Drushbametrics Project

Third Winter School on Chemometrics

Modern Methods of Data Analysis



Russia, Pushkinskiye Gory, February 16-20, 2004

Российское хемометрическое общество (РХО) Университет Ольборг (ACACSRG) Институт геохимии и аналитической химии им. В.И.Вернадского (ГЕОХИ РАН) Институт химической физики им. Н.Н.Семенова (ИХФ РАН) Санкт-Петербургский государственный университет (СПбГУ) Компания "Люмэкс" (Lumex)



Третья международная школа-конференция

Современные методы анализа многомерных данных

Организационный комитет

Сопредседатели проф. Kim Esbensen (университет Ольборг) член-корр. РАН Грибов Лев Александрович (ГЕОХИ РАН) Ученый секретарь к.ф.-м.н. Родионова Оксана Евгеньевна (РХО) Члены Оргкомитета проф. Paul Geladi (NIR Nord) к.ф.-м.н. Крашенинников Анатолий Александрович (Люмэкс) к.г.н. Коробова Елена Михайловна (ГЕОХИ РАН) к.х.н. Легин Андрей Владимирович (СпбГУ) д.ф.-м.н. Померанцев Алексей Леонидович (РХО)

Адрес Организационного комитета

Россия 119991, г. Москва, ул. Косыгина, 4, Телефон/факс: (095) 939-74-83

http://rcs.chph.ras.ru/WSC3 e-mail: rcs@chph.ras.ru

Thanks

The WSC-3 organizers and participants wish express greatest appreciation to the following conference sponsors for their valuable economic and friendly help:

Ålborg University Esbjerg (Denmark)

Applied Chemometrics Analytical Chemistry & Sampling Research Group (Esbjerg, Denmark)

Polycert Company (Moscow, Russia)

Lumex Company (St. Petersburg, Russia)

Institute of Geo and Analytical Chemistry (Moscow, Russia)

Institute of Chemical Physics (Moscow, Russia)

Chemometrics and Intelligent Laboratory Systems journal (Elsevier Company)

For a hearty welcome we also want to thank

State Memorial Museum-Preserve of A.S. Pushkin (Pushkinskiye Gory, Russia)

Hotel Drushba (Pushkinskiye Gory, Russia)

Restaurant Lukomorye (Pushkinskiye Gory, Russia)

Our personal thanks are to the following people for their important contributions to the conference success:

Casper Dahl and Lars Petersen (for WEB support)

Finally, we are grateful to all the WSC-3 attendees, lecturers, accompanying persons and visitors for their interest to the conference.

See you again at the next WSC-4 conference!

Oxana Ye. Rodionova (secretary)

Mour

Alexey L. Pomerantsev (manager)

Helpful Hints

Foreign Visitors Registration

In accordance with the Russian law, each foreigner should be registered in police within 3 days after the entry. Most of you have already been registered by the St. Petersburg (Moscow) hotel, where you spent your first night in Russia. However, please, check it once more and contact with the conference manager, Alexey Pomerantsev, in case of any doubt.

Conference Venue

WSC-3 will be held in the *Pushkin Center*, which is situated in some 5 minutes walk from *Drushba* hotel. To reach the WSC-3 auditorium one should use the left-side entrance in the building. Actually, it will be the right-side entrance while walking from *Drushba* hotel, which is behind the Center.

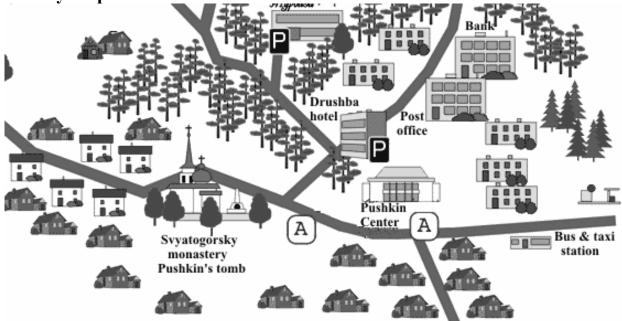
Bank and Currency Exchange

The bank is situated behind Drushba hotel in 10 minutes walk along the main (Lenin) street. Its name is СБЕРЕГАТЕЛЬНЫЙ (Sberbank). Here you can change euros and dollars. Cash dispenser machine is also available. Small sums (up to 50 USD) could be exchanged by the conference manager, Alexey Pomerantsev.

Communications

Mobile phones are working in Pushkinskiye Gory. You will find public telephone in the hotel Drushba lobby near the reception desk. Ask about the phone cards at the reception. The trunk-call office is situated in the building, which is close behind the hotel. E-mail connection will be available in the Pushkin Center Library, during the coffee breaks.

PushGory Map



Meals and Coffee

All meals will be served in the hotel *Drushba* restaurant. Coffee breaks will be held in the *Pushkin Center* in the hall, close to the WSC-3 auditorium.

Scores & Loadings

These traditional meetings will be held in the hotel *Drushba* restaurant every evening. As usual, tea, coffee, sugar, and biscuits are available <u>for free</u>, while more serious beverages are <u>for charge</u>. These drinks could be obtained in the restaurant bar.

Conference Badges

You are kindly asked to wear your badge all the day, both during the scientific and social activities.

Language

The official conference language is English.

Proceedings

Conference proceedings will be published as a book by Nova science Publishers, NY. The size of contribution is up to 25 pages. Printed manuscript as well as the corresponding files (MS Word 2000) on two diskettes should be submitted to the conference secretary, Oxana Rodionova, <u>before</u> the WSC-3 closing at February 20, 2004.

Posters

Time for pasting up your posters is Tuesday, February 17, 11:00-13:00. The place for Poster Session will be the hall near the WSC-3 auditorium.

Excursion at Wednesday

This will be a combined trip that includes bus transportation and short distance walks. We plan to visit three memorial country estates: Petrovskoye, Mikhailovskoye, and Trigorskoye. In Mikhailovskoye, we plan to divide all participants in two parts. The first group will go to Trigorskoye by bus (12 km by-pass), while the second group will make a direct 3 km hike. These people will enjoy the beautiful landscapes but could surmount some difficulties, like hill climbing, so only the strong persons are invited in the second group.

Weather and Clothing

Normal February temperature in PushGory is about -5°C at day, and about -10°C at night. We cannot expect extremely high temperature in the WSC-3 auditorium, so some warm wearing, like wool sweater, will be appropriate to this place.

Long Distance Trips

We plan two long distance trips (Tuesday and Thursday) for the accompanying persons. Each trip begins at 09:00 in the morning and ends about 18:00 in the evening. Please, contact with the social program manager, Katya Pomerantseva, about the details.

Useful Mobile Phone Numbers

Conference secretary	Oxana Rodionova	+7 9031504855
Conference manager	Alexey Pomerantsev	+7 9035618227
Social program manager	Katya Pomerantseva	+7 9037140646

Timetable

Monday, February 16, 2004 12:00–19:00 Arrival, registration 19:30–20:30 Dinner 20:30– Scores & Loadings

Tuesday, February 17, 2004

08:00-09:00

Breakfast

Session 1 09:00–10:00	Chair: Oxana Rodionova Conference opening L1 <i>Kim Esbensen</i> Reliable sampling, - chemical analysis and - chemometric data analysis	
10:00-10:30	T1 <i>Rolf Ergon</i> Informative score-loading biplot for monitoring of process with two response variables	
10:30-11:00	T2 Andrey Bogomolov Software for interactive curve resolution using SIMPLISMA	
11:00-11:30	Coffee Break	
11:30-13:00	Free time. Trip to the monastery. Poster mounting	
13:00-14:00	Lunch	
Session 2 14:00–15:00	Chair: Alexey Pomerantsev L2 Tormod Næs Measuring, monitoring and optimising food pro- cesses	
15:00-15:30	T3 <i>Igor Aristov</i> The "Electronic nose": a multivariate approach	
15:30-16:00	T4 <i>Alisa Rudnitskaya</i> Application of chemometrics for data processing of the electronic tongue	
16:00–16:30	Coffee Break	
Session 3 16:30–17:30	Chair: Tormod Næs L3 Oxana Rodionova Introduction to Simple Interval Calculation (SIC) method and its application for representative subset selection	
17:30-18:00	T5 Juan A. Fernandez Pierna Hyperspectral infrared technique: the use of Support Vector Machines (SVM) in large data sets	
18:00–19:00	Free time	
19:00–20:00	Dinner	
20:00-	Scores & Loadings	

Wednesday, February 18, 2004

08:00-09:00

Breakfast

09:00-13:00

Excursion (bus & walk) around Pushkin's places



small chapel in Mikhailovskoye park

13:00-14:00

Lunch

Session 4 14:00–15:00	Chair: Satu-Pia Reinikainen L4 <i>Kurt Varmuza</i> Global and local chemometric models of spectra- structure relationships
15:00-15:30	T6 Sergei Kucheryavski The analysis of fracture surfaces of porous metal materials using AMT and fractal geometry methods
15:30-16:00	T7 Alexander Bondarenko Inverse problem in potentiodynamic electrochemical impedance spectroscopy
16:00-16:30	Coffee Break
Session 5	Chair: Kurt Varmuza
16:30-17:30	L5 <i>Vladimir Palyulin</i> The joint application of 3D-QSAR and mo- lecular modelling in the design of new ligands of AMPA and NMDA receptors
17:30-18:00	T8 <i>Rustem Sunyaev</i> Chemometric analysis of benzine
18:00–19:00	Free time
19:00-20:00	Dinner
20:00-	Scores & Loadings together with museum staff

