**Drushbametrics Project**

**Tenth Winter Symposium on Chemometrics**

# Modern Methods of Data Analysis



Russia, Samara, February 29–March 4, 2016

Samara State Technical University
Institute of Chemical Physics
Russian Chemometrics Society



Tenth Winter Symposium on Chemometrics

# Modern Methods of Data Analysis

## The organizing committee

**Co-chairmen**

Yuri Zolotov

Dmitry Bykov

**Secretary**

Anastasiia Melenteva

Natalia Ryumina

**Members**

Andrey Bogomolov

Kim Esbensen

Sergey Kucheryavskiy

Federico Marini

Alexey Pomerantsev

Oxana Rodionova

Sergey Zhilin

## Address

SSTU, Molodogvardeyskaya St. 244, 443100 Samara, Russia

http://wsc.chemometrics.ru/wsc10        e-mail: wsc10@chemometrics.ru

# Thanks

The WSC-10 organizers and participants wish express greatest appreciation to the following conference sponsors for their valuable economic and friendly help:

art photonics GmbH

Global modelling

LLC FOSS Electric

American elements

CSort company

EnSpectr LCC

Finally, we are grateful to all the WSC-10 attendees, lecturers, accompanying persons, and visitors for their interest to the conference.

**See you again at the next WSC-11 conference!**

# Useful information

## Conference and activities

Conference sessions will be placed at the main building. Russian banya (sauna) and cross-country ski will be available by request.

## Meals

Breakfast, lunches and dinners as well as conference banquet will take place at the dining room in main building.

## Scores & Loadings

Traditional "Scores and Loadings" meeting, will again be organized in the old-fashioned, self service Kostroma style and will take place in main building.

## Communication

The main three Russian cellular networks, Beeline, MTS and Megafon, have a proper coverage around the hotel. WiFi Internet access is available in the main building.

## Money

You may exchange EURs and US dollars to Russian rubles (RUR) in Samara banks, at the railway station, or in the airport.
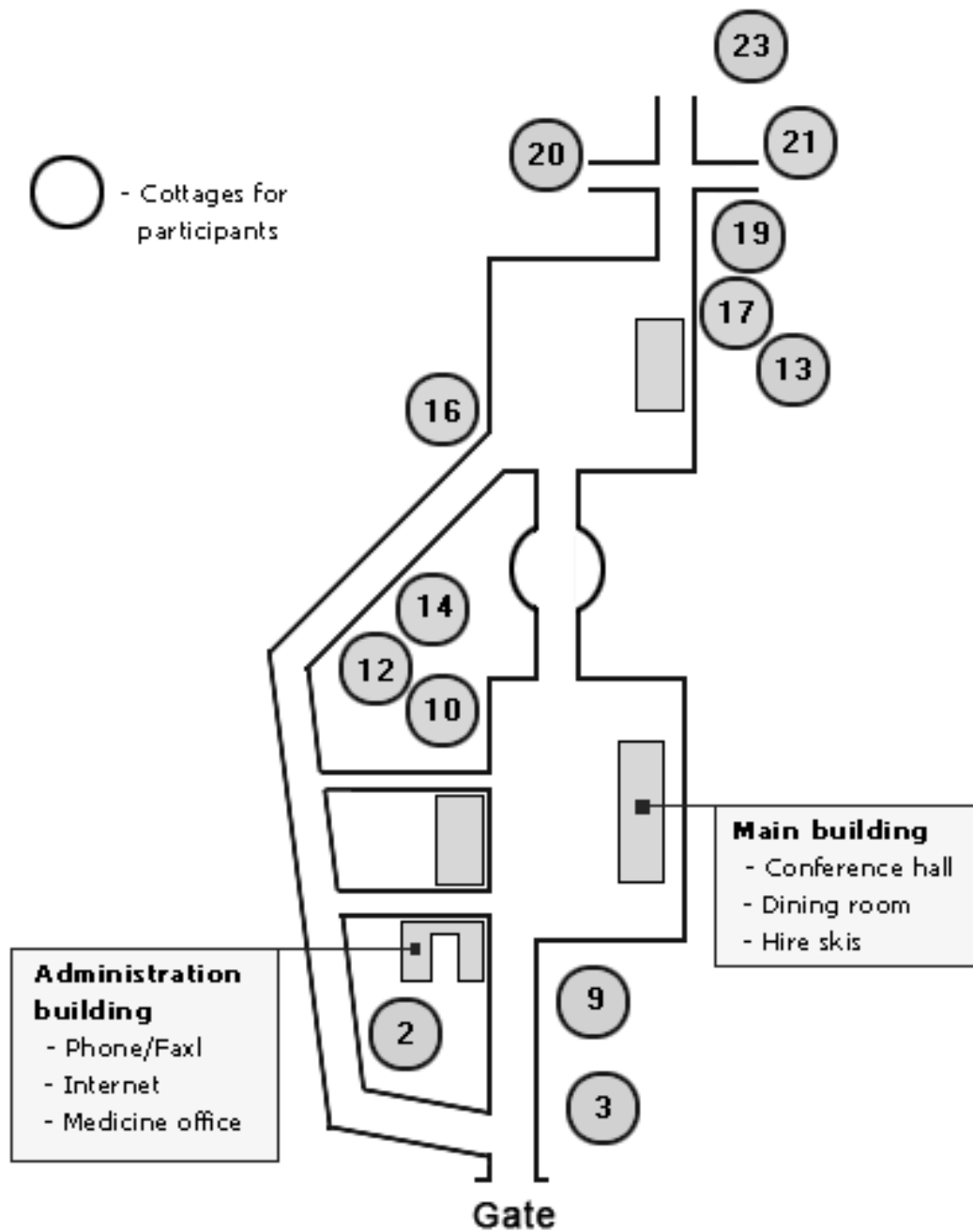
## Excursion

Friday morning, there will be an excursion to the famous Stalin's bunker from the World War II. The bunker was built in 1942 in nine months by 800 engineers and 2900 workers who worked secretly.  The bunker remained hidden from the locals until it was declassified in 1991 and turned to a museum.

## Miscellaneous

The conference official language is English.

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

# Scheme of "Polytechnic" campus

- Cottages for participants

23
20
21
19
17
13
16
14
12
10

**Main building**
- Conference hall
- Dining room
- Hire skis

**Administration building**
- Phone/Faxl
- Internet
- Medicine office

2

9

3

Gate

## Useful phone numbers

Anastasiia Melenteva,  *conference secretary*          +7 927 745 49 32. (mobile)

## Monday, February 29, 2016

| Time | Session |
|------|---------|
| 09:00–12:30 | **Registration** |
| 13:00–14:00 | **Lunch** |
| 14:00–14:30 | **Opening** |
| **Session 1** | **Process chemometrics** <br> **Chair:** Andrey Bogomolov |
| 14:30–15:30 | **L1** *Kim Esbensen* MSPC is not enough for proper Process Monitoring and Control — Chemometrics can do better: TOS variography |
| 15:30–16:00 | **T1** *Leon Rusinov* Online diagnosis of incipient faults at technological processes on the basis of nonlinear PCA and linear discriminant analysis |
| 16:00–16:30 | **Coffee break** |
| **Session 2** | **Methods developing** <br> **Chair:** Kim Esbensen |
| 16:30–17:00 | **T2** *Oxana Rodionova* How to solve authentication problems |
| 17:00–17:30 | **T3** *Yulia Monakhova* Hybrid chemometric approaches to increase efficiency of classification and data fusion techniques |
| 17:30–18:00 | **T4** *Federico Marini* Non-linear modeling revisited: advances in locally linear models |
| 18:00–18:30 | **T5** *Alisa Rudnitskaya* Applications of the ANOVA-Simultaneous Component Analysis to the multivariate data with underlying experimental design |
| 18:30–19:00 | **Free time** |
| 19:00–20:00 | **Dinner** |
| 20:00–00:00 | **Scores & Loadings** |

## Tuesday, March 1, 2016

| | |
|---|---|
| 08:00–09:30 | **Breakfast** |
| **Session 3** | **Methods developing**<br>**Chair:** Riccardo Leardi |
| 09:30–10:30 | **L2** *Jose Manuel Amigo* Multi-way analysis. Past, present and future |
| 10:30–11:00 | **T6** *Dmitry Kirsanov* An approach to calibration transfer based on latent variable space |
| 11:00–11:30 | **Coffee break** |
| 11:30–13:00 | **Free time** |
| 13:00–14:00 | **Lunch** |
| **Session 4** | **Curve resolution**<br>**Chair:** Oxana Rodionova |
| 14:00–15:00 | **L3** *Ludovic Duponchel* When MCR-ALS pushes back the limits of instrumentation |
| 15:00–15:30 | **T7** *Bruno Debus* Quantification of lanthanides in complex mixtures with X-ray fluorescence: three-point calibration models by correlation constrained MCR-ALS |
| 15:30–16:00 | **T8** *Evgeny Karpushkin* Using IR spectroscopy and multivariate curve resolution to elucidate mechanism of heat-induced decomposition of an organic complex |
| 16:00–16:30 | **Coffee break** |
| **Session 5** | **Software and algorithms**<br>**Chair:** Leon Rusinov |
| 16:30–17:00 | **T9** *Andrey Bogomolov* TP$^\mathrm{T}$ cloud: a full-featured chemometrics software online |
| 17:00–17:30 | **T10** *Vladislav Galyanin* Joint variable selection and preprocessing optimization approach to the multivariate calibration |
| 17:30–18:00 | **T11** *Alexandr Dokuchaev* Software for regression linear equations |
| 18:00–18:30 | **T12** *Sergey Kucheryavskiy* Speeding up PCA in R |
| 18:30–19:00 | **Free time** |
| 19:00–20:00 | **Dinner** |
| 20:00–00:00 | **Scores & Loadings** |

# Wednesday, March 2, 2016

| | |
|---|---|
| 08:00–09:30 | **Breakfast** |
| **Session 6** | **Spectroscopy**<br>**Chair:** Ludovic Duponchel |
| 09:30–10:30 | **L4** *Alexey Pomerantsev* Looking through a cover: Diffuse reflectance spectroscopy of hidden objects |
| 10:30–11:00 | **T13** *Yuri Kalambet* Peak integration: finally, how many points? |
| 11:00–11:30 | **Coffee break** |
| 11:30–13:00 | **Free time** |
| 13:00–14:00 | **Lunch** |
| **Session 7** | **Spectroscopy**<br>**Chair:** Oxana Rodionova |
| 14:00–14:30 | **T14** *Natalia Nechaeva* Multivariate data analysis in biomedical spectroscopy: cholinesterase activity determination |
| 14:30–15:00 | **T15** *Janos Elek* Method development and validation for the quantitative application of mid-range infrared spectroscopy in GLP environment |
| 15:00–15:30 | **T16** *Anastasiia Melenteva* Determination of fat and protein content in milk using optical spectroscopy in the range 400-1100 nm |
| 15:30–16:00 | **T17** *Jordi Cruz* Monitoring chemical and sensory parameters of tomato product with near infrared spectroscopy |
| 16:00–16:30 | **Coffee break** |
| **Session 8** | **Chair:** |
| 16:30–18:30 | **Poster Session** |
| 18:30–19:00 | **Free time** |
| 19:00–20:00 | **Dinner** |
| 20:00–00:00 | **Scores & Loadings** |

## Thursday, March 3, 2016

| | |
|---|---|
| 08:00–09:30 | **Breakfast** |
| **Session 9** | **Applications** <br> **Chair:** Alexey Pomerantsev |
| 09:30–10:30 | **L5** *David Hopkins* Using the derivative quotient method in regression |
| 10:30–11:00 | **T18** *Andrey Kotenko* Multi-criteria stochastic optimization of complex chemical production |
| 11:00–11:30 | **Coffee break** |
| 11:30–13:00 | **Free time** |
| 13:00–14:00 | **Lunch** |
| **Session 10** | **Applications** <br> **Chair:** Dmitry Kirsanov |
| 14:00–15:00 | **L6** *Riccardo Leardi* Application of N-way methods to environmental data from Genoa |
| 15:00–15:30 | **T19** *Vitaly Panchuk* X-ray fluorescence analysis of lanthanide mixtures using partial least squares regression |
| 15:30–16:00 | **T20** *Valeriy Babushkin* Application of NIR and MIR Foss solutions for food analysis |
| 16:00–16:30 | **Closing** |
| 16:30–17:00 | **Coffee-break** |
| 17:00–19:00 | **Free time** |
| 19:00–00:00 | **Banquet** |

## Friday, March 4, 2016

| | |
|---|---|
| 08:00–09:30 | **Breakfast** |
| 09:30–14:00 | **Excursion** |

# Abstracts

# L01. MSPC is not enough for proper Process Monitoring and Control — Chemometrics can do better: TOS variography

*Kim H. Esbensen[a,b]*

*[a]KHE Consulting, Copenhagen, Denmark*

*[b]ACABS research group, Department of Chemistry and Bioscience, Aalborg University, campus Esbjerg, Denmark*

Process monitoring and control in technology and industry is incomplete without full understanding of all sources of variation. It is not enough to be in command of Multivariate Statistical Process Control (MSPC) if the process data are affected by significant errors to an unknown degree. Process data are affected by both measurement errors as well as sampling or PAT sensor acquisition errors. These categories can dominate over analytical errors by factors 10-20-50(sic) when process sampling principles and methods for reduction are not taken into proper account. It is not sufficiently known in chemometrics that PAT signal acquisition gives rise to identical error types as physical sample extraction; the latter is well understood and solutions abound in the Theory of Sampling (TOS). This presentation will present the intimate analogy between the popular PAT approach and more conventional physical sample extraction (which is needed for calibration). In the process and production context it is essential to eliminate, or reduce maximally, all unnecessary contributions to the Total Measurement Uncertainty (MUtotal) in order to be able to perform valid process monitoring and control. This task is standard in the Theory of Sampling (TOS), which forms the only reliable scientific framework from which to seek resolution. A variographic process description allows quantification of the sum-total effect of all unwanted, unnecessary sampling/sensor acquisition errors (plus the conventional analytical error). Continuous variographic process data analysis, with problem-dependent updating, is able to issue warnings that a particular process measurement system is not, or is no longer fit-for-purpose, and must therefore be improved. This presentation calls for a process monitoring paradigm shift - variography is a missing link to MSPC, which needs a complete process monitoring system augmented by an updatable variogram facility. The example of self-contradictionary FDA regulatory sampling demands for critical pharmaceutical mixing processes is used to illustrate the above issues and claims.

