

Drushbometrics Project

Seventh Winter Symposium on Chemometrics

Modern Methods of Data Analysis



Russia, Saint Petersburg, February 15–19, 2010

Saint Petersburg State Technological University of Plant Polymers
Scientific Consul for Analytical Chemistry RAS
Russian Chemometrics Society
Aalborg University Esbjerg
Institute for Geo and Analytical Chemistry
Institute of Chemical Physics

with the support of
Russian Foundation for Basic Research
Polycert



Seventh Winter Symposium on Chemometrics

Modern Methods of Data Analysis

The Organizing Committee

Co-chairmen

Lev Gribov
Victor Kurov
Alexey Pomerantsev

Secretary

Olga Derkacheva

Members

Andrey Bogomolov
Kim Esbensen
Nadezhda Khodyreva
Sergey Kucheryavskiy
Pavel Lukanin
Satu-Pia Reinikainen
Oxana Rodionova
Leon Rusinov

Address

SPb STUPP, 4, Ivan Chernykh str., St. Petersburg 198095, Russia

<http://wsc.chemometrics.ru/wsc7>

e-mail: wsc7@chemometrics.ru

Thanks

The WSC-7 organizers and participants wish to express their greatest appreciation to the following conference sponsors for their valuable economic and friendly help:

Russian Foundation for Basic Research (RFBR)

Russian Academy of Sciences

Saint Petersburg State Technological University of Plant Polymers

CAMO Software Company

John Wiley & Sons, Ltd.

Polycert Ltd

Finally, we are grateful to all the WSC-7 attendees, lecturers, accompanying persons, and visitors for their interest to the conference.

See you again at the next WSC-8 conference!

Useful information

Conference and activities

Conference sessions will be held in the conference-hall on the third floor of main building (no. 7 in the scheme below). Several saunas with swimming pools, as well as the cross-country skies, are available for the sport and leisure activities.

Meals

All meals (buffet) will be served in the dining room on the second floor of the main building.

Scores & Loadings

Traditional “Scores and Loadings” meeting will be held in the main building on the first floor (Disco Club). A bar offers you the drinks for very reasonable prices (e.g. 2\$ for a glass of wine or bottle of beer, 2.5\$–3\$ for 50 ml of whisky or cognac, etc.). Tea, coffee, cookies are free. Banquet will also be here.

Communication

Cellular network has a proper coverage around the hotel. WiFi Internet access is available on the third (near the conference-hall) and second floor of the main building.

Money

There is an ATM at the reception to get Russian rubles. Credit cards are not accepted in the hotel. The organizing committee could also change the reasonable amount of currency.

Excursion

On February 19, a tour over Saint Petersburg in the English and Russian languages (two buses) will be arranged for WSC 7 participants. These are the buses with which you leave Raivola. The tour comprises visiting the city’s historical centre with observing its numerous places of interest. Experienced guides will help you to know interesting facts of the Saint Petersburg history.

Connection with St. Petersburg

The hotel is located about 60 km from St. Petersburg. There is a public bus No. 680 between the hotel and the metro station Prospect Prosvescheniya (north part of blue line No. 2). The trip takes approximately 1.5 hours. You can also get to/from St. Petersburg by train; the train station Roshchino is about 2 km from the hotel.

Miscellaneous

The conference official language is English.

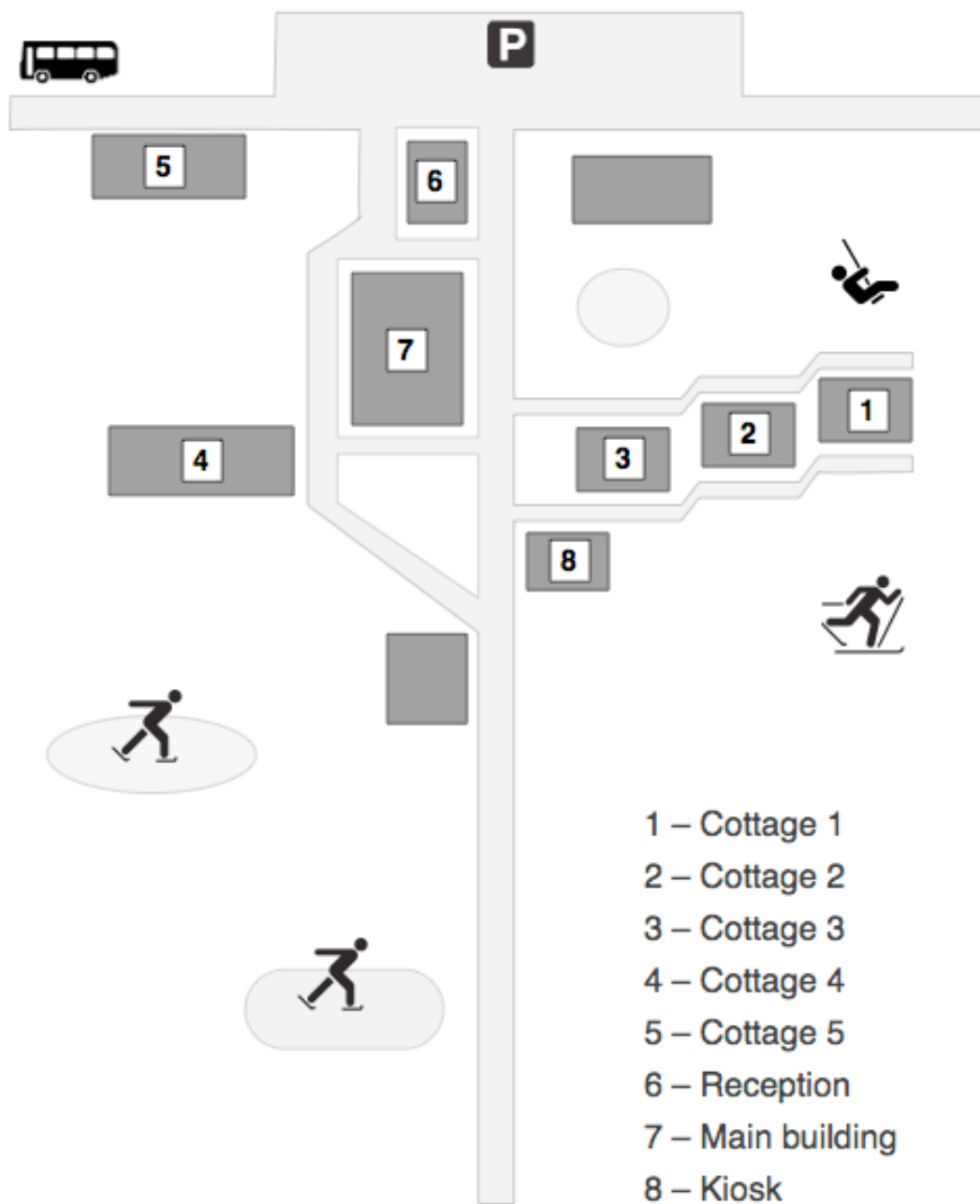
Everyone is encouraged to have his/her badge attached, both during the symposium sessions and social activities.

Useful Phone Numbers

Olga Derkacheva, *conference secretary*
Hotel Raivola

+7 909 590 7802 (mobile)
+7 812 322 5060 (local)

Scheme of hotel Raivola



Timetable

Monday, February 15, 2010

09:00–12:30	Registration
13:00–14:00	Lunch
14:00–14:30	Opening
Session 1	Chair: Oxana Rodionova
14:30–15:00	T1 <i>Nadezhda Khodyreva</i> St. Petersburg State Technological University of Plant Polymers — the host of WSC-7
15:00–15:30	T2 <i>Olga Derkacheva</i> The use of FTIR-spectroscopy and chemometrics for analysis of kraft pulps
15:30–16:00	T3 <i>Leon Rusinov</i> Statistical methods of process monitoring and diagnostics
16:00–16:30	Coffee break
Session 2	Chair: Leon Rusinov
16:30–17:00	T4 <i>Mikhail Elyashberg</i> Chasing molecules that were never there: misassigned molecular structures and the role of computer-aided systems in structure elucidation
17:00–17:30	T5 <i>Yuri Kalambet</i> Adaptive polynomial approximation of chromatographic peaks
17:30–18:00	T6 <i>Alejandra Urtubia</i> Predictive power of LDA to discriminate abnormal wine fermentations
18:00–18:30	T7 <i>Yulia Monakhova</i> Analysis of complex mixtures using self-modeling decomposition of different spectral data
18:30–19:00	Free time
19:00–20:00	Dinner
20:00–00:00	Scores & Loadings

Tuesday, February 16, 2010

08:00–09:30	Breakfast
Session 3	Chair: Marcel Maeder
09:30–10:30	L1 <i>Johan Trygg</i> The OPLS method and recent extensions — unlocking interpretation in two block modeling and multivariate calibration
10:30–11:00	T8 <i>Lennart Eriksson</i> O2PLS® and its applicability in hierarchical modeling, preference mapping and QSAR
11:00–11:30	Coffee break
11:30–13:00	Free time (skiing)
13:00–14:00	Lunch
Session 4	Chair: Johan Trygg
14:00–15:00	L2 <i>Marcel Maeder</i> Advances in hard-modelling of chemical processes
15:00–15:30	T9 <i>Alexey Pomerantsev</i> PAT solution to the drug release prediction
15:30–16:00	T10 <i>Jens Bo Holm-Nielsen</i> Process Analytical Technology — monitoring of biogas processes at meso scale test biogas plants
16:00–16:30	Coffee break
Session 5	Chair: Alexey Pomerantsev
16:30–17:00	T11 <i>Dmitry Kirsanov</i> Beer quality assessment by hybrid spectroscopic-electrochemical technique
17:00–17:30	T12 <i>Michael Madsen</i> PLS2 regression as a tool for selection of optimal analytical modality — a closer look at beer brewing process analysis with alternative spectroscopic data types
17:30–18:00	T13 <i>Andrey Bogomolov</i> In-line monitoring of yeast fermentation with 2D-fluorescence probe
18:00–18:30	T14 <i>Pentti Minkkinen</i> Comparison of independent process analytical measurements — a variographic study
18:30–19:00	Free time
19:00–20:00	Dinner
20:00–00:00	Scores & Loadings

Wednesday, February 17, 2010

08:00–09:30	Breakfast
Session 6	Chair: Karoly Heberger
09:30–10:30	L3 <i>Richard Brereton</i> One class classifiers in chemometrics
10:30–11:00	T15 <i>Oxana Rodionova</i> Another proof that chemometrics is usable: NIR confirmed by HPLC-DAD-MS and CE-UV
11:00–11:30	Coffee break
11:00–13:00	Free time (skiing)
13:00–14:00	Lunch
Session 7	Chair: Richard Brereton
14:00–15:00	L4 <i>Michal Daszykowski</i> Challenges in handling complex metabolomic data
15:00–15:30	T16 <i>Andris Jankevics</i> Quality control for large-scale LC-MS studies at runtime
15:30–16:00	T17 <i>Ivana Stanimirova</i> Simultaneous processing of data from multiple sources
16:00–16:30	Coffee break
Session 8	Chair: Pentti Minkkinen
16:30–18:30	Poster Session
18:30–19:00	Free time
19:00–20:00	Dinner
20:00–00:00	Scores & Loadings

Thursday, February 18, 2010

08:00–09:30	Breakfast
Session 9	Chair: Federico Marini
09:30–10:30	L5 <i>Kim Esbensen</i> Cross-validation — a chemometric dinosaur going extinct
10:30–11:00	T18 <i>Kaj Heydorn</i> The impact of metrology on chemometrics — a historical challenge
11:00–11:30	Coffee break
11:00–13:00	Free time (skiing)
13:00–14:00	Lunch
Session 10	Chair: Michal Daszykowski
14:00–15:00	L6 <i>Federico Marini</i> Particle swarm optimization (PSO): a potentially useful tool in chemometrics?
15:00–15:30	T19 <i>Anna Shelpakova</i> Calibration set design in spectrophotometric analysis of multicomponent mixtures
15:30–16:00	T20 <i>Karoly Heberger</i> Comparison of methods and models by consensus
16:00–16:30	Coffee break
Session 11	Chair: Kim Esbensen
16:30–17:00	T21 <i>Andrey Samokhin</i> Application of principal component analysis for extraction of pure mass spectra from gas chromatography/mass spectrometry data
17:00–17:30	T22 <i>Ralf Marbach</i> Message from the EE department: calibration of multivariate measurements is easy
17:30–18:00	Discussion
18:00–18:30	Closing
18:30–20:00	Free time
20:00–00:00	Banquet

Friday, February 19, 2010

08:00–09:30

Breakfast

09:30–14:00

Excursion

Abstracts

Lectures

L01. The OPLS method and recent extensions — unlocking interpretation in two block modeling and multivariate calibration

Johan Trygg

*Research group for Chemometrics, Computational Life Science cluster (CLIC),
Umeå University, Sweden*

Multivariate analysis (MVA) has been crucial in translating spectral data into information, mainly for predictive purposes. Orthogonal partial least squares (OPLS), is a novel development of the partial least squares (PLS) method with similar prediction properties. However, the distinct advantage of the OPLS method and its recent extensions (K-OPLS, O2PLS, OPLS-DA and BIF-OPLS) is that it provides an intrinsic improvement for model interpretation. This is achieved through sharper modeling of both the response matrix Y , provided by OPLS predictive components and uncorrelated profiles (i.e. orthogonal components in both X and Y). This partitioning of the X -data (and Y -data) into two sets of scores and loadings has a profound effect on model interpretation, some of which are listed below,

- Model diagnostics, e.g. how much of X is correlated to Y , and how much is not?
- Interpretation of scores and loadings
- Outlier detection and diagnosis
- Variable selection
- Multi-block modelling
- Pure profile estimation

The impact of the OPLS method from several applications areas will be discussed. These include pharmaceutical industry, plant biotechnology and medicine.

L02. Advances in hard-modelling of chemical processes

Marcel Maeder

Department of Chemistry, University of Newcastle, Australia

Hard-modelling of chemical processes is a neglected area in the field of chemometrics. While soft-modelling methods are powerful and the only option in many applications, hard-modelling is significantly more robust and importantly provides useful quantitative information such as rate or equilibrium constants. For that reason, it is much more commonly used by the wider community of chemists.

Hard-modelling consists of determining the appropriate chemical model that describes the process under investigation, together with the determination of the optimal values for a set of parameters defined by that chemical model. Several crucial aspects of hard-modelling that have been addressed in the recent past will be introduced and discussed in this presentation:

- For multivariate data potentially a very large number of parameters have to be fitted; in such instances, methods that differentiate between linear and non-linear parameters are absolutely crucial
- The determination of the correct model (chemical process) is often not easy and thus time consuming. Genetic algorithms can help.
- Many chemical processes are complex and their parameters cannot be determined from one single experiment; global analysis of many individual experiments taken under different conditions is the solution.
- Traditional chemical experimentation is done under strictly controlled conditions: thermostating to control temperature, buffering to control pH, and inert salts to control ionic strength; it is possible to allow changes in these conditions during the experiment by accommodating them into the computation. This can significantly experimental design.

The application of the above principles will be demonstrated using appropriate chemical systems. In particular, reference will be made to our present research which is directed towards the investigation of the chemistry of CO₂ in aqueous solution, particularly in the presence of organic amines. This research is relevant for PCC (Post Combustion Capture), the attempt of removing CO₂ from the exhaust gases of fossil fuel based power production. It could be a significant contribution to the abatement of the greenhouse effect.

L03. One class classifiers in chemometrics

Richard G. Brereton

Centre for Chemometrics, School of Chemistry, University of Bristol, UK

One class classifiers involve modelling each class in a dataset independently, in contrast to two class classifiers that divide dataspace into two regions. The difference between one and two (multi) class classifiers, between hard and soft models and between disjoint and conjoint PC models will be described.

Many one class classifiers in chemometrics derive from QDA (quadratic discriminant analysis), and attempt to form boundaries at a given level of confidence from the centroid of a class. If PC reduction is performed first then this can be either conjoint (on all data) or disjoint (on each class separately — the principle of SIMCA). For disjoint models it is advisable first to see how well the data fit into the PC model, often using the Q statistic. The D statistic, based on Hotelling's T² can be employed subsequently. It is possible to derive joint Q and D confidence limits. We illustrate the use of these statistics using iterative methods for formulating models to define Predictive Ability and Model Stability.

If data do not well fit into a normal distribution, these methods may not be appropriate, and an alternative, Support Vector Domain Description (SVDD) allows for more complex boundaries, and will be illustrated. In addition to the advantage of not requiring multinormality, SVDD do not require all models to be concentric, overcoming a problem of least squares solutions that can be influenced unduly by outliers. Movies will be shown to illustrate the change in boundaries according to the values of the penalty error and gaussian radius (for Radial Basis Functions).

Problems of optimising and validation of one class classifiers will be discussed and strategies for overcoming these. The use of ROC curves and Class Membership plots will be introduced.

The methods will be applied to a variety of datasets including simulations, metabolomics, polymer characterisation, forensics and environmental. They will be compared to two class and multiclass models. One class approaches are especially useful where there are many groups in the data or alternatively e.g. in multivariate process control where only one class (e.g. the Normal Operating Conditions region) can be adequately characterised.

L04. Challenges in handling complex metabolomic data

Michał Daszykowski

*Department of Chemometrics, Institute of Chemistry, The University of Silesia,
Katowice, Poland*

Over the past few years there has been a growing interest in understanding complex biological processes. Owing to the development of powerful analytical instruments, now, this objective seems to be within one's grasp. This is supported by many examples of successfully conducted omics studies described in the literature. In metabolomic studies, one is mainly interested in discovering a unique set of target compounds, often called biomarkers, indicating an on-going process, e.g., development of a disease. Usually non-targeted approaches are firstly used in order to collect as much chemical data as possible about the studied system. At present, hyphenated chromatographic techniques such as liquid chromatography coupled with mass spectrometry (LC-MS) are frequently adopted for metabolomic profiling [1, 2]. They can provide easily a huge amount of chemical information about each sample. The sample, analyzed by the LC-MS technique, can be represented by a two-dimensional chromatographic fingerprint (abundance measured at a given retention time and mass-to-charge ratio). Accordingly, the data collected are multi-dimensional and can contain many irrelevant features and some signals of questionable quality. These issues pose a real problem in data handling (exploration and modeling) and therefore, drawing general conclusions about the experiment and/or testing a hypothesis could be very sophisticated or even impossible. Similar problems related to data explosion and handling were also encountered many years ago, when near infrared (NIR) data appeared and stimulated an extensive chemometric research.

An appropriate use of well-suited chemometric methods can greatly facilitate the analysis of complex proteomic/metabolomic data. Specifically, with chemometric methods an enhancement of signal-to-noise quality, alignment of peak shifts observed among samples due to difficulties in maintaining the same conditions during analysis, can be achieved [3].

References:

1. G. Theodoridis, H.G. Gika, I.D. Wilson, *TrAC*, **27** (2008) 251-260
2. J.C. Lindon, J.K. Nicholson, *TrAC*, **27** (2008) 194-204
3. M. Daszykowski, B. Walczak, *TrAC*, **25** (2006) 1081-1096

L05. Cross-validation — a chemometric dinosaur going extinct

Kim H. Esbensen

ACABS research group, Aalborg University, campus Esbjerg, Denmark

Validation is concerned with assessing the performance of a specific data analytical model, be it for prediction, classification, time-series forecasting, or similar... In statistics, data analysis and chemometrics, a much favoured method of validation is cross-validation, in which a subset of the training set apparently performs as an 'independent test set' in a sequential manner. Depending on the fraction of training set samples (N) temporarily held out of the contemporary modelling, a range of no less than $(N-1)$ potential cross-validation segments will always exist, the specific number of segments falling in the interval $[2, 3, 4, \dots, (N-1), N]$. Various 'schools of thought' of cross-validation have developed within chemometrics, some favouring 'full cross-validation' (one object per segment; N segments in total), some defining 10 (or 4) as the canonical number of segments — with still other, more complex schemes are also offered. Usually however there is more focus on *strict adherence* to some form of cross-validation procedure, or other, than openness to investigate what exactly are the precise assumptions and prerequisites behind cross-validation. This comprehension has hitherto been mostly lacking — as has indeed also been a desire to discuss and comprehend the full suite of in-depth issues involved in a debate which has ranged throughout the entire history of chemometrics (sometimes a very heated discussion, often more emotional than rational).

