

Drushbametris Project

FOURTH WINTER SCHOOL ON CHEMOMETRICS

Modern Methods of Data Analysis



Russia, Chernogolovka, February 15—18, 2005

Институт проблем химической физики (ИПХФ РАН)
Российское хемометрическое общество (РХО)
Университет Ольборг (ACACSRG)
Институт химической физики им. Н.Н.Семенова (ИХФ РАН)



Четвертая международная школа-конференция

Современные методы анализа многомерных данных

Организационный комитет

Сопредседатели

член-корр. РАН Разумов В. Ф. (ИПХФ РАН)
д.ф.-м.н. Померанцев А. Л. (ИХФ РАН)
проф. Эсбенсен К. (ACABS)

Ученый секретарь

к.ф.-м.н. Родионова Оксана Евгеньевна (РХО)

Члены оргкомитета

член-корр. РАН Грибов Л. А. (ГЕОХИ РАН)
проф. Вармуза К. (Австрия)
к.ф.-м.н. Кучерявский С.В. (АлтГУ)
к.х.н. Богомоллов А.Ю. (РХО)

Адрес Организационного комитета

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

<http://www.chemometrics.ru/wsc4/>

e-mail: rcs@chph.ras.ru

Thanks

The WSC-4 organizers and participants wish express greatest appreciation to the following conference sponsors for their valuable economic and friendly help:

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

Polycert Company (Moscow, Russia)

Institute of Problems of Chemical Physics (Chernogolovka, Russia)

Institute of Chemical Physics (Moscow, Russia)

Chemometrics and Intelligent Laboratory Systems journal (Elsevier Company)

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

Casper Dahl (for WEB support)

Finally, we are grateful to all the WSC-4 attendees, lecturers, accompanying persons and visitors for their interest to the conference.

See you again at the next WSC-5 conference!



Oxana Ye. Rodionova (secretary)



Alexey L. Pomerantsev (manager)

Useful information

Location

The present symposium is held in Chernogolovka, a young town founded in 1960s by Nikolay Semenov, a Noble Prize winner and the head of the Institute for Chemical Physics (ICP). Initially established as a campus for a new branch of ICP, during the next two decades Chernogolovka had grown into a real “science city” consolidating a number of institutes for natural sciences and over 20,000 inhabitants. The ICP subsidiary has turned into an independent institution called the Institute for Problems of Chemical Physics (IPCP) that is the WSC-4 host organization.

Just an interesting fact, the share of Academicians (members of the Russian Academy of Sciences) among the townsfolk is the largest in the whole Russia.

Chernogolovka is enclosed in the ring of picturesque forests and lakes that let both aboriginals and guests feel being in touch with the nature. Local people are very friendly; their average education level is rather high so if you apply a stranger in the street with a question in English you have good chances to get necessary help.

The symposium sessions will be held at KON, which is a generally used abbreviation for the General Purpose Building (refer to the Vocabulary below) at IPCP.

Transportation

The city is small and can be walked through in about 20 minutes, so there is no public transport in it. Taxi is always at your service 24 hours a day and can be called by phone (Useful Phone Numbers section below). A local trip by taxi normally costs 40 rubles (~\$1.5) while long trips to Moscow are around 600 rubles to the “Shchelkovskaya” metro station, 1000 rubles to the downtown, and 1100 rubles to the Sheremet’evo (SVO) airport. Therefore, hiring local taxi is cost-effective if you go to the airport directly from Chernogolovka. However, the bus service (routes No.320 and No.360) is generally recommended to get to Moscow. Buses go every 30 min in the daytime from the bus station. Also, there are shuttles (minivans) that leave one by one as filled up. Bus tickets can be bought at the station 10 – 20 min prior to the departure. In the shuttle you pay to the driver. Ticket price (one way) for both kinds of transport is about 40 rubles. The bus destination point is Shchelkovskaya metro station. Driving time to Moscow is about 1 hour.

Accommodation

Symposium participants will mostly stay at the city hotel. A minor part of guests will be accommodated at a small hotel, Third Street.

Meals

Breakfast will be served at the hotels. A small restaurant called Akademia Club next to the marketplace is reserved for lunches and dinners. The symposium banquet will take place at this restaurant as well.

Scores&Loadings

Every night after the dinner everyone is invited to stay at the restaurant for Scores&Loadings. This is the name of a famous traditional party where all WSC participants meet to discuss things (that may go far beyond the area of chemometrics) in a relaxing friendly atmosphere and in immediate proximity to a source of beer and stronger alcohol. This ancient tradition is an important and inherent part of the Drushbametries program.

Communication

The main two Russian cellular nets, BeeLine and MTS, have excellent coverage in Chernogolovka and around. Even in the forest within 4-5 km mobile phones keep steady connection. The Internet can be accessed from KON and from Akademia Club restaurant. Public long-distance phone booths are available at the city hotel. Intercity and international telephone cards are present in most of supermarkets and stores.

To reach a local phone number from a distant place one should dial

+ *[international line]* 7 *[code of Russia]* (09652) *[local code]* x-xx-xx *[five-place local number]*

For calls from Moscow the local area code (252) can be used instead.

Money

You may exchange Euros and US dollars to Russian rubles (RUR) in banks and some supermarkets. Sberbank is a bank that is closest to the city hotel (indicated in the scheme below). There are several bank machines around the city designated by the \$ sign in the scheme. A cash dispenser in the KON lobby is one of the most opportune one.

Shopping

There are several supermarkets and a lot of stores in the city. The city marketplace is a right place to buy agricultural products. There you may also find some home craft such as valenki (Russian national felt woolen shoes).

Good-quality alcohol of any type is widely represented in a supermarket that is about 200 m from the bus station as shown by an arrow in the scheme below.

Leisure

In accordance with the conference program you are not supposed to have too much free time☺.

In case you were able to find it somehow, we recommend you to spend it walking. A beautiful Southern lake is only 5 min from the city hotel. A circled 2 km walkway through the forest starts from the lake. It is suitable both for skiing and walks and stays illuminated all night through.

Nightclubs, dance halls, casinos are not among the attractions of Chernogolovka.

Program for Accompanying Person

Accompanying persons could spend time in Moscow on their own, or ask Inna Tatarintseva, the manager of CityServiceTour for any assistance.

Weather

Temperature in the range from -5°C to -10°C is typical for the mid February. It is usually a little bit colder than in Moscow. Chernogolovka is often subjected to winds so wearing a warm cap and gloves is strongly recommended.

Miscellaneous

The conference official language is English.

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

Vocabulary of Names and Terms

«Щелковская» [shchelkovskaya]	Shchelkovskaya, metro station at the bus terminal in Moscow
«КОН» [kon]	General Purpose Building, symposium place
«Академ-Клуб» [akadem klub]	Academia Club, restaurant
«Гостиница» [gostinitsa]	Hotel, the main symposium hotel
«Первая улица» [pervaya ulitsa]	First street
«Вторая улица» [vtoraya ulitsa]	Second street
«Третья улица» [tret'ya ulitsa]	Third street
«Институтский проспект» [institutski]	Institute avenue