Thursday, February 19, 2004

08:00-09:00

Session 6	Chair: Vladimir Palyulin
9:00-10:00	L6 <i>Paul Geladi</i> Near infrared spectroscopy and calibration in in- dustrial applications
10:00-10:30	T9 Leon Rusinov Adaptation of calibration models while transferring within series of IR-spectrometers
10:30-11:00	T10 Elena Sulima Features of practical realization of transferring of calibration models from a master - device on working NIR - analyzers, for the base wheat indexes analysis
11:00–11:30	Coffee Break
Session 7	Chair: Paul Geladi
11:30–12:30	L7 Satu-Pia Reinikainen Wavelets: useful tool for compressing spectral data
12:30-13:00	T11 Jaap van der Weerd FTIR imaging and multivariate analysis of drug release from controlled release pharmaceutical formulations
13:00–14:00	Lunch
Session 8	Chair: Kim Esbensen
14:00-14:30	T12 David Kubicka Application of pattern recognition methods in analysis of complex hydrocarbon mixtures to improve the quality of kinetic data
14:30-15:00	T13 Taghi Khayamian Application of wavelet neural networks in multivariate data analysis
15:00-16:00	L8 Alexey Pomerantsev Multivariate Statistical Process Optimiza- tion
16:00–16:30	Kim Esbensen. Conference closing
16:30-18:00	Poster Session
18:00–20:00	Free time
20:00-	Conference banquet

Breakfast

Friday, February 20, 2004

08:00-09:00	Breakfast
09:00-	Departure to St-Petersburg and to Moscow (via Pskov)

Lectures

L1. Reliable sampling, - chemical analysis and - chemometric data analysis

<u>Kim Esbensen</u>, Lars P. Houmoller, Lars Petersen, Casper K. Dahl, Hans Henrik Friis Pedersen, Peter Paasch Mortensen, Jens Bo Holm-Nielsen, Lotte Hojbjerg, Julie Johnsen, Anette Ørnskov

ACACSRG, Ålborg University, Esbjerg, Denmark

We have inducted the Theory of Sampling (TOS), developed since 1950 by Pierre Gy, into the chemometric regimen. An overview of the intricate relationships between sampling, chemical analysis and the possibilities for reliable data analysis is presented. This contribution will highlight a few, but significant consequences not on the current chemometric agenda. What is "correct sampling for analysis and data analysis?", and why can it only be addressed based on TOS - contrary to many chemometric and analytical myths of a wishful "constant bias" situation. Critical in this context is the fact that sampling errors typically can attain 100-1000% of the analytical errors (s.s.), but usually go unnoticed! There are two types of "representativity" involved in these matters, one regarding the individual samples themselves and how they are sampled physically (the TOS issue) and one regarding the ensemble representativity of a data set w.r.t. the population it is supposed to represent (the statistical sample issue). We give exemplifications of "TOS-at-work" from a broad range of full-scale scientific and industrial applications, as well as from the analytical laboratory, to illustrate these representativity issues. They are totally interrelated when the issues surrounding "proper validation" is at hand.

L2. Measuring, monitoring and optimising food processes

Tormod Næs

Matforsk and University of Oslo, Norway

Increased competition is causing organisations to find ways to better meet the needs of the consumers, to reduce the costs and to increase productivity. This has a number of different aspects such as defining quality by the use of consumer tests and sensory analysis, on-line process monitoring and control and optimal use of the potential of the raw materials.

In the present lecture, we will discuss a number of these aspects with examples from the food industry. Most of the results and examples provided are related to an ongoing joint effort among Norwegian food manufacturers to utilise chemometric and statistical techniques (see www.ibion.no) for the purpose of improving product quality and economy in the production.

Among the topics covered are the following:

• how to use raw material measurements in order to adjust process variables in an optimal way. Alternatives related to both sorting and continuous updating will be covered

- how to design and analyse process experiments with focus on quality control and process optimisation
- how to utilise spectroscopy for process monitoring and prediction

The presentation will focus both on method developments and on real life examples from the food industry.

L3. Introduction to Simple Interval calculation (SIC) method and its application for representative subset selection

Oxana Rodionova and Alexey Pomerantsev

Institute of Chemical Physics, Moscow, Russia

The main theoretical issues of the SIC method that is proposed for linear modelling and for prediction interval estimation are considered. The roots of the method, based on ideas of Kantorovich, relates to applying linear programming to data analysis. The current SIC-approach is based on the singular assumption that the combined datamodel error is finite: For prediction modelling, this leads to results that are in a convenient interval form, and which account for all uncertainties present (X-measurement errors, Y-measurement errors, bilinear modelling errors). The SIC-approach also provides new possibilities for comprehensive object status classifications, i.e. a leverage-type classification of relative importance of the samples with respect to the specific calibration data set configuration and the mathematical model.

The ideas of the SIC method are applied for the representative subset selection in the calibration transfer problem. Calibration transfer relates the response of the slave instrument to that of the master instrument, using a special subset selected from a large calibration master set. Such a subset should satisfy two opposite requirements: be of the maximal representativity with respect to the total set, and be essentially less than the total set. The SIC object status classification defines among the calibration data set boundary samples that form a most representative subset.

A real-world data of whole wheat corn were analyzed. The X-measurements are NIR spectrum in the range of 908-1120 nm recorded at 118 wavelengths; the reference data (Y matrix) are protein, water, and gluten content, quantified in the laboratory by standard analytical methods.

L4. Global and local chemometric models of spectra-structure relationships

Kurt Varmuza

Vienna University of Technology, Vienna, Austria

The recognition of chemical structure information from spectral data challenged chemometricians since the pioneering time of this discipline. A lot of work has gone into the development of spectral classifiers based on a variety of multivariate classification methods. Actually such classifiers are capable to detect with low error rates the presence or absence of some substructures automatically from IR or MS data. Together with the known molecular formula and using appropriate software for isomer generation a systematic structure elucidation of organic compounds is possible in some cases [1,2]. However, because of the complexity of spectra-structure relationships, the high diversity of chemical structures, and the large amount of data this global approach is very limited in practical applications.

Routinely used for the identification of unknowns are spectra similarity searches in spectroscopic databases. Result is a hitlist (a subset of the database compounds) which contains the reference compounds exhibiting spectra that are most similar to the spectrum of the unknown. A method - based on binary substructure descriptors - has been developed for evaluating the structural similarity between the unknown and the hitlist compounds [3]. This method has been applied to improve search methods in IR and MS databases with the aim to yield subsets of compounds that contain useful chemical structure information about the unknown, even in cases the unknown is not in the database. Spectral and structural data from the hitlist compounds can be used to create local models by PCA, PLS, or other methods [4,5]. Resulting scatter plots may support the spectroscopist in spectra interpretation and finding useful relationships between spectra and chemical structures.

- 1. Varmuza K., Werther W.: J. Chem. Inf. Comput. Sci., 36, 323-333 (1996)
- 2. Penchev P. N., Andreev G. N., Varmuza K.: Anal. Chim. Acta, 388, 145-159 (1999)
- 3. Varmuza K., Karlovits M., Demuth W.: Anal. Chim. Acta, 490, 313-324 (2003).
- 4. Werther W., Demuth W., Krueger F. R., Kissel J., Schmid E. R., Varmuza K.: J. Chemometrics, 16, 99-110 (2002)
- 5. Scsibrany H., Karlovits M., Demuth W., Muller F., Varmuza K.: Chemom. Intell. Lab. Syst., 67, 95-108 (2003)

L5. The joint application of 3D-QSAR and molecular modelling in the design of new ligands of AMPA and NMDA receptors

Vladimir Palyulin, Igor I. Baskin, Irina G. Tikhonova, and Nikolai S. Zefirov

Moscow State University, Moscow, Russia

Molecular modeling and QSAR techniques being the powerful instruments for the design of new bioactive compounds in most cases are used separately while their joint application could provide deeper understanding of both structure-activity relationships and ligand-receptor interactions.

Molecular models of the NMDA receptor glycine binding site have been built using homology modeling technique on the basis of an X-ray structure of the watersoluble AMPA-sensitive receptor. The docking of agonists and antagonists to these models was used to reveal binding modes of ligands and to explain known structureactivity relationships. Two types of quantitative models, CoMFA (Comparative Molecular Field Analysis, 3D-QSAR technique based on PLS analysis of steric and electrostatic parameters of 3D ligand structures) and a regression model based on docking energies, were built for a series of antagonists. The CoMFA steric and electrostatic maps were superimposed on the homology-based model, and a close correspondence was marked. The derived computational models have permitted the evaluation of the structural features important for high glycine binding site affinity.

The approach for the analysis of binding selectivity of ligands was suggested. For analyzing the pair-wise selectivity, we propose to use the difference between biological activities (expressed in energy scale) of ligands with respect to different receptor subtypes as a dependent variable for building CoMFA models. The resulting "selectivity fields" indicate the ways of increasing binding selectivity. The applicability of this approach was demonstrated using Glycine/NMDA and AMPA receptors. The mapping of "selectivity fields" onto the molecular models of the corresponding receptors makes it possible to reveal the reasons of experimentally observed selectivity as well as to suggest additional ways of increasing it.