This contribution discusses these issues in depth. The general conclusion arrived at is that cross-validation is only a *simulation* of test set validation, in form strikingly similar but not with regard to the essential characteristics. The crucial fact is that there is only one data set involved in cross-validation, namely the training set. This precludes any possibility for more than one realization of the sampling errors involved; there are both statistical as well as physical sampling errors in the general case, the latter collectively termed TSE (Total Sampling Errors). Given the fact that the physical sampling errors overwhelmingly dominate, typically 10–50–100 X analytical errors alone, it is evident that any singular N -object data set constitutes only one specific realization of these sampling error materializations. The main lesson from TOS' more than 50 years of practical experience is that there is no such thing as a constant sampling bias - the physical sampling bias changes with every new sampling from heterogeneous materials as well as from similar

measurement scenarios. From this it follows that there can never be any guarantee that one specific training set realization will also be representative of all future similar data sets. Taking a second data set, the test set, becomes an absolute necessity. Incorporating this second data set is the only way in which to incorporate information from both TSE materializations, for example, in prediction performance validation. From this discussion, one can conclude with complete generality that all variants or schematics of the cross-validation type are inferior, indeed unscientific. Cross-validation must logically be discontinued — unless absolutely no option for test set validation can be demonstrated. It should be stated that there does exist a (very) minor class of scenarios (only) in which cross-validation still has merit, but absolutely no generalizations can be made on this basis. Only test set validation can stand up to the logical demands of the general validation imperative.

L06. Particle swarm optimization (PSO): a potentially useful tool in chemometrics?

Federico Marini¹, Beata Walczak²

¹*Department of Chemistry, Sapienza University of Rome, Italy*

²*Institute of Chemistry, University of Silesia, Katowice, Poland*

Particle swarm optimization techniques are a family of optimization methods based on the concept of swarm intelligence, that has been developed in the last 10–15 years. In particular, it is a population-based stochastic optimization technique inspired by social behavior of bird flocking or fish schooling. If compared to other evolutionary optimization techniques as GA, PSO has a simpler algorithmic structure and fewer parameters to be adjusted, both desirable characteristics in the light of their possible application to the solution of practical problems.

However, while successful applications have been reported in different engineering, computer science or signal processing problems, yet their potential for solving chemometric problems remains almost unexplored.

In this communication, the possibility of using PSO in different chemometric ambits will be explored both theoretically and by means of selected representative examples (robust regression, clustering, optimization of learning parameters).

Talks

T01. St. Petersburg State Technological University of Plant Polymers: incorporation of chemometrics into its research and educational activities

Nadezhda Khodyreva

Saint Petersburg State Technological University of Plant Polymers, Russia

Saint Petersburg State Technological University of Plant Polymers (SPb STUPP), the WSC7 host, is the only higher educational establishment in Russia providing integrated training of multiskilled specialists for industrial enterprises engaged in in-depth chemical processing of wood (i.e. for the pulp and paper industry).

The use of chemometrics approach can assist in solving of a number of both industrial and scientific problems arising in pulp production and paper making. First, this industrial sector is among those where on-line automated process control together with fault detecting is important and strongly required. Being coupled with some chemometrics approach versions such as PCA, methods of fuzzy logics, neural networks, etc., the process monitoring becomes more efficient. It makes possible early detection and identification of process faults as well as their elimination at early stages, not bringing protective systems into operation and increasing thus production reliability and efficiency. This has been shown in numerous studies on the use of statistical methods in process monitoring and diagnostics carried out by Prof L. Rusinov and his colleagues. Prof. Rusinov's presentation is incorporated in the WSC 7 programme.

At the Physics Department of our university chemometrics methods find their use as a mathematical tool for interpreting FTIR spectra of different wood species and different pulps. The approach developed by scientists of the Laboratory of Spectroscopy has made possible construction of calibration models usable for fast estimation of chemical parameters, reflecting changes in major wood components caused by different treatments of wood and its constituents (such as pulping, pulp delignification and bleaching, artificial and light ageing of pulp fibers, etc.). Some results of these studies are presented in the report of Dr. O. Derkacheva. In the poster of Master's Degree student A. Dynina (Dr. Derkacheva is a supervisor) it is shown that the application of FTIR spectroscopy together with PCR makes it possible to construct models for predicting paper brightness. This approach is of considerable promise for paper restoration studies.

Chemometrics approach was also used by Department scientists for decomposing vibrational spectra of the majority of oxide glass forming systems and particularly leads metaphosphate — TeO_2 glasses for the purpose to investigate intermediate-range order. Results obtained when applying factor analysis have confirmed those based on the constant stoichiometry groupings (CSG) concept. The concept has come from the finding that vibrational spectra of glasses can be interpreted as a superposition of relatively small number of unchangeable spectral forms belonging to CSGs. This concept opens the promising way for developing glasses characterized by the record Kerr coefficient and low scattering losses to be usable for Raman fiber lasers and amplifiers. These results obtained under the leadership of Prof O. Yanush will be also presented in the poster session.

The fact that chemometrics methods can be fruitfully used in research carried out in different fields of university specialization is among the reasons why it is important to include chemometrics as a discipline into university curricula especially at the Master's level. Recognizing this, SPb STUPP is making efforts to integrate chemometrics training components into its educational process. For example, the university together with other partners: LUT, Finland, AAUE, Denmark, ICP RAS, RF, having wide experience in chemometrics education has recently implemented the project aimed at determining features and basics of a distance learning course in chemometrics for technological and natural science mastership education

T02. The use of FTIR-spectroscopy and chemometrics for analysis of kraft pulps

O. Yu. Derkacheva, D. A. Sukhov

Saint Petersburg State Technological University of Plant Polymers, Russia

Wood is a complex, chemically heterogeneous material. Its components can be divided in two groups: structural components (cellulose, hemicelluloses, lignin) and non-structural ones (extractives and inorganic compounds). It is known that during delignification both content of components and their structure can change. Development of new technology for production of half-finished fiber materials needs new knowledge about transformation of basic wood components in cooking processes.

At studying structure of the components two problems occur. The one is searching isolation methods which do not change strongly structure of component. The

second problem is connected with selecting experimental method of analysis. A number of chemical and physical methods used for lignin structure characterization are destructive.

We develop approach for studying variation of wood components in the process of kraft cooking of different wood species without their extraction from fibers by using FTIR-spectroscopy. When cellulose, lignins, hemicelluloses are studied in situ by the spectral method there is a risk of interference from wood constituents. The structural complexity of major wood polymers and their possible transformation contribute to the uncertainty of spectroscopic data. To decide this problem different mathematical tools are used.

In this report the factor analysis was applied to study the spectra of four wood species and kraft pulps from soft- (spruce and fir) and hardwood (birch and aspen) obtained at different degrees of delignification (Klason lignin: 2–30%) by laboratory conventional kraft cooks. Each sample was characterized by lignin content and yield value. These two characteristics are important parameters for this processing. After carrying out some treatment of IR-spectra the calibration models for definition of Klason lignin and yield value in investigated woods and pulps with using PLS1 method for each series and for all samples have been constructed. Calibrations were built by using all the spectral data or selected regions. For lignin content, the calibrations present good performances in terms of R^2 (> 0.96). The yield parameter (range of 45–100 %) was more difficult to predict.

The factor analysis allowed to reveal the number of principal spectra, see correlations of bands and do some interpretations. For each group of pulps from four wood species there was evaluated a number of factors to describe experimental spectra with good precision. The main factors were critically analyzed by comparing with spectra of "pure" wood components: isolated cellulose, hemicellulose and hard- and softwood lignins.

Thus, the use of chemometrics allowed to create calibration models for a fast estimation of the important chemical and economical parameters, and also to reveal the some information to interpret spectra of different origin pulps.

T03. Statistical methods of process monitoring and diagnostics

L.A. Rusinov, I.V. Rudakova

Saint-Petersburg State Technological Institute, Russia

Many processes in the chemical industry are potentially dangerous and require a permanent monitoring of their conditions. In case of an abnormal situation, it is highly important to determine its origin, and provide control engineers with all the necessary instructions on the recovery operation. The process monitoring and diagnostics is carried out on the basis of diagnostic models, which are able to detect the faults based on symptoms. Taking into account, that most of the processes have no analytical model describing their behaviour, especially in abnormal situations, the statistical modelling becomes important, as it does not need any knowledge of process chemistry.

The processes data is commonly multivariate, so the application of Principal Component Analysis (PCA) becomes effective. It allows not only reducing the dimensionality of a monitoring problem, but also the use of Q and T2 statistics to carry it out. There are a number of method modifications, e.g. "moving" PCA. With their help, it is possible to monitor fault processes with significant speed differences, like batch processes.

Furthermore, many production processes in the chemical industry are essentially non-linear. Therefore, the use of conventional linear PCA for their control may lead to monitoring errors. At the same time, the application of non-linear PCA methods, including kernel PCA, may result in significant computational difficulties and still cannot guarantee a sufficient sensitivity for exact detection of abnormal situations.

The origin identification of an abnormal situation in the space of principal components is a complicated task. Usually, it is executed by the estimation of variable contributions in the fault statistics. However, it is quite difficult to determine the fault, especially if it involves changes in many variables, as it frequently happens in practice. In this situation combined diagnostics methods are of use. Thus, attracting an expert information allows more effective identification. For this purpose, the two-level frame-based production diagnostic model, connecting symptoms with faults by means of fuzzy production rules can be constructed. The root frames on the upper level of the model contain information on the faults in the structural units that are selected during the decomposition of the

controlled process. The daughter frames of the model contain fuzzy production rules, which determine certain faults in the actual structural unit. The origin of the fault depends on the situation similarity degree that can be determined according to some criterion.

If the expert data, connected to an actual structural unit, appears to have a lack of information, there comes the fuzzy clustering of the data for the benefit of the diagnostics. Using the clusters, the fuzzy diagnostic model, for example, the one with production rules of the Takagi-Sugeno type, can be built.

The two-level neural network diagnostic model, constructed and trained on expert analysis results can be a good choice. The upper-level network serves for the faults isolation, while the bottom-level networks determine the origin of the faults. The PCA is used for significant dimensionality reduction of the upper-level network. It allows a considerable reduction of the networks retraining time and to keep the main benefits of the expert system without inserting it into the structure.

T04. Chasing molecules that were never there: misassigned molecular structures and the role of computer-aided systems in structure elucidation

M. Elyashberg

Advanced Chemistry Development, Moscow, Russia

The presentation was encouraged by the excellent review [1] of Nicolaou and Snyder entitled "Chasing Molecules That Were Newer There: Misassigned Natural Products and the Role of Chemical Synthesis in Modern Structure Elucidation". The title of our presentation was obviously chosen taking into account to underline connection between the problems discussed in both reports.

According to the review [1], around 1000 articles were published between 1990 and 2004 where originally determined structures were revised. Figuratively speaking, it means that 40-45 issues of the imaginary "Journal of Erroneous Chemistry" were published where all articles contained only mistakenly recognized structures and, consequently, at least the same number of articles was necessary for revision of these structures. The labor (and not only labor) expenses necessary for structural misassignments and subsequent reassignments become at least twice greater than in the case of getting initially correct solution.

It is evident that the stream of publications in which structures of new natural products and new products of organic synthesis determined incorrectly is large enough and there is a problem how to diminish this stream. The authors [1] comment that "there is a long way to go before natural product characterization can be considered a process devoid of adventure, discovery, and, yes, even unavoidable pitfalls".

We suppose that application of modern Computer-Aided Structure Elucidation (CASE) systems (see review [2]) can frequently help the chemist to avoid falling into a pitfall or, if nevertheless the researcher found himself in the pitfall, the expert system can give a signal "caution!" ("it seems you are in pitfall", "be carefully!"). Our hope is based on the fact that the molecular structure elucidation can be formally described as deducing all (without any exclusions) logical corollaries from a system of statements which ultimately form a partial axiomatic theory related to a current spectrum-structural problem. These corollaries are all conceivable structures that meet the initial system of axioms.

The history of CASE systems development convincingly confirmed the point of view suggested 40 years ago [3] that, as a matter of fact, the process of molecular structure elucidation is reduced to logical inferring the most probable structural hypothesis from a set of statements reflecting the interrelation between a spectrum and a structure. This methodology was implicitly used for long time before computer methods appeared. Independently on application or ignoring computer-based methods the way to the target structure is the same. CASE expert systems mimic considerations of human expert. The main advantages of CSASE systems: 1) all statements about interrelation between spectrum and structure ("axioms") are expressed explicitly; 2) all logical consequences (structures) following from the system of "axioms" are deduced completely, without any exclusions; 3) the process of the computer-based structure elucidation is very fast, which gives a tremendous saving of time and labor of the scientist; 4) if the chemist has several alternative sets of axioms related to a given structural problem, an expert system allowed rapid generating all consequences from each of the sets and identify the most probable structure by comparing the solutions obtained.

In our presentation the main kinds of "axioms" used during the molecular structure elucidation are discussed. The axioms are classified in the following three groups: 1) Axioms and hypotheses reflecting characteristic spectral features; 2) Axioms and hypotheses of 2D NMR spectroscopy; 3) Structural axioms necessary for

structure assembling. When the structure elucidation of a new chemical compound is performed with assistance of expert systems all axioms are expressed explicitly. Therefore it becomes possible to investigate dependence of the solution of a structural problem on any change in the initial set of axioms.

We will consider a series of examples in which the original structures suggested by researchers were revised later in successive works. All examples were taken from recent publications in respectable international journals. In each case, we show how the structure could be quickly and correctly identified if reliable MS and 2D NMR data were available and expert system Structure Elucidator [4] was employed. We will also show that if only 1D NMR spectra are available, empirical calculation of ¹³C chemical shifts [5] for suggested structures frequently allow researcher to realize that some structural hypothesis is most probably incorrect. Figuratively speaking, Structure Elucidator can be used as an analytical tool resembling a "polygraph detector".

The considered approach and presented examples allow one to come to conclusion that application of an expert system similar to the Structure Elucidator for the structure elucidation of new complex organic compounds, particularly natural products, can prevent inferring incorrect structures, which is not excluded even for highly qualified and experienced organic chemists. It is possible to expect that worldwide application of CASE systems will reduce the stream of publications containing erroneously elucidated chemical structures.

References:

1. K.C. Nicolaou, S.A. Snyder. *Angew. Chem. Int. Ed.*, 2005, **44**, 1012-1044
2. M. E. Elyashberg, A.J. Williams, G. E. Martin. *Prog. NMR Spectrosc.* 2008, **53**, (1/2), 1-104
3. M.E. Elyashberg, L.A. Gribov, V.V. Serov. *Molecular Spectral Analysis and Computers* (in Russian). Nauka, Moscow, 1980, 308 P.
4. M. E. Elyashberg, K. A. Blinov, A. J. Williams, S. G. Molodtsov, G. E. Martin. *J. Chem. Inform. Model.*, 2006, **46**, 1643-1656
5. K. A. Blinov, Y. D. Smurnyy, T.S. Churanova, M. E. Elyashberg, A. J. Williams. *Chemom. Intell. Lab. Syst.*, 2009, **97**, 91-97

T05. Adaptive polynomial approximation of chromatographic peaks

*Yuri A. Kalambet, Sergey Maltsev, Dmitry Gladkov
Ampersand Ltd., Moscow, Russia*

Savitsky-Golay algorithm is one of the most frequently used noise filtration algorithms in chromatography. It is based on polynomial approximation of the chromatogram and provides excellent reduction of noise without big modification of peak shape. Besides, in some case it is difficult to implement it, as in the case of e.g. triangular peaks, like in capillary electrophoresis, optimal numbers of points for effective noise reduction for left and right slopes of the peak are significantly different.

We implemented peak approximation using different number of points for either slope of the peak, depending on the slope “width”. This approximation allows accurate calculation of peak parameters even for very asymmetric peaks. This kind of peak approximation is complemented by effective noise filtration, also based on Savitsky-Golay algorithm.

Filtration and approximation results are compared using several examples from different kinds of chromatography.

T06. Predictive power of LDA to discriminate abnormal wine fermentations

Alejandra Urtubia¹, Jean-Michel Roger²

¹Department of Chemical and Environmental Engineering Universidad Técnica Federico Santa María, Valparaíso, Chile

²UMR ITAP (Information & Technologies for Agro-Processes), Montpellier, France

Wine fermentation is a critical step of winemaking. Abnormal behaviors can affect seriously the quality of the final product. However it is difficult to know this information with anticipation. In this study, the predictive power of Linear Discriminant Analysis (LDA) was evaluated in order to discriminate the behavior of wine fermentation. Different chemical concentrations were estimated from MIR spectroscopy yielding temporal profiles of concentrations from eighteen industrial wine fermentations of *Cabernet Sauvignon*. The statistical procedure consisted in a pre-treatment with an exponential curve fitting, then the parameters of these curves were used to feed LDA and finally the evolution of the prediction error as a

function of the time was calculated, giving an idea of the predictive power of the method. The methodology was applied to different times between the beginning and the end of fermentation (72, 100, 150, 250 and 400 hours). The results were similar in all the times studied; they revealed that seven of twenty eight variables minimized the Standard Error of Cross Validation (SECV) for the different times. The slopes of the ethanol and density profiles had high influence in the performance of the model. Before 100 hours, it is possible to predict if the fermentation will have problems or not. The objective was to predict as soon as possible the behaviour of fermentation, thus the enologist could correct the problems in time and the process would finish optimally.

Acknowledgements:

A. Urtubia thanks the financial support from FONDECYT Project (No 11070251).

T07. Analysis of complex mixtures using self-modeling decomposition of different spectral data

Yu. B. Monakhova, S.P. Mushtakova, S.S. Kolesnikova

Department of Chemistry, Saratov State University, Saratov, Russia

A system where no a priori information about chemical components is available is classified as a black system. The aim of performing curve resolution on black systems is to estimate the spectra (qualitative analysis) and concentration profiles (quantitative analysis) of the entire chemical component present in the mixture.

In this work we have compared performances of different self-modeling curve resolution techniques in identification and quantitative analysis of substances of different nature. Traditional self-modeling (SIMPLISMA, ALS) and independent component analysis (MILCA, SNICA, RADICAL, FASTICA and JADE) algorithms were applied to experimental standardless analysis of complex mixtures and real objects. Results are presented for several types of spectroscopic signals: UV-VIS, IR, and fluorescence.

Consideration of the composition of waste waters in the petroleum industry and widespread environmental contaminants has defined our choice to study certain organic compounds. We have investigated up to five component mixtures with various concentrations of benzene, toluene and o-xylenes in the UV-region. We have also examined various multicomponent mixtures of different compositions (up to six substances) of PAHs in their solid and liquid states by means of IR, UV

spectroscopy and fluorescence. We have analyzed complex mixtures of water and fat-soluble vitamins B6, B9, B12, PP, C, E, D, A as well.

Application of ICA decomposition algorithms on the data from real objects is of great practical interest, especially where the mixture composition is not known exactly. We have applied ICA and ALS to the analysis of vitamins and metals in ten complex multivitamin drugs. Spectra of individual vitamins were extracted and compound concentrations in the initial drug were obtained with 10% relative error. We also examined different fuels (gasoline, diesel fuel) and carried out quantitative and qualitative analysis of five aromatic compounds and three sulfur-containing organic substances. Their detection limits were comparable to the chromatographic determination.

The study has shown that the nature and the number of components present in mixtures, as well as selected spectral region and other instrumental factors, like noise, step, and speed of scan, severely influenced the results. In most cases, however, MILCA, SNICA and ALS provided superior results than other techniques used.

T08. O2PLS® and its applicability in hierarchical modeling, preference mapping and QSAR

Lennart Eriksson

MKS Umetrics AB, Umeå, Sweden

O2PLS® is a generalization of PLS and OPLS. Its objective is to enhance visualization and understanding of complex data sets. In contrast to PLS and OPLS, O2PLS is bidirectional, i.e. $X \leftrightarrow Y$. In other words, X can be used to predict Y (as with PLS and OPLS), but at the same time Y can be used to predict X (unlike PLS and OPLS).

O2PLS allows the partitioning of the systematic variability in X and Y into three model compartments: the X/Y joint predictive variation; the Y-orthogonal variation in X; and the X-unrelated variation in Y. The first part of the O2PLS model expresses the joint information overlap in X and Y, the second the information that is unique to X (and not found in Y) and the third the information that is unique to Y (and not found in X).

Each part of the O2PLS model comprises scores and loadings with a familiar meaning. Hence, from a conceptual point of view, interpreting an O2PLS model is

no different from interpreting a classical PLS model. From a practical point of view, however, the upshot is that the analyst has a better idea of what is being interpreted, since there is no confounding of predictive and non-predictive information.

The objective of this communication is to disseminate the O2PLS method to a broader chemometrics community. To this end, we shall investigate three multi-response data sets covering classical multivariate problems (hierarchical modeling, preference mapping and QSAR).