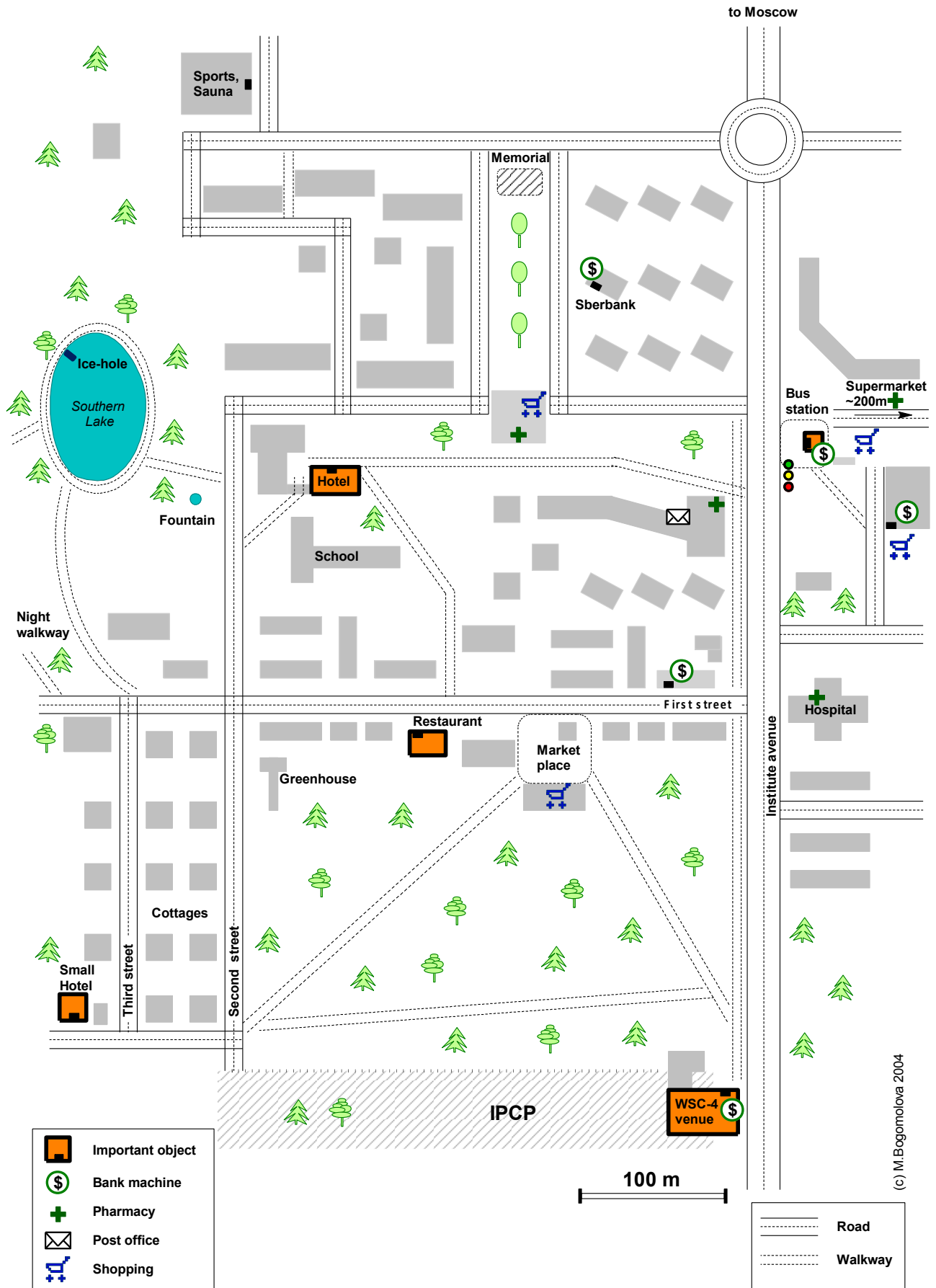
проспект]	
«ШКОЛЬНЫЙ БУЛЬВАР» [shkol'nyi bul'var]	School boulevard
«АВТОСТАНЦИЯ» [avto stantsiya]	Bus station
«СБЕРБАНК» [zber bank]	Sberbank, bank
«РЫНОК» [rynok]	Marketplace
«ХИМФИЗИКА» [khim fizika]	Chemphysics, Institute for Problems of Chemical Physics
«МАГАЗИН» [magazin]	Any shop, store, supermarket, etc.
«ВОДКА» [vodka]	Vodka
«ПИВО» [pivo]	Beer

Useful Phone Numbers

Codes: 7 – Russia, 095 – Moscow, 09652 – local area

Vladimir Rasumov, <i>conference chair</i>	7857046 (Moscow) 49676 (local)
Oxana Rodionova, <i>conference secretary</i>	+7 (903) 1504855 (cell)
Andrey Bogomolov, <i>local org. committee member</i>	+7 (916) 3464983 (cell) 42270 (local)
Alexey Pomerantseva, <i>conference manager</i>	+7 (903) 561 82 27 (cell)
Inna Tatarintseva, City Tour, travel agency	+7(095) 2587453 (Moscow)
City hotel	49266 (local) 42461 (local)
Small hotel	24270 (local)
Academia Club restaurant	24135 (local)
Taxi	48800, 43800, 46699, 49699 (local) +7 (905) 518 69 79 (cell)
Information	09 (local)
Ambulance	03 (local)
Rescue (from cellular phone)	911 (BeeLine); 112 (MTS)

Scheme of main conference places



Timetable

Monday, February 14, 2005

14:00–18:30 **Arrival, registration**

19:00–20:00 **Dinner**

20:30–00:00 **Scores & Loadings**

Tuesday, February 15, 2005

08:00–09:30 **Breakfast**

Session 1 Chair: Kim Esbensen

09:30–10:00 Conference opening

10:30–11:30 **L1** *Vladimir Razumov* Principal component analysis in photochemistry

11:00–11:30 **T1** *Juan Fernandez Pierna* Uncertainty in multivariate calibration: application to embedded NIR data

11:30–12:00 **Coffee break**

12:00–13:00 **Free time**

13:00–14:00 **Lunch**

Session 2 Chair: Kurt Varmuza

14:00–15:00 **L2** *Roma Tauler* Gray modeling approaches to investigate chemical processes

15:00–15:30 **T2** *Igor Vasiliev* Software package for optical emission spectrometry with arc discharge

15:30–16:00 **T3** *Elena Shabanova* Application of OES-AD software package for developing automated techniques

16:00–16:30 **Coffee break**

Session 3 Chair: Alexey Pomerantsev

16:30–17:00 **T4** *Andrew Kalach* Determination of nitromethane in the air with piezosensors array application

17:00–17:30 **T5** *Alisa Rudnitskaya* Comparison of PLS regression and artificial neural network for the processing of the electronic tongue data from fermentation growth media monitoring

17:30–18:00 **T6** *Andrey Bogomolov* Interactive series baseline correction algorithm

18:00–18:30 **T7** *Alexander Ryabenko* Application of principal component analysis and parameterized matrix self-modeling to the UV-Vis absorption spectra of the Zirconocene/Polymethylalumoxane catalytic systems

18:30–19:00 **Free time**
19:00–20:00 **Dinner**
20:00–00:00 **Scores & Loadings**

Wednesday, February 16, 2005

08:00–09:30 **Breakfast**

Session 4 Chair: John Kalivas
09:30–10:30 **L3** *Agnar Höskuldsson* The H-principle of mathematical modeling
10:30–11:00 **T8** *Sergei Kucheryavski* Multivariate image classification: comparison of AMT and transform based methods
11:00–11:30 **Coffee break**
11:30–13:00 **Free time**
13:00–14:00 **Lunch**

Session 5 Chair: Roma Tauler
14:00–15:00 **L4** *Kurt Varmuza* Exploration and classification: applications from archaeometry to spectroscopy
15:00–15:30 **T9** *Vladimir Palyulin* QSAR/QSPR: the universal approach to the prediction of properties of chemical compounds and materials
15:30–16:00 **T10** *Dmitry Bykov* Multilevel approach to analysis and processing of industrial wastes
16:00–16:30 **Coffee Break**

Session 6 Chair: Agnar Höskuldsson
16:30–17:30 **L5** *Alexey Pomerantsev* Simple view on Simple Interval Calculation
17:30–18:00 **T11** *Oxana Rodionova* Subset selection problem
18:00–18:30 **T12** *Sergei Zhilin* Simple method for outlier detection in fitting experimental data under interval error
18:30–19:00 **Free time**
19:00–20:00 **Dinner**
20:00–00:00 **Scores & Loadings**

Thursday, February 17, 2005

08:00–09:00	Breakfast
09:00–13:00	Excursion to Zvezdny Gorodok (training center of cosmonauts)
13:00–14:00	Lunch
Session 7	Chair: Lev Gribov
14:00–15:00	L6 <i>John Kalivas</i> Defining multivariate calibration model complexity for model selection and comparison
15:00–15:30	T13 <i>Leon Rusinov</i> The extended algorithm of adaptation of calibration models while transferring within series of IR-spectrometers
15:30–16:00	T14 <i>Boris Bravy</i> Remote recognition of aerosol chemicals
16:00–16:30	Coffee break
16:30–18:30	Poster Session
18:30–20:00	Free time
20:00–00:00	Conference banquet

Friday, February 18, 2005

08:00–09:30	Breakfast
Session 8	Chair: Oxana Rodionova
09:30–10:30	L7 <i>Lev Gribov</i>
10:30–11:00	T15 <i>Jaap van der Weerd</i> Controlled drug release from pharmaceutical tablets
11:00–11:30	Coffee break
Session 9	Chair: Vladimir Razumov
11:30–12:00	L8 <i>Kim Esbensen</i> Fundamental interrelationship between sampling-analysis-chemometrics
12:30–13:00	Conference closing. WSC-5 presentation
13:00–14:00	Lunch
14:00–16:00	Traveling to Moscow

Lectures

L1. Principal component analysis in photochemistry

Vladimir Razumov

Institute of Problems of Chemical Physics, Chernogolovka, Russia

The method of singular decomposition is one of the most effective way for treatment a spectrophotometric data, which very often use in photochemical experiment. In the present lecture an application of this method is shown by some examples. One of them concerns with the identification of the rotational isomers of flexible organic molecules from fluorescence and absorption spectra. In the others examples the potentialities of this method is illustrated on an investigation of different photochemical processes. From it follows that the method of singular decomposition makes possible revealing very fine details of kinetics and mechanism of photochemical reactions.

