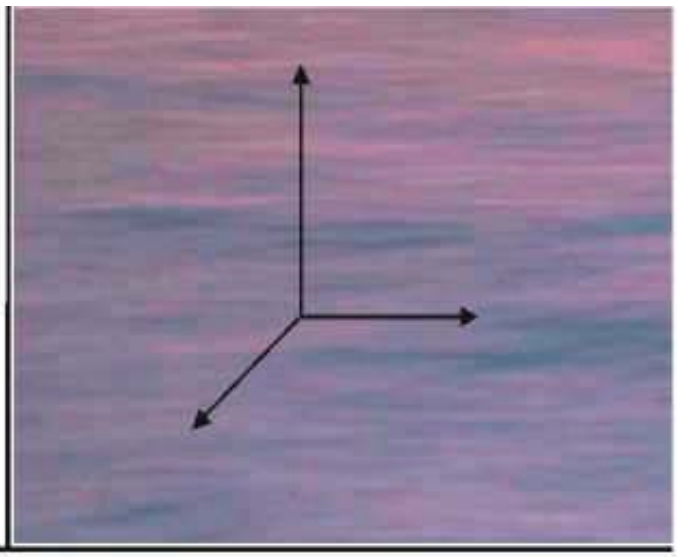
SSC7

Copenhagen

August 19 - 23

2001



7th Scandinavian Symposium on Chemometrics

BOOK

OF

ABSTRACTS

7TH SCANDINAVIAN SYMPOSIUM

ON

CHEMOMETRICS

COPENHAGEN, DENMARK, 19–23 AUGUST 2001

Committees

Organizing Committee

Lars Nørgaard (Chairman)

Carsten Ridder

Claus A. Andersson

Kaj Heydorn

Ivan Christensen

Joan Grønkjær Pedersen

Rolf Singer

Scientific Committee

Olav M. Kvalheim, Bergen

Pentti Minkkinen, Lappeenranta

Michael Sjöström, Umeå

Agnar Höskuldsson, Lyngby

Kim H. Esbensen, Esbjerg/Porsgrunn

Sponsors

- H. Lundbeck A/S
  http://www.lundbeck.com
- Foss Electric A/S
  http://www.foss.dk
- Radiometer Medical A/S
  http://www.radiometer.com
- Danish Meat Research Institute
  http://www.dmri.dk
- City of Copenhagen
  http://www.kbhbase.kk.dk
  (Copenhagen International)

Contents

General Information

Dates

2001 August 19-23

Venue

Ingeniørforeningen i Danmark
Kalvebod Brygge 31-33
DK-1780 København V

Secretariat and General Information

Birgitte Magnér-Egeberg
Ingeniørforeningen i Danmark
Kalvebod Brygge 31-33
DK-1780 København V
Telephone +45 3318 4848
Fax +45 3318 4899
E-mail: bme@ida.dk
Website: http://www.ssc7.dk

Language

The symposium language is English

Publication

The proceedings will be published as a special issue of *Journal of Chemometrics*, John Wiley & Sons, Ltd.

Registration during the Symposium

Sunday August 19: 15.00 – 18.00
Monday August 20: 08.00 – 17.00
Tuesday August 21: 08.00 – 09.00
Wednesday August 22: 08.00 – 09.00
Thursday August 23: 09.00 – 09.40

Social Programme

Sunday, August 19	19.00 Get-together at Tycho Brahe Planetarium starting with an Omnimax Show (1h) followed by a reception & buffet
Tuesday, August 20	18.00 Welcome reception & buffet at the City Hall of Copenhagen
Wednesday, August 21	17.00 Symposium banquet at Trekroner fortress including the Herman Wold Medal Award Ceremony. Boats leave from the IDA building and return (several departures) to the same place later that night

Scientific Programme

The scientific programme involves

- Invited lectures
- Contributed lectures
- Poster session

The subjects are

- Applications
- Process Control
- New Algorithms
- Variable Selection
- Uncertainty
- QSAR
- Multi-way Methods
- New and Old Trends in Chemometrics

Conference Programme

Monday August 20, 2001

08.00 – 09.40 Registration
09.40 – 10.00 Official Opening of the Symposium including Practical Details

Opening lecture
Chairman: Lars Nørgaard

10.00 – 10.40 **Lars Munck**
Exploratory data analysis - Addressing the context in which chemometrics work (A1)

10.40 – 11.20 Coffee break

Applications
Chairman: Kim Esbensen

11.20 – 11.40 **Christian Airiau**
Resolution of proton LCNMR by multivariate means (A2)

11.40 – 12.00 **Harald Martens**
Multivariate analysis of quality (A3)

12.00 – 12.20 **Maurice O´Sullivan**
The reliability of naïve assessors in sensory evaluation using multivariate data analysis (A4)

12.20 – 13.20 Lunch

13.20 – 14.20 **Poster Session (Posters 1-20)**

Process Control I
Chairman: Dora Kourti

14.20 – 15.00 **Age Smilde**
A framework for multiblock component models (A5)

15.00 – 15.20 **Elaine Martin**
Process performance monitoring in the presence of confounding variation (A6)

Monday August 20, 2001 (continued)

15.20 – 15.40 **Jonas Röttorp**
Data mining issues on modern multivariate online industrial process control (A7)

15.40 – 16.20 Coffee break

QSAR
Chairman: Bjørn Alsberg

16.20 – 16.40 **Lennart Eriksson**
Multivariate biological profiling and constrained principal response profile regions of PCB´s (A8)

16.40 – 17.00 **Tarja Rajalahti**
Analysis of short protein sequences using multivariate batch modelling (A9)

17.00 – 17.20 **Rudolf Kiralj**
QSAR of progestogens: Use of a priori and computed molecular descriptors and molecular graphic (A10)

17.20 – 17.40 **Torbjörn Lundstedt**
A combinatorial approach for drug discovery - based on multivariate methods (A11)

Conference Programme

Tuesday August 21, 2001

New Algorithms I
Chairman: Barry Wise

09.00 – 09.40 **Klaas Faber**
Use of Monte Carlo simulations in chemometrics (A12)

09.40 – 10.00 **Anton Belousov**
Applicational aspects of support vector machines (A13)

10.00 – 10.20 **Rolf Ergon**
Static PLSR optimization based on Kalman filtering theory and noise covariance estimation (A14)

10.20 – 11.00 Coffee break

11.00 – 11.20 **Rasmus Bro**
MILES: A general approach to maximum likelihood estimation (A15)

11.20 – 11.40 **Satu-Pia Reinikainen**
Strategies in modelling dynamic systems (A16)

11.40 – 12.00 **Martin Høy**
Combining bilinear modeling and ridge regression (A17)

12.00 – 12.20 **Per Anker Hassel**
Non-linear partial least squares (estimation of the weight vector) (A18)

12.20 – 13.20 Lunch

13.20 – 14.20 **Poster Session (Posters 21-37)**

Multi-way Methods
Chairman: Claus Andersson

14.20 – 15.00 **Olav Kvalheim**
Automated curve resolution of multi-way data and prediction of biological properties from the resolved profiles (A19)

Tuesday August 21, 2001 (continued)

15.00 – 15.20 **Søren Balling Engelsen**
Rapid and unique curve resolution of low field NMR T2-components (A20)

15.20 – 15.40 **Philip Hopke**
A modified alternating least-squares (MALS) algorithm: Draining the swamps in multiway analysis (A21)

15.40 – 16.20 Coffee break

Variable Selection
Chairman: Carsten Ridder

16.20 – 16.40 **Marlon M. Reis**
PARAFAC with splines: A case study (A22)

16.40 – 17.00 **Andreas Niemöller**
Variable selection based on genetic algorithms. Study on wavelet-coefficient- and principal-component-regression (A23)

17.00 – 17.20 **Henrik Öjelund**
Optimal choice of multiple filters for measuring chemical substances (A24)

18.00 - **City Hall Reception**

Conference Programme

Wednesday August 22, 2001

Process Control II
Chairman: Olav Kvalheim

09.00 – 09.40 **John MacGregor**
Latent variable methods in chemometrics: Theoretical foundations and practical implications (A25)

09.40 – 10.00 **Ingunn Berget**
Sorting of raw materials based on the predicted end product quality (A26)

10.00 – 10.20 **Kim Esbensen**
Development of on-line image analytical industrial process monitoring calibrated against structurally correct sampling of heterogeneous materials (A27)

10.20 – 11.00 Coffee break

11.00 – 11.20 **Dora Kourti**
Batch process monitoring using multivariate analysis. Recent developments and acceptance in industry (A28)

11.20 – 11.40 **Pia Jørgensen**
On-line batch fermentation process monitoring - Prediction of "Biological Time" (A29)

11.40 – 12.00 **Alberto Ferrer**
Dealing with missing data in MSPC: Several methods, different interpretations, some examples (A30)

12.00 – 12.20 **Anders Björk**
Spectra of wavelet scale coefficients of process acoustic measurements as input for PLS modelling of pulp quality (A31)

12.20 – 13.20 Lunch

13.20 – 14.20 **Poster Session (Posters 38-53)**

Wednesday August 22, 2001 (continued)

New & Old
Chairman: Agnar Höskuldsson

14.20 – 15.00 **Svante Wold**
New and old trends in Chemometrics - How to deal with the increasing data volumes in research, development, and production with examples from pharmaceutical research and process modelling (A32)

15.00 – 15.20 **Jürgen von Frese**
100 years old and still a cutting edge method: Principal component analysis in gene expression analysis (A33)

17.00 - **Conference Dinner**

Conference Programme

Thursday August 23, 2001

Uncertainty
Chairman: Kaj Heydorn

09.40 – 10.00 **Peter Wentzell**
Case studies in the application of maximum likelihood principal components analysis (A34)

10.00 – 10.20 **Oxana Rodionova**
Simple interval calculation - A method for linear modelling (A35)

New Algorithms II
Chairman: Rasmus Bro

10.20 – 10.40 **Agnar Høskuldsson**
Important and influential variables and samples in latent structure models (A36)

10.40 – 11.00 **Rasmus Larsen**
Decomposition of spectra using maximum autocorrelation factors (A37)

11.00 – 11.40 Coffee break

11.40 – 12.00 **Frank Westad**
Independent Component Analysis - A new valuable tool in data analysis or the Emperor's new clothes? (A38)

12.00 – 12.20 **Johan Trygg**
Orthogonal-PLS (O-PLS) for removing irrelevant variation in X variables (A39)

12.20 – 12.40 **Johan Westerhuis**
A discussion on orthogonal signal correction (A40)

12.40 – 13.00 **Closing of SSC7**

13.00 – Lunch

Poster presentations

A41	**P1**	QSAR modeling of polychlorinated dibenzofurans using a 3D structure representation using quantum topology (StruQT)
A42	**P2**	Investigation of the autofluorescence from cod extracts
A43	**P3**	Design, synthesis and QSAR evaluation of a chemical library directed towards the melanocortin receptors
A44	**P4**	Batch statistical processing of 1H NMR-derived urinary spectral data
A45	**P5**	Omeprazole and analogue compounds: A QSAR study of activity against *helicobacter pylori* using theoretical descriptors
A46	**P6**	Level of validation controlled by the choice of segmentation, in the cross-validation/jack-knifing of bi-linear regression models
A47	**P7**	Analysis of residuals: Statistical method in QSAR studies
A48	**P8**	Quantitation of the active substance in a pharmaceutical tablet using Near Infrared (NIR) transmittance spectroscopy and chemometrics
A49	**P9**	Multivariate image regression (MIR) & image regression validation
A50	**P10**	Multivariate methods in the development of a new tablet formulation
A51	**P11**	Near Infra-Red spectroscopy for brain studies? An early attempt at monitoring responses in the human orbito-frontal cortex to smell stimuli, by the use of multi-channel multi-wavelength diffuse NIR spectroscopy
A52	**P12**	Non-linear partial least squares (the error based non-parametric PLS algorithm)
A53	**P13**	Development of a software sensor for estimation of phosphorus in municipal wastewater
A54	**P14**	Evaluation of three methods for assessor and descriptor analysis
A55	**P15**	Implementation and validation of on-line models for monitoring of wood-chips properties
A56	**P16**	Frequency characterization of the chemical periodicity - The problem of assessing missing values
A57	**P17**	Correlations between biodegradation rates of alkyl sulphosuccinates and their physicochemical parameters
A58	**P18**	Estimation of uncertainty of concentration estimates obtained by image analysis
A59	**P19**	Expert system shell for evaluation of bond dissociation energies of organic compounds

A60	**P20**	Application of genetic algorithm - PLS to the determination of wine parameters from FTIR spectra
A61	**P21**	Improved discrimination of sea ice types: AMT and MIR applied to satellite images from ERS SAR
A62	**P22**	Sensory analysis of MRI pictures: Using human perception and cognition to segment and assess the interior of potatoes
A63	**P23**	Sampling noise and measurement noise: Two sources of uncertainty in the assessment of food quality
A64	**P24**	Pre-processing of input data for simplified GLS modelling, as applied to PLSR
A65	**P25**	Applied electrical DC-potential for flow improvement in power plant turbine inlet pipe-lines chemometric intercalibration between acoustics and PIV-laser velocimetry
A66	**P26**	Relationship between the conditions of fermentation and the laccase production of four strains of Lentinus edodes. Comparison of principal component analysis and spectral mapping technique
A67	**P27**	A quantitative assay of intact tablets by transmittance near-infrared spectroscopy
A68	**P28**	Simultaneous determination of water constituent concentrations and partial least squares
A69	**P29**	Spectral transformation and range-selection in multivariate calibration
A70	**P30**	Quantifying catecholamines using multiway modeling
A71	**P31**	Dioxin contamination of fish oil. PARAFAC and N-PLS analysis of fluorescence spectra
A72	**P32**	Application of simple interval calculation method
A73	**P33**	Successive estimating of reaction rate constants from spectral data: A case study of two-step kinetics
A74	**P34**	Generalization of pair-correlation method (PCM) for nonparametric variable selection
A75	**P35**	Application of PLS and back propagation neural networks for the determination of soil samples properties
A76	**P36**	A methodological multi-way analysis of Cassava Starch properties
A77	**P37**	Supervisory control of wastewater treatment operation by PC-space control
A78	**P38**	Implementation of multivariate real-time methodologies for industrial process control
A79	**P39**	Characterisation of noise uncertainty: Allan variance and sample variance
A80	**P40**	Development of a new fermented fish food product

A81 **P41** Performance of multivariate calibration methods for determination of the active ingredient and impurity in a pharmaceutical process solution analysed by near infrared spectroscopy

A82 **P42** Covariate challenge in multivariate statistical process monitoring

A83 **P43** On spurious models in process monitoring and fault identification - industrial experiences

A84 **P44** Modelling time series of Low Field 1H NMR relaxation curves of starch retrogradation. Comparison of PARAFAC, TUCKER3, slicing and multi-exponential fitting

A85 **P45** Multiblock models for explorative data mining in food technology

A86 **P46** Uniaxial compression data for predicting potato quality parameters

A87 **P47** Selection of optimal process analyzer for monitoring

A88 **P48** Simultaneous processing of raw data of samples and standards for the enhancement of selectivity and sensitivity in the liquid chromatographic analysis of cocaine

A89 **P49** Simultaneous processing of data obtained by high performance liquid chromatography and capillary electrophoresis with diode array detection

A90 **P50** An example showing the use of qualitative variables in experimental design applied to a process of making cheese

A91 **P51** Calibration transfer by generalized least squares

A92 **P52** Analysis and display of historical Stehekin river flow data with robust PCA methods

A93 **P53** Automated LC-MS analysis of natural products: extraction of UV, MS and retention time data for compound identification and chemometric analysis

Abstracts (Oral & Posters)

Exploratory data analysis - Addressing the context in which chemometrics works

Lars Munck, lmu@kvl.dk, Royal Veterinary and Agricultural University, Chemometrics Group, Food Technology, Department of Dairy and Food Science. Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark

Keywords: exploratory data analysis, multivariate screening, induction, data mining, induction

The instrumental revolution in chemistry and physics has produced a wide range of destructive and non-destructive identification and separation methods, which can be used as fast and inexpensive multivariate screening methods in considering laboratory as well as real life data. Within the limits of the analysis these methods may give a chemical-physical data fingerprint of a product or process which may contain more information than scientists as a collective may hypothesize.

The hypothesis-generating exploratory data analysis was developed in the social and economic sciences, starting with the principal component analysis (PCA) algorithm. Its application is discussed in examples from the food industry based on data from multivariate screening methods including multiblock/multiway extensions of PCA such as PARAFAC and Tucker which have more recently been applied in chemometrics.

Exploration can be pursued as early as in the formulation phase of a problem as a pre-project, dynamically employing a multivariate screening method (e.g. based on spectroscopy) covering the problem space with the generally held hypothesis that this strategy will produce a data base which could embrace the problem. The covariate latent structure of the resulting data set containing different blocks of data at different context levels of biological or technological organization is presented graphically as principal components (PCs) in a cognitively interpretable form. The identification and the naming of the PCs acts as a stimulus for the intuition of the investigator, occasionally generating a new fresh hypothesis which later could be tested by experimental design based on the factors revealed.

It is concluded that the exploratory inductive strategy using modern technology now available makes a new, largely independent channel of information accessible to classical, deductive, normative science, making possible a fruitful dialogue. The role of exploratory chemometrics at present and in the future with the aim to economize the input of resources in research is discussed using examples. There is a tendency that the prevailing normative strategy leads to specific problems getting solved, while multifactorial problems acccumulate. A much higher success rate should be guaranteed by allowing the financing of exploratory pre-projects based on multivariate screening methods and exploratory data analysis as a complement. The challenge of how to launch such exploratory chemometrics is discussed.

Resolution of proton LCNMR by multivariate means

Christian Airiau, c.airiau@bristol.ac.uk, Bristol University, School of Chemistry. Cantcock, United Kingdom
Hailin Shen, School of Chemistry, University of Bristol, Cantock's Close, BRISTOL BS8 1TS, UK
Richard G. Brereton, School of Chemistry, University of Bristol, Cantock's Close, BRISTOL BS8 1TS, UK

Keywords: LC-NMR, deconvolution, resolution, PCA

Proton liquid chromatography nuclear magnetic resonance spectroscopy is a relatively new approach for the analysis of mixtures. This method is highly quantitative because signals depend on concentration and number of protons so eliminating the need for calibration curves, and also results in very diagnostic spectra which can be directly correlated to structures. The method is conventionally limited by the problems of expense (deuterated solvents), slow analysis times (stopped flow) and high concentrations.

In this paper we show that it is possible to rapidly and economically record complex mixtures at acceptable concentrations, using chemometrics approaches for resolution, based on partial selectivity in the chromatographic and spectroscopic dimensions, employing PCA and morphological operators to determine the best variables. First steps involve preprocessing the Fourier transforms including alignment of spectra. A typical NMR spectrum may consist of 8192 possible variables which can be reduced to 50 or so variables. Because of the unique features of NMR it is possible to increase the signal to noise of contiguous variables by summing measurements over small windows of the spectrum further reducing the dataset. Factor analysis is then performed on these variables to reconstruct high quality concentration profiles which are in turn used to obtain improved spectra and quantitative profiles. The method is illustrated by spectra of closely overlapping aromatic isomers, with overlap both in the spectroscopic and chromatographic dimensions. The potential of LCNMR as against LCDAD or LCMS is unlocked using chemometrics methods.

Multivariate analysis of quality

Harald Martens, Harald.Martens@mail.tele.dk, DTU/NTNU/KVL . Teglgaardstr 12A , DK-1452 , Denmark
Magni Martens, mma@kvl.dk, The Royal Veterinary and Agricultural University, Chemometrics Group. The Royal Veterinary and Agricultural University , 1958, Denmark

Keywords: exploratory factor analysis, PLS, quality assesment

Data analysis is a vital part of science today, and in assessing quality, multivariate analysis is often necessary in order to avoid loss of essential information. Complex systems such as food call for multivariate data analysis.

There is no 'best' method in data analysis. But today there are too many data sets and not enough professional data analysts. Moreover, it is important that contextual background knowledge is used creatively and critically in the data analysis. So it is important that the sensory scientist learns to analyse his or her own data. This requires that the researcher chooses a flexible and understandable method, and learns not to misuse it.

This presentation outlines the statistical and cognitive concepts behind a compact, reasonably simple and safe approach to data analysis – soft multivariate modelling. This approach is presented in a recent book (Martens & Martens 2001). It builds on a previous book in chemometrics (Martens & Naes 1989), but is much less mathematical and has a wider scope. In particular, it focuses on the definition and measurement of quality.

With the present approach, researchers with no formal training of statistics are provided with a powerful and versatile methodology that allows them to design their investigations and analyse their own data effectively and safely.

The methodology used is the graphically oriented, interactive multivariate 'soft modelling' based on bi-linear modelling by cross-validated PLS Regression. This is a rather complete data analytical approach, in the sense that it has a wide applicability, ranging from purely explorative factor analysis, via a flexible set of regressions and classifications to confirmative analyses of effects in designed experiments.

This method for extracting useful information from data tables is demonstrated for various types of quality assessment, ranging from human quality perception via industrial quality monitoring to environmental quality and its molecular basis.

References:
1. Martens, H. and Naes, T. (1989) Multivariate Calibration. J.Wiley & Sons Ltd.
2. Martens, H. and Martens, M. (2001) Multivariate Analysis of Quality. An Introduction. J.Wiley & Sons Ltd.

The reliability of naïve assessors in sensory evaluation using multivariate data analysis

Maurice O´Sullivan, mos@kvl.dk, The Royal Veterinary and Agricultural University. Sensory Science, Rolighedsvej 30, DK-1958, Denmark
S. Boberg, The Department of Human Nutrition, The Royal Veterinary and Agricultural University
Harald Martens, Harald.Martens@mail.tele.dk, DTU/NTNU/KVL . Teglgaardstr 12A , DK-1452 , Denmark
L. Kristensen, The Department of Dairy and Food Science, The Royal Veterinary and Agricultural University
Magni Martens, mma@kvl.dk, The Royal Veterinary and Agricultural University, Chemometrics Group. The Royal Veterinary and Agricultural University , 1958, Denmark

Keywords: sensory evaluation, PLS, jack-knifing

This study was part of a larger study designed to determine if pork meat consumption in specially formulated test meals improved non-heme iron bio-availability in a selection of Danish women (n = 45) test subjects with low iron status. The aims of the present study were to assess the sensory characteristics of these test meals, composed of bread, rice, pea puree, tomato and meat (pork) and to determine the reliability of these naïve experimental subjects as sensory panelists using multivariate data analytical methods. As well as a control meal (meal without meat) 3 test meals were monitored and each test meal was evaluated in two replicates by 3 different groups of subjects, 25g meat in the meal (Assessor set 1), 50g meat in the meal (Assessor set 2), 75g meat in the meal (Assessor set 3). The S/N (Signal to Noise) ratios for the two replicates of the meals were good indicating reliability in the sensory evaluation. As all the meals were not assessed by all subjects The ‘Jack-knife` Estimated Stability data verified the combining of assessor sets 1, 2 and 3 and allowed interpretation of the resulting data. The sensory metallic and bitter descriptors were associated with the meals containing meat. The Salt descriptor was associated with the tomato component of the meal and the Sweet descriptor was associated with the pea component of the test meal. The results imply that untrained subjects can be used for basic sensory evaluation in a reliable way. They were able to discriminate between the different meals and adding meat to the meals significantly increased the taste of bitter and metallic.

From a statistical methodology point of view, the presentation demonstrates how the Procrustes Rotation in the bi-linear “Jack-knifing” simplifies the comparison of PLS Regression models from different data sets.

A framework for multiblock component models

Age Smilde, asmilde@its.chem.uva.nl, University of Amsterdam, Department of Chemical Engineering, Process Analysis and Chemometrics. Nieuwe Achtergracht 166, NL-1018 WV Amsterdam, Netherlands, http://www.uva.nl
Johan Westerhuis, westerhuis@its.chem.uva.nl, Chemical Engineering, University of Amsterdam, Process Analysis and Chemometrics. Nieuwe Achtergracht 166, 1018 WV AMSTERDAM, The Netherlands, http://www-its.chem.uva.nl/research/pac/
Frans van den Berg, fb@kvl.dk, The Royal Veterinary and Agricultural University, Food Technology, Chemometrics Group. Rolighedsvej 30, DK-1958, Denmark, http://www.models.kvl.dk

Keywords: multiblock, multiway, stationary phases, chromatography

In more and more areas of chemistry, biology and food technology problems translate into analyzing multiple sets of data. Usually, these sets of data have something in common (e.g. the samples) and, hence, analyzing them in a simultaneous fashion is more efficient than analyzing each data set individually. In this paper, the focus is on component models. Such models try to capture the 'commoness' of the data sets and to what extent this commoness is reflected in the individual data sets.

Several methods are available for making multiblock component models, such as Consensus-PCA, Hierarchical-PCA and Multiblock Covariates-PCA. When using multiblock models, some fundamental choices have to be made, e.g. do all blocks have to contribute to some extent to the commoness? All available models deal with these choices in a particular way.

A framework will be given for multiblock models, based on how each model deals with the fundamental choices. This will also point to alternative models. Extensions to multiblock multiway models will be given. The differences and similarities of the models will be illustrated with a three-block problem from food technology.

Process performance monitoring in the presence of confounding variation

Baibing Li, Centre for Process Analytics and Control Technology, University of Newcastle, Newcastle upon Tyne, NE1 7RU, United Kingdom
Elaine Martin, e.b.martn@ncl.ac.uk, University of Newcastle. Merz Court, NE1 7RU , United Kingdom
Julian Morris, julian.morris@ncl.ac.uk, University of Newcastle. Centre for Process Analytics and Control Technology (CPACT), NE1 7RU , United Kingdom

Keywords: process chemometrics, PLS, MSPC

When underlying process behaviour is masked by variability that is an inherent part of routine operation, applying the ordinary partial least squares algorithm will result in multivariate statistical process control (MSPC) models that are strongly influenced by these known sources, leading to models with reduced sensitivity materialising in unwanted process changes not being detected. The aim of this paper is to present a new methodology, two-stage partial least squares, for use in process performance monitoring situations where the processes is confounded by known sources of variation that are introduced as part of routine process operation and that mask other more subtle process changes.