T09. PAT solution to the drug release prediction

A. L. Pomerantsev^{1,2}, O. Ye. Rodionova¹

¹*Institute of Chemical Physics RAS, Moscow, Russia*

²*State South Research & Testing Site RAS, Sochi, Russia*

Numerous studies are devoted to mathematical modeling of drug release. The term "release" encompasses several processes that contribute to the API transfer from the dosage form to the bathing solution. In our case, the drug release kinetics has a sigmoid form with pronounced induction period for moderate and final coating thickness. Such kind of models is rarely presented in the literature. Several attempts have been made to explain the S-shaped profiles but till now there is no generally recognized theory. For this purpose, we propose the autocatalytic kinetic model, which still has no obvious physical background.

Two types of experiments were studied. The first one is the dissolution test, which shows the API release from the pellet core. Nonlinear regression analysis with the successive Bayesian estimation of the regression parameters is applied for processing of the dissolution profiles.

The second type of experiment is the NIR spectra acquired in the course of coating process. As the thickness of the pellet shells and the drug process dissolution are tightly connected, the spectra can be used for prediction of the drug release profiles. NIR data set is considered in PLS analysis as the predictor matrix X. The response matrix Y consists of the kinetic constants values estimated at the dissolution curves' fitting.

We have showed that such combination of the hard kinetic model and the soft PLS regression provides a unique possibility for the on-line prediction of the drug release profiles in the course of the pellets' production.

T10. Process Analytical Technology — monitoring of biogas processes at meso scale test biogas plants

Jens Bo Holm-Nielsen and Kim H. Esbensen

ACABS research group, Aalborg University, campus Esbjerg, Denmark,

Biogas plants converting animal manure materials and other biomass feedstocks by anaerobic digestion (AD) are among the cheapest and most effective tools for achieving post-Kyoto targets concerning reduction of the emission of greenhouse gases of CO₂, CH₄, N₂O. Focus in this study is on a mesoscale biogas test plant implementation of Process Analytical Technologies (PAT) to develop chemometric multivariate calibration and prediction models for on-line monitoring and control of the anaerobic digestion process in a re-current loop modus. The larger goal is to prepare for implementation of on-line monitoring and control applications in the biogas and biorefinery sectors, in order to be able to adjust the process according to the demands of the bioenergy supply for producing combined heat and power when needed over the 24 hours or for the biorefinery sector to optimize the product yielding.

Most studies reported in the literature have investigated near infrared spectroscopy (NIR) in laboratory-scale or minor pilot biogas plants; only very few studies exist for on-line process monitoring of full-scale biogas plants of meso-scale biogas plants. It is here necessary to obtain a fairly constant level of VFA concentration which leads to a stable biogas production. VFA concentration levels should not exceed 4–5000mg/l. On-line control and management of VFA concentration levels can speed up or slow down the AD-process.

By comparing pilot plant NIR-spectra to laboratory VFA reference concentrations at the experimental locality Bygholm, it was possible to develop highly satisfactory calibration models by Partial Least Squares (PLS) regression, i.e. acceptable to very good PLS-prediction models for total VFA as well as for all essential individual acids. The average statistics assessing prediction performance, accuracy (slope-value) and precision: correlation (r^2), were both 0.92. VFA contents had a significant impact on biogas production. The Bygholm studies took place in a meso-scale 150L bioreactor supplemented with a recurrent loop of the TENIRS modus.

T11. Beer quality assessment by hybrid spectroscopic-electrochemical technique

Dmitry Kirsanov¹, Andrey Legin¹, Irina Papieva¹, Alisa Rudnitskaya^{1,2}, Anna Kartsova¹, Alexey Bocharnikov³, Viacheslav Artyushenko³ and Andrey Bogomolov⁴

¹St. Petersburg University, Chemistry Department, St. Petersburg, Russia

²University of Aveiro, Chemistry Department, Aveiro, Portugal

³Fibre Photonics Ltd, Berlin office, Berlin, Germany

⁴J&M Analytik AG, Essingen, Germany

A combination of two powerful analytical techniques, optical spectroscopy (OS) and electrochemical potentiometric multisensor system (electronic tongue, ET) has been studied within a joint multi-sensor technology for beer quality assessment. Based on different physical principles, ET and OS are not just different, but also complementary techniques that derive different kinds of chemical and physicochemical information about a sample. Such a hybrid optical-electrochemical analyzer appears to be very attractive for express quality control. It could offer an alternative to the standard physicochemical analysis as well as the sensory taste assessment, which are time-consuming and may be too expensive for small- to middle-scale breweries.

Twenty different brands of beer produced in Germany, Russia and Czech Republic were analysed. Most of them were Pilsener and other high-quality blond beers, having a very similar composition. It was necessary to test the method's capability of recognising delicate taste-related differences. Some other different beer types like dark, alcohol-free, strong, yeast-containing and filtered white beer were tested as well. All the samples were analyzed with the potentiometric electronic tongue and with a number of different spectroscopic techniques. The latter included: attenuated total reflectance infrared (ATR-IR), near infrared (NIR), visible (Vis), and ultraviolet (UV) spectroscopy as well as fluorescence at two different excitation wavelengths and 2D-fluorescence spectroscopy (full excitation-emission spectra). In most cases the spectra were measured directly in a sample through an appropriate immersion probe, operating in ATR, transmittance, transreflectance or diffuse reflectance mode. The measurement geometry and conditions were optimised for the beer analysis. Reproducible data acquisition procedures, suitable for sample quantitative analysis, were developed in each case.

Independent reference analyses of the same beer samples were carried out by certified laboratories in Germany. The analysis has covered standard

characterisation of the beer quality attributes in accordance with German regulations, e.g. extractivity, alcohol content, colour, pH, foam resistance, turbidity, concentrations of bitter-forming components and other standard parameters. Besides, the samples were assessed by a trained commercial sensory panel to estimate taste and flavour details as well as the overall quality in accordance with a standard procedure.

In the present work we report the results of multivariate analysis performed on the collected data. The sensitivity of individual techniques to certain beer components and properties is presented. The main accent is made on the feasibility analysis of the suggested hybrid ET-OS technique to detect the main taste-related attributes of the beer quality.

T12. PLS2 regression as a tool for selection of optimal analytical modality — a closer look at beer brewing process analysis with alternative spectroscopic data types

Michael Madsen and Kim H. Esbensen

ACABS research group, Aalborg University, campus Esbjerg, Denmark

Intelligent use of modern process analysers allows process technicians and engineers to look deep into the dynamic behaviour of production systems. This opens up for a plurality of new possibilities with respect to process optimisation. Oftentimes, several instruments representing different technologies and price classes are able to decipher relevant process information simultaneously. The question then is: *how to* choose between available technologies without compromising the quality and usability of the data. We apply PLS2 modelling to quantify the relative merits of competing, or complementing, analytical modalities.

We here present results from a feasibility study, where Fourier Transform Near InfraRed (FT-NIR), Fourier Transform Mid InfraRed (FT-MIR), and Raman laser spectroscopy were applied on the same set of samples obtained from a pilot-scale beer brewing process. Quantitative PLS1 models were constructed for the most interesting parameters, namely ethanol, maltose, and total sugar. Models were naturally validated with appropriate test sets. The same three set of multivariate spectra were subsequently related to one another using a sequential PLS2 approach allowing to quantify the proportion of *unique* vs. *joint* information in each spectral data type. These results reveal new quantitative interrelationships between NIR, MIR and Raman spectral information on biological systems.

T13. In-line monitoring of yeast fermentation with 2D-fluorescence probe

Andrey Bogomolov¹, Thorsten Grasser² and Martin Hessling²

¹*J&M Analytik AG, Essingen, Germany*

²*University of Applied Sciences, Ulm, Germany*

Yeast fermentation is the basis of various biotechnological food and beverage productions. Traditional sensor-based monitoring cannot provide necessary level of process control and do not fulfill today's quality requirements. To meet them, new in situ methods of process analysis are actively developed [1].

2D-fluorescence spectroscopy has a great potential as an in-/ on-line analytical technique for biotechnological process monitoring. Due to its high sensitivity and selectivity it enables identification and quantification of (co-)enzymes, both extra- and intra-cellular, which typically fluoresce. The latter may act as markers indicating the concentration and metabolic state of the culture [2]. At the same time, fluorescence spectra, obtained during the running fermentation, are prone to multiple interferences, including excitation light intensity drift, scattering and absorbance effects, overlapped fluorophor signals, cascade fluorescence etc. Therefore, the analysis of two-dimensional excitation-emission process spectra may be challenging and require advanced chemometrics [3].

Lighthouse Probe™ (LHP) [4] was adapted to acquire 2D-fluorescence spectra and tested in the *Saccharomyces cerevisiae* yeast fermentation. Data collection, analysis and interpretation issues are discussed. Some new approaches are suggested to separate useful signal from the background. Taking additional information into the model, i.e. NIR-spectrometric scattering sensor data, may significantly improve the method performance.

References :

1. S. Hisiger and M. Jolicoeur. *J. Biotechnology* **117** (2005), 325-336.
2. C. Lindemann, S. Marose, H. O. Nielsen and T. Scheper. *Sensors and Actuators B* **51** (1998), 273-277.
3. P. P. Mortensen and R. Bro. *Chemom. Intell. Lab. Syst.* **84** (2006), 106-113.
4. A. Bogomolov, M. Melichar, A. J. Wigmore and K. Grau. CAC-2008, 30th June-4th July, 2008, Montpellier, France, Conference Materials, 2 (201).

T14. Comparison of independent process analytical measurements — a variographic study

Pentti Minkkinen

Lappeenranta University of Technology, Finland

Process analytical measurements can be based on samples cut from the process streams and analyzed in the laboratory. On-line and in line process analyzers and sensors are nowadays widely used to control and monitor chemical processes. Process analyzers need regular checking and updating of their calibration and all analytical systems should also go through proficiency testing which sometimes are mandatory. Both the calibration check and proficiency tests provide independent measurement sets from the same target. There are two main reasons which make comparisons of the mean values of independent measurements difficult. One is weighting error [1], which can cause systematic errors, if the flow-rate varies and correlates with the analyte concentration. To eliminate this error source the mean of the measurement set should be calculated as weighted mean by using flow-rates of sampling time as weights. Weighting error is the reason why often the mean estimated by process analyzers and by laboratory measurements do not agree.

Another reason is that the process data are often autocorrelated. In this case the standard deviation of the mean of the measurement set depends on sampling mode: whether *random*, *stratified* or *systematic* sample selection is used. Consequently, the statistics based on the assumption on normality cannot be used to estimate the uncertainty of the mean. Instead some form of chronostatistics has to be used. Pierre Gy has developed a variographic technique [2, 3] especially process analytical measurements in mind. Variography is a practical tool for characterizing the process heterogeneity. It also provides variance estimates for the different sampling modes for different sample intervals (lags) which take into account the effect of autocorrelation. The method is elucidated by using real and simulated data sets. The simulations are also based on real data sets.

References:

1. Minkkinen, P., Weighting Error — Is it Significant in Process Analysis?, WCBS3 Conference Proc., Publ. Ser. Fundação Luis Englert 1/2007, 59-68
2. Gy, P.M., *Sampling of Heterogeneous and Dynamic Material Systems*, Elsevier, Amsterdam, 1992
3. Pitard, F.F., *Pierre Gy's sampling theory and sampling practice* (2nd edn.), CRC Press LLC: Boca Raton, USA, 1993

T15. Another proof that chemometrics is usable: NIR confirmed by HPLC-DAD-MS and CE-UV

O. Ye. Rodionova¹, A. L. Pomerantsev^{1,2}, L. P. Houmøller³, A. V. Shpak⁴, O. A. Shpigun⁴

¹*Institute of Chemical Physics RAS, Moscow, Russia*

²*State South Research & Testing Site RAS, Sochi, Russia*

³*Arla Foods amba, Videbæk, Denmark*

⁴*Moscow State University, Moscow, Russia*

The study is devoted to the recognition of the 'high quality fakes' and addresses two important issues. The first one is the possibility of applying the NIR/chemometric approach for detection of the injection remedy, which is a water solution with low concentration of active ingredients, directly in the closed ampoules. The second issue is the comparison of the NIR/chemometric conclusions with the detailed chemical analysis. In this study we have shown that a low selective method such as NIR spectrometry can discriminate very similar objects and may be applied for a rapid screening. NIR results were confirmed by HPLC-DAD-MS and capillary electrophoresis analysis. These methods are much more time consuming and may be applied to the suspicious objects revealed by NIR/chemometric approach.

The study was caused by a strange particular case when a counterfeited remedy was revealed by manufacturer with the help of special hidden marks on a package, but the standard GC pharmacopie test could not recognize the forgery.

T16. Quality control for large-scale LC-MS studies at runtime

A. Jankevics¹, R.A. Scheltema¹, S. Grinberga², O. Pugovics², R. Breitling¹

¹*Groningen Bioinformatics Centre, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, The Netherlands*

²*Latvian Institute of Organic Synthesis, Riga, Latvia*

Liquid Chromatography Mass Spectrometry (LC-MS) is a powerful and widely applied method for the study of biological systems, biomarker discovery and pharmacological studies. LC-MS measurements are, however, significantly biased by several factors, including: (1) ionization suppression/enhancement, interfering with the correct quantification of analytes; (2) detection of large amounts of derivative ions, increasing the complexity but not the information content; and (3) machine drift during extensive sample sequences, altering mass and quantification

accuracy. Traditionally, quality control (QC) samples (e.g., pooled samples of aliquots taken from all samples) are analyzed randomly throughout the analytical sequence. Afterwards the analyst can use the results of the QC samples to determine if there was a gradual change during the analysis or whether a sudden deterioration had occurred at some point midway through the analysis. Based on the analysis of the QC samples it can be decided in the worst case to discard the dataset and repeat the whole experiment, resulting in additional time and money spent.

Here we present a method capable of detecting instrument performance at measurement runtime, offering the possibility of early stop of the whole experiment, saving both time and samples. The method works simply by extracting information from the raw data of a set of analytes known to be present in all samples. These ubiquitous analytes might serve as quality markers. Several mathematical methods are then applied to the RT-values, m/z and intensity data of these analytes, resulting in the detection of significant changes. The method also results in meta-information (achieved mass accuracy, quantification dispersion, adducts, fragments, contaminants, etc.), which will assist in the further identification of unknown metabolites.

The methodology is currently being tested on batch of urine samples measured on a UPLC Q-TOF MS instrument.

T17. Simultaneous processing of data from multiple sources

I. Stanimirova and B. Walczak

Department of Chemometrics, Institute of Chemistry, University of Silesia, Katowice, Poland

Much attention has recently been paid to the comprehensive characterization of samples in order to obtain a more complete understanding of the system being studied. Very often a set of samples is described by measurements obtained from different instrumental methods. In food quality control, for example, the product composition is often analysed by the use of a separation method such as gas chromatography (GC) or liquid chromatography (LC), whereas the various sensory characteristics are evaluated by experts trained in sensory analysis or by electronic nose or tongue. The aim then is to find the underlying relationship among variables of different sets of data. Several chemometric approaches can be adopted to explore such blocks of data simultaneously, among which are SUM-PCA [1],

STATIS (structuration des tableaux a trois indices de la statistique) [2, 3] and MFA (multiple factor analysis) [4, 5]. For a successful data exploration, preprocessing to correct for scale differences among variables and for variation differences among blocks is crucial.

References:

1. A. Smilde, J. Westerhuis, S. de Jong, *J. Chemometrics* **17** (2003) 323–337.
2. Ch. Lavit, Y. Escoufier, R. Sabatier, P. Traissac, *CSDA* **18** (1994) 97–119.
3. I. Stanimirova, B. Walczak, D.L. Massart, V. Simeonov, C.A. Saby, E. Di Crescenzo, *Chemom. Intell. Lab. Syst.* **73** (2004) 219–233.
4. B. Escofier, J. Pages, *CSDA* **18** (1994) 121–140.
5. I. Stanimirova, B. Walczak, D.L. Massart, *ACA* **545** (2005) 1–12.

T18. The impact of metrology on chemometrics — a historical challenge

Kaj Heydorn

Technical University of Denmark, Roskilde, Denmark

A generation ago two separate disciplines set out to deal with important scientific problems associated with multivariate data processing and with measurement uncertainty, neither of which were handled in a satisfactory manner by classical statistics. Since then mature scientific tools have developed under the names of chemometrics and metrology, however more or less independently of each other. Today's challenge is to stimulate interaction between these two important fields of science, and this lecture will present examples on areas, where such efforts are most urgently needed.

In 2008 the Bureau Internationale des Poids et Mesures issued a completely new vocabulary covering terminology to be used in metrology for many years to come; it presents definitions of metrological concepts to clarify exactly what is meant by various terms. The most significant changes from previous terminology will be presented and discussed in relation to terms currently used in chemometrics.

Co-operation between chemometrics and metrology to ascertain mutual understanding is a condition for further scientific development of both these important disciplines.

T19. Calibration set design in spectrophotometric analysis of multicomponent mixtures

A.S. Shelpakova, I.V. Vlasova, S.M. Dobrovolsky

Omsk State University, Omsk, Russia

In this work we have studied designing the calibration set both in MLR and PLS methods applied to the spectroscopic analysis of multicomponent mixtures at small deviations from spectral additivity. Absorbance spectra were obtained in the range of 220–350 nm with a step of 0.2 nm, using SF-2000 spectrophotometer equipped with a one-centimeter quartz cavity. The MLR and PLS regression were conducted in Microsoft Excel and the Unscrambler 9.8 respectively.

Binary mixtures of Papaverine (P) and Dibazole (D) as well as four- to six-component mixtures of water-soluble group B vitamins were studied. The components were mixed in a wide concentration range at different component ratios.

MLR has been applied to the P-D system. Five mixtures were used in the calibration set. The test set included 35 mixtures. Three methods of the calibration set design were used. The first one was random sampling. In the second method, we used two mixtures with maximum component ratios, one mixture with the maximum component concentrations, and two mixtures with intermediate ratios (2:1, 1:2, 3:1, 1:3, and 20:1). The third set was designed using the nominal mixture (1:1) and four mixtures, where concentrations were shifted from the nominal one by 25% and 50% (0.75:1.25, 0.5:1.5, 1.25:0.75, 1.5:0.5). The test set validation has shown that such a factorial design gives the smallest relative error in prediction (REP), which is less than 5% for all components. REP was calculated by the formula

$$\bar{\delta}_X = \frac{1}{J} \sum_{j=1}^J \frac{|C_{Xj} - \hat{C}_{Xj}|}{C_{Xj}}$$

where J is the number of mixtures in the test set, and X = P or D.

This kind of design was applied to B vitamin mixtures containing up to six components. For example, a calibration set for the four-component system was designed using the nominal mixture and eight mixtures with $\pm 50\%$ deviations from it. The MLR and PLS regression models were built and validated with a test

set of fifty mixtures. The obtained REP values have demonstrated that the PLS model predicted all component concentrations better than MLR.

The proposed method of calibration set design has also been employed for the analysis of the real-world objects, namely, for some medicines and for bird vitamin supplements.

T20. Comparison of methods and models by consensus

Károly Héberger¹ and Klára Kollár-Hunek²

¹*Chemical Research Center, Hungarian Academy of Sciences, Budapest, Hungary*

²*Budapest University of Technology and Economics, Budapest, Hungary*

Numerous different methods have been elaborated in modern analytical chemistry. One of the most important key problems in chemometrics is to compare various methods in a fair way. If the data can be arranged in a matrix form without empty cells the ranking of methods (or models) can be carried out. Features (variables, columns) and samples (cases, rows) characterizing the methods form the input matrix. The novel ranking procedure is based on sum of ranking differences (SRD) [1]. The features should be expressed on the same scale. The absolute values of differences for the ideal and actual ranking are summed up and the procedure is repeated for each (actual) feature. The SRD values obtained such a way order and group the features simply. If the ideal ranking is not known, it can be replaced by the average (consensus modeling).

The background idea is the same as that for collaborative tests: if the systematic errors of a given method are not known, it is expedient to measure it with various methods and assuming that the errors from different sources follow normal distribution. Whereas the noise is random, the signal is systematic. i.e. the noise cancels out, but the signal accumulates.

The validation can be performed with a kind of permutation test. Random features should be simulated and ordered by SRD procedure. The comparison of random and real SRD values shows unambiguously the reliability of ranking. Random features (>100 000) were generated for each number of samples $13 < n < 61$, and for each fifth number of samples (65, 70,... etc.) if $n > 60$. An approximation with Gaussian distribution has been used for these larger number of samples, the error term was less than 10^{-5} . For small number of samples ($n < 13$) the true discrete distribution has been used.

The novel ranking method can be applied in many different instances from determination of panel consistency [2] via column selection in chromatography to the selection of the best models [1]. Several examples, determination of number of latent variables [3], selection of the best method for measuring antioxidant capacity in bears [4] and comparison of chemometric methods for classification [5], show the usefulness of the procedure unambiguously in diverse fields.