## L02. Multi-way analysis. Past, present and future

*José Manuel Amigo*

*Department of Food Sciences, Faculty of Sciences, University of Copenhagen*

Multi-way methods have been applied in chemistry for many years now. From the first CANDECOM/PARAFAC proposals to the newest PARALIND, many algorithms and applications have been developed and published. Nevertheless, there are still many issues to face and aspect to improve when dealing with the application of multi-way methods in the real world. This presentation does not pretend to be an exhaustive revision of all multi-way methods and their algorithmic features. Instead, the presentation is aimed at showing the fundamental methods on which Multi-way analysis is based on (past), their current application field, benefits and drawbacks (present), and, most importantly, the improvements that would be needed to make Multi-way methods more reliable, fast and adaptable to any data-driven issue (future). This flow-chart in multi-way analysis will be conducted by showing real-world applications in various scientific fields (fermentation monitoring, counterfeit detections in foodstuff, metabolomics and pharmaceutical tablets monitoring) using different analytical methodologies (fluorescence landscapes, hyperspectral imaging and hyphenated chromatography).

## L03. When MCR-ALS pushes back the limits of instrumentation

*L. Duponchel*

*Laboratoire de Spectrochimie Infrarouge et Raman, Lille University, Lille, France*

Molecular spectroscopy is an essential tool for characterization of complex physico-chemical systems. This is mainly explained by the wealth of information available in spectra but also by the opportunity of *in situ* analysis with time and / or space resolution. Our insistence on seeking the truth often encourages us to push the limits of instrumentation sometimes without obtaining answers.

In such a case, many scientists consider that what cannot be done today will be possible tomorrow with the development of a new instrumental concept. Hopefully, this work will show that we can do many things right now with

chemometrics and instrumentation. Three different case studies will enable us to demonstrate the potential and the adaptability of MCR-ALS method. First, hydration shell of molecules will be probed with terahertz spectroscopy. In a second part, photochromic molecules will be explored with time-resolved spectroscopy. The latest case will consider mid-infrared spectroscopic imaging at the Soleil French Synchrotron Facility for single HeLa cancer cell characterization.

## L04. Looking through a cover: Diffuse reflectance spectroscopy of hidden objects

_A.L. Pomerantsev_, O.Ye. Rodionova

_Semenov Institute of Chemical Physics RAS, Moscow, Russia_

Investigation of samples' properties through layers of material of different thickness is demanded for the process control, for biochemical analysis, and for many other applications. Investigation of objects, which are located under a cover is based on the analysis of spectra acquired using the Diffuse reflectance near infrared (DR NIR) spectroscopy. Each spectrum is a composition of the useful (target) spectrum and the spectrum of the nuisance layer. Unfortunately, DR NIR is a more complicated technique than the conventional transmission spectroscopy. Therefore, we cannot simply subtract the spectrum of the cover from the cumulative spectrum. To separate these signals, we suggest using a new phenomenological approach, which employs the Multivariate Curve Resolution methods. We consider that the nuisance spectrum of the layer can be presented as a superposition of three signals $c_t s_t(\lambda) + c_a s_a(\lambda) + c_s s_s(\lambda)$, which are responsible for transmission, absorption and scattering of the light beam inside the layer. Spectra $s_s(\lambda)$ are not affected by the layer depth $h$, or the nature of the target. On the contrary, coefficients $c$ depend both on $h$, and on the target reflection factor $r$. Applying the conventional transmission spectroscopy a researcher is always interested to remove reflection and scattering effects from the spectra. The peculiarity of our approach is that we do not remove the scattering and reflection components, but study them in details with the aim to use this knowledge for the target spectrum recovering.

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

This concept is then utilized to obtain a hidden target spectrum in the real-world systems. In particular, we succeed in reconstruction of the target using a single spectrum obtained through several PE layers.

## L05. Using the Derivative Quotient Method in Regression

*David W. Hopkins*

*NIR Consultant, Battle Creek, MI, USA*

The original work of Karl Norris and William Hruschka in applying multiple linear regression (MLR) on derivative quotients has been reviewed using new software developed using Matlab. The method starts with full spectra for a set of samples used for calibration, and results in one or more terms in a prediction equation (Eq. 1) where the numerators (num) and denominators (den) of each term are the values of derivatives at individually selected wavelengths.

$Y_{pred} = B_0 + B_1 \, der(W_{num1})/der(W_{den1}) + B_2 \, der(W_{num2})/der(W_{den2}) + \ldots \text{Eq. 1}$

The derivatives are calculated by the Gap method pioneered by Norris, and each derivative is optimized for the center wavelength (W), the width of the derivative gap, and the width of "boxcar" smoothing, in both the numerator and the denominator of each term. The original work has been extended to utilize derivatives up to fourth order. The numerator and denominator derivatives do not have to be of the same order. The method is particularly effective for samples exhibiting spectra with a high range of baseline offsets and background slopes due to particle size variability.

Principles of developing calibration models and validating them will be presented. For measuring protein in wheat samples using NIR reflection spectroscopy, a prediction equation using a single term was sufficient to achieve

optimal prediction. It is interesting that even in a more complex system measuring the hemoglobin concentration in whole blood samples, the optimum performance of models is obtained with models using only 2 terms.

The Derivative Quotient models are very easy to implement. It is suggested that coupling the models with current chemometric methods for qualifying samples for analysis provides a powerful new technique for routine analysis of samples. This regression method appears to be capable of achieving greater accuracy than can be obtained utilizing the standard methods of pretreating the spectra and regression using PCR or PLS, possibly because multiplicative spectral effects are directly compensated by the quotient terms, while the chemical differences are efficiently calibrated by the MLR.

## L06. Application of N-way methods to environmental data from Genoa

*R. Leardi, E. Martini, E. Scalabrini*
*Department of Pharmacy, University of Genoa*

The amount of environmental data available becomes more and more impressive. Unfortunately, in the very great majority of cases they are analyzed one datum (not even one variable) at a time. This very crude approach leads to decisions that, in spite of having quite relevant effects on daily life, such as stopping the private car traffic or reducing the temperature in private flats during winter, do not have a strong effect on the reduction of pollution.

Obviously, a multivariate data analysis would lead to a much better understanding of the underlying phenomena and therefore to much more effective actions.

The data set under study has been collected in Genoa during 6 years (2007-2012), by measuring 3 variables (CO, NO and NO2) in 5 different sites once per hour. This makes a total of 789.120 data (6 years * 365/366 days * 24 hours * 3 variables * 5 sites).

It is obvious that N-way PCA is the best way of dealing with such a complex and highly structured data set. The data can be rearranged in different ways, in order to produce different structures with a different number of modes (three, four and five).

The application of several N-way models allowed to detect in a very clear way the effects related to sampling site, sampling year, seasonality, day of the week and hour of the day. Some conclusions were quite obvious and in total agreement with the previous knowledge, but some other results were quite different from what was given for sure by the "experts" of the problem.

It has been noticed that, in spite of the high complexity of the mathematical approach, the results can be shown by very simple and highly informative plots that can be understood also by people with no chemometric knowledge.

# T01. Online diagnosis of incipient faults at technological processes on the basis of nonlinear PCA and linear discriminant analysis

*M.R.Galiascarov, V.V.Kurkina, L.A.Rusinov*

*St.Petersburg State Technological Institute (Technical University), Saint Petersburg, Russia*

The Fisher linear discriminant analysis (LDA) was widely used in pattern recognition, but only recently it began to be used in diagnostic systems to identify faults of technological processes. However, the majority of technological processes in the chemical industry are non-linear and LDA works poorly. In these cases, the improving of classification can be expected when using nonlinear discriminant analysis (NDA).

NDA similarly to the nonlinear principal component analysis (NPCA) suggests at the first stage the mapping of the original data in a feature space of large dimension and then, at the second stage, the use of conventional LDA. In order to reduce the amount of computation the kernel functions are used as nonlinear functions for mapping.

However, the use of the NDA in on-line mode gives a low percentage of correct classifications, especially in the case of incipient faults due to strong elongation of their classes.

The combined method of diagnosis is offered in order to reduce these elongations. The continuous monitoring of the process is carried out by NPCA and the fault identification by LDA is started only after the detecting of their occurrence. Thus, the identification of a fault (its assignment to a particular class) is produced in the vicinity of the point of detection in compliance with the minimum distance between the vector of the current sample and the centroid of that class. For the confident classification the decision of the presence of the particular fault is taken after several consecutive identical definitions. At the same time, if the operator for any reason does not take any action to return the process to normal operation, the blocking of the LDA after successful classification of faults is proposed. It prevents further blurring of classes if the incipient faults will continue to develop.

These measures have enabled to obtain more compact classes and significantly increase the percentage of correct classifications especially in the case of incipient faults. The efficiency of the method has been studied while diagnosing slowly developing faults in the pyrolysis of hydrocarbons.

## T02. How to solve authentication problems

*O.Ye. Rodionova*, *A.L. Pomerantsev*

*Semenov Institute of Chemical Physics RAS, Moscow, Russia*

Pattern recognition encloses a variety of different methods and techniques. Each type of problem requires an application of relevant methods.

Authentication is the process of determining whether an object is, in fact, what it is declared to be. Typical authentication problems, relevant for analytical chemistry and chemometrics, are counterfeit drug detection, food adulteration detection, identification of illegal additives in fuels, and confirmation of geographical origin of products. Class-modeling methods also called one-class classifiers are well suited for solving such type of problems. These methods develop the acceptance area around the target class, and, thus, delimit the target objects from any other objects and classes.

Analysts often substitute authentication task with solving discrimination problems. A well constructed discrimination method will perfectly classify a new sample only if this sample is a member of one of the predefined classes. However, in case the new sample does not belong to any of such classes, the discriminant analysis is unable to properly define the membership of the sample. Thus, such techniques do a poor authentication job.

Our considerations are illustrated by a comparison of the results provided by the PLS-DA and DD-SIMCA in application to a real-world example.

# T03. Hybrid chemometric approaches to increase efficiency of classification and data fusion techniques

*Yulia B. Monakhova[a,b], Monika Hohmann[c,d], Svetlana P. Mushtakova[b], Norbert Christoph[d], Helmut Wachter[d], Bernd Diehl[b], Ulrike Holzgrabe[c], Douglas N. Rutledge[e]*

[b] *Institute of Chemistry, Saratov State University, Saratov, Russia*

[a] *Spectral Service AG, , Cologne, Germany*

[c] *Institute of Pharmacy and Food Chemistry, University of Würzburg, Würzburg, Germany*

[d] *Bavarian Health and Food Safety Authority, Würzburg, Germany*

[e] *UMR Ingénierie Procédés Aliments, AgroParisTech, Inra, Université Paris-Saclay, France*

In chemometrics it is common to apply a cascade of multivariate approaches to the dataset under study. For example, PCA and cluster analysis are often applied to detect outliers and provide exploratory analysis of the data before the application of more advanced chemometric approaches such as discriminant analysis. Another promising but yet unexplored research direction is to increase the efficiency of classical chemometric methods by synergetic combination of their advantages within a new approach.

In this regard, independent components analysis (ICA) as a preprocessing tool was combined with linear discriminant analysis methods (LDA and FDA). To illustrate the performance of the ICA/DA methodology, four representative nuclear magnetic resonance (NMR) data sets of wine samples were used. The classification was performed regarding grape variety, year of vintage and geographical origin. The average increase for ICA/DA in comparison with PCA/DA in the percentage of correct classification varied between 6±1-and 8±2%.

Second, a new algorithm was proposed for the supervised analysis of multiblock data structures. It associates the advantages of the ability of common components and specific weights analysis (CCSWA) to explain the maximum of variance in the data and PLS-DA framework to shift results towards the targeted property. The robustness of the classification results comparing with each single methodology (PLS-DA and CCSWA) was shown on the example of

the determination of tomato production type (organic/conventional) based on the fused $^1$H NMR, IR and stable isotope data ($^{13}C/^{12}C$, $^{15}N/^{14}N$, and $^{18}O/^{16}O$).

The use of hybrid chemometric approaches such as ICA/DA and PLSDA/CCSWA resulted in improved classification performance comparing with single approaches and increased efficiency of chemometric modelling.