The first stage of the two-stage PLS algorithm selectively removes those sources of variation, associated with operational requirements that are not of interest to operational personnel, from a process monitoring perspective, resulting in the latent variables obtained through the second stage being uncorrelated with this nuisance source of variation. This leads to more sensitive MSPC models that are more able to detect undesirable changes in process operation. The idea of the two-stage PLS algorithm is thus to decompose total variation into:
Common Cause Variation + Assignable Cause Variation + Confounding Variation

Potential areas of application for the two-stage PLS include the building of unified MSPC models for different production modes, for example different recipes, the monitoring of processes where recipes are changed on a 'rolling' basis, or improving the sensitivity of MSPC models by removing the influence of other non-process variables such as those associated with environmental influences. The algorithm will be demonstrated on a simulation of an industrial process.

Data mining issues on modern multivariate online industrial process control

Franck Torre, ftorre@stats.warwick.ac.uk, University of Warwick, UK
H. P. Wynn, University of Warwick, UK
P. Corbett, University of Warwick, UK
J. Rottorp, IVL, Sweden

Keywords: data mining, process control, data filtering, data cleaning

In the context of a global competitive economy and reinforced public environmental policies, more elaborated industrial process control strategies are needed. The current EC project MAPP aims to define an integrated approach for online statistical process control. In one hand, the information technology revolution gives unlimited real-time access to most of the production process variables. On the other hand, processes consist in huge amounts of data. Data mining is devoted to the pre-processing of the online captured data. From our point of view, the data mining task consists in two different things:

1. The cleaning of the online multivariate data as soon as they are captured on the process and before storage in the database;

2. The extraction and reduction of the stored information for the purpose of modelling;

Both kinds of data mining will be illustrated on some real online industrial applications.

Multivariate biological profiling and constrained principal response profile regions of PCB´s

Lennart Eriksson, lennart.eriksson@umetrics.com, Umetrics AB. Umetrics AB , S-907 19 , Sweden
Patrik L. Andersson, Environmental Chemistry, Umeå University, Umeå, Sweden
Erik Johansson, Umetrics AB, Umeå, Sweden
Mats Tysklind, mats.tysklind@chem.umu.se, Environmental Chemistry, Umeå University, Umeå, Sweden

Keywords: multivariate design, multivariate biological profiling, principal toxicity scale

The polychlorinated biphenyls (PCBs) comprise a group of 209 congeners varying in the number of chlorine atoms and substitution pattern. The structural characteristics of the PCBs influence their potency and mode of biological action. In an early stage, 52 physico-chemical descriptors were compiled and evaluated by PCA, which lead to the postulation of four principal properties (i.e., PC-scores) summarizing important molecular properties of PCBs. Statistical molecular design (SMD) in these principal properties was then utilized to select a set of 20 representative PCBs for further experimentation.

The 20 selected PCBs were tested in several bioassays to investigate structure-specific biochemical responses. Additionally, biomagnification in fish was determined. In total, seven test systems were used to assess the biological properties of the PCBs, thus creating a biological response profile for each compound. We here call this approach multivariate biological profiling. It is the objective of this contribution to examine the complexity of these biological profiles and to see how they relate to the corresponding physico-chemical characterization.

Using PCA, it will be shown that the PCBs exhibit three characteristic biological profiles. These profiles are different and unique, but there is a gradual transition from one profile to another. Using PLS, the quantitative relations between these biological profiles and the physico-chemical data are established. It will be shown that there exist strong relationships between these two types of data.

Sometimes, there will only be a few tested compounds representing a certain biological profile. In such circumstances, it might be pertinent to select additional compounds for biological testing in order to better map that profile. Such supplementary choices of PCB-congeners, focusing on a particular profile of responses, can be accomplished in the following way: The QSAR(s) between physico-chemical data and response profiles are interrogated to predict multivariate biological profiles for the large number of yet untested PCBs. From such predicted response profiles interesting molecules are drawn using D-optimal design. These may improve the resulting QSARs.

Analysis of short protein sequences using multivariate batch modeling

Tarja Rajalahti, tarja.rajalahti@chem.umu.se, Umeå University, Organic Chemistry, Research Group for Chemometrics. SE-90187, Umeå, Sweden
Maria Edman, mariae@dbb.su.se, Biochemistry & Research Group for Chemometrics, Department of Chemistry, Umeå University, SE-90187 Umeå, Sweden
Michael Sjöström, michael.sjostrom@chem.umu.se, Umeå University. Umeå University , SE-901 87 , Sweden
Åke Wieslander, ake.wieslander@dbb.su.se, Department of Biochemistry, Stockholm University, SE-10691 Stockholm, Sweden
Svante Wold, svante.wold@umetrics.com, University of Umeå. , Sweden

Keywords: batch modeling, protein sequences, bioinformatics

Biological data are accumulating at an enormous rate as the sequencing techniques are becoming faster and better. There is an increasing need for understanding the relationships between sequences, structure, biological activity and chemical properties. Since protein sequences has an univariate direction from the beginning (N-terminal) to the end (C-terminal) one sequence can be considered as one batch. In this paper the possibility to use multivariate batch modeling as a tool for analyzing sequence data is discussed.

Wold, et al. (1998) have developed methods for modeling and diagnostics of batch processes using multivariate techniques. This technique can handle situations where the length of the batches is varying. Multivariate batch modeling is performed in two steps:
1. Observation level – modeling the batch evolution (models relating each amino acid in the sequence to its position in the sequence)
2. Batch level – modeling the whole batch, predicting results of new batches (models relating the whole peptide sequence to the properties of the peptide).

Signal peptides are N-terminal amino acid sequence extensions on proteins. Different classes of Escherichia coli signal sequences with known location were studied in the present work. Each individual amino acid in the sequences was first translated to numbers using 29 chemical measurements describing the properties of the amino acid. A set of peptides gives a 3-way data table, which was first unfolded to give a 2-way matrix where one row corresponds to one individual amino acid (the direction of the variables is preserved). The position of the amino acid in the peptide sequence was used as the batch maturity index in the PLS modeling. The score vectors from the obtained PLS model were reorganized as row vectors, giving a new matrix where one row now corresponds to one whole peptide. Peptides were divided into known classes and a PLS-DA model was computed.

The studied signal sequences contained patterns related to the final location of the protein, and the peptide classes could be clearly separated using a combination of multivariate batch modeling and PLS-DA.

References:
Wold, S., N. Kettaneh, H. Fridén and A. Holmberg (1998). Modelling and diagnostics of batch processes and analogous kinetic experiments, Chemom. Intell. Lab. Syst. 44, 331-340.

QSAR of progestogens: Use of a priori and computed molecular descriptors and molecular graphics

Rudolf Kiralj, rudolf@iqm.unicamp.br, Instítuto de Química. UNICAMP, 13083-970, Brazil
Márcia M. C. Ferreira, marcia@iqm.unicamp.br, Universidade Estadual de Campinas, Instituto de Química. Campinas, SP, 13083-970, Brazil

Keywords: progestogens, QSAR, molecular graphics

There are only a few recent attempts to describe progesterone and progestogens on a quantitative level. Altough widely known as contraceptives, these compounds also showed to be potential drugs for hormone and anti-cancer therapies, and for other clinical treatments, what justifies studies of their molecular properties and intermolecular interactions with various molecules. The lack of larger number and homogenicity of progestogens activity data makes it difficult to have a clear picture of the progestogens behaviour at atomic level. Recently, the one and only up to date crystal structure of progesterone-receptor complex[1] gave much insight into this intermolecular interaction.

Continuing the SAR idea[2] to relate molecular descriptors to contraceptive activity for two sets of progestogens, we employed molecular graphics and QSAR analysis (Principal Component Analysis and Partial Least Squares) here using both *a priori*[3] and various computed descriptors at a semi-empirical and an *ab initio* level. The smaller set of six progestogens (derivaties of progesterone with changes of side chains at and of the bonds of the rings A, C, and D) was successfully described by shape, sterical and electronic (as heteroatomaticity, etc.) descriptors. The presence of hetero atoms and multiple bonds that can contribute to electron delocalization *via* conjugation and hyperconjugation shows to be important for the progestogen activity. The other set of progestogens, comprising nineteen progesterone derivatives with small spherical substituents (mainly halogens, methyl and hydroxy group), exhibits parabolical activity dependence on sterical parameters (size of substitutents) for each substitution position.

Detailed analysis of the data will be disscussed. The *a priori*[3] (simple hand or pocket calculator made) descriptors will be compared with computed descriptors. The method of data transformation for the nineteen progestogens to build a linear model will be presented.

References:
1. S. P. Williams, P. B. Sigler, Nature, **393** (1998) 392-396.
2. R. Vendrame, M. C. Ferreira, C. H. Collins, Y. Takahata, J. Mol. Graph., **19** (2001) in press.
3. R. Kiralj, M. M. C. Ferreira, J. Mol. Graph., submitted for publication.

A combinatorial approach for drug discovery - based on multivariate methods

Torbjörn Lundstedt, torbjorn.lundstedt@melacure.com, Melacure Therapeutics. Ulleråkersvägen 38, SE 756 43, Sweden
Per Andersson, per.andersson@melacure.com, Melacure Therapeutics. Ulleråkersvägen 38, SE 756 43, Sweden, http://www.melacure.com
Arne Boman, Melacure Therapeutics AB, Ulleråkersvägen 38, SE-756 43Uppsala, Sweden
Elisabeth Seifert, Melacure Therapeutics AB, Ulleråkersvägen 38, SE-756 43Uppsala, Sweden
Maria Flärdh, Melacure Therapeutics AB, Ulleråkersvägen 38, SE-756 43Uppsala, Sweden
Anna Skottner, Melacure Therapeutics AB, Ulleråkersvägen 38, SE-756 43Uppsala, Sweden

Keywords: PCA, QSAR

We are presenting a strategy for constructing combinatorial libraries with optimal information while still taking experimental feasibility into account. The objective is to provide optimal chemical diversity with a moderate number of compounds, plus adequate depth and width of the response testing including not only activity but also bioavailability. The strategy is based on a multivariate characterisation of the synthesis starting materials (building blocks), Principal Component Analysis (PCA), multivariate design, and Multivariate Quantitative Structure-Property Relationships (M-QSPR). The strategy applies to solid phase synthesis, libraries in solution, polymers, tablet formulation, catalyst development, bioinformatics and functional genomics.

Examples from combinatorial molecular biology and combinatorial chemistry for lead identification and optimisation will be presented.

Use of Monte Carlo simulations in chemometrics

Klaas Faber, n.m.faber@ato.wag-ur.nl, ATO, Agro & Industrial Production Chains, Production & Control Systems. P.O. Box 17, NL-6700 AA Wageningen, Netherlands, http://www.ato.wageningen-ur.nl

Keywords: Monte Carlo simulation, calibration standard error of prediction, limit of detection, trilinear decomposition

Monte Carlo simulations have acquired a poor reputation in Chemometrics, because often the results cannot be extrapolated to the real world. A good example is the “simulation” of so-called spectro-chromatographic data (e.g. HPLC-UV) by adding random noise to a sum of outer products of Gaussian profiles and experimentally obtained spectra: depending on the specific target application, reality is oversimplified. Possibly as a result of this bad reputation, Monte Carlo simulations are generally under-utilised. The focal point of the current presentation is that Monte Carlo simulations can be extremely useful for testing the correctness of an approach. By contrast, the practical utility of an approach is demonstrated only on real examples. In other words, the tasks of determining the correctness and practical utility of a method should be approached differently. Since developments in Chemometrics are mainly driven by applications, demonstrating the practical utility often receives more attention than proving the intrinsic correctness. The following examples are treated where Monte Carlo simulations have proved to yield valuable information with respect to correctness:

1-quantifying the uncertainty in estimates of root mean squared error of prediction (RMSEP), which can be useful for determining the size of an adequate test set in multivariate calibration;

2-testing the adequacy of approximate expressions for standard error of prediction when using partial least squares (PLS) regression (including multiway versions);

3-testing the adequacy of a limit of detection estimator when using the generalized rank annihilation method (GRAM) for the calibration of second-order data;

4-comparison of trilinear decomposition methods.

Applicational aspects of support vector machines

Anton Belousov, a.belousov@icb-online.de, Institut fuer Chemo- und Biosensorik. Mendelstr. 7, D-48149, Germany, http://www.icb-online.de
Serguei Verzakov, verzakov@uni-muenster.de, Institut fuer Chemo- und Biosensorik. Mendelstr. 7 , D-48149 , Germany
Jürgen von Frese, j.von-frese@icb-online.de, Institut für Chemo- und Biosensorik, Chemometrics. Mendelstr. 7, D-48149, Germany, http://www.icb-online.de

Keywords: classification, SVM, MIR

A primary goal when implementing classifiers for practical applications is an optimal performance on future unknown samples, i.e. a high generalisation ability. Choosing the right balance between maximal classifier flexibility and minimal overfitting to a limited training set poses one of the most difficult obstacles for obtaining a good generalisation ability.

Support vector machines (SVM) have been established as an important contribution to classification. Their special emphasis on generalisation ability makes SVM particularly interesting for real world applications with limited amounts of training data. But for this classifier to be adopted and widely used the conditions under which SVM outperforms ‘classical’ methods are to be revealed.

In this contribution we present the results of the analysis of this topic. As an illustration we show the step-by-step construction of a classifier of polymers by their mid-infrared spectra. With this example we show how SVM can manage the main difficulties that a typical classification task delivers: 1) complexity (e.g. multimodality) of the class conditional distributions, 2) contamination with outliers/noise, 3) limited number of samples. General theoretical estimations of the sensitivity of the method to each of these factors are derived, the corresponding control parameters are found and an application of these considerations to the real classification task is shown. Weak points of the method are discussed.

Static PLSR optimization based on Kalman filtering theory and noise covariance estimation

Rolf Ergon, rolf.ergon@hit.no, Telemark University College. Telemark University College, P.O.Box 203, Porsgrunn, Norway, http://www.hit.no/
Kim Esbensen, kes@aue.auc.dk, Ålborg Universitet Esbjerg. Norgesgade 31, 1.th , DK-6700 , Denmark

Keywords: PLSR, optimization, noise, covariance, estimation

Due to practical and economical reasons many industrial applications of partial least squares regression (PLSR) have to do with quite few Y observations. At the same time it may often be possible to obtain a much larger number of X observations. The paper discusses how these extra observations may be used in order to improve the PLSR predictor.

The standard PLSR predictor may be expressed by the data matrices X and Y and the loading weight matrix W [1]. The optimal W is a transposed Kalman gain K, which could be determined if the X and Y noise covariances were known [2]. From the data it is possible to obtain estimates of the noise covariances, and with a large number of X observations this may lead to an improved W estimate and thus an improved predictor.

Simulation results and industrial and analytical laboratory examples will be given in the contribution.

References:
[1] Helland I.S.: On the structure of partial least squares estimation. Communications in statistics, 17(2), 581-607 (1988)
[2] Ergon R.: Dynamic System Multivariate Calibration for Optimal Primary Output Estimation. PhD thesis, the Norwegian University of Science and Technology /Telemark University College, Trondheim/Porsgrunn, Norway, 1999

MILES: A general approach to maximum likelihood estimation

Rasmus Bro, rasmus@optimax.dk, The Royal Veterinary & Agricultural University, Dept. of Dairy and Food Science, Food Technology. Rolighedsvej 30, 1958, Frederiksberg, Denmark, http://www.models.kvl.dk/users/rasmus/
Nicholas D. Sidiropoulos, nikos@ece.umn.edu, Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455, USA
Age Smilde, asmilde@its.chem.uva.nl, University of Amsterdam, Department of Chemical Engineering, Process Analysis and Chemometrics. Nieuwe Achtergracht 166, NL-1018 WV Amsterdam, Netherlands, http://www.uva.nl

Keywords: iterative majorization, PARAFAC, PCA, weighted least squares

A general algorithm is provided for maximum likelihood fitting of deterministic models subject to Gaussian distributed residual variation. By deterministic models is meant models in which no distributional assumptions are valid (or applied) on the parameters. The algorithm is called MILES (Maximum likelihood via Iterative Least squares EStimation). It is shown that for any model with residual variation that follows a Gaussian distribution (including any type of nonsingular covariance), maximum likelihood parameters can be computed using simple least squares (LS) algorithms in an iterative fashion. The maximum likelihood algorithm is based on iterative majorization. The suggested algorithm is shown to include e.g. the algorithms for maximum likelihood principal component analysis suggested recently.

The MILES algorithm is simple and can be implemented as an outer loop in any least squares algorithm e.g. for analysis of variance, regression, response surface modeling etc. Thus, with MILES, it is possible to handle efficiently, the modeling of data where the error covariance structures is known or can be estimated with reasonable certainty. Several examples on the use of MILES are given to highlight its properties.

Strategies in modelling dynamic systems

Satu-Pia Reinikainen, satu-pia.reinikainen@lut.fi, Lappeenranta university of Technology, Department of Chemical Technology, Chemometrics Group. PO Box 20, Lappeenranta, Finland, http://www.lut.fi
Agnar Høskuldsson, ah@akp.dtu.dk, Danish Technical University. Danish Technical University, Lyngby, Denmark

Keywords: PLS, variable selection, lags, periodicity, trend

It is often difficult to model data by dynamic models. The reason is that data, which describe dynamic situations, often show irregular behaviour. Here we present a collection of strategies that can be used to judge the behaviour of data with the purpose of obtaining best possible predictions.

The strategies include how we want to handle data, types of models to choose among and requirements to the model that is to describe data. They include: 1) Past observations. How much of the past observations should be used if we want best possible predictions? 2) Lags in X-data. How many and which lags of the X-data should be used? 3) Lags in Y-data. How many and which lags of the Y-data should be used? 4) Variables to use. Which variables should be used in the modelling task? 5) Single or multiple latent structure. Should we use the same latent structure for all response variables or separate ones? Should we use the same latent structure for some of the response variables? 6) Dimension. What dimension should we use in the modelling task? 7) Predictions. What types of predictions are you looking for? One step ahead, two steps etc. Do you want prediction with restrictions on the solution vector? 8) Important response values. Do you want to include in the modelling procedure that some samples are more important than the others? 9) Do you want to model data with a window of important response values?

These strategies can be applied to different types of mathematical models: 1) Linear least squares. Linear models like e.g., autoregressive models.
2) Kalman filtering. Sequential updates of solutions, when new samples arrive. Resulting figures like e.g., Kalman gain are computed. 3) Balanced fit and solutions. Besides fit there can be requirements to the changes of the solution vector from sample to sample. There can also be conditions on e.g., the size of the solution vector.

Using these strategies we can detect and include in the model special features in data like: 1) Periodicity. We can detect significant periodic behaviour of the data. 2) Trend. We can control for special types of trend, like polynomial behaviour. 3) Non-linearity. We can estimate the non-linearity in the latent structure of the data.

These methods can be applied to any kind of data. Thus, they can be applied to e.g., NIR data or other optical measurement technology. They can be used for process control or for supervision of production processes. Clients on Internet can use the software program. I.e., analysis of data can be carried out from a PC that is connected to Internet.

Combining bilinear modeling and ridge regression

Martin Høy, martin.hoy@pvv.ntnu.no, NTNU. Fak. kjemi, Inst. kjemi , 7491, Norway

Keywords: PLS

A method is presented for making Principal Component Regression (PCR), Partial Least Squares Regression (PLSR), and other regressions based on Bilinear Modeling (BLM) less sensitive to overfit. The idea is to use generalized ridge regression to calculate the y-loadings, in order to prevent small, uncertain values of the score vectors from causing inflation of variance in the regression coefficients. Thus, we combine the stabilizing power of ridge regression with the modeling power and interpretability of bilinear models. The method is intended to provide better predictive ability and improved stability for regression models.

Non-linear partial least squares (estimation of the weight vector)

Per Anker Hassel, p.a.hassel@ncl.ac.uk, University of Newcastle. Centre for Process Analytics and Control Technology, Newcastle upon Tyne, United Kingdom, http://ncl.ac.uk/
Elaine Martin, e.b.martn@ncl.ac.uk, University of Newcastle. Merz Court, NE1 7RU , United Kingdom
Julian Morris, julian.morris@ncl.ac.uk, University of Newcastle. Centre for Process Analytics and Control Technology (CPACT), NE1 7RU , United Kingdom

Keywords: PLS algorithm, weight vector, weighted average, projections, nonlinear PLS

Given the standard partial least squares (PLS) projections, $\mathbf{t} = \mathbf{Xw}$ and $\mathbf{u} = \mathbf{Yq}$, non-linear PLS has been achieved by fitting a non-linear function $f(\mathbf{t})$ between the **X** and **Y** scores, $\mathbf{u} = f(\mathbf{t}) + \mathbf{e}$. In previously reported work, the weight vector **w**, has been either calculated as for the linear estimation of **w** or through an optimisation routine as in error based non-linear PLS. One exception is the framework presented in the SPLINE-PLS of Wold, where the weight vector is computed based on the covariance criterion but is adjusted due to the presence of non-linearities. This approach gives equal weight to the improvement in fit and the square root of the variance of the score vector (**t**). In situations where the variance of the response variable, **Y**, is not adequately explained by the process data, **X**, the covariance criterion is sub-optimal. Whilst in linear PLS by including additional latent variables the solution tends towards the least squares solution, this is not normally the case for non-linear PLS. Thus it is critical to find the 'optimal' weight, that identifies the 'true' non-linear relationship between the vectors **t** and **u**.

A novel extension of the estimation of **w** for any non-linear function is proposed. As for the covariance or correlation criteria, it is based on a measure of fit. In this case, the measure is taken from the theory of the weighted average and is the inverse of the variance of the error. In PLS, the error is estimated separately for each variable, thus an **X** variable with a small error will be given a large weight whilst a large error will be given a small weight, since it is calculated as the inverse of the variance of the error.

The methodology is illustrated using a number of data sets. It is shown that for noisy, non-linear data containing collinear or highly correlated data, the proposed algorithm for calculating **w** outperforms both the covariance and the correlation criteria of PLS in terms of performance on a validation data set.

Automated curve resolution of multi-way data and prediction of biological properties from the resolved profiles

Olav Kvalheim, olav.kvalheim@kj.uib.no, University of Bergen, Department of Chemistry. Allégaten 41, N-5007 Bergen, Norway
Bjørn Grung, Department of Chemistry, University of Bergen, Allégaten 41, N-5007 Bergen, Norway
Hailin Shen, Department of Chemistry, University of Bergen, Allégaten 41, N-5007 Bergen, Norway
Ingvar Eide, Statoil Research Centre Trondheim, N-7005 Trondheim, Norway

Keywords: multiway data, curve resolution

This lecture presents an attempt to develop a solution to the automated resolution of complex multicomponent data. Several constraints on concentration and spectral profiles are used to determinine the number of components, and for cutting complex peak clusters into smaller sub matrices. Our results show that the proposed procedures resolve complex overlapping systems without human intervention.

They are applied to multicomponent data from chromatography-mass spectrometry (GC-MS and LC-MS) with several hundred components. Typically the raw data needs more than 50Mbytes of disc space. The resolved data can then be used to predict properties, e.g., mutagenacity.

Rapid and unique curve resolution of low field NMR T2-components

Søren Balling Engelsen, se@kvl.dk, The Royal Veterinary and Agricultural University. Rolighedsvej 30, 1958, Denmark
Henrik Toft Pedersen, het@kvl.dk, The Royal Veterinary and Agricultural University, Department of Dairy and Food Science, Food Technology. Rolighedsvej 30, 1958 Frederiksberg C, Denmark
Rasmus Bro, rasmus@optimax.dk, The Royal Veterinary & Agricultural University, Dept. of Dairy and Food Science, Food Technology. Rolighedsvej 30, 1958, Frederiksberg, Denmark, http://www.models.kvl.dk/users/rasmus/

Keywords: DECRA

The established method for analysing nuclear magnetic resonance relaxation decay profiles is exponential curve fitting, resolving the decay into a number of mono-exponential functions equal to the number of resolved factors. The curve fitting is performed by minimising the squares of the residuals, typically by a variation of the Levenberg-Marquardt equation . However the linarization approximation in multi-exponential fitting algorithms has severe problems with the non-linear exponential functions, and the solutions found may not be unique.

A brilliant idea was put forward by Windig and Antalek when studying first order reaction kinetics by high-resolution n.m.r. This method called Direct Exponential Curve Resolution Algorithm (DECRA) takes advantage of the fact that exponential decay functions (time´intensity), when translated in time, retain their characteristic relaxation times while only their relative amounts or concentrations change. By such simple translations (slicing) it is possible to create a new “pseudo” direction in the relaxation data (time´intensity´slice) and thus facilitate application of trilinear (multiway) data-analytical methods, which in turn provide mathematically unique recovery of the underlying T2 components.

In the original work, by Windig and Antalek, the exponential decay was found in the sample direction (first order kinetics). In this study the method is applied to the analysis of low-field n.m.r. data in the time domain by shifting in the variable direction (time) instead of the sample direction. Results obtained on time domain NMR data of meat samples is discussed and compared to results from multi-exponential curve fitting.

A modified alternating least-squares (MALS) algorithm: Draining the swamps in multiway analysis

Ji-Hong Wang, Department of Chemical Engineering, Clarkson University, Potsdam, NY 13699-5705, USA
Philip Hopke, hopkepk@clarkson.edu, Clarkson University. Box 5705 , 13699-5705 , USA
Thomas M. Hancewicz, Unilever Research US, Edgewater, NJ 07020

Keywords: alternating least squares, multiway factor analysis, mixture resolution, swamps, collinearity

One of the most widely used approaches to solving the factor analysis problem is alternating least-squares. It has been used particularly with respect to multiway analysis problems (Tucker2, Tucker3, etc). In many cases, it has been found that there can be very slow convergence that requires many, many iterations before a final approach to the solution and these regions of slow convergence have been turned "swamps." Prior efforts have been made to move more rapidly through the swamps using ridge regression (RR). However, ridge regression is a biased estimator and thus, there is concern regarding the properties of the solution. The situation is even worse when we try to constrain factors. There are times when solutions can not be obtained even in simple examples. An alternative approach, the Modified Alternating Least-Squares (MALS) algorithm, has been developed to provide an unbiased estimator and at the same time, greatly increases the speed with which the solution is obtained. MALS is designed to address the two problems within the same approach. MALS works better than the typical RR in this kind of problem, and avoids the bias created by RR. With respect to the computational load, the time and memory requirements are close to the original ALS in each step. That is to say, we can use MALS instead of ALS in all the situations. The nature of the algorithm and its application to a number of test data sets will be presented.