References:

1. K. Heberger, *TrAC*. **28** (2009), doi:10.1016/j.trac.2009.09.009
2. K. Kollar-Hunek, J. Heszberger, Z. Kokai, M. Lang-Lazi, E. Papp, *J. Chemom.* **22** (2008) 218-226.
3. R. Todeschini *ACA* **348** (1997) 419-430.
4. Haifeng Zhao, Wenfen Chen, Jian Lu, Mouming Zhao, *Food Chemistry* **119** (2010) 1150-1158.
5. R. Todeschini, D. Ballabio, V. Consonni, A. Mauri, M. Pavan, *Chemom. Intell. Lab. Syst.* **87** (2007) 3-17.

T21. Application of principal component analysis for extraction of pure mass spectra from gas chromatography/mass spectrometry data

A.S. Samokhin, I.A. Revelsky

Division of Analytical Chemistry, Chemistry Department, Moscow State University, Moscow, Russia

Gas chromatography/mass spectrometry (GC/MS) is a hyphenated analytical technique that allows performing qualitative and quantitative determination of components in different matrices (drugs, biological and environmental samples and etc.). However, overlapping of chromatographic peaks is possible when unknown multicomponent complex mixture is analyzed. In this case some approaches should be used to obtain pure mass spectra of components and identify unknown substances. To solve this problem background subtraction and methods based on comparison with library spectra, using certain peak shape or reconstruction from mass chromatograms are applied. These approaches have different limitations (especially in case of complete overlapping of components).

In this work we have investigated possibility for resolution of overlapped peaks and purification of respective mass spectra using principal component analysis (PCA). Simulated and real GC/MS data were considered. Real GC/MS data were

obtained using Agilent 7890A (Agilent Technologies) gas chromatograph coupled with TruTOF (Leco) mass spectrometer. Scan range — 50–350 a.m.u., scan rate — 20 scan/sec.

Simulated data were obtained in Maple 13.0 software. Gaussian shaped chromatographic peak profiles were used. Data matrix was calculated by multiplying matrix of chromatographic peak profiles and matrix of pure noise-free mass spectra. Then matrix with randomly distributed noise was added to data matrix. Resolution of chromatographic peaks ranged from 0.075 to 0.75 and components concentration ratio ranged from 0.01 to 100.

Rows of data matrix corresponded to mass spectra and columns of data matrix corresponded to mass chromatograms. PCA was performed in order to calculate scores and loadings. In the case of GC/MS data first n columns of loading matrix are abstract mass spectra which represent linear combination of mass spectra of n coeluting substances. Loading matrix was changed by multiplying at transformation matrix to obtain pure mass spectra of coeluting components.

Applicability of PCA to purifying of coeluting substances mass spectra containing a lot of common ions was shown. It has been shown that this method gives satisfactory results even when chromatographic resolution (R_s) less 0.05.

Acknowledgements:

The authors are grateful to A.V. Garmash (Lomonosov Moscow State University) for fruitful discussion.

T22. Message from the EE department: calibration of multivariate measurements is easy

Ralf Marbach

Optical Instruments Center, Oulu, Finland

Multivariate measurement and calibration is not fundamentally different from the univariate case, rather, the two correspond one-to-one in every aspect. The scientific "mother discipline" of both cases is time signal processing, not statistics, and everything becomes simple when looked at from the angle of time signal processing. There is a simple general formula that describes the solutions of all multivariate calibration methods. There are only two inputs into this formula, namely, the spectral signal (response vector) and the spectral noise (covariance matrix). The two variables are easily recognized when looking from the point of

view of time signal processing. The measurement results ("predictions") are optimized, in the mean-square error sense, when both variables are "matched" i.e., when both are good estimates of the true parameters. This is a direct analog to the famous matched filter a.k.a. Wiener filter you are carrying around in your mobile phone. In spectroscopy, the method has become known as "science based calibration" (SBC).

The SBC method is shortly reviewed and the advantages for calibration of spectroscopic analyzers are given. Compared to PLS, cost and time of calibration is drastically reduced, often by as much as 80%, at typically better results.

The correct mathematical definitions of sensitivity and specificity in the multivariate case are given. Both limits are testable from first principles, i.e., from measurable pieces of spectroscopic data. Current standards for testing specificity (ASTM 1655 etc.) are shown to be wrong and misleading. The importance of applying spectroscopic expertise and application knowledge to the calibration process is stressed, as is the need to scientifically estimate both the "spectral signal" and the "spectral noise" parts required in calibration. Both estimates are important because proof of specificity is a two-step process. The user must first prove that the multivariate measurement measures "the right thing" and then, second, that this "correct" measurement is not affected by any unspecific correlations.

Posters

P01. Estimation of volume fractions in two-phase flow systems based on acoustic chemometric

Benjamin Kaku Arvoh, Maths Halstensen

Applied Chemometric Research Group (ACRG), Telemark University College, Porsgrunn, Norway

Application of new techniques in real time estimation of volume fractions of two-phase flow systems in the oil industry is still being developed. Volume fractions in gas-liquid and liquid-liquid real time systems are estimated based on principles of acoustic chemometrics. The feasibility of applying acoustic chemometrics for on-line monitoring of volume fractions of trace concentrations of liquid-liquid mixtures has been investigated.

Acoustic chemometrics is a relatively new technique for on-line process monitoring based on characterization of system vibrations generated by industrial processes. Gas-liquid and liquid-liquid mixtures were pumped through a rig. In order to generate acoustic signal, two orifice plates were mounted (horizontal and vertical pipeline). The acoustic sensors were mounted on these plates since earlier feasibility study demonstrated this position as optimum. The spectra from the acoustic sensor were transformed from time domain to frequency domain by Fast Fourier Transform (FFT) together with other signal processing techniques prior to multivariate calibration. Subsequent applications of Partial Least Squares regression PLS-R were used to calibrate a multivariate prediction model for online monitoring of volume fractions in the pipeline.

In both gas-liquid and liquid-liquid two phase systems, concentrations from 0—100% levels were investigated. The pressure drop in the system was approximately 1.5bar and was kept constant throughout the process. The temperature in the rig was also maintained at a constant value throughout the experimental process. One model each based on test set validation was developed for both the gas-liquid and liquid-liquid systems. These models were then used to estimate on-line volume fractions in both systems. The average Root Mean Square Error of Prediction (RMSEP) for both the gas-liquid and liquid-liquid systems did not exceed ± 5 and 8% respectively. The correlation between the predicted and measured in both systems was approximately 0.98.

The results from applying acoustic chemometric techniques in estimating on-line volume fractions of gas-liquid and liquid-liquid systems were promising but more experimental work is needed to explore the effect of all relevant process conditions/variations in such systems.

P02. Multivariate analysis and mass spectrometry for drug testing

*S. A. Appolonova, P. A. Baranov, N. V. Mesonzhnik, and G. M. Rodchenkov
Antidoping Centre, Moscow, Russia*

The 2010 prohibited list of the World Anti-Doping Agency (WADA) covered nine classes of substances (S1–S9), three categories of prohibited methods (M1–M3) and two groups of substances (P1 and P2) prohibited in particular sports disciplines. All prohibited substances and methods of doping are subject of doping controls.

The constantly increasing numbers of compounds that potentially risk the integrity of sport the complexity of objects and problems in doping control also requires modern multivariate approach for data analysis. Combinations of various chemometric techniques with mass spectrometry should give possibility for efficient investigations in different areas of doping control such as

- Metabolites identification (The tool where the identified metabolic differences observed can be visualized);
- Structure classification (The tool which will help maximize the separation between classes);
- Target analysis;
- Non-target analysis.

Examples of the activities of various chemometric models, illustrating achievements in doping control will be shown. The aims and application problems of multivariate analysis with mass spectrometry for drug testing in doping control successfully, and the variants to accomplish this, will be outlined.

P03. ESI-MS/MS and multivariate analysis application for structural classification of 3-Oxosteroids

S.A. Appolonova, P.A. Baranov and G.M. Rodchenkov

Antidoping Centre, Moscow, Russia

The collision-induced dissociation (CID) fragmentation pathway data in Electrospray ionization mode coupled with principle component analysis (PCA) and Agglomerative Hierarchical Clustering (AHC) were used for structural analysis and classification of 3-oxosteroids.

All 3-oxosteroids analytes have been divided into 7 groups according to the number and position of double bonds and the presence of functional groups: 4-ene-3-keto; 1-ene-3-keto; 1,4-diene-3-keto; 4,9,11-triene-3-keto; poly-ene-3-keto; unconjugated 3-keto and hydroxylated 3-keto steroids. Tandem mass spectrometry measurements were used at the collision energy of 20eV to obtain product ions for each analyte. Experimental design and multivariate analysis were performed in XLSTAT (Addinsoft, NU, USA version 2009.3.02). All mass spectra data were converted into text files and then imported into XLSTAT as an array of m/z and intensity values. For all analytes the spectral data were referenced by finding out the common mass shifts due to the $[M+H]^+$ ion. PCA and AHC analysis were applied to all previously referenced data transferred into XLSTAT matrix for the structural model evaluation. Analysis of the obtained model showed that the compounds were divided into 7 groups corresponding to the numbers and position of double bonds and the presence of functional groups.

Application of multivariate analysis to the ESI-MS/MS data provides to be potentially efficient method of structure analysis for 3-oxosteroids. The model was verified on the 3-oxosteroids analytes that were not previously used in model evaluation. The method outlined here can be used for the structural analysis of other steroids and other classes of analytes. This technique gives a powerful tool for the rapid identification of analytes in complex matrixes.

P04. The data processing method of sensor system for pH monitoring

*T.Y. Rusanova , N.A. Levina, N.A. Burmistrova, T.V. Dikanev
Saratov State University, Saratov, Russia*

Determination of the pH has critical importance for a wide range of applications, namely in the medical and environmental fields. Optical pH sensors (small and simple device) are suitable for remote monitoring in hazardous environments and insensitive to electromagnetic interferences. The main disadvantage of optic pH sensors is narrow dynamic range. The use of several indicators can overcome this problem but requires special data processing. The sol-gel technology was used for preparation thin films with immobilized two pH-indicators (pyrocatechol violet and bromophenol blue) at the surface of glass plates. The UV-Vis spectrum of the films with indicators for number of buffers solutions (pH 1.0–8.0) and standard samples were obtained. It has been shown that detection of the analytical signal (absorbance) at one wavelength does not give unique solution about pH value. The possibility of using multidimensional data was investigated and the multifactor analysis was used. To evaluate the precision and the predictive capabilities of the model, the errors of calibration (RMSEC) and prediction (RMSEP) were calculated. Cross validation for calibration models was used.

P05. Mathematical simulation and strategy for optimization of ion chromatographic analysis

*Dolgonosov A.M.
Vernadsky Institute of Geochemistry and Analytical Chemistry, RAS, Moscow,
Russia*

Theoretically well-founded, adequate simulation of the chromatography is possible only on the base of the sorption dynamic theory with description of the thermodynamic equilibrium, hydrodynamics and kinetics of the sorption separation process. The mathematical model level mentioned was developed, and on its base the software IONCHROM was created. This simulator of IC allows to predict behavior of ionic mixtures in complicated conditions of the IC-separation with satisfactory precision, to optimize conditions of separation of given mixtures and to solve the inverse task for analysis of ionic composition of solutions. There are the sorption dynamic equations, the author's researches in the field of kinetic of high-performance IC, and developed by authors the method of optimization of

separation based on construction of the dynamic map of chromatographic system, which are the base of the mathematical model.

The model is adequate in wide range of varying parameters and can be used for the estimation of chromatographic behavior of different substances on the columns packed with various ion-exchangers. In the software for simulation of IC there is possibility to calculate isocratic and gradient modes.

A scheme allowed to adjust the theoretical calculation with chromatographic experiments and to minimize the time and expenses for the analysis of the unknown ionic mixtures was developed.

P06. The estimation of age of human remains using soft modeling approach

Ivan Belyaev¹, Sergey Fominykh², Vladimir Ynakovskiy², Sergey Kucheryavskiy³

¹Altai State University, Barnaul, Russia

²Altai State Medical University, Barnaul, Russia

³ACABS research group, Aalborg University, campus Esbjerg, Denmark

For many objects in nature there are certain relations between their structure and various properties, e.g. physical, chemical, biological and many others. In most of the cases, revealing of such relations is quite complex, but new analytical methods, based on soft modeling approach, allow to make descriptions of dependencies between structure of an object and its properties. In the present work the analysis of skeleton bones structure has been used for estimation of age of victims' remains in forensic medicine.

The skeleton bones are stable to putrefying transformations and have different age-related information features, first of all connected with osteoporosis process. However the quantitative estimation of related structure properties is not trivial and required a lot of experiments. The other issue here is that the relations between structural characteristics and age of individuals are not linear, as it has been shown by experiments with light absorbance.

The preliminary investigations showed that the combination of different methods for structure features extraction (like Angle Measure Technique and statistics of wavelet coefficients) and Partial Least Squares for modeling the relations between these features and age gives quite promising result. In this work we would like to present the developing of the method. The new results obtained using a novel

method for quantitative structure analysis, based on fractal dimension, and Support Vector Machines regression. Three sample sets were used to evaluate the model - calibration, validation and a control set for prediction. The comparison with previous results as well as with one of approaches used in forensic medicine will be shown.

P07. Acoustic chemometrics and sampling for volume fractions determination in oil/gas flow

Ihunegbo Felicia Nkem, Maths Halstensen

Applied Chemometric Research Group (ACRG), Telemark University College, Porsgrunn, Norway

Estimation of volume fractions in multiphase flow is typically a major concern in process sectors especially the oil industries. The aim of this study was to develop an online technique to address the demands of monitoring the volume fractions of the most challenging multiphase fluid, oil/ gas, flowing in a pipeline based on process sampling and acoustic chemometrics. An experimental rig, on which acoustic sensor and a newly constructed sampling device were mounted, was utilized for this investigation. The sampler was mounted on a bypass loop in the rig and reference samples of the fluid were collected automatically and evaluated. The volume fractions of each sample were compared with a priori known input volumes. It was concluded that the sampling error was quite low. Also, the acoustic signals of various mixtures of oil/gas flowing through the rig were recorded. Thereafter, multivariate calibration (Partial Least Square-Regression, PLS-R) method, based on test set validation, was utilized for the prediction of various volumes of oil, as well as gas in the oil/gas mixtures flowing in the rig. The pressure drop of the system was 2.5 bar and a constant temperature was maintained throughout the experiment. This study does not cover temperature variations and static pressure variations. The experimental results from the feasibility study were promising, but more experimental work is needed to use this technique for online monitoring in the oil industry.

P08. Analysis of MRI images of a brain of patients with dementia using multivariate approach

Alexander Dmitriev¹, Sergey Kucheryavski²

¹Altai State University, Barnaul, Russia

²ACABS research group, Aalborg University, campus Esbjerg, Denmark

Alzheimer disease (a common form of dementia) is one of the heaviest and incurable disease of a human brain, resulting in intellectual loss and low mental activity of a person. The diagnosis of the disease implies a full inspection of the patient including analysis of mental activity and studying of a brain by series of MRI images. Inspecting of MRI images first of all allows to estimate the form of a brain, to reveal infringements in his structure. However on early stages of dementia the morphological changes in a brain structure are not very obvious and often hidden completely.

In the present work a new method for analysis of MRI tomograms of patients with early stages of Alzheimer's disease is proposed. The method is based on soft modeling approach, where different integral and local features of tomogram structure are calculated and used further as predictors in classification models. A degree of illness (0 – patient without dementia, 0.5 – pre-dementia, 1.0 – early dementia), has been used as a class variable.

A priori knowledge of changes in a brain structure during development of dementia allowed us to narrow a range of potential features on a first step of this work. Than a set of numerical experiments have been carried out to the most relevant features. Finally the best results were achieved using statistics of 2D Fourier transform and fractal properties of brain convolutions.

P09. Classification of crystal drops images

Sergey Kucheryavski¹, Ivan Belyaev², Jose A. Marquez³

¹ACABS research group, Aalborg University, campus Esbjerg, Denmark

²Altai State University, Barnaul, Russia

³The High Throughput Crystallization Laboratory EMBL, Grenoble, France

Automatic crystallization systems allows to make about 100–1000 and even more crystallization experiments per hour. However only part of the experiments result in crystals and only some of the crystals are of the special interest. Usually such cases are detected manually by a person who recognizes crystals in drops by their

appearance. In order to improve the effectiveness of crystallization experiments, the automatic detection systems, based on image analysis approach, have been invented recently, but the recognition quality of these systems still needs significant improvements.

In the present work we have discovered the problem of recognition of crystal drops in general and some special cases, which are of a special interest, in particular. Several approaches, based on such methods as eugenfaces, wavelet transformation and texture analysis, have been used to get relevant image features. Principal component analysis was employed for exploratory data analysis, soft independent modeling of class analogue (SIMCA) and support vector machines (SVM) have been utilized for pattern recognition. The detailed comparison of the results is shown.

P10. Pattern recognition methods for profiling microbial communities

Agnieszka Lemanska¹, Karlheinz Trebesius², Karl Grammer³, Dustin J Penn⁴, Richard G Brereton¹

¹*Centre for Chemometrics, School of Chemistry, University of Bristol, United Kingdom*

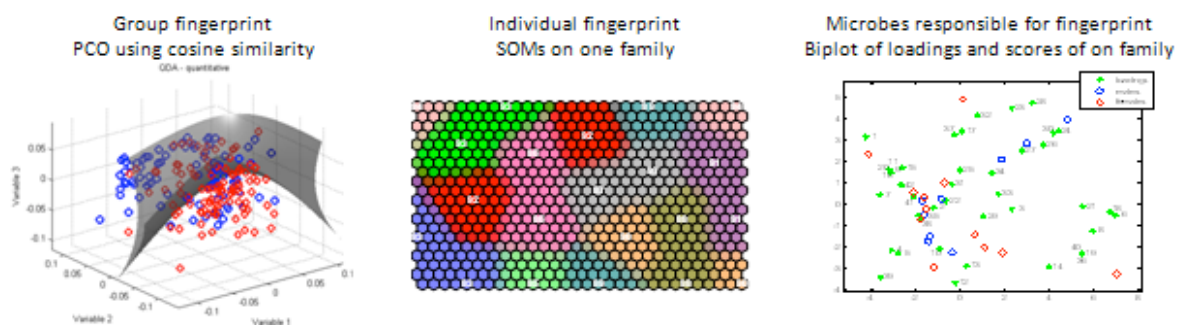
²*Vermicon AG, Munich, Germany*

³*Ludwig Boltzmann Institute for Urban Ethology, Department for Anthropology, Vienna, Austria*

⁴*Konrad Lorenz Institute for Ethology, Austrian Academy of Sciences, Vienna, Austria*

Human secretions contain large numbers of microbes. In this study carried out in a remote village in Austria microbial samples from armpits of almost 200 individual donors (88 males and 108 females) grouped into families were investigated. Several pattern recognition techniques were applied to reveal trends in microbial fingerprint of subjects.

DGGE (Denaturing Gradient Gel Electrophoresis) band tables that require accurate alignment of all the unique bands (corresponding to different microbes) among all the included samples may be difficult to produce. For smaller studies (one family) they are feasible. PCA (Principal Component Analysis) can be applied to band tables.



For larger studies automated pair-wise similarity measures are more feasible alternatives. In this approach both qualitative (presence or absence of unique bands using the Jaccard distance) and quantitative (presence and position using the cosine distance) measures combined with fuzzy matching of peak positions were applied. Patterns represented by dissimilarity matrices were revealed by PCO (Principal Co-ordinates Analysis).

Pair-wise dissimilarity and PCO analysis revealed considerable separation between genders. Rank analysis showed that the within individuals variation is significantly less than between individuals variation. This suggests that the individuals have unique microbial fingerprint. SOMs (Self Organising Maps) deliver information about specific microbes characteristic for individuals and demonstrate the existence of an individual fingerprint. The SOMDI (SOM Discrimination Index) together with supervised SOMs is described as a mechanism for determining which microbes are characteristic of individuals.

Group fingerprints (e.g. to differentiate males and females) can be determined using PCO components as input to supervised classifiers, including one class methods such as Quadratic Discriminant Analysis (QDA) and Support Vector Domain Description (SVDD) and two class methods such as PLS-DA (Partial Least Squares Discriminant Analysis) and Support Vector Machines (SVM). The validation of classification models is described.