## T04. Non-linear modeling revisited: advances in locally linear models

*Federico Marini*

*Dept. of Chemistry, University of Rome "La Sapienza", Rome, Italy*

The evolution of the analytical instrumentation, especially through the increase in the application of high throughput fingerprinting platforms, has led to the possibility that many sources of variation, other than the one(s) of interest, affect the instrumental signals, resulting in a non-linear relationship between the dependent and independent blocks or, in the case of pattern recognition, in a non-linear separability of the categories in the feature space.

In such cases, the implementation of non-linear modeling through a locally linear approach [1-2] can help tackling with the complex relationships between the data blocks to be analyzed, reducing the probability of overfitting but, at the same time, allowing a high flexibility, as the extent of non-linearity can be tuned by a proper selection of the model parameters (above all, the number of nearest neighbors to be included). In this communication, attention will be posed on Locally Weighted-PLS approach originally proposed by Centner and Massart [3] and to its classification analog (Locally Weighted PLS-DA) recently proposed by our group [4]. Indeed, although the concept of locally linear modeling has been present in the literature for almost thirty years, there are many issues which are still open. Accordingly, in the present communication key points such as the evaluation of the influence of the weighting and/or scheme adopted and the possibility of interpreting the model through the inspection of the local loadings, or the definition of which could be the best cross-validation scheme, especially in the case of highly structured data, will be addressed, and some possible solutions will be presented.

**References**

1. W.S. Cleveland, S.J. Devlin, J. Am. Stat. Assoc. 1988, 83, 596-610.

2. T. Naes, T. Isaksson, B.R. Kowalski, Anal. Chem. 1990, 62, 660-673.

3. V. Centner, D.L. Massart, Anal. Chem. 1998, 70, 4206-4211.

4. M. Bevilacqua, F. Marini, Anal. Chim. Acta 2014, 838, 20-30.

## T05. Applications of the ANOVA-Simultaneous Component Analysis to the multivariate data with underlying experimental design

*Alisa Rudnitskaya*

*CESAM and Chemistry Department, Aveiro University, 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. Generalization of Analysis of Variance (ANOVA) to the multiple variables – multivariate ANOVA or MANOVA is not applicable to the typical omics data sets as it breaks down due to the singularity of covariance matrices. Several approaches to the analysis of multivariate design data were suggested in the literature. Most of them are based on the decomposition of the data by PCA or PARAFAC, which is followed by the analysis of effects. This approach allows effectively reduce number of the variables and therefore number of the parameters to estimated and leads to the simpler, more parsimonious and more interpretable models. On of such methods ANOVA- Simultaneous Components Analysis (ASCA) enables the decomposition of the original data matrix into sub-matrices containing variance related to the design factors and their interactions [1]. PCA is run on each of the matrices and an ANOVA-like model is constructed from the PCA models of all effects and interactions. Sub-matrices are defined to have orthogonal column spaces therefore the column spaces of the sub-models fitted on these matrices are also mutually orthogonal. This allows interpreting all sub-models independently of each other. Since loadings are calculated for

each sub-model independently, contribution of the variables to every source of the variation in the data may be identified.

Applications of the ASCA to the analysis of the lipidomics, plant and fungal metabolomics and electronic tongue data sets will be discussed.

**References.**

1. A.K. Smilde, J.J. Jansen, H.C.J. Hoefsloot, R.-J.A.N. Lamers, J. van der Greef, M.E. Timmerman, Bioinformatics 21 (2005) 3043.

## T06. An approach to calibration transfer based on latent variable space

*V.V. Panchuk [1], <u>D.O. Kirsanov [1,2]</u>, E. S. Oleneva[1], M.M. Khaydukova[1,2], A.V. Legin[1,2]*
*[1]Institute of Chemistry, St.Petersburg State University, St. Petersburg, Russia*
*[2]Laboratory of artificial sensory systems, ITMO University, St. Petersburg, Russia*

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

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

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

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

The performance of the suggested method was tested with simulated data sets and with real data. In the latter case calibration models constructed for quantification of Co, Ni and Cu in their triple mixtures from energy-dispersive X-ray fluorescence measurements were successfully employed for prediction of these three metals from UV-Vis spectrometry data. Moreover the suggested method can be employed for traditional calibration transfer, e.g. between two NIR instruments.

It appears that the method can have a broad range of possible applications in analytical chemistry.

**References:**

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

# T07. Quantification of lanthanides in complex mixtures with X-ray fluorescence: three-point calibration models by correlation constrained MCR-ALS

*B. Debus[1], D.O. Kirsanov[1,2], V. V. Panchuk[1,2], A.A. Goydenko[1], V.G. Semenov[1], A. Legin[1,2]*

*[1]Institute of Chemistry, St. Petersburg State University, St. Petersburg, Russia*
*[2]Laboratory of Artificial Sensory Systems, ITMO University, St. Petersburg, Russia*

Total reflection X-ray fluorescence (TXRF) is a technique of choice for various tasks where high sensitivity and selectivity are required. From a single TXRF spectrum both qualitative and quantitative information can be derived based on the position and the intensity of the X-ray fluorescence radiation, respectively. Despite its high sensitivity, the interpretation of TXRF spectra can be challenging in situations where several chemical elements are analyzed simultaneously. Experimentally, the absorption of fluorescence radiation from one element by other chemical elements of the matrix and the overlapping of characteristic lines from neighboring elements may occur. One consequence of such so-called matrix effects is that low detection limit cannot be achieved which hinders the use of TXRF in certain applications.

In an attempt to improve the precision in quantification of individual elements in complex mixtures, first-order calibration methods have been proposed. Recently, partial least-squares (PLS) was proposed for the quantification of a six elements lanthanide mixtures from TXRF spectra [1]. The main advantage of multivariate methods like PLS compared with classical approaches such as ordinary least-squares is that matrix effects can be accommodated by the model yielding adequate calibration with better precision. The only pre-requisite to perform PLS is the number of samples which should be large enough to select a relevant calibration and test sets. This can be a limitation in situation where measurement costs are high or access to the measurement set-up is limited.

To overcome this limitation, we propose to use the correlation constrained MCR-ALS method recently suggested in [2]. This implies the implementation of known concentration values of the target analyte in calibration set during

the ALS iteration to reach unique solution. Compared with PLS, correlation constrained MCR-ALS yields similar precision but enables the recovery of pure spectra of individual lanthanides. The most intriguing feature is the possibility to simultaneously predict the concentration of all analytes from a set of three-point calibration models with RMSEP values around $1.10^5$ mol.L$^{-1}$. This promising result indicates that accurate quantification of chemical elements in TXRF spectroscopy can be achieved with a minimal number of calibration samples whereas larger amount of samples is required for PLS.

**References**

1. D. Kirsanov et al. SpectrochimicaActa Part B 113 (2015) 126–131

2. G. Ahmadi et al. Chemometrics and Intelligent Laboratory Systems 142142 (2015) 143–150

## T08. Using IR spectroscopy and multivariate curve resolution to elucidate mechanism of heat-induced decomposition of an organic complex

*Evgeny Karpushkin[1], Nataliya Gvozdik[1], Sergey Kucheryavskiy[2], Andrey Bogomolov[3]*
*[1]Moscow State University, Moscow, Russia*
*[2]Aalborg State University, Esbjerg, Denmark*
*[3]Art Photonics GmbH, Berlin, Germany*

Many practically important processes occur upon elevated temperatures, and the induced chemical transformations are accompanied by gases evolution: heat-induced degradation of plastics, conversion of acrylic fibers into carbon ones, and fossil fuels pyrolysis – to name just a few. Generally, the evolved compounds structure is relatively similar, and therefore their analysis by means of vibrational spectroscopy is a convenient approach towards elucidation of the underlying reactions mechanism as well as optimization of the process conditions. Modern analytical instruments provide the opportunity to carry out simultaneous thermogravimetry/differential scanning calorimetry analysis and IR/Raman/mass spectrometry investigation of the evolving gaseous products.

However, elucidation of the mechanism of the reactions occurring upon heating is not completely straightforward, due to a number of experimental limitations. They include, for instance, poor resolution of the spectral bands, their overlap with these of traces of water and carbon dioxide in the atmosphere, the lag between the TGA/DSC signal and the spectra registration, and the specimen inhomogeneity.

In this contribution we present and discuss the results of TGA-DSC-IR study of heat-induced decomposition of tris(acetylacetonato)manganese(III) complex. Using this process as an example, we have indicated the complications of the evolved gas analysis and demonstrated that they can be partially overcome taking advantage of multivariate curve resolution methods. Moreover, we have shown that the conventional methods based on TGA and IR analysis of the solid decomposition products produce incomplete data that is not enough for reliable elucidation of the decomposition mechanism. Finally, we have developed a workflow to extract the meaningful information from the set of the evolved gas spectra. The elaborated approach is fairly general and applicable for study of a variety of the processes accompanied by evolution of gaseous products.

## T09. TP$^T$ cloud: a full-featured chemometrics software online

_Andrey Bogomolov[1,2,3], Vladislav Galyanin[1], Vadim Belikov[1], Valeria Belikova[1]_
[1] _Samara State Technical University, Samara, Russia_
[2] _art photonics GmbH, Berlin, Germany_
[3] _Global Modelling, Aalen, Germany_

Chemometrics is one of the modern scientific disciplines that cannot be imagined without software. TP$^T$ cloud presented here is a new chemometrics toolbox that has been conceived and _ab ovo_ developed online using up-to-date web technologies.

The main advantages of having chemometrics "on a cloud" are related to the central hosting of the data and models that facilitates their use and share. The data can be easily uploaded from a number of spectroscopic and general-purpose formats and analyzed using a conventional web browser, independently on the hardware platform and operating system. The data analysis work is therefore automatically backed-up and synchronized. The model cen-

tral location enables its training using multiple data sources that can be spread all over the world, i.e. for a global modelling. Powerful computational resources that can be involved into the remote computing are necessary to solve modern "big data" problems. In the cloud chemometrics word heavy files are replaced by weightless links; the data and models can be shared, distributed and reported without cloning. This opens new capabilities for the team analysis, corporate work and scientific publishing.

TP$^T$ cloud is a rapidly developing chemometrics platform that is not limited to widely spread multivariate regression and classification tools. The university-based software team develops and implements their own novel multivariate algorithms. One of them is an intensive model optimization approach based on generalized interval selection with simultaneous preprocessing optimization.

Chemometrics software is sophisticated and should be created by scientists, but it should not be intended for scientists only. The final product of multivariate data analysis is a model that is aimed at solving various practical problems by different people, who may be not aware of factor spaces. Connecting specialists and users can be effectively performed in the internet and this is one of the main missions of tptcloud.com.

## T10. Joint variable selection and preprocessing optimization approach to the multivariate calibration

*Vladislav Galyanin[1], Andrey Bogomolov[1,2,3]*

*[1] Samara State Technical University, Samara, Russia*

*[2] art photonics GmbH, Berlin, Germany*

*[3] Global Modelling, Aalen, Germany*

Variable selection is a one of the most common and important treatments preceding the multivariate calibration, particularly, of spectral data. Irrelevant and noisy data parts are commonly removed from the calibration set in order to simplify multivariate models and improve their performance.

Nowadays, data scientists usually determine an optimal preprocessing method and perform variable selection separately, as two independent data pretreatment steps preceding the calibration. In general, however, the prepro-

cessing method, the optimal variables and even the model complexity (the number of latent variables, LVs) are related to each other. Sequential optimization may result in the selection of variables, which are only optimal for the chosen preprocessing method for a given number of LVs.

Variable selection is usually performed using the binary variable coding, as in conventional genetic algorithm (GA) or using a sequential way, like interval partial least-squares (iPLS) regression. Conventional GA optimization with the binary coding leads to an excessive complexity of the optimization problem, and sequential interval selection cannot guarantee finding the global optimal solution.

In the present work a generalized optimization approach for simultaneous determination of an optimal preprocessing method, variable intervals (including their widths) and the number of LVs is presented. The proposed approach is "time-dependent" but well generalized, has lower optimization complexity than common variable selection algorithms and converges to a global optimal solution in many cases.

Here, GA has been used as a parametric optimization routine in order to obtain optimal values of the merit function. Various calibration statistics are considered for quantification and discrimination cases. Problems of chromosome coding, genetic algorithm adaptation and constraints, interval width and spectral resolution, performance issues and merit function selection are discussed.

Code for Interval selection was written in GNU Octave/MATLAB and used in TPT cloud software (www.tptcloud.com) for online optimization of variable selection "on the cloud". All m-files are designed in a way to be optionally used as an independent standalone GNU Octave/MATLAB toolbox. The software (Interval Selection toolbox) has been registered in Russian Federal Service on Intellectual Properties.

# T11. Software for regression linear equations

*A.V. Dokuchaev*, *A.A. Kotenko*

*Samara State Technical University, Samara, Russia*

Prediction by means of multivariate regression analysis allows to solve the problem of multi-criteria optimization of road bitumen production with a set of contradictory characteristics of modern standards. Mathematical formulation of the problem is described in [1]. While deterministic problem is unsolvable because of the opposing claims of some objective functions, the standard ranges of product's characteristics of complex chemical production allow us to find the confidence intervals to control the technological parameters. As a result, one can determine the parameters of the production process and the composition of the raw material to produce bitumen with specified characteristics. Conflicting criteria for multi-criteria optimization can be described as a system of interdependent regression equations. Then the algebraic insolvability is possible to overcome by indication of the confidence intervals for the parameters of technology to ensure the release of non-defective product.