PARAFAC with splines: A case study

Marlon M. Reis, marlon@iqm.unicamp.br, Chemistry Institute - UNICAMP. Cidade Universitária Zeferino Vaz, s/n, Campinas, Brazil
Márcia M. C. Ferreira, marcia@iqm.unicamp.br, Universidade Estadual de Campinas, Instituto de Química. Campinas, SP, 13083-970, Brazil

Keywords: PARAFAC, smoothing splines constraint, carbon monoxide

The PARAFAC model has been used in several applications in chemistry, e.g. for overlapped spectra curve resolution, second order calibration and others. In general, PARAFAC approach considers the decomposed multilinear components as being vectors. This work presents a PARAFAC approach where the decomposed multilinear components are considered as functions.

The functional objects used to constrain the PARAFAC decomposition are Splines. The methodology used to promote the Spline-PARAFAC decomposition is based on Bro-Sidiropoulos' approach(1) for the unimodality constraint.

The Spline-PARFAC requires an additional calculation of a penalty parameter or the number basis functions, which were found in this work by using an Ordinary Cross Validation OCV(2). The Spline-PARAFAC was applied on a data set, which corresponds to the concentration of carbon monoxide measured every hour during one year in the São Paulo city in Brazil. The data was arranged as a Three Way Array having the modes: (Hours of the Day-Days of Week-Weeks of the Year). The Spline-PARAFAC presented a good performance.

The authors acknowledge the financial support from FAPESP for carrying out this work and CETESB for kindly supplying the data set.

References:
1. Bro, R.; Sidiropoulos, N. D "Least Squares Algorithms Under Unimodality and Non-Negativity Constraints", Journal of Chemometrics 1998, 12, 223-247.
2. Silverman, B. W. "Some Aspect of the Spline Smoothing Approach to Non-parametric Regression Curve Fitting", Jornal of Royal Statistcal Society, 1985, 47, (1), 1-52.

Variable selection based on genetic algorithms. Study on wavelet-coefficient- and principal-component-regression

U. Depczynski
Volker J. Frost, volker.frost@sensologic.de, SensoLogic GmbH, Software + Sensor Systems. Hummelsbuetteler Steindamm 78a, 22851 Norderstedt, Germany, http://www.sensologic.com
K. Molt
Andreas Niemöller, andreas.niemoeller@bruker.de, Bruker Optik GmbH, FT-NIR Application. Rudolf-Plank-Str. 23, Ettlingen, Germany, http://www.bruker.com

Keywords: chemometrics, multivariate calibration, genetic algorithms, wavelets, PCR

Variable selection is a frequently emerging type of combinatorial problem. This article focuses on the skills and benefits but also the problems that has to be faced by using stochastical methods like genetic algorithms in order to solve such tasks.

One of the main questions raised in chemometrics is how to find a satisfactory description (model) of the relation between a set of independent variables, e.g. factors derived from spectral analysis by principal component analysis (PCA), and a known chemical or physical property. This article points out that genetic algorithms can solve such an optimisation problem reliably and precisely. Therefore results of studies on wavelet coefficient regression (WCR) and principal component regression (PCR) are presented and compared to commonly used methods like partial least squares (PLS). The studies were carried out on near infrared spectroscopic data.

Optimal choice of multiple filters for measuring chemical substances

Henrik Öjelund, hoe@imm.dtu.dk, IMM. IMM, DTU, DK-2800, Denmark, http://www.imm.dtu.dk/
Jan Nygaard Nielsen, jnn@imm.dtu.dk, Informatics and Mathematical Modelling, Technical University of Denmark DK-2800 Lyngby, Denmark
Jesper Falden Offersgaard, DELTA Light &Optics, Hjortekærsvej 99 DK-2800 Lyngby, Denmark
Henrik Madsen, hm@imm.dtu.dk, Informatics and Mathematical Modelling, Technical University of Denmark DK-2800 Lyngby, Denmark
Niels Henrik Eisum, DHI, Science Park Århus, Gustav Wieds vej 10 DK-8000 Århus C, Denmark

Keywords: optical filters selection, nitrate, nitrite

Increasingly spectroscopic methods are being used for quantitative analysis in chemistry, biology, and medicine. While advances in instrumentation and computer technology allow for higher resolution and greater sensitivity even in online applications, there is still a need for inferring useful, low-dimensional information from high-dimensional data. A number of multivariate calibration methods and variable selection methods have been proposed to select only the most informative subsets of the measured absorption coefficients subject to different definitions of the concepts of optimality and informative data.

In this paper it is proposed to introduce a number of optical filters in the optical system in order to take absorption coefficients close to the center wavelength of the filters into account with the objective of developing more robust measuring devices. In this context each optical filter is assumed to be characterized by a center wavelength and a halfwidth, and the shape of the filter is modelled by a Gaussian shape (assuming that transmittance spectra are considered).

Given an appropriate optimality criterion, an important problem is to determine the optimal number of filters, the properties of these filters and the performance of a sensor based on the application of these filters. Obviously, the cost of the filters, the necessary circuitry etc. will also affect the design choice in the final analysis. The optimal choice of selecting a number of filters for measuring the concentration of an equal number of substances simultaneously is a challenging problem.

The methodology proposed in this paper is to solve the problem by performing Monte Carlo simulation studies of a mathematical model based on a complete description of the optical components of the system. This approach has at least two interesting implications: It makes it possible to determine the achieveable performance of any given design provided that the technical specifications are available for the constituent optical elements of the sensor. It also makes it possible to take into account the transformations of the spectra introduced by the constituent optical elements of the system. In other words, this approach rules out expensive design, development and in situ testing of prototypes that would not perform optimally anyway.

The proposed methodology is widely applicable, but a particular application for measuring the NO_x-concentration in wastewater treatment plants is used as a case study throughout the paper.

Latent variable methods in chemometrics: Theoretical foundations and practical implications

John MacGregor, macgreg@mcmaster.ca, McMaster University, Department of Chemical Engineering. Hamilton, ON, Canada L8S 4L7, Canada

Keywords: PCA, PLS, latent variables

Latent variable models provide the theoretical foundation for nearly all analysis and inference using multivariate data. The concept that the measured variable spaces (X, Y) are not fundamental, but rather are dependent on a much more fundamental and lower rank set of latent variables (T) is a major paradigm shift from classical statistical analysis and inference methods. The tremendous power and utility of multivariate methods, such as PCA and PLS, in treating problems in science and engineering stem directly from the assumption of the latent variable model structure. It is this structure that allows us to handle missing data, test for outliers, monitor processes, invert models, and simplify the analysis of large data sets.

Unfortunately, literature on statistical inference based on these non-full rank latent variable models, is almost non-existent. To revert to the use of classical full-rank statistical inference methods in these situations is fraught with danger. This presentation provides an overview of these concepts and illustrates them with several examples.

Sorting of raw materials based on the predicted end product quality

Ingunn Berget, ingunn.berget@matforsk.no, MATFORSK. Oslovn. 1, 1430, Norway, http://www.matforsk.no
Tormod Næs, tormod.naes@matforsk.no, MATFORSK, Oslovegen 1, 1430 Ås, Norway

Keywords: sorting, fuzzy clustering, process optimisation, raw material quality, end product quality

Considerable variation in raw material quality often leads to an unstable end product quality. This is a common problem in the food industry and in other industries where the raw material quality can not be controlled.

There are several ways of handling such unwanted variation, for instance by robust process design or by statistical process control (SPC) together with feedback or feedforward adjustments. In this contribution we will present another methodology that can reduce the impact of unwanted raw material variation by sorting into homogenous categories. Sorting may lead to an easier operation of the process, because the need for frequent process adjustments is reduced.

The most obvious method for sorting raw materials is to use the similarity between the raw materials to define the categories. Sorting should however lead to a better and more stable end product quality. We have therefore based the sorting criterion on the predicted end product quality's closeness to the target. The first step in the sorting algorithm is therefore to estimate a model for prediction of end product quality from raw material and process variables. After this model is found the categories are found by cluster analysis of the raw material data. For this purpose fuzzy clustering is used. Fuzzy clustering is a partitioning method that assigns memberships to each object. The memberships indicate the degree of belonging an object has in each cluster. Fuzzy clustering is chosen because the memberships give useful information about cluster validity and it is flexible with respect to distance measure used. As a part of the cluster analysis the process variables are optimised for each category.

In this presentation the basic methodology is explained and illustrated by examples both for the single response and the multiresponse case.

Development of on-line image analytical industrial process monitoring calibrated against structurally correct sampling of heterogeneous materials

Kim Esbensen, kes@aue.auc.dk, Ålborg Universitet Esbjerg. Norgesgade 31, 1.th , DK-6700 , Denmark
Dawn Field Karlsrud, Telemark University College (HIT/TF), Institute of Process Technology (PT), Porsgrunn, Norway
Thorleif Simonsen, Telemark University College (HIT/TF), Institute of Process Technology (PT), Porsgrunn, Norway

Keywords: industrial, sampling, tutorial, principles

In cooperation with the Porsgrunn-based engineering company IDE-CON, which a.o. produces industrial mixers, internal mapping of a complete mixer volume as represented by 60 sub-samples taken with modified non-segregation sampling instrument/procedure has been carried out. This novel sampling scheme has allowed to depict the degree of (in)homogeneity in the entire 3-D mixer volume, which has lead to quantification of the degree of deviation from the stipulated "complete, homogenous mixing" result. The 3-D mapping resulted in determination of the structurally correct mean value of the effective mixer volume, which allowed estimation of the individual sampling bias for all 250 sub-specimens. These results can be expressed as interesting 3-D contour plots of the relevant concentrations/deviations in the mixer volume.

Two types of materials were used in the experiments, both directly related to ongoing large-volume industrial food production mixing processes: 1) minced meat (beef + fat) and vegetable mixtures (carrots, peas, maize). While minced meat is a very high-viscous material, experiencing very little, if any, segregation after "complete mixing", vegetable mixtures present a significant post-mixing transportation segregation problem (also when sub-sampling the 3-D mixer volume). This work also paid particular attention to solving this critical problem; we present two novel solutions, freeze-drying and gelatinuous fixation.

A continuous mixing process layout was also studied. Sampling was now performed only on the extruded product string on a conveyor belt in order to perform structurally correct 1-D sampling (sensu Pierre Gy). This layout was subjected to continuous image analytical characterisation of the top-surface of the extruded string (complete imaging), as a means of quantifying the relative sampling bias for the mixing/extruding/sampling process, for both types of sampling (image analytical vs. correct belt sub-sampling).

A set of "correct" reference samples from the conveyor belt made up the basis for an absolute calibration of the image analytical system (multivariate calibration), based on the AMT-transform + MAR (Multivariate AMT Regression).

Batch process monitoring using multivariate analysis. Recent developments and acceptance in industry

Dora Kourti, kourtit@mcmaster.ca, McMaster University. 1280 Main Street West, Hamilton, Canada

Keywords: batch process monitoring, multivariate process analysis

The foundations for monitoring batch processes using multivariate statistical process control (MSPC) methodologies were presented 10 years ago. Industry has adopted these methodologies to certain processes with remarkable success. However in other processes these ideas cannot be adopted yet, due to peculiarities related to the frequency of the process and quality data collection, and further research is needed.

In this paper we present the state of the art in batch process monitoring. Recent developments that allow the applicability of the MSPC methods to a wider range of batch processes are presented. Industrial applications are used to illustrate key issues in this area.

On-line batch fermentation process monitoring - Prediction of "Biological Time"

Pia Jørgensen, pia@bioteknologisk.dk, Biotechnological Institute. Holbergsvej 10, DK-6000, Denmark
Joan Grønkjær Pedersen, jgp@bioteknologisk.dk, Biotechnological Institute, Integrated ProcessOptimisation. Holbergsvej 10, DK-6000, Denmark, http://bioteknologisk.dk
Ejner Paaske Jensen, Biotechnological Institute, Holbergsvej 10, DK-6000 Kolding, Denmark
Kim Esbensen, kes@aue.auc.dk, Ålborg Universitet Esbjerg. Norgesgade 31, 1.th , DK-6700 , Denmark

Keywords: batch monitoring, MSPC, MPCA, MPLS, biological time, on-line sensors

On-line Near InfraRed (NIR) Spectroscopy, Multichannel Fluorescence and Membrane Inlet Mass Spectrometry (MIMS) enable detailed description of industrial fermentation processes to document progress and to ensure end-product quality. Multivariate statistical control charts allow easy and efficient identification of abnormal fermentations - often even at an early stage of the fermentation, which is critical for industrial production. This contribution focuses on the time variable in general and prediction of "biological time" in particular.

Fermentations are complex biological processes in which different kinds of biological matter are transformed into high value products (e.g. enzymes, antibiotics and starter cultures). Production by fermentation is widely used in both the food and the pharmaceutical industries.

In fermentation processes each batch run should theoretically have the same duration and the monitoring variables should follow closely some predetermined "optimal" trajectories w.r.t. time. However, in real-world industrial fermentation processes the duration of the process often varies significantly from batch to batch due to variation in the culture and in the raw materials etc. The basic shape of the time trajectories from batch to batch are similar, but their time duration vary adversely. In order to analyse such data reliably, it is necessary to account properly for this varying batch duration. Furthermore critical attention has to be made to the fact that all batch monitoring data are to be analysed prospectively in real time, i.e. before the actual duration of the batch is known. In such cases it is necessary to predict the equivalent "biological time" before proper alignment, and comparison, of batch data is possible.

In this presentation it will be shown that the three analytical techniques (NIR, Fluorescence and MIMS) supplement one another advantageously both in describing several semi-industrial on-going fermentation processes as well as in predicting the specific chemical compounds involved, metabolites and substrates in the fermentation. Special attention will be made on how the time variable is handled making it possible to construct multivariate statistical control charts with significance levels based on the delineation of a relevant set of historical trajectories for normal fermentations.

Dealing with missing data in MSPC: Several methods, different interpretations, some examples

Francisco Arteaga, fjarteaga@fee.edu, Facultad de Estudios de la Empresa. Dpto. Métodos Cuantitativos C/ Guillén de Castro 175, 46007 Valencia, Spain
Alberto Ferrer, aferrer@eio.upv.es, Universidad Politécnica de Valencia, Dpto. Estadística, Quality Improvement Group. Camino de Vera S/N Edif. I-3, 46022 Valencia, Spain, http://www.upv.es/deio

Keywords: missing data, PCA, MSPC

In modern process industries Principal Component Analysis (PCA) is widely used to develop models from data sets with large numbers of highly correlated variables collected continually from sensors hooked up to continuous and batch processes. Once a PCA model has been built, it can be applied in multivariate statistical process control (MSPC) schemes to monitor and diagnose future process operating performance.

One problem in this environment is that missing measurements are a common occurrence due to different causes: sensor failure, sensor routine maintenance, gross measurement errors, samples not collected at the appropriate time, sensors having different sampling periods, and so on. In batch process monitoring some method for filling in the unknown data between the current time interval and the end of the batch is needed. So, since some future multivariate observations will also have missing values, the PCA models would be of limited value unless methods were available to handle missing data. From the fact that only a few underlying events are driving a process at any time, and measurements on all the process variables are simply different reflections of the same underlying events, missing values can often be estimated from the measured data.

This work deals with estimating scores from an existing PCA model when new observation vectors are incomplete. Several methods are analysed: a method, termed trimmed scores; the single component projection method, derived from the NIPALS algorithm for model building with missing data; a method of projection to the model plane; a method based on iterative imputation of missing values; a method who minimise the squared prediction error; a Hotelling´s T2 statistic minimisation method; a conditional mean replacement method; and several least squares methods based on the training data set.

Several methods lead to the same analytical solution. Expressions for the errors in the estimated scores for the different methods are developed. Using simulated data, an experimental design is carried out to study the influence of different factors (loading vector characteristics, number of high variance dimensions, estimating method, ...) on the estimated scores errors. The performance of the different methods analysed in this work is evaluated on an industrial data set.

References:
1. P. R. C. Nelson, P. A. Taylor, and J. F. MacGregor, Chemometrics and Intelligent Laboratory Systems 35, 45-65 (1996).
2. B. Grung, and R. Manne, Chemometrics and Intelligent Laboratory Systems 42, 125-139 (1998).
3. H. A. L. Kiers, Psychometrika 62, 251-266 (1997)

Spectra of wavelet scale coefficients of process acoustic measurements as input for PLS modelling of pulp quality

Anders Björk, anbj@analyt.kth.se, Royal Inst. of Technology (KTH), Dep. of Chemistry, Div. of Analytical Chemistry. Royal Inst. of Technology (KTH), SE-100 44 Stockholm, Sweden, http://www.analyt.kth.se/department/
Lars-Göran Danielsson, lgd@analyt.kth.se, Royal Inst. of Technology (KTH), Dep. of Chemistry, Div. of Analytical Chemistry. Royal Inst. of Technology (KTH), SE-100 44 Stockholm, Sweden, http://www.analyt.kth.se/department/

Keywords: acoustics, wavelets, FFT, PLS, pulp

Acoustic and vibration signals are captured by simple standard accelerometers. These can be mounted either directly on operative process equipment creating a completely non-invasive measurement system or in connection with a constriction designed to give a strong and reproducible disturbance of the process flow. The signal from the accelerometer is then amplified, digitised by an analogue to digital converter and stored in some suitable format in a PC. Constantly cheaper and more powerful computers facilitate the digital signal processing needed.

Before trying to correlate the acoustic signals to process parameters the signals will have to be transformed into a more useful form. The method most often used has been to apply variants of Fast Fourier Transform, FFT, on sampled data to produce a frequency domain representation. An alternative way could be to use the Fast Wavelet Transform, FWT, in combination with FFT. The FWT has the drawback that it produces time resolved representations but the advantage that on each scale different features are extracted. The wavelet step can be seen as a pre-filtering before FFT but instead of making FFT on the complete time series it is done on coefficients at each wavelet scale.

We have used spectra of wavelet scale coefficients in an attempt to model pulp quality with PLS. In this case the number of points in the resulting spectrum from the FFT can be limited to a low number, e.g. 16 compared to 2048 with direct FFT on the time series. In the PLS modelling step the advantage is that the first two components describe Y better than with conventional approach, e.g. 84 % explained Y-variance compared to 36 %. A second advantage is that the model requires fewer coefficients, e.g. 238 compared to 1025. This treatment of acoustic time series data also opens a possibility to use three-way methods. With such methods at least partially separated coefficients corresponding to the wavelet scale and to the frequencies in the FFT step respectively could be obtained.

The choice of mother wavelet is an important issue in this case because some wavelets capture transient behaviour of signals while other suppress such features.

New and old trends in Chemometrics - How to deal with the increasing data volumes in research, development, and production with examples from pharmaceutical research and process modelling

Svante Wold, svante.wold@umetrics.com, University of Umeå. , Sweden
Nouna Kettaneh, Umetrics Inc., 17 Kiel Ave, Kinnelon, NJ 07405, USA

Keywords: large data sets, data mining non-linear PLS, scalability, interpretability

Chemometrics was started around 30 years ago to cope with and utilize the rapidly increasing volumes of data produced in chemical laboratories. The methods of early chemometrics were mainly focused on the analysis of data, but slowly we came to realize that it is equally important to make the data contain reliable information, and methods for design of experiments (DOE) were added to the chemometrics toolbox. This toolbox is now fairly adequate for solving most R&D problems of today in both academia and industry, as will be illustrated with a few examples.

However, with the further increase in the size of our data sets, we start to see inadequacies in our multivariate methods, both in their efficiency and interpretability. Drift and non-linearities occur with time or in other directions in data space, and models with masses of coefficients become increasingly difficult to interpret and use.

Starting from a few examples of some very complicated problems confronting chemical researchers today, possible extensions and generalizations of the existing chemometrics methods, as well as more appropriate preprocessing of the data before the analysis, will be discussed. Criteria such as scalability of methods to increasing size of problems and data, increasing sophistication in the handling of noise and non-linearities, interpretability of results, and relative simplicity of use, will be held as important.

The discussion will be made from a perspective of the evolution of the scientific methodology as driven by new technology, e.g., computers, and constrained by the limitations of the human brain, i.e., our ability to understand and interpret scientific and data analytic results.

100 years old and still a cutting edge method: Principal component analysis in gene expression analysis

Jürgen von Frese, j.von-frese@icb-online.de, Institut für Chemo- und Biosensorik, Chemometrics. Mendelstr. 7, D-48149, Germany, http://www.icb-online.de
Anton Belousov, a.belousov@icb-online.de, Institut fuer Chemo- und Biosensorik. Mendelstr. 7, D-48149, Germany, http://www.icb-online.de
Eric A. Frauendorfer, Institut für Chemo- und Biosensorik, Mendelstr. 7, D-48149 Münster, Germany

Keywords: PCA, bioinformatics, microarrays, cancer

In 1901 K. Pearson published his article "On lines and planes of closest fit to systems of points in space" [1]. This was the first description of the method which lies at the very heart of chemometrics: Principle component analysis. It was used in chemistry even before the advent of the discipline of "chemometrics". A large number of our methods are ultimately based on PCA, e.g. SIMCA, PCR, factor analysis; its application can certainly be considered as routine in chemometrics. Does that mean that there is nothing of interest about PCA anymore?

The advent of the "post genomic era" has resulted in an enormously increased importance of data analysis for molecular biology. For example, microarrays for monitoring gene expression have replaced the traditional "one gene at a time" approach by an overwhelming 10^3 - 10^4 genes for every sample. The current flourishing of bioinformatics has even led to a reorientation in supercomputing and an engagement of computer hardware companies into the field.

In the search for "new" methods, which can be applied to the analysis of gene expression data, the bioinformatics community recently adopted principle component analysis for "Making the most of microarray data" [2].

We introduce the microarray technique and the data analysis problems connected with it. Exemplary an analysis of current publicly available cancer data sets is shown with an emphasis on the use of PCA. Even after 100 years, PCA is still vivid enough to help in making a considerable progress in current cancer diagnostics and the understanding of cancer.

References:
[1] K. Pearson, Phil. Mag. 2 (1901) 559
[2] T. Gaasterland, S. Bekiranov, Nature Genetics 24 (2000) 204

Case studies in the application of maximum likelihood principal components analysis

Peter Wentzell, peter.wentzell@dal.ca, Dalhousie University, Department of Chemistry. Halifax, Nova Scotia, B3H 4J3, Canada, http://www.dal.ca

Keywords: maximum likelihood, PCA

Maximum likelihood principal components analysis (MLPCA) is a generalization of principal components analysis which is based on the premise that incorporation of measurement error information can lead to better multivariate models. While PCA is based on modeling variance in the data, MLPCA utilizes information about measurement uncertainty to develop models which attempt to distinguish the variance in the data from the variance in the noise. This encompasses not only the case of heteroscedastic noise, but also, and perhaps more importantly, the case of correlated measurement errors. Correlated measurement errors, often characterized by terms such as “drift”, “offset”, “flicker noise”, “pink noise”, and “cross-talk”, are ubiquitous features of analytical measurements, but are rarely treated in data analysis. MLPCA is a general approach which can accommodate virtually any measurement error structure, although in practice a number of simplifying constraints can greatly expedite the implementation of the algorithm.

While MLPCA has been applied to problems ranging from the estimation of missing data to calibration transfer, its most successful application has been in multivariate calibration. The derived regression methods, maximum likelihood principal components regression (MLPCR) and maximum likelihood latent root regression (MLLRR), have been applied to a number of problems and have demonstrated a significant improvement in prediction errors. Nevertheless, applications to date have not been extensive, and so the generalization of these results remains an open question. This is particularly true since the improvement observed is closely tied to the underlying error structure in the data, which can vary from application to application. In this presentation, a variety of applications of MLPCR will be described in an attempt to draw some general conclusions about its advantages and limitations. Data will be drawn from a variety of sources, ranging from near-infrared to fluorescence, so that a broad window of error structures are encompassed. Additionally, observations related to the practical implementation of MLPCR will be discussed.

Simple interval calculation - A method for linear modeling

Oxana Rodionova, rcs@chph.ras.ru, Polycert, Semenov Institute of Chemical Physics. 4, Kosygin Str., Moscow, Russia, http://polycert.chph.ras.ru
Alexey Pomerantsev, polycert@chph.ras.ru, Semenov Institute of Chemical Physics, Polycert. 4, Kosygin Str, Moscow, Russia, http://polycert.chph.ras.ru

Keywords: prediction uncertainties, non-regression approach, multivariate calibration, interval estimations, linear programming

A method of simple interval calculation (SIC) is proposed for linear analysis of chemometric data. This method provides with the results of modeling in the interval form. It should be emphasized that SIC-method has nothing in common with interval mathematics. The current approach assesses the uncertainty of predicted values in such a way that each point of yielding interval has equal likeliness and helps to establish realistic prediction intervals in practical situation. This is a non-regression approach because it does not use an objective function (e.g., sum of squares) for estimate search. In regression analysis the estimates are the values of unknown parameters that agree with experiment in the best way. In the current method any parameter value that does not contradict the experimental data is accepted as a feasible estimate. It is natural that such an approach has sense only under some assumptions. The only assumption of the method is that a measurement error is limited. The roots of the method are in the old ideas of Kantorovich to apply the linear programming to the data analysis. The calculation aspects of SIC-method are rather simple since they founded on the well-designed Simplex algorithm.

No doubt that multivariate problems where data matrix is rank- deficient are of great practical interest. To apply SIC-method to such kind of problems we join it with traditional projection methods (e.g., principal component analysis or partial least squares).

We consider that the criteria of quality of interval prediction used in SIC-procedure allow to look at the old problems of multivariate data analysis from a new point of view. These problems are optimum number of PCs, outlier detection, missing data, and insignificant observations.