P11. Analysis of multicomponent medicines without separation: concentration calculations for nonadditive analytical signals

I.V. Vlasova, V.I. Vershinin, A.S. Shelpakova
Omsk State University, Omsk, Russia

Spectrophotometry has been used for a long time for analysis of the unseparated multicomponent medicines. This easy, fast and sensitive method has the only drawback — low selectivity, which is connected with overlap of spectra and/or

interaction of components. Components interaction leads to the spectral nonadditivity and therefore to systematic errors (ΔC) in determination of components. Our investigation aims at estimation of this error. Three chemometric algorithms were applied: the Firordt method (FM), multiple linear regression (MLR), and regression by projection on latent structures (PLS). Individual organic components were determined both in the simulated mixtures and in the real medicines. Samples contained 2–7 analytes (n) with concentration ratio from 1:1 to 1:20. The spectra of aqueous solutions of model components (10^{-6} – 10^{-4} M), their mixtures and real medicines were obtained in the UV-region without addition of any reagent. Some analyzed mixtures had significant deviation from additivity (ΔA). The data were processed using Microsoft Excel and The Unscrambler.

The compared algorithms differ in their analytical abilities even in the trivial case when $\Delta A = 0$. It was shown that the demanded accuracy of analysis ($\Delta C < 5\%$ rel.) may be obtained with FM only for the simplest mixtures ($n = 2$ – 3); MLR is suitable for the more complicated systems ($n \leq 5$). The PLS method leads to the desirable results even for $n = 6$ – 7 . Small deviations from additivity do not prohibit using FM and MLR, but in such cases the analytical wavelengths (AWLs) had to be selected in a special way. A new criterion to select AWLs for FM was proposed. Calculation formulas were proposed relating the deviations of the absorption of a random binary mixture from additivity and systematic errors of determining components by the Firordt method. An algorithm was developed and validated using the model mixtures. This technique aims at simultaneous determination of two components mixtures with errors, which below a specified limit. Within the limits MLR it is expedient to calculate absorption factors for all components with spectral data concerning mixtures with well-known concentrations of components. Such approach compensates the effect of nonadditivity, and decreases the average ΔC values in 3–5 times. When $\Delta A \neq 0$, PLS provides better results than FM and MLR. In the case of multicomponent medicines all components may be determined by PLS with $\Delta C < 2\%$ rel.

These findings were applied in the express methods for analysis of polyvitamines and multicomponent medicines, such as papaverine + dibasol, etc. The results coincide with concentrations obtained by the other methods (HPLC) and with the nominal composition of examined mixtures.

P12. Instrument-oriented wavelets for data-processing of the analytical devices

Novikov L.V.

Institute for Analytical Instrumentation RAS, Saint Petersburg, Russia

The use of wavelet transformation in chemometrics, started in the early 1990-ies, has proven to be very useful for noise suppression, data shortening, and quantitative analysis based on linear models [1, 2]. In this work, wavelet method was used for increase of resolution and sensitivity that are the most important characteristics of analytical instruments. Wavelet processing was performed by utilization of synthesized or modified wavelets oriented on the treatment of signal for certain class of instruments or adapted for specific type of instrument. Synthesis is based on information a priori on the apparatus function, noise correlation function, and specifications for apparatus on the stage of primary treatment [3, 4, 5]. Presented approach allows to broaden wavelet processing on the calculation of noise estimations for not only signal but also for linear transformation of this signal. Demonstrated fast algorithms can be realized in real time by compact computing devices. Application of the method for analytical instruments allows to increase sensitivity three or four times and, simultaneously, resolution for rises up to the separation of completely overlapping peaks.

References:

1. Jetter K., Depezynski U., Molt K., Niemöller A. *ACA*. (2000). **420**.169–180.
2. Rodionova O.Ye. , Pomerantcev A.L. *Uspekhi Khimii* (2006) **75** 302-321
3. Novikov. L.V. *Instruments and Experimental Techniques*. 2007,**50**, 1.
4. Novikov. L.V. *Izv, Vuzov. Priborostroenie* (Russia). 2007, **50**, 3–10.
5. Novikov. L.V. *J. Communic. Technology and Electronics*. 2006. **51**. 1337–1346.

P13. Application of chemometric tools for coal calibration by DRIFT spectroscopy

Yu. M. Possokhov, V. K. Popov, V. I. Butakova

Eastern Research & Development Institute of Coal Chemistry, Yekaterinburg, Russia

The traditional determination of bituminous coal properties is time-consuming, requires special equipment and sample preparation routines. The innovation consists in use of diffuse reflectance mid-infrared Fourier-transform (DRIFT)

spectroscopy which provides rapid molecular analysis of coals and coal blends in the 350–7500 cm^{-1} region from laboratory-comminuted coal powders and thus reveals the most valuable information about the constitution of organic matter. It also stands for elaboration of the methodology able to determine coal properties with high reliability attending only to its DRIFT spectral data. Thus the coal calibration is a problem of QSPR/QSAR modeling.

Summarizing our long-term experience in coal calibration we mark out the next stages: (a) single spectrum preprocessing, (b) preprocessing of the spectrum population formed, (c) regression model building, and (d) regression model tuning.

Single spectrum preprocessing consists in choosing intensity units, specular and scatter correction, background estimation. None of these may be applied independently towards the coal property to model and the model accuracy. The choice of intensity units among theoretically valid log-scaled, Kubelka-Munk and reflectance ones is exposed as a non-linear transformation of predictor variables in terms of the Generalized Additive Model technique.

In order to eliminate specular and scatter artifacts, adversely affecting DRIFT spectra of coal powders, several data treatments may be applied. But neither traditional Multiplicative Scatter Correction (MSC), nor modern Extended Multiplicative and Extended Inverse Scatter corrections suit for single spectrum preprocessing being a part of population-level preprocessing routines. Moreover these treatments assume the mean spectrum to represent a reference (ideal) one and we have showed that it lacks of regards for FTIR-spectrometer peculiarities: its light source energy distribution and detector sensitivity. In exchange for above-mentioned corrections we have proposed algorithms for the hallmarked selection [1]. Also Standard Normal Variate (SNV) or sub-spectrum area normalization are then may be applied.

Background estimation is a very difficult problem for coal spectra as they have broad bands. For a long time we are using the expert-dependent technique of linear baselines built upon our earlier findings in bituminous coal structure [2]. Being at most automated this technique brings high selectivity in peak identification of structural groups. Such extraction of analytical signal from raw spectra may be implemented by difference spectra approach too [3].

There are two main tasks in preprocessing of the spectrum population. The first is to reach a uniform-like distribution of the modeled property while forming a

training subset. The second is to identify multivariate outliers. Manipulating populations of thousands spectra we can achieve much success in uniforming the distributions of the modeled properties. The second task is the outstanding problem. Traditionally the studentized residuals over ± 3 are considered as outliers. But in practice we usually observe the outlier masking and swamping effects well described in [4]. Therefore we apply the precalibration identification methods based on Mahalanobis distance and robust regression for outlier unmasking.

Regression model building is the culmination procedure in coal calibration. We have arrived at a conclusion that the competitive approach is the best for coal calibration as it combines several regression approaches into the unified predictive algorithm. Also we underline two types of regression techniques: the first based on spectrum-like data (functional data) and the second based on expert grouping of predictors. PLS-like approaches with non-linear effects such as Orthogonal Signal Correction (OSC) are powerful in the first case, and in the second case several machine learning methods (usually referred to as stochastic gradient boosting trees) are preferred.

Regression model tuning is that we first propose to use. The idea comes from the fact that there are obvious errors in a coal property determination. We can iteratively variate these errors in a valid tolerance range to reach a minimum of the model RMSEC. For that purpose a some kind of genetic algorithm may suit.

As a matter of preceding facts we have developed the on-line coal analyzer called SPEK-TROTEST employed in laboratories of coke and byproduct producers, cleaning plants, and mines in Russia and Kazakhstan [5]. The SPEKTROTEST is introduced in the National Standard Grading System of these countries. This commercial analyzer can predict up to 25 coal properties characterizing its quality. For example, ash is the hardest coal property to predict. Applying the above-mentioned chemometric tools in the SPEKTROTEST elaboration we have reached the predictive accuracy for Ekibastuz coals essentially close to required for ash range 10–70 abs. %. [3].

Further accuracy improvements of coal calibration is a conjunction of chemometric tools and our findings in coal structure. If the latter is not abandoned the prediction of coal properties by DRIFT spectroscopy will be extremely important in the near future.

References:

1. V. K. Popov, Yu. M. Posokhov, *Coke and Chemistry* **52** (11) (2009).
2. N. D. Rus'yanova, *Coal Chemistry* [in Russian], Nauka, Moscow, 2000
3. V. K. Popov, Yu. M. Posokhov, *Coke and Chemistry* **52** (12) (2009).
4. Ben-Gal I., Outlier detection, In: Maimon O. and Rockach L. (Eds.) *Data Mining and Knowledge Discovery Handbook*: Kluwer Academic Publishers, 2005
5. V. K. Popov, *Coke and Chemistry* **3** (2006) 13.

P14. Probing cisplatin reactivity towards biological nucleophiles by UV-spectroscopy and multivariate analysis

A.N. Skvortsov

Institute of Cytology, RAS, St. Petersburg, Russia

Cisplatin and its analogues are anticancer platinum coordination compounds, widely used in clinic. The biological activity of platinum drugs is caused by reaction with DNA bases, subsequent structural change of DNA, interference with normal DNA functioning, and cell death or rapidly dividing cells. One of the main problems of our current understanding of platinum drug metabolism and structure-activity relationships is the fact that platinum drugs may easily react with a vast variety of nucleophilic groups. Some of them, e.g. methionine thioether group, react with platinum drugs much faster, than DNA bases, and the future role of methionine-bound platinum is still unknown. It becomes accepted, that the most important properties of platinum drug are the rates of the reactions with different nucleophiles and their ratios, which control kinetically the distribution of the drug between DNA (therapeutic target) and byproducts. So kinetic studies of platinum drug reactions are of great interest.

Platinum drugs lack specific spectral properties and act at low concentrations, while their targets have large molecular weight, so the number of applicable research methods is limited. UV absorption spectroscopy may partially circumvent these problems, but the spectra need thorough numeric analysis, because of low resolution and low magnitude of changes. Model systems for studying cisplatin reactivity towards biomolecules by UV spectroscopy are a good application field for multivariate curve resolution techniques.

In the present work we studied cisplatin hydrolysis (the primary process, underlying most cisplatin reactions) monitoring the weak CT bands of the complex. Hydrolysis is generally a two-step reversible process, and its rates studied by various methods, unfortunately literature data are somewhat

controversial here. In biological media the second step is not realized, so we focused on the first step. We also studied cisplatin reactions with high molecular weight DNA and L-methionine, monitoring the specific absorption bands of biomolecules. Reaction of cisplatin with DNA in the presence of methionine (mimicking the DNA-protecting anti-platinum action of methionine in the cell) was tested. The use of UV-spectroscopy allowed to use low concentrations of platinum and make the models more relevant, than the ones previously studied by NMR. Multivariate analysis approaches, including PCA, EFA and matrix augmentation were used to analyze the obtained data matrices. Reaction rates were found by hard modeling. As reaction rates were more important for us, we did not focus on complete spectra reconstruction, and fitted curve profiles directly to primary components, obtained from PCA of the data matrices. This simplified approach still yielded good results, consistent with accepted mechanisms of the reactions and with reference data. The results show very good applicability of multivariate techniques for the system under study. Yet, reaction of cisplatin with methionine, displaying many well-reproduced steps was too complex for simple modeling.

P15. Multivariate classification and discrimination of oil-slime depositories in accordance with their condition

¹*V.V. Ermakov, ²A Bogomolov, ¹D.E. Bykov*

¹*Samara State Technical University, Samara, Russia*

²*J&M Analytik AG, Essingen, Germany*

Oil-slimes appear on all the stages of oil mining and processing. As a rule, waste accumulation is performed in specially designed places and sludge tanks. Depositories of large petrochemical complexes occupy hundreds of hectares and present places of danger. During the accumulation and preservation of wastes in oil-slime depositories some natural transformations are observed. Usually, light liquid hydrocarbons concentrate in upper layers, while middle layers tend to contain the water. Heavy hydrocarbon fractions, tars and mineral particles gather in the bottom.

Structure investigation of depositories in connection with the oil-slime composition was carried out for several places of oil extraction, transportation and preparation in Samara region. Analysed depositories had different age and slime formation sources. Geometrical dimensions of depositories, thicknesses and

densities of layers, as well as their chemical composition (water, diesel fraction, tars, chats and sulfur content) were determined. A quantitative characteristic describing quality changes of hydrocarbons is the ratio of diesel fraction to tars concentration. Gathered information, containing 34 variables, was analysed by means of Principal Component Analysis (PCA).

Four PCs were found to be enough to adequately describe the data. On the scores and loadings plots reveal the internal data structure that can be interpreted. One can distinguish groups of samples typical to the northern and southern clusters of oil fields of Samara region, as well as the areas typical to depositories of definite types. PC1 correlates with the sulfur content in all the layers and strippant, diesel content in the upper layer, layer density, middle layer thickness, mineral impurities content in all the layers and in the depository itself. PC2 is basically determined by water and tar content of the bottom layer, its thickness and age. PC3 is influenced by the ratio of diesel fraction to tars concentration. PC4 depends on the thickness of the bottom and upper layers and general water content. PCA and its interpretation was also carried out for each of the layers separately.

With the help of the developed model, unknown waste samples can be identified, e.g. assigned a geographic region or depository type. The principal components, derived from the data analysis of various oil slimes, are put into the basis of a new oil-slime classification and discrimination system, suggested by the authors.

P16. Sampling strategies for hyperspectral image models

James Burger

BurgerMetrics SIA, Jelgava, Latvia

The application of Partial Least Squares regression (PLS) and PLS Discriminate Analysis (PLS-DA) to NIR spectra for quantification and classification purposes are well established and proven technologies. However the creation and validation of calibration models requires a known Y value for every spectrum included in the model training and test sets — either a quantity or class member value. This makes extension of PLS to hyperspectral imaging difficult, since accurate Y values for each pixel based spectrum are typically not known.

In conventional spectroscopy it is common to use a single bulk sample value — an average value representing a homogenous sample, matched to a single sample spectrum. In the case of hyperspectral imaging, how should the individual pixel

spectra be used? A single spectrum can be computed from the set of spectra selected within a defined region of interest (ROI), but is this the optimal choice? Should the average spectra from multiple ROI's be used? Or perhaps sample variation should be incorporated into the calibration model by including all spectra within a single ROI, mapped to a single Y value. This poster presents various sampling strategies used for pixel spectra selection to optimize PLS and PLS-DA hyperspectral image model building.

P17. The voltammetric identification of water solutions using PLS-DA

*A. V. Sidelnikov, D. M. Bikmееv, F. Kh. Kudasheva, V. N. Maistrenko
Bashkir State University, Ufa, Russia*

Today the developing of multisensor systems for identification of the nature of multicomponent mixtures is of a great interest. The main goal here is not to identify the concrete substance in a complex mixture but to find out does the mixture meet some requirements, e.g. standards. Such systems also allow to analyze some general quality properties of multicomponent solutions either in-line or on-line.

In the present work the combination of voltammetric system of tubular electrodes and projection on latent structure discriminant analysis (PLS-DA) were applied for identification and classification of investigated electrolyte solutions. The results of the investigations show, that the identification is possible for groups of solutions, but not for individual salts.

P18. Identification of multicomponent water solutions using voltammetric system of tubular electrodes

*A. V. Sidelnikov, D. M. Bikmееv, F. Kh. Kudasheva, V. N. Maistrenko
Bashkir State University, Ufa, Russia*

Multisensor systems together with methods for multivariate data analysis allow to solve quite difficult analytical problems, including express identification of multicomponent mixtures.

In the present work a particular implementation of such approach based on voltammetric system of tubular electrodes and principal component analysis (PCA) is proposed. The method has been applied for investigation of voltammetric

behavior of aromatic nitrocompounds in mineral water samples. Differential voltammograms of the reduction of nitrocompounds obtained at three different scan rates were used as a raw data. PCA was applied for exploratory data analysis. The identification and classification of samples was carried out using soft independent modeling of class analogy (SIMCA).

The results of classification shows, that the identification of investigated solutions with proposed method is possible for groups of mineral waters, and in some cases for individual samples.

P19. A simple method for estimation of optimal dimensionality in PLS models based on regression vector features

A. A. Gowen¹, G. Downey², C. Esquerre², C. P. O'Donnell¹

¹School of Agriculture, Veterinary Medicine and Food Science, UCD, Ireland

²Teagasc, Ashtown Food Research Centre, Ireland

PLS regression (PLSR) models are prone to overfitting through inclusion of redundant latent variables that model noise inherent in a given dataset. This can adversely affect model predictive ability with future datasets. Numerous methods have been proposed for prevention of overfitting in PLSR, including Wold's criterion, Monte Carlo cross-validation, smoothed partial least-squares regression [1] and the randomisation test [2]. One well-known sign of overfitting is the appearance of noise in regression vectors; this often takes the form of a reduction in apparent structure and the presence of sharp peaks with a high degree of directional oscillation, features which are currently estimated subjectively. We propose a simple method for objectively quantifying the shape of a regression vector; this measure can be combined with an indicator of model performance, such as root mean square error of cross-validation (RMSECV), to produce a new criterion for PLS model dimensionality estimation. The consistency of this new method is demonstrated on simulated and real datasets and compared with existing methods for estimation of optimal model dimensionality.

References:

1. S. Gourvenec, J. A. Fernandez Pierna, D. L. Massart, D. N. Rutledge, 2003. *Chemom. Intell. Lab.Syst.* **68**, 41-51.
2. M.P. Gomez-Carracedo, J.M. Andrade, D.N. Rutledge, N.M. Faber, 2007. *ACA*, **585**, 253-265.

P20. Regularities of phase transition temperatures within homologous series of organic compounds

A. N. Marinichev

Chemical Department, Saint-Petersburg State University, Russia

Formulas of neighbour members in homologous series are known to differ in the same fragment. For n -alkanes having formula C_nH_{2n+2} and increment $\Delta n = 1$, it is the so called homologous difference CH_2 . One of important properties of organic compounds is the normal boiling temperature (T). Its dependence on the number, n , of carbon atoms in the chain length is assumed to be given by the equation

$$T = C_0 + C_1 \exp(-k_1 n) + C_2 \exp(-k_2 n) \quad (1)$$

where the parameters C_0 , C_1 , C_2 , k_1 and k_2 are determined using the temperature data of series. This equation is a solution of the inhomogeneous linear difference equation of second order

$$T(n + 2\Delta n) + aT(n + \Delta n) + bT(n) + c = 0,$$

where the constant coefficients a , b , c are calculated as the least squares estimates defined on the basis of some restricted number of the experimental data. If z_1 and z_2 are the positive roots of the quadratic equation $z^2 + az + b = 0$, it follows that $k_1 = \ln(z_1)/\Delta n$, $k_2 = \ln(z_2)/\Delta n$. Additionally, we can use the obvious identity: $C_0 = c/(1 - a - b)$. In an effort to use equation (1) for interpolations we deal with the $T(n)$ data under condition that $\Delta n > 1$.

For the n -alkanes, precise experimental data are available [1], providing determination of coefficients a , b , c , as well as of the parameters k_1 and k_2 . In particular, for the n -alkanes with $n \in [3 \div 18]$ $\Delta n = 3$ we obtain $k_1 = 0.054 \pm 0.028$, $k_2 = 0.3556 \pm 0.0053$. The mean absolute difference between interpolated and experimental values of boiling temperatures does not exceed 0.061^0 with standard deviation $\pm 0.050^0$. An accuracy of the boiling temperatures of the n -alkanols ($n \in [4 \div 14]$ and $\Delta n = 2$), the alkancarboxylic acids ($n \in [4 \div 10]$ and $\Delta n = 2$) and the alkyl arenes ($n \in [6 \div 14]$ and $\Delta n = 2$) [2] is lower as compared to that of the n -alkanes. Nevertheless, numerical analysis of these series revealed that equation (1) is also suitable for approximation of the $T(n)$ data, the above-mentioned values of the parameters k_1 and k_2 being applied. In all these cases, the mean absolute difference between interpolated and experimental values of boiling temperatures does not exceed 1.0^0 with standard deviation $\pm 0.7^0$.

Results demonstrate, that the parameters k_1 and k_2 may be treated as latent invariants which are identical for the $T(n)$ data within the homologous series under consideration.

References:

1. Zwolinsky B.J., Wilhoit R.C. Handbook of vapor pressures and heats of vaporization of hydrocarbons and related compounds. Thermodynamics Research Center, Texas A&M University, Publication N101 (1971),.
2. Zenkevich I.G. *J. Gen. Chem.* (Russ.). 2006, **76**, 1821–1833.

P21. The study of the interaction between dimethyl formamide and water in binary mixtures by means of Raman spectroscopy and chemometric techniques

S.O. Paul

Department of Chemistry, University of South Africa, Pretoria, South Africa

In both liquid dimethyl formamide (DMF) and water, some self-association between the molecules takes place. When the two liquids are mixed, some of this association is expected to change, and instead some interaction between the two liquids is anticipated.