The solution is implemented in the software package [2]. It takes into account the experimental data of large samples and simulates various forms of interconnections between the control regressors and the objective functions. The developed software package allows to quickly carry out large series of experiments and analyze the relevant data.

We present the details of the mathematical problem and the features of designed software for submitting of a large number of control parameters and complex effect on the standardized characteristics of product.

## References

1. Kotenko A.P., Sharapova E.A. Bengina P.M. Multi-criteria optimization using systems of linear regression equations // International scientific-technical conference "Advanced information technologies", 28-30 April 2015 - Samara: Samara Publishing Science Centre, 2015 - P.276-277. (In Russian.)

2. Dokuchaev A. V., Kotenko A. A. The software package for multi-criteria optimization of production systems using regression equations. ROSPATENT. Application for state registration of computer programs №2015660647 from 09.11.2015. (In Russian.)

# T12. Speeding up PCA in R

*Sergey Kucheryavskiy*

*Department of Chemistry and Bioscience, Aalborg University, Esbjerg, Denmark*

Principal component analysis (PCA) is a widespread method for multivariate data exploration. In many cases, PCA is the first tool an analyst uses to get a feeling of the data structure. Analysis of hyperspectral images is not an exception here and many image analysis packages provide, among other tools, the possibility for exploring images with PCA.

Most of multivariate datasets used in chemometrics have a number of variables (significantly) larger than a number of objects (observations) and many algorithms are optimized for this case. When PCA is applied to a hyperspectral image, each pixel of the image is an individual observation (spectrum). Nowadays many laboratory hyperspectral cameras can obtain an image with hundreds of thousands of pixels. In remote sensing the images can easily have resolution up to several million pixels. Working with such images can easily lead to memory and computational issues if standard algorithms for estimation of principal components are employed.

In the present work author is going to share his experience of creating an interactive tool for exploratory analysis of relatively large hyperspectral images with PCA using R and JavaScript based user interface. The main goal was to make the implementation of PCA decomposition as fast as possible without substituting standard R tools (e.g. library for linear algebra operations), so the package is easily available for non-experienced users. For the same reason it was decided to write all code in R without using C/C++ at the current stage.

Three PCA state-of-art algorithms have been considered — singular value decomposition (SVD), eigenvectors of variance-covariance matrix and non-iterative partial least squares (NIPALS). Each algorithm was investigated in order to find the steps critical for working with datasets containing huge number of rows. The steps were then optimized (both algorithmically as well as an R implementation) to reduce the computational time. The results of optimization will be discussed.

# T13. Peak integration: finally, how many points?

*Kalambet Y.[1], Kozmin Y.[2], Samokhin A.[3]*
*[1]Ampersand Ltd., Moscow, Russia*
*[2]Institute of Bioorganic chemistry RAS, Moscow, Russia*
*[3]Chemistry Department, Moscow State University, Moscow, Russia*

Accuracy of peak integration algorithms for the case of stand-alone peak-wise function shape and constant data rate is revisited. It is shown, that the simplest rectangle rule gives the unbiased estimate of the true peak area, if we consider frame start position as a random variable. Most other formulas, such as Simpson's rule, either provide biased area for periodic noise or their area is equivalent to that obtained by rectangle rule. Number of points required for peak description with sufficient accuracy depends on the type of measurement (instantaneous or integrating) and on parameter under consideration (area, height, width, asymmetry, baseline width, …). Most important factor for area and height is the influence of proper baseline subtraction. Adaptive noise filtering significantly improves baseline detection and provides benefit to estimates made with the help of filtered data. Integration rules influence area in the case of unseparated peaks only, so that adjacent peaks exchange small parts of areas, but in this case, area errors cannot be compensated by such exchange.

It also has to be noted, that integration parameters, such as area or height, are not used directly; they first are used to construct calibration curve and then quantification is made. Many algorithmic distortions are compensated on this step.

The best results can be obtained by modelling peaks according to their shape, known in advance (e.g. Gaussian or EMG). The more is known about the shape of peaks in advance, the better the accuracy of approximation.

# T14. Multivariate data analysis in biomedical spectroscopy: cholinesterase activity determination

*Nechaeva N.L., Kurochkin I.N.*
*Chemical enzymology department, Faculty of Chemistry, Lomonosov Moscow State University, Russia*

Surface-enhanced Raman spectroscopy is a modern powerful tool for the analysis of complex biological samples. The use of silver pastes is one of the most common SERS applications on the practice. In the current work silver screen-printed SERS-substrates were used. This developed technique is inexpensive, easy-to-use and do not required special equipment. The application of the SERS-effect is expedient in the case of the analysis of complex biological objects, for example human blood enzymes. Acetylcholinesterase and butyrylcholinesterase catalyze the hydrolysis of many physiologically important compounds such as neurotransmitter acetylcholine, caffeine, and cocaine. The activity of these enzymes in the blood may be lowered due to poisoning by organophosphorus compounds, pesticides, and also due to various diseases. Therefore it is very important to measure cholinesterase activity in blood with high accuracy to assess toxicity and for diagnosis of diseases at early stages. The best known method for determining the activity of AChE and BChE is to measure butyrylthiocholine (BTCh) and thiocholine (TCh) – substrate and product of the enzymatic reaction, respectively. We used a silver print-screen surface as the SERS-active structure for the analysis of BTCh and TCh, with subsequent determination of enzyme activity.

Raman spectra are a large amount of data, which includes information about the composition of the substance, as well as various noises. To assess the contribution of each peak in the spectrum and make quantitative measurements of the spectrum is a very difficult task due to the sheer amount of information obtained from the spectra. For statistically significant allocation of contributions and for accurate quantitative analysis we used multiparametric statistical methods. In our work we used the principal component analysis (PCA) method for the separation of spectra of substrate and product of the enzymatic reaction, and the method of PLS to construct a calibration curve and further define the unknown samples. Ellman's assay was used as reference method to validate enzymatic activity values, measured by Raman spectroscopy.

# T15. Method development and validation for the quantitative application of mid-range infrared spectroscopy in GLP environment

*János Elek [a,b], Angéla Simon [b], János Török [b], Angéla Csontos [b], Csaba Ballai [b]*
*[a] Science Port Kft, Debrecen, Hungary.*
*[b] Fumoprep Kft, Budapest, Hungary*

The main analytical role of the mid-range infrared spectroscopy was the structural investigation of organic compounds for decades until the cheap, versatile and easily interpretable mass spectroscopic methods slowly occupied the area. Nowadays new perspectives are open for the MIR users due to the multiplied computing capacity of the personal computers. It does not only mean stronger hardware but the development of chemometric data evaluation methods and software. Nevertheless, thanks to the new sample introduction techniques that provide good repeatability with minor or no sample pretreatment. The significantly quicker method development, the shorter analysis time, the smaller cost of chemicals and labor-intensive work phases compared to the conventional chromatographic methods are the other reasons of this "renaissance".

The laboratories operating in GLP/GMP environment must fulfill rigorous requirements. The analytical performance must be demonstrated for any methods applied in quantitative determinations. Depending on the application field different validation procedures exist (such as FDA or ICH[1] guidelines) but they are almost exclusively set up for univariate — mainly chromatographic — methods. As probably the only exemption, there exist methodology for the validation of near infrared systems[2], but due to the different nature of NIR and MIR not all the parameters correspond to each other.

In our presentation the development and the validation of a Fourier-transform mid-range infrared method will be demonstrated with special focus on the *assay specificity* parameter.

## References

1. *ICH Quality Guidelines* Q2A, Text on Validation of Analytical Procedures; Q2B, Validation of Analytical Procedures: Methodology (http://www.ifpma.org/ich5.html) (1996).

2. Handbook of Vibrational Spectroscopy, John Wiley & Sons Ltd, Chichester, 2002

## T16. Determination of fat and protein content in milk using optical spectroscopy in the range 400–1100 nm

*Anastasiia Melenteva[1], Andrey Bogomolov[1,2,3]*
*[1]Samara State Technical University, Samara, Russia*
*[2]art photonics GmbH, Berlin, Germany*
*[3]Global Modelling, Aalen, Germany*

The rapid growth of the food industry requires new solutions in the quality control of processes and products. The development of new methods enabling quick and accurate determination of the analyzed components is an important task especially for mass consumption products such as milk. Due to the presence of fat globules and protein micelles milk has pronounced light-scattering properties. Economically attractive visible and short-wave near infrared (Vis/SW-NIR) spectroscopy can be successfully used for the quantitative analysis of milk. In the present work a new scatter-based approach to the determination of fat and protein content in milk has been developed. The difference in scattering spectral patterns by fat and protein particles can be utilized for their accurate quantitative determination using multivariate modelling methods, i.e. partial least squares (PLS) regression analysis.

In the first step the dependence of diffuse transmittance spectra in the Vis/SW-NIR region has been systematically investigated using gradual homogenization of raw milk samples. The homogenization results in significant spectral changes, which were explained in terms of the representative layer theory and other scattering theories. The fundamental feasibility of scatter-based quantitative analysis of milk fat and protein content using their low-selectivity diffuse transmission spectra has been proved in a series of designed experiments. This new analytical technique is resistant to an essential variability of fat globule sizes that may occur in raw milk samples. The applicability of the approach to the homogenized milk analysis has been tested on a large set of different samples, which are available in trade networks in Russia and Germany. Finally, the global (i.e. resistant to seasonal, genetic, regional and other milk variations) models for determination of fat and total

protein content in raw milk based on historical spectroscopic data collected during a year has been developed. In was also shown that full-range spectroscopy can be replaced by a set of light-emitting diodes (LEDs) as a light source and conventional digital camera as a detector. This simplified technology has shown an acceptable determination accuracy of fat and protein content in raw natural milk.

The presented results have significant practical value and the developed approach can be put into the basis of different devices such as portable inexpensive analyzers or analytical systems for in-line monitoring.

## T17. Monitoring chemical and sensory parameters of tomato product with near infrared spectroscopy

*D. Sun[a], J. Cruz[b], J. Casals[d], J. Sabate[c], J. Simó[c], Manel Alcal [a],*
*[a]Departament de Química, Facultat de Ciències, Universitat Autònoma de Barcelona, Spain*
*[b] Escola Universitària Salesiana de Sarrià, Barcelona, Spain*
*[c]Escola Superior d'Agricultura de Barcelona, Universitat Politècnica de Catalunya, Spain*
*[d]Fundació Miquel Agustí, Sabadell (Barcelona), Spain*

Tomato is one of the most important crops in the world and its quality will affect the heath of people all round the world. The near infrared spectroscopy (NIRS) was adopted to monitor the Critical Quality Attributes (CQAs) of tomato, such as the soluble solids, dry matter content, concentration of glucose and fructose, sweetness, taste intensity, aroma intensity, mealiness, acidity, crunchiness, juiciness and explosiveness. Quantitative models were constructed by partial least squares regression (PLSR). The validation models showed that the NIRS had a good capability for the quality control of tomato product both in chemical and sensory parameters. For the chemical parameters, root mean square error of prediction (RMSEP) was lower than 0.5. For sensory parameters, RMSEP was from 0.2063 to 1.0445 which was low enough for the determination of sensory parameters in factory. The NIRS provides us a simple and rapid tool to realize the application of process analysis technology (PAT) in tomato manufacturing.

# T18. Multi-criteria stochastic optimization of complex chemical production

*A.P. Kotenko, D.A. Pshenina*

*Samara State Technical University, Samara, Russia*

The production of road bitumen is an example of the problem of multi-criteria optimization of petrochemical production. Conflicting demands of modern standards for these products make it difficult to control the parameters of complex technological processes. In addition, the variable composition of raw materials coming from different fields of oil, does not allow to fine-tune the production process by the methods of feedback control theory. Under these conditions, it gives good results of the forecasting characteristics of the product by means of multivariate regression analysis of the systems of equations describing some standardized parameters [1, 2]. While deterministic formulation of the problem is unsolved because of the opposing claims of some objective functions, acceptable standards of performance ranges of properties of product allow us to find the confidence intervals of the control technological parameters.

The proposed mathematical model is implemented by the set of programs for data analysis of large experimental samples with a variety of analytical dependence of control covariates and objective functions. It was investigated the influence of partial managing regressors on the spread of monitored parameters derived road bitumen. Among the technological parameters (temperature, warm-up time) and the composition of the incoming raw materials, marked significant and non-significant regressors.

**References**

1. Kotenko A.P., Bukarenko M.B. The geometry of linear regression equations // Bulletin of Samara Scientific Center of the Russian Academy of Sciences, t.15, №6 (3), 2013 - P.820-823. (In Russian.)

2. Kotenko A.P., Sharapova E.A. Bengina P.M. Multi-criteria optimization using systems of linear regression equations // International scientific-technical conference "Advanced information technologies", 28-30 April 2015 - Samara: Samara Publishing Science Centre, 2015 - P.276-277. (In Russian.)