L6. Near infrared spectroscopy and calibration in industrial applications

Paul Geladi

Near Infrared Center of Excellence (NIRCE), Sweden

The near infrared region (780-2500 nm) is situated between the better known visual (400-780 nm) and infrared (2500-25000 nm) spectral regions. Near-infrared (NIR) spectrometry and calibration are very powerful tools for monitoring, regulating and controlling processes and for determining product quality within many branches of industry and the public sector. The major ingredients in all biological and organic materials can easily be determined by NIR spectroscopy. Examples of the use of NIR are found in clinical measurement, environmental measurement, waste management, forest industry, food industry, pharmaceutical industry, chemical industry and biotech industry.

Fast, non-contact, non-destructive and cost effective measurements can be done by the NIR-technique and move the chemical laboratory to the process line. Many industrial processes are regulated by relatively slow feedback mechanisms. The reasons for this are the time-consuming and expensive wet chemical analyses and tests of product quality. By constructing good calibration models between NIR spectra and chemical or quality tests, the NIR-technique offers high accuracy prediction of the outcome of these tests. This allows the delivery of almost real-time results and sometimes quality parameters can be calculated even before the product comes out of the process. If the process is starting to go wrong or if the product is showing bad quality, operators can immediately begin to make corrections and minimize or completely eliminate bad products. This gives savings of time and energy and a better environmental efficiency (less waste produced).

Calibration on near infrared spectral data is on of the important testing grounds and success stories of calibration in chemometrics.

- 1. B. Osborne, T. Fearn & P. Hindle, *Practical NIR Spectroscopy with Applications in Food and Beverage Analysis* (2nd ed.), Longman Scientific & Technical, London, UK, 1993.
- 2. D. Burns & E. Ciurzak eds., *Handbook of Near-Infrared Analysis*. Marcel Dekker, New York, NY, 1992.
- 3. H. Siesler, Y. Ozaki, S. Kawata & H. Heise eds., *Near Infrared Spectroscopy*. Principles, Instruments, Applications, Wiley VCH, Weinheim, 2002.

L7. Wavelets: useful tool for compressing spectral data

Satu-Pia Reinikainen

Lappeenranta University, Lappeenranta, Finland

Wavelet methods provide powerful tools for analyzing, encoding, compressing, reconstructing, and modeling large spectral data, and images. Wavelet analysis may offer added insight in data analysis situations where Fourier techniques have previously been used. It is well known from Fourier theory that a signal can be expressed as the sum of a, possibly infinite, series of sines and cosines. This sum is also referred to as a Fourier expansion. The disadvantage of the Fourier expansion is that, unlike wavelet transformations, the FT's basic functions are continuous over time and are, therefore, not suitable for representing time-related information of the transformed signal. The wavelet transform or wavelet analysis is probably the most recent solution to overcome the shortcomings of the Fourier transform. In wavelet analysis the use of a fully scalable modulated window solves the time scale problem. The window is shifted along the signal and for every position the spectrum is calculated. Then this process is repeated many times with a slightly shorter (or longer) window for every new cycle. In the end the result will be a collection of time-frequency representations of the signal, all with different resolutions.

The use of the wavelet transform has tended to focus on the compression and noise reduction of separate signals. Wavelet analysis and compression, followed by PCA or PLS and their generalizations, form a simple and practical tool, allowing computations to be made using a reduced set of variables, but without any loss of information. Results can be interpreted in the original variable space, using the inverse wavelet transform on resultant model parameters such as loadings. Additionally, the time-frequency nature of many types of signals can facilitate separation of different features, thus increasing the detection limit and improving fault detection. Some common aspects, related to denoising, smoothing, and compression of data in applying wavelet transformations and multivariate data analysis in 1-D and 2-D spectral data is considered and illustrated by using, e.g., NIR and NMR data.

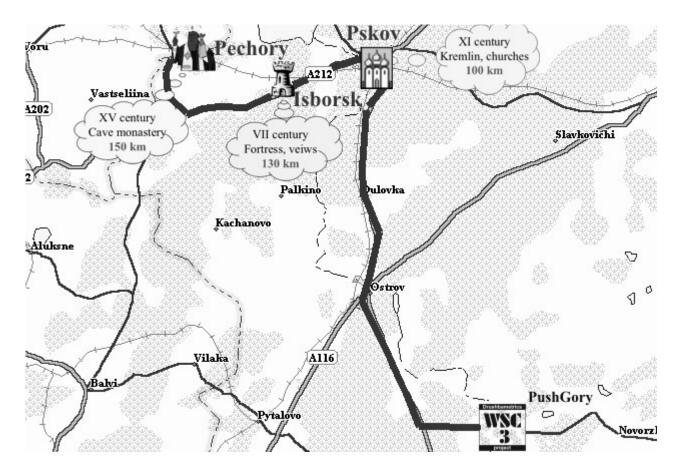
L8. Multivariate Statistical Process Optimization

Alexey Pomerantsev and Oxana Rodionova

Institute of Chemical Physics, Moscow, Russia

The multivariate statistical process control (MSPC) is a popular approach that helps to understand and to run the real-world processes. However, we think that the general aim of any statistical analysis of technology is to improve the quality of the final product or/and to reduce the production costs. The main MSPC concept is to apply the historical process X-data for construction of linear (calibration) model, which explains how the final results (i.e. quality, y) depend on X-variables. Apparently, studding this model, one can suggest the program of actions that could improve the performance in general. However, the most important thing in production is the local optimization, which means the immediate actions that are performed in the course of process (i.e., inline). The main problem in optimization of linear models is not to find a solution, but to restrict the area where the solution is found. Badly limited area could give the unreasonable results. We suggest to use the main concept of the MSPC approach and to pose that for the each state of a process those actions are admissible that are similar to ones that sometimes happened in the historical data in the similar situation.

The presentation contains two new ideas, which (we believe) are interesting in process control. The first one comes within the conventional MSPC approach. We suggest constructing a series of MSPC models that describes the process in order of its cycle, stage by stage. This Expanding MSPC (EMSPC) method can be used for prediction of future final product quality at the early process stages while production cycle is not over yet. This may be considered as a passive optimization as we do not suggest the correcting actions, but just predict the expediency of the offered actions. The second idea is not usual in the standard MSPC approach as it implies the novel SIC method to find the best in-line corrections. This technique may be called the Multivariate Statistical Process Optimization (MSPO) and it also applies the idea of EMSPC. The MSPO approach may be considered as an active optimization as we always suggest the correcting actions in the course of the process.



Long distance trips

Talks

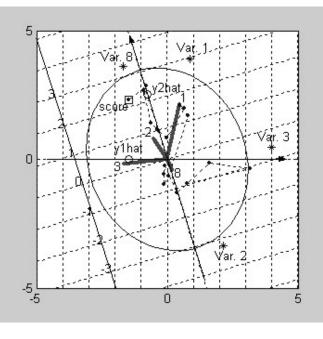
T1. Informative score-loading biplot for monitoring of process with two response variables

Rolf Ergon

Telemark University College, Porsgrunn, Norway

For a process with many x-variables and two y-variables, it is possible to construct a single score-loading biplot such that

- there is an exact correspondence between the directions for loadings and scores
- the score for a new object is the exact sum of contribution vectors obtained by multiplication of loadings with the corresponding new variable values
- new predictions of the two y variables can be read directly on axes for that purpose
- unusual score movements can be detected by use of a confidence ellipse according to statistical considerations.



A process operator will thus get all relevant information presented in one plot only. The plot is based on the general theory for score-loading correspondence [1] and the 2PLS algorithm [2], and it has some features in common with the PCP method [3]. In the figure the new method is illustrated by use of data from a mineral processing plant [4]. Although the number of x-variables in this data set is 12, the important contributions to the score in the projection used are given by four variables. The sum of the contribution vectors 1, 2, 3 and 8 thus approximately result in the score. The length of a contribution vector shows the absolute

value of the corresponding x-variable, while the *-marked normalized loadings make it possible to see the sign. The score in the figure is thus caused by a positive value of variable 1, and negative values of variables 2, 3 and 8. Some old scores are shown by solid dots connected by dotted line.

- 1. Ergon, R., J. Chemometrics, 16 (2002), 368-373
- 2. Ergon, R., J. Chemometrics, 17 (2003), 303-312
- 3. Langsrud O, Næs T., to appear in Chemom, Intell. Lab. Syst., 68 (2003), 61-74.
- 4. D. Hodouin, J.F. MacGregor, M. Hou, M. Franklin, Can. Inst. Mining Bull. 86 (1993), 23-34.

T2. Software for interactive curve resolution using SIMPLISMA

Andrey Bogomolov, Michel Hachey, and Antony Williams

ACD Moscow, Russia

In the world of modern analytical data, hyphenated techniques and multidimensional experiments are now fairly standard in the analytical laboratory. Indeed, for maximum data content they dominate classical forms of analysis and the resulting data existing in the form of a matrix is now a general output. At the same time, new methods of multivariate data analysis have migrated from the domain of published theory to routine laboratory procedures, with intuitive software being a necessary element of this shift.

A wide number of applications require the ability to resolve a data matrix into the underlying set of physical response curves. This problem, known as curve resolution, can be illustrated by extracting concentration profiles and the related spectra of pure components from a series of spectra, which have been generated as a function of time. Such arrays of data can be generated by monitoring a chemical reaction spectroscopically or through the process of chromatographic separation.