L2. Gray modelling approaches to investigate chemical processes

Roma Tauler

Institute of Chemistry and Environmental Research, CSIC, Barcelona, Spain

Data modeling and data fitting in chemical sciences has been traditionally done by hard modeling techniques, i.e., data are tested against a known model based on physical and chemical laws and the parameters of this model are obtained by least squares curve fitting optimization techniques. This approach may be also called *white modeling* and it is valid for well known phenomena and laboratory data, where the variables of the model are under control during the experiments and only the phenomena under study affect the data. However, this ideal situation is not obeyed in many circumstances in Chemistry, especially in Analytical Chemistry, when natural samples or unknown processes are investigated. Complex phenomena like those involving macromolecular compounds or industrial processes, where physical parameters cannot be appropriately fixed are typical examples not solved by the traditional *white model* based data treatments. Alternative approaches have been proposed. In particular, soft or *black modeling* approaches attempt the description of a system without the need of an a priori model postulation, physical or/and chemical. The goal of these methods is the explanation of data variance using the minimal or softer assumptions about data. Some of these soft modeling approaches are based on Factor Analysis decompositions of experimental data. These decompositions are done by pure mathematical means and allow the identification of the number of data variance sources and often their qualitative and, eventually, quantitative estimation. Results of soft modeling data analysis are useful to validate hard modeling results and also for investigation of complex chemical systems. Pure soft-modeling (black modeling) approaches however do not provide a full characterization and knowledge of the systems and therefore a mixed soft- hard modeling approach, the so-called *grey modeling* approach, is desirable. In this communication, some attempts related to grey modeling using Multivariate Curve Resolution will be shown.

L3. The H-principle of mathematical modeling

Agnar Höskuldsson

IPL, DTU, Kgs Lyngby, Denmark

The H-principle (the Heisenberg principle of mathematical modeling) is a recommendation of how mathematical models should be solved, when data are uncertain. It is recommended to carry out the modeling task in steps, where at each step the data (variables) are weighted according to their contribution to the aim of the modeling task. The primary interest in science and industry is the prediction aspect of the model. There are many ways to evaluate the prediction ability of a model. In the present work a cross-validation is used. 10% of the samples are selected at random and eliminated. The estimation of unknown parameters is carried out for the 90% of the samples. The response values of the 10% left out are then estimated using the estimated parameters. This is carried out 30 times. This procedure will tell fairly well of the prediction quality of the model.

Associated with the H-principle there have been developed a large collection of algorithms that have a common name of the 'H-method'. The H-method looks at the weight vectors and chooses the ones that contribute to the modeling task. The remaining weights are set to zero. In the case of linear regression the H-method gives results that have been optimized with respect to the prediction aspect of the model. Thus it gives better prediction than other traditional methods (PLS regression, Principal Component Regression, least squares regression and others).

Many methods in applied sciences use a positive definite matrix a starting point for finding the solution. Examples are Ridge regression, variance components in experimental design, Kalman filter methods and others. The H-method has been extended to these areas. It will be shown by examples that the H-method is superior to the traditional methods that use the full rank solution. Also, it is better to work with the solution obtained by the H-method, when carrying out the significance testing.

Estimation in non-linear models is traditionally carried out by the Gauss-Newton method, supplemented by Marquardt regularization. It will be shown by an example that the H-method gives better results than the traditional method. This holds both for the predictions derived from the model and also for the interpretation of parameters.

In many areas the classification of samples are important. New type of weight vectors have been defined such that the H-method can be used for classification of data. Applications of this method to spectral data have given superior results compared to traditional methods.

There is a tradition to use differentials and differential equations, when it is needed to describe changes of systems in time. The H-method has been extended to Path Modeling, where at each 'knot' (time point, status etc) there is given a data block, a matrix, that describes the situation. The H-method gives regression coefficients between the data blocks in a similar way as in linear regression, where there are only two data blocks. It is more efficient and more informative to use Path Modeling than the traditional approach by differential equations. An example of the usage of the H-method will be shown.

An important aspect of the H-method is that the vectors computed at each step have some special interpretations. Thus, using graphic analysis the user can get important insight into the latent variation of the data. These graphic features are available for all areas, where the H-method has been applied, both linear and non-linear models. The H-method has been extended to many areas of applied sciences. For orientation visit the author's personal website.

L4. Exploration and classification: Applications from archaeometry to spectroscopy

Kurt Varmuza

Vienna Technical University, Vienna, Austria

Evaluation of multivariate data often starts with exploratory data analysis by linear or nonlinear mapping with the aim to obtain a widely unbiased insight into the data structure. Multivariate classification is one of the roots of chemometrics. Applications in some recent projects will be discussed.

An example from archaeometry deals with fatty acid concentration data, measured on tissue samples from human mummies, including the 5000-years old Tyrolean Iceman, as well as 2000-years old corpses found in a permafrost area in Siberia. In another example, PLS mapping of concentration profiles of triterpenoids has been applied to obtain information about the botanical origin of a black glue material found on Neolithic weapons and tools.

Since March 2004 the Rosetta mission of the European Space Agency is cruising towards a comet. One of the instruments on board is a secondary-ion time-of-flight mass spectrometer (TOF-SIMS) dedicated to investigate comet dust particles. Currently available laboratory mass spectra from organic and inorganic reference compounds - presumably relevant for the analysis of comet material - have been evaluated by chemometric methods.

L5. Simple View on Simple Interval Calculation (SIC)

Alexey Pomerantsev, Oxana Rodionova

Institute of Chemical Physics, Kosygin Str. 4, 119991, Moscow, Russia

Keywords: multivariate calibration, interval estimations, SIC-method, linear programming, object status classification.

Simple Interval Calculation (SIC) is a method for linear modeling and for prediction interval estimation in the multivariate calibration (MVC) problem

$$y = \mathbf{X}\mathbf{a} + \boldsymbol{\varepsilon}.$$

It is shown that SIC leads to results that are in a convenient interval form, and which account for all uncertainties present (X measurement errors, y measurement errors, bilinear modeling errors). The SIC approach also provides wide possibilities for leverage-type object status classification. This method is based on the single assumption that all errors involved in MVC are finite. In this aspect SIC differs from traditional

chemometric methods used for multivariate data analyses and therefore is hardly apprehended by analytics.

In the presentation we discuss the finiteness of error. It is known, that assumption of normal error distribution is a commonplace in the conventional data analysis. Sometimes this is expressed implicitly, but often this is assumed by default. However, the researchers do not connect the normality of error distribution with its unboundedness. Does anybody take into account the data points, which are located beyond four standard deviations (4σ)? On the other hand, the amount of data in the modern data mining is often greater than 10^6 . Therefore, from the statistical point of view, there should be 20-30 values that lie beyond 4σ . Where are they? The answer is that if such values occurred, they are excluded before data processing. We consider that in a real case study it should be assumed that all error distributions are truncated on 4σ , or may be even on 3σ . Just this simple idea leads to the very drastic outcomes that throw a new light on the old MVC problem.

We are explaining the SIC approach using the simplest and familiar examples. Our goal is to introduce the SIC method in parallel with the traditional regression approach in order to emphasize both common and extraordinary SIC features. In the description we use a simulated data set that illustrates the SIC technique and adumbrate the main SIC concepts and postulates. We also demonstrate the SIC application to a well-known real world data set that is the octane rating procedure based on the near infrared spectroscopy.

L6. Defining Multivariate Calibration Model Complexity for Model Selection and Comparison

John Kalivas

Idaho State University, Pocatello, USA

In the analysis and comparison of multivariate calibration models, the concept of degrees of freedom (fitting degrees of freedom, prediction rank, model complexity, etc.) has an important role. This concept is often related to the number of respective basis vectors (latent vectors, factors, etc.) when using principal component regression (PCR) or partial least squares (PLS). Comparisons between PCR and PLS models for a given data set are often made with the prediction rank to determine the more parsimonious model, ignoring the fact that the values have been obtained using different basis sets. Additionally, it is not possible to use this approach for determining the prediction rank of models generated by other modeling methods such as ridge regression (RR). Measures are presented of what will be called the effective rank for a given model that can be applied to all modeling methods, thereby providing inter-model comparisons in terms of model complexity. With a proper definition of effective rank, a better assessment of degrees of freedom for statistical computations is possible. Additionally, the true nature of variable selection for improved parsimony over full variable models can be properly assessed. Spectroscopic and quantitative structure activity relationship (QSAR) data sets are used as examples with PCR, PLS, and RR.

L7. Lecture of Lev Gribov

Lev Gribov

Vernadsky Institute of Geochemistry and Analytical Chemistry, Moscow, Russia

L8. Fundamental interrelationships between sampling—analysis—chemometrics

Kim H. Esbensen, ACABS, Aalborg University Esbjerg (AUE), Denmark

Chemometrics is all about multivariate data analysis, data modeling, statistics etc. The issue of "data representativity" has never received proper attention however. What is the relevant breakdown of the total uncertainty embedded in empirical data? How does this uncertainty impact on data models and predictions? A very often met with myth is that the responsible agent(s) only causes measurement errors, which can be parsed in the statistical tradition, but this concept is never adequately defined nor analyzed in the proper context of material heterogeneity - and it therefore completely lacks the singularly most dominating factor - sampling errors. There is no satisfactory answer from within traditional chemometrics.

Analytical chemistry deals with accurate and precise determination of salient analyte concentrations etc. Within this science there is complete control of the attendant analytical uncertainty, but this is only a reflection of the miniscule analytical mass/volume however. What is the relationship to the lot from which the samples were taken? The lot is typically $10E3$ to $10E6$ times larger than the analytical mass - so the issue of how to conduct proper representative mass reduction comes to the fore, but there is no answer from within traditional analytical chemistry.