Important and influential variables and samples in latent structure models

Agnar Høskuldsson, ah@akp.dtu.dk, Danish Technical University. Danish Technical University, Lyngby, Denmark

Keywords: latent structure, causality, importance, extremes, inverse

Here we present a collection of methods that can be used to judge the importance and influence of variables and samples (objects) on the latent structure that has been developed by a modelling procedure. The latent structure T is the part of data that the modelling is based upon. In many types of linear models it can be derived from X as T=XR. The matrix R shows how the latent structure T is derived from X, the first column of R gives the first score vector and so on. Thus, the r-vectors, columns of R, can be interpreted as 'causal' vectors that show how the variables generate the latent variables (corresponding to the columns of T). We show how we can work with the r-vectors with the purpose of revealing the causal structure in data. We show some methods how we can rank the variables and samples according to how much they describe of T. These methods can be used to identify outliers or groups of outliers.

We discuss some basic issues concerning importance or requirements to variables. The literature is often imprecise or incorrect concerning many related issues, e.g., statements like: Variables that do not show significant regression coefficients (e.g., by re-sampling procedures), can or should be eliminated from the analysis.

In many cases we can write **X**=TDP', where T has orthogonal columns and D is a diagonal matrix of scaling constants that can be defined such that T'T=inv(D). In these cases we can write P=(X'X)R. It shows that we can study the correlation structure using the R matrix. We show how we can work with the r-vectors with the purpose of studying the correlation structure in data.
An important issue is the case, where there are many variables that have no or very little importance. We discuss what we should do in such situations.

Regression methods are often judged by the size of $s^2(X'X)^{-1}$, where s^2 is the residual variance. The residual variance s^2 gets usually stable when some dimensions have been selected. The size of the inverse, inv(X'X) (the diagonal elements) on the other hand may increase rapidly, when more dimensions (score/variables) are selected. We show methods that can be used to keep the inverse small, while the modelling is carried out. Everything else equal, the smaller the inverse is the better is the model to predict using future samples.

Decomposition of spectra using maximum autocorrelation factors

Rasmus Larsen, rl@imm.dtu.dk, DTU, Informatics and Mathematical Modelling. IMM, building 321, DK-2800, Denmark, http://www.dtu.dk

Keywords: maximum autocorrelation factors, independent components, principal components, PCA

This paper addresses the problem of generating a low dimensional representation of the variation present in a set of spectra, e.g. reflection spectra recorded from a series of objects. The resulting low dimensional description may subsequently be input through variable selection schemes into classification or regression type analyses. A featured method for low dimensional representation of multivariate datasets is Hotellings principal components transform.

We will extend the use of principal components analysis incorporating new information into the algorithm. This new information consists of the fact that given a spectrum we have a natural order of the input variables. This is similar to Switzers maximum autocorrelation factors, where a natural order of observations (pixels) in multi-spectral images is utilized. However, in order to utilize an ordering of the input variables we need a non-trivial reformulation of the maximum autocorrelation problem in Q-mode. We call the resulting transformation for Q-MAF. The resulting new variables can be interpreted as a frequency decomposition of the spectra. But contrary to ordinary Fourier decomposition these new variables are located in frequency as well as well wavelength.

Independent Component Analysis - A new valuable tool in data analysis or the Emperor's new clothes ?

Frank Westad, frank.westad@matforsk.no, MATFORSK. Arildsvingen 12, Oslo, Norway,
http://www.matforsk.no

Keywords: independent components analysis, PCA, ICA

One frequently common objective in data analysis is to extract "pure sources" from a complex set of signals. Examples of such are how to find the pure spectra from a mixture, or profiles from sources of pollution. Principal Component Analysis (PCA) is often applied in chemometrics for explorative data analysis, sometimes combined with some sort of rotation of the axis to get the underlying structures. However, since the criterion in PCA is to maximise variance, one can not expect to perfectly extract the pure spectra from a set of mixtures.

Independent Component Analysis (ICA) is a method that tries to recover the original signals by finding a linear transformation, which minimises the mutual entropy between them. This recently developed method has attracted attention the past years in fields such as signal processing, image analysis and blind source separation. The theory of ICA is briefly presented; the statistical assumptions and ICA's connection to neural networks. The performance is compared to PCA for different applications, and ICA is discussed in terms of model validation and stability.

Orthogonal-PLS (O-PLS) for removing irrelevant variation in X variables

Johan Trygg, johan.trygg@chem.umu.se, University of Umeå, Research Group for Chemometrics. 901 87, Umeå, Sweden, http://www.chem.umu.se/dep/orgchem/forskning/chemometrics/index.stm
Svante Wold, svante.wold@umetrics.com, University of Umeå. , Sweden

Keywords: preprocessing, quantitative interpretation, O-PLS, PLS

Data measured on complicated samples and in complicated processes contain contributions from many sources and several types of noise. This presents difficulties in the analysis and modelling of such data. Preprocessing methods can be applied in those situations, such as variable selection techniques that search for optimal sets of variables. These methods are susceptible to chance correlation unless apriori physical information is used to select variables that are known to be important to the analysis. Stepwise regression, genetic algorithm, and simulated annealing selection of variables, are all prone to chance correlations due to their search of very large numbers of possible solutions. Sometimes variable selection methods are warranted, for instance when each variable is associated with a cost. However, the interpretation of the resulting multivariate calibration models can suffer and future detection of outliers are often affected when only a small subset of the original variables are used.

Alternatively, it can be advantageous to use a method that does not remove variables but removes the influence of the irrelevant variation in variables. In this paper, a generic preprocessing method called orthogonal-PLS (O-PLS) is described. O-PLS (1) removes variation from X (descriptor variables) that is not related to Y (response variables, for example yield, cost or toxicity). In mathematical terms, this is equivalent to removing systematic variation in X that is orthogonal to Y. In an earlier paper (2), Wold et al. described the orthogonal signal correction (OSC). O-PLS is a method with the same objective, but with different means of achieving that. The non-relevant systematic variation in X is removed, improving interpretation of the resulting model and with the additional benefit that the non-correlated variation can be studied and analyzed further. This approach is reviewed and exemplified. Other OSC related methods are also discussed in this context.

References:
1. Trygg J and Wold S, Orthogonal projections to latent structures, O-PLS. Journal of Chemometrics, submitted 2000
2. Wold S, Antti H, et al., Orthogonal signal correction of near-infrared spectra. Chemometrics and Intelligent Laboratory Systems 44(1-2): 175-185 (1998)
3. Fearn T, On orthogonal signal correction. Chemometrics and Intelligent Laboratory Systems 50(1): 47-52 (2000)

A discussion on orthogonal signal correction

Johan Westerhuis, westerhuis@its.chem.uva.nl, Chemical Engineering, University of Amsterdam, Process Analysis and Chemometrics. Nieuwe Achtergracht 166, 1018 WV AMSTERDAM, The Netherlands, http://www-its.chem.uva.nl/research/pac/
Sijmen de Jong, Unilever Research Vlaardingen, POBox 114, 3130 AC Vlaardingen, The Netherlands
Age Smilde, asmilde@its.chem.uva.nl, University of Amsterdam, Department of Chemical Engineering, Process Analysis and Chemometrics. Nieuwe Achtergracht 166, NL-1018 WV Amsterdam, Netherlands, http://www.uva.nl

Keywords: preprocessing, orthogonal signal correction, direct orthogonalization

In this presentation, the concept of orthogonal signal correction (OSC) as a spectral preprocessing method is discussed and a number of OSC algorithms that have appeared [1-5] are compared from a theoretical viewpoint. Since all of these algorithms had some problems concerning the orthogonality towards Y, non-optimal amount of variance removed from X, or a non-attainable solution, a new direct OSC algorithm (DOSC) is introduced. DOSC was originally developed as a direct method solely based on least squares steps that had none of the problems mentioned above. The first practical results with the new method, however, were not encouraging due to the complete orthogonality constraint. If this orthogonality constraint is loosened, the method improves considerably and simplifies the calibration model for the prediction of Y. Besides the theoretical comparison, some applications will be presented where the methods are compared will be presented.

References:
1. Wold,S., Antti H., Lindgren,F. & Ohman,J. Orthogonal signal correction of near-infrared spectra. Chemometrics and Intelligent Laboratory Systems 44 (1998) 175-185.
2. Sjöblom,J., Svensson,O., Josefson,M., Kullberg,H. & Wold,S. An evaluation of orthogonal signal correction applied to calibration transfer of near infrared spectra. Chemometrics and Intelligent Laboratory Systems 44 (1998) 229-244.
3. Andersson,C.A. Direct orthogonalization. Chemometrics and Intelligent Laboratory Systems 47 (1999) 51-63.
4. Fearn,T. On orthogonal signal correction. Chemometrics and Intelligent Laboratory Systems 50 (2000) 47-52.
5. Wise,B.M. & Gallagher,N.B. http://www.eigenvector.com/MATLAB/OSC.html. 1999.

QSAR modeling of polychlorinated dibenzofurans using a 3D structure representation using quantum topology (StruQT)

Bjorn Alsberg, bka@aber.ac.uk, University of Wales, Aberystwyth, Department of Computer Science, Computational Biology Group. Llandinam bldg, Ceredigion, SY23 3DB, United Kingdom, http://www.aber.ac.uk

Keywords: QSAR, quantum topology, atoms in molecules, dibenzofurans

The recently developed 3D molecular structure representation using quantum topology (StruQT) is applied to a QSAR modeling of a set of polychlorinated dibenzofurans. Central to StruQT is the use of critical points in scalar fields where the gradient is zero. Commonly used fields are the electron density field, its Laplacian and the electrostatic potential of the molecules. It has been demonstrated that in particular the bond critical point (BCP) of the electron density field is useful in QSAR/QSPR modeling. The BCP effectively summarises the electronic properties of a bond in a molecule and can usually be characterised using only three parameters (in addition to their spatial coordinates). Some of the parameters that can be used to describe the BCP are the electron density, the Laplacian at the BCP and the ellipticity. StruQT has the advantage that there is no need to use a large number of 3D lattice points in the QSAR model to capture the important parts of the 3D electronic structure as is the case for e.g. the comparative field analysis (CoMFA) technique.

Investigation of the autofluorescence from cod extracts

Charlotte Møller Andersen, chm@dfu.min.dk, Danish Institute for Fisheries Research. DTU, build. 221 , 2800, Denmark
Bo Jørgensen, boj@dfu.min.dk, Danish Institute for Fisheries Research, Department of Seafood Research, Raw-material and product technology. DTU build. 221, DK-2800, Denmark, http://www.dfu.min.dk
N. Bøknæs, Danish Institute for Fisheries Researchy, Department of Seafood Research, Søltofts Plads, Technical University, Building 221, DK-2800 Kgs. Lyngby
Rasmus Bro, rasmus@optimax.dk, The Royal Veterinary & Agricultural University, Dept. of Dairy and Food Science, Food Technology. Rolighedsvej 30, 1958, Frederiksberg, Denmark, http://www.models.kvl.dk/users/rasmus/

Keywords: fluorescence, PARAFAC, cod, multiway, validation

The potential of fluorescence spectroscopy for estimating the quality of fish products will be investigated by using multi-way methods. This study is an attempt in understanding the fluorescent properties of cod, which in the longer run should be used to develop fast on-line/at-line methods. The development of on-line/at-line methods is important for registrating quality attributes ensuring that the final products meet the expectations of the consumers.

Muscle foods are complex materials with many compounds contributing to the autofluorescence. Therefore, to increase the simplicity and ease the interpretability the experiment has been performed on aqueous extracts. Extracts with a pH of 5.2 have been prepared from chill stored thawed cod fillets of varying quality packed in modified atmosphere.

Fluorescence landscapes are obtained and analysed with the multi-way method PARAFAC. The uniqueness of the PARAFAC model makes it possible to decompose the complex signals into contributions from individual chemical components. Relating this information to the knowledge about the cod extracts will make it possible to analyse, understand and describe factors influencing the fluorescence of cod.

The development and validation of the PARAFAC model will be described. Problems of finding the optimal PARAFAC model will make up an important part of the presentation. This includes choosing the optimal number of components, the use of constraints and problems with missing values, scatter and other non-trilinearities. Furthermore, an interpretation of the model with respect to the process parameters and chemistry of the extracts will be presented.

Design, synthesis and QSAR evaluation of a chemical library directed towards the melanocortin receptors

Per Andersson, per.andersson@melacure.com, Melacure Therapeutics. Ulleråkersvägen 38, SE 756 43, Sweden, http://www.melacure.com
Torbjörn Lundstedt, torbjorn.lundstedt@melacure.com, Melacure Therapeutics. Ulleråkersvägen 38, SE 756 43, Sweden
Elisabeth Seifert, Melacure Therapeutics AB, Ulleråkersvägen 38, SE-756 43, Uppsala, Sweden
Anna Skottner, Melacure Therapeutics AB, Ulleråkersvägen 38, SE-756 43, Uppsala, Sweden
Arne Boman, Melacure Therapeutics AB, Ulleråkersvägen 38, SE-756 43, Uppsala, Sweden

Keywords: DOE, QSAR, melanocortin

Five melanocortin receptors have to date been identified and there are evidence that two of them are important for modulating the immunosystem and the feeding behaviour [1], hence there is a great interest to identify small organic compounds which influence these receptors. By using Statistical Molecular Design (SMD) [2-5] a number of chemical libraries has been planed, synthesised and tested. One example presenting the whole drug discovery process from the design to the Quantitative Structure-Activity Relationship evaluation is presented.

References:
1. Wikberg, J.E., Melanocortin receptors: perspectives for novel drugs. Eur J Pharmacol, 1999. 375(1-3): p. 295-310.
2. Linusson, A., et al., Statistical Molecular Design of Building Blocks for Combinatorial Chemistry. J. Med. Chem., 2000. 43(7): p. 1320-1328.
3. Andersson, P.M., et al., Design of Small Libraries for Lead Exploration. Molecular Diversity in Drug Design, ed. R. Lewis and P.M. Dean. 1999, Dordrecht, The Netherlands: Kluwer Academic Publishers. 197-220.
4. Wold, S., et al., eds. Multivariate Design and Modelling in QSAR, Combinatorial Chemistry, and Bioinformatics. Molecular Modeling and Prediction of Bioactivity, ed. F.S.J. K. Funddertofte. 2000, Kluwer Academic/Plenum Publishers: New York. 27-45.
5. Lundstedt, T., et al., Intelligent Combinatorial Libraries. Computer-Assisted Lead Finding and Optimization, ed. H.v.d. Waterbeemd. 1997, Weinheim: VCH Verlagsgesellshaft mbH. 189-208.

Batch statistical processing of 1H NMR-derived urinary spectral data

Jahanara K. Azmi, Biological Chemistry, Biomedical Sciences Division, Sir Alexander Fleming Building, Imperial College of Science Technology & Medicine, London, SW7, 2AZ, UK
J. L. Griffin, Biological Chemistry, Biomedical Sciences Division, Sir Alexander Fleming Building, Imperial College of Science Technology & Medicine, London, SW7, 2AZ, U.K.
H. Antti, Biological Chemistry, Biomedical Sciences Division, Sir Alexander Fleming Building, Imperial College of Science Technology & Medicine, London, SW7, 2AZ, U.K.
S. D. Jones, BG Technology, Loughborough, Leics., UK.
R. F. Shore, Centre for Ecology and Hydrology, Monks Wood, Abbots Ripton, Huntingdon, Cambs., PE28 2LS, U.K.
J. K. Nicholson, Biological Chemistry, Biomedical Sciences Division, Sir Alexander Fleming Building, Imperial College of Science Technology & Medicine, London, SW7, 2AZ, U.K.
E. Holmes, elaine.holmes@ic.ac.uk, Biological Chemistry, Biomedical Sciences Division, Sir Alexander Fleming Building, Imperial College of Science Technology & Medicine, London, SW7, 2AZ, U.K.

Keywords: NMR, PLS, visualisation

Multivariate statistical batch processing (BP) analysis of 1H NMR urine spectra was employed to establish time-dependent metabolic variations in animals treated with the model hepatotoxin, a-Naphthylisothiocyanate (ANIT). ANIT was administered orally to rats (n=5) at 100 mg/kg and urine samples were collected from dosed rats and from matched control rats at time-points up to 168h post-dose. Urine samples were analysed via 1H NMR spectroscopy and Partial Least Squares (PLS) based batch processing analysis was used to investigate the 1H NMR spectra, treating each rat as an individual batch comprising a series of timed urine samples. A model defining the mean urine profile was established for the control group and samples obtained from ANIT treated animals were assessed using this model.

Time-dependent deviations from the control model were evident in all ANIT treated animals and hepatotoxicity was manifested by glycosuria, marked reduction of Tricarboxylic acid (TCA) cycle intermediates, bile aciduria, and elevated taurine. Furthermore, BP plots showed a characteristic pattern for ANIT, allowing the time-points at which there were maximum metabolic differences to be determined and provided a means of visualising the net ANIT-induced metabolic movement of urinary metabolism. BP may prove to be a powerful metabonomic tool in defining time-dependent metabolic consequences of toxicity and is an efficient means of visualising inter-animal variations in response as well as defining multivarite statistical limits defining normal physiological variation in terms of biofluid composition.

Omeprazole and analogue compounds: A QSAR study of activity against *helicobacter pylori* using theoretical descriptors

Aline Thaís Bruni, alinetb@df.ibilce.unesp.br
Márcia M. C. Ferreira, marcia@iqm.unicamp.br, Universidade Estadual de Campinas, Instituto de Química. Campinas, SP, 13083-970, Brazil

Keywords: Omeprazole, PLS, Hartree-Fock, QSAR

Helicobacter pylori usually lives in the stomach and requires urease enzyme to colonize mucus layer.[1] It plays an important role in peptic ulcer disease, and the bacterium eradication decreases the ulcer recurrence.[2,3] These facts motivates a search to find new treatments. As an example, omeprazole and analogues have been studied against *Helicobacter pylori* action.

Kühler *et al.*[4] presented a chemometric model to predict new compounds activity using experimental descriptors, however most of its features are not explicit. In present work, omeprazole and some analogues were studied theoretically, and the results were compared to Kühler's[4]. Theoretical descriptors were calculated for all drugs and were used to construct a new predictive model. Initially, conformational analysis was performed for all compounds. The novel methodology for systematic search coupled with PCA was used to find all energy minima structures for each drug. The PM3 semi-empirical method implemented in Gaussian 98 package was used. The descriptors were calculated by using the Hartree-Fock method at *ab initio* level (6-31G**), implemented in Spartan Pro package.

The properties used as descriptors were: electronic energy, heat of formation, atomic charges on the principal substituents on the basic structure, dipole moment (total and x, y, z components), HOMO and LUMO energy, electronegativity, hardness, molecular mass, volume, area, and ovality among others.

The PLS regression method was used to build the QSAR models. The obtained models are shown to be much better suited for prediction than those presented in the literature.[4] The theoretical descriptors included were important to improve the models' efficiency.

The authors acknowledge the financial support from FAPESP.

References:
1. B. J. Marshall, Helicobacter pylori, *Am. J. Gastroenterol.*, **89**, S116 (1994).
2. Y. Glupczynski, A. Burette, *Am. J. Gastroenterol.*, **85**, 1545 (1990).
3. N. Chiba, B. V. Rao, J.W. Rademaker, R. H. Hunt, *Am. J. Gastroenterol.*, **87**, 1716 (1992).
4. T. C. Kühler, J. Fryklund, N. A. Bergaman, J. Weiliotz, A. Lee, H. Larsson, *J. Med. Chem.*, **38**, 4906 (1995).

Level of validation controlled by the choice of segmentation, in the cross-validation/jack-knifing of bi-linear regression models

Derek Byrne, dby@kvl.dk, The Royal Veterinary and Agricultural University, Dairy and Food Science, Sensory Science. Rolighedsvej 30, 5 sal., 1958, Denmark
Frank Westad, frank.westad@matforsk.no, MATFORSK. Arildsvingen 12, Oslo, Norway, http://www.matforsk.no
Harald Martens, Harald.Martens@mail.tele.dk, DTU/NTNU/KVL . Teglgaardstr 12A , DK-1452 , Denmark

Keywords: ANOVA, PLS, validation, jack-knifing

Tools are demonstrated for assessing the predictive ability and the parameter stability of bilinear regression models at different levels of validity, ranging from repeatability via interpolation ability to reproducibility. This is illustrated for APLSR method - the ANOVA-like use of Partial Least Squares Regression (PLSR). The data in the application concerns relating a sensory response y to experimental design indicator variables X in a study of flavour quality in warmed-over meat (Byrne et al. 2001).
The estimated Mean Square Error of Prediction (MSEP) for y-predictions ^y=X^b is assessed statistically at different validation levels (Martens & Martens 2001). This is attained by changing the method of segmentation of the sample set in the cross-validation. It is shown how this choice of validation level affects both the estimated prediction error itself, as well as the uncertainty in estimating this prediction error.
The estimated PLSR parameters (scores, loadings and regression coefficients) are validated by the extended cross-validation/Jack-knifing technique. While the estimated model parameters are the same at the different validity levels, their estimated uncertainty was found to vary with validation level.

References

D. V. Byrne, W. L. P. Bredie, L. S. Bak, G. Bertelsen, H. Martens and M. Martens (2001). Sensory and chemical analysis of cooked porcine meat patties in relation to warmed-over flavour and pre-slaughter stress. (accepted for publication in Meat Science)

Martens, H. and Martens, M. (2001). Multivariate Analysis of Quality. An Introduction. J. Wiley & Sons Ltd., U.K.

Analysis of residuals: Statistical method in QSAR studies

A. Dmitriev, a_v_dmitriev@mail.ru, Department of Theoretical Physics, Lipetsk State Pedagogical University, Lipetsk 398020, Lenin Str., 42, Russia
G. Isaeva, kgpu@kosnet.ru, Department of Theoretical Physics, N.A. Nekrasov Kostroma State University, Kostroma 156000, Pervogo Maya Str., 14, Russia
P. Isaeva, kgpu@kosnet.ru, Department of Theoretical Physics, N.A. Nekrasov Kostroma State University, Kostroma 156000, Pervogo Maya Str., 14, Russia

Keywords: residuals, QSAR

This work is devoted to the analysis of QSAR reliability, in which, predictors are calculated additive molecular descriptors. For the analysis of QSAR reliability, is offered the method of the analysis of medicinal preparations molecules by structural fragments molecules. For the structural fragments molecules by the methods regression and correlation analysis compared the experimental and calculated additive molecular descriptors.

The analysis of variance of residuals and analysis of residuals on the Durbin-Watson statistics this regression equation allows to find the best computational method of additive molecular descriptors estimates, a reason of outliers existence. By this method is obtained a reliability regression equations linking minimum blocking concentration with a projection of a dipole moment vector on the plane perpendicular axes of a maximal moment of inertia, electrical polarizability, molecular and optical anisotropy. The molecular descriptors were calculated by a CNDO/2, INDO, MINDO/3, ZINDO/1, MNDO, MNDO/AM1, and MNDO/PM3 methods. The minimum blocking concentration at a conduction anesthesia was defined by a Skou method. Out-going from outcomes of the analysis of residuals is obtained that the CNDO/2 method gives the best estimation of electrical dipole moment. The MNDO, MNDO/AM1, and MNDO/PM3 methods gives the best estimations of electrical polarizabilities, molecular and optical anisotropy.

This work was supported by the CCFE Grant (code 34.17.15. - Biophysics, cipher 97-0-10, 0-167).

Quantitation of the active substance in a pharmaceutical tablet using Near Infrared (NIR) transmittance spectroscopy and chemometrics

Marianne Dyrby, md@kvl.dk, The Royal Veterinary and Agricultural University, Department of Dairy and Food Science, Chemometrics Group. Rolighedsvej 30, 1958 Frederiksberg, Denmark, http://www.models.kvl.dk
Søren Balling Engelsen, se@kvl.dk, The Royal Veterinary and Agricultural University. Rolighedsvej 30, 1958, Denmark
Lars Nørgaard, lan@kvl.dk, The Royal Veterinary and Agricultural University, Chemometrics Group, Department of Dairy and Food Science. Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark, http://www.models.kvl.dk
Mette Bruhn, meb@lundbeck.com, H. Lundbeck A/S, Ottiliavej 9, 2500 Valby, Denmark
Line Lundsberg-Nielsen, lilu@lundbeck.com, H. Lundbeck A/S, Ottiliavej 9, 2500 Valby, Denmark

Keywords: NIR, transmittance, pharmaceutical, PLS, quantitative

The introduction of Good Manufacturing Practice (GMP) in the pharmaceutical industry has raised a quest for rapid altervative methods to perform at-line/on-line quality control. Spectroscopic techniques in combination with chemometrics have potential for replacing many existing chemical methods in the pharmaceutical industry (see e.g. Jedvert et al. 1998). In this study, Partial Least Squares Regression (PLSR) models were investigated based on Near Infrared (NIR) transmittance spectra of a pharmaceutical tablet for the prediction of active substance content.

The tablet under investigation was Escitalopram (R) from the pharmaceutical company H. Lundbeck A/S, a product that is produced in four different dosages (5, 10, 15 and 20 mg active substance per tablet). The sample set consisted of full scale and pilot scale batches as well as laboratory made batches covering the range 85 - 115 % of the nominal content for each dosage form.

Models made with all four dosages together resulted in a prediction error expressed as Root Mean Squared Error of Cross Validation (RMSECV) of 0.33 % w/w - the active substance content in the tablets being between 4.8 and 9.1 % w/w. For comparison, models were also made using NIR reflectance and Raman spectroscopy (see Dyrby et al. 2001) but both these methods gave markedly higher prediction errors - 1.5 to 2 times the error using NIR transmittance. Separate models for each dosage gave prediction errors between 0.22 % w/w and 0.31 % w/w, corresponding to relative prediction errors (RMSECV/y_nom) of 3.3 - 4.0 %. This error is relatively low when compared to the error of the chromatographic reference method, which was estimated to 3.5 %.

References:
1. Jedvert, I., Josefson, M. and Langkilde, F. (1998). Quantification of an active substance in a tablet by NIR and Raman spectroscopy. Journal of Near Infrared Spectroscopy 6, 279-289.
2. Dyrby, M., Engelsen, S. B., Nørgaard, L., Bruhn, M. and Lundsberg-Nielsen, L. (2001). Chemometric quantitation of the active substance (containing CN) in a pharmaceutical tablet using Near Infrared (NIR) transmittance and NIR FT-Raman spectroscopy. Applied Spectroscopy, Submitted.

