusions was determined by a coulometric titration with electrogenerated bromine. A linear mixed model was fit using cathodic and anodic peak currents of

individual sensors based on polyaniline-DNA with three above mentioned coatings. The classification of the five tea brands based on the peak currents recorded on the electrodes with polymer coatings was performed using linear discriminant analysis. In this method, a linear combination of features (peak currents) that separate points belonging to specific tea brands is established by linear regression. The resulting combination was used as a linear classifier to predict the tea brand of unknown samples.

The electrochemical sensors based on combination of various electropolymerized materials can be then applied for the assessment of the quality of foodstuffs and beverages in accordance with their antioxidant capacities.

Acknowledgement

P13. Real-life application of the drift correction protocol for water quality monitoring

M.M. Khaydukova [1,2], _V.V. Panchuk [1]_, _A.V. Legin [1,2]_, _D.O. Kirsanov [1,2]_

[1]_Laboratory of chemical sensors, Institute of Chemistry, St. Petersburg State University, Universitetskaya nab 7/9, Mendeleev Center, St. Petersburg, Russia_

[2]_ITMO University, Kronverkskiy pr. 49, St. Petersburg, Russian Federation_

The analytical systems for real-time monitoring of air and water quality are highly demanded to ensure the ecological safety. Multisensor systems based on different principles are the promising analytical tools for monitoring of chemical composition of water. The most popular sensors for such systems are electrochemical ones due to their low cost and simplicity of handling. The analytical signals obtained from the array of such sensors have to be processed with chemometric techniques for quantitative or qualitative assessment of water. The multisensor approach is attractive since these systems can be implemented directly in the natural water source.

The response of the potentiometric sensor is defined by two components, related to the sensor, according to Nikolsky-Eisenmann equation. It contains the standard potential of the sensor and the slope of the sensor's calibration curve. Both

components are not constant since sensor's sensitive membrane can be affected by the interaction with sample solution (e.g. by surface absorption of the lipophilic components, by leaching of the membrane components). Therefore potentiometric sensors have to be calibrated right before the application. It is impractical to perform calibration daily in case of the monitoring task especially when the system is installed out of the laboratory. It is even more complicated in case of the multivariate calibration since the required number of samples with representative composition is quite large.

Several mathematical approaches were suggested for the correction of the sensors array response and it helped to exclude everyday calibration. However, most of them were tested with the static protocol of the measurement. Their applicability and efficiency in case of online measurements were not analyzed. This study addresses the application of the drift correction algorithms in case of long-term monitoring of the natural water composition.

The following experimental setup was used. The plastic vessel (4.0 L) was equipped with pump for flow simulation. It was filled with water from the river. Water in the vessel was changed every 3-4 days. Array with 19 cross-sensitive potentiometric sensors and standard reference electrode was immersed into the vessel. The responses of the sensors were registered every 10 seconds, except of the period for the calibration and water replacement. Aliquots of the water from the vessel were collected in parallel to the potentiometric measurements. These samples were analyzed with atomic emission spectrometry with inductively coupled plasma (ICP-AES) for quantitative assessment of their elemental composition. Every 7-10 days potentiometric sensors were calibrated with two standard solutions (Na_2SO_4 and $Ca(NO_3)_2$) in the concentration range 10^{-5} to 10^{-2} M. The overall duration of the experiment was 3 months.

The part of the data from potentiometric measurements and corresponding results of ICP-AES were used for PLS modeling to relate the response of the sensor array with chemical composition of the samples. The rest of the data were employed for drift correction studies. Descriptions of algorithms and obtained results will be presented.

Acknowledgement

P14. Soft sensor method for the prediction of the hydration state of Na-bentonite using multivariate analysis of low-field NMR data

M. Csontos[1], _D. Nyul[1]_, _I. Bányai[1]_, _J. Elek[2]_

[1]_University of Debrecen, Department of Physical Chemistry_

[2]_Science Port Kft_

One of the key features of clay minerals which makes them potential engineered barriers is their exceptional water isolation ability due to swelling. During hydration the clay mineral expands, and many physical properties change by the change of its pore structure.

In this study, water transverse T_2 relaxation times were measured by low-field NMR relaxometry, which is a versatile form of the NMR devices providing unique structural information of practically any proton containing materials. The tests were carried out both in homogenous and permanent gradient magnetic fields at different water-bentonite ratios. The relaxation times depend on numerous effects, but the T_2 change of the absorbed water protons provides us information about the hydration state and the structural change of the bentonite. While the results obtained in homogenous field give information about the fast chemical exchange processes, the measurements in gradient field allow the determination of the diffusion coefficient of the water beyond the wetting processes. The other advantage of the used permanent gradient probe head compared to the high field instruments is the portability. Since one of the most promising applications of the bentonites is the use as engineered barriers, the possibility of the on-field measurements by the NMR-MOUSE (Mobile Universal Surface Explorer) can be very helpful.