The Theory of Sampling (TOS) is the only complete scientific approach to representative sampling of heterogeneous materials. TOS provides a fundamental understanding of all the factors involved in producing sampling errors (there are seven fundamental types), which are by far the largest contributor to the total uncertainty budget of any analytical method. Typically the total sampling error are some 100 times larger than the analytical error(sic). In this context the specific analytical errors (which are often taken to equate the elusive "measurement errors") are reduced into oblivion! What is the impact of such overwhelming sampling errors on chemometrics and data analytical modeling and prediction etc?

It is actually necessary to invoke the science of metrology (the science of measurement) for a complete, systematic understanding of these complex interrelationships, but for a first-order introduction, this lecture presents a new multivariate perspective on the Theory of Sampling.

Talks

T1. Uncertainty in multivariate calibration: application to embedded NIR data

J.A. Fernández Pierna¹, V. Baeten², G. Sinnaeve², P. Dardenne²

¹*University of Agronomical Sciences, Statistics and Informatics Department, Gembloux, Belgium*

²*Walloon Agricultural Research Centre (CRA-W), Quality of Agricultural Products Department, Gembloux, Belgium*

The primary goal of using a regression model in multivariate calibration, for instance PLS, is to predict the value of a property of interest, the so-called predictand, and its uncertainty. The uncertainty of a calculated value is defined as a parameter, associated with the result of a measurement that characterizes the dispersion of the values that could reasonably be attributed to the measurand. In most of the cases this uncertainty is calculated as a function of the different sources of uncertainty present in the model.

In the univariate context, prediction uncertainty is quantified by a sample-specific standard error of prediction. Unfortunately, multivariate models are inherently much more complex than their univariate analogues. Monte Carlo simulation techniques such as the bootstrap and the noise addition method can give an estimate of this uncertainty but also some approximate mathematical expressions have been proposed in the literature. One of these proposals is the correction made by De Vries and Ter Braak (1995) on the expression derived by Martens (1989) and used in the Unscrambler[®] software package. Another proposal is the simplification of Faber and Bro (2002) of an expression derived earlier under the errors-in-variables (EIV) model (1996).

The purpose of this study is to show how these two proposals work and to assess their results from the ones obtained using the bootstrapping and the noise addition methods. This study has been performed using embedded near-infrared (NIR) data sets. This kind of data is produced by a NIR instrument installed on a forage plot harvester. Thanks to this instrument the collection, compression and scanning of forage samples is performed during harvesting. In such a way, properties as the dry matter content in forages, for instance, can be measured without having to handle the samples or transport them to a laboratory.

T2. Software package for optical emission spectrometry with arc discharge

I.L. Vasiliev¹, I.E. Vasilyeva² and E.V. Shabanova²

¹*Institute of System Dynamics and Control Theory, SB RAS, Irkutsk, Russia*

²*Vinogradov Institute of Geochemistry SB RAS, Irkutsk, Russia*

This work considers some mathematical models and algorithms, which are used in developing of expert system for automatic spectra processing in direct atom-emission analysis (AEA) of solid samples. The attraction of direct AEA consists of its universality and low price. Using visual interpretation, in a few minutes an experienced analyst

can enumerate 50-70 elements in the samples with unknown element structure. The visual interpretation allows making qualitative analysis and semiquantitative analysis with 10-100 % of relative error. The visual interpretation together with the instrumental measurement and computer data processing can provide the results of quantitative analysis with ≤ 30 % of relative error.

The process automation of direct AEA is a complex and insufficiently formalized problem. The spectrograph computerization allows us using 30-100 of spectral lines. But there is no a software, which provides us with handling of this exhaustive data in the knowledge base, like the analyst does. Analyzing the visual interpretation properties we can conclude that this procedure can automatized using the expert system paradigm. Actually, the direct AEA includes all the characteristics the expert system asks, i.e.:

1. An analyst has to solve a lot of diagnostic problems arising in the sample classification.
2. The complete and adequate mathematical models have not been developed for the direct AEA yet. Therefore, there is no a steady-state theory in this field.
3. Changing to the instrumental measurement leads that the small number of analysts remain to be able to make the visual analysis.
4. Noise pollution is a natural feature of the direct AEA data since it operates with samples of heterogeneous composition.

Our report discusses a general scheme of the expert system and we try to give mathematical models and algorithms for some expert system components. In particular, we describe a methodology for fitting the optimal parameters of analysis based on multi-objective optimization. We also consider questions concerned with the calibration using methods of multivariate analysis such that the principal compound analysis and neuron networks. This report focuses on the formal mathematical method description illustrated with practical examples.

T3. Application of OES-AD Software Package for Developing Automated Techniques

E.V. Shabanova¹, I.E. Vasilyeva¹, and I.L. Vasiliev²

¹*Vinogradov Institute of Geochemistry SB RAS, Irkutsk, Russia*

²*Institute of System Dynamics and Control Theory SB RAS, Irkutsk, Russia*

OES-AD SOFTWARE PACKAGE is a convenient and useful tool for creating automated techniques of multielement atomic-emission analysis of solid samples if there is multichannel spectrometer in laboratory.

This software contains knowledge base and database, which are connected with one another by calculation modules. Registered spectra of calibration, CRM and unknown samples are collected in database for every analytical technique. Knowledge base and calculation modules do rapidly the primary methodological studies such as:

1. selection of analytical lines (determination of upper and low bounds of intensity measuring);
2. selection of analytical parameter for each spectral lines (calculation method of intensity);
3. type selection of n -dimensional calibration for every analyte or its group.

Optimization of spectral information process is realized with performance criteria of analytical results.

OES-AD SOFTWARE PACKAGE is applied to obtain and process for techniques of impurity determining in quartz and crystalline silicon and silanes; Au, Ag and noble metals in nonsoluble carbonaceous substance from black shales; lanthanum and yttrium in barium fluoride crystals; boron and tin in granitoids.

T4. Determination of Nitromethane in the Air with Piezosensors Array Application

Andrew Kalach

Physical department, Institute of Ministry of Internal Affairs, Patriots ave., 53, 394065 Voronezh, Russia

The qualitative analysis of multicomponent gas mixtures with application of classical methods, as a rule, conjugates to known by difficulties that are connected to presence at analyzable test of many substances of unknown nature.

Responses of multi-sensor system to the exposure of nitromethane vapours in a wide range of concentrations (0.005 – 0.075 g/m³) have been investigated; this interval includes maximum permissible concentration of nitromethane in the working area (0.03 g/m³).

It is possible to make a precise determination of nitromethane concentrations in the air using a set of sensors composed of piezosensors. Responses from a set of sensors are applied to the input layer of artificial neural network. In order to reduce the error of determination the structures of an artificial neural network and its parameters have been optimized. Conditions of determination of nitromethane with the use of an artificial neural network have been found in the work: these are 7 neurons in the buried layer; impulse is 0.9; learning rate is 0.1; algorithm of preliminary processing is $\lg(|F_{gas} / F_{mod}|)$. To minimize an error of nitromethane determination learning algorithm for neural network has been applied called as “error backpropagation”.

A set of sensors is quite portable and it can be applied in the intelligence systems possessing abilities for independent estimation of the environment, detection of certain components, and their recognition and instruction issue for corresponding execution units.

T5. Comparison of PLS regression and Artificial Neural Network for the processing of the Electronic Tongue data from fermentation growth media monitoring

A. Legin¹, D. Kirsanov¹, A. Rudnitskaya¹, B. Seleznev¹, K. H. Esbensen², J. Mortensen³, L. P. Houmøller², Yu. Vlasov¹

¹ *Laboratory of Chemical Sensors, Chemistry Department, St. Petersburg University, Russia*

² *Applied Chemometrics, Analytical Chemistry, and Sampling Research Group (ACACSRG), Aalborg University Esbjerg, Denmark;*

³ *Department of Life Science and Chemistry, Roskilde University Centre, Denmark.*

Different kinds of biotechnological conversion processes, e.g. fermentations, are widely used in industrial scale for production of food, beverages, enzymes, organic acids, pharmaceuticals, biogas, etc. Continuous monitoring of biotechnological liquids is urgently needed for improving efficient management and optimization as well as quality control to obtain high standard final products. Despite rapid current development of biotechnology industry, cost effective, real-time and convenient process monitoring means are still lacking. Commonly used analytical techniques, such as HPLC, GC, IR or UV spectroscopy, are expensive, time-consuming, demand experienced operators and, as a rule, do not provide comprehensive real-time information about the process.