Multivariate image regression (MIR) & image regression validation

Kim Esbensen, kes@aue.auc.dk, Ålborg Universitet Esbjerg. Norgesgade 31, 1.th , DK-6700 , Denmark
Thorbjørn T. Lied, Telemark University College, Department of Technology, Porsgrunn, Norway and Applied Chemometrics Research Group (ACRG)

Keywords: MIA, MIR

We have recently developed a comprehensive, stand-alone software system for Image-PLS regression [1], which includes facilities for prediction cross-validation [4].

This completes our efforts of developing a complete multivariate image analysis analogue to the multivariate calibration concept for two-way matrices. We briefly refer to the distinctions between (OOV) and (VVO) 3-way decomposition.

We illustrate with several real-world industrial and laboratory application studies [2-6].

References:
[1] "Multivariate image regression (MIR) for quantitative predictions - Prototype software implementation and selected industrial-technological pilot studies". Ph.D. thesis, Telemark University College & NTNU, Feb. 09, 2001.
[2] Lied, T.T., Geladi, P. & Esbensen, K.H. (2000): Multivariate image regression (MIR): implementation of image PLSR - first forays. Jour. Chemometrics, 14, 585-598.
[3] Lied, T.T. & Esbensen, K.H. (2001): Principles of MIR, Multivariate Image Regression - I: regression typology and representative application studies. Jour. Chemometrics "Special Issue: PLS 2000" (in print).
[4] Lied, T.T. & Esbensen, K.H. (2001): Principles of MIR, Multivariate Image Regression - II: Cross validaton - what you see is what you get. Jour. Chemometrics (submitted).
[5] Esbensen, K.H., Lied, T.T., Lowell, K. & Edwards, G. (2001): Principles of Multivariate Image Analysis (MIA) in remote sensing, technology and industry. International Journal of Remote Sensing (submitted).
[6] Lied, T.T., Matveyev, I.H., Karlsrud, D. Huang, J. & Esbensen, K.H. (2002): Image-analytical quantitative monitoring of heterogenous mixture processes: Angle Measure Technique (AMT) vs. Multivariate Image Regreession (MIR). Chemometrics and Intel. Lab. Systems (submitted).

Multivariate methods in the development of a new tablet formulation

Jon Gabrielsson, jon.gabrielsson@chem.umu.se, Umeå University. Organic chemistry, Umeå University, 901 87 Umeå, Sweden, http://www.chem.umu.se
Nils-Olof Lindberg, Pharmacia AB, Consumer Healthcare, Box 941, SE-251 09, Helsingborg, Sweden
Magnus Pålsson, Pharmacia AB, Consumer Healthcare, Box 941, SE-251 09, Helsingborg, Sweden
Torbjörn Lundstedt, torbjorn.lundstedt@melacure.com, Melacure Therapeutics. Ulleråkersvägen 38, SE 756 43, Sweden

Keywords: FT-IR, NIR, DOE

A new tablet formulation was developed. As a part of the objective a vast number of potential excipients for this new formulation were screened. The screening of the excipients proceeded with the aid of multivariate characterization and multivariate design.

Approximately 100 samples of excipients, 21 different excipients of varying quality types and from different producers, were characterized by FT-IR and NIR spectroscopy. The combined spectra form the basis for the multivariate characterization, which is an integral part of the multivariate design.
The excipients are divided into different classes according to their potential use, diluents, binders or disintegrants. Batches of active substance from two manufacturers were also characterized.

The FT-IR and NIR spectra for the excipients in the different classes were SNV pretreated separately in Simca-P 8.0 (Umetrics AB, Umeå, Sweden) and then combined to form the basis for the multivariate characterization. PCA was performed and the different classes were described by three principal components, except for the active substance. The 6 batches of active substance were described by one principal component.

The multivariate design includes ten factors that describe the excipients and the active substance. The other 4 factors are part of or related to a mixture design with filler. The mixture design consists of only three constituents, two excipients - buffer and disintegrant - and filler. Since the ratio between them is more interesting than the actual amount a separate factor describes the ratio of binder and diluent in the filler in the design. One factor describes the type of buffer.

35 experiments were used for the screening of the excipients. Excipients were chosen according to the original design from Modde 5.0 (Umetrics AB, Umeå, Sweden) to form the multivariate design. The design is of resolution IV, which means main effects are clear of two-factor interactions.

The results enabled the identification of excipients that gives the formulation suitable qualities. The validation experiments also show that the multivariate designs yield rather crude models and that a lot of work still remains for the multivariate characterization to be fully reliable.

Near Infra-Red spectroscopy for brain studies? An early attempt at monitoring responses in the human orbito-frontal cortex to smell stimuli, by the use of multi-channel multi-wavelength diffuse NIR spectroscopy

T. Hansen, The Royal Vet. and Agric. University, Dept. of Dairy and Food Science, DK-1958 Frederiksberg C, Denmark
Harald Martens, Harald.Martens@mail.tele.dk, DTU/NTNU/KVL . Teglgaardstr 12A , DK-1452 , Denmark
P. Møller, The Royal Vet. and Agric. University, Dept. of Dairy and Food Science, DK-1958 Frederiksberg C, Denmark
Magni Martens, mma@kvl.dk, The Royal Veterinary and Agricultural University, Chemometrics Group. The Royal Veterinary and Agricultural University , 1958, Denmark

Keywords: NIR, smell, brain, orbito-frontal cortex

The study of the human brain and its response to external and internal stimuli is a hot topic in contemporary research. In the sensory science group at KVL, we are interested in the physiological and psychological basis for olfaction. A variety of methods are available for studying brain responses. Functional Magnetic Resonance Imaging (fMRI) is a very useful method for studying patterns of variation in the brain's haemodynamic (blood flow and blood oxygenation) response to e.g. smell stimuli.

However, fMRI is expensive and cumbersome. So there is a need for faster, cheaper non-invasive methods of monitoring the brain. Diffuse NIR spectroscopy based on specialized equipment has recently been reported to allow relatively low-cost monitoring of brain activity. This poster presents some preliminary experiences with using a dedicated, commercially available fiber-optical NIR instrument (Hamamatsu NIRO 300, with 2 channels, each with 4 wavelength channels x 3 detector positions). One channel was positioned near the orbitofrontal cortex, where strong olfaction response has been found by fMRI. For control, the other sensor was in some cases positioned on top of the chewing muscle, for referencing.

More than 10 different persons were tested for responses to smell stimuli. For almost every person the same pattern of response was seen: An increase in apparent blood volume and a change in the degree of oxygenation of the blood, lasting for several minutes after the smell stimulus had been removed. This pattern of response is similar to the one previously published by others. Multivariate modelling of the experimental results seems to support this, although our conclusions are somewhat uncertain.

However, it is at the present time not clear whether this response pattern is a specific signal associated with local neuronal activity in the cortex. It could also be due to changes in brain blood pressure due to general awareness of the stimuli. It could even be sensitive to changes in extra-cranial muscle and skin. Therefore, more research is needed before a final verdict on this promising methodology can be given. But the complexity of the signal indicates that the multi-wavelength/multi-detector NIR instrumentation yields informative and interesting physiological data, irrespective of its physiological origin.

Non-linear partial least squares (the error based non-parametric PLS algorithm)

Per Anker Hassel, p.a.hassel@ncl.ac.uk, University of Newcastle. Centre for Process Analytics and Control Technology, Newcastle upon Tyne, United Kingdom, http://ncl.ac.uk/
Baibing Li, Centre for Process Analytics and Control Technology, University of Newcastle, Newcastle upon Tyne, NE1 7RU, United Kingdom
Elaine Martin, e.b.martn@ncl.ac.uk, University of Newcastle. Merz Court, NE1 7RU , United Kingdom
Julian Morris, julian.morris@ncl.ac.uk, University of Newcastle. Centre for Process Analytics and Control Technology (CPACT), NE1 7RU , United Kingdom

Keywords: non-parametric regression, PLS algorithm, non-linear error based PLS

An error based non-parametric partial least squares (PLS) algorithm is presented where the mapping between the scores of the corresponding latent variables of the **X** and **Y** data matrices is fitted using kernel regression. The kernel regression family forming the basis of the analysis is that of "local polynomial kernel estimators". Within this work, the local linear kernel estimator was investigated due to its simplicity, mathematical properties and widespread appeal. Linearity is controlled through the bandwidth, i.e. the smoothing parameter. If the smoothing parameter is very large, the solution approaches that of linear least squares. In contrast as the bandwidth becomes very small, overfitting will result. The bandwidth can be determined either directly by calculating the optimal value, as determined by large sample size assumptions, or by cross-validation.

The motivation for this approach is that the accuracy of the inner fit is comparable to that achieved by a feed forward neural network. However since the fitting of the kernel regression model is not iterative, computational times are considerably reduced. In addition the smoothing factor can be estimated automatically. For the neural network approach, the number of nodes, corresponding to the degree of smoothing, needs to be estimated using cross validation. This results in a further reduction in computational costs compared to neural networks.

The method is illustrated using a number of data sets that exhibit different properties including linear and non-linear behaviour. In particular the methodology will be compared with other non-linear PLS algorithms on an industrial data set.

Development of a software sensor for estimation of phosphorus in municipal wastewater

Åsa Jansson, asa.jansson@ivl.se, IVL - Swedish Environmental Research institute. Box 210 60, 100 31, Sweden, http://www.ivl.se
Jonas Röttorp, jonas.rottop@ivl.se, IVL - Swedish Environmental Research institute. Box 210 60 , Stockholm, Sweden

Keywords: software sensor, PLS, FIR model, wastewater treatment, phosphorus

Chemical precipitation of phosphorus is a method used world wide in wastewater treatment. Today most treatment plants lack elaborate control of the addition of precipitation chemicals. Common control strategies are to use a chemical dose proportional to the water-flow through the plant or to simply vary the dose between a few pre-set levels according to a time-schedule. A better approach is to control the dose of precipitation chemicals based on the phosphorus content in the wastewater. Even though on-line instruments for phosphorus exist on the market, an appealing idea is to create a software sensor from instruments already at use in the treatment plants to estimate the phosphorus concentration.

As a part of a project financed by the European commission (IST 11990), an initial investigation concerning the development of a software sensor for phosphorus has been carried out at Borlaenge WWTP, Sweden. The parameters considered as inputs were water-flow, conductivity, pH, chemical oxygen demand (COD) and suspended solids (SS). Both total and soluble phosphorus were modelled as outputs.

Three different modelling techniques were used in the development of the software sensor: multiple linear regression, MLR, principal component regression, PCR, and projection to latent structures, PLS. Static models were compared with finite impulse response, FIR, models. The best result was obtained using a FIR model created with PLS. The model for the soluble phosphorus was based on all input variables.
The results showed that the phosphorus content in wastewater could be described using a software sensor. All parameters in the model improved the estimations of phosphorus. However, more data for external validation is needed to give information about the applicability for using the software sensor for this approach. Using a software sensor could be a powerful approach to control dosage of precipitation chemicals and phosphorus removal in wastewater treatment plants.

Evaluation of three methods for assessor and descriptor analysis

Ina Jensen, inaj@dsr.kvl.dk, Sensory Science Group, Department of Dairy and Food Science, Royal Veterinary and Agricultural University, Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark
Michael Bom Frøst, Sensory Science Group, Department of Dairy and Food Science, Royal Veterinary and Agricultural University, Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark
Garmt Dijksterhuis, Sensory Science Group, Department of Dairy and Food Science, Royal Veterinary and Agricultural University, Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark
Per Bruun Brockhoff, Department of Mathematics and Physics, Royal Veterinary and Agricultural University, Thorvaldsensvej 40, DK-1871 Frederiksberg C, Denmark
Magni Martens, mma@kvl.dk, The Royal Veterinary and Agricultural University, Chemometrics Group. The Royal Veterinary and Agricultural University , 1958, Denmark

Keywords: assessor differences, performance, reliability, descriptors

Information about the performance of individual assessors in a panel during descriptive analysis is important for panel monitoring. The different use of scale, use of descriptors and the 'signal-to-noise ratio' are parameters that are interesting in the description of a sensory panel. A plethora of different methods have been developed for this purpose. In this work three methods are compared: The Panmodel, an Anova-based model [1], the Manual for Sensory Data Analysis, a PLS-based model [2] and Generalised Procrustes Analysis (GPA)[3].

The purpose of this work was to examine the different methods to evaluate the assessors, both with regard to the type of information that the methods give, and scrutinise differences in conclusions about assessor performance. The data set analysed was obtained from a standard sensory descriptive analysis of 9 cream cheeses.

Regarding the reliability of the assessors in general, the same conclusions are given by the three methods. However, the rank order of "best" to "worst" assessor for a given descriptor was not always the same. In these cases there was the highest agreement between the Panmodel and the PLS-based model. This was also the case when analysing for differences in use of scales. The analysis of which descriptors discriminate between products shows a very high agreement between the PLS-based model and GPA.

The three different methods do not give exactly the same type of information regarding assessor performance and descriptors. However, they all provide measures of panellist reliability that can be used as an overall measure of their performance. The analysis of the three different methods reveals some minor differences in results in comparable measures. However, the overall conclusions regarding reliability of descriptors are the same.

References:
[1] Brockhoff, P.B Statistical testing of individual differences in sensory profiling. Invited paper for 5th European Conference on Food Industry and Statistics, December 3-5, 1997, Versailles, France
[2] Martens, H., Wedøe, S., Bredie, W. L. P. & Martens, M Manual for Sensory Data analysis in the Unscrambler, Dept. Food Science and Technology, Royal Veterinay and Agricultural University, Frederiksberg, Denmark (1999)
[3] Dijksterhuis, G.B., Gower, J.C. (1991/2). The Interpretation of Generalized Procrustes analysis and Allied Methods. Food Quality and Preference, 3, 67-87.

Implementation and validation of on-line models for monitoring of wood-chips properties

Pär Jonsson, par.jonsson@chem.umu.se, Umeå University. Organic chemistry, SE-901 87, Sweden, http://www.chem.umu.se
Henrik Antti, h.antti@ic.ac.uk, Imperial College. London, United Kingdom
Lars Wallbäcks, lars.wallbacks@telia.com. Barrgränd 15 , SE-944 71 Piteå , Sweden
Michael Sjöström, michael.sjostrom@chem.umu.se, Umeå University. Umeå University , SE-901 87 , Sweden

Keywords: NIR spectroscopy, PLS, on-line, validation

The pulpwood chips entering a pulp mill have been characterised using an on-line NIR instrument (NIRSystems 6500). Collected NIR spectra were used in multivariate calibration models to predict the moisture and the between and the within variation of species. Statistical experimental design was used to form a calibrationset that includes most of all variation occurring in real situations. NIR spectra for all designed samples were measured at-line and the estimated calibration models were used on-line.

Three PLS models were used one for predicting moisture, expressed as dry-content, the second one for predicting the amount of pine-chips, measured in percent of total amount of chips, and the third one for predicting the amount of sawmill-chips. On-line NIR spectra were measured once every minute. The time-series of predictions were noisy; to reduce that problem a moving average was applied. This gave a smoother time-series of predictions, more valid for the process.

To validate the quality of the predictions, wood-chips from the conveyer belt were sampled. These wood-chips were analysed in laboratory. The analysis of moisture is easy, but the other y-variables are more difficult to validate. To solve that problem the samples from the conveyer belt were ground to a powder and dried. For each sample a NIR spectrum was measured. A calibration model was done from a mixture design so that the amount of pine and the amount of sawmill can be predicted.

The smoothed on-line predictions were compared with the laboratory measurements. This validation approach works well for the "pine" model but not for the "sawmill" model. The validation proved that the predictions on-line were as good as at-line for moisture and amount of pine.

Frequency characterization of the chemical periodicity - The problem of assessing missing values

Nikolay N. Khramov, V.G. Khlopin Radium Institute, 2nd Murinski 28, 194021 St. Petersburg, Russia
S. A. Bartenev, V.G. Khlopin Radium Institute, 2nd Murinski 28, 194021 St. Petersburg, Russia
G. S. Markov, V.G. Khlopin Radium Institute, 2nd Murinski 28, 194021 St. Petersburg, Russia
V. N. Romanovskii, V.G. Khlopin Radium Institute, 2nd Murinski 28, 194021 St. Petersburg, Russia
Kim Esbensen, kes@aue.auc.dk, Ålborg Universitet Esbjerg. Norgesgade 31, 1.th , DK-6700 , Denmark

Keywords: missing data, periodic system, PCA, FA

The properties of chemical elements depend periodically on element group order sequence or charge of nucleus. The character of chemical periodicity has been studied thoroughly ever since the Periodic System was established. Chemical periodicity is characterised by the pertinent chemical group length a.o.

Generel periodical processes are characterized by the length of periodic components as well as their specific frequencies. Nevertheless frequency characterization of the Periodic System chemical periodicity is not widely in use.

To derive a possibly new didactic view on chemical periodicity we have modeled various physical and chemical propertjes of the elements as a function of atomic number by a sum of sinusoidal functions. Such a model facilitate comparison of the contributing sine functions frequencies (relative "spectral densities"). Comparison is made with spectral density functions, which are well known e.g. from time series analysis. The spectral density of orbital radius in the frequency domain depicts completely the overall structure of the Periodic Table, i.e. components having periods 8, 18, 32 atomic numbers. But interestingly, this spectral density functionality is SPECIFIC for each property modeled. In total we have studied 26 important element properties. We speculate on reasons why...Not all values of the studied properties are known however. As important examples, the values of the fifth and further ionization potentials for rare earth elements are absent in all available reference books.

We have tried alternative approaches to model replacement of missing data in the first ionization potential and selected those which do not the change spectral densities significantly. Apparently best choices were: replacement by the overall mean, interpolation from adjacent values and predicted values from linear trend regressions. To systematize the presentation of chemical periodicity with frequency characteristics we use bilinear projections (PCA/FA). Differences as well as similarities in the periodicity of the chemical element properties are depicted in a way which lends itself easily to synoptic perception and interpretations.

Correlations between biodegradation rates of alkyl sulphosuccinates and their physicochemical parameters

Gergely Csiktusnásnádi Kiss, Institute of Chemistry, Chemical Research Centre, Hungarian Academy of Sciences, P.O.Box 17, 1525 Budapest, Hungary
Esther Forgács, forgacs@cric.chemres.hu, Institute of Chemistry, Chemical Research Centre, Hungarian Academy of Sciences, P.O.Box 17, 1525 Budapest, Hungary
Alena Vrbanová, Institute of Microbiology, Slovak Academy of Sciences, Stefanik 3, 81434 Bratislava, Slovakia

Keywords: Sulfosuccinate anionic surfactants, QSAR, PCA, thin-lazer chromatography

The relationship between rates of primary degradation of a seies of alkyl sulphosuccinates by bacterium Comamonas terrigena N3H and their structural characteristics was determined by the stepwise regression and principal component analysis. It was found that the electronic constant characterising the inductive effect of substituent and the parameter related to the complex forming capacity of surfactant molecules (determined by reversed phase thin -layer chromatography) had significant impact on the degradation rate. The results indicated that the degree of polarisation of sulpho-group and the ability of alkyl sulphosuccinates to form inclusion complexes may govern their transport to the bacterial cell and the binding to the hydrolytic enzyme catalysing the process of primary degradation.

Estimation of uncertainty of concentration estimates obtained by image analysis

Maaret Korpelainen, maaret.korpelainen@lut.fi, Lappeenranta university of Technology, Lappeenranta, Finland
Satu-Pia Reinikainen, satu-pia.reinikainen@lut.fi, Lappeenranta university of Technology, Department of Chemical Technology, Chemometrics Group. PO Box 20, Lappeenranta, Finland, http://www.lut.fi
Jukka Laukkanen, jukka.laukkanen@vtt.fi, VTT Mineral Processing. Tutkijankatu 1 , 83500, Finland
Pentti Minkkinen, Pentti.Minkkinen@lut.fi, Lappeenranta university of Technology. Lappeenranta University of Technology. , FIN-53851 , Finland

Keywords: image analysis, measurement uncertainty, sampling theory, SEM

Images of scanning electron microscopy (SEM) and optical microscopy are widely used in mineral processing and in metallurgy to estimate the concentration of mineral or metallic species of the materials produced or processed. Even though the method is widely used the uncertainty of the results is seldom determined. The purpose of this study was to optimize the image analysis procedure so that the uncertainty of the measurement can be estimated during the progress of the image analysis so that the analyst can estimate when the required precision of the analysis is achieved.

In SEM both solid and powder samples can be used in image analysis. From solid samples e.g., from drill cores) polished sections are made. Powders are first mixed with polyester resin and after solidification polished sections are prepared for analysis. From the polished sections images are taken and then the images are analyzed, e.g., by using frames of 512 x 512 pixels. The method involves segmenting, identification and evaluation of the total area of the particles of interest in each analyzed frame. Here the polished surface presents is a two-dimensional sample from the original three-dimensional material and each analyzed frame a sample from the two-dimensional surface; hence the sampling theory can be applied to estimate the uncertainty of this procedure.

Pierre Gy's sampling theory was used to identify the different error sources of the determination procedure. Gy's theory and results derived by using binomial distribution were used to estimate the fundamental sampling error. Sample preparation for powder sample was also investigated to overcome the difficulties caused by segregation during the sample preparation procedure. Theory of stratified sampling was applied to optimize the number and position of frames to be analyzed. Simulations, synthetic mixtures, and natural drill core samples were used to validate the theoretical results.

Expert system shell for evaluation of bond dissociation energies of organic compounds

E. Kromkin, tve@icp.ac.ru, Institute of Problems of Chemical Physics, Russian Academy of Sciences
V. E. Tumanov, tve@icp.ac.ru, Institute of Problems of Chemical Physics, Russian Academy of Sciences

Keywords: expert system, bond dissociation energies, organic compounds

The bond dissociation energy (BDE) changes associated with making and breaking bonds between atoms in a molecule are important practical concepts used everywhere in chemistry. Chemists have accumulated a vast amount of information on free radical abstraction reactions. These reactions may be subdivided on groups. It is seem that there is a parameter that is constant for every group. This parameter was used for development statistical model for prediction of BDE on constant rate of free radical reactions [1]. The model was incorporated in computer program.
In this empirical model every elementary abstraction radical reaction may be described by the following variables. 1) the activation energy of reaction Ei; 2) the distance of atom transfer re; 3) BDE of attacked bond Di; 4) forming bond Df; 5) their vibration frequencies ni and nf; 6) the reduced masses of the bonds mi and mf; 7) rate constant k and 8) temperature T. These variables are presented the state space of reactions. The statistical parameter bre = (Ei - 0.5(hLni -RT))1/2 - ni/nf (mi/mf)1/2(Ei - 0.5(hLni -RT)) - Di - Df -0.5hL(ni -nf))1/2 , where L is Avogadro's Number, h is Plank's constant, R is Gas constant. The parameter bre is the measure of similarity of the free radical abstraction reactions on the state space of reactions defined above. bre = const on the groups of structurally isotypical reactions. There are about 110 clusters of free radical abstraction reaction in Knowledge Base of our expert systems.

An expert system consists of the few program components: knowledge base, the training tool, the inference engine and the user interface. The facts of knowledge base contain information about reaction clusters and their bre. The production rules characterize a decision making in selection the path of evaluation of BDE value. There are 30 production rules. The inference engine contains the code required to interpret the fact and rules and to provide the general problem solving techniques. It initiate a consultation, implementing the search procedure and determining that a conclusion has been reached or no. The backward-chaining strategy has been used in engine. The user interface facilitates the dialogue based on the question-answer relations. The expert system was been used for the evaluation of the BDE's in the hydrocarbons successfully. The expert system has been realized with MS VB 6.0 and MS Access97. There are export/import interface with MS Access databases and MS Excel.

References:
1. E.T. Denisov New empirical models of radical abstraction reactions. Russian Chemical Reviews 66 (10) 859-876(1997).

Application of genetic algorithm - PLS to the determination of wine parameters from FTIR spectra

Riccardo Leardi, riclea@dictfa.unige.it, University of Genova, Department of Pharmaceutical and Food Chemistry and Technology, Section of Analytical Chemistry. via Brigata Salerno (ponte), I 16147, Italy, http://www.unige.it
Claus-Dieter Patz, cdpatz@web.de, Research Institute of Geisenheim, Department for Wine Analysis and Beverage Research - Rüdesheimer Str. 28, D-65366 Geisenheim, Germany
Anika Blieke, Research Institute of Geisenheim, Department for Wine Analysis and Beverage Research - Rüdesheimer Str. 28, D-65366 Geisenheim, Germany

Keywords: genetic algorithms, wavelength selection, PLS, FTIR spectroscopy, wine

With the application of multivariate calibration to FTIR spectra of wine samples it is possible to determine up to 12 components, such as alcohol, extract, tartaric, lactic, malic, total and volatile acid, pH, reducing sugars, fructose, glucose and glycerol in one minute[1].

In the quality control of winemaking, during the fermentation process and in routine analyis this instrument could therefore replace the "classical" chemical analysis. The great advantages of this method over other conventional methods (e.g. HPLC, photometric methods), is the saving in terms of analysis time, human effort and its low ecological impact, because of the very few chemicals needed for determination.

It is commonly accepted that a PLS calibration model should not include the whole spectrum, and that a feature selection is highly beneficial. The application of Genetic Algorithm - PLS[2] to these data sets lead to quite interesting results, since a good predictive ability has been obtained for each response. Furthermore, since the selected variables almost always correspond to well defined spectral regions, the results can be useful also for spectral interpretation.

Some examples will be shown, and the results will be compared with those of a commercial in-built spectroscopic software. These results have been obtained on a "global" data set made by 143 international wine samples (100 samples in the training set, 43 samples in the evaluation set). Of course better results could be obtained if more "specific" models (one for each type of wine) would be built. Owing to the great variety of wines, this would anyway require a huge sampling effort. In future, it is likely that the same approach will also be applied to fruit beverages.