# T19. X-ray fluorescence analysis of lanthanide mixtures using partial least squares regression

_V.V. Panchuk[1,2]_, _D.O. Kirsanov[1,3]_, _A.A. Goydenko[1]_, _M.M. Khaydukova[1,3]_, _S.M. Irkaev[2]_, _A.V. Legin[1,3]_, _V.G. Semenov[1,2]_

[1]_Institute of Chemistry, St. Petersburg State University, St. Petersburg, Russia_
[2]_Institute for Analytical Instrumentation RAS, St. Petersburg, Russia_
[3]_Laboratory of artificial sensory systems, ITMO University, St. Petersburg, Russia_

X-ray fluorescence (XRF) is a very convenient tool of express multicomponent analysis and it does not require complex sample pretreatment procedures. Nevertheless there are certain issues hindering wide acceptance of the method in analytical practice. These issues are considerable sample matrix influence and rather high detection limits. During simultaneous analysis of metals in solutions these matrix effects are mainly twofold: 1) absorption of fluorescence radiation from analyzed element by other chemical elements of the matrix; 2) overlapping of characteristic lines from analyzed elements. While matrix effects can be easily taken into account in aqueous solution by choosing an appropriate standard samples, the sensitivity is still insufficient for certain analytical tasks. A possible way to circumvent these issues can be the application of appropriate multivariate data analysis techniques.

There are certain analytical tasks requiring simultaneous quantification of several lanthanides in complex mixtures, e.g. in spent nuclear fuel reprocessing where lanthanides are fission products that must be removed from the media; during processing of monazite — an important ore for rare earth elements and during technological monitoring of various industrial solutions.

To address this task we employed two different X-ray fluorescence techniques: energy-dispersive (EDX) and total reflection (TXRF). Comparison of this techniques (EDX and TXRF) in simultaneous quantification of six neighboring lanthanides in complex water solutions was performed with two regression methods (OLS and PLS). Due to low signal/noise ratio and overlapping signals in EDX measurements OLS fails to provide for reasonable precision in lanthanide determination especially in low concentrations. PLS processing of these data allowed for significant improvement of accuracy. Moreover, simple XRF method such as EDX can provide for the same precision as a

more sophisticated TXRF, when the measurement's results are fitted by PLS regression. Substantial and reliable lowering of detectable concentration of lanthanides in complex mixtures can be observed. The proposed approach can be applied in similar situations with XRF measurements in other analytes and it significantly broadens possible application range of the EDX method.

# P01. Multivariate regression techniques and NMR spectroscopy to quantitatively determine falsification of pharmaceutical and food products

*Yulia B. Monakhova*[1,2], *Bernd W.K. Diehl* [2]

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

[2] *Spectral Service AG, Köln, Germany*

There is no doubt that multivariate analysis is required to authenticate complex pharmaceutical samples and food products and, especially, to quantitatively predict the amount of adulteration in blends. In this report we consider two projects aimed at authentication of porcine heparin and sunflower lecithin by NMR spectroscopy. For multivariate modelling several statistical methods such as partial least squares regression (PLS), ridge regression (RR), stepwise regression with variable selection (SR), stepwise principal component regression (SPCR) were utilized for modeling NMR data of in-house prepared blends (n>80 in each case).

In case of porcine heparin blended with bovine species PLS and RR showed the best performance with the limit of detection (LOD) and root mean square error of prediction (RMSEP) below 2% w/w and 1% w/w, respectively. Reproducibility expressed in coefficients of variation was estimated to be below 10%. The developed method was found to be applicable also to heparinoid matrix (not purified heparin).

The models based on phospholipid ($^{31}$P NMR), fatty acid ($^1$H NMR) and saccharide ($^1$H NMR) distributions were validated to distinguish pure sunflower lecithin from that blended with soy species. PLS based on saccharide composition is able to estimate lecithin falsification regarding its vegetable origin with the sensitivity and RMSEP below 1% w/w and 3.5% w/w, respectively. Prediction error was improved by fusing three available NMR profiles.

The developed approaches allowed to evaluate the composition of lecithin and heparin blends on the basis of multiple chemical components in parallel with official NMR tests.

# P02. Diagnostics of sintering processes on the basis of PCA and two-level neural network model

*E.G.Egorova*, *I.V.Rudakova, L.A.Rusinov*
*St.Petersburg State Institute of Technology (Technical University)*

The manufacturing process of iron ore sinter flows under the influence of the set of disturbances, including changes in chemical and mineralogical composition and particle size of the charge, the conditions of its dosing and placing on the sintering belt, variations in the characteristics of the fuel, ignition and sintering conditions, etc. Control systems stabilize the main process parameters, without considering the frequently occurring deviations from the nominal values of these factors. This leads to wastage of sinter quality, and thus reduces the effectiveness of the process.

Survey of experts and analysis of operating regulations has allowed identifying a number of abnormal situations, each of which is caused by various faults. The task of the diagnostic system is to identify these faults in on-line mode and giving out the recommendations to operators for returning the process in normal state. The system operates in parallel with the traditional control system.

From the various types of diagnostic models applied for solving the problems of diagnosing complex systems, to which the sintering process also relates, neural network model was selected. The forming of abnormal situations, each of that brings together a group of faults with similar symptoms, led to the formation of the hierarchy in the structure of the diagnostic model. An important advantage of such a hierarchical neural network model is a significant reduction in dimensions of networks forming the model, which means a sharp reduction in network learning time and, more importantly, the time of relearning, that is sometimes necessary to adapt the model to the changing conditions of the process.

The top-level network performs the task of fault localization. Fault identification is made by low-level networks. To further reducing of the dimension of the top-level network the PCA is used. The training of the top-level network is carried out by the data of the matrix of PCA scores. The low-level networks, in principle, have small dimensions, so the initial variables can be used  there as

input network variables that facilitate to operators the perception of diagnostic results. To prevent a situation when the network loses already defined fault due to lack of results of the operator actions, i.e. when the situation continues to evolve, the changing of process parameters on the inputs of networks should be restricted.

The PCA allows using the classic monitoring of the process, by controlling the statistics T2 and Q, and starting the diagnostics after the detection of the fault occurrence, that increases system reliability.

Experimental study of the system by simulation on the real data showed efficiency of proposed method. The system confidently defines faults within 20% of their maximum development.

## P03. Combined analysis of X-ray diffraction patterns and IR spectra of LiFePO$_4$

*Anna Yarchuk[1], Evgeny Karpushkin[1], Andrey Bogomolov[2]*
*[1]Moscow State University, Moscow, Russia*
*[2]Art Photonics GmbH, Berlin, Germany*

Lithium iron phosphate (LiFePO$_4$) has been recognized as a promising cathode material for application in lithium-ion rechargeable batteries. This chemistry provides long cycle life, very constant discharge voltage, and the energy density comparable to that of the commonly used LiCoO$_2$-based cathodes. Additional advantages of LiFePO$_4$ over other cathode materials include lower cost and improved safety. On top of that, high-quality LiFePO$_4$ in the form of nanosized crystalline particles can be prepared via a simple hydrothermal synthesis method.

Importantly, the electrochemical properties of LiFePO$_4$ are highly sensitive to its particles size and shape and the presence of impurities, especially at the crystals surface. These material characteristics depend on the preparation procedure. Luckily, the hydrothermal synthesis is fairly flexible and allows for preparation of LiFePO$_4$ particles of different micromorphology.

X-ray diffraction analysis has been accepted as a straightforward tool to evaluate the quality of LiFePO$_4$ product. However, our trials have indicated that lithium iron phosphate specimens prepared via significantly different proto-

cols exhibit similar XRD patterns. Therefore, we have applied IR spectroscopy as a supplementary technique for the product study. The detailed analysis of the IR spectra (including the bands assignment) is fairly complicated; however, we have demonstrated that the vibrational spectra are very indicative of the material purity and (likely) morphology. In this contribution we present the first attempt of combined analysis of XRD patterns and IR spectra of various forms of LiFePO$_4$ assisted by electron microscopy images analysis in order to develop an approach towards fast and reliable characterization of LiFePO$_4$ and similar materials. In particular, we have performed cluster and similarity analysis of the IR spectra and X-ray diffraction patterns and demonstrated that IR spectroscopy is advantageous over X-ray diffraction method to evaluate the LiFePO$_4$ product quality.

## P04. Optimization of information support to enhance diagnosability of technological processes

*S.A.Kosarevskaya, V.V.Kurkina, L.A.Rusinov*
*Saint-Petersburg State Institute of Technology (Technical University), Saint Petersburg, Russia*

A large number of sensors are generally aspired to use to ensure the effective monitoring of the process, since it considered that the more the sensors, the better quality of monitoring is achieved, and the faults will be detected more confidently. However, economic and technical possibilities impose significant limitations on the number of sensors required for monitoring and diagnosis of the process. First of all, it is the cost of purchasing and installing the sensors, in addition the technical limitations can be, for example, due to the impossibility of direct measurement of a variable, etc.

It is important not only to minimize the number of sensors, but also effectively place them on the process for increasing the observability of process states, detection and identification of its possible faults. Different approaches are used for determining the optimum number of sensors and their location (creating a network of sensors) for ensuring early detection and correct identification of all potential process faults. The most popular approach uses directed graphs. The graph nodes represent the variables of the process, and the edges show a causal effect between nodes.

Analysis of the graph starts from the node, where it is assumed the appearance of a fault (the root node), and then the search for other nodes reachable from the root one is carried out. If on this way on the graph at least one node with the sensor will be found, it can be assumed that this fault will be detected. So the sensor network will be formed, but it will not be optimal.

Some sensors of this network are obligatory and cannot be removed, for example sensors that are used for control the process parameters; only the position and the number of sensors that will be use for improving the observability of the process and increasing the depth of diagnostics can be optimized.

In this case the classic genetic algorithm is preferable among the stochastic algorithms for the solution of multiextremal optimization tasks. The solution (chromosome) here is a network of sensors; the population is a set of chromosomes. Goodness of solutions in the population is defined upon the fitness function that is calculated as the inverse value of the global penalty function. Penalties are assigned in proportion to the cost of the installed sensor and losses owing to not detecting the fault. After assessing fitness of all the chromosomes in the population, a new generation is created from this population by using the selection, crossover, and mutation operators. The stop of the algorithm is defined by the absence of improvements in several successive populations. The obtained solution comprises a minimum sufficient sensor network, the use of which allows obtaining the effect of early-detection of faults without costly measuring.

## P05. Voltammetric identification of antiarrhytmic drugs using principal component analysis

*R. A. Zilberg, Yu. A. Yarkaeva, A. I. Khabletdinova, A. V. Sidel'nikov, V. N. Maistrenko.*
*Bashkir State University, Ufa, Russia*

Voltammetric methods are widely used for the quality control of pharmaceuticals, because many medicinal substances are electrochemically active in the available potential region. In this work, possibilities of voltammetry are considered as applied to problems of the identification of antiarrhythmic drugs from the group of $\beta$-adrenoblockers (atenolol, propranolol, sotalol, metopro-

lol, amiodarone) from different producers using an electronic tongue based on glassy carbon electrodes modified with polyarylene phthalides (PAFs).

To assess the similarities and distinctions between the cyclic voltammograms, we used their PCA modeling. In most cases, in the score plots of PCA models on the coordinates PC1–PC2, the voltammograms of medicines from different producers belonged to different clusters, depending on the nature and concentrations of auxiliary substances. The optimum number of principal components was selected so that explained dispersion was not lower than 90% (2–5 principal components). The stronger was the difference between the compositions of auxiliary substances in some medicines, the longer was the distance between the corresponding clusters in the plane of principal components. The score plots of PCA modeling were used to gain information about the compliance of medicines to the standard samples and also for revealing the distinctions between the products from various manufacturers. In addition, the voltammograms of medicines transformed by principal component analysis can serve as "virtual prints" and be used for their identification and quality control.

It should be noted that the identification of medicines differing in the composition of auxiliary substances using three-factor images (registration of voltammograms on three glassy carbon electrodes modified by PAFs) allowed us to considerably increase the percentage of correctly recognized samples compared to the registration of voltammograms on only one electrode. In the latter case, clusters of medicines from various producers overlapped even in the case when medicines contained different auxiliary substances.

# P06. Mathematical modeling of "composition-current" phase diagrams for three-component heterogeneous alloys of cadmium-tin-bismuth

*Moshchenskaya E.Yu.*
*Samara State Technical University, Samara, Russia*

The phase diagram is one of the main tools in solving most of the problems of modern physical and chemical analysis in materials science, metallurgy and chemical engineering. Currently, methods of experiment planning are widely used in physical and chemical analysis particularly for charting "structure-property". Experimental design by thermodynamic modeling of invariant points on the chart "structure-property" greatly simplifies the experimental study of multicomponent systems.

The present work is dedicated to the development of methods and algorithms for the simulation of phase complex of physical and chemical systems as well as to automation of the research and analysis process. The method of local electrochemical analysis (LEA) uses "composition-current" phase diagrams as a source of information. Dependencies of the partial currents phase dissolution on the alloy composition are used as calibration curves.

I have offered equations for modeling "composition-current" diagrams for different alloy systems. An algorithm, which calculates the partial currents within the composition of the alloy components, has been developed. This algorithm was implemented in a visual programming environment Delphi.

The equations for computing the current partial dissolution for cadmium, tin and bismuth are obtained for the dissolution process from matrix metal ternary alloy Cd-Sn-Bi. Points for each of the three surfaces were calculated using a program written in a visual programming environment Delphi.