Of promising analytical tools for such applications are the 'Electronic Tongue' multi-sensor systems (ET). The idea of the electronic tongue is based on utilization of an array of cross-sensitive chemical sensors combined with multivariate data analysis. Such approach gives an opportunity of performing simultaneous quantitative determination of various substances in multi-component media as well as of integral monitoring (follow-up) of industrial processes as a whole. Since cross-sensitive sensors produce complex non-selective responses in multicomponent media, the use of proper data processing techniques is of paramount importance for successful application of the system. Different pattern recognition and multivariate calibration methods can be employed for these purpose, Principal Component Analysis (PCA) and Partial Least Square regression (PLS) being probably the most common. In the present paper two multivariate calibration methods that is PLS regression and Back-propagation artificial neural network (ANN) were applied and their performance was compared for the processing of the electronic tongue data from monitoring of the fermentation growth media.

Measurements with the electronic tongue were made in simulated fermentation media closely resembling real-world samples typical of the production process involving *Aspergillus niger* with the aim of simultaneous determination of ammonium, oxalate and citrate content. Two multivariate calibration techniques namely PLS regression and back-propagation neural network (ANN) were employed to produce calibration models. Optimization of the ANN configuration for the given task was performed, which included optimization of number of input signals and hidden neurons and choice of data preprocessing method. It was found that ANN produced somewhat better results than PLS in the data fitting, most likely due better consideration of significant non-linearity in the sensor responses particularly at low concentration levels of detected components.

The average prediction errors for independent test set solutions were in the range from 5 to 8% for the three target components. The electronic tongue shows promise for fermentation monitoring and industrial applications based on good precision and reproducible behavior of the sensor system and adequate data processing.

T6. Interactive Series Baseline Correction Algorithm

Andrey Bogomolov¹, Willem Windig², Susan M. Geer³, Debra B. Blondell³, and Mark J. Robbins³

¹ *Russian Chemometrics Society, Moscow, Russia*

² *Eigenvector Research Inc., East Coast Office, Rochester NY, USA*

³ *Eastman Kodak Company, Rochester NY, USA*

A new interactive approach to the baseline correction problem for hyphenated techniques has been suggested. It allows adapting traditional automated single-scan baseline correction routines or performing manual correction on bilinear data as if it were a single curve. Advantages of the method include “transparency” of the process and the means for extensive operator’s interaction. The method has passed a long-term testing in an industrial laboratory, and was integrated into a professional software package.

In spite of simplicity of the algorithm, it allows to successfully eliminate the background even in complex cases. The approach can be effectively applied to hyphenated chromatographic data such as HPLC-DAD or, more generally, to any hyphenated technique that apply a spectroscopic detector to control processes evolving in time. In the present work the method performance as a preprocessing stage for resolving mixture components is shown using several experimental datasets from two hyphenated techniques: HPLC-DAD and TGA-IR.

T7. Application of Principal Component Analysis and Parameterized Matrix Self-Modelling to the UV-Vis Absorption Spectra of the Zirconocene/Polymethylalumoxane Catalytic Systems

Alexander Ryabenko, E.E. Faingold, E.N. Ushakov, N.M. Bravaya

Institute of Problems of Chemical Physics, Chernogolovka, Russia

Keywords: Principal Component Analysis, Parameterized Matrix Self-Modelling

The changes in the UV-Vis absorption spectra of the zirconocene-based catalytic systems $\text{rac-Me}_2\text{Si}(2\text{-Me},4\text{-PhInd})_2\text{ZrCl}_2/\text{polymethylalumoxane}$ (MAO) and $\text{Ph}_2\text{CCpFluZrCl}_2/\text{MAO}$ in toluene, as observed upon the variation of the AlMAO/Zr molar ratio over the range 0-3000 mol/mol were studied using statistical methods. Application of principal component analysis made it possible to determine the number of the light-absorbing reaction products in each system as well as to ascertain the general trends in the chemical processes taking place in these systems. The UV-Vis absorption spectra of the intermediate reaction products were estimated using the parameterized matrix self-modelling method on the basis of the supposed reaction model.

T8. Multivariate Image Classification: Comparison of AMT and Transform Based Methods

Sergei Kucheryavski, Kim Esbensen

Altai State University, Barnaul, Russia

ACABS group, Aalborg University Esbjerg, Esbjerg, Denmark

Keywords: image processing and analysis, wavelet transform, Fourier transform, angle measure technique

Image processing and analysis is used in the wide range of scientific and industrial applications. Often, it is possible to obtain some information about different properties of materials, liquids, mixtures etc., with the help of analysis of their images bypassing the instrumental measurements. Thereby the development of image classification methods is very actual. Image classification methods, that used widely at the present time use different approaches: statistics, morphology, fractal analysis, Fourier and wavelet transforms and so on. But the choice of specific classification methods depends on the types of images very much.

In this work, the problem of using multivariate data analysis methods for image classification is considered. The results of classification of different types of images (heterogeneous, homogeneous non-periodic and homogeneous quasi-periodic) using Fourier and wavelet transforms and Angle Measure Technique are compared.

T9. QSAR/QSPR: the universal approach to the prediction of properties of chemical compounds and materials

Vladimir A. Palyulin, Igor I. Baskin, Nikolai S. Zefirov

Department of Chemistry, Moscow State University, Moscow, Russia

QSAR/QSPR (Quantitative structure-activity/property relationships) approaches can be considered as universal techniques for the modeling and prediction of nearly any properties of chemical compounds and many properties of materials. These approaches are based on the automatic analysis of structures and property values for a series of known chemical compounds with known properties, the chemical structures being described numerically with a series of parameters (descriptors). The structure-property relationships are usually evaluated using artificial neural networks. After creation of structure-property model it can be used for the prediction of properties for new chemical compounds for which these properties were never studied or compounds themselves were never synthesized. Some properties of materials can be predicted as dependent on the structure of small molecules used as additives (e.g. antioxidants, etc.). We have correlated successfully, e.g. the properties of tire rubbers with the structure of accelerators of vulcanization. Good results of modeling had been obtained for the diffusion coefficients of small molecules in some polymers. A number of properties of polymers had been modeled as dependent on the chemical structure of a monomeric unit (e.g. glass transition temperature, molar heat capacity).

A large number of properties were modeled for various organic compounds basing on their structural formulas, e.g. density, boiling points, viscosity, surface tension,

magnetic susceptibility, lipophilicity (octanol-water distribution coefficient), critical temperature, flash points, polarizability, enthalpy of evaporation, etc.

Prediction of toxicity is a challenging problem which until now is not completely solved. However the toxicity for many industrially important compounds can be predicted basing on the computation of lipophilicity, taking into account the presence of toxophores and using fragmental descriptors.

T10. Multilevel Approach to Analysis and Processing of Industrial Wastes

D.E. Bykov

Technical State University of Samara, Samara, Russia

Industrial wastes, if not treated properly, may be dangerous for the environment. Modern manufacturing processes often result in the waste of a complex multiphase composition, processing of which is complicated. The most expedient approach is an integrated utilization that converts the waste into raw materials, reagents or bulk additives.

A multilevel approach to the development of integrated waste utilization technologies has been suggested. It includes informational, physicochemical, technological, and integral blocks. Two basic principles are used for the creation of integrated utilization methods in the most complicated cases such as multicomponent heterogeneous wastes: the principle of phase redistribution and that of composite component correspondence. The principle of phase redistribution implies phase transformation processes resulting in the products, composition of which is satisfactory for further utilization of serves for neutralization of dangerous components. The principle of composite component correspondence requires that compounds of all material flows, when playing their roles in the process of the waste transformation or acting as a ballast, to fully correspond by their properties to the general transformation route and the quality of the final product. The multilevel approach to the problem of analysis and processing of wastes makes it possible to convert multicomponent heterogeneous toxic wastes into ecologically safe and utilizable products.

T11. Subset Selection Problem

Oxana Rodionova, Alexey Pomerantsev

Institute of Chemical Physics, Kosygin Str. 4, 119991, Moscow, Russia

Keywords: projection methods, SIC method, object status classification, representative subset selection.

Whenever samples are extracted from a large set of samples, the representativity of the subset should be achieved. It is known that representativity is a vague term, which can be interpreted in different way. The results for representativity depend not on the data itself but also on the working variable-subspace, which in its turn depends on the calibration model, i.e. the number of PLS-components. We will consider two different situations. First, a set of samples should be split into training and test sets and we want

to verify that these subsets are representative for each other. Different statistical tools are used for this purpose.

On the other hand, the test set is intensively used for calculation of the Root Mean Square Error of Prediction, which serves for evaluation of the prediction ability of the model. However, this quality measure should be used with a care as it greatly depends on the “quality” of this test set.

In the second case we want to select the most important objects among the training set and use this subset for model construction without significantly compromising the prediction ability of the model. Such a subset shall satisfy two opposing requirements: 1) it should be of maximal representativity with respect to the entire set, but 2) it should simultaneously be noticeably smaller than the total set. Here, the SIC object classification is used as a main tool. The presented results are compared with Kennard-Stone algorithm.

T12. Simple Method for Outlier Detection In Fitting Experimental Data Under Interval Error

Sergei Zhilin

Altai State University, Barnaul, Russia

Keywords: interval-bounded error, fitting experimental data, outliers, outlier detection, linear programming

The analysis of experimental errors shows that in both cases of direct and indirect measurements the resultant error should be considered as a bounded value. The model of interval-bounded errors has become a subject of intensive study in model building and design of experiment during last decades.