References:
[1]Patz CD, David A, Thente K, Kürbel P, Dietrich H; "Wine Analysis with FTIR Spectrometry", Vinicultural and Enological Science, 54, 80-87 (1999).
[2]Leardi R; "Application of genetic algorithm - PLS for feature selection in spectral data sets"; Journal of Chemometrics; 14, 643-655 (2000)

Improved discrimination of sea ice types: AMT and MIR applied to satellite images from ERS SAR

Maria Lundhaug, maria.lundhaug@nrsc.no, Nansen Environmental and Remote Sensing Center, Edvard Griegsvei 3A, N-5059 Bergen, Norway and Telemark University College (HiT), Department of Technology, Box 203, N-3901 Porsgrunn, Norway
Kim Esbensen, kes@aue.auc.dk, Ålborg Universitet Esbjerg. Norgesgade 31, 1.th , DK-6700 , Denmark

Keywords: AMT, MIR, artic sea, ice

Arctic sea ice plays an important role in the global climate system, as well as for shipping and offshore operations in ice covered waters. Monitoring of sea ice in the Pechora and Kara Sea region is very important both for economic and environmental reasons, as extensive oil and gas resources are found in these areas. Cargo transport along this part of the Northern Sea Route may cause environmental problems, which can be prevented by improved knowledge of the ice conditions. Earth Resources Satellite 1 and 2 (ERS-1/2) Synthetic aperture radar (SAR) is a valuable tool for monitoring of difficult sailing areas due its independence of cloud cover and sunlight.

For characterization of sea ice by the single channel ERS SAR sensor, the mean and standard deviation of the backscattering values are used in combination with meteorological data, latitude/longitude, and angle of incidence. Separation of water and ice, as well as separation among the different ice classes is difficult due to overlap in the distributions of the different classes.

There are many interesting parallels between industrial chemometric image analysis and satellite image analysis. The Angle Measure Technique (AMT) is applied to selected sea ice image samples in order to assess the possibilities of morphological characterization to improve sea ice separation. This method creates, in combination with the original variables an alternative multivariate data set, which we study in detail by Partial Least Square Regression (PLS).

Sensory analysis of MRI pictures: Using human perception and cognition to segment and assess the interior of potatoes

Harald Martens, Harald.Martens@mail.tele.dk, DTU/NTNU/KVL . Teglgaardstr 12A , DK-1452 , Denmark
A. K. Thybo, Danish Institute of Agricultural Sciences, Dept. of Horticulture, DK- 5792 Aarslev, Denmark
H. J. Andersen, Danish Institute of Agricultural Sciences, Dept of Animal Product Quality, DK-8830 Tjele, Denmark
A. H. Karlsson, Danish Institute of Agricultural Sciences, Dept of Animal Product Quality, DK-8830 Tjele, Denmark
S. Dønstrup, Aarhus University Hospital, MR Research Centre, DK-8200 Aarhus N, Denmark
H. Stødkilde-Jørgensen, Aarhus University Hospital, MR Research Centre, DK-8200 Aarhus N, Denmark
Magni Martens, mma@kvl.dk, The Royal Veterinary and Agricultural University, Chemometrics Group. The Royal Veterinary and Agricultural University , 1958, Denmark

Keywords: image analysis, sensory, segmentation, MRI, potatoes

Magnetic Resonance Imaging (MRI) is a non-destructive method measuring the abundance of free and bound water. This methods is useful in determination of texture characteristics in animal and vegetable products. The aim of this study was to investigate the MRI pictures of the interior of potato tubers by descriptive sensory image analysis and by conventional computer image analysis.

Materials and Methods: Two potato varieties Sava and Berber were sampled in November 1999 and in May 2000. For each of these 4 conditions, 15 tubers were analysed. Each of the 60 tubers was scanned by MRI (Sisco 300/183, Varian Inc. Palo Alto, USA) in order to get a picture of its intact interior.The MRI pictures were analysed by computer image analysis MaZda (ver. 2.21, Politechnika Lodzka, Lodz, Poland) in terms of means, variance, skrewness, curtosis and percentiles of greytone histograms. Sensory image analysis of the MRI same pictures was done by nine trained assessors using 16 sensory visual attributes. Data were analysed by cross-validated Partial Least Squares Regression.

Results: Potato variety Berber is associated with fig-like interior and Sava with passionfruit-like interior. Further separation in e.g. White Centre, White Peel and Number of White Grain in Berber versus e.g. Sharp Rest, Size of White Grains, Channels Distinct and Whiteness of Grains in Sava. Effect of harvest times: (Fig. 2, PC2): November 1999 storage is correlated with high levels of Image Background Blackness, Tuber Roundness and two internal potato descriptors, and inversely for May 2000 storage. The sensory and the computer image analysis of the MRI pictures both discriminated well between varieties and storage times. Compared to the sensory image analysis, the computer image analysis was more sensitive to MRI measurement artifacts.

These results indicate that descriptive sensory analysis is an alternative to conventional computerised image analysis for e.g. MRI picture sets. The sensory analysis may offer an advantage, due to people's ability to interpret the images into meaningful segments, and to describe each of these segments qualitatively and quantitatively.

Sampling noise and measurement noise: Two sources of uncertainty in the assessment of food quality

Harald Martens, Harald.Martens@mail.tele.dk, DTU/NTNU/KVL . Teglgaardstr 12A , DK-1452 , Denmark
Anette Thybo, Danish Institute of Agricultural Sciences, Dept. of Horticulture, DK- 5792 Aarslev,Denmark
Rasmus Bro, rasmus@optimax.dk, The Royal Veterinary & Agricultural University, Dept. of Dairy and Food Science, Food Technology. Rolighedsvej 30, 1958, Frederiksberg, Denmark, http://www.models.kvl.dk/users/rasmus/
Magni Martens, mma@kvl.dk, The Royal Veterinary and Agricultural University, Chemometrics Group. The Royal Veterinary and Agricultural University , 1958, Denmark

Keywords: uncertainty, variance, power, estimation, Monte Carlo

In order to optimise experimental designs with respect to cost/benefit, it is important to have some information about the relative contributions of the two main sources of uncertainty - biological sampling errors and instrumental measurement errors. The present poster gives an example of how to estimate various components of variance, for a series of rheological texture measurements in potatoes Thybo & Martens (1999).

The effects in a certain designed experiment are estimated and assessed w.r.t. significance by jack-knifed "Anova-like PLS Regression" (APLSR) (Martens & Martens 2001).

From the same experimental data, the two uncertainty-variance components, 1) due to measurement noise and 2) due to sampling noise are then estimated.

These two uncertainty estimates are finally used in order to optimize the plan for a future experiment w.r.t. statistical power, based on Monte Carlo simulation (Martens et al 2000), in order to answer the question: How much biological sampling replicates and how many measurement replicates are needed in order to reveal a certain effect at a certain significance level?

References:
1. Box. G., Hunter, S. and Hunter. J. (1978) Statistics for experimenters. J. Wiley & Sons Inc.
2. Martens, H. and Martens, M. (2001) Multivariate Analysis of Quality. J.Wiley & Sons Ltd. 420 pages.
3. Martens, H., Byrne, D. V. and Dijksterhus, G. (2000) Power of experimental designs, estimated by Monte Carlo simulation. J.Chemometrics, 14, 441-462.
4. Thybo, A.K. and Martens, M. (1999) Instrumental and sensory characterization of potato texture. J. Texture Studies, 30, 259-278.

Pre-processing of input data for simplified GLS modelling, as applied to PLSR

Harald Martens, Harald.Martens@mail.tele.dk, DTU/NTNU/KVL . Teglgaardstr 12A , DK-1452 , Denmark
Frank Westad, frank.westad@matforsk.no, MATFORSK. Arildsvingen 12, Oslo, Norway, http://www.matforsk.no
Barry M. Wise, bmw@eigenvector.com, Eigenvector Research, Inc.. 830 Wapato Lake Road, Manson, WA 98831, USA, http://www.eigenvector.com/
Rasmus Bro, rasmus@optimax.dk, The Royal Veterinary & Agricultural University, Dept. of Dairy and Food Science, Food Technology. Rolighedsvej 30, 1958, Frederiksberg, Denmark, http://www.models.kvl.dk/users/rasmus/
Per Brockhoff, The Royal Vet. and Agric.l University, Dept. of Dairy and Food Science, DK-1958 Frederiksberg C, Denmark

Keywords: preprocessing, generalized least squares, rank reduction, bi-linear modelling, covariance

The success of applied multivariate data analysis, especially in complex materials like food, depends on how easy it is for the user to gain visual overview of the estimated model parameters, e.g. bilinear principal components (PCs). When the number of valid and important PCs turns out to be >2, then the graphical inspection of the modellinga results becomes confusing. Moreover, if the subject matter is complex and needs a model with many PCs to be estimated, this requires lots of good input data in order for the model to become statistically stable. That may be expensive.

This poster outlines a general method for simplifying multivariate modeling, applicable e.g. for the assessment of food quality and for improved calibration of e.g. NIR and NMR multi-channel instruments. It reduces the effect of known but undesired structures in the input data tables, already at the pre-processing stage. The method consists of multiplying the input data matrices by the square root of the a priori approximately known covariance matrix of errors or irrelevant structures. The method makes Generalized Least Squares (GLS) estimators out of Ordinary Least Squares (OLS) or Weighted Least Squares (WLS) estimators, without having to change the OLS/WLS algorithms.

The effect of the method is sometimes similar to that of other pre-processing methods such as OSC and Direct Orthogonalization. But in contrast to these, the GLS preprocessing reduces the effect of unwanted phenomena by shrinking, rather than by subtraction.

The method is illustrated here for the simplification of multivariate calibration of a spectrophotometer by a stabilized cross-validated Partial Least Squares Regression, PLSR(X,Y). But the method is equally applicable as pre-processing for other methods, e.g. Principal Component Analysis, PCA(X).

References:
1. Martens, H. and Martens, M. (2001) Multivariate Analysis of Quality. An Introduction. J.Wiley & Sons Ltd.

Applied electrical DC-potential for flow improvement in power plant turbine inlet pipe-lines chemometric intercalibration between acoustics and PIV-laser velocimetry

Inger Hedvig Matveyev, inger.h.matveyev@hit.no, EMT. Bergsbygda , N-3914 , Norway
Magne Waskaas, EMT, Electromagnetic Research, Porsgrunn & Applied Chemometric Research Group, Telemark University College, Kjølnes Ring 56, N3914 Porsgrunn, Norway
Kim Esbensen, kes@aue.auc.dk, Ålborg Universitet Esbjerg. Norgesgade 31, 1.th , DK-6700 , Denmark

Keywords: DC-potential, flow-friction, reduction, acoustic, chemometrics

Multivariate calibration methods are used to study means for flow improvement in water flow in a stainless steel pipe-line. The objective is to study the effect of an applied electrical DC-potential to the pipeline on the flow velocity profile, which is measured by PIV laser over a cross-section in a plexiglas section of the pipe-line. In addition acoustic chemometric measurements are made to monitor velocity changes in the near wall parts of the same section.

Results (1999-2000): Initial flow improvement studies (1999) were carried out in a full factorial regimen with turbulent flow (Rn: 50,000) under four different mean flow rates (1,2,3,and 4 l/s) and five temperatures (15,18, 21, 24 and 27 (C). During each experiment - with and without exposure to the DC- potential, simultaneous flow velocity profiles and acoustic measurements are recorded. Results show a significant increase in the velocity profile when a very particular electric potential is introduced only; applied potentials below and above this theoretical target potential showed no effect. This result proves the principle and the feasibility of the reduced wall-friction by application of this type of electrical DC-potential. We are able to establish a 42 Y-variables PLS2 inter-calibration, allowing the near-wall velocity profiles, which can ONLY be observed in transparent sections of pipe-lines, to be predicted directly from the simple acoustic chemometric sensor which is ultimately designed to be applied on the outside of the pipe-line wall.

In accordance with the underlying hydromagnetic electrochemical theory it would appear that there is a so-called window-effect for the applied potential to achieve real flow improvement. Optimal application and stability of the applied DC-field remains the greatest challenge for moving towards industry-level implementations.

We now also have field evidence from an existing power plant in Norway where a full-scale test installation has been operative for one year. We document friction-loss reductions up to 4.8 m (in relation to some 40 m without our DC-effect), which is of significant economic importance for the hydro-power industry in Norway and elsewhere.

Relationship between the conditions of fermentation and the laccase production of four strains of Lentinus edodes. Comparison of principal component analysis and spectral mapping technique

Helena Morais, Instituto Nacional de Investigacao Agraria, Oeiras, Portugal
A. C. Ramos, Instituto Nacional de Investigacao Agraria, Oeiras, Portugal
Tibor Cserháti, Institute of Chemistry, Chemical Research Centre, Hungarian, Academy of Sciences, P.O.Box 17, 1525, Budapest, Hungary
Esther Forgács, forgacs@cric.chemres.hu, Institute of Chemistry, Chemical Research Centre, Hungarian, Academy of Sciences, P.O.Box 17, 1525, Budapest, Hungary

Keywords: two dimensional PCA, three dimensional PCA

The effect of cation type and concentration and fermentation time on the lacase production of four strains of Lentinus edodes has been determined using 28 different fermentation media and 60 days of fermentatiton time. Principal component analyis (PCA) was employed for the assessment of similarities and dissimilarities between the laccase activities of the samples taken from each culture at each ten days. As PCA is not suitable for the separation of the strengh (potency) and selectivity of the effects they were separated by the spectral mapping technique (SPM). The di mensionality of the matrices of PC loadings and variables and the selectivity maps were reduced to two by the nonlinear mapping technique. The results of PCA and SPM were compared by calculating linear relationships among the potency values and the corresponding coordinates of PCA and SPM maps.

It was established that neither the charge of the cations nor their concentration in the fermentation media exert a marked effect on the laccase production. It was found that the type of strains of Lentinus edodes and the fermentation time exerts a considerable impact on both the strength and selectivity of enzymatic activity. The results of PCA and SPM were considerably different, therefore, their simulataneous application in QSAR studies is highly recommended.

A quantitative assay of intact tablets by transmittance near-infrared spectroscopy

Kent Tram Møller, kmm@nycomed.com, Nycomed Pharma, QC Laboratories, NIR Department, Langebjerg 1, DK-4000 Roskilde, Denmark,
Thomas Würtz Jensen, twj@nycomed.com, Nycomed Pharma, Pharmaceutical Development, Analytical Development. Langebjerg 1, DK-4000, Denmark, http://nycomed.dk
Magnus Tolleshaug, mwt@nycomed.com, Nycomed Pharma, Pharmaceutical Development, Analytical Development. Langebjerg 1, 4000, Denmark

Keywords: NIR, tablets, pharmaceutical, chemometrics, validation

Near infrared spectroscopy is ideal suited for the pharmaceutical quality control analysis of intact tablets. The analysis is fast and easy to use because no sample preparation is required and the method is non-destructive. In order to reduce the cost for the release of finished products near infrared (NIR) transmittance spectroscopy were used to determine the content of the active ingredient in intact tablets.

A problem with intact tablet assay is that normal production batches do not encompass a sufficiently wide range for setting up a reliable calibration equation. Therefore tablets that are both under- and overdosed with respect to the active ingredient need to be produced in lab scale, without altering other factors that may affect the physical and pharmaceutical properties.

A quantitative NIR calibration model has been developed on a MB 160 FT/NIR spectrometer from ABB Bomen inc. equipped with a Tablet SamplIR autosampler with an InGaAs detector. The calibration model is based on eight lab scale batches with nominal value of 80 - 105% m/m active and ten production batches containing 98-100% m/m active. NIR absorbans spectres are recorded from 4.000-12.000 cm-1 on 10 tablets per batch. The spectres are pre-treatment with Savitzky-Golay first derivative and an absolute normalisation followed by a partial least-square regression, (PLS).

The quantitative NIR model has been validated and shows good accuracy compared to the reference assay method. Twenty batches, not included in the calibration model, were predicted by the NIR method and compared with the UV spectroscopic reference method. This gave a standard error of prediction of 0,4 %.

The NIR method shows good repeatability, the relative standard deviation is 0,1 % based on six NIR assays on the same 10 tablets and on the same day.
The intermediate precision is 0,1 % based on six NIR assays on the same 10 tablets analysed over two days and by two technicians.

The validation shows that NIR can be used for prediction of the active in intact paracetamol tablets. By using the NIR method the analysis time is reduced from ½ day to 15 minutes. The Norwegian regulatory authorities have approved the NIR method in 2001.

Simultaneous determination of water constituent concentrations and partial least squares

Allan Aasbjerg Nielsen, aa@imm.dtu.dk, Technical University of Denmark, Informatics and Mathematical Modelling, Image analysis. IMM/DTU building 321, DK-2800 Kgs. Lyngby, Denmark, http://www.imm.dtu.dk/~aa

Keywords: PLS, canonical covariance, water constituents

This paper describes the application of partial least squares (PLS) methodology to determine the weights in a physically based weighted least squares regression model for simultaneous determination of concentrations of the constituents chlorophyll (CHL), total suspended matter (TSM) and coloured dissolved organic matter (CDOM) in Danish coastal waters. CDOM is also known as yellow substance or *gelbstof*. The regression model is based on the relation between specific inherent optical properties of the water and the constituents determined by *in-situ* water sampling and measured reflectance spectra in the visible region. The simultaneous determination of all three constituent concentrations constitutes a marked improvement over previously used *ad-hoc* methods for marginal determination of the concentration of one constituent at a time. As a theoretical contribution the connections between PLS methodology and canonical correlations and covariance analyses are described.

Spectral transformation and range-selection in multivariate calibration

Henrik Aalbrog Nielsen, han@imm.dtu.dk, Technical University of Denmark, Informatics and Mathetical Modelling, Statistics. Richard Petersens Pl., Bygn. 321, 2800 Lyngby, Denmark, http://www.imm.dtu.dk
Michael Rasmussen, Technical University of Denmark, Informatics and Mathetical Modelling, Statistics, Denmark
Henrik Madsen, hm@imm.dtu.dk, Technical University of Denmark, Informatics and Mathetical Modelling, Statistics, Denmark

Keywords: basis functions, spectral measurements, regularization

We consider multivariate calibration where the goal is to predict a quantity characterizing a sample based on a spectrum measured on the sample. To achieve this the method is calibrated on samples where reliable measurements of the characteristic quantity are available. Let y_i denote the characteristic quantity for sample number $i=1,...,N$ and let $a_i(w_j)$ be the corresponding spectrum measured at wavelengths $w_j; j=1,...,m$. We consider the case where m is larger than N. The standard approach to this problem is to express y_i as a linear combination of $a_i(w_j); j=1,...,m$ plus an error term e_i. To solve this singular problem partial least squares (PLS) is often used.

Often the spectra are measured at so many wavelengths that the measurements can be considered continuous over wavelengths. A consequence of this is that very little further information is gained by increasing m. We believe that the underlying model of the method for predicting y should reflect this. It can be achieved by replacing the linear combination of $a_i(w_j); j=1,...,m$ with an integral of $b(w)a_i(w)$ over the range of wavelengths $[w_0 , w_1]$ for which measurements are performed. Here $b(w)$ is a function, acting as a weight on the spectra.

To make the approach feasible we approximate $b(w)$ using a basis function expansion, i.e. a sum over $k=1,...,p$ of $c_kB_k(w)$. Here c_k are unknown coefficients and $B_k(w)$ are the basis functions. This approximation transforms the integral mentioned above into a linear combination of $B_k(w)a_i(w); k=1,...,p$, all integrated over $[w_0 , w_1]$. Given the basis functions, these integrals contain only known quantities and can be evaluated using the trapezoid rule of integration. Consequently, the dimension of the problem is p, which is independent of the number of wavelengths m.

If the number of basis functions p is less than N a unique OLS solution exists, but generally it is advantageous to allow $p > N$ and use methods such as PLS to obtain regularized estimates of the coefficients. The method is applied to some well known datasets and it is shown that the method may outperform the standard method described above quite significantly. Furthermore, we apply LASSO (penalty on the sum of $|c_k|; k=1,...,p$) instead of PLS and show that the resulting estimate of $b(w)$ can be used to select ranges of the most important wavelengths.

Quantifying catecholamines using multiway modeling

Rikke Nikolajsen, rn@ami.dk, The Royal Veterinary and Agricultural University, Department of Dairy and Food Science, Chemometrics Group. Lersø Parkalle 105, 2100 Ø, Denmark
Aase Marie Hansen, The National Institute of Occupational Health, Denmark
Karl S. Booksh, Department of Chemistry and Biochemistry, Arizona State University, USA
Rasmus Bro, rasmus@optimax.dk, The Royal Veterinary & Agricultural University, Dept. of Dairy and Food Science, Food Technology. Rolighedsvej 30, 1958, Frederiksberg, Denmark, http://www.models.kvl.dk/users/rasmus/

Keywords: PARAFAC, fluorescence kinetics, multiway data, catecholamines

Objective: The overall objective is the development of a fast and inexpensive analytical method for the determination of nanomolar concentrations of the two catecholamines, adrenaline and noradrenaline, in human urine. Measurements of catecholamines in urine are e.g. used in occupational health investigations, as catecholamine excretion is reported to be sensitive to mental stress.

Theory: Adrenaline's and noradrenaline's native fluorescence landscapes are practically identical. A reaction that generates different fluorescence landscapes and/or kinetic profiles for the analytes is therefore necessary. The rate at which the fluorescing 3,5,6-trihydroxyindole derivatives (lutines) are formed is different for adrenaline and noradrenaline. Further, the two derivatives have slightly different excitaiton- and emission maxima. The formation of the two derivatives have been monitored by measuring excitation-emission landscapes over time. This results in 4 way data: 1. excitation wavelength, 2. emission wavelength, 3. time, and 4. concentrations of analytes. The intensity of fluorescence landscapes in dilute solutions ideally follow a trilinear PARAFAC model. If further time is used as a variable, the data can be expected to follow a quadrilinear PARAFAC model.

Results: In the work presented here standard solutions of the catecholamines were used; both pure solutions and mixtures of the two. For each sample 60 fluorescence landscapes were collected within 25 minutes. The excitation range was 360-420 nm and the emission range 450-610 nm, while concentrations ranged from 30 to 1400 nM. A two component four way PARAFAC model seems to fit the data. Especially the time profiles look as could be expected, with adrenaline having a very steep curve that falls off because of degradation of the formed derivative, while the noradrenaline curve shows a much slower increase in intensity. The concentration mode regressed against reference values shows good linear relationship for both analytes. Potentially this reaction, as well as a clean-up step to separate the catecholamines from the urine matrix, can be implemented in an automated system, which will result in big savings of both time and reagents compared to e.g. the HPLC methods currently used.

Dioxin contamination of fish oil. PARAFAC and N-PLS analysis of fluorescence spectra

Dorthe Kjær Pedersen, dkp@kvl.dk, The Royal Veterinary and Agricultural University, Department of Dairy and Food Science, Food Technology. Rolighedsvej 30, 1958 Frederiksberg C, Denmark, http://www.mli.kvl.dk/foodtech/index.htm
Lars Munck, lmu@kvl.dk, Royal Veterinary and Agricultural University, Chemometrics Group, Food Technology, Department of Dairy and Food Science. Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark
Søren Balling Engelsen, se@kvl.dk, The Royal Veterinary and Agricultural University. Rolighedsvej 30, 1958, Denmark

Keywords: multiway, fluorescence, PARAFAC

Food and feed may contain residues of environmental contaminants. Monitoring programs are required to analyse food items for the presence of trace amounts of toxic substances such as heavy metals, pesticides, polyaromatic hydrocarbons (PAH), polychlorinated biphenyls (PCB), dioxins, flame retardants such as polybrominated diphenylethers (PBDE), and estrogenic compounds such as nonyl phenols. Normally the analytically methods are based on hypersensitive (sub ppb) physicochemical separation techniques such as ICP-MS, GC-MS and HPLC-MS. In case of complex organic molecules such methods are often laborious and very expensive and as a result, only limited monitoring can be performed.

Fluorescence spectroscopy has the potential to rapidly measure sub ppm levels of complex organic molecules due to the normally low background fluorescence signal (few molecules exhibit fluorescence). This poster presents the most recent spectroscopic determination of the dioxin content in fish oil by fluorescence measurements. Since most of the above-mentioned organic substances do not fluoresce the method is based on indirect correlations. We pursue two possible hypotheses: (1) That fluorescent indicator substances[1] follow the dioxin through the trophic levels and (2) that the presence of sub ppb levels of PCBs and dioxin exhibit sufficiently strong quenching effect that the "normal" background fluorescence is diminished. The two hypotheses are enlightened by application of N-PLS[2] and PARAFAC[3].

References:
[1] Munck,L., Nørgaard, L., Engelsen, S. B., Bro, R., and Andersson, C. A. Chemometrics in food science - a demonstration of the feasibility of a highly exploratory, inductive evaluation strategy of fundamental scientific significance. Chemometrics and Intelligent Laboratory Systems 1998, 44, 31-60
[2] Bro,R. Multiway calibration. Multi-linear PLS. Journal of Chemometrics 1996, 10, 47-61
[3] Harshman,R.A., and Lundy, M. E. PARAFAC: Parallel Factor Analysis. Computational Statistics and Data Analysis 1994, 18(1), 39-72

Application of simple interval calculation method

Alexey Pomerantsev, polycert@chph.ras.ru, Semenov Institute of Chemical Physics, Polycert. 4, Kosygin Str, Moscow, Russia, http://polycert.chph.ras.ru
Oxana Rodionova, rcs@chph.ras.ru, Polycert, Semenov Institute of Chemical Physics. 4, Kosygin Str., Moscow, Russia, http://polycert.chph.ras.ru

Keywords: prediction uncertainties, multivariate calibration, interval estimations, octane number, simplex algorithm

A simple interval calculation (SIC) is a method [1] of multivariate data analysis which yields the results of modeling and prediction in the interval form. We demonstrate the results of application of SIC-procedure to real world data. We use a classical example "Prediction of Gasoline Octane Number" [2] to show the feasibility of SIC-method.

In this example we use PCR projecting and apply both validation methods such as full cross-validation and prediction testing. To advocate the current approach we compare the confidence intervals for predicted values constructed for PCR results with SIC-intervals. The size of assessing intervals are quit sensible, and they correspond to the empirical demands and the traditional statistics. So, the proposed method for interval prediction construction may be useful for practical applications.

All SIC-calculations were made with the specially designed software that includes the following base algorithms: NIPALS algorithm for matrices decomposition, Standard Simplex algorithm [3] for optimization, and special procedure for preliminary data analysis that yields to reduce any initial problem to the form that is suitable for efficient application of simplex procedure.