Sets of coordinates (x, y, z) in the form of an array of numerical data for each of the three surfaces were obtained. The program allows to calculate the coordinates with steps specified by a user. The resulting sets of coordinates were visualized using software Algebrus 3.1.

It has been established that the intersection point of three surfaces on the "composition-current" diagram in a ternary system Cd-Sn-Bi corresponds to a relative minimum in the total dissolution of the components surface currents.

46

## P07. Recognition of physiological states using stabilometric data: a feasibility study

*Andrey A. Muraenko, <u>Sergei I. Zhilin</u>*
*Altai State University, Barnaul, Russia*

The principle of stabilometry consists in translating the mechanical oscillations of <u>man</u>'s physiological gravicentre into electrical signals by means of force plates, signals amplification, recording, and analysis [1]. Traditionally, the stabilometry was considered as a technique for measuring only the balance of human beings under certain conditions and diagnosing balance-related disorders. As the stabilometry is simple and rapid to use, stabilometric data is potentially suitable for research and applications not only in clinical medicine but in broader context as well. The presented work is aimed at the discovering features of stabilometric time series and studying their informativity and usefulness for physiological states detection and physical activity for the development of systems for monitoring and control of physical activity in sport and rehabilitation.

The first contribution is the way to construct features of stabilometric signals suitable for distinguishing of physiological states. The idea is to collect and aggregate scalar and histogram statistics of signals for running time-window: amplitude, standard deviation, etc). The resulting feature vector is obtained by concatenation of features for several window widths. Discrimination of different object (physiological states) in principal components space was studied for the following groups of persons: 1) patients of neurological clinical unit after hypertension stroke survived in 2-5 days before stabilometric measurements; 2) adult patients with infantile cerebral paralysis which are able to stay at force plate without assistance; 3) adults in good health (without affirmed health problems). The results of the study showed principal feasibility of the proposed features of stabilometric time series for pairwise distinguishing of mentioned groups of person.

The second contribution concerns the feasibility of stabilometric data statistics for estimation of sportsman organism response to exercise stress. Stabilometric measurements were collected during power training of sportsmen with different level of preparation before and after each exercise. It is concluded that the mean density of stabilometric trajectory is correlated to the change of physical activity and can be substantially interpreted by experts.

**References**

Chiari L. Stabilometry. In: Binder M., Hirokawa N. and Windhorst U. Encyclopedia of Neuroscience. Berlin Heidelberg: Springer, 2009. pp. 3830-3833.

## P08. Identification of bisoprolol from different manufacturers using voltammetric "electronic tongue"

_Yu. A. Yarkaeva_, R. A. Zilberg, E. I. Maxyutova, A. V. Sidelnikov, and V.N. Maistrenko

*Bashkir State University, Ufa, Russia*

Due to the increased requirements for the quality of medicines, the importance of the use of modern standardized methods for their analysis grows. Along with the use of chromatographic, optical methods of determining organic compounds electrochemical analysis methods are commonly used, in particular voltammetry.

In recent years, the formation and practical application of chemically modified electrodes became interesting. Sensors based on these electrodes are used for determination of a large number of different substances, particularly medicinal compounds. For modifying different chemical compounds, polymer films are applied to the surface of the electrode material. They significantly change the ability of the electrode to the voltammetric response. Such modified electrodes are often used in multisensory systems as "electronic tongue".

New three-electrode multisensory system are proposed for voltammetric analysis of antiarrhythmic drugs using three polyarylenephthalides as modifiers (TO, TOO, TOOO). The use of three different modifiers provides cross-sensitivity condition of electrodes required for the functioning of the voltammetric systems as the "electronic tongue".

In this work original drug of bisoprolol Konkor and its generics (Aritel, Biprol, Niperten, Bisogamma, Bisoprolol, Bidop, Biol, Konkor Kor, Kardinorm) were investigated by voltammetric method. Processing of the obtained voltammograms of antiarrhythmic drugs was carried out using chemometric methods: principal component analysis (PCA) and method of soft independent modeling of class analogy (SIMCA).

The use of polyarylenephthalides as modifiers provides a clear separation of the response signals of antiarrhythmic drugs on the plot of PCA-modeling. SIMCA-classification showed that the proportion of correctly identified antiarrhythmic drugs varies between 67-100%. Thus, express-identification of medicines using a voltammetric "electronic tongue" can be carried out with high accuracy.

## P09. Software implementation of the Data-Driven SIMCA method

*Y.V. Zontov[a], O.Ye. Rodionova[b], A.L. Pomerantsev[b]*
*[a]National Research University Higher School of Economics (HSE), Moscow, Russia*
*[b]Semenov Institute of Chemical Physics RAS, Moscow, Russia*

We present a software implementation of the Data-Driven SIMCA method (DD-SIMCA) [1, 2]. The method is implemented in Matlab scripting language in the DD-SIMCA Toolbox, which provides the means for data pre-processing, interpretation, validation and visualization. The two main modules of the toolbox are named in accordance with their basic functionality, DD_SIMCA and DDS_PLOT respectively. The first one is used for the training and testing. The second one is employed for graphical representation and analysis. Both modules rely on the fixed structured representation of the classification model as well as the validation results. This approach simplifies the communication between the software modules and the reusability of the model.

The main features of the software are as follows:

- The DD-SIMCA method is implemented using the best Matlab programming practices. The command-line interface provides numerous options, which are well documented and exemplified.

- The Acceptance plot provides a graphical representation of the decision area, normal samples, extremes and outliers in a user-friendly manner. The Extreme plot shows the observed vs. the expected number of extreme objects and provides graphical means for the data analysis.

- Calculation of the rate of the false accepted samples (type II error β) for the new classes.

- The toolbox may be use without the Matlab Statistics Toolbox.

The software functionality is illustrated with a real-world example of quantitative risk assessment in classification of drugs with identical API content [3].

**References**

1. A.L. Pomerantsev, O.Ye. Rodionova, J. Chemometrics, 28, 429–438 (2014)

2. A.L. Pomerantsev, O.Ye. Rodionova, J. Chemometrics, 28, 518-522 (2014)

3. O.Ye. Rodionova, K.S. Balyklova, A.V. Titova, A.L. Pomerantsev, J. Pharm. Biomed. Anal. 98, 186-192 (2014)

## P10. The use of PLS-DA and pulsed electric stimulation method for assessing the quality of biological objects

*Aleksey G. Zarubin[a,b], Yuliya N. Dolgova[b,c], Anatoliy I. Mamaev[b,c]*
*[a]Department of Oil and Gas Storage and Transportation, Tomsk Polytechnic University, Russia*
*[b]Department of Analytical Chemistry, Tomsk State University, Russia*
*[c]Sibspark Limited Liability Company, Tomsk, Russia*

Food products have to conform to modern requirements for quality and safety of food stuffs. In this regard the development of identification methods for fresh biological production and biological production after freezing and subsequent thawing represents academic interest.

As biological production the samples of a fresh crucian and an unfrozen crucian were selected (the frozen meat of the crucian after freezing was defrosted in the thickness of muscular tissue up to the temperature not lower than minus 1 °C). To receive an analytical signal the method of pulse electric impact on biological fabric with measurement of electric current force values depending on the time period of electric impact was chosen [1]. The PLS-DA method was chosen for a formal classification. Selection of experimental data was expanded with the method of numerical statistical modeling (the Monte-Carlo method) on the basis of aprioristic statistical information about the analytical signal. On the basis of the obtained model data with the use of

CHEMOMETRICS library [2] the PLS-DA model for two classes of biological production – the fresh crucian and the crucian after freezing and thawing – was constructed and checked. It was established that the PLS-DA model for two hidden variables was characterized by the lack of errors of the first and second kind for prediction of classes for the fresh and unfrozen crucian.

As a result of the conducted research it was established that the method of pulse electric impact on biological fabric with measurement of electric current force values depending on the time period of electric impact in combination with the PLS-DA method allows carrying out unambiguous identification of the fresh and unfrozen crucian.

**References**

1. Dolgova, Yu. N., A. K. Chubenko, and A. I. Mamaev. Pulsed electric stimulation as a method for assessing the quality of biological objects. Russian Journal of General Chemistry 84.5 (2014): 1061-1064.

2. Pomerantsev A. L., Chemometrics in Excel, John Wiley & Sons, 2014.

## P11. Weight selection and rotation ambiguity in multivariate curve resolution of absorption spectra with weighted least squares

*Skvortsov A.N.*

*Saint Petersburg Polytechnic University, St. Petersburg, Russia*

Multivariate curve resolution (MCR) is set of chemometric techniques, which aim to resolve data matrices into spectra and concentrations of individual chemical components, given a set of physicochemical constraints. Given a data matrix X an MCR method searches for a best bilinear model $CS^T$, where C and S are constrained matrices of spectra and concentrations. In most MCR methods this is done by minimizing the sum of squares $\sum_{mn}e^2_{mn}$ of the elements of the residual matrix $E=X-CS^T$. Such approach implies that the residuals are distributed normally, similarly, and independently. This is not always the case in real data, where the uncertainties may differ from point to point and/or outliers may occur. One of the classic solutions is to use weighted least squares (WLS) approach, which minimizes $\sum_{mn}e^2_{mn}w_{mn}$, where $\{w_{mn}\}$ is a weight matrix. Typically, the weight matrix is constructed from the maximum likelihood princi-

ple. However, WLS is a very flexible technique, it can be adapted for MCR of sparce matrices, robust MCR (by iterative reweighting procedure), and some nonlinear MCR tasks.

A complicacy of MCR with WLS is the inapplicability of principal components for finding MCR factors, so many good MCR methods cannot be used with WLS. The problem can be solved by weighted alternative least squares (termed MCR-WALS). Also a faster positive matrix factorization (PMF) algorithm was proposed (*Paatero, Tapper, 1994*), which is currently a widely used WLS method. The possibilities of MCR-WALS, PMF and some other methods were recently revisited in (*Stanimirova, Tauler, Walczak; 2011*). MCR-WLS solutions, similarly to common MCR, may have residual rotation ambiguity (RA). While RA in common MCR was extensively studied, RA in MCR-WLS and its relation to weight matrix is much less known.

In the present work we focused on MCR of 'saturated' absorption spectra (optical density). Optical density $D$, which is bilinear with rellllspect to spectra and concentration, is a log-ratio measurement. Together with the nature of uncertainties in light intensity measurements this results in strong dependence of measurement uncertainty $\delta D$ on $D$ with typical sharp increase of $\delta D$ at large $D$. So large $D$ values are unstable, effectively nonlinear (hence the term 'saturated spectra') and compromise the MCR model. Common MCR is unable to handle this situation, but MCR-WLS may succeed if proper weights are selected, which would damp or even zero out the contribution of large $D$ values. Still, the analysis of residual RA of the solution is required.

To test the relative efficiency of different weighting schemes and estimate RA in MCR- WLS several simulated examples and a real-life data matrix were analyzed. MCR-WLS factors and analogs of principal components (for RA analysis) were calculated by MCR-WALS and weighted version of NIPALS respectively. RA was estimated by Borgen plots, tracking, and charged particle optimization cPSO. Simple threshold weighting ($w=0 \mid D > D_{max}$) or accurate weighting based on uncertainty function and maximum likelihood principle were used.

Both weighting schemes displayed better performance in MCR of saturated spectra than unweighted MCR (regarding the determination of the number of

components, retrieving spectra and concentrations, and satisfying the constraints). Most of the existing methods of RA estimation can be generalized to MCR-WLS. RA in simple MCR and MCR-WLS for slightly saturated spectra showed similar behavior, although the shapes of the feasible areas were different. Strong saturation in the spectra increased rotation ambiguity; it corresponded well to the decrease in the relative count of 'reliable' and 'extreme' data points. It was also shown that accurate weighting of measurements, which are close to zero, can expand the power of Borgen plot RA analysis (as an alternative to generalized Borgen plot algorithm).

## P12. The method of hyperspectral image classification for identifying oil pollution on the ground

*M. Kuznetsov[1], E. Ramenskaya[2], V. Ermakov[2]*
*[1]Moscow Institute of Physics and Technology, Moscow, Russia*
*[2]Samara State Technical University, Samara, Russia*

The problem of environmental pollution with oil and oil products is not new. Questions of rapid detection and liquidation of oil spillages remain relevant for many years. It is proposed that the approach of using remote sensing, in particular, the analysis of hyperspectral image of the Earth's surface will allow to solve the problem of environmental monitoring of oil pollution. Some difficulties arise in the analysis of hyperspectral images. We address the problem of the oil spill identification on the land surface using the hyperspectral images processing methodology.

The considered data are the hyperspectral images of the land surface obtained from the airborne lab, with the expert-given allocation of the regions of interest. That is, the experts allocate the oil spill areas together with the areas that don't contain any oil spills with certainty. The second-type allocations include the dark regions similar to the spills: infrastructure shadows, small water ponds, cloud shadows etc.

Using the given data, we construct a two-class classification model of the image spectra. The first class contains the allocated oil spill spectra, the second class consists of spectra of the remaining objects.

The developed classification method consists of the two stages. On the first stage we estimate a Gaussian mixture model describing the given spectral data. Using the mixture we perform data clustering and compute the cluster centers. On the second stage we construct the classification algorithm using the distances to the cluster centers as the new feature description of the spectra.