Using the "black box" model we assume that the measurement error for the output variable is interval-bounded and the errors for the input variables are negligible.

Each record in the table of measurements with bounded error in output variable allows us to formulate a constraint on model parameters. All the values of model parameters which are consistent with all constraints form the set of possible parameters values also called uncertainty set or informational set. The emptiness of the uncertainty set means inconsistency of the collected empirical information. The presence of outliers in the processed data is one of possible reasons of contradictions in dataset.

The core idea of the proposed outlier detection method is as follows. An outlier caused by a blunder may be treated as a value measured with the underestimated error, i.e. the actual measurement error is greater than the declared error. In order to correct the outlier, it is necessary to find out the lowest bound of its possible actual error, which makes the corrected observation consistent with the others. Comparing the values of the lowest bound of possible actual error to the values of the declared observation error allows us to make some inferences concerning the degree of inconsistency of outlier to the whole dataset.

In the case of the linear parameterized model the problem of finding the lowest bound of possible actual error may be stated as a linear programming problem.

The proposed technique is applied to solving the problem of geometric correction of satellite images.

T13. The extended algorithm of adaptation of calibration models while transferring within series of IR-spectrometers

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

Technological University & Lumex Company, St. Petersburg, Russia

The near-infrared (NIR) spectroscopy methods are widely used for the analysis of complex organic matters, in particular, in agriculture. However essential limitation on the wide use of this type of the analysis is connected with problems of large complexity of calibration and requirements of individual calibration of analyzers over the whole spectrum.

At present, this problem is avoided by fulfilling individual calibrations of analyzers. At the same time, the development of methods enabling the calibration models transfer within instruments of the same type essentially diminishes labor expenditures and, thus, allows widening ranges of application of NIR-analyzers. Now only one group out of different methods of calibration models transfer is used namely the methods, in which the spectra measured on a secondary instrument are corrected so that they conform to spectra measured on the master instrument. Then begins possible to use the calibration model of the master instrument at the secondary one. In all methods of calibration models transfer the spectra correction is executed over a partial set of samples that is significant smaller, than at usual calibration.

However, at such transfer in case of appearance samples of new types of organic matters (for example the wheat of new crop or new sort) in the laboratory, where the secondary instruments are installed, it is necessary to retry their calibration on the master instrument which is, as a rule, unfortunately placed very far from the laboratory (usually at the vendor place over hundreds kilometers from the laboratory).

Therefore it is offered to carry out the correction not of a spectrum of the secondary instrument for adaptation it to a spectrum of the master instrument, but on the contrary, to correct spectra of the master instrument, adjusting them to spectra of the same samples, measured on the secondary one. Then in case of appearance of new samples their spectra have to be simply measured on a secondary instrument and added to revised spectra of the master instrument delivered by the vendor together with calibration model. The calculation of improved calibration model is carried out then by a method of PLS or PCA over this aggregated set of spectra. The results of research of the metrology characteristics of the considered algorithm of adaptation of calibration models are discussed.

T14. Remote recognition of aerosol chemicals

B. Bravy, V. Agroskin, G. Vasiliev

Institute of Problems of Chemical Physics, Chernogolovka, Russia

Keywords: multifrequency lidar, aerosol recognition, genetic algorithm

The recognition of composition and of microphysical characteristics of aerosol impurities is one of the most urgent tasks in atmosphere monitoring. Physical basics giving an opportunity of aerosol identification in remote sensing and some methods of solving this problem are considered. The concrete results of recognition and the dependence of recognition efficiency on the number of frequency channels and on signal to noise ratio in received multifrequency lidar signals are given.

T15. Controlled drug release from pharmaceutical tablets

Jaap van der Weerd

Imperial College London, London, UK

The current standard test for drug release from pharmaceutical formulations is the dissolution test, in which a tablet is immersed in water and the concentration of released drug in this water is monitored. In recent years, a growing number of imaging techniques have been applied to tablets to improve the understanding of processes inside the tablets rather than only the final results of these processes, i.e. drug release.

We will explain the application of Fourier transform infrared-attenuated total reflection (FTIR - ATR) imaging in tablet research, and show a number of applications. The emphasis will be on the coupling of FTIR - ATR imaging with conventional dissolution test, the behavior of soluble and poorly soluble drugs in tablets, the application of different ATR crystals, and of course the multivariate techniques used to process to large quantities of data obtained.

Posters

P1. How long ago chemometrics was appeared in Russia?

Granovsky Y.V., Markova E.V.

Moscow State University, Moscow, Russia

Keywords: development domestic chemometrics, infrastructure new directions, the Scientific Council of problem

The paper deals with the questions of development domestic chemometrics within two decades (1960-1970-es). In these years an infrastructure such new directions were supporting by the Scientific Council of problem cybernetics of the Academy of Science USSR.

The great informal collective successfully working in different directions has been created: application of new methods in laboratory and industrial researches, extensive publishing and educational activity, transformation of the education system in chemical higher schools, etc. Applied statistical methods were regularly used in analytical, organic, physical and in other areas of chemistry. Really it was one, that named as chemometrics today.

P2. Multisensor analysis of amino acids

Igor Aristov

Voronezh State University, Russia, Voronezh

In this work, two main tasks are performed: 1) multivariate physicochemical analysis of lysine-glycine water solution and 2) multivariate calibration for specimen's analysis of this solution.

Chemometrics methods appropriate to these tasks were used: MANOVA, multiply regression, factor analysis and Structural Equation modeling of Latent variables (SEPATH, LISREL).

In mixed aqueous solution of glycine and lysine the ionization of both amino acids increases due to proton transfer from glycine zwitterions to lysine zwitterions. This is manifested by a considerable nonlinear (pairwise interaction of amino acids) increase in the mixed solution electroconductivity at high amino acid concentration. The solution pH varies negligibly, as the amino acids in aqueous solutions demonstrate buffering. The refraction coefficient additively depends on glycine and lysine concentrations.

P3. Analysis of apples varieties – comparison of Electronic Tongue with different analytical techniques

A.Legin¹, D. Kirsanov¹, A.Rudnitskaya¹, K.Beullens², J. Lammertyn², B. Nicolai², J. Irudayaraj³, Yu. Vlasov¹

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

²*Department of Agrotechnique and Economics, Catholic University of Leuven, Heverlee, Belgium*

³*Agricultural and Biological Engineering Department, Pennsylvania State University, PA, USA*

Quality assessment and classification of food products is necessary in the modern markets for buyers and producers alike. Conventional analytical techniques used for such measurements are often time-consuming, expensive and cannot be used in the field. The present study was aimed at application of rapid analytical techniques such as Electronic Tongue and FTIR spectroscopy to recognition and quantitative analysis of different apple varieties.

Five varieties of apples were studied using three different analytical techniques: HPLC, electronic tongue multisensor system based on potentiometric chemical sensors and FTIR spectroscopy. Twenty samples (apples) of each variety were measured. Juice was pressed from each fruit, clarified and deep-frozen before measurements. Juice samples were stored at -80°C and thawed before measurements. Concentrations of organic acids such as malic, citric, galacturonic etc. and sugars were measured by HPLC, which is a conventional method for fruit analysis in such case. HPLC data were used as reference for calibration of the electronic tongue and also FTIR.

Different aspects of data processing were addressed. Recognition of the apples according to the variety using data from three different analytical instruments was performed by PCA and PLS discrim. Quantitative calibration of the electronic tongue and FTIR with respect to organic acids and sugars content was done using PLS regression. Issues of obtaining complimentary information from electronic tongue and FTIR spectroscopy as well as merging of the data of different nature were considered.

Acknowledgements

This work was partly supported by the NATO Linkage Grant.

P4. Data Warehousing and Data Mining In Thermochemistry of Free Radical Reactions

V. E. Tumanov

Institute of Problems of Chemical Physics RAS, Chernogolovka

Keywords: Data warehouse, bond Dissociation energies, rate constants, free radical reactions, expert system

Chemists have accumulated a vast amount of data on rate constants of free radical reactions. The bond dissociation energy of organic molecules represents one of the important characteristics of compounds of radical reactions. These data can be structured and are assembled in electronic data collections.

The concepts Data Warehouse can be used for representation of experimental data radical kinetic and thermochemistry of organic connections. The basic assignment Data Warehouse of the scientific data is an organization and support given for processing with the purpose of extraction of the new data or generalization available.

The storehouse given on kinetic of radical reactions was developed and thermochemistry of organic connections (cumulative volume is estimated in 40000 records). It allows to put and to solve tasks of the statistical and comparative analysis of the kinetic data on various groups of such reactions (Knowledge Discovery).

Knowledge Discovery in data warehouse is a process of non-trivial extraction of implicit, previously unknown and potentially useful data about BDE organic compounds and its reactivity. Data mining is a step such process.

For the analysis of the kinetic data was used patterns - empirical model of crossed parabolas prof. E. T. Denisova. Within the framework of this model enthalpy of radical abstraction reaction of atom of hydrogen is connected to energy of activation of reaction by parabolic dependence.