References:
1. O. Ye Rodionova, AL Pomerantsev. SIC (Simple Interval Calculation) - a new approach for linear modelling. *J. Chemometrics*, submitted.
2. K. H. Esbensen. *Multivariate Data Analysis - In Practice* 4-th Ed., CAMO, (2000), 598p
3. G. Dantzig. *Linear Programming and Extensions*, Princeton University Press, Princeton, New Jersey, 1963.

Successive estimating of reaction rate constants from spectral data: A case study of two-step kinetics

Alexey Pomerantsev, polycert@chph.ras.ru, Semenov Institute of Chemical Physics, Polycert. 4, Kosygin Str, Moscow, Russia, http://polycert.chph.ras.ru

Keywords: kinetics, successive Bayesian estimating, spectroscopy, fitter, non-linear regression

Nowadays spectral data are often used to determine the kinetic parameters. Computerized spectroscopy provided us with a rapid on-line method of measurements [1]. There are many "soft" methods available to solve this problem where the employed approach consists of performing separate soft calibration between spectra and concentration units. Sometimes "soft" methods fail to provide the proper accuracy of results. On the other hand, "hard" kinetic models built on the main principles of chemical physics give the strong basement for data analysis. Such methods can be easily applied when spectra of pure components are known. But when they are unknown - some essential problems oppose their usage. In this work we suggest a new way of "hard" analysis of spectral data, which is suitable even in cases when pure component spectra are unknown.

The aim of this work is to show the feasibility of the method of successive Bayesian estimating in application to non-linear models with a large number of unknown parameters[2]. The kinetics of spectral data gives the excellent example where this technique may be used most naturally and fruitfully. We also wish to demonstrate that this algorithm can be carried out with the help of FITTER software [3], designed for non-linear regression analysis [4].

This method is of general nature and it can be used for any kind of kinetic models. This algorithm can deal with a strong spectral overlap and with an extremely small number of time points. It was demonstrated also that this method agrees with the OLS method in case of short-range spectra. From the real-world example it can be concluded that the successive method leads to lower deviations and correlations of reaction rate constants estimates in comparison with some known methods. This method is Bayesian only in its form not in its idea. No subjective a priori information is actually used in the approach. Each a priori information element is produced from experimental data during previous steps and only the form of its application is dictated by the Bayes theorem. No extra assumptions (like the number of PCs or a time-shift) are needed for this method.

References:
[1] S Bijlsma, DJ Louwerse, AK Smilde. Estimating rate constants and pure UV-VIS spectra of a two-step reaction using trilinear models. *J. Chemometrics*, 1999; **13** :311-329
[2] AL. Pomerantsev. Successive estimating of reaction rate constants from spectral data: a case study of two-step kinetics. *J.Chemometrics*, 2001, submitted
[3] Fitter AddIn. [Online]. http://polycert.chph.ras.ru/fitter.htm, [1 July 2001]
[4] EV Bystritskaya, AL Pomerantsev, OYe Rodionova. Non-linear regression analysis: new approach to traditional implementations. *J.Chemometrics*, 2000; **14**: 667-692

Generalization of pair-correlation method (PCM) for nonparametric variable selection

Róbert Rajkó, rajko@sol.cc.u-szeged.hu, University of Szeged, College Faculty of Food Engineering, Department of Unit Operations and Environmental Engineering. Mars tér 7., H-6725, Hungary, http://www.szef.u-szeged.hu
Károly Heberger, heberger@chemres.hu, Hungarian Academy of Sciences, 292. H-1025 Budapest, Pusztaszeri ut 59/67., postal address: H-1525 Budapest, P.O. Box 17, Hungary

Keywords: nonparametric variable, model selection, generalized pair-correlation method

Pair-Correlation Method (PCM) has been developed for choosing between two, correlated predictor variables provided that the scatter is caused not only by random effects. After a heuristic use of PCM, we have developed several test statistics. PCM is, in part, analogous to a 2x2 contingency table. We have investigated and compared the Conditional Exact Test, the McNemar's test, the Chi-square test, and the Williams' t-test. A macro based on the MS Excel 8.0 Visual Basic for Application (VBA) was also constructed which yielded a user-friendly and easy-to-use application because of the spreadsheet properties.

PCM can easily be generalized for variable selection purposes using more than two variables. The comparison of factors can be made pair-wise in all possible combinations. Every comparison can mark a factor as superior, inferior or no decision can be made. Then the factors are ordered according to the number of their superiority. If a given statistical test indicates a significant difference between the descriptors, we use the terms: superior - inferior or winner - looser for the overwhelming and subordinate descriptors, respectively.

The following step is the ordering of variables. Three ways of ordering have been elaborated: (i) simple ordering, (ii) ordering of differences and (iii) ordering according to probability weighted differences. (Difference here means wins minus losses). Moreover, PCM can be generalized for any fixed nonlinear model.

This scientific research was supported by the *Hungarian Science Foundation (No. OTKA T035125).*

References:
1. K. Héberger, R. Rajkó: Discrimination of statistically equivalent variables in quantitative structure-activity relationships. *In Quantitative Structure-Activity Relationships (QSAR) in Environmental Sciences-VII*, Ed. Fei Chen & Gerrit Schüürmann, SETAC Press, 1997, Ch. 29, 423-431
2. R. Rajkó, K. Héberger: Conditional Fisher's exact test as a selection criterion for pair-correlation method. Type I and Type II errors. *Chemometrics and Intelligent Laboratory Systems*, 57/1, 1-14, 2001
3. K. Héberger, R. Rajkó: Variable selection for environmental data using pair-correlation method. *SAR and QSAR in Environmental Research*, 2001, accepted for publication

Application of PLS and back propagation neural networks for the determination of soil samples properties

Ziad Ramadan, Departments of Chemical Engineering and Chemistry, Clarkson University, Potsdam, NY 13699-5705
Philip Hopke, hopkepk@clarkson.edu, Clarkson University. Box 5705 , 13699-5705 , USA
Mara Johnson, Department of Land, Air, and Water Resources, University of California, Davis, CA 95616
Kate Scow, Department of Land, Air, and Water Resources, University of California, Davis, CA 95616

Keywords: PLS, artificial neural network, soil properties, DNA profiles, crop type

Two different multivariate calibration methods, Partial Least Square (PLS) and Back Propagation Neural Networks (BP-ANN) were applied to Microbial community DNA to predict soil properties (%Sand, %Silt, %Clay, %Nitrogen, %Organic Carbon, %DNA) in environmental soil samples. The microbial community DNA was extracted from 48 environmental soil samples derived from different soils. After amplification of bacterial ribosomal RNA genes by polymerase chain reaction (PCR), the products were separated by gel electrophoresis.

Characteristic complex band patterns were obtained, indicating high bacterial diversity. Two hundred and fifty-six DNA-band patterns produced in the gels of the soil samples were used for the determination of the soil property, after removal of included DNA standard markers. Based on the brightness of the bands, densitometric curves of the selected DNA band pattern were extracted from the gel images. The curves were smoothed using Savitsky-Golay method and scaled to the DNA standard markers. The predictive power of the two methods (PLS and AP-ANN) will be presented and compared.

A methodological multi-way analysis of Cassava Starch properties

Marlon M. Reis, marlon@iqm.unicamp.br, Chemistry Institute - UNICAMP. Cidade Universitária Zeferino Vaz, s/n, Campinas, Brazil
Márcia M. C. Ferreira, marcia@iqm.unicamp.br, Universidade Estadual de Campinas, Instituto de Química. Campinas, SP, 13083-970, Brazil
Silene B. S. Sarmento, Department of Agroindustry, Food and Nutrition, "Escola Superior de Agricultura -Luiz de Queiroz" Universidade de São Paulo, Piracicaba Brazil

Keywords: multiway exploratory analysis, Cassava starch, constrained, TUCKER

The original methods proposed by Ledyard R. Tucker during the 1960s for Multi-Way Analysis present the rotational freedom problem, making the interpretation of its results rather difficult to be carried out. With the goal of making the multi-way data analysis more straightforward, this work uses a methodology of extracting meaningful information from the data set. The present methodology is based on the decomposition of data set in 3-way blocks by using Constrained Tucker Model. The aim of using this approach is to keep in one block all the similar information about data properties. The decomposition used is due to a Constrained Tucker Model, where the core array has some of its elements fixed to zero.

This work deals with a data set formed by the properties of four cassava cultivars, harvested at different ages during the usual harvesting of cultivars for industrial usage (age-properties-cultivars). The formulation of the Constrained Tucker Model considered independence among blocks. The inertia function, which gave information about how the data slices are described by the core slice, was fundamental for the final adjustment of the Constrained Tucker Model. This methodology is interesting since the vectors on the A and B modes, which show the correlation between properties and age, are directly related in one block, making its analysis quite easy. Although the correlation among the considered properties, the starch structure, the age stage and the cultivars varieties presents itself as a complex puzzle, the three-way analysis carried out showed to be able to provide useful information about the data, helping to choose the best harvesting period, considering the starch's properties which are important for industry.

The authors acknowledge the financial support from FAPESP for carrying out this work.

Supervisory control of wastewater treatment operation by PC-space control

Christian Rosen, christian.rosen@iea.lth.se, Lund University, Industrial Electrical Engineering and Automation. Box 118, SE-211 00, Lund, Sweden, http://www.lu.se/
Ulf Jeppsson, Industrial Electrical Engineering and Automation, Lund University, Box 118, SE-221 00 Lund, Sweden

Keywords: PCA, wastewater, process control

The requirements on wastewater treatment operation have increased in terms of effluent quality, efficient use of resources and process and personnel safety. This has led to an increase in investments of sensors and control equipment as well as new process configurations with extended control capabilities to optimise the operation. As a consequence, the need for co-ordination of the local controllers (supervisory control) has increased to avoid counteractive control and to find the appropriate controller set points to obtain a desired output.

In this paper, we present a multivariable steady-state control approach, based on control of the process location in the principal component (PC) space. The approach was originally proposed by Piovoso and Kosanovich in 1994 (Int. J. Control. (59) 3). A PCA model is identified on a set of controlled variables, a number of process variables and one or several output/target variables. New data is projected onto the model and the difference between the desired location (set points for the output/target variables expressed in the PC-space) and the current location is computed. The control law can then be expressed in terms changes in controlled variables by mapping the difference in PC-space onto the measurement space. The resulting controller is a multivariable controller with integral action only, describing the steady state relation between the controlled variables and the output/target variables.

An important objection to the above approach is that when the loop is closed, the system characteristics will change, including the process poles and gain. The changes in the poles are not crucial as the aim of the controller is to control the process to a steady-state (or rather in quasi steady-state). The change in process gain is more important. Techniques based on pole placement may be used to compensate for this, but this requires knowledge on the process, which is not always attainable. Moreover, if the process is non-linear, pole placement may be impossible. In our approach we instead add a compensation term to the control law. This term expresses the difference between the current target value and the desired value. By introducing this term in a PI fashion, errors due to non-linearities and stationary gain changes can be overcome.

Implementation of multivariate real-time methodologies for industrial process control

Jonas Röttorp, jonas.rottorp@ivl.se, Sweden

Keywords: SPC, data mining

Industrial processes are complex dynamic multi-variable systems. In the context of a global competitive economy and reinforced public environmental policies, more elaborated industrial process control strategies are needed. Traditional univariate methods for process monitoring and control do not satisfy evaluation of these complex data.

A new approach designated to optimise such complex systems in real-time is proposed which integrates recent information technologies along side with new statistical theory developments. The procedure covers the different steps from on-line data capture to statistical process control (SPC), going through database management systems, data mining as well as multivariate modelling. In each area several aspects such as process dynamics, adaptive control etc., have to be taken into account. The different steps in these methodologies will be discussed and a case study will be presented.

Characterisation of noise uncertainty: Allan variance and sample variance

Tatiana Siraya, skt703@atom.nw.ru, Khlopin Radium Institute, CSRI «Elektropribor», St Petersburg, Russia
Nikolay Khramov, Khlopin Radium Institute, CSRI «Elektropribor», St Petersburg, Russia

Keywords: variance, noise, Allan variance, Hilbert space

In practice the most popular characteristic of data scatter is a classic sample variance, which is the best estimate of variance in case of random sample with Gaussian distribution. Otherwise one can use other estimates, such as truncated sample variance, median absolute deviation, mean absolute deviation, quartile deviation estimate. Lately another scatter characteristic was proposed - Allan variance.

Allan variance proved to be very useful in complicated data processing problems, especially in case of white noise, flicker or 1/f type noise.It is highly important in the measurements of time and frequency. However, Allan variance is usually introduced as just an empirical value. In the report two ways are proposed for the formalisation of Allan variance as an important scatter characteristic of data. Allan variance is also compared with classic sample variance for revealing the scope of each estimate.

The Statistical approach is based on the testing of statistical hypotheses. The test statistic is the ratio of Allan variance to the sample variance. It is used to test the null hypothesis of a random sample with constant mean and variance against the alternatives of systematic shift in data, or the time series with non-correlated increments. The properties of the ratio under hypotheses are investigated for comparing the fields of application of Allan variance and sample variance.

The second approach is based on the reproducing kernel Hilbert space H(R), which provides an isomorphic representation of time series (random process) with correlation function R (s, t). An important advantage of H(R) representation is that it is applicable for both stationary and non-stationary processes, and also for generalized random processes (such as white noise or flicker noise). The properties of Allan variance are investigated for various types of noise, including white, flicker noise and some others. This representation is also related with the canonic innovation representations by H.Wald and H.Cramer for non-deterministic time series (random processes).

It appeared that for the time series with non-correlated increments, the corresponding space H(R) has the norm, which is calculated just according the formula of Allan variance. This fact slightly elucidates the nature of Allan variance as noise scatter characteristic, and explains its efficiency for the cases of non-correlated increments, white and 1/f type noise.

Development of a new fermented fish food product

Erik Slinde, erik@imr.no, Institute of Marine Research, Department of Aquaculture. Postboks 1879 Nordnes, N-5098 Bergen, Norway, http://imr.no
Tom Johannessen, MATFORSK, Norwegian Food Research Institute, Osloveien 1, N-1430 Ås, Norway
Bjørg Narum Nilsen, bjorg.narum.nilsen@matforsk.no, MATFORSK. Osloveien 1, N-1430, Norway
Brit Oppegård Pedersen, MATFORSK, Norwegian Food Research Institute, Osloveien 1, N-1430 Ås, Norway
Frank Westad, frank.westad@matforsk.no, MATFORSK. Arildsvingen 12, Oslo, Norway, http://www.matforsk.no
Per Morten Kjærnes, TINE Norwegian Dairies BA, P.O.Box 7 Kaldbakken, 0902 Oslo, Norway
Berit Nordvi, berit.nordvi@tine.no, TINE Norwegian Dairies BA, P.O.Box 7 Kaldbakken, 0902 Oslo, Norway

Keywords: fermentation, salami, salmon, saithe, NIR

Lactic acid bacteria have been used for centuries for the preservation of food. Smoked and graved salmon are two common fish products. Smoked salmon are characterized by a relatively high water content, while graved salmon has a characteristic fermented taste. Salami sausage is a traditional fermented meat product. By combining the good properties of graved and smoked salmon with the fermentation technology from the salami production, it has been possible to obtain a new fermented, smoked and dried seafood product with good storage properties.

Batch fermentation technology is used for the production of fermented foods. The ingredients in the recipe are the main contributors to the final product. The mixing of the ingredients has to be determined and adjusted on-line. We therefore chose NIR (Near Infrared Reflectance) to determine the amount of protein, fat, water and colour properties of the batter to be fermented, and PCA (Principal Component Analysis) to describe the start and end properties of the product. The growths of the bacteria are determined by the amount of sugar added, and the final pH and water content of the product are determined by the process conditions, temperature (20-25 ºC) and loss of water. Addition of spices and antioxidants are important for the taste of the final product.

A product has been developed that has acceptable taste and texture properties. The main problem was to stabilize the polyunsaturated marine fat. This has been achieved by immobilization of the fat with protein, and different protein sources have been studied.

Performance of multivariate calibration methods for determination of the active ingredient and impurity in a pharmaceutical process solution analysed by near infrared spectroscopy

Laila Stordrange, nkjlt@kj.uib.no, University of Berge. Allégt. 41, 5007, Norway
Fred Olav Libnau, Nycomed Amersham Imaging, Nycovn. 2, N-0401 Oslo, Norway
Dick Malthe-Sørenssen, Nycomed Amersham Imaging, N-4510 Spangereid, Norway
Olav Kvalheim, olav.kvalheim@kj.uib.no, University of Bergen, Department of Chemistry. Allégaten 41, N-5007 Bergen, Norway

Keywords: near-infrared spectroscopy, preprocessing, multivariate calibration

Near Infrared (NIR) spectroscopy has been used in multivariate calibration models for measuring the initial compound and impurity in a pharmaceutical primary production of active ingredient used in contrast media.

Presently the reaction is monitored offline by HPLC. The goal is to replace the time-consuming reference method with on-line NIR monitoring. NIR is a fast and non-destructive method that needs no sampling and gives continuous real time surveillance of the reaction.

The calibration models developed are based on spectra measured at-line. This is to enhance the knowledge of the predictive ability of the multivariate model of the process.

The process solution is chemically complex. In addition variation of physical properties affecting the spectra make the modelling a challenging task. Different data pre-treatment methods and variable selections reported in the literature have been tested to enhance the chemical information and reduce irrelevant variability due to physical influence on measured spectra, as scattering and temperature. In addition to standard pre-treatment methods, as normalisation, differentiation and multiplicative scattering correction, newer methods such as orthogonal signal correction and optimized scaling have been tested. The modelling were performed by using Partial Least Squares (PLS), except for optimized scaling which uses Principal Component Regression (PCR). Normalisation gave the highest error of prediction while optimized scaling on differentiated data gave the lowest error of prediction. The data show signs of a non-linear behaviour. Non-linear calibration models were developed to model these non-linearities.

Covariate challenge in multivariate statistical process monitoring

Pekka Teppola, pekka.teppola@basf-ag.de, BASF AG, Scientific Computing. ZDP/C-C13, 67056 Ludwigshafen, Germany
Gerhard Krennrich, gerhard.krennrich@basf-ag.de, BASF AG. BASF AG , 67056, Germany
A. Schreieck, anna.schreieck@basf-ag.de, BASF AG, Scientific Computing, ZDP/C-C13, 67056 Ludwigshafen (Rh), Germany
G. Jones, geoff.jones@basf-c-s.co.uk, BASF Computer Services (UK) Ltd., Seal Sands, P.O. Box 62, Middlesbrough, TS2 1TX, United Kingdom
D. Kratz, BASF AG, 67056 Ludwigshafen (Rh), Germany
R. Krokoszinski, BASF AG, 67056 Ludwigshafen (Rh), Germany

Keywords: chemical industry, process chemometrics, MSPM, covariates, direct orthogonalization, OSC

A present application at BASF Ludwigshafen exhibits a new challenge in multivariate statistical process monitoring (MSPM). This challenge is due to covariate variables measured on a continuous scale. This covariate is used in tuning the process to meet an almost infinite number of different customer settings and correspondingly different normal operating conditions specific and optimal to different final product formulations. By definition, we say that covariate is a variable that may affect relationships between variables of interest, but is not of intrinsic interest itself.

In this work, we apply partial least squares (PLS) with two pretreatment techniques, namely, direct orthogonalization (DO) and orthogonal signal correction (OSC) to monitor a continuous industrial process with continuously changing product formulations, respectively. In this approach, DO extracts and removes first a source of variation from process variables (X) and responses (Y) which is related to covariate information (Z) and which is not of interest in this application. After removing the effect of covariates (Z) and forming new corrected X and Y matrices, OSC is used to prune the process data (X) and to remove a part of variation in a new X which is not related to responses (Y). These two pretreatment steps, DO and OSC, result in conditioned data which will be then monitored. It will be exemplified by using two simple simulated data sets with a discrete/continuous covariate how a direct orthogonalization procedure with PLS is easier to interpret and more sensitive than a standard PLS. After this, examples are given to demonstrate industrial use of a combined DO, OSC, and PLS procedure.

On spurious models in process monitoring and fault identification - industrial experiences

Pekka Teppola, pekka.teppola@basf-ag.de, BASF AG, Scientific Computing. ZDP/C-C13, 67056 Ludwigshafen, Germany
Gerhard Krennrich, gerhard.krennrich@basf-ag.de, BASF AG. BASF AG , 67056, Germany
A. Schreieck, anna.schreieck@basf-ag.de, BASF AG, Scientific Computing, ZDP/C-C13, 67056 Ludwigshafen (Rh), Germany
G. Jones, geoff.jones@basf-c-s.co.uk, BASF Computer Services (UK) Ltd., Seal Sands, P.O. Box 62, Middlesbrough, TS2 1TX, United Kingdom
D. Kratz, BASF AG, 67056 Ludwigshafen (Rh), Germany
R. Krokoszinski, BASF AG, 67056 Ludwigshafen (Rh), Germany

Keywords: chemical industry, process chemometrics, MSPM, PLS, EW-PLS, contributions, residuals

In multivariate statistical process monitoring (MSPM), process models are normally based on data with only common-cause variation representing either nominal or target operating conditions. In principle these models detect non-directionally any observable deviation from the nominal/target data set with respect to common-cause variation. This involves three steps which are fault detection, isolation, and identification. In this particular work, we show that fault identification tools, often providing valuable information, still suffer from what we call spurious model behavior. We also suggest several routes how this kind of behavior can take place. In fact all these routes boil down to 'calibration' data and covariance matrices.

Usual problems in MSPM are a lack of stationary and/or representative data with respect to future data and similarly a tendency towards clustered data. Both of these easily introduce problems. It seems that the former is more like a rule than exception. The latter imposes very often either masking or leverage effects or otherwise broadens the confidence limits of MSPM models if some clusters are undesirable. Compared to single outliers, outlying clusters are much more problematic as of being very difficult to detect. Another issue due to clustering is that unequal clusters can cause unwanted weighting effects, i.e. large clusters downweight small clusters. Similarly this weighting problem concerns not only samples but also variables. Besides these there are other phenomena very different from each other such as selection of variables, scaling of data, treatment of missing values, and covariates, which all have an impact on the calibration data, covariance matrices, and models.

All the above reasons can lead to problems and especially in online fault identification. In this work, we show with a very simple example how spurious model behavior propagates from raw data to model scores and from these to residuals. Thus occasionally both model and residual diagnostics in fault identification become corrupted. Contrary to standard PLS algorithms, we compute PLS as successive singular value decompositions(1).

1. Kaspar, M.H., Ray, W.H., Partial least squares as successive singular value decompositions, Computers Chem. Eng., 17 (1993) 985-989.

Modelling time series of Low Field 1H NMR relaxation curves of starch retrogradation. Comparison of PARAFAC, TUCKER3, slicing and multi-exponential fitting

Lisbeth G. Thygesen, lit@kvl.dk, The Royal Veterinary and Agricultural University, Department of Food and Dairy Science, Food Technology. Rolighedsvej 30, 1958 Frederiksberg C, Denmark, http://models.kvl.dk
Søren Balling Engelsen, se@kvl.dk, The Royal Veterinary and Agricultural University. Rolighedsvej 30, 1958, Denmark
Andreas Blennow, Plant Biochemistry Laboratory, Department of Plant Biology, The Royal Veterinary and Agricultural University, Thorvaldsensvej 40, DK-1871 Frederiksberg C, Denmark

Keywords: NMR, time series, PARAFAC, TUCKER3, slicing

Starch retrogradation is a phenomena influencing the shelf life of a number of food products.The aim of the present study was to study the process and if possible find correlations between retrogradation behaviour and structure characteristics such as the degree of phosphorylation, the polymorph type and the relative contents of amylose and amylopectin. Low Field NMR relaxometry was used to monitor starch retrogradation and in this presentation we focus on comparing four different ways of modelling the NMR time series: PARAFAC[1], TUCKER3[2], slicing (Decra)[3] and multi-exponential fitting[4].

A set of 26 starches from different botanical sources and with a wide range in the degree of phosphorylation was gelatinized in NMR tubes, incubated at 98 ^{o}C for 30 min., and cooled to 35 ^{o}C. CPMG (Carr-Purcell-Meibomm-Gill) relaxation curves at 35 ^{o}C were obtained using LF ^{1}H NMR at 11 different times after the cooling. The data cube obtained comprised 51 samples (25 of the samples were measured in replicates) x 3250 point CPMG curves x 11 times during retrogradation.

The models are compared with regard to a number of important parameters including the number of latent variables/components suggested and the grouping of the samples, and it is discussed which models best explain known differences among the samples.

References:
[1] Harshman, R.A. Foundations of the PARAFAC procedure: Models and conditions for an 'explanatory' multi-modal factor analysis. UCLA working papers in phonetics 1970, 16, 1-84.
[2] Tucker, L.R. Implications of factor analysis of three-way matrices for measurement of change, in (C.W. Harris, ed.) Problems of Measuring Change, University of Wisconsin Press, Madison, MI, 1963, pp.122-137.
[3] Winding, W. and B. Antalek, Direct exponential curve resolution algorithm (DECRA): A novel application of the generalized rank annihilation method for a single spectral mixture data set with exponentially decaying contribution profiles.Chemometrics and Intelligent Laboratory Systems 1997, 37, 241-254.
[4] Bechmann, I.E., H.T. Pedersen, L. Nørgaard and S.B. Engelsen, Comparative Chemometric Analysis of Transverse Low-field ^{1}H NMR Relaxation Data, in (P.S. Belton and G.A. Webb, eds.) Advances in Magnetic Resonance in Food Science, The Royal Society of Chemistry, Cambridge, UK, 1999, pp.217-225.

Multiblock models for explorative data mining in food technology

Frans van den Berg, fb@kvl.dk, The Royal Veterinary and Agricultural University, Food Technology, Chemometrics Group. Rolighedsvej 30, DK-1958, Denmark, http://www.models.kvl.dk
Anette Thybo, anette.thybo@agrsci.dk, Danmarks JordbrugsForskning, Hortikulturinstituttet, Årslev, Danmark
Rasmus Bro, rasmus@optimax.dk, The Royal Veterinary & Agricultural University, Dept. of Dairy and Food Science, Food Technology. Rolighedsvej 30, 1958, Frederiksberg, Denmark, http://www.models.kvl.dk/users/rasmus/

Keywords: multiblock, explorative, data mining

Within the framework of the Advanced Quality Monitoring (AQM) project a study is initiated to investigate the possibilities of advanced models, collectively known as multiblock methods.