The SIMPLe-to-use Interactive Self-modeling Mixture Analysis (SIMPLISMA) by Windig [1] is a well-known approach for multivariate curve resolution. It provides a powerful means of analyzing samples that undergo changes with respect to time without requiring prior information of either the mixture composition or component spectra and concentrations. Even with the widespread application possibilities for SIMPLISMA, it still remains a tool for the experts due, in part, to the learning curve for applying such algorithms. Specifically, expert generic software and macros are not necessarily suited to mainstream users. In this work, we will present the first commercial implementation of SIMPLISMA in a ready-to-use intuitive interface designed to make SIMPLISMA more readily accessible to chemists as a routine analysis tool.

1. W. Windig, J. Guilment, Anal. Chem. 63 (1991) 1425.

T3. The "Electronic nose": a multivariate approach

Igor Aristov¹ and T.A. Kuchmenko²

¹Voronezh State University and ²Voronezh State Technological Academy, Russia

Solution of the problems of qualitative and quantitative analysis of multicomponent systems is possible on the basis of three approaches: 1) separation of the mixture into different components and their analysis; 2) application of selective sensors for detection of each of the components in the mixture and observing of the individual responses; 3) application of multi-element system of relatively non-selective sensors with the following extraction of the information from the summarized response of sensor array.

The first approach is characterized by inevitable losses of substances under separation thus resulting on the quality of analysis. Second approach is limited by insufficient selectivity of sensors in multi-component mixtures thus resulting in overlapping of signals. The third approach is connected with the use of the methods of multi-dimensional statistics and image recognition under analysis of the total response of poly-sensor array to the multi-component object.

In the report application of the methods of discrimination and classification under analysis of the vapor of volatile organic substances is discussed when using detector of "electron nose" kind. Detector is an array of sensors based on piezoelectric quartz resonators with different polymer coatings on the electrodes. The array is adjusted for multi-component gas mixtures of organic compounds, and certain sensors in the array are characterized by cross sensitivity to the main components in a mixture. Under exposure of sensors in the analyzed gas mixture each element in the array has an individual response to the main components of mixture.

T4. Application of chemometrics for data processing of the electronic tongue

<u>Alisa Rudnitskaya</u>, Andrey Legin, Andrey Ipatov, and Yuri Vlasov St. Petersburg University, St. Petersburg, Russia

Chemometrics is widely used for processing and interpreting of chemical data from various analytical instruments. Chemometric tools are essential part of multisensor systems for both liquid and gas analysis (electronic tongues and electronic noses). The underlying idea of such systems is application of a set of semi-selective or cross-sensitive chemical sensors, which produce response to a number of chemical species simultaneously. It is arguable that this is analogous to the mechanism of human taste and smell perception. The use of chemometrics is therefore vital for this type of analytical instruments.

Electronic tongue based on potentiometric chemical sensors was successfully applied to different analytical tasks including recognition and classification of foodstuffs, taste assessment, quantitative analysis of food, natural and waste waters etc. Chemometric methods appropriate to the task were used: PCA for data evaluation and recognition, SIMCA and PLS-regression for classification, PLS-regression for multi-variate calibration and neural networks for all of the above.

The present paper is devoted to the application of chemometrics tools to the processing of data from flow-injection electronic tongue based on potentiometric chemical sensors. By the nature of sensors used each of them gives only one signal in each measured sample – value of stationary potential vs. reference electrode. Flow-injection design of the electronic tongue presumes continuous measurements when background solutions, then samples then again background solutions are passing through measuring cell. Thus, for each sensor a number of points can be collected producing 3-dimensional data set (sensors × potentials × time). Traditionally only one value of all is used for data processing namely the maximum change of sensor potential in sample compared to background solution. However, response of a sensor in time might also contain useful information e.g. about interaction of a sensitive membrane with solutions' components. In the present paper two types of data sets of flow-injection electronic tongue were used for data processing and results were compared. Different approaches to the data processing of time response of the electronic tongue (three dimensional data set) are evaluated and discussed.

T5. Hyperspectral infrared technique: the use of Support Vector Machines (SVM) in large data sets

Juan A. Fernandez Pierna, V. Baeten, A. Michotte Renier, and P. Dardenne Agricultural Research Centre of Gembloux, Belgium

Support Vector Machines (SVM) is currently a very active research area within machine learning. SVM is a learning approach, which utilizes the concept of kernel substitution in order to make the task of learning more tractable by exploiting an implicit mapping into a high dimensional space. SVM has been successfully applied to numerous tasks within data mining, computer vision, bioinformatics and multivariate calibration, for example.

In this work, two main tasks are performed: first the Near-infrared camera coupled to a wavelength selector is presented as a fast technique in order to obtain a large number of NIR spectra in few minutes. The use of hyperspectral infrared techniques becomes necessary in order to treat all the information contained in the data. As a second task the application of SVM as a classification method is studied. For this part, different kind of samples measured using the NIR camera are studied in order to discriminate them according to the differences on particle origin.

T6. The analysis of fracture surfaces of porous metal materials using AMT and fractal geometry methods

Sergei Kucheryavski, Viktor Polyakov, and Artem Govorov

Altai State University, Barnaul, Russia

The investigation of deformation structure that is occurring in metal materials under the plastic deformation and fracture allows us to understand the deformation processes and obtain the most effective operative conditions. The study of fracture surfaces of those materials gives us the ability to determine the fracture conditions and dependencies of deformation characteristics of materials on different properties. The most difficult thing in such investigations is to obtain the quantitative description of fracture surfaces that, usually, have not uniform structure. This is still more complicated when one studies the complex multiphase materials with a very heterogeneous structure such as, for example, porous materials.

In this work the fractal analysis and the AMT (Angle Measure Technique) has been considered. These two methods empower us to obtain the estimated value of the "complexity" of fracture surfaces structure and to classify the fractured materials depending on the different criteria. We applied both methods to the classifying of simulated surfaces with the different values of roughness and noise. In spite of the similarity of algorithms and existence of the correlation between fractal dimension and AMT spectra, the results show that the applying of concrete method depends on the image parameters. For example, if we have noised image or image of not self-similar surface the better results can be obtained with using the AMT.

T7. Inverse problem in potentiodynamic electrochemical impedance spectroscopy

Alexander Bondarenko, G.A. Ragoisha

Belarusian State University, Minsk, Belarus

Potentiodynamic electrochemical impedance spectroscopy (PDEIS) is a new digital technique for the characterisation and analysis of the electrochemical response of interfacial processes and structures [1-6]. PDEIS belongs to the latest generation of the digital techniques that employ virtual instruments both for the implementation of the experiment and analysis of the acquired data. Common electrochemical techniques, due to their analog origin, investigate various aspects of the electrochemical response separately. Thus, with common techniques, ac and dc responses, transfer functions and effects of the variable potential have to be analysed separately, each by its own technique, and this impedes greatly the electrochemical characterisation, especially in the case of nonstationary systems. PDEIS acquires the frequency response and dc response in a single potential scan and represents the data in a 3D plot as the variation of the frequency response and dc current with the potential. We have developed the computer program for the analysis of these 3D spectra. The program decomposes the total response of the interface into the individual responses of the constituent processes. The analysis is based on different frequency responses of the double layer capacitance, components of Faraday impedance and other constituents of the total response. Each of the constituents is represented by one or several parameters of the variable equivalent electric circuits. The analyser uses the complex nonlinear least squares fitting routine to find the equivalent circuit and represents the parameters as functions of the potential. In order to minimise the ambiguity of the analysis, three fitting sub-routines based on Box, Newton and Powell algorithms were implemented, with three different minimisation functions available to each of them. The analyser was built into the program of the virtual spectrometer, so that data acquisition and analysis can be implemented simultaneously to give constituent responses from the measurements of the transfer function evolution in the potential scan.

- 1. G.A. Ragoisha, A.S. Bondarenko, *Solid State Phenom.* **90-91** (2003) 103-108; free preprint
- 2. G.A. Ragoisha, A.S. Bondarenko, *Electrochem. Comm.* 5 (2003) 392.
- 3. G.A. Ragoisha, A.S. Bondarenko, World Scientific 2003, 373-376; free preprint
- 4. G.A. Ragoisha, A.S. Bondarenko, Minsk, 2003, 138-150; free preprint
- 5. G.A. Ragoisha, A.S. Bondarenko, Surface Science, 2003, in press; free preprint
- 6. G.A. Ragoisha, A.S. Bondarenko, N P. Osipovich, E.A. Streltsov, J. Electroanal. Chem., 2003, in press.

T8. Chemometric analysis of benzine

<u>Rustem Sunyaev¹</u>, R.Z. Safieva¹, O.G. Chulyukov², and S.S. Demygin² ¹Oil and Gas University and ² "Lumex-Centrum" Ltd, Moscow, Russia

The chemometric analysis was made by the PLS-method on the base of reference and NIR-spectrum data for the graduation series of benzines. First being disstillated 18 under atmospheric pressure benzine distillate fractions (48 samples) were kindly given by oil refinery (Orsk, Russia). The NIR-analysis of benzines was provided by NIRanalyser "INFRALUM FT-10" due to special software of Russian firm "Lumex".

The calibration models with high predictable ability for some properties of benzine (density, fractional composition, octane number, contents of resins) were created. Cross-validation of these models was made. The Standard error of cross-validation (SECV) was calculated. The repeatability data from the spectrum analysis in all cases was higher than repeatability of reference methods. The validation of test benzine samples showed precision of the calibration models that were made also.