The most often applied evaluation method for relaxometry data is the inverse Laplace transformation by Multi Exponential Relaxation Analysis (MERA). The usual Tikhonov regularization of the relaxation decay data allows us to obtain the relaxation time distribution, but MERA is time consuming, and in some cases – especially at low Echo-times – mathematical artefacts appear which make the analysis more difficult and needs experienced analyst to avoid false conclusions.

Our goal is to develop alternative, less ambiguous data evaluation methods. Multivariate data analysis is a potential powerful tool to build a soft-sensor model

for the prediction of the hydration state. From the matrix of the T_2 relaxation decays measured at each water-bentonite ratio, Partial Least Square regression model was built in order to predict the hydration degree, and as a consequence, additional physical properties, such as water content or diffusion coefficient of the absorbed water can be obtained from a simple and fast measurement.

Acknowledgement

P15. Identification of origin of paracetamol tablets by potentiometric multisensor system and chemometrics

Regina Belugina[1], Yulia Monakhova[2,3], Alexander Becht[4], Curd Schollmayer[4], Ulrike Holzgrabe[4], Dmitry Kirsanov[5]

[1]Laboratory of Artificial Sensory Systems, ITMO University, St. Petersburg, Russia
[2]Spectral Service AG, 50996 Köln, Germany

[3]Institute of Chemistry, Saratov State University, Astrakhanskaya Street 83, 410012 Saratov, Russia

[4]University of Würzburg, Institute for Pharmacy and Food Chemistry, 97074 Würzburg, Germany

[5]Institute of Chemistry, St. Petersburg State University, St. Petersburg, Russia

Paracetamol is on the essential drug list and WHO Model List of Essential Medicines. It is a common medication which is used to treat pain and fever in many countries under a number of different brand names. Generic versions of the paracetamol are widely available. Generic drugs are considered as a good way to allow access to safe, effective and high-quality drugs at affordable prices to a vast majority of patients. However, the growth of the number of generics is a problem, because their efficiency is not often the same, as different source materials employed for production may have different properties influencing final product quality. In addition, many pharmaceutical companies produce medicine in different countries to decrease the costs, which can change drug properties. A certain part of medicines supplied in developing countries is counterfeit. With respect to the mentioned issues the development of fast and simple approach for identification of the origin of medicinal

preparation or manufacturers is demanded. Numerous methods were proposed to detect individual components and properties of a drug, however, most of them are time-consuming and require sophisticated equipment (various modes of chromatography, NMR, IR- and UV- spectroscopy, etc.).

The purpose of this study was to explore the potentiometric multisensor systems as a possible tool for identification of paracetamol formulations. 72 paracetamol formulations produced by 53 companies from 20 different countries were measured by potentiometric multisensor system. The data were analyzed by means of various chemometric methods to assess the potential of the multisensor system in correct classification of the samples. The results obtained with the suggested approach will be given in presentation.

P16. Towards programming the potentiometric sensor sensitivity patterns using multi-ionophore sensor membranes

J. Ashina[1], V. Babain[2], A. Legin[1], D. Kirsanov[1]

[1]Institute of Chemistry, St. Petersburg State University, Mendeleev Center, Universitetskaya nab.7/9, 199034 St. Petersburg, Russia

[2]ThreeArc Mining Ltd., 115184 Moscow, Russia

Potentiometric "electronic tongue" comprised of polymeric membrane sensors is among the most extensively employed analytical instruments in many fields related to analysis of aqueous solutions in food industry, technological monitoring, waste waters etc. [1]. The variability of the sensitivity patterns within the sensor array, combined with multivariate data processing techniques, can provide for simultaneous determination of mixture constituents, see e.g. [2].

In this work, we consider an approach of mixing several ionophores within one polymeric membrane for programming the potentiometric sensor sensitivity pattern. This approach is aimed at developing of time-saving and robust membrane composition design procedure.

The proposed multi-ionophore sensor array contains 5 types of ligands previously used as extracting agents and capable of binding rare-earth element (REE) cations.

Further details will be demonstrated in the presentation.

References

1. D. Kirsanov, D.S. Correa, G. Gaal, A. Riul, M.L. Braunger, F.M. Shimizu, O.N. Oliveira, T. Liang, H. Wan, P. Wang, E. Oleneva, A. Legin, Electronic Tongues for Inedible Media, Sensors 19(23) (2019) 5113.

2. J. Saurina, E. López-Aviles, A. Le Moal, S. Hernández-Cassou, Determination of calcium and total hardness in natural waters using a potentiometric sensor array, Anal Chim Acta 464(1) (2002) 89-98.