On the basis of this model the expert system (ES) for an estimation of BDE of organic molecules on rate constants of radical abstraction reactions was developed. Were calculated of BDE more than for 500 organic connections.

One more ES allows to predict reactionary ability compounds in elementary radical abstraction reaction. The testing ES on known experimental data has shown satisfactory results.

Both expert systems work on one knowledge base and use a small set of production rules. Both the knowledge base and data warehouse is realized as set of databases of special structure.

Thus, the created infrastructure of databases allows to put a task of development information portal for support of knowledge management in thermochemistry of radical reactions.

P5. Optimization of a Structure of a Composite Material by a Method of Data Analysis

Sulejmanov A.M.

Kazan State Academy of Architecture and Civil Engineering

Durability and the operating performances of composite materials depend on a set of the external factors and internal parameters of a material. To the external factors it is possible to refer various climatic effects in their diverse combination and sequence. To internal parameters concern variety of various structures of composite materials. On an example of coated textiles for soft envelopes the attempt is made to apply a Method of Data Analysis to optimization of a structures of composite materials. For these purposes the computer structural - simulation model of an composite material with varied parameters of a structure was developed. On the developed model the numerical experiments on forecasting durability of an composite material for want of various parameters of a structure were conducted. The analysis was further made with the purpose of revealing the most significant parameters of a structure on durability of an composite material.

P6. The multisensor system for voltammetric analysis of multi-component mixtures of aromatic nitrocompounds

Sidelnikov A.V., Maystrenko V. N., Kudasheva F. Kh., Kuzmina N.V., Sapelnikova S.V.

Bashkir State University, Ufa, Russia

Keywords: multisensor system, voltammetry, nitrocompounds, “Maple” package, modeling, polynomial models, film coated electrodes

The majority of studying environmental objects are complex systems. One of the research problems of such a systems lies in the determination of the relationship between system components' concentration and analytical signal. The traditional ways of solution of this problem is not always possible because of insufficient selectivity of the sensors. The development of chemometric multivariate data processing allows application of non-selective sensor arrays to the analysis of complex solutions. In this work multisensor system for voltammetric analysis of multi-component mixtures of aromatic nitrocompounds was developed. With the help of “Maple” package the algorithm for processing of sensor arrays responses was approbated. It includes the following steps: the construction of polynomial model of voltammetric behavior of nitroaromatic compounds mixtures, models testing and simultaneous determination of the mixture components. The multisensor system allows the quantitative determination of three-components mixtures with the concentration ratios (one component to the sum of two others) from 1:1 up to 1:5. The advantages of the suggested method are the fast analysis, low cost of analysis and data processing, versatility and the possibility of use for the classification of the analytical objects.

P7. Mathematic modeling of swelling degree for plastified rubbers in oil

Petrova N.N.

Institute of Non-Metallic Materials, Yakutsk, Russia

Injection of large amounts of plasticizers into rubber mixtures is one of the most widespread methods to develop rubbers having increased level of frost resistance. However, in-cite testing of such rubbers showed that simultaneous impact of low temperatures and oil causes intense washing out of plasticizers, resulting in sharp degradation of low-temperature characteristics. Model systems (rubber based on butadiene-nitrile rubber containing different amounts of dibutylphtalate) made it possible to control rubber swelling kinetics in oil, to determine coefficients of diffusion of hydrocarbons and the plasticizer, to reveal the influence of the plasticizer dose on glass-transition temperature of the elastomer material after its exposition in the hydrocarbon medium. Mathematic modeling applied the model of multi-component diffusion and "FITTER" software.

Investigations of swelling degree for rubbers containing 5 to 20 wt per cent of dibutylphtalate performed extreme character of dependence of diffusion coefficients on the amount of injected plasticizer, which allowed optimizing the material composition aiming at reduction of negative impact of hydrocarbon media on the properties of the materials. It is necessary to note that application of this approach (investigation of

rubber swelling in hydrocarbon medium at room temperature and determination of diffusion coefficients by means of "FITTER" software) can be very useful while developing new recipes of rubbers performing controlled diffusion rate and swelling degree.

P8. Prediction of the heating value of plant biomass from the elemental composition

A. Friedl, E. Padouvas, H. Rotter, K. Varmuza

Vienna Technical University, Vienna, Austria

The heating value of biomass is an important parameter for the design and the control of power plants using this type of fuel. The so called higher heating value, HHV, is the enthalpy of complete combustion of a fuel including the condensation enthalpy of formed water.

The 154 biomass samples, available for this study, have very different origin, such as wood, grass, reed, brewery waste, or poultry litter. Each sample has been characterized by the contents of carbon, hydrogen, nitrogen, oxygen, sulfur, chlorine, and ash. PCA of these data shows a good clustering according to the origin of the samples.

A subset of 122 samples, all consisting of plant materials, has been used to develop regression models for a prediction of HHV from the elemental composition. OLS- and PLS-models with best predictive ability have been obtained using the contents of carbon, C, hydrogen, H, and nitrogen, N, with the variables defined by C, C*C, H, C*H, and N. The standard errors of prediction of the new models are considerably smaller than those obtained with the many models reported in literature.

P9. What is a right approach to the educational process in area of chemometrics in universities?

Yu. Adler, Yu. Granovsky

Moscow Institute of Steel and Alloys, Moscow, Russia

Chemometrics is need for chemists. The methods of chemometrics have a positive impact in all the characteristics of scientific production. There are some difficulties in the way to implement a systematic and full course of chemometrics in Russian`s higher schools. First from the troubles is strong need to understand for students the probabilistic approach to interpretation of the World. The second trouble which have real beginning is need to teach of students so variety of topics. We suggest two-stage process of chemometrics education in higher schools. Introductory course (15-20 hours) is illustration of broad application and effectiveness the chemometrics. Basic course (more than 100 hours) is include lectures and seminars with multimedia. In education of chemometrics can sense introduce of Total Quality Management approach.

P10. Program complex for simulation of chemical reactions in micellar systems

Vladimir Razumov, Galina Lubimova and Anastasia Losikova

Institute of Problems of Chemical Physics, Chernogolovka, Russia

In present report we submit the program complex for statistical modeling kinetics of chemical reactions of an ion exchange and initial stages of nucleation and growth of nanosize particles which are created as a result of these reactions in the micellar solutions and reverse microemulsions. The algorithm of the program complex is based on Monte-Carlo method. The ensemble consisting of 105-106 subsystems, each of which includes about 100 micelles, is considered. In the subsystems the certain number of the acts of micellar impact are simulated. In everyone the random exchange of reacting ions takes place. After the predetermined number of impacts an averaging on all subsystems of ensemble is made, and by this means we calculate the time-depending micellar distribution of reagents and nanoparticles. The computation results are compared with the experimental data of silver iodide nanoparticles growth in reverse micelles AOT and gold nanoparticles in water/oil microemulsion stabilized by Triton-X100.

This work is supported by Russian Foundation for Basic Research, grant 04-03-32177.

P11. Application of two-channel FTIR spectrometer ISD-206 to dairy produce quality control.

Boldyrev N.U., Vinogradov E.A., Kalinin A.V., Krivtsun V.M., Sadovsky S.I.

Institute of spectroscopy RAS,

The dairy quality control systems consists: first, an exit testing equipments for fat, protein and lactose quantifying in each party of raw milk, second, an adulterants classification technique.

Two models of portative two-channel hi precision FTIR spectrometers for near and middle spectral range were developed and tested by the Institute of Spectroscopy at 2003. The main parameters of them are shown in the site: www.isan.troitsk.ru

A method of middle range FTIR-ATR spectrometry with fiber-optic sensor and ZnSe cuvette was used for previous study of fat, protein and lactose simultaneous quantifying procedure in raw milk and fat, soya and palm oils identification procedure. The calibration was based on series of standard samples, produced by Institute of dairy industry, RAAS, and adapted statistical regression analysis. It is concluded that FTIR-ATR method offers a simple and efficient quality control method of milk analysis.