The interaction between DMF and water has been investigated by recording Raman spectra of DMF-water mixtures over the entire composition range. Interaction between DMF and water seems likely, because the maxima of several of the bands in the spectra shift as the composition of the mixtures is varied; the shapes of several of the bands change as well. Such a changes are particularly evident for the -C=O stretching vibration of DMF. The plot of the band maximum of this vibrational band versus mixture composition exhibits unusual features: first it shifts to a lower wavenumber, then to a higher wavenumber as the water component of the mixture increases. The exact positions of the band maxima have been determined by evaluating the second derivative of the Raman spectra.

The spectra of the binary mixtures have been analysed using chemometric techniques, in particular multivariate curve resolution, to validate the observations of the spectral shifts. The results which were obtained assisted in the elucidation of some of the unusual behaviour which has been observed, and some explanations will be presented.

P22. Evaluation of the number of factors needed for residual bilinearization in BLLS and UPLS models to achieve the second-order advantage

E. Sanina

Pharmaceutical Company ZAO GlaxoSmithKline Trading, Moscow, Russia

Bilinear least squares (BLLS) and unfold partial least squares (UPLS) are second-order multivariate calibration methods, which require the application of the residual bilinearization (RBL) algorithm to achieve the second-order advantage.

The poster presents a study of the choice of the number of RBL factors, in BLLS and UPLS models, for two different datasets based on treatment with anthracyclines or taxanes in the adjuvant or metastatic setting in subjects with breast cancer. Confidence limits for mean calibration residuals, based on a student-t distribution, are proposed as a criterion for determination of the number of RBL factors. Feasible results were obtained based on the proposed confidence limits, but divergences were observed in some situations in the dataset due to either differences in the models or characteristics of the used study assessments.

These results suggest, whenever possible, that the number of RBL factors should be checked with a dataset composed by selections where values of the property of interest are known from a reference method.

P23. Principal Component Analysis and Multi-way Analysis of tropospheric trace gases data in Lake Como area (Italy) during the period 1992–2007

Simone Roncoroni, Barbara Giussani, Alessio Beri, Andrea Pozzi

Department of Chemical and Environmental Sciences, University of Insubria, Como, Italy

The study of the troposphere represents one of the most important issues in today's atmospheric research as it is the region of the Earth's atmosphere in which we live and into which chemical compounds are generally emitted as a result of human activities. In particular, the monitoring of trace gases is of great concern as they can affect the health of humans and they can have damaging effects on flora and fauna. Trace gases lead to serious phenomena such as the acid rains, the photochemical smog, and the greenhouse effect as well.

In this work near-surface pollutants concentrations data in Lake Como district (Lombardy, northern Italy) from 1992 to 2007 were considered. Three sampling stations were used, two of them located in industrialized areas and the third located in a rural area.

The work shows how, from the same data matrix, it is possible to emphasize different information depending on the data handling performed on the array. Both bidimensional (unfold-PCA) and multidimensional methods were employed. Among them, the Tucker3 method was chosen due to its fitting flexibility. Bidimensional and multidimensional modeling were applied with the aim to compare the results and the benefits achieved using a more complex model that allows considering all the variability of the system simultaneously with respect to a simpler better-known one.

Through this modeling, a characterization of the Lake Como Area troposphere and its evolution in time was obtained. The models showed a distinction between the two "hot spot" stations of Como and Lecco, and the rural site of Varenna. The evidence of a tendency of pollution reduction and a change of contaminants during the investigated years clearly emerge from the modeling. The seasonal cycle with ozone peak in summer emerges as well, as expected. There is also the evidence of a week cycle with a reduction of pollutants during the week-end. The work demonstrated that multivariate techniques and in particular multiway analysis should be considered an essential and powerful device for environmental monitoring data treatment.

P24. Straight-line calibration when both variables are subject to error

Emese Vágó, István Kollár, Sándor Kemény

Budapest University of Technology and Economics, Budapest, Hungary

In analytical chemistry the straight line calibration is a general method for defining the connection between the concentration of a substance and an analytical signal. Usually a set of standards is prepared for the calibration (their concentration is known). The analysis of these standards produces a series of measurements. The data (i.e. the concentration of the standards vs. the analytical response) can be fit to a straight line, using linear regression analysis.

When doing linear regression the regressors assumed to be error-free. In case of calibration it means that the concentration of the standards is known without error. In practice this assumption is not always fulfilled. The two major problems that may arise concerning the standards: i) their concentration is subject to error ii) their errors are not independent. The typical example for the second case is, when the standards are made from the same solution with dilution.

The aim of our work was to estimate the calibration line when these errors occur.

P25. The oil treating process diagnostic system

E. V. Yakimova, I. V. Rudakova, L. A. Rusinov

Saint-Petersburg State Technological Institute (Technical University), Russia

When the oil emulsion ingresses, it contains sand, gas, water and other substances. Before the oil is pumped into commercial pipelines, it undergoes depuration through an oil treating process. The developed process monitoring and diagnostic system operates in the principal component space, using Q and T^2 statistics. For this purpose, the Principal Component Analysis (PCA) model has been constructed on the data of a normally running process. Then the statistic threshold values must be calculated, and the faults can be determined as exceeding the thresholds. It provides both incipient and abrupt faults detection in their early stage.

A two-level frame-based production diagnostic model that determines the fault's origin has been developed based on the expert data, on the results of the HAZOP-analysis and on the process decomposition. When an abnormal situation is detected, the system isolates the fault, using upper-level root frames of the model. It leads to the activation of corresponding daughter frame. This frame contains fuzzy production rules, which describe the real faults that can happen in a particular process unit. The fault is determined by comparing the similarity degree of a current situation with that, written in a rule. The similarity degree is defined, as an inner product of the respective vectors.

For the diagnostics purpose of such loaded devices as cut-off and control valves, whose faults can be masked by a control system, the following technique can be applied. The statistics and mathematical descriptions of basic faults in the work of electropneumatic positioners, which is the most widespread kind of valves, has been recently published (European project DAMADICS). Based on this data, fuzzy diagnostic models, describing the valve faults by fuzzy rules, such as Takagi-

Sigeno, have been developed. The faults can be connected to the measured variables, i.e. flow rate before and after the valve, pressure in a stream before and after the valve, output signal of controller and position of the valve rod. The models were developed, using traditional methods, including fuzzy clustering of the statistical data for all the faults that happen during the valve functioning.

A preliminary PCA model has been constructed in accordance with the normal process statistics. The Q and T² threshold values were then defined and a fuzzy model of the valve faults was specified. The actual work of the system begins from the monitoring in the space of the principal components, using Q and T² statistics. In case of a fault, the expert system controlling the process, isolates the fault and identifies its origin. If the valves are the cause of the fault, the fuzzy valve fault models are activated and the proximity degree of their outputs to the current process state is examined. It helps to improve the quality of diagnostics.

P26. KNIME extension for building multivariate regression under interval error

Sergei I. Zhilin

Altai State University, Barnaul, Russia

KNIME [1] is a modular open source data analytics platform that enables the user to visually create data workflows, selectively execute some or all analysis steps, and investigate the results through interactive views on data and models. Because of open interfaces KNIME allows easy integration of different data loading, processing, transformation, analysis and visual exploration modules without the focus on any particular application area. Such popular tools as WEKA (machine learning environment), R Project (statistical programming language), Python (scripting language), ImageJ (image processing and analysis program) can be integrated in KNIME data processing and analysis workflows. A standard functionality of KNIME can be extended with user-defined plugins. A great number of additional workflow nodes are developed by user community as well as by commercial vendors for chemical applications, image and text analysis, data mining, etc.

We provide a set of KNIME nodes intended for building and analyzing multivariate regression under interval error. Applying these nodes a user can detect outliers, construct outer and inner interval estimations of regression parameters,

make interval predictions, and use techniques of Simple Interval Calculations (SIC) [2].

Acknowledgements:

This research work was supported by departmental analytical program «Development of Scientific Potential of Higher Education» №2.2.2.4/4278.

References:

1. KNIME - Konstanz Information Miner, <http://www.knime.org>.
2. Rodionova O.Ye., Pomerantsev A.L. *Chem. Intell. Lab. Syst.* **97**, 2009, 64-74.

P27. Chemical and isotopic tracers in copper deposits and ancient artefacts: a multivariate analysis approach to establish the provenance of materials

Barbara Giussani¹, Marcello Marelli¹, Sandro Recchia¹, Simone Roncoroni¹, Ivana Angelini², Gilberto Artioli², Ilaria Giunti²

¹*Dipartimento di Scienze Chimiche e Ambientali, Università dell'Insubria, Como, Italy*

²*Dipartimento di Geoscienze, Università di Padova, Padova, Italy*

The study of the provenance of ancient copper objects on a scientific basis embraces a number of challenging issues that can be summarised in the following question: is it possible to link a copper object to the minerals that were used to produce it? In this respect this work reflects one of the most important problem related to this question, namely to find out if Cu-minerals coming from different historical Cu-ores can be traced as a function of their provenance, giving information about the origin of the metal used by the chalcolithic metalworkers.

To aid metal provenance studies, a database of fully characterized Alpine copper mineralization is being developed as the fundamental reference frame for metal extraction and diffusion in the past. In the early stages of the project, some of the most well known copper deposits in the Western Alps were selected and compared with very different minerogenetic deposits from the French Queyras (Saint Veran) and the Ligurian Apennines (Libiola, Monte Loreto).

The fully characterized samples were then analysed by ICP-QMS (Inductively Coupled Plasma-Quadrupole Mass Spectrometry). The abundances of about 60 minor and trace elements, including the rare earths, were measured in all samples.

Furthermore, the feasibility of the routine reliable measurement of the $^{65}\text{Cu}/^{63}\text{Cu}$ isotope ratio and its eventual use as a possible ore tracer was tested.

Principal Component Analysis (PCA) was initially employed for this purpose. However, the PCA models were able to separate only a few numbers of different Cu-ores according to their geochemical and isotopic features. A complete description of all the investigated Cu-ores was obtained using a different approach which is based on Partial Least Squares Regression – Discriminant Analysis (PLS-DA). The advantage of this method is related to the regression coefficients that the calculation provides. They allow establishing which variables mainly characterize a Cu-ore with respect to the others. A variable selection was thus performed both using PLS-DA, and studying the correlation loadings calculated from the PCA models.

As we will show, the combination of both methods, PCA and PLS-DA, allows separating the investigated Cu-ores efficiently. Finally, applications of these models to ancient copper objects provenance determination will be also treated.

P28. Identification of unknowns based on structure data bases and computer simulation of mass-spectra

*I.A. Revelsky, I.V. Gulyaev, A.I. Revelsky, D.A. Chepelyansky
Chemical Department, Moscow State University, Moscow, Russia*

Database searches based on EI-MS, where the unknown spectrum is compared with spectra of known compounds in database, are well-established. EI spectra are relatively reproducible, so finding a matching spectrum is often a relatively good indication of the compound structure. NIST and Wiley mass-spectra databases are widely used for these searches. But they are applicable only for identification of substances contained in these databases. There are about 220 000 for NIST and 310 000 substances for Wiley that gives hardly any great opportunities for identification of unknowns in real samples.

Structure databases (e.g. PubChem, Merck Index, ChemIndex) contain much more compounds than mass-spectra databases (for example PubChem Database contains more than 6 000 000 structures). Using the chemical formula as an input for a compound database search is becoming a viable tool to provide indications or tentative identifications for unknown compounds. In this approach chemical formula predicted from GC-AED or HR-MS data is used for separation of possible

structure candidates that then can be matched due to spectral classifiers and spectral data, so unknown compounds can be determined.

In this research three compounds (afobazol and its metabolites M-3 and M-11) were used to simulate unknown substances to be determined by structure databases search. It was supposed that their chemical formulas ($C_{15}H_{21}N_3O_2S$, $C_{13}H_{17}N_3O_2S$, $C_{15}H_{19}N_3O_3S$) were determined by GC-AED. Then all structures possessing these formulas were separated from ChemExper database (in average 138 possible structure candidates per each unknown compound). Next prediction of spectra of structure candidates and comparison with real EI-MS spectrum of unknown compound were used to choose the most appropriate structures for each unknown compound. Computer program Mass Frontier 3.0 was applied to predict EI-mass-spectra of structure candidates using their structural formula and general fragmentation rules common for all candidates. After that these generated spectra were compared with real mass-spectra of unknown compound, match value (MV) for every structure candidate was calculated, and candidates were ranged. NIST MS Search 2.0 software was used for mass-spectra comparison and calculation of MV. It is based on dot product calculation and used to be applied for mass-spectra database search. The candidates with the highest MV are to be the most probable structures of unknown compounds.

It was found that right candidate structures possessed the highest MV in all three cases, so they were isolated from wrong candidate structures, and structure of unknown compounds were determined unequally.

Proposed approach includes only available in most laboratories and relatively non-cost techniques GC-AED and GC-MS-EI. Furthermore these techniques can be applied for identification of impurities in complex mixtures without their isolation and concentrating (usually most time-consuming steps) and without using of other techniques such as IR and NMR spectroscopy. So it provides fast and effective tentative identification.

P29. Prediction of problematic wine fermentations using Artificial Neural Networks

Alejandra Urtubia¹, Alik Abakarov¹, Gonzalo Hernández², César Román²

¹Department of Chemical and Environmental Engineering, Universidad Técnica Federico Santa María, Valparaíso, Chile

²Ingeniería Civil Industrial. Facultad de Ciencias Económicas y Administrativas, Universidad de Valparaíso

The problems in wine fermentations have a direct impact on wine quality. Early detection of problematic fermentations would be very useful for the winemakers, because they could take the corrective and appropriate actions in order to guarantee the success of winemaking.

Artificial Neural Networks (ANNs) have been used for recognition of non linear pattern, which is characteristic of bioprocess like wine production. In addition, it is a very useful tool for fault detection and process control. ANNs are capable to detect complex relationships between the inputs and outputs and they can reproduce the behavior of a system after of previous training with known data.

In order to predict problems of wine fermentation ANNs was tested. A database of 20,000 data (sugars, alcohol, organic acids and nitrogen compounds) from 22 fermentations of *Cabernet Sauvignon* was used. Of 22 fermentations, 9 were normal and 13 were problematic fermentations.

A model multilayer-perceptron was developed. The predictor variables were total sugar, alcohol, glycerol, acetic acid, lactic acid, malic acid and tartaric acid. In order to process the information, one hidden layer and neurons between 5 and 10 were used. Hyperbolic tangent function was the communication medium, and one output was used with Softmax function, which represents to dependent variables (1: normal fermentations; 0: problematic fermentations).

Over 70% of data were well classified using the neural architecture developed and trained by optimization algorithm Gradient Descent. This represents an alternative model non parametric for prediction of problem fermentations. Using this model, the measurements of the first three days are enough to detect over 70% of problematic wine fermentations.

Acknowledgements:

A. Urtubia thanks the financial support from FONDECYT Project (No 11070251).

P30. Internal standard calculations for non-linear detectors

Yuri A. Kalambet¹, Sergey Maltsev¹, Yuri P. Kozmin²

¹*Ampersand Ltd., Moscow, Russia*

²*Shemyakin Institute of Bioorganic Chemistry, Moscow, Russia*

Internal Standard (ISTD) is a well-known chromatographic technique, where known amount of a component, called internal standard is added to both standard and unknown samples. The classic Internal Standard quantification method plots the response ratio (analyte to internal standard) versus amount ratio (again analyte to internal standard). Internal standard component itself does not have any calibration curve. Quantification procedure uses this plot to get concentration ratio from response ratio.

This approach may cause systematic errors in the case of non-linear (or linear, not going through origin) detector response to concentration of internal standard. Typically it is recommended to keep amount of internal standard in this case precisely constant.

Offered calculation scheme allows wide variations of standard and analyte concentrations. In the case of non-directly proportional dependencies it requires that External standard dependencies of both internal standard component and analyte are measured from time to time. Internal standard calculations are split into two parts:

1. Calculation of Relative concentration, i.e. concentration of analyte, provided concentration of Internal Standard is known.
2. Improvement of calibration curves of the analytes.

The calculation scheme can be further extended to the case of linear dependencies and can successfully replace the method with response ratio calculations.

The described calculation scheme is successfully implemented for Internal Standard calculations in commercial chromatographic software.

P31. Reconstruction of out-of-range peaks using Exponentially Modified Gaussian peak shape

Yuri Kalambet¹, Ksenie Mikhailova¹, Yuri P. Kozmin², Igor Nagaev³, Pavel Tikhonov⁴

¹*Ampersand Ltd., Moscow, Russia*

²*Shemyakin Institute of Bioorganic Chemistry, Moscow, Russia*

³*Institute of Molecular Genetics, RAS, Moscow, Russia*

⁴*The Russian Research Institute of Metrological Service, Moscow, Russia*

Exponentially Modified Gaussian (EMG) peak shape [1] is widely used for peak approximation in chromatography. We constructed EMG peak deconvolution routine for chromatography, using combination of two EMG formulas [1, 2] and linear optimization methods. This routine accounts for maximum linear range of the detector and can work with out-of range peaks.

Optimization routine is applied to reconstruction of out-of-range peaks, so that analyst can get an idea about area/height and concentration of such peaks. We found, that in many cases such reconstruction provides reasonable prediction error. This information helps in reducing number of chromatographic runs while method development and routine work. A possibility of reconstructing out-of-range peaks using pre-defined peak shape, obtained while calibration, is also discussed.

References:

1. McWilliam, I. G.; Bolton, H. C. *Anal. Chem.* 1969, **41**, 1755–1762.
2. Delley, R. *Anal. Chem.* 1985, **57**, 388

P32. Self-modelling factor analysis combined with analysis of the signal derivative for separation of incompletely separated chromatographic peaks

Yuri A. Kalambet¹, Sergey Maltsev¹, Yuri P. Kozmin²

¹*Ampersand Ltd., Moscow, Russia*

²*Shemyakin Institute of Bioorganic Chemistry, Moscow, Russia*

Factor analysis is widely used for separation of incompletely separated chromatographic peaks. We combined this technique with analysis of the angle derivative of the multi-channel chromatogram to get a better choice of basic spectra. Improvement in algorithm is demonstrated by examples.

P33. Confidence interval for weighted polynomial calibrations

*Sergey Maltsev, Yuri A. Kalambet
Ampersand Ltd., Moscow, Russia*

In many chromatographic software weighted polynomial regression is used for calibration curves. Weighting is a useful mean to account for measurement error that may depend on the detector response value. Confidence intervals show, how accurate is a measurement with the help of a calibration curve. Besides, theory of confidence intervals was not extended to weighted polynomial regressions.

We extended confidence interval theory to the frequent case of weights, expressed as a function of Y value, in particular $1/Y$ and $1/Y^2$. This extension allows accurate calculation of concentrations with confidence intervals for calibration curves, constructed using point weighting. Examples are shown that demonstrate applications of a weighted calibration curves with confidence intervals in chromatography.

P34. Renaissance of colorimetry: scientometrical study

*Bulianitsa A.L., Shcherbakov A.P., Belenkii B.G., Arkhipov D.B.
Institute of Analytical Instrumentation RAS, Saint-Petersburg, Russia*

Modern scientometrics borders with chemometrics and bibliographical statistics [1]. Scientometricists measure the number of articles and construct time series. However, mathematical apparatus of time series can not explain attractors and bifurcations that have place in trends of analytical instruments.

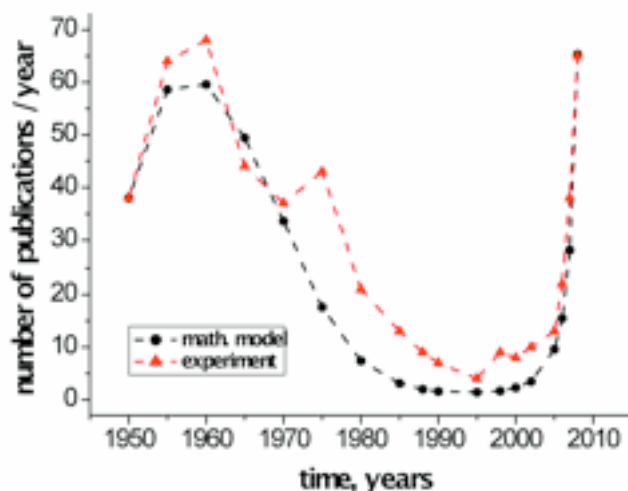


Figure shows dynamics of yearly number of publications in the journal Analytical Chemistry published by American Chemical Society. One can easily see stages of increase, decrease and new increase that we named as the stage of "second birth".