The proposed method allows to achieve high classification accuracy through the use of the cluster data assumption. The constructed model is simple and interpretable: the oil spills correspond to the spectra with the similar shape to the cluster centers of the allocated spills. Furthermore, the training stage is computationally efficient: it is linear over the total objects (the allocated hyperpixels) number. This computational requirement is essential for the hyperspectral data processing.

Our current and further investigations are devoted to the collecting and processing the satellite hyperspectral images. Upon achieving acceptable recognition quality, we will provide recommendations on modifications of the ecological monitoring systems in oil and gas companies.

## P13. Raman spectrometry for pharmaceutical water solutions analysis

*Maxim Khannanov[1,2], Oleg Volkov[2], Roman Galeev[3], Leonid Kulik[1,2], Nail Khannanov[2], Ramil Ibragimov[2], Sergey Dorofeenko[2].*
*[1]Institute of Solid State Physics Russian Academy of Sciences, Chernogolovka, Russia*
*[2] EnSpectr LCC, Chernogolovka, Russia*
*[3] Federal service of health care, Moscow, Russia*

Certification, counterfeit revealing, raw material inspection and so on are tasks related to drug solutions identification. The problem in this case is to perform it routinely and during short time as defined by the controlling authorities.

Chromato-mass-spectrometry does not allow practically to perform the identification routinely for the complexity of the method itself, lack of portable spectrometer modification, required sample preparation as well as the need of a highly qualified operator. Chromato-mass-spectrometry analysis requires

as well ampoules opening that is generally forbidden without the manufacture's permission.

Raman spectrometry does not have these drawbacks being a rapidly expanding non-destructive method for solid and liquid samples analysis. Raman spectroscopy allows measurements of liquid and powder substances through a transparent or semi-transparent package without its destruction, the identification of pure substances and mixes in a matter of seconds. The main problem though about quantitative Raman analysis of water solutions with portable setup is the detection of water itself.

We report on experimental results of Raman spectra analysis of pharmaceutical water solutions with the concentration of about one percent with unprecedented accuracy in an intact ampoule using a specially developed portable Raman spectrometer. The investigation of water Raman band depending on a pH-value, the luminescence and Raman scattering from a glass package, the analyte spectra chemical shift in water are performed in order to archive the mentioned characteristics.

In conclusion a new technique is given for the distinguishing between different medicines in water solutions and following calculation of the concentration of the main and additional components. This technique was successfully tested by Russian federal service of surveillance in healthcare and proved all the mentioned characteristics.

## P14. Application of Chemometrics in Geoecology

_K. Chertes_, O. Tupitcina, V. Ermakov
_Samara State Technical University, Samara, Russia_

Samara State Technical University has a great experience in solving geoecological problems. Geoecology is a fundamental and applied science that deals with the engineering objects and geologic environment cooperation. Both of the elements generate the destructed geological systems (DGS). For example, off-site landfills, borrow pits, contaminated and polluted lands, uncompleted construction projects and their agglomerations.

The important geoecological problem is the DGS recovery. Since it is impossible to recover some of the DGS, a great number of parameters should be esti-

mated beforehand. Though it's practically impossible to identify the degree of influence of the parameters.

DGS are the multivariate data systems. So the dominant component analysis (DCA) is the most useful instrument for their estimation. This method was used for the estimation of more than hundreds of DGS: open-casts, waste deposits, sludge tanks, sludge draining beds and other artificial geo-structures in the water industry, waste handling services, agriculture, industrial and civil construction.

Chemometric methods are constantly used as the mathematic and practical foundation of the PhD thesis of SSTU scientists, such as IR spectroscopy and PCA applied for polymer sorting in the domestic waste, application of scanning probe spectroscopy in the monitoring of oil pollution, ecological assessment of landfills with multivariate analysis, principles of managing ecologically safe architectural reconstruction, etc.

As the result of the conducted research we concluded that the main problems of multivariate data analysis usage in geoecology are the following:

1. The difficulty of obtaining raw data. The test objects are of big size, and it is necessary to carry a great number of laboratory work.

2. The inability to use ordinary analyzers. The studied objects are multicomponent and multiphase.

3. The studied systems are open. The objects are inextricably linked to all the components of the environment.

4. The dynamics of the study object does not depend on the researcher. The rate of change in the properties of air and water is high. The geological environment is changing slowly. The abrupt changes (rainfall, volley emission, etc.) are possible.

Notwithstanding the above-mentioned problems we should say that the study of complex geoecological systems without the use of multivariate data analysis is impossible.

# P15. Environmental pollution assessment using PCA/PLS analysis (through the example of soil and natural water of Karaidel' district of Bashkortostan republic)

_Kabirova L.R._[1], _Sidelnikov A.V._[2], _Lygin S.A._[1]
_[1]Birsk branch of Bashkir State University, Birsk, Russia_
_[2]Bashkir State University, Ufa, Russia_

Biotesting is a determination of toxicity with the help of testers. The testers give signals of danger regardless of what substances and in what combination cause changes in vital signs of the investigated organisms and, therefore, for revealing the main factors of contamination further investigation is necessary.

In the present work we analyzed the influence of soil composition and natural water on the indicators of environmental pollution (using water from river Karaidel' and a melt water). Principal Component Analysis (PCA) was applied to establish chemical factors of soil heavy metals pollution and determine the behavior of such test organisms as buttercup (lead pollution) and watercress (copper, zinc, cobalt pollution) growing at the Akbulyak-Baiki-Birsk highway.

Partial least squares (PLS) discriminant analysis was employed to investigate the influence of certain indicators (e.g., ions of heavy metals, COD, BOD and ammonium groups) on the level of pollution in the sampling areas: villages of Berdyash and Horoshaevo located near the river Karaidel'. K-criterion was used as a generalized criterion of the natural objects contamination level. It was equated to 1 for objects with maximum permissible concentration (MPC) contamination level, and to 0 for objects with contamination level lower than MPC. The PLS model was built using the K-criterion as a response variable. It was fount out that the concentration of harmful substances in the areas under study does not exceed the MPC (the predicted K values did not exceed 0.2).

## P16. Calibration transfer between different types of multisensor systems

_M.M. Khaydukova_ [1,2], D.O. Kirsanov [1,2], V.V. Panchuk [1], C. Medina-Plaza [3], M. L. Rodriguez-Mendez [3], A.V. Legin [1,2]

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

[2] ITMO University, St. Petersburg, Russia

[3]Sensor Group, Department of inorganic chemistry, E.T.S. Industrial Engineering, University of Valladolid, Valladolid, Spain

Multisensor systems, also known as "electronic tongues", are promising analytical tools for express analysis of various complex parameters in liquids, like e.g. taste, toxicity, etc. These systems are a combination of hardware and software parts. Hardware part consists of chemical sensor array and devices for sensor response registration. Two most commonly used sensor types are potentiometric and voltammetric. Software part is related to chemometric techniques. Typically multivariate regression modelling is employed to relate the response of sensor system with parameter of interest. A good model requires considerable efforts and large number of samples, which can be unique, expensive and hard to get. It would be very practical to have a method capable of multivariate calibration transfer between various multisensor systems working on different principles.

In our previous research a simple algorithm based on PCA decomposition has been suggested for calibration transfer between different analytical methods. It can convert the data from one instrumental format to another [1]. In this study we apply the proposed strategy to transfer calibration models between voltammetric and potentiometric multisensor systems. Sample set consisted of 8 types of grape musts. All the samples were measured at least three times with two sensor arrays: voltammetric of 4 sensors with metal oxide modified surface operated in cyclic voltammetry mode, and potentiometric of 26 plasticized polymeric membrane sensors. Cyclic voltammograms were compressed by kernel method, 10 kernels for each sensor were used. Resulting voltammetric (VA) data were 8×40 matrix, while potentiometric one was 8×26. The VA data were used to build PLS regression models to predict total acidity, potassium concentration, tartaric acid concentration and other parameters of

grape must. RMSEP values of these models for total acidity were around 0.08 (for the range 0.66-0.96 lg C(TA)). Potentiometric data were converted into voltammetric response shapes with suggested algorithm and were used further to predict the same parameters. Corresponding RMSEP for TA was 0.05 (this decrease can be attributed to the higher sensitivity of potentiometric sensors towards TA which is confirmed by RMSEP values of the original potentiometric model equal 0.02). Similar results were obtained for several other parameters.

These results show that data conversion between different sensor arrays based on different working principles is possible and this fact can be employed for building cross-platform multivariate regression models.

**References**

1. V.V. Panchuk, D.O. Kirsanov, M.M. Khaydukova, V.G. Semenov, A.V. Legin. Latent variable based calibration transfer between different analytical methods // Book of abstracts CAC-2015, XV Chemometrics in Analytical Chemistry, June 22-26, 2015 Changsha, China – p.25

## P17. Method of impedance spectroscopy for titration of surfactants using principal component analysis

*D. I. Dubrovski, A. V. Sidelnikov, F. Kh. Kudasheva, and V. N. Maistrenko*
*Bashkir State University, Ufa, Russia*

Recently, interest in controlling the content of surfactant in the oil increased. In this regard, there is the need for improved methods for determination of surfactants in complex organic mixtures - oil and water-oil emulsions. Among them spectrophotometry and potentiometry are used most widely. However, these methods require prior separation, purification, preparation of two-phase systems. This often complicates the analysis.

In this paper, electrochemical systems and chemometric approaches that extract information about the content of surfactants in aqueous solutions and water-oil emulsions using impedance spectroscopy were proposed. An important advantage of the method is the high sensitivity of the measurement,

the absence of requirements to the selectivity of the electrodes, to the presence of colored components and heterogeneous phases and others. The spectra of the real and imaginary components of impedance were recorded during the titration. The obtained data were processed using principal component analysis. Spectra were converted into points on a plane of principal components. It is shown that the inflection point corresponds to the end point of titration and characterizes the state of equivalence reaction.

## P18. Multivariate discrimination analysis of multispectral fiber sensor data for kidney cancer diagnostics

_Valeria Belikova[1]_, Urszula Zabarylo[2], Viacheslav Artyushenko[2], Franziska Schulte[2], Iskander Usenov[3], Vladislav Galyanin[1], Andrey Bogomolov[1,2,4]
[1] Samara State Technical University, Samara, Russia
[2] art photonics GmbH, Berlin, Germany
[3] TU Berlin, Institute for Optics & Atomic Physics, Berlin, Germany
[4] Global Modelling, Aalen, Germany

According to the statistics, cancer remains one of the most common causes of death throughout the world. Surgical intervention is still the main method of medical treatment of the disease. Operation success strongly depends on the doctor's ability to distinguish tumor from healthy tissues. Visual assessment of tissue sections and biopsy are the common methods of histopathological analysis. Its main disadvantages are the high costs and considerable time that is needed to obtain results. During a surgical treatment this puts the patient at greater risk due to prolonged exposure to general anesthesia.

Over the past two decades, the use of optical spectroscopy for biomedical applications has grown significantly. Its attractiveness comes from its ability to provide quantitative information about the biochemical and morphological states of the tissue in a minimally invasive or noninvasive manner. In accordance with the literature, different fiber-optic probes and spectroscopic techniques are used in tissue analysis of skin, breast, lungs, brain, gastric and some other types of tumor.

In this work we have investigated *ex-* and *in-vivo* samples of malignant and healthy kidney tissues of eleven patients using four spectroscopic techniques simultaneously: mid-Infrared, near infrared, fluorescence and Raman spectroscopy. The measurements were performed with various fiber-based probes transmitting the light in the respective spectral range. Here we report on the development of diagnostic multivariate algorithms capable of cancer of the human kidney using this multispectral data.

## P19. Simulation of spectral data for regression methods validation

*Vadim Belikov[1], Vladislav Galyanin[1], Anastasiia Melenteva[1], Andrey Bogomolov[1,2]*
*[1]Samara State Technical University, Samara, Russia*
*[2]art photonics GmbH, Berlin, Germany*

Chemometricians tend to use many technical tricks to obtain an optimal multivariate model: various regression or classification algorithms, data preprocessing techniques, variable/ sample selection etc. The development of an optimal data analysis algorithm is important for the future model performance, in particular, in the spectroscopic applications. In order to make it sure that the selected algorithms and data treatments are efficient, they should be validated using spectral data of a necessary nature. The algorithm validation should take all relevant spectral characteristics such as noise level and background, into account and be based on a sufficiently large and representative sample set. Often, these requirements can hardly be met using the experimental data only. In such cases, the algorithm can be verified on a simulated data that can be generated in a sufficient volume and with the specified parameters. During the algorithm verification it may be important to find out and to understand individual influences of different factors related to the data nature, such as peak overlap (selectivity), signal-to-noise ratio, the presence of a non-linearity and the baseline, on the final result of analysis.

To facilitate the task of algorithm validation we have developed an application that generates spectra of individual components and their mixtures (based on the Beer's law) with preliminary specified properties. The spectral data generator enables the adjustment of the following parameters: the number of

samples and variables, the number of peaks, their shapes and positions (and hence, the degree of their overlap), the nature and the level of noise, the background intensiveness and nature (multiple scattering, fluorescence, chromatographic baseline drift) etc.

Here, we present the spectral generator software and show several modelling examples using simulated data sets illustrating its advantages for the multivariate algorithm development and for the optimization of data treatment routines.