T9. Adaptation of calibration models while transferring within series of IR-spectrometers

Leon Rusinov, K.A. Zharinov, A.L. Sulima, and V.A. Zubkov

Technological University, St. Petersburg, Russia

The series of transference methods of calibration models of a quantitative analysis for spectrometers working in short-range infrared range of a spectrum from a master device on working ones is considered. The comparative assaying of a method of direct transfer and transference method with linearly - piecewise regression is resulted.

T10. Features of practical realization of transferring of calibration models from a master - device on working NIR - analyzers, for the base wheat indexes analysis

<u>Elena Sulima</u>, K.A Zharinov, V.A. Zubkov, and L.A.Rusinov Technological University & Lumex Company, St. Petersburg, Russia

The transference method of initial calibration model, which is generated by PLSmethod, from a master - device on working NIR-analyzers "INFRALUM FT-10 " is considered. The methods of sampling for models transfer on values of reference data, and also with the help of the scores parameter of the first principal component are considered. On an example of the wheat indexes analysis the results of an estimation of quality of models transfer from a master - device on 14 working NIR-analyzers placed on elevators of Krasnodar Territory of Russia are resulted.

T11. FTIR imaging and multivariate analysis of drug release from controlled release pharmaceutical formulations

Jaap van der Weerd and Sergei G. Kazarian

Imperial College London, UK

Formulation of controlled release pharmaceuticals is one of the most challenging issues in modern drug research. The ability to control the drug release from a tablet is of crucial importance for drug administration. A drug concentration that is too low might not be effective at all, while a high concentration, e.g. an initial 'burst', can be dangerous and reduces the efficiency of the drug.

The most frequently used technique to study drug release is the monitoring of the increase of drug concentration in a fluid in which the tablet is immersed. While highly informative, this method does not provide any information on the mechanisms of release inside the tablet. Recently, FTIR imaging has been proposed to study drug release from tablets. The high chemical specificity of FTIR imaging allows the complete quantification of all materials involved, i.e. tablet material, drug and water. Besides, the relatively high spatial resolutions that can be obtained (~15 μ m for the used set-up) allows a detailed insight in the relevant mechanisms. FTIR imaging normally yields relatively big data-sets, and as several measurements are performed as a time-series, the amount of data is enormous. Data-reduction can be achieved this using multivariate analysis, e.g. partial least squares (PLS). The final goal of these multivariate and subsequent calculations is the transformation of the FTIR imaging results into pharmaceutically relevant information.

T12. Application of pattern recognition methods in analysis of complex hydrocarbon mixtures to improve the quality of kinetic data

David Kubicka, Dmitry Yu. Murzin

Åbo Akademi, Åbo, Finland

Detailed analysis and identification of reaction products are essential for interpreting the kinetic data in order to disclose relevant reaction mechanisms. In the case of very complex reaction mixtures, the identification of individual products is often impossible and classification into product groups, based on the similarity of the reaction components, is then the only possibility to extract relevant information about the reaction mechanism. Pattern recognition methods constitute a power tool to achieve classification into individual classes derived from the structural similarities between chemical components.

Soft Independent Modeling of Class Analogies (SIMCA) has been successfully applied in the classification of C10 monocyclic alkylnaphthenes into classes using their mass spectra. Seven classes, reflecting the number of carbon atoms in the naphthenic ring and the length of the longest alkyl-substituent, have been defined. The pretreatment of the raw spectral data is of crucial importance for reliable classification. Therefore, several pretreatment procedures, including the data centering, scaling and autotransformation, have been investigated and their effect on the separation of individual classes will be discussed, together with the classification results, in detail in the presentation.

T13. Application of wavelet neural networks in multivariate data analysis

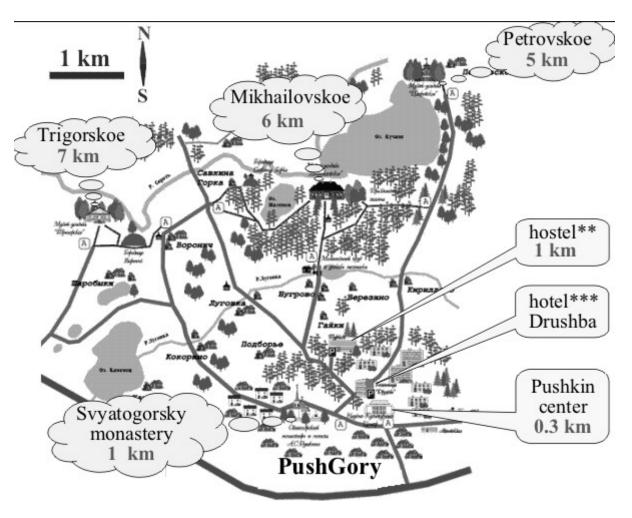
Taghi Khayamian and M. Esteki

Isfahan University of Technology, Isfahan, Iran

Wavelet neural network (WNN) is an alternative soft modeling method for learning at multiple resolutions based on wavelet functions. In other words, WNN is an artificial neural network with wavelet functions as an activation function for the nodes in the hidden layer.

In this work the application of WNN in quantitative structure property relationship (QSPR) is investigated for prediction of critical micelle concentration (cmc) of nonionic surfactants, for the first time. The model was constructed based on three topological descriptors and two quantum chemical descriptors. The topological descriptors were the Kier and Hall index of zero order of the hydrophobic segment of the surfactant, the average information content of order 2 for the hydrophobic segment and the relative number of nitrogen and oxygen atoms. The quantum chemical descriptors were the polarizability and the highest occupied molecular orbital energy level, which were obtained using the Hyperchem software (semi-empirical AM1 method). The quantum chemical descriptors were selected in a stepwise manner, from many different descriptors using multiple linear regression (MLR) method.

The data set consists of 73 molecules, which were randomly divided into two groups, the training set and the prediction set. The training set and the prediction set consist of 54 and 19 molecules respectively. The performance of the model was evaluated by plotting the prediction of cmc by the model against the experimental values for the prediction set. The large correlation coefficient (0.997) a large value of F (5765) and a small standard error (0.083) revealing the capability of the WNN model.



Surroundings of PushGory

Posters

P1. Critical survey of mass reduction devices in practical sampling (industry and laboratory)

Lars Petersen, Casper K. Dahl and Kim H. Esbensen ACACSRG, Ålborg University, Esbjerg, Denmark

As shown several times in the history of sampling, often samples need to be of large mass due to the fact that many sub-samples from different regions of the sampled materials (frequently) have to be combined into forming the representative primary sample. This often calls for the need of a heavy mass reduction to obtain the final sample. The reduction in mass often needs to be of a factor 20 or (much) more. The present work presents various theoretical aspects of mass reduction and various practical examples to illustrate these. The effective characterizing property of mass reduction is called representativity, and is defined as:

$$r^2(SE) = m^2(SE) + \sigma^2(SE),$$

where the selection error, SE, is defined as:

$$SE = \frac{a_{sample} - a_{Lot}}{a_{Lot}}$$

and a is the grade of the critical component.

A large experiment was carried out to characterize 16 mass reduction techniques and devices with regard to representativity. The material system used consisted of wheat (89.9% W/W), rape seed (10% W/W) and fine glass particles (0.1% W/W). The reduction ratio was 1 to 20 or 1 to 16 depending on the device/method.

Large improvements in representativity can be obtained by using the correct device or procedure for mass reduction. The present work will be presented along with other practical examples of sampling in real world industrial settings, where mass reduction is essential.

Correct mass reduction provides the individual samplers with a great degree of freedom!!!

P2. Image analytical sampling - prospects and limitations

Casper K. Dahl and Kim H. Esbensen

ACACSRG, Ålborg University, Esbjerg, Denmark

Sampling is the most important unit operation for product and/or process monitoring and control. Structurally correct sampling, based on the Theory of Sampling (TOS) must be used wherever analysis plays a role in the decision-making process. In many industrial cases it is difficult and/or deemed too expensive to use correct TOSmethodology however and "remote sensing" in the form of robot vision or image analysis has been put forward as an interesting alternative. This has been termed "sampling by proxy" ... There is one major limitation when using image analysis however:

Image analysis only records a rendition of the surface of the object being imaged, while the necessary reference sampling necessarily must be based on the total volume

of the sample(s) being imaged by a camera. Thus Image Analytical Sampling (IAS) combines traditional image analysis (univariate, multivariate) and TOS and is therefore a potential method for analyzing correctly delimited samples only relying on surface features (texture, morphology, chemistry manifestations etc.). This Ph.D. study is aimed at delineating where is there a limit to how - and to which material types - the IAS methodology can be applied!