P17. Influence of social aspects on the disposal of municipal solid waste

Elena Gubar, Nikolay Gladyshev, Andrey Bogomolov

Samara State Technical University, Samara, Russia

The purpose of the work is to check the hypothesis of the influence of the efficiency of the financial sector and the level of public welfare of the state on the volume and ways of utilizing municipal solid waste (MSW).

A multivariate analysis of the relationship between various social aspects and the process of utilization of MSW has been performed. It has shown the influence of economic, cultural, legislative and environmental factors on this relationship.

Statistical indicators for 26 countries were the starting point for the principal component (PCA) analysis.

The PCA scores plot reveals groups of countries formed in accordance with their population density.

The study has shown some important trends. For example, material processing, composting and anaerobic digestion of MSW are mainly used in the countries with a high population density and the countries with lower level of corruption use energy recovery to a large extent.

The results obtained suggest that there is a relationship between the quality of education, the level of corruption, the management efficiency, the degree of citizen participation in the political life of the country, and the methods of utilizing MSW.

P18. Modelling of emission spectra of homogeneous and inhomogeneous laser plasma for its composition determination

S.M. Zaytsev, I.N. Krylov, T.A. Labutin

Lomonosov Moscow State University, Moscow, Russia

Laser-induced breakdown spectrometry (LIBS) is an actively developing technique for rapid elemental analysis. The emission of laser-induced plasma resulting from interaction of a high-power pulsed laser radiation with a sample is used for both quantitative and qualitative analysis. The optical excitation and registration of plasma radiation allows a high-performance remote elemental analysis of any substance, including those in aggressive environments. Since the analysis of unknown remote objects is the key advantage of the technique the use of standard samples is not appropriate solution. In this regard, the development of robust variants of LIBS analysis without CRMs is highly desirable.

In this work, we considered the possibility of the use of multi-parametric (plasma parameters and plasma elemental composition) optimization for spectra approximation. We used the NLopt library of nonlinear optimization algorithms to obtain a solution. In this work we have considered only derivative-free optimization algorithms, since in our implementation of the model the synthetic spectrum cannot be represented in analytical form as a function of plasma parameters, and it is not possible to provide adequate initial values for the most varied parameters.

To verify the suitability of the suggested algorithm, we approximated the specially generated synthetic spectra of steel (range 190–470 nm) and aluminum alloy (range 250–480 nm) at $N_e = 10^{16,7} cm^{-3}$ and $T = 0,8 eV$. Since pronounced self-reversed lines are present in the spectrum of the aluminum alloy, a two-zone model was also tested to thoroughly describe both the hot core and the cold periphery of plasma. Slight variation of each parameter (T-0.005 eV; logarithm of electron density - 0.005; radius of the plasma zone - 0.1 mm; the value of element concentration - 1%) served as stopping criterion for successfully completed optimization.

Convergence of the multi-parametric optimization has been shown in the case of synthetic spectra, which allows the use of the algorithms for studying plasma and determination of its composition.

Acknowledgement

Participants

Aidene Soraya
St. Petersburg State University
PhD student
St. Petersburg, Russia
soraya.aidene@gmail.com

Akhmetzhanov Timur
Lomonosov Moscow State University
PhD student
Moscow, Russia
atimur@laser.chem.msu.ru

Ashina Julia
St. Petersburg State University
Postdoc
St. Petersburg, Russia
ashina.julia91@gmail.com

Belugina Regina
PhD student
ITMO University
St. Petersburg, Russia
reginabelugina@yandex.ru

Bian Xihui
Tiangong University
Associate professor
Tianjin, China
Radboud University
Research scholar
Nijmegen, Netherlands
bianxihui@163.com

Bogomolov Andrey
Endress+Hauser Conducta GmbH+Co
Chemometrician
Aalen, Germany
ab@globalmodelling.com

Borovkova Ekaterina
Saratov State University
Associate professor
Saratov, Russia
rubanei@mail.ru

Bratchenko Ivan
Samara National Research University
Associate professor
Samara, Russia
iabratchenko@gmail.com

Brazovsky Konstantin
Tomsk Polytechnic University
Professor

Tomsk, Russia
mbc@tpu.ru

Burmistrova Natalia
Saratov State University
Professor
Saratov, Russia
naburmistrova@mail.ru

Cardoso Victor
University of Campinas
Researcher, Master student
Campinas, Brazil
v212089@dac.unicamp.br

Csontos Máté
University of Debrecen
PhD student
Debrecen, Hungary
csontos.mate@science.unideb.hu

Elek János
Science Port Ltd
Manager
Debrecen, Hungary
elek@scienceport.hu

Ferreira Victor Hugo
PhD student
University of Campinas
Campinas, Brazil
vhcferreira@outlook.com

Gubar Elena
Samara State Technical University
PhD student
Samara, Russia
e.v.gubar@mail.ru