## P20. Comparison of univariate and multivariate calibration for lead determination in soils by laser-induced breakdown spectroscopy

*Ivan N. Krylov, Sergey M. Zaytsev, Timur A. Labutin, Andrey M. Popov and Nikita B. Zorov*

*Lomonosov Moscow State University, Department of Chemistry, Moscow, Russia,*

Laser-induced breakdown spectroscopy (LIBS) is a type of atomic emission spectroscopy which employs laser emission for sampling via ablation while the resulting laser-induced plasma acts as both excitation source and analytical signal source. Ecological monitoring of lead in soils (for example, near roads or metallurgical/glass factories) and geochemical prospecting require fast elemental analysis of solids suitable for being conducted in the field. LIBS is a promising technique for such purposes.

However, spectral noises caused by interfering elements worsen accuracy of optical atomic emission spectroscopy techniques. Since multivariate calibration allows improving accuracy of analysis when signals overlap, we checked whether the reduction of uncertainty was possible for low lead concentrations (32 ppm is MPC according to Russian norms / 16 ppm is the mean content in Earth crust) in soils of different types (black, red, gray soil and others).

A set of 15 samples was involved into construction of calibration models. A spectral range containing lead line at 405.78 nm was selected. Leave-one-out cross-validation was employed to validate the model. Different baseline correction methods were tried; subtraction of minimal value within spectral range for each spectrum provided the best accuracy for PCR. A comparison of

PCR results with univariate regression model results was performed. The use of PCR has lowered the RMSEC, thus the determination of lead in soils down to 15 ppm became available. The uncertainties of predicted lead content at the required concentration range were in agreement with the confidence band given for certified values.

## P21. Geoecological evaluation of the development of housing and communal services sludge storages by handling multidimensional data

*V.N. Pystin*, *N.A. Safonova, V.I. Zuber*
*Samara State Technical University, Samara, Russia*

Activities of housing and communal services is accompanied by the formation of low-hazard waste. Directions of processing, recovery and recycling of waste of housing and communal services are well-known. They are mainly associated with the production of building materials and secondary reagents for wastewater treatment. However, high energy costs and material costs hinder their implementation. As a result, the waste of housing and communal services are deposited in a storage.

The basic characteristics of housing and communal services sludge storages were considered from the perspective of their target development. The system of condition appraisal of sludge storages was represented by the multidimensional data analysis. The waste disposal facilities that were considered in this study were classified into three groups from the perspective of the feasibility and the practicability of their abandoning. The ex-ample of the digital matrix of the waste disposal facilities condition was given. The constructive and technological design of the system of recultivating materials production was given.

Using composites on the basis of waste of housing and communal services was applied in the preparation of areas disturbed by object dismantling of inactive buildings and enterprises installations in the Samara region. The eco-

nomic effect of the developed technologies is achieved by reducing the required volumes of natural soil and return of the recovered territories in economic use.

## P22. Attenuated total reflection infrared spectroscopy and multivariate analysis: a tool for determination of oil contamination in soils

*Alena Guryanova[a], Vasiliy Ermakov[a], Dmitry Bykov[a], Viacheslav Artyushenko[b], Andrey Bogomolov[a,b]*
*[a]Samara State Technical University, Samara, Russia*
*[b]art photonics GmbH, Berlin, Germany*

Current laboratory methods for the determination of oil content in contaminated soils provide necessary accuracy of analysis, but they are not effective enough for the purposes of environmental monitoring. Analysis of the published data revealed that infrared spectroscopy techniques generally satisfy the accuracy requirements. Being implemented in portable spectrometers they can also provide high speed and frequency of analysis in the field.

Here we report on a study, where the method of attenuated total reflection mid infrared (ATR MIR) spectroscopy has been used to quantify the degree of oil-contamination in the soil. Partial least squares (PLS) regression was used as a multivariate calibration tool. The analysis and evaluation of ATR MIR-spectra of more than one hundred samples of the real and model soils with different contents of oil and water has been performed. To improve the calibration accuracy PLS models were separately constructed for the soil samples with low, medium and high levels of the oil pollution. It has been proved that suggested method meets the regulatory requirements and can be used for the initial determination of oil contamination level in the soil.

## P23. Geoecological estimation of solid waste landfills with the use of multivariate data analysis methods

*Saveliev A.A, Ramenskaya E.V. Martynenko E.G.*
*Samara State Technical University, Samara, Russia*

The numerous methods of land development of landfills include recreational, forestry-based, landscape, etc. However, the creation of forest plantations and parks on the landfills territories is problematic because of the ecological specific character of landfills. As an alternative to the widespread methods of development, the construction and economic direction of territory development can be suggested, that is the construction of waste sorting stations, areas of ballast and recycled fractions briquetting, enterprises on the construction waste recycling into the secondary gravel, sites of stacking or grate contaminated soil decontamination in the surface of waste disposal facilities.

The existing approaches to the assessment of the solid municipal waste condition usually take into account their geometry, humidity, organic matter content, methane content, but the potential ability of array solid municipal waste to serve as a construction base of the lowered stability and responsibility buildings is not considered.

The authors suggest the fundamental principles that prove the possibility of construction and economic development of derelict territories and define the measure groups necessary for the full assessment of solid municipal waste conditions.

The use of the methods of the multivariate data analysis and the results interpretation in the form of digital state matrixes allow to distinguish heterogeneous objects of various geo-mechanical structure and chemical properties in the municipal solid waste structure depending on the conditions of object formation and the geo-environment stay period.

Differentiation of solid waste landfills on the heterogeneous parts and research of the geo-mechanical characteristics of the landfill soil allow to distinguish constant areas suitable for the construction of the solid municipal waste recycling complex.

# P24. Study of the variations in recovered paper properties with the help of multivariate analysis

*Ruslan O. Shabiev*

*Saint Petersburg State University of Industrial Technologies and Design, Russia*

It is well known that the pulp and paper properties correlate with each other, but the traditional one-to-one-factor analysis does not establish the exhaustive relationship between them because of a vast number of different properties being counted and measured.

The pulp and handsheet quality parameters of the recovered paper from different sources have been obtained based on laboratory means. The recovered paper is mainly represented by the used corrugated case materials namely old corrugated containers grade (OCC). As much as ten recovered paper samples have been withdrawn and analyzed for the variations of nine independent responses. The approach being used in the research was fully based on Chemometrics in Excel – the course and book written by Professor Alexey Pomerantsev [1].

The study found that there are two large groups in the scores plot – one is represented by the best samples and another by the samples which have been being influenced by some unknown factors. The loadings plot allowed to interpret the presence of the groups. The study has proven a possibility of using PCA for analysis of pulp and paper properties.

**References**

1. Pomerantsev A., Chemometrics in Excel, Tomsk, 2014, 435 p.

# Participants

**Amigo Jose Manuel**
University of Copenhagen
Associate professor
Copenhagen, Denmark
jmar@life.ku.dk

**Babushkin Valeriy**
LLC FOSS Electric
Application engineer
Moscow, Russia
vb@foss.dk

**Belikov Vadim**
Samara State Technical University
Ph student
Samara, Russia
mymailbox-desu@yandex.ru

**Belikova Valeria**
Samara State Technical University

Samara, Russia
valerya.belickova@yandex.ru

**Bogomolov Andrey**
art photonics GbmH

Berlin, Germany
ab@globalmodelling.com

**Chertes Konstantin**
Samara State Technical University
Professor
Samara, Russia
chertes2007@yandex.ru

**Cruz Jordi**
Escola Universitària Salesiana de Sarrià
Lecturer
Sabadell, Spain
jcruz@euss.cat

**Debus Bruno**
Saint Petersburg State University
Postdoc
Saint Petersburg, Russia
b.debus87@gmail.com

**Dokuchaev Alexandr**
Samara State Technical University
Software engineer
Samara, Russia
docuhaevrud@gmail.com

**Dubrovskiy Dmitriy**
Bashkir State University
student
Ufa, Russia
du315@mail.ru

**Duponchel Ludovic**
Lille1 University
Professor
Lille, France
ludovic.duponchel@univ-lille1.fr

**Egorova Ekaterina**
Saint Petersburg State Institute of Technology
PhD student
Saint Petersburg, Russia
katerinka7@list.ru

**Ermakov Vasiliy**
Samara State Technical University
Head of laboratory
Samara, Russia
wassiliy@rambler.ru

**Esbensen Kim**
KHE Consulting
Professor
Copenhagen, Denmark
ke@geus.dk

**Filippov Maxim**
CJSC "Koudijs MKorma"
QA/QC manager
Moscow, Russia
mfilippov@kmkorma.ru

**Galkin Evgenii**
CSort, Ltd.
Head of R&D department
Barnaul, Russia
tech9@csort.ru

**Galyanin Vladislav**
Samara State Technical University
Engineer
Samara, Russia
v.galyanin@gmail.com

**Hopkins David**
Battle Creek, USA
NIR.Consultant@sbcglobal.net

**Kabirova Liana**
Bashkir State University
student
Ufa, Russia
Kabirova.lian@yandex.ru

**Kapustin Denis**
CSort, Ltd.
Engineer
Barnaul, Russia
denkapustin@csort.ru

**Khannanov Maxim**
EnSpectr LCC
Head of department
Chernogolovka, Russia
mk@enspectr.com

**Kirsanov Dmitry**
Saint Petersburg State University
Associate professor
Saint Petersburg, Russia
d.kirsanov@gmail.com

**Kotenko Andrey**
Samara State Technical University
Associate professor
Samara, Russia
ako1959@mail.ru

**Krylov Ivan**
Moscow State University
Student
Moscow, Russia
krylov.ivan@gmail.com

**Guryanova Alena**
Samara State Technical University

Samara, Russia
belki.da@gmail.com

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

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

**Karpushkin Evgeny**
Moscow State University
Senior researcher
Moscow, Russia
eukarr@gmail.com

**Khaydukova Maria**
Saint Petersburg State University
PhD student
Saint Petersburg, Russia
khaydukova.m@gmail.com

**Kosarevskaya Svetlana**
Saint Petersburg State Institute of Technology
PhD student
Saint Petersburg, Russia
kossas13@mail.ru

**Kozhukhar Alexey**
LLC FOSS Electric
Senior application engineer
Moscow, Russia
ak@foss.dk

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

**Kurkina Victoria**
Saint Petersburg State Institute of Technology
Associate professor
Saint Petersburg, Russia
victoria.kurkina@gmail.com

**Leardi Riccardo**
University of Genoa
Associate professor
Genova, Italy
riclea@difar.unige.it

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

**Monakhova Yulia**
Saratov State University
Senior researcher
Saratov, Russia
yul-monakhova@mail.ru

**Nechaeva Natalia**
Moscow State University
PhD student
Moscow, Russia
Nechaeva.n.l@yandex.ru

**Podchasov Anton**
CSort, Ltd.
Software engineer
Barnaul, Russia
podchasov@csort.ru

**Pystin Vitaly**
Samara State Technical University
Engineer
Samara, Russia
vitaliy.pystin@yandex.ru

**Rodionova Oxana**
Institute of Chemical Physics RAS
Leading researcher
Moscow, Russia
rcs@chph.ras.ru

**Labutin Timur**
Moscow State University
Senior researcher
Moscow, Russia
timurla@laser.chem.msu.ru

**Lykina Anastasia**
Samara State Aerospace University
Student
Samara, Russia
zuum63@mail.ru

**Melenteva Anastasiia**
Samara State Technical University
Engineer
Samara, Russia
melenteva-anastasija@rambler.ru

**Moshchenskaya Elena**
Samara State Technical University
Assistant professor
Samara, Russia
lmos@rambler.ru

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

**Pomerantsev Alexey**
Institute of Chemical Physics RAS
Principal research fellow
Moscow, Russia
forecast@chph.ras.ru

**Ramenskaya Ekaterina**
Samara State Technical University

Samara, Russia
iljushinakatja@rambler.ru

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

**Rusinov Leon**
Saint Petersburg State Institute of Technology
Chief of department
Saint Petersburg, Russia
lrusinov@yandex.ru

**Savinkov Maxim**
CSort, Ltd.
CEO
Barnaul, Russia
smv@csort.ru

**Skvortsov Alexej**
Saint Petersburg State Polytechnical University
Head of department
Saint Petersburg, Russia
colbug@mail.ru

**Yarchuk Anna**
Moscow State University
student
Moscow, Russia
anechka1955@yandex.ru

**Zarubin Aleksey**
Tomsk Polytechnical University
Assistant professor
Tomsk, Russia
zagtpuru@gmail.com

**Zhilin Sergey**
Altai State University
Associate professor
Barnaul, Russia
szhilin@gmail.com

**Zontov Yury**
National Research University Higher School of Economics
Assistant professor
Moscow, Russia
yury.zontov@gmail.com

**Saveliev Aleksey**
Samara State Technical University
PhD student
Samara, Russia
alekssaveliev@mail.ru

**Shabiev Ruslan**
Saint Petersburg State University of Industrial Technologies and Design
Associate professor
Saint Petersburg, Russia
ncjob@yandex.ru

**Spiridonov Alexander**
LLC FOSS Electric
Application manager
Moscow, Russia
asp@foss.dk

**Yarkaeva Yulia**
Bashkir State University
PhD student
Ufa, Russia
julijajarkaeva05@gmail.com

**Zaytsev Sergey**
Moscow State University
PhD student
Moscow, Russia
sergz@laser.chem.msu.ru

**Zilberg Rufina**
Bashkir State University
Lecturer
Ufa, Russia
ZilbergRA@yandex.ru