P3. Peculiarities of rubber-oil interaction under the conditions of cold climate

<u>Natalya Petrova</u>, A.F Popova, O.V Startsev¹

Institute of Non-Metallic Materials, Yakusts, ¹Altai State University, Barnaul, Russia

Injection of plasticizers - low molecular weight substances, which promote flexibility and elasticity of the material - is the most widespread method to improve frost resistance of rubbers. However, investigations of climatic stability of rubbers under the conditions of sharply continental cold climate showed that their interaction with hydrocarbon media (oil, mineral oil, lubricants) causes intense wash out of the plasticizers resulting in sharp deterioration of low-temperature characteristics. Since diffusion processes extensively determine efficiency of elastomer units under real operation conditions, it is necessary to carry out their detailed study and modeling. Taking into account complexity of occurring processes, their dependence on temperature, the first stage of investigations should include modeling at constant temperature. The present paper covers the process of swelling of elastomer materials in oil at room temperature using the model of multicomponent diffusion and FITTER software. Peculiarities of rubber-oil interaction, which were performed in the form of successive penetration of hydrocarbons of different composition into the elastomer material (quick diffusion of light hydrocarbons during the first day, long-term process of penetration of heavier hydrocarbons) and extraction of the plasticizer, were revealed. Correlation between diffusion coefficient values computed at room temperature and changes in basic operational properties of rubbers after open-air exposition in oil was brought to light, which allowed determining new approaches to optimize the rubber recipes at the stage of designing. It was shown that injection of absorbers, such as natural zeolites into the material composition, application of two-component elastomer mixtures, of which one performs high oil and benzene stability, allows slowing down oil diffusion and washing out plasticizers from the rubber mixture

P4. Acoustic Chemometrics for determination of skim milk in water - preliminary results

<u>Andriy Kupyna</u>, Elling-Olav Rukke, Anton Bryhn, Maths Halstensen¹ and Tomas Isaksson

Agricultural University, Oslo and ¹Institute of Process Technology, Porsgrunn, Norway

Acoustic chemometrics (AC) [1] is a new on-line measurement method, which can be used to monitor all processes, which generate sound/vibrations. The principle of AC is simple: obtaining acoustic signals induced by the process, followed by some form of signal transformation and chemometric data analysis.

The aim of this preliminary study was to evaluate the use of AC to analyze aqueous food systems. Skim milk in water has been chosen as a model system. Such systems are very common in food industry. For example it may be a problem when changing from one product to another or when pipe-lines have been filled with water for washing or rinsing according to a CIP (clean in place)-system. 21 levels of skim milk in water with a skim milk concentration from 0 to 50 weight-% (0-4,29% dry matter) in 2,5% steps were evaluated by on-line acoustic chemometrics. Measurements of each concentration level were performed in 5 replicates giving 105 samples per trial. Six trials were carried out giving totally 630 samples. All the measurements were done at 30,2°C. A non-cavitating experimental rig constructed from a 31 mm copper pipes, positive pump, mixing tank and orifice plate (6,5 mm constriction hole) at measurement zone [2], was used in this study. Acoustic "noise" from the turbulent flow was sampled and transformed by Fast Fourier Transform giving the frequency spectra. PLSregression was performed on these spectra. Test-set was used as a validation method. Each of the 6 trials was used as a test set giving 6 PLS-models. O-PLS (Orthogonal projections to latent structures) [3] was used as a preprocessing method in order to improve interpretation of the models.

Preliminary results demonstrated that AC has a potential as an on-line measurement method in food industry. The acoustic frequency spectra changed already at adding of small amounts of milk to water, however the change was not as visible at higher concentrations as it was at 0-2,5%. There was possible to predict the concentration of milk (up to 50%) with an RMSEP = 3,465 - 4,535 and correlation 0,935 - 0,980.

- 1. Esbensen, K. H., Halstensen, M., et al. Chemom. Intell. Lab. Syst. 44 (1998) 61-76
- 2. Esbensen, K.H, Hope, B., et all J. Chemometrics, 13 (1999) 209-236
- 3. Trygg, J. and Wold, S. J. Chemometrics, 16 (2002.) 119-128

P5. Chemometric investigation of polarization curves: further adventures

Christopher A. Marks

University of Virginia, Charlottesville, 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 have 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. Polarization curves from a family of Al-Co-Ce alloys in 3.5 % NaCl solution are investigated in this work.

P6. Band and point statistical estimations of channel-forming polypeptide potential

<u>Andrei Dmitriev</u>, I.V. Markov, V.G. Baryshnikov, and V. A. Tverdislov¹ Lipetsk State Pedagogical University and ¹Moscow State University, Russia

The adequate statistical estimation algorithm of electrostatic potential distribution in channel-forming polypeptides of biological and model membranes was developed.

The formula, defining the potential confidential band, which is explained due to experimental errors of X-ray diffraction and NMR measurements of polypeptide atomic coordinates, was obtained. The offered approach allows to:

- conduct demonstrative band comparison of polypeptides potential distribution
- conduct demonstrative band comparison of polypeptides potential distribution
- make demonstrative averaging of local maximums and minimums of polypeptide's potential

P7. Ecology and toxicology monitoring pesticides in agrosystem: structuring and analysis

Galina Larina and Yu.Ya. Spiridonov

Institute of Phytopathology, Moscow, Russia

Ecology and toxicology monitoring includes field, green house and laboratory experiments. As a result researcher has a set to heterogeneous information: the number and kind of a weed plant in sowing of the different cultures, length of the conservation remainder herbicide in components agrosystem, ecological load and other. Received of the monitoring information requires the conservations and systematizations. So was designed scheme of the structuring and analysis of data for example pesticides.

The working results (initial data) are entered in working area of the tables formed in the program Excel and STATISTICA. Here they are subjected to mathematical processing. The basic methods: dispersion and correlation analysis. Further information is imported to a database "Herbicide 99" (is created on the basis of the program MS ACCESS 97). Now in database the information about 304 active ingredients (a.i.): being the active beginning herbicides (162 preparations), fungicides (90 preparations) and insecticides (87 preparations). Database represents a set is systematized connected among themselves of help and experimental data, accompanying conditions of experiments, designed parameters of persistence, toxicity, efficiency of preparations and results of the quantitative and graphic analysis. It allows (with the help of the integrated tools) to spend additional mathematical processing of data, to form searches and to receive as the reports the necessary information. For example, it is possible to create the report about average values of productivity, increase of a crop and ecological loading on different cultures and in view of a place of application herbicides. This method helps the researcher to apply herbicides in practice as reception of operative management status of agrosystem and the greatest possible degree of safety for an environment. The

fresh data in the reports updated automatically on the basis of the information contained in database "Herbicide 99" on the concrete moment.

P8. Chemometric analysis of oils

<u>*Rimma Safieva¹*</u>, R.Z. Suinuaev¹, O.G. Chulyukov², and S.S. Demygin² ¹Oil and Gas University and ² "Lumex-Centrum" Ltd, Moscow, Russia

Some of crude oils from different oilfields (31 samples) were chosen for chemometric analysis. Before it all investigated crude oils (with contents of water from 0,5 to 64 %) were deemulsified under laboratory conditions. The NIR-analysis of deemulsified oils was provided by NIR-analyzer "INFRALUM FT-10" due to special software of Russian firm "Lumex".

It was shown that NIR- spectra of prepared oils were so different in their form and optical density that it was possible to divide them into three groups: A, B and C. The most representative group of oils (28 samples) from group C was elected for further investigation. For creation the calibration models for some properties of oils (density, fractional composition, temperature of freezing, kinematic viscosity under 20,30,50°C) PLS-method was used. Cross-validation of these models was made. There are some unrealized possibilities for improving the created calibration models for oils, the investigations shall be continued

P9. System analysis and chemometrics of modern problems in common chemistry

Yevgeny Slyusarenko and V.T. Mosidze

Moscow State University, Moscow, Russia

Chemometrics is the important part of system analysis so as "... our knowledge's about chemical phenomenon's have been expressed by numbers, geometrical images and in during time (*V.I. Vernadsky*)". System analysis a modern problems in principal sections of common chemistry: structure of atoms, periodic low, theory of chemical bond, thermodynamics and kinetics of the processes in multi-component systems has been carried out and quantitative correlation's have been established.

As results of analysis a correlation between the atomic masses of the most abundant isotopes of the noble gases and the charge in their nuclei has been established. A shell model of nucleus consisting of proton and neutron sublevels has been offered. The full filling sublevels leads to periodic changes of atomic masses in periods, and also in groups of s-, p-, d- and f-elements. Mathematical equation for Mendeleev's periodic low the change of atomic masses of elements has been obtained.

A modern theory of chemical bond cannot prognoses a direction and character chemical bond between reacting atoms in molecules and crystals. In present investigation a new approach for describing of chemical bond - geometric conception of chemical bond (GCCB) - have been used. Besides electronic excitation of atoms in GCCB are used a concept orbital excitation of atoms. Obtained configurations of excited atoms and use as elements for design of molecules and crystals. Geometry of the some more interesting molecules and crystal structures have been calculated

The experimental investigation of multi-component systems is a very labourconsuming task. The analysis of the phase equilibria in multi-component systems: Ni-V-Cr-Mo-Re; Ni-Mn-V-Cr-Mo-W: Ni-V-Nb-Ta-Cr-Mo-W have been realised using the method of graphs.

Calculations the processes of interaction on the phase boundary in multicomponent inorganic materials by discrete computer model have been carried out.

P10. Textural and mineralogical classification soils in national park 'Plescheevo lake'

Valentina Slyusarenko and Yu.N. Zborischuk

Moscow State University, Moscow, Russia

The investigation of texture and mineralogical composition soils in national park 'Plescheevo Lake' has been carried out for different elements of relief within of the lake hollow. Here the parent materials have the glacial origin. The soils in the sloping parts are forming on the morainal loams, which are covered by fluvioglacial sands. On the steep slopes the parent materials are ancient-alluvial sands.