The phenomenon of "second birth" can be explained by equation, which based on classic logistic growth model [2]. The main modification is connected with including basic parameters as linear decreasing functions, i.e. variant values. This equation is presented as $dy/dt = k_0(1 - at)y(N_0(1 - bt) - y)$; $y(0) = n$, where n number of publications (articles) at the first year after birthing colorimetry as scientific direction, t — time, a , k_0 , N_0 and b — basic parameters of mathematic model. The estimation of this parameters values was the general problem. We used traditional numerical iterations method (co-ordinate or spiral descent) for searching optimum of quality functional. But search of point of first iteration is based on solution the system of four single (but unlinear) equations. For comparison, we present the first iteration of parameters values $a^* = 0.122$; $b^* = 0.022$; $N_0^* = 97.1$; $k_0^* = 0.0134$ and the final solution: $a = 0.1280$, $b = 0.02242$, $N_0 = 120.0$, $k_0 = 0.0020$

Another examples of the "second birth" stage in analytical instrumentation trends are time-of-flight mass spectrometers, paper chromatography and, perhaps, potentiometry. But the most types of analytical instruments don't have the stage of "second birth" (or "death of scientific direction").

References:

1. Malinetskii G.G., Kurdyumov S.P. *Herald of RAS*. 2001. **71**, 210–224.
2. Arkhipov D.B., Bulianitsa A.L., Scherbakov A.P. Simulation of Trends of Analytical Instrumentation Based o Logistic Equation with Coefficients Depending on Time Proc. of 3-d All-Russian Conf. "Analytical Instruments". St. Petersburg. 2008. P. 20–21.

P35. Determining of general water hardness by digital image-based titration

E.A. Cherednik, A.G. Zarubin

Faculty of Chemistry, Tomsk State University, Tomsk, Russia

Water quality is now a significant problem. Therefore, it is important to work out a technique for determination of natural waters general hardness. This method aims

at obtaining of the more accurate values of hardness and at automatization the of the water quality monitoring process.

We suggest using a digital indicator that is the value of red color channel of photographic image of an analyzed solution for determining of the general water hardness.

At complexon III titrations of the model solutions ($n = 5$, $P = 0.95$, pH 9, indicator is eriochrome T black) with MgSO_4 concentration equals 0.05M and 0.1M, the following results have been obtained: $C(\text{Mg}^{2+}) = (0.050 \pm 0.002)\text{M}$, $s_r = 0.025$ и $C(\text{Mg}^{2+}) = (0.100 \pm 0.003)\text{M}$, $s_r = 0.002$, accordingly.

The proposed method let us to determine the cold water-pipe general hardness. At titration of 40 ml solutions of analyte ($n = 5$, $P = 0.95$) it gives (6.3 ± 0.1) mmol/l, $s_r = 0.015$.

P36. Improved process understanding of solid dosage formulations through spectroscopy and multivariate data analysis

Niels Peter Aae Christensen¹, Claus Cornett¹, Jukka Rantanen¹, Poul Bertelsen² and Søren Stig Nielsen²

¹Department of Pharmaceutics and Analytical Chemistry, University of Copenhagen, Denmark

²Nycomed, Pharmaceutical Development, Roskilde, Denmark

There is a growing recognition of the need for a thorough understanding of the solid state properties of pharmaceuticals. A complex interplay exists between both the active pharmaceutical ingredients' and excipients' physical and chemical properties and the related unit operations. One important unit operation is wet granulation where the physical characteristics of a powder are optimized for tablet compression. During wet granulation the powder is exposed to solvent, shear and changes in temperature. These conditions make wet granulation likely to induce phase transitions in the solid state. These phase transitions may have a great impact on the quality and properties of the final pharmaceutical product.

In this study the solid state of an active pharmaceutical ingredient during wet granulation is investigated. An elucidation of the physicochemical phenomena taking place during processing is approached by process analytical technologies and off-line testing using near-infrared spectroscopy, Raman spectroscopy,

infrared spectroscopy, x-ray powder diffraction, thermal methods and multivariate data analysis.

P37. Processing of mass-spectrometer's signals utilizing integrated spectral analysis in adaptive basis

V. Manoilov, I. Zarutskiy

Institute of Analytical Instrumentation RAS, St. Petersburg, Russia

The new mathematical method for processing is proposed. The method is based on factorization the source data in an orthogonal basis adapted to the standard mass-spectra data. Orthogonal transformations can be represented in the form of matrix equation $Y = HX$ (1), where X — initial vector of source data, H — operator of the orthogonal transformation, Y — vector of data after factorization. Theoretical justification of such transformations has been developed in works made by prof. Alexey Solodovnikov and his successors [1]. The system of adapted basis functions can be build on the base of a standard R_{st} . An operator of orthogonal transform H is considered to be adapted if the one meets the condition: $H R_{st} = Y_t = [y_{t1,0}, \dots, 0]$ (2), Y_t — target vector, taken as operator adjustment criterion, $R_{st} = [r_{st,1}, \dots, r_{st,N}]$ — the vector of standard mass spectrum. Standard mass spectrum is ideal mass spectrum of the pure product without noise, without contaminations and with peaks of well known shape. Standard mass spectrum (line 1), standard mass spectrum with contaminations (line 2) and mass spectrum of contaminations (line 3) are shown on the Fig.1. In Fig.1 the axis of abscissas is mass.

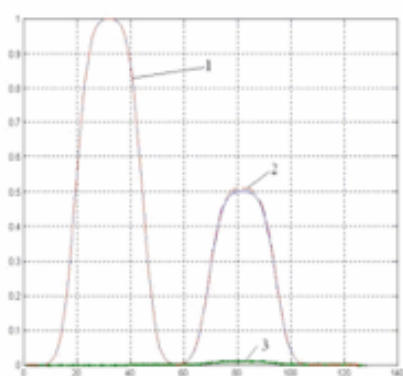


Fig. 1



Fig. 2

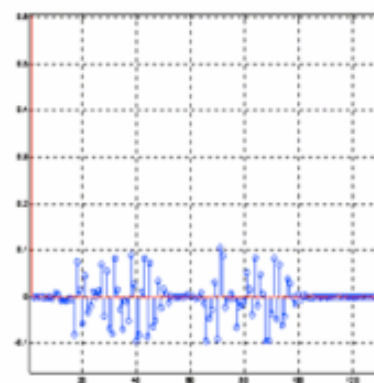


Fig. 3

The presentation of factorization of standard mass-spectrum in adapted basis function according to equation (2) is shown in Fig. 2. The presentation of the factorization of the standard mass spectrum with contaminations in adapted basis

function according to equation (1) is shown in Fig 3. In the Fig. 2 and Fig.3 the axis of abscissas is spectral coefficient k .

One can distinguish pure mass spectrum from mass spectrum with contaminations by comparing the transformed signals according to equation (1). Those comparison can be done by testing the inequalities

$$\begin{cases} \|Y - Y_{st}\| \leq \delta \\ |Y(1)^2 - Y_{st}(1)^2| \leq \varepsilon \end{cases} \quad (3)$$

Y — vector of source data transformed according to equation (1) Y_{st} — vector of standard data transformed according to equation (2). $Y - Y_{st}$ — norm of difference of two vectors. $Y(1)$, $Y_{st}(1)$ — elements of vectors for spectral coefficient $k = 1$. If inequalities (3) are true, then our source data are pure mass spectrum, else mass spectrum has contaminations. Parameters δ and ε are chosen in learning process considering the shape of peaks and the variance of the noise.

Main advantages of proposed method consist in the independence from shape of mass spectrum peak, in simplicity of achievement, and possibility to detect multiplets in mass spectrum data.

References:

1. Solodovnikov. A.I. Spivacovsky A.M. Fundamentals of theory and methods of spectral processing information. School-book Publisher Leningrad University, 1986. 272 pages (in russian)

P38. Chemometric investigation of polarization curves

John R. Quagliano and Christopher A. Marks
Richmond, USA

Polarization curves are measurements of net current versus potential for an electrochemical system. The measured current is the sum of currents from several anodic (+) and cathodic (–) reactions, which can not be measured independently. It is hoped that chemometric models can be used to determine important characteristics of the underlying reactions. Several difficulties due to the nature of polarization data have been encountered. The domain of potentials for which the current can be measured depends on the variables under study, leading to non-random missing values at the extreme values of potential. The uncertainty in the measured current depends on the current, not the potential. In this work, missing

values have been replaced by their estimates as the model is being generated, and iteratively re-weighted least squares has been used to overcome the latter problem. There is also noise in the potential (independent variable), including random missing values. A Savitzky-Golay filter has been used to interpolate the missing values and smooth the data prior to analysis.

P39. Comparative assessment of different chemometrics techniques in multicomponent determination of metals without preliminary separation

Yu. B. Monakhova, S. P. Mushtakova

Department of Chemistry, Saratov State University, Saratov, Russia

Quantitative spectroscopic methods have given birth to a number of widely used analytical tools, yet being a subject of active research and development. In spite of numerous advantages, some difficulties still exist. These include, but are not restricted to broad and highly overlapping bands, typical for multicomponent mixtures. These issues present particular difficulties for exploratory studies and the analysis of unknown systems.

Chemometrics and closely related approaches such as independent component analysis (ICA) can offer a solution in this situation. One branch of the research in this field focuses on the development and practical application of algorithms aimed at the decomposition of experimental spectra X of a multicomponent mixture onto a matrix of concentration profiles A and a spectral profile matrix S . Methods of formal modeling that does not require any prior knowledge about the system under study (its composition, concentration range, etc.) are employed in such an analysis.

We have proposed a simple, fast and reliable method of spectroscopic determination of metals in complex multicomponent mixtures. The analysis is based on decomposition of UV-VIS spectra of EDTA-metal complexes by means of chemometrics. We have compared the performance of Alternating Least Squares (ALS) with that of ICA methods (MILCA, SNICA, SIMPLISMA, RADICAL, FASTICA and JADE) in the identification and quantitative analysis of metals (Ca, Zn, Mg, Co, Cu and Mn) in complex mixtures. The results are presented as a series of experimental case studies, including spectroscopic analysis of metal model mixtures as well as real objects (multivitamin drugs, juices, etc.) in the UV-VIS region.

The results have shown that MILCA and MCR-ALS algorithms outperform other methods under study and are capable of providing robust results in well-defined accuracy ranges. By our estimate, the analytical errors in recovered concentrations are at the level of several percent, while the localization error of peak positions is comparable to instrumental uncertainties (below 1 nm for the techniques used).

P40. Statistical data evaluation for trace element analysis in nuclear safeguards

K. Zhao, M. Penkin, M. Ryzhinskiy

International Atomic Energy Agency, Vienna, Austria

Trace element analysis is now receiving more and more attention in strengthened safeguards to verify the completeness of the State declaration with respect to the origin and production of Uranium materials. For the trace element analysis, samples are usually taken from the feed, intermediate, and product materials at Uranium concentration and/or conversion plants, the trace element compositions are analyzed using advanced techniques such as ICP-MS, and the analytical results are evaluated to assess the consistency with the declared origin of the given Uranium materials.

Statistical data evaluation is one of the most important steps in trace element analysis. Two sequential modules, detection and classification, are applied for the evaluation. The detection module is to determine whether a new sample is significantly different from the historical dataset for the samples collected from the given Uranium material at a facility. If an abnormal sample is detected, the classification module will then be triggered to determine which material in the available reference database has the highest probability to belong to.

Linear techniques have been pursued for the statistical data evaluation. Principal Component Analysis (PCA) is used as a dimensionality reduction technique in the detection module. PCA is optimal in dimensionality reduction in terms of capturing the variance of the data while the correlations among trace elements are taken into account. The decision making for detection is based on T^2 statistics to determine whether a new sample is out-of-control in terms of systematic variation. In the classification module, Partial Least Square (PLS) is used as a modelling tool to predict the class memberships of a new sample. By using data decomposition method for maximizing the covariance between the input and the output, the PLS model developed based on historical dataset is able to capture the relationship

between the trace element compositions and the class memberships for different origins.

The performance of the procedure described above is demonstrated using real data. In addition, special difficulties encountered in the implementation are also discussed, which include statistical treatment of missing values, results reported as less than detection limits, and removal of susceptible results due to poor quality in the sampling or measurement process. It is recommended that new statistical methods be studied such that uncertainties of trace element results reported by analytical laboratories can be effectively used in the statistical evaluation.

P41. Intelligent monitoring subsystem of information security with mandatory access control

Gorkovenko E. V.¹, Sokolova S. P.²

¹Institute of Problem Informatics and Control MES RK, Almaty, Republic of Kazakhstan

²St. Petersburg Institute for Informatics and Automation RAS, St. Petersburg, Russia

Information Security policies for the information systems with confidential information are based on the mandatory access control. It constrains the rights of a subject to access the objects in the system. To refer to the information at the appropriate level of confidentiality a subject must have an official permission (mandate).

The system with such policy of information security contains the following subsystems:

- mandatory access differentiation;
- intelligent information security monitoring with mandatory access control.

Subsystem of mandate access differentiation protects the multilevel security from the unauthorized access. The construction procedure of this subsystem has been presented in [1].

Intelligent information security monitoring subsystem aims at decision of the following tasks: authentication of the security subjects; classification and clustering of the security objects; analysis of the properties of the restriction rights

access matrix. This subsystem contains the following modules: learning and self-learning, pattern recognition, risk index generation, interpretation of the results. Mathematical models and calculation procedures of Immunocomputing and PCA have been realized in the above-listed modules.

The efficiency of the proposed intelligent subsystems is demonstrated on the example of the multilevel delimitation access to thematic information resources in the satellite information-telecommunication system in the Republic of Kazakhstan.

References:

1. Sokolova S.P., Gorkovenko E.V. The system with mandate policy of information security. Proc. of XI St. Petersburg Interregional conf. "Regional Informatics — 2008". St. Petersburg, October 22–24 2008. PP. 159–163.

P42. Application of chemometrics method and constant stoichiometry groupings concept to the Raman spectra of lead metaphosphate — TeO_2 glasses

Inna E. Marusina¹, Olga Yu. Derkacheva¹, Vladimir K. Goncharuk², Irina G. Maslennikova², Irina G. Polyakova³, Anatolii V. Anan'ev⁴, Leonid V. Maksimov⁴, Tatiana S. Markova¹, Oleg V. Yanush¹

¹*St. Petersburg State Technological University of Plant Polymers, Russia*

²*Institute of Chemistry of Far-Eastern Branch of Russian Academy of Sciences, Vladivostok, Russia*

³*Institute of Silicate Chemistry, Russian Academy of Sciences, St. Petersburg, Russia*

⁴*Research and Technological Institute for Optical Material Science, St. Petersburg, Russia*

Lead metaphosphate — TeO_2 glasses in the whole range of glass content were first obtained and their properties (refractive index, density, glass transition temperature and light scattering losses and others) were determined. Based on the vibrational spectroscopy data a new approach has been proposed to investigate the interactions of initial oxides in melts resulted in so-called constant stoichiometry groupings (CSG) formation. Vibrational spectra of glasses are interpreted as a superposition of relatively small number of unchangeable spectral forms belonging to CSGs. For lead metaphosphate — TeO_2 glass series, CSGs extracted from Raman spectra analysis are $\text{PbO} \times \text{P}_2\text{O}_5$, $\text{TeO}_2 \times 2\text{PbO} \times 2\text{P}_2\text{O}_5$, $\text{TeO}_2 \times \text{PbO} \times \text{P}_2\text{O}_5$, $2\text{TeO}_2 \times \text{PbO} \times \text{P}_2\text{O}_5$ and TeO_2 .

In this work MVDA has been applied as the independent mathematical tool to decompose Raman spectra of glasses and reveal the number of PSCs. It is shown that application of factor analysis results in five PSCs that confirms our data obtained from the concept of CSGs. These principal spectral components were selected taking into account dependences of ERV parameter and eigenvalues on the number of PSCs.

The composition dependences of the concentrations of the CSGs have afforded an opportunity to explain and calculate the compositional dependences of refractive index and density of glasses in a good agreement with experimental data and even to predict refractive index and density values for vitreous TeO_2 and $\text{PbO} \times \text{P}_2\text{O}_5$. The Raman spectra analysis developed allowed also predicting of the existence of unknown compounds. The CSGs $\text{TeO}_2 \times 2\text{PbO} \times 2\text{P}_2\text{O}_5$, $\text{TeO}_2 \times \text{PbO} \times \text{P}_2\text{O}_5$, $2\text{TeO}_2 \times \text{PbO} \times \text{P}_2\text{O}_5$ discovered using Raman spectroscopy data exclusively has been verified by X-ray powder diffraction of crystallized glasses.

The concept of CSG opens the way to elaboration of low scattering glasses as candidates for Raman amplifiers. It is shown that glasses of lead metaphosphate - TeO_2 series containing small additive of tellurium dioxide are of interest to photonics technology. These glasses are uncolored, weather-resistant, the level of light scattering losses of glasses is comparable with that in fused silica. The minimal optical losses give us reasons to consider these glasses as an efficient active medium for Raman fiber lasers and amplifiers. In the case of using excitation laser source at 1480 nm the Raman amplifier range will correspond to wavelength between 1500 and 1790 nm.

P43. Using infrared spectroscopy and principal component regression for study of the bleaching and ageing influence on paper brightness

A. V. Dynina, O. Yu. Derkacheva

St. Petersburg State Technological University of Plant Polymers, Russia

The aim of the study was to evaluate using FTIR-spectroscopy the changes in bleached cotton, flax and kraft pulp papers after aqueous light bleaching and accelerated light ageing and to reveal influence some absorbance bands on paper brightness.

To solve this task the principal component regression (PCR) was applied to two sample sets. First set included only papers from cotton fibers and second one consisted of samples produced from different fibers. The FTIR spectra of the samples in the range 1460–1840 cm^{-1} were used as predictors.

Principal component analysis allowed to decompose the spectra and to get some interesting information about absorbance bands in the frequency range of interest. Thus for the paper samples from cotton fibers it was found out that intensities of the bands near 1512, 1540, 1577, 1590, 1656, 1690 and 1740 cm^{-1} are quite different for treated and untreated samples. For the cotton, flax and kraft pulp papers the treatments resulted in changing of the same spectral bands with addition of the band near 1480 cm^{-1} . Some correlations in behavior of the bands values and changes in content of functional groups were also revealed.

Then the PCR was used to predict the paper brightness. Good models, with correlation between predicted and measured values about 0.95, were obtained for the both sample sets. It is worthy to notice that if in the first sample set only bands from functional groups vibrations gave main input to the paper brightness, for the second one a band connected with vibration of cellulose macromolecules situated in ordered ranges also had a strong influence.

Applying PCR to FTIR-spectroscopic data allowed to clear changing spectral curves in the range of C=O groups vibration at bleaching and ageing and to calibrate the models for prediction of paper brightness. Such approach has a real perspective for its application in paper restoration studies.

P44. Mathematical models and frequency histograms as a tool to study the structure of geochemical fields

Romanov S.¹, Korobova E.²

¹*UE "Geoinformation Systems", Minsk, Belarus*

²*Vernadsky Institute of Geochemistry and Analytical Chemistry, Moscow, Russia*

Spatial distribution of chemical elements presented on geochemical maps frequently exhibits a polycentric character of its structure. Such a structure opens up possibilities for mathematical reproduction of its pattern. In our studies a simple model of the normal cone was used to model the structure of Cs-137 digital field. Our research has revealed that it can be used to analyze and reproduce the digital geochemical fields of mono- or polycentric structure of different complexity.

Analysis of the obtained frequency histograms has shown that its form could reflect specificity of the spatial distribution of radiocesium concentration in monocentric anomaly and is even related to its parameters. The maximum frequency is divisible by the area of anomaly while the maximum value presents the maximum possible concentration of this anomaly. Results of modeling different cases of superposition of anomalies of different size and intensity were considered. It was shown that a comparative analysis of conic models and frequency histograms allowed describing and explaining of the shape of frequency histograms. It was demonstrated that a lognormal character of histograms that were frequently obtained in geochemical studies could have only shown the sampling within fragments of the monocentric or polycentric anomaly.

P45. Using the neuron network based classification technique for modeling the nuclear power plant water chemistry

Kritsky V.G., Zakharova S.V., Nikolaev F.V.

JSC "Leading institute "VNIPIET", Saint-Petersburg, Russia

Generation of effective functional control of water chemistry at Nuclear Power Plant (NPP) is one of the guidelines of NPP safeguarding. This problem can be solved by means of water chemistry automatic monitoring system. This system provides the on-line monitoring of quality indexes. Analyzing and estimating the water chemistry parameters is the prerogative of the operator as an expert of the system.

Formalization and automation of these functions can be realized only by comprehensive approach. This approach supposes the use of the statistical information and expert experience for diagnosing and forecasting.