Ishbulatov Yurii
Saratov State University
Junior researcher
Saratov, Russia
Ishbulatov95@mail.ru

Kalambet Yuri
Ampersand Ltd.
General director
Moscow, Russia
kalambet@ampersand.ru

Karavaev Anatoly
Institute of Radio Engineering and
Electronics of RAS

Senior researcher
Saratov, Russia
karavaevas@gmail.com

Khaydukova Maria
St. Petersburg State University
Assistant
St. Petersburg, Russia
khaydukova.m@gmail.com

Kirsanov Dmitry
St. Petersburg State University
Professor
St. Petersburg, Russia
d.kirsanov@gmail.com

Krylov Ivan
Lomonosov Moscow State University
PhD student
Moscow, Russia
ikrylov@laser.chem.msu.ru

Kucheryavskiy Sergey
Aalborg University
Associate professor
Esbjerg, Denmark
svkucheryavski@gmail.com

Kulikova Tatiana
Kazan Federal University
PhD student
Kazan, Russia
wefy2009@ya.ru

Labutin Timur
Lomonosov Moscow State University
Associate professor
Moscow, Russia
timurla@laser.chem.msu.ru

Majek Pavel
Slovak University of Technology
Teacher
Bratislava, Slovakia
Pavel.majek@stuba.sk

Maksyutova Elza
Bashkir State University
PhD student
Ufa, Russia
elzesha@gmail.com

Małyjurek Zuzanna
University of Silesia

PhD student
Katowice, Poland
zuzanna.mitrega@op.pl

Marini Federico
Sapienza University of Rome
Professor
Rome, Italy
federico.marini@uniroma1.it

Martynko Ekaterina
St. Petersburg State University
Master student
St. Petersburg, Russia
ekaterina.martynko@gmail.com

Monakhova Yulia
Spectral Service
Researcher
Cologne, Germany
yul-monakhova@mail.ru

Oleneva Ekaterina
St. Petersburg State University
Researcher
St. Petersburg, Russia
ekaterina.oleneva@inbox.ru

Olivieri Alejandro
National University of Rosario
Professor
Rosario, Argentina
olivieri@iquir-conicet.gov.ar

Pakhomov Andrey
Ltd. "Chromos Engineering"
General director
Moscow, Russia
director@has.ru

Panchuk Vitaly
St. Petersburg State University
Associate professor
St. Petersburg, Russia
vitpan@mail.ru

Pidenko Pavel
Saratov State University
PhD student
Saratov, Russia
pidenkops@gmail.com

Pomerantsev Alexey
Institute of Chemical Physics RAS

Principal researcher
Moscow, Russia
alexey.pomerantsev@gmail.com

Rodionova Oxana
Institute of Chemical Physics RAS
Leading researcher
Moscow, Russia
oxana.rodionova@gmail.com

Romanenko Sergey
Tomsk Polytechnic University
Professor
Tomsk, Russia
svr@tpu.ru

Rudnitskaya Alisa
University of Aveiro
Researcher
Aveiro, Portugal
alisa@ua.pt

Rusanova Tatyana
Saratov State University
Head of department
Saratov, Russia
tatyanarys@yandex.ru

Rutledge Douglas
AgroParisTech
Emiritus Professor
Paris, France
utledge@agroparistech.fr

Samokhin Andrey
Lomonosov Moscow State University
Researcher
Moscow, Russia
andrey.s.samokhin@gmail.com

Shestopalova Natalia
Saratov State Medical University
Associate professor
 Saratov, Russia
shestopalovanb@yandex.ru

Sherstobitov Danil
Samara State Technical University
Student
Samara, Russia
sherstobitovdn@gmail.com

Skvortsov Alexej
Peter the Great St. Petersburg

Polytechnic University
Professor
St. Petersburg, Russia
colbug@mail.ru

Soboleva Polina
Saratov State University
student
Saratov, Russia
sobpolina@yandex.ru

Surkova Anastasiia
St. Petersburg State University
Postdoc
St. Petersburg, Russia
melenteva-anastasija@rambler.ru

Sushkov Nikolai
Lomonosov Moscow State University
PhD student
Moscow, Russia
nikolaisushkov@yandex.ru

Teixeira Carlos
University of Campinas
PhD student
Campinas, Brazil
calb.teixeira@gmail.com

Titova Anna
Information and Methodological Center for
Expertise, Stocktaking and Analysis of
Circulation of Medical Products, Roszdravnadzor
Head of Department
Moscow, Russia
titova1701@yandex.ru

Walczak Beata
University of Silesia
Full professor
Katowice, Poland
beata.walczak@us.edu.pl

Zagitova Liana
Bashkir State University
PhD student
Ufa, Russia
kabirova.lian@yandex.ru

Zhilin Sergey
Altai State University
Engineer
Barnaul, Russia
szhilin@gmail.com