The soil texture was estimated by the close-packed coefficient (Kcp), which is a ratio a number of sand to a number of dust in investigated soils. Kcp changes from 2,13 to 2,85 for ideal close-packed soils. This coefficient allows to estimate the degree of transformation both loams and sands under weathering and podzol-forming processes.

More detailed information gives the use of dust coefficient (Kf) and cryogenic contrast coefficient (KKK). These coefficients give the information about mineralogical differentiation and weathering of soils.

P11. A data-driven classification of solvents used in extraction of organic compounds

Vladimir Zernov and Igor V. Pletnev

Moscow State University, Moscow, Russia

The organic solvents classification resulting from exploration of set of partition coefficients (LogP) comprising 49 water-immiscible solvents (including an ionic liquid, BMIM PF6) and about 4000 solutes is described. Because of missed data in given dataset, the pairwise data were used for classification (cluster analysis and Sammon non-linear mapping) and estimation of the number of factors (principal component analysis) that determine solute partition data.

Resulting classification shows the two stable groups of solvents. The first contains various hydrocarbons and their halo- and nitro- substitutes. The latter contains alcohols, ethers, esters, and tributylphosphate. The location of two solvents is specially discussed: ionic liquid is situated between two above-mentioned groups; chlorobenzol lies far away from both groups.

Given fully empirical classification is compared with the others described in literature

P12. Application of multivariate statistical methods to classification of metal ions according to their complexing properties

Vladimir Zernov and Igor V. Pletnev

Moscow State University, Moscow, Russia

We obtained a classification of metal ions according to their complexing properties using a large experimental data set of stability constants of metal ion complexes, logK₁, with correction to the ionic strength. The data taken from IUPAC Stability Constants Database comprise publications to 2000 (inclusively) and cover complexes of Ag⁺, Al³⁺, Ba²⁺, Be²⁺, Ca²⁺, Cd²⁺, Ce³⁺, Co²⁺, Cu²⁺, In³⁺, Fe²⁺, Fe³⁺, Ga³⁺, H⁺, Hg²⁺, K⁺, La³⁺, Mg²⁺, Mn²⁺, Na⁺, Ni²⁺, Pb²⁺, Sr²⁺, UO²⁺, VO²⁺, Y³⁺, Zn² with 4560 distinct ligands. Corresponding matrix contains blank cells, the total number of valid entries being 18660. A correlation matrix of size 27×27 was built out for this data set and metal similarity series for every ion were drawn. Further, we performed factor and cluster analyses of obtained correlation matrix. There exists six clusters of metal ions; the classification does not contradict with the well-known schemes by Arhland-Chatt-Davies/Pearson but complements them and allows to examine the complexing properties more closely. Factor loadings of the five most significant factors were also used to build a Kohonen map, which visually summarizes the properties of metal ions.

P13. Multivariate methods for data processing in organic geochemistry

Elena Belitskaja and Yury P. Turov

Institute of Petroleum Chemistry, Tomsk, Russia

In organic geochemistry one operates with abundant data arrays (factual data) about composition, properties and spatial distribution of organic matter (OM) in geological objects. Multi-component compositions of both dispersed organic matter and the concentrated hydrocarbon accumulations demand using special expedients of multivariate statistics to search and reveal latent regularities in the experimental data and to make up qualified conclusions on the nature of the observed phenomena.

The paper presents some opportunities of applying variance, correlation and factorial analyses and other statistical expedients to investigate genetic regularities of OM transformation and migration in geological time scale.

The results obtained on gas chromatography-mass spectrometry analyses of hydrocarbon compositions of oils, petroleum fluids, underground waters, cores, natural bitumen and of other geologic samples serve as the initial information for this purpose.

The conclusions obtained may be used to forecast oil-and-gas content in the investigated areas, to simulate feed of layers in the operating oil fields and to obtain additional arguments in the continuing discussion on the nature of petroleum.

P14. Range of relative bias as a criterion of calibration quality

Elena Shabanova and Irina Vasilyeva

Institute of Geochemistry SB RAS, Irkusts, Russia

Comparison of techniques for estimating precision of unknown regression parameters is the subject of wide speculation. Nevertheless, the choice of estimation and a technique evaluating remains the complex task as before.

A common conception about validation of multivariate calibration in AES supposes their stability to spectral and matrix influences for a wide area of analyte content. We have applied the range of relative bias (RRB) on CRM set, that was used in training and test samples to evaluate validity of n-dimensional regression model. Utilizing the RRB as a criterion of regression quality, we have orientated on the nature of characteristic under study and assessed, i.e. bias of analytical result. The RRB sensitivity to random errors is decreased by normalized bias on certified value of element for each CRM. Relative bias estimation is distinguisher feature mainly for major composition in CRM set. The RRB for the CRM set is capable to account differences of chemical composition of CRMs.

If the relative bias of every determined elements in j-th calibration sample is estimated by formula

 $R_j = (C_{calculated} / C_{certified} - 1) \times 100\%$

where $C_{calculated}$ and $C_{certified}$ are contents calculated and certified accordingly, then RRB for the CRM set on interval of contents $\Delta C = C_{max} - C_{min}$ will be equal to $\Delta R = R_{max} - R_{min}$. Optimal condition of any n-dimensional calibration model is written as ΔR =>min. When calibration model has good validation, RRB is to be minimally for training and test samples, that is $\Delta R_{training} => \min$ and $\Delta R_{test} => \min$ simultaneously.

P15. Modeling of data structure for multivariate calibration in atomic emission spectrometry

Elena Shabanova and Irina Vasilyeva

Institute of Geochemistry SB RAS, Irkusts, Russia

Application of the multivariate analysis for atomic emission spectrometry (AES) bases on the dependence of analyte line intensity from: (i) its content in sample; (ii) total chemical composition, i.e. concentrations of major elements; (iii) contents of interferents; (iv) presence of molecular bands in spectra.

Table of initial data has to include information that reflects these connections. The rows of the table are spectra of certified reference materials (CRM). The columns of the table are independent (content) and dependent (intensities of lines of major, interference and determined elements) variables. Selecting measurement range is crucial stage to be appropriated the area of changing these characteristics. In this case the direct and latent connections of measured values and determined analyte contents are taken into account when constructing regression multivariate models.

This study presents the influence of structure of spectral data in training samples on the validation of multivariate calibration for different types of spectral lines.

P16. Investigation of interrelationship between parameters of the phenomenological and physicochemical equations describing anodic peaks in case of stripping voltammetry

Ekaterina Larionova, A.G. Stromberg, S.V. Romanenko, and D.M. Karachakov

Tomsk Polytechnic University, Tomsk, Russia

In this work the interrelation between two approaches to research of peak-shaped analytical signal in stripping voltammetry is investigated: on the one hand, physicochemical simulation of the electrode processes underlying an analytical signal, and with another phenomenological (empirical) modeling of the analytical signal. In processing of signal interpretation comparison between physicochemical and phenomenological model allows to effectively solve a number of practical problems of analytical chemistry (the resolution of overlapped analytical signals, detection, filtration, etc.) as well as to investigate mechanism of electrode processes, electrochemical behavior of substances. Theoretical anodic peaks for reversible electrode process on a mercury film electrode in the case of liner sweep anodic stripping voltammetry are obtained by physicochemical modeling. For construction of phenomenological functions it was used several approaches: modifications of the base functions, their combinations and construction signal model from the simple function [1]. The interrelationship was established between the parameters of the phenomenological equations and some physicochemical parameters (film thickness, rate of potential changes, etc.) in case of liner sweep anodic stripping voltammetry.

1. Stromberg, A.G.; Romanenko, S. V., J. An. Chem. 55, (2000), 1144-47

P17. Adaptive mathematical models of composition in analytical chemistry and earth sciences

Yury P. Turov and Elena Belitskaja

Institute of Petroleum Chemistry, Tomsk and Yugorsky University, Khanty-Mansiysk, Russia

A methodological approach is proposed and discussed for adaptive mathematical model generating in chemical analysis with multichannel detectors. The methodology is oriented for multidimensional data sets manipulation, such as GC/MS, diodes matrix detectors, and combined data sets – MS + spectrophotometry + elementary composition etc. Model dimension (i.e. number of components) and white noise level are considered as adaptive parameters. The algorithms are based on dispersions analysis and iterative target transformation factor analysis. The target mathematical model connects the data sets of experimental characteristics – the multi-channel detector output signals – with component contents and their individual characteristics.

The report is illustrated by examples from environment chemistry, petrochemistry and geochemistry, concerned with research of genesis, migration and biodegradation of organic pollutants, oil systems, oil fractions, and also with analysis and comparison of the compositions of dispersed organic matter in recent and ancient sediments. Moreover there are the examples illustrating capabilities of mass spectral and optical spectroscopy methods in research of paleovolcanism which was effected on the structure of organic matter and, in particular, on that of polycyclic aromatic hydrocarbons in ancient sediments.

P18. Explanatory data analyses of fish-breeding in Siberian lakes. First forays

<u>Alexey Semenchenko¹</u>, O.Ye. Rodionova² and A.L. Pomerantsev²

¹Ecological Monitoring University and ²Institute of Chemical Physics, Moscow, Russia

Wide breeding of whitefish (*Coregonus peled*) began in the lakes of Western Siberia in the 1960-s. A lot of lakes were stocked yearly. Total catch of fish exceeded 5 000 tons during fishery arrangement with the share of whitefish about 90%. Now the amount of 1 400 tons of whitefish are planting yearly. We study the data set collected for the last 40 years about whitefish culture in the lakes of Ural and Western Siberia (11 different regions). Till now, MDA methods have not been applied for data processing of Russian fish culture industry.