List of Participants

Name	Affiliation	Contacts
Adler, Yuri Адлер, Юрий	Moscow Institute of Steel and Alloys Moscow, Russia Professor	+7 (095) 954 47 20 Adler@astroinform.ru
Aristov, Igor АРИСТОВ, Игорь	Voronezh State University Voronezh, Russia Assistant	+7(0732) 49 04 09 igor@chem.vsu.ru
Bogomolov, Andrey БОГОМОЛОВ, Андрей	Russian Chemometrics Society Moscow, Russia	a.bogomolov@mail.ru
Bravy, Boris БРАВЫЙ, Борис	Institute of Problems of Chemical Physics Chernogolovka, Russia Laboratory manager	bgbrav@icp.ac.ru
Brichkin, Sergei БРИЧКИН, Сергей	Institute of Problems of Chemical Physics Chernogolovka, Russia Department manager	brichkin@icp.ac.ru
Bud, Anton Будь, АНТОН	Donetsk University Donetsk, Ukraine Ph.D. Student	+380 (44) 428-0831 nadc@i.kiev.ua
Bud, Vladimir Будь, Владимир	National Antidope Center Kiev, Ukraine Director	+380 (44) 428-0831 nadc@i.kiev.ua
Budyka, Mikhail Будыка, Михаил	Institute of Problems of Chemical Physics Chernogolovka, Russia Laboratory manager	+7 (096) 522-1265 budyka@icp.ac.ru

Вуков, Dmitri БЫКОВ, ДМИТРИЙ	Technical University of Samara Samara, Russia Head of the Chair of Chemical Technology and Industrial Ecology	
Chulyukov, Oleg Чулюков, Олег	Lumex-Centrum company Russia Laboratory manager	Chulukov@newmail.ru
Esbensen, Kim Эсбенсен, КИМ	Ålborg University ACACS-RG Esbjerg, Denmark Professor	+45 7912 7691 kes@aaue.dk www.acabs.dk
Fernandez, Juan Фернандес, Хуан	University of Agronomical Sciences Gembloux, Belgium Researcher	+32 (0)81 62 03 55 fernandez@cra.wallonie.be
Granovsky, Yuri Грановский, Юрий	Moscow State University Moscow, Russia Researcher	zpch@rambler.ru
Gribov, Lev Грибов, Лев	Institute of Geo and Analytical Chemistry Moscow, Russia Professor	+7 (095) 1374147 elkor@geokhi.msk.su www.geokhi.ru
Hoskuldsson, Agnar Хоскульдсон, Агнар	Denmark Technical University Kgs. Lyngby, Denmark Professor	ah@ipl.dtu.dk www.predict.ws
Julius, Anne Джулиус, ЭНН	Gørding, Denmark Accompanying person	lars@julius-petersen.dk
Kalach, Andrew Калач, Андрей	Ministry of Home Affairs Institute Voronezh, Russia Lecturer	+7 (073)2 47 64 85 a_kalach@mail.ru

Kalivas, John Калайвас, Джон	Idaho State Univeristy Pocatello, USA Professor	+1 (208) 282-2726 kalijohn@isu.edu
Kirsanov, Dmitry Кирсанов, Дмитрий	St. Petersburg State University St. Petersburg, Russia Ph.D. student	+7 (812) 3282835 eth500@mail.ru
Korobova, Elena Коробова, Елена	Vernadsky Institute of Geo and Analytical Chemistry Moscow, Russia Science secretary	+7 (095) 137 24 84 Korobova@geokhi.ru
Kucheryavski, Sergei Кучерявский, Сергей	Altai State University Barnaul, Russia Docent	+7 (3852) 363571 svk@asu.ru
Kurandina, Mariya Курандина, Мария	Institute of Problems of Chemical Physics Chernogolovka, Russia Junior Researcher	max2004@icp.ac.ru
Lobach, Anton Лобач, АНТОН	Moscow State Institute of Steel and Alloys Chernogolovka, Russia Ph.D. Student	+7 (252) 233-93 lobach@icp.ac.ru
Lobach, Anatoly Лобач, АНАТОЛИЙ	Institute of Problems of Chemical Physics Chernogolovka, Russia Senior Researcher	+7 (252) 233-93 lobach@icp.ac.ru
Marks, Christopher Маркс, Кристофер	University of Virginia Charlottesville, USA Laboratory Supervisor	+1(434) 982-5794 cam8h@virginia.edu www.virginia.edu/cese/
Marks, Laurel Маркс, Лорел	USA Accompanying person	laurelandchris@adelphia.net

Osipov, Vadim ОСИПОВ, Вадим	Miross company Nizhni Novgorod, Russia Director	+7 (312) 77 52 99 vad@miross.ru
Palyulin, Vladimir Палюлин, Владимир	Moscow State University Moscow, Russia Leader Researcher	+7 (095) 9393557 vap@org.chem.msu.su
Pastukhov, Alexander Пастухов, Александр	Institute of Problems of Chemical Physics Chernogolovka, Russia Researcher	avp31@icp.ac.ru
Pedersen, Carina Педерсен, Карина	Aalborg University Esbjerg Esbjerg, Denmark Ph.D. student	+45 79127702 cjp@aaue.dk
Penkova, Lyudmila Пенкова, Людмила	National Antidope Center Kiev, Ukraine Expert	+380 (44) 428-0831 nadc@i.kiev.ua
Petersen, Lars Петерсен, Ларс	Ålborg University ACABS Esbjerg, Denmark Assistant Professor	+45 79127691 lape@aaue.dk www.acabs.dk
Petrova, Natalya Петрова, Наталья	Institute of Non- Metallic Materials Yakutsk, Russia Laboratory manager	pnn2002@mail.ru
Pisarev, Evgeniy Писарев, Евгений	Ukrainian research institute of al- cohol and biotechnology of foods Kiev, Ukraine Department manager	nadc@i.kiev.ua
Pomerantsev, Alexey Померанцев, Алексей	Institute of Chemical Physics Moscow, Russia Senior Researcher	+7 (095) 9397483 forecast@chph.ras.ru rcs.chph.ras.ru/ICP/paleng.htm

Pomerantseva, Катя Померанцева, Екатерина	Institute of Gene Biology Moscow, Russia Researcher	+7 (095) 135-9734 katy.sol@mail.ru
Razumov, Vladimir Разумов, Владимир	Institute of Problems of Chemical Physics Chernogolovka, Russia Deputy director	+7 (095) 785-7046 razumov@icp.ac.ru
Rodionova, Охана РОДИОНОВА, Оксана	Institute of Chemical Physics Moscow, Russia Senior Researcher	+7 (095) 9397483 rcs@chph.ras.ru rcs.chph.ras.ru/ICP/royeng.htm
Rudnitskaya, Alisa Рудницкая Алиса	St. Petersburg University, St. Petersburg, Russia Researcher	+7(812)282835 eth500@mail.ru www.electronic tongue.com
Rusinov, Leon Русинов, Леон	Technological University St. Petersburg, Russia Professor	+7 (812) 316-3526 rusinov@tztel.spb.ru
Рябенко, Александр Ryabenko, Alexander	Institute of Problem of Chemical Physics Chernogolovka, Russia Researcher	+7 (096) 2522 2514 aryabenk@icp.ac.ru
Sadovsky, Sergei Садовский Сергей	Institute of Spectroscopy Troitsk, Russia Student	kalinin@isan.troitsk.ru
Shabanova, Elena Шабанова, Елена	Institute of Geochemistry Irkutsk, Russia Researcher	+7 (395)2 42 92 40 shev@igc.irk.ru
Shireeva, Dilyara Ширеева, Диляра	Kazan State Academy of Archi- tecture and Civil Engineering Kazan, Russia Senior Lecturer	+7 (843)2 72 80 09 sulejmanov@ksaba.ru

Sidelnikov, Artem Сидельников, Артём	Bashkir State University Ufa, Russia Assistant	+7 (347)2 23-67-21 artsid2000@mail.ru
Spirin, Maxim Спирин, Максим	Institute of Problems of Chemical Physics Chernogolovka, Russia Researcher	max2004@icp.ac.ru
Spitsyna, Natalia Спицына, Наталья	Institute of Problems of Chemical Physics Chernogolovka, Russia Senior Researcher	+7 (252) 233-93 spitsina@icp.ac.ru
Suleimanov, Alfred Сулейманов, Альфред	Kazan State Academy of Archi- tecture and Civil Engineering Docent Russia, Kazan	+7 (843)2 72 80 09 sulejmanov@ksaba.ru
Sulima, Elena Сулима, Елена	Lumex company St. Petersburg, Russia Researcher	stasenkoel@lumex.ru www.lumex.ru
Tauler, Roma Таулер, Рома	Institute of Chemistry and Envi- ronmental Research Barcelona, Spain Professor	+34 93 4006140 rtaqam@iiqab.csic.es
Tumanov, Vladimir Туманов, Владимир	Institute of Problems of Chemical Physics Chernogolovka, Russia Researcher	+7 (096) 522-5156 tve@icp.ac.ru
Ushakov Evgeny Ушаков, Евгений	Institute of Problems of Chemical Physics Chernogolovka, Russia Senior Researcher	+7 (096) 524-4346 eushakov@icp.ac.ru
Varmuza, Kurt Вармуза, Курт	University of Technology Vienna, Austria Professor	+431-58801-16060 kvarmuza@email.tuwien.ac.at www.lcm.tuwien.ac.at

Vasiliev, Igor Васильев, Игорь	Institute of System Dynamics and Control Theory, SB RAS Irkutsk, Russia Senior Researcher	+7 (395)2 51 13 98 vil@icc.ru
van der Weerd, Jaap ван дер Вирд, Йапп	Imperial College London, UK Research Associate	+442075895111 ext 55671 jvdw@ic.ac.uk www.ic.ac.uk
Yaroshenko, Mariya Ярошенко, Мария	National Antidope Center Kiev, Ukraine Expert	+380 (44) 428-0831 nadc@i.kiev.ua
Zhilin, Sergei Жилин, Сергей	Altai State University Barnaul, Russia Docent	+7 (385)2 36 70 18 sergei@asu.ru