Large multidimensional data arrays of quality indexes referring to different operating conditions of NPP allow to analyze complex relations between different factors. For analyzing these relations it's necessary to carry out the decomposition of all data volume to the subsets with similarity of system factors and characteristics.

Classification and modeling of water chemistry was realized by Kohonen neuron network. In general, this network is an algorithm, based on estimating of proximity of classification objects.

The facilities of identification of water chemistry conditions by neuron network were verified by algorithms described in normative documents and by the data of the diagnostic system based on expert knowledge.

P46. Comparative analysis of chemical compositions of *Artemisia* L., growing in Central Asia

S.V. Zhigzhitzhapova¹, T.E. Soktoeva¹, L.D. Radnaeva^{1,2}, O. Grahl-Nilsen³

¹Buryat state university, Ulan-Ude, Russia

²Baikal institute of nature management SB RAS, Ulan-Ude, Russia

³Bergen university, Bergen, Norway

In the present work data on chemical composition of essential oils of plants of genus *Artemisia* L., growing in the Central Asia.

Artemisia L. as proved by many researchers, is rich in essential oils, which can be used in folk and official medicine as an antibacterial, anti-inflammatory, antiulcer, antiviral, antispasmodic, and other means [1–4].

The material for analysis has been collected in the second and third decades of July and in the first decade of August, 2008–2009, in different districts of Buryatiya, Irkutsk region and Mongolia. Essential oil received from an above-ground part of a plant with hydrodistillation from air-dry mass, in a year of the collecting of raw materials [5].

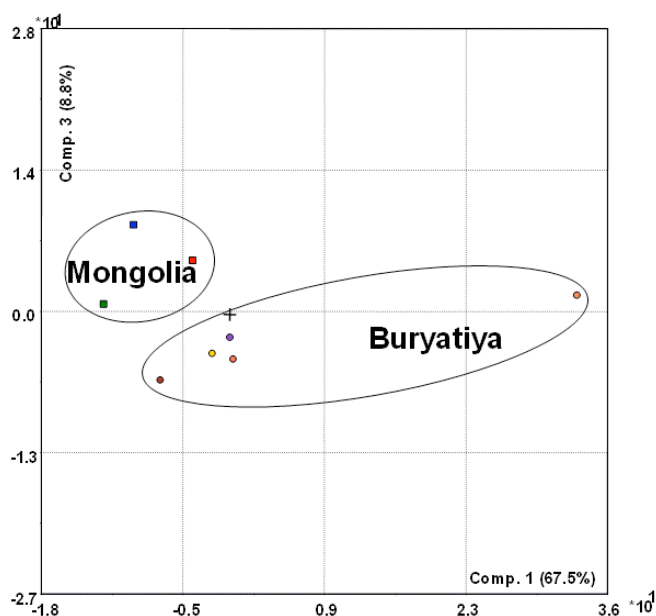


Figure 1. Dependence chemical composition of essential oils of *Artemisia sieversiana* and *Artemisia macrocephala* from districts growth

Chemical composition of essential oils had been researched by chromatomass spectrometry.

Application of PCA has allowed to show distinctions of chemical composition of essential oils and microelements of genus *Artemisia* L. growing in the Central Asia.

All samples of essential oils investigated by us can

be divided into three groups. Species of the wormwood concern the first group, a dominating which component is chamazulen (*A.sieversiana*, *A.macrocephala*), to the second group species, with prevalence as a part of acetylene hydrocarbons (benzylodiacetyllene, capillene and capilline), to them concern *A. glauca*, *A. commutata*, *A. dolosa*. 1,8-cineole, camphor, germacrene D, caryophyllene and spathulenol occupy considerable quantity in the composition to the third group (*A. gmelinii*, *A. anethifolia*, *A mongolica*, *A.dolosa*, *A. tanacetofolia*, *A. scoparia*, *A. subviscosa*, *A. dracunculus*).

Chemical composition of essential oils of plant of genus *Artemisia* L are depend not only from a species, but also on a districts growth (Figure 1) and weather conditions.

Work is executed at Russian Federal Property Fund support (the grant № 08-04-98037-r_Sib_a)

References:

Blinova K.F., Kuvayev V. B. Medicinal plants of the Tibetan medicine of Transbaikalia//Pharmacognosy Questions. **3**, 1965, 163–178.

Atazhanova G. A. Essential oils of plants of genus *Artemisia* L. Properties and application//Development of phytochemistry and prospect of building of new medical products - Karaganda, 2006. P.18-37.

Wormwood of Siberia: systematisation, ecology, chemistry, hemosystematization, use prospects / T.P. Berezovsky, etc. Novosibirsk, 1991. 125p.

Khanin M. A. Wormwood of Siberia and the Far East (pharmacognosy research and use prospects in medicine): autoref. dis. dock. pharm. sciences/M.A.Khanin. - Perm, 1999. 46p.

The state pharmacopoeia of the USSR. XI the edition. Moscow, 1990. 398p.

P47. Study of heavy metals distribution in organs and tissues of freshwater fishes from Kichera river (lake Baikal feeder) using principal component analysis

Pintaeva E.Ts.¹, Bazarsadueva S.V.¹, Radnaeva L.D.^{1,2}, Petrov E.A.³

¹Baikal Institute of Nature Management of SB RAS, Ulan-Ude, Russia

²Buryat State University, Ulan-Ude, Russia

³JSC "Vostsibribcenter", Ulan-Ude, Russia

Principal component analysis (PCA) is a favorite tool in environmetrics for data compression and information extraction. It has shown a great efficiency in various applications both in chemistry and biology.

The content and character of accumulation of heavy metals (Pb, Cd, Zn, Mn and Co) which are the most dangerous and well studied pollutants of superficial waters, in three kinds of fish of different trophic levels were investigated. There are pseudobenthophage (roach), both predators — a perch and a pike (in a food of the perch, along with fish, the important place belongs to bottom dwellers; the pike diet consists exclusively of fish).

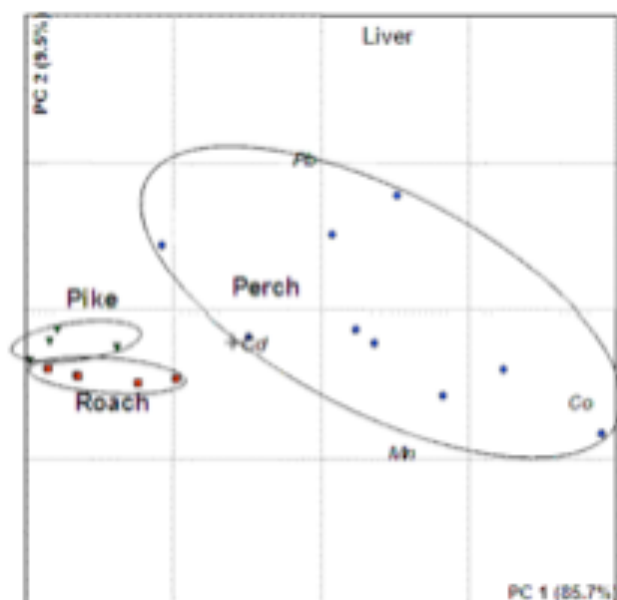


Fig. 1. Principal component analysis. Distribution of samples in liver of fish.

Metals in fish are not evenly distributed. It is associated with biochemical features of organs and tissues, specificity of physiological functions, non-identical behavior of different forms of metals, and etc. It was revealed that the analyzed organs and tissues accumulate various metals in different degrees.

Liver, as an organ where occurs detoxication of xenobiotics, represents an interest (Fig. 1). Pb, Cd, Mn and Co are more accumulated in the liver.

Application of PCA has allowed to distinguish different kinds of fish of different trophic levels on the content of heavy metals in the liver (Fig. 1). At the first it depends on differences in their ecology, mainly in the feeding nature.

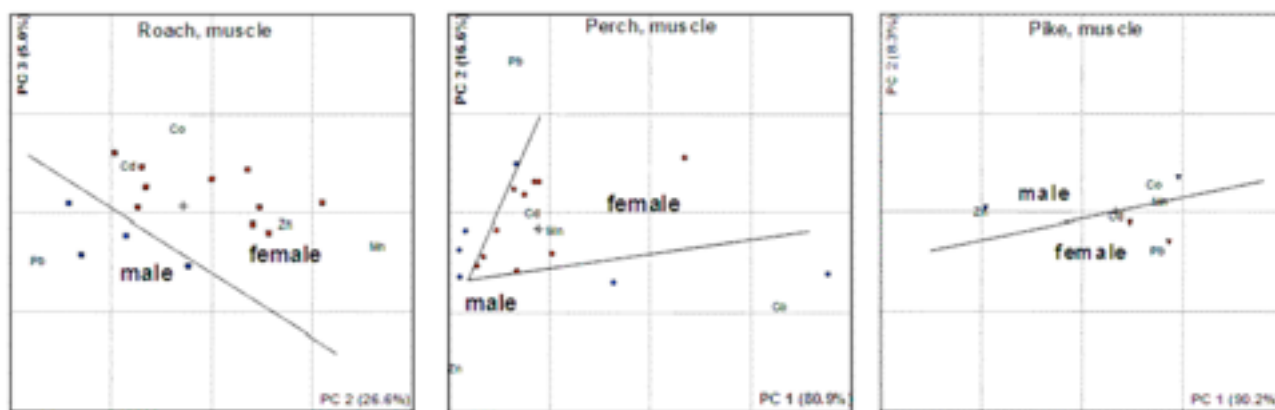


Fig. 2. Principal component analysis. Distribution of samples in muscles of fish depending on gender.

Muscles also serve as indicators of pollutions, but in lesser extent than the liver. Muscle, like the liver, can be attributed to the deposition agencies because it constitutes a large percentage of body weight. The PCA revealed sex differences in deposition of metals in the muscles of all investigated fish whereas in a liver the similar was not observed (Fig. 2).

Work is executed at Russian Federal Property Fund support (the grant № 08-05-98029-r_Sib_a)

Notes

Notes

List of participants

Appolonova Svetlana

Moscow Antidoping Centre
Head of Horse Doping Laboratory
Moscow, Russia
appolosa@yandex.ru

Baranov Pavel

Moscow Antidoping Centre
Research assistant
Moscow, Russia
baranov@dopingcontrol.ru

Belyaev Ivan

Altai State University
PhD student
Barnaul, Russia
iab@asu.ru

Blinov Kirill

Advanced Chemistry Development
Leading expert
Moscow, Russia
kirill.blinov@gmail.com

Bondarenkova Irina

Saint Petersburg State Technological University
of Plant Polymers
Senior lecturer
Saint Petersburg, Russia
bondarik@nwlk.spb.ru

Bulianitsa Anton

Institute of Analytical Instrumentation RAS
Head of department
Saint Petersburg, Russia
antbulyan@yandex.ru

Arvoh Benjamin Kaku

Telemark University College
PhD student
Porsgrunn, Norway
benjamin.k.arvoh@hit.no

Bazardueva Selmeg

Baikal Institute of Nature Management of SB
RAS
PhD student
Ulan-Ude, Russia
bselmeg@gmail.com

Bikmeev Denis

Bashkir State University
PhD student
Ufa, Russia
limm@list.ru

Bogomolov Andrey

J&M Analytik AG
Project manager
Essingen, Germany
Bogomolov@j-m.de

Brereton Richard

University of Bristol
Professor
Bristol, UK
r.g.brereton@bris.ac.uk

Burger James

BurgerMetrics SIA
Director
Jelgava, Latvia
james.burger@burgermetrics.com

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

Christensen Niels Peter Aae
University of Copenhagen
PhD student
Copenhagen, Denmark
npac@farma.ku.dk

Derkachova Olga
Saint Petersburg State Technological University
of Plant Polymers
Associate professor
Saint Petersburg, Russia
derkachevaou@rambler.ru

Dolgonosov Anatoly
Vernadsky Institute of Geochemistry and
Analytical Chemistry RAS
Professor
Moscow, Russia
amdolgo@mail.ru

Elyashberg Mikhail
Advanced Chemistry Development
Leading specialist
Moscow, Russia
elyas@acdlabs.com

Ermakov Vasiliy
Samara State Technical University
Engineer
Samara, Russia
wassiliy@rambler.ru

Cherednik Ekaterina
Tomsk State University
MS student
Tomsk, Russia
analytica@xf.tsu.ru

Daszykowski Michal
Institute of Chemistry, The University of
Silesia
Adiunkt
Katowce, Poland
mdaszyk@us.edu.pl

Dmitriev Alexander
Altai State University
Student
Barnaul, Russia
dmitriev@asu.ru

Dynina Alena
Saint Petersburg State Technological University
of Plant Polymers
Student
Saint Petersburg, Russia
afka88@mail.ru

Eriksson Lennart
MKS Umetrics AB
Chief Scientific Officer
Umea, Sweden
lennart.eriksson@umetrics.com

Esbensen Kim
Aalborg University
Professor
Esbjerg, Denmark
kes@aaue.dk

Folkmane-Burgere Aelita

BurgerMetrics SIA
Project manager
Jelgava, Latvia
daelitad99@hotmail.com

Gorbatsova Jelena

Tallinn University of Technology
PhD student
Tallinn, Estonia
jgorbatsova@yahoo.com

Gulyaev Ivan

Moscow State University
PhD student
Moscow, Russia
avmsu@mail.ru

Heberger Karoly

Chemical Research Center, Hungarian
Academy of Sciences
Scientific advisor
Budapest, Hungary
heberger@chemres.hu

Holm-Nielsen Jens Bo

Aalborg University
Head of Bioenergy and Green Engineering
Centre
Esbjerg, Denmark
jhn@bio.aau.dk

Jankevics Andris

Groningen Bioinformatics Centre, University of
Groningen
PhD student
Groningen, The Netherlands
a.jankevics@rug.nl

Giussani Barbara

University of Insubria
Researcher
Como, Italy
barbara.giussani@uninsubria.it

Gowen Aoife

University College Dublin
Postdoc researcher
Dublin, Ireland
aoife.gowen@ucd.ie

Halstensen Maths

Telemark University College
Associate Professor
Porsgrunn, Norway
maths.halstensen@hit.no

Heydorn Kaj

Technical University of Denmark
Consultant
Roskilde, Denmark
heydornk@gmail.com

Ihunegbo Felicia Nkem

Telemark University College
PhD student
Porsgrunn, Norway
felicia.n.ihunegbo@hit.no

Kabanova Natalja

Tallinn University of Technology
PhD student
Tallinn, Estonia
nataljakabanova@yahoo.com

Kalambet Yuri

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

Kirsanov Dmitry

Saint Petersburg State University
Senior researcher
Saint Petersburg, Russia
d.kirsanov@gmail.com

Kucheryavski Sergey

Aalborg University
Assistant professor
Esbjerg, Denmark
svkucheryavski@gmail.com

Lemanska Agnieszka

University of Bristol
PhD student
Bristol, UK
a.lemanska@bris.ac.uk

Maeder Marcel

University of Newcastle
Associate professor
Callaghan, Australia
Marcel.Maeder@newcastle.edu.au

Marbach Ralf

Optical Instruments Center
Chief Research Scientist
Oulu, Finland
Ralf.Marbach@vtt.fi

Khodyreva Nadezhda

Saint Petersburg State Technological University
of Plant Polymers
Vice-rector
Saint Petersburg, Russia
zsv@gturp.spb.ru

Korobova Elena

Vernadsky Institute of Geochemistry and
Analytical Chemistry RAS
Science secretary
Moscow, Russia
Korobova@geokhi.ru

Kurkina Victoria

Saint Petersburg State Institute of Technology
(Technical University)
Docent
Saint Petersburg, Russia
no@email.net

Madsen Michael

Aalborg University
PhD student
Esbjerg, Denmark
mima@bio.aau.dk

Mannoylov Vladimir

Institute of Analytical Instrumentation RAS
Head of department
Saint Petersburg, Russia
manoilov_vv@mail.ru

Marini Federico

Sapienza University of Rome
Researcher
Rome, Italy
fmmonet@hotmail.com

Marinichev Anatoly

Saint Petersburg State University

Docent

Saint Petersburg, Russia

amar@evergreen.spb.ru

Marks Chris

Chemistry teacher

Richmond, USA

camarks@yahoo.com

Mazina Jekaterina

Tallinn University of Technology

MS student

Tallinn, Estonia

jatak@msn.com

Monakhova Yulia

Saratov State University

PhD student

Saratov, Russia

yul-monakhova@mail.ru

Novikov Lev

Institute of Analytical Instrumentation RAS

Leading researcher

Saint Petersburg, Russia

novik38@mail.ru

Penkin Maxim

International Atomic Energy Agency

Senior safeguards analyst

Vienna, Austria

m.penkin@iaea.org

Markova Tatiana

Saint Petersburg State Technological University

of Plant Polymers

Associate professor

Saint Petersburg, Russia

t.s.markova@mail.ru

Marusina Inna

Saint Petersburg State Technological University

of Plant Polymers

PhD student

Saint Petersburg, Russia

O.V.Yanush@inbox.ru

Minkkinen Pentti

Lappeenranta University of Technology

Professor

Lappeenranta, Finland

Pentti.Minkkinen@lut.fi

Nikolaev Fedor

JSC "Leading institute "VNIPIET"

Senior scientific employee

Saint Petersburg, Russia

f_nikolaev@mail.ru

Paul Sylvia

University of South Africa

Professor

Pretoria, South Africa

Paulso@unisa.ac.za

Pomerantsev Alexey

Institute of Chemical Physics RAS

Leading researcher

Moscow, Russia

forecast@chph.ras.ru

Possokhov Yuri

Eastern Research & Development Institute of
Coal Chemistry
Senior researcher
Yekaterinburg, Russia
possokhoff@gmail.com

Roncoroni Simone

University of Insubria
PhD student
Como, Italy
simone.roncoroni@uninsubria.it

Rusinov Leon

Saint Petersburg State Institute of Technology
(Technical University)
Professor
Saint Petersburg, Russia
rusinov@ztl.spb.ru

Sanina Elena

Pharmaceutical Company ZAO
GlaxoSmithKline Trading
Management accountant
Moscow, Russia
alencas@bk.ru

Skvortsov Alexej

Institute of Cytology RAS
Associate professor
Saint Petersburg, Russia
colbug@mail.ru

Soktoeva Touyana

Buryat State University
PhD student
Ulan-Ude, Russia
stuyana85@mail.ru

Rodionova Oxana

Institute of Chemical Physics RAS
Senior researcher
Moscow, Russia
rcs@chph.ras.ru

Rudakova Irine

Saint Petersburg State Institute of Technology
(Technical University)
Associate professor
Saint Petersburg, Russia
riv-lilu@yandex.ru

Samokhin Andrey

Moscow State University
PhD student
Moscow, Russia
andrey.s.samokhin@gmail.com

Shelpakova Anna

Omsk State University
PhD student
Omsk, Russia
shelik85@mail.ru

Sokolova Svetlana

Saint Petersburg Institute of Informatics and
Automation RAS
Leading researcher
Saint Petersburg, Russia
sokolova_sv@mail.ru

Stanimirova Ivana

Institute of Chemistry, The University of
Silesia
Adiunkt
Katowce, Poland
istanimi@us.edu.pl

Styazhkin Pavel

JSC "Leading institute "VNIPIET"

Deputy chief of department

Saint Petersburg, Russia

f_nikolaev@mail.ru

Taivosalo Anastassia

Tallinn University of Technology

PhD student

Tallinn, Estonia

anastassia-t@mail.ru

Urtubia Alejandra

Universidad Tecnica Federico Santa Maria

Researcher

Valparaiso, Chile

alejandra.urtubia@usm.cl

Vlasova Irina

Omsk State University

Associate professor

Omsk, Russia

vlaso-iri@yandex.ru

Yakimova Eugenia

Saint Petersburg State Institute of Technology
(Technical University)

PhD student

Saint Petersburg, Russia

kimovaz@mail.ru

Zakharova Svetlana

JSC "Leading institute "VNIPIET"

Engineer-technologist

Saint Petersburg, Russia

s_zakharova@hotmail.ru

Sysoeva Inna

Saint Petersburg State Institute of Technology
(Technical University)

Student

Saint Petersburg, Russia

sysinna@gmail.com

Trygg Johan

Umea University

Associate professor

Umea, Sweden

johan.trygg@chem.umu.se

Vago Emese

Budapest University of Technology and
Economics

Teaching assistant

Budapest, Hungary

evago@mail.bme.hu

von Frese Juergen

Data Analysis Solutions DA-SOL GmbH

Managing Director

Greifenberg, Germany

jvf@da-sol.com

Yanush Oleg

Saint Petersburg State Technological University
of Plant Polymers

Professor

Saint Petersburg, Russia

o.v.yanush@inbox.ru

Zhilin Sergei

Altai State University

Docent

Barnaul, Russia

sergei@asu.ru