The goals of research are: the explanatory analyses of lake features and weather conditions, which influence on effectiveness of lake fish culture and prediction of whitefish culture effectiveness in the Tobol region lakes.

Steps of analyses. 1) Construction of the PLS2 model on the whole data set. The results show essential difference between lakes located in different regions. Distinct clusters may be seen in the score plots. 2). For one cluster (Tobol region lakes) the PLS1 model is used for prediction of fishery yield. 3). For the groups of lakes with total square less then 60 sq. km, the PCA model helps to analyze the influence of lake conditions on productivity. Software Unscrambler 8.0 was used for data analyses.



In Trigorskoe park

List of Participants

Name	Affiliation	Contacts
Aristov, Igor Аристов, Игорь	Voronezh State University Voronezh, Russia Assistant	+7(4722) 85 59 87 igor@anch.vsu.ru
Belitskaja, Elena Белицкая, Елена	Institute of Petroleum Chemistry Tomsk, Russia Researcher	+7(3822) 492597 7 helene@ipc.tsc.ru
Bogomolov, Andrey Богомолов, Андрей	ACD Company Moscow, Russia Project Leader	+7(095) 4382153 bogomol@acdlabs.ru
Bondarenko, Alexander Бондарекнко Александр	Belarusian State University Minsk, Belarus PhD Student	+375 172 247043 bondarenkoas@bsu.by
Dahl, Casper Даль, Каспер	Ålborg University ACACS-RG Esbjerg, Denmark PhD Student	+45 79127691 ckd@aue.auc.dk www.acacsrg.aue.auc.dk
Dmitriev, Andrei Дмитриев, Андрей	Lipetsk Pedagogical University Lipetsk, Russia Assistant	+7 (0742)242913 a_v_dmitriev@mail.ru
Ergon, Rolf Эргон, Рольф	Telemark University College Porsgrunn, Norway Head of department	+47 35 57 51 60 Rolf.Ergon@hit.no www-pors.hit.no/~rolfe/
Ergon, Malfrid Эргон, Малфрид	Norway Accompanying person	

Name	Affiliation	Contacts
Esbensen, Kim Эсбенсен, Ким	Ålborg University ACACS-RG Esbjerg, Denmark Professor	+45 (7912) 7688 kes@aue.auc.dk www.acacsrg.aue.auc.dk
Fernandez, Juan Фернандес, Хуан	Agronomical Research Center Gembloux, Belgium Research Assistant	+32 (0)81 62 03 55 fernandez@cra.wallonie.be
Geladi, Paul Желади, Пол	Swedish University of Agricultural Sciences Umea, Sweeden Researcher	+46 907869473 paul.geladi@chem.umu.se
Gribov, Lev Грибов, Лев	Institute of Geo and Analytical Chemistry Moscow, Russia Professor	+7 (095) 1374147 elkor@geokhi.msk.su www.geokhi.ru
Holm-Nielsen, Jens-Bo Холм-Нильсен, Иенс-Бо	Ålborg University ACACS-RG Esbjerg, Denmark Researcher	+45 7912 7715 jhn@aue.auc.dk www.acacsrg.aue.auc.dk
Khayamian, Taghi Хайамьян, Таги	Isfahan University of Technology Isfahan, Iran Associate Professor	+98 311 391 2351 taghi@cc.iut.ac.ir
Amiri, Maryam Амир, Мариам	Iran Accompanying person	
Korobova, Elena Коробова, Елена	Instituteof Geo and Analytical Chemistry Moscow, Russia Science Secretary	+7 (095) 1374127 Korobova@geokhi.ru www.geokhi.ru

Name	Affiliation	Contacts
Kubicka, David Кубичка, Давид	Abo Akademi University Abo, Finland PhD Student	+358-2-215-4559 dkubicka@abo.fi
Kucheryavski, Sergei Кучерявский, Сергей	Altai State University Barnaul, Russia Docent	+7 (3852) 363571 svk@asu.ru
Kupyna, Andriy Купина, Андрий	Agricultural University Aas, Norway PhD Student	+47 64 94 77 15 andriy.kupyna@ikbm.nlh.no
Larina, Galina Ларина, Галина	Institute of Phytopathology Moscow, Russia Senior Researcher	+7 (233) 40902 galina_larina@vniif.rosmail.com
Larionova, Ekaterina Ларионова, Екатерина	Tomsk Polytechnic University Tomsk, Russia PhD Student	+7 (8322)563832 evs@anchem.chtd.tpu.ru
Marks, Christopher Маркс, Кристофер	University of Virginia Charlottesvile, USA Laboratory Supervisor	+1(434)982-5794 cam8h@virginia.edu www.virginia.edu/cese/
Marks, Laurel Маркс, Лорел	USA Accompanying person	
Næs	Oslo University	47 22 25 50 27

Næs, Haec, Тормод Oslo University Matforsk Oslo, Norway Adjunct Professor

+47 22 85 59 87 tormod.naes@matforsk.no

Name

Palyulin, Vladimir Палюлин, Владимир

Petersen, Lars Петерсен, Ларс

Petrova, Natalya Петрова, Наталья

Pomerantsev, Alexey Померанцев, Алексей

Pomerantseva, Katya Померанцева, Екатерина

Reinikainen, Satu-Pia Райникайнен, Сату-Пиа

Reinikainen, Ville Райникайнен, Виле

Affiliation

Moscow State University Moscow, Russia Leader Researcher

Ålborg University ACACS-RG Esbjerg, Denmark PhD Student

Institute of Non-Metallic Materials Yakutsk, Russia Laboratory manager

Institute of Chemical Physics Moscow, Russia Senior Researcher

Institute of Gene Biology Moscow, Russia Researcher

Lappeenranta University Finland Researcher

Finland Accompanying person

Rodionova, Oxana Родионова, Оксана Institute of Chemical Physics Moscow, Russia Senior Researcher

Contacts

+7 (095) 9393557 vap@org.chem.msu.su

+45 79127691 lape@aue.auc.dk www.acacsrg.aue.auc.dk

pnn2002@mail.ru

+7 (095) 9397483 forecast@chph.ras.ru rcs.chph.ras.ru/ICP/paleng.htm

+7 (095) 135-9734 katy.sol@mail.ru

+358-5-621 2112 satu-pia.mujunen@lut.fi www.lut.fi/kemia

+7 (095) 9397483 rcs@chph.ras.ru rcs.chph.ras.ru/ICP/royeng.htm

Name	Affiliation	Contacts
Rudnitskaya, Alisa Рудницкая Алиса	St. Petersburg University, St. Petersburg, Russia Researcher	+7(812)282835 andrew@sensor.chem.lgu.spb.su www.electronictongue.com
Rusinov, Leon Русинов, Леон	Technological University St. Petersburg, Russia Professor	+7 (812) 259-4753 rusinov@tu.spb.ru
Safieva, Rimma Сафиева, Римма	Oil & Gas University Moscow, Russia Professor	+7 (095) 930-9063 rustem@garnet.ru
Semenchenko, Alexey Семенченко, Алексей	Ecological Monitoring University Moscow, Russia Student	
Shabanova, Elena Шабанова, Елена	Institute of Geochemistry Irkutsk, Russia Researcher	+7 (3952)425837 shev@igc.irk.ru
Slyusarenko, Yevgeny Слюсаренко, Евгений	Moscow State University Moscow, Russia Researcher	+7 (095) 939-4061 vacht@yachoo.com
Slyusarenko, Valentina Слюсаренко, Валентина	Moscow State University, Moscow, Russia PhD Student	+7 (095) 939-3578 vacht@yachoo.com
Sulima, Elena Сулима, Елена	Lumex Company St. Petersburg, Russia Researcher	+7 (812) 118-6862 StasenkoEL@lumex.ru www.lumex.ru

Name

Sunyaev, Rustem Сюняев, Рустем

Turov, Yuri Туров, Юрий

van der Weerd, Jaap ван дер Вирд, Джап

Varmuza, Kurt Вармуза, Курт

Zernov, Vladimir Зернов, Владимир

Zharinov, Anatoly Жаринов, Анатолий Affiliation

Oil & Gas University Moscow, Russia Professor

Institute of Petroleum Chemistry Tomsk, Russia Researcher

Imperial College London, UK Research Associate

University of Technology Vienna, Austria Professor

Moscow State University Moscow, Russia Researcher

Lumex Company St. Petersburg, Russia Researcher +431-58801-16060 kvarmuza@email.tuwien.ac.at

Contacts

+7 (095) 930-9063

rustem@garnet.ru

+7 (3822) 492597

tur@ipc.tsc.ru

+44 20 5895111

jvdw@ic.ac.uk

www.lcm.tuwien.ac.at

+7 (095) 9395464 zernov@analyt.chem.msu.ru

+7 (812) 118-68-62 Zharinov@lumex.ru www.lumex.ru Krasheninnikov, Anatoly Крашенинников, Анатолий

Krasheninnikov, Anatoly Крашенинников, Анатолий Lumex Company St. Petersburg, Russia Manager

Lumex Company St. Petersburg, Russia Manager +7 (812) 3151517 krash@lumex.ru www.lumex.ru

+7 (812) 3151517 krash@lumex.ru www.lumex.ru