These techniques are extensions of better-known multivariate data analysis and regression methods like Principal Component Analysis (PCA) and Partial Least Squares (PLS).

Multiblock methods can be beneficial when analysing systems where the variables or measurements are organized in conceptually meaningful blocks. Examples of such 'natural' blocks are different measurements used on the same sample, or data collected at different stages in a production pathway (raw material > intermediate > product).

Multiblock methods strive to maintain this natural ordering in the data. They are considered exploratory in that they focus on explaining the relation between blocks, and the block's relative contribution in the final model.

Although multiblock techniques have been around for quit some time, to our knowledge there have been very few applications in the realm of food technology.

Uniaxial compression data for predicting potato quality parameters

Frans van den Berg, fb@kvl.dk, The Royal Veterinary and Agricultural University, Food Technology, Chemometrics Group. Rolighedsvej 30, DK-1958, Denmark, http://www.models.kvl.dk
Anette Thybo, Danmarks JordbrugsForskning, Hortikulturinstituttet, Årslev, Danmark
Rasmus Bro, rasmus@optimax.dk, The Royal Veterinary & Agricultural University, Dept. of Dairy and Food Science, Food Technology. Rolighedsvej 30, 1958, Frederiksberg, Denmark, http://www.models.kvl.dk/users/rasmus/

Keywords: uniaxial compression, sensory, texture

Texture is very important for the consumer's perception of potato quality. Dry matter (specific gravity) is known to be an important factor for texture development.

Within the framework of AQM a study is initiated to predict important physical parameters and sensory attributes from the full uniaxial compression curves on raw and/or cooked potato samples.

Uniaxial compression is widely used for texture determination of fruits, vegetables, gels, cheeses and potatoes to determine mechanical properties. For namely potatoes the relation between curve parameters as stress, strain and moduli and the sensory quality is only sparsely treated in the literature.

Selection of optimal process analyzer for monitoring

Frans van den Berg, fb@kvl.dk, The Royal Veterinary and Agricultural University, Food Technology, Chemometrics Group. Rolighedsvej 30, DK-1958, Denmark, http://www.models.kvl.dk

Keywords: process-analyzers, selection, position

An ever-increasing number of process analyzers are implemented in industry. At the same time the diversity in techniques suitable for harsh process conditions - e.g. chromatography, (near)infrared-, Raman- or (low field) nuclear magnetic resonance spectroscopy, mass spectrometry, flow injection analysis, ultrasonic analysis, to name just a few - grows steadily. The implementation and operation of analytical in-process measurements is, however, still relatively expensive. The cost of purchase and maintenance often limits the number of analyzers that can be implemented for monitoring and/or control purposes to one or a few key-components of the process. This naturally leads to the following questions: what is the added value of process analyzers, what is the better choice from the wide selection of process analyzers, and what is the best location to place this limited number of instruments? All these questions are related and can only be answered adequately by simultaneously looking at the process under observation. The 'information content' of measured process variables is a function of the underlying process dynamics, the external process disturbances and of the process analyzer measuring these variables. The dynamic behavior of various important process variables e.g. reactant versus product can be quite distinct. An important objective is thus to sample the process variable with the most information in its measured signal, at the most informative position in the process (e.g. reactor inlet versus outlet). The characteristics of a process analyzer - e.g. slow but precise GC-analysis versus fast but relative imprecise spectroscopic-measurements - determine which technique is best suited for the analysis task at hand.

A state observer can be used to estimate process variables from the (discrete) in-process measurements. The process state is a collection of all the important process variables e.g. concentration of all constituents participating in a reaction (both measured and unmeasured). State observers also provide an expected estimation error in the form of a covariance uncertainty matrix of the state estimate. The optimal process analyzer type and position is selected by minimizing this state estimation error.

Simultaneous processing of raw data of samples and standards for the enhancement of selectivity and sensitivity in the liquid chromatographic analysis of cocaine

Paul van Zomeren, p.v.van.zomeren@farm.rug.nl, University of Groningen, Pharmacy, Pharmaceutical Analysis. PO Box 196, NL-9700 AD, The Netherlands, http://www.farm.rug.nl/interact/pars.html
H. J. Metting, Department of Pharmaceutical Analysis, Groningen University Institute for Drug Exploration, Antonius Deusinglaan 1, 9713 AV, Groningen, The Netherlands
Pierre Coenegracht, p.m.j.coenegracht@farm.rug.nl, University Groningen Dep. Pharmacy . A.Deusinglaan 1 , 9713 AV , The Netherlands
G. J. de Jong, Department of Pharmaceutical Analysis, Groningen University Institute for Drug Exploration, Antonius Deusinglaan 1, 9713 AV, Groningen, The Netherlands
D. Pol, Department of Pharmaceutical Analysis, Groningen University Institute for Drug Exploration, Antonius Deusinglaan 1, 9713 AV, Groningen, The Netherlands

Keywords: liquid-chromatography, diode array detection, cocaine, PARAFAC, PARAFAC2

A series of cocaine standards and a number of seized drug samples, containing cocaine, related compounds and cutting agents, were analysed by high performance liquid chromatography with diode array detection. In order to take advantage of the different selectivity of various separation conditions and of the different composition of various samples and standards, the data-matrices, obtained from all measurements on all samples and standards, were handled simultaneously.

Combination of the data-matrices, obtained for a single sample or standard but with various separation conditions, was achieved by augmentation. Matrices can only be augmented, when one of their dimensions is equal. Since the same detector, wavelength range and sample interval were used, the number of wavelength points was equal. Therefore, the matrices were augmented in the time direction. Combination of the augmented data-matrices, originating from various samples and standards, was achieved by stacking. Matrices can only be stacked, when both of their dimensions are equal. This requirement was met, since the samples and standards were analysed under similar conditions. Finally, the PARAFAC2 algorithm was used to decompose the array that was formed by augmenting and stacking of the individual data-matrices.

One of the resulting loadings contained the relative concentrations of the sample components, while the other loadings contained the spectra and chromatographic profiles. Since the concentration of cocaine was known for the standards, it could be calculated for the samples. Furthermore, the relative concentrations of the other sample components could be used to classify the cocaine samples.

Simultaneous processing of data obtained by high performance liquid chromatography and capillary electrophoresis with diode array detection

Paul van Zomeren, p.v.van.zomeren@farm.rug.nl, University of Groningen, Pharmacy, Pharmaceutical Analysis. PO Box 196, NL-9700 AD, The Netherlands, http://www.farm.rug.nl/interact/pars.html
H. J. Metting, Department of Pharmaceutical Analysis Groningen University Institute for Drug Exploration Antonius Deusinglaan 1, 9713 AV, Groningen, The Netherlands
H. Darwinkel, Department of Pharmaceutical Analysis Groningen University Institute for Drug Exploration Antonius Deusinglaan 1, 9713 AV, Groningen, The Netherlands
D. Pol, Department of Pharmaceutical Analysis Groningen University Institute for Drug Exploration Antonius Deusinglaan 1, 9713 AV, Groningen, The Netherlands
P. M. J. Coenegracht, Department of Pharmaceutical Analysis Groningen University Institute for Drug Exploration Antonius Deusinglaan 1, 9713 AV, Groningen, The Netherlands
G. J. de Jong, Department of Pharmaceutical Analysis Groningen University Institute for Drug Exploration Antonius Deusinglaan 1, 9713 AV, Groningen, The Netherlands

Keywords: liquid-chromatography, capillary-electrophoresis, diode array detection, alternating-least-squares, augmentation

Mixtures of nitrazepam, clonazepam and lorazepam were analysed by high performance liquid chromatography (HPLC) with diode array detection (DAD) and by microemulsion electrokinetic chromatography (MEEKC) with DAD. A number of seized drug samples, containing cocaine, related compounds and cutting agents, were analysed by HPLC-DAD and by capillary zone electrophoresis (CZE) with DAD. In order to take advantage of the different selectivity of the liquid chromatographic and capillary electrophoretic separation systems, the data-matrices, obtained from measurements under various conditions with both systems, were handled simultaneously. This simultaneous data-processing was performed for each sample separately. Combination of the data-matrices was achieved by augmentation. Matrices can only be augmented, when one of their dimensions is equal.

Since the same type of detector was used with the same wavelength range and sample interval, the number of wavelength points was equal. Therefore, the matrices were augmented in the time direction. Finally, multivariate curve resolution methods were used to process the augmented data-matrices. Modified forms of alternating least squares (ALS) and iterative target testing factor analysis (ITTFA) were used, which took the multimodal character of the augmented chromatographic profiles into consideration. Differences in peak height and noise level and calibration of the wavelength scale proved to be critical aspects. ITTFA led to lower detection limits and a higher quality of the resolved chromatographic profiles and spectra than ALS.

An example showing the use of qualitative variables in experimental design applied to a process of making cheese

Malin Wikström, malin.wikstrom@chem.umu.se, Umeå Universitet, Organic Chemistry, Chemometrics.
Organisk kemi, Umeå Universitet, 901 87 Umeå, Sweden

Keywords: design of experiments, PLS, qualitative variables, cheese, off-flavour

The use of experimental design is a well know procedure for investigation or optimization of a product or a process. The factors in a process are classified as either quantitative or qualitative. A quantitative variable can be seen as a continuous factor that can take any number between the predefined levels in the design. A qualitative variable on the other hand is called qualitative in respect of the fact that the factor may only be varied at two distinct levels such as present/not present, on/off, method A/B and so on. A qualitative variable may also be termed as discrete. This project concerns the work with experimental design and qualitative variables in a process of making cheese. The objective was to find out which stage of the process that cause off-flavor in the cheese. A full factorial design in five qualitative variables, representing the location (A or B) at which different steps in the process takes place, was constructed using Modde 5.0 (Umetrics). The design resulted in a total of 35 experiments (25 + 3 replicates).

Each experiment in the design represents a unique combination of production of cheese. Milk from the same batch is used at both locations, guaranteeing the same starting conditions. The cheeses are thereafter moved between the locations and hence creating the possibility to compare if the rate of off-flavor is affected by the treatment of the cheese.

The experiments were evaluated by a trained sensory panel consisting of 6 assessors, which scored the cheese as 1 if off-flavor was detected and as 0 if not. The average score for each sample was used as response, i.e. the rate of off-flavor. This means that if 5 out of 6 assessors scored a sample as 1, the response was calculated as 5/6 = 0,83.

Partial Least Square regression (PLS) was used as regression method to find the relationship between the treatments (X) and the rate of off-flavor (Y). A two-component PLS-model with R2 at 0.763 and Q2 at 0.588, indicate a good model. Evaluation of the design and the regression coefficients clearly showed that one of the factors, x3 for location B, was responsible for a high rate of off-flavor. The other factors had little or no effect.

Calibration transfer by generalized least squares

Barry M. Wise, bmw@eigenvector.com, Eigenvector Research, Inc.. 830 Wapato Lake Road, Manson, WA 98831, USA, http://www.eigenvector.com/
Harald Martens, Harald.Martens@mail.tele.dk, DTU/NTNU/KVL . Teglgaardstr 12A , DK-1452 , Denmark
Martin Høy, martin.hoy@pvv.ntnu.no, NTNU. Fak. kjemi, Inst. kjemi , 7491, Norway
Rasmus Bro, rasmus@optimax.dk, The Royal Veterinary & Agricultural University, Dept. of Dairy and Food Science, Food Technology. Rolighedsvej 30, 1958, Frederiksberg, Denmark, http://www.models.kvl.dk/users/rasmus/
Per B. Brockhoff, The Royal Vet. and Agric.l University, Dept. of Dairy and Food Science, DK-1958 Frederiksberg C, Denmark

Keywords: calibration transfer, instrument standardization, generalized least squares

The instrument standardization/calibration transfer problem has been addressed by a wide variety of methods. In this work, we investigate the ability of Generalized Least Squares (GLS) preprocessing methods to deal with artifacts causes by changes in spectroscopic instrumentation that would normally require the building of complete calibration models. GLS preprocessing works by measuring a number of transfer samples on two or more instruments. These samples can be used to estimate an offset and shift in the covariance structure of the data due to instrument differences. The method of GLS proprocessing shown recently by Harald Martens et. al. can then be used to remove variation in the data which is not common to both instruments. Calibration models can then be built on data from one of the instruments and used on the other, with the GLS preprocessing applied prior to predictions on new samples.

The GLS method is tested on two data sets from near infrared spectroscopy, one being pseudo-gasoline mixtures and the other corn measured on three instruments. Comparisons are made to other calibration transfer methods. In particular, the similarities and differences to orthogonal signal correction are discussed.

Analysis and display of historical Stehekin river flow data with robust PCA methods

Barry M. Wise, bmw@eigenvector.com, Eigenvector Research, Inc.. 830 Wapato Lake Road, Manson, WA 98831, USA, http://www.eigenvector.com/

Keywords: robust PCA, least median squares, trimmed least squares

Principal Components Analysis (PCA) continues to be a mainstay of multivariate data analysis. New applications continue to be developed. In this poster, PCA is applied to the daily flow rate of the Stehekin River (Chelan County, Washington, USA) as a method to characterize its natural variation over the course of a year. Daily average flows are available starting in 1927 and continuing to the present day. While PCA is quite useful in elucidating the typical types of variation seen in the flow, it is also quite influenced by anomalies in the data, namely, floods.

Robust PCA methods, less influenced by outlier data, are used to characterize "typical" river variability. PCA based on a least median squares criteria is used as well as a trimmed PCA variant. Iterative re-weighting with generalized least squares preprocessing is also considered. The methods are contrasted using the Stehekin River data as the primary example. Methods for display of the data are highlighted, with the goal being to develop an display that is accessible to the layman.

Automated LC-MS analysis of natural products: extraction of UV, MS and retention time data for compound identification and chemometric analysis

Deborah Zink, debbie_zink@merck.com, Merck & Co Inc., Natural Products Chemistry. PO Box 2000 R80Y335, Rahway NJ 07065, USA
Claude Dufresne, claude_dufresne@merck.com, Merck & Co., Inc., Natural Products Chemistry, Automation and Informatics Group. R80Y-360, POB 2000, Rahway, NJ 7065, USA, http://www.merck.com
Jesus Martin, Merck Research Laboratories Natural Products Chemistry Rahway New Jersey 07065 USA
Valdimir Svednik, Merck Research Laboratories, Biometric Research, Rahway New Jersey 07065 USA
Andrew Liaw, Merck Research Laboratories, Biometric Research, Rahway New Jersey 07065 USA

Keywords: LC-MS, multidimensional searching, MDS

LC-MS analysis has become a valuable tool in our natural products drug discovery program. As the number of screening samples increases it became important to be able to quickly analyze LC-MS data with a minimal amount of specialist interaction. As a result, we are developing general, automated LC-MS analysis tools for high throughput applications.

Our initial application (written in Visual Basic) allows one to quickly identify previously characterized natural products in semi purified samples and provides data sets for evaluation with chemometric tools for the analysis of mixtures. For compound identification UV spectral data is extracted directly from the raw Agilent diode-array file. The mass spectral data is extracted using AMDIS dll's (AMDIS was developed by NIST for GC-MS data). The data is combined based on retention time of the extracted peaks producing a summary of the characteristic data for each peak in the LC-MS run. A molecular weight interpretation algorithm is applied to assign a molecular weight using both positive and negative ion data. The full data set is then searched in a custom database. For mixture analysis, a matrix of the full UV data and MS data is produced for further analysis.

References:
1. An Integrated Method for Spectrum Extraction and compound Identification from Gas Chromatography/Mass Spectrometry Data. S. E. Stein, Journal of The American Society for Mass Spectrometry, vol. 10, Num 8, August 1999, pg. 170-781
2. Automated Data Massaging Interpretation, and E-Mailing Modules for High Throughput Open Access Mass Spectrometry. Hui Tong, et al, Journal of The American Society for Mass Spectrometry, vol. 10, Num 11, August 1999, pg. 1174-1187.
3. Optimization and Testing of Mass Spectral Library Search Algorithms for Compound Identification. Steven E. Stein, Donald R. Scott; Journal of The American Society for Mass Spectrometry, vol. 5, 1992, pg. 859-866.

Keyword Index

Author Index

List of Participants

Name	Affiliation	Address	Phone/Fax number
Mr. Christoffer Abrahamsson	Lund Institute of Technology Atomic Physics	P.O. Box 118 SE-221 00 Lund Sweden	+46-46-2223120
Mr. Christian Airiau	Bristol University School of Chemistry	Cantcock United Kingdom	
Mr. Kazeem Alabi Abiodun	Polygon Investment CC Ltd.	P. O. Box 2625 Pretoria South Africa	27731999347 27123245568
Dr. Bjorn Alsberg	University of Wales, Aberystwyth Department of Computer Science Computational Biology Group	Llandinam bldg, Ceredigion SY23 3DB United Kingdom	00441970622537 00441970622455
Ms. Charlotte Møller Andersen	Danish Institute for Fisheries Research	DTU, build. 221 2800 Denmark	45252542 45884774
Mr. Ulf Andersen	Arla Foods Amba Innovation Product Functionality	Rørdrumvej 2 Brabrand Denmark	+4587466751 +4587466688
Ms. Tina Birgitte Argonne Andersen	Novo Nordisk Engineering	Færgevej 65 3600 Denmark	44427635
Mr. Per Andersson	Melacure Therapeutics	Ulleråkersvägen 38 SE 756 43 Sweden	+4618530088 +461818530070
Dr. Claus A. Andersson	BriSense Innovation ApS On-line Monitoring Systems	Gammeltoftsgade 12B DK-1355 Copenhagen K. Denmark	+4522203335 +4522203335
Dr. Henrik Antti	Imperial College	London United Kingdom	
Mr. Börkur Arnvidarson	ChemoMetec A/S	Gydevang 43 DK-3450 Allerød Denmark	(+45)48131020 (+45)48131021
Mr. Tom Artursson	Linköping University	IFM, Linköpings universitet 58183 Sweden	+4613281357 +4613288969
Mr. Søren Balling Engelsen	The Royal Veterinary and Agricultural University	Rolighedsvej 30 1958 Denmark	+4535283205
Dr. Meir Bar	Government	20 Zivoni 34651 Haifa Israel	972-48340140 972-48795386
Mr. Suleman Bashiru	KMC	10 Arthur Rive, Bakau Banjul Gambia	
Dr. Dorrit Baunsgaard	Novo Nordisk A/S Applied Trinomics Bio-NMR	Novo Nordisk Park DK-2760 Måløv Denmark	+4544448888 +4544663939
Ms. Iben Ellegaard Bechmann	Novo Nordisk Engineering A/S	Krogshøjvej 55 DK-2880 Denmark	+4544447777 +4544443777
Dr. Anton Belousov	Institut fuer Chemo- und Biosensorik	Mendelstr. 7 D-48149 Germany	+4902519802892 +4902519802890
Ms. Nancy Bendwell	Tembec Inc.	PO Box 3000 J0Z 3R0 Canada	(819)627-4315 (819)627-9908

Mr. Frank Berg Rasmussen	Haldor Topsøe A/S	Nymøllevej 55 Lyngby Denmark	45272766
Ms. Ingunn Berget	MATFORSK	Oslovn. 1 1430 Norway	+4764970322 +4764970333
Mr. Eric Bernard	Nexfor Technology	240 Hymus Blvd. Pointe-Claire Canada	514-630-9503 514-630-9379
Mr. Anders Björk	Royal Inst. of Technology (KTH) Dep. of Chemistry Div. of Analytical Chemistry	Royal Inst. of Technology (KTH) SE-100 44 Stockholm Sweden	+4687908216 +468108425
Dr. Rasmus Bro	The Royal Veterinary & Agricultural University Dept. of Dairy and Food Science Food Technology	Rolighedsvej 30 1958, Frederiksberg Denmark	+4535283296
Mr. Derek Byrne	The Royal Veterinary and Agricultural University Dairy and Food Science Sensory Science	Rolighedsvej 30, 5 sal. 1958 Denmark	+4535283174
Mr. Jakob Christensen	The Royal Veterinary and Agricultural University Food Technology	Rolighedsvej 30 1958 Danmark	+4535283323 35283245
Mr. Jan Christensen	National Environmental Research Institute Department of Environmental Chemistry	Frederiksborgvej 399 4000 Roskilde Denmark	46301200
Mr. Ivan Christensen	Teknologisk Institut	 Denmark	
Mr. Pierre Coenegracht	University Groningen Dep. Pharmacy	A.Deusinglaan 1 9713 AV The Netherlands	+3150363-3348/3336 +3150363-7582
Dr. Eigil Dåbakk	Q-Interline AB	Box 2086, Lövängsvägen 8 SE-194 02 Upplands Väsby Sweden	+46859072000 +46859072010
Mr. Casper K. Dahl	Aalborg University Esbjerg	Torvegade 68B, 2, th. Esbjerg Denmark	+4575181956
Mr. Therese Dahlström	Ovako Steel AB	TAA 303 S-813 82 Sweden	+4629025598 +4629025670
Mr. Lars-Göran Danielsson	Royal Inst. of Technology (KTH) Dep. of Chemistry Div. of Analytical Chemistry	Royal Inst. of Technology (KTH) SE-100 44 Stockholm Sweden	+46(0)87908215 +46(0)8108425
Mr. Rolf Danielsson	Uppsala University Institute of Chemistry Dept. of Analytical Chemistry	Box 531 SE - 751 21 Sweden	+46184713678 +46184713692
Dr. Onno de Noord	Shell International Chemicals	P.O. Box 38000 Amsterdam Netherlands	
Ms. Gunvor Dingstad	Matforsk	Osloveien 1 AAs Norway	+4764970314 +4764970333
Dr. Claude Dufresne	Merck & Co., Inc. Natural Products Chemistry Automation and Informatics Group	R80Y-360, POB 2000 Rahway, NJ 7065 USA	732-594-5783 732-594-6880
Ms. Marianne Dyrby	The Royal Veterinary and Agricultural University Department of Dairy and Food Science Chemometrics Group	Rolighedsvej 30 1958 Frederiksberg Denmark	+4535283564 +4535283245
Dr. Tim Ebbels	Imperial College	London United Kingdom	

Mr. Max Egebo	FOSS Electric R & D Center of Chemometrics	Slangerupgade 69 3400 Denmark	+4570103370 +4548208070
Dr. Rolf Ergon	Telemark University College	Telemark University College, P.O.Box 203 Porsgrunn Norway	++4735575160 ++4735575250
Mr. Lennart Eriksson	Umetrics AB	Umetrics AB S-907 19 Sweden	+46.90.184852 +46.90.184899
Prof. Kim Esbensen	Ålborg Universitet Esbjerg	Norgesgade 31, 1.th DK-6700 Denmark	
Mr. Richard Escott		United Kingdom	
Dr. Klaas Faber	ATO Agro & Industrial Production Chains Production & Control Systems	P.O. Box 17 NL-6700 AA Wageningen Netherlands	+31317475311 +31317475347
Dr. Márcia M. C. Ferreira	Universidade Estadual de Campinas Instituto de Química	Campinas, SP 13083-970 Brazil	5501937883102
Dr. Alberto Ferrer	Universidad Politécnica de Valencia Dpto. Estadística Quality Improvement Group	Camino de Vera S/N Edif. I-3 46022 Valencia Spain	0034963877007Ext.49 32 0034963877499
Mr. Henrik Fodgaard	Radiometer Medical A/S Research and Sensor Development	Åkandevej 21 2700 Danmark	38273249
Ms. Ragnhild Frank	Novozymes A/S Production and Procurement Granulation Denmark	Krogshøjvej 36 (Att: RaFr, EC1.31) 2880 Bagsværd Denmark	+4544435305 +4544435611
Dr. Jens Christian Frisvad	Technical University of Denmark Biocentrum	Lyngby Denmark	
Ms. Stina Frosch	Danmarks Fiskeriundersøgelser	DTU, byg. 221 2800 Danmark	45254922 45884774
Mr. Volker J. Frost	SensoLogic GmbH Software + Sensor Systems	Hummelsbuetteler Steindamm 78a 22851 Norderstedt Germany	+49/40529567-41 +49/40529567-99
Mr. Jon Gabrielsson	Umeå University	Organic chemistry, Umeå University 901 87 Umeå Sweden	46907865359 4690138885
Dr. Bjørn Grung		Norway	
Mr. Alf Gustafsson	MoRe Research	MoRe Research S-891 80 Örnsköldsvik Sweden	+46(0)66075022 +46(0)66075981
Mr. Per Waaben Hansen	FOSS A/S Research & Development	Slangerupgade 69 DK-3400 Hillerød Denmark	+4548208524 +4548208070
Mr. Per Anker Hassel	University of Newcastle	Centre for Process Analytics and Control Technology Newcastle upon Tyne United Kingdom	441912225382 441912225748
Dr. Károly Heberger	Hungarian Academy of Sciences 292	H-1025 Budapest, Pusztaszeri ut 59/67. postal address: H-1525 Budapest, P.O. Box 17 Hungary	+364380411ext.426 +363257554

Mr. Kaj Heydorn	IBX International	Department of Chemistry, Technical University of Denmark Kgs. Lyngby Denmark	45252342 45883136
Mr. Peter R. Hillestrøm	University of Copenhagen DHI - Water & Environment	Ålandsgade 8; 1tv 2300 Kbh. S. Denmark	32576915
Mr. John Holm	Danisco Cultor	Edwin Rahrs Vej 38 DK-8220 Denmark	+4589435206 +4586251077
Mr. Martin Holmberg	Linköping University S-SENCE	IFM/Linköping University S-581 83 Sweden	
Mr. Philip Hopke	Clarkson University	Box 5705 13699-5705 USA	3152683861 3152686654
Mr. Agnar Høskuldsson	Danish Technical University	Danish Technical University Lyngby Denmark	+4545255643 +4545931577
Mr. Lars Houmøller	Aalborg University Esbjerg Chemistry Chemical Analysis	Niels Bohrs Vej 8 DK-6700 Denmark	4579127645 4575453643
Mr. Martin Høy	NTNU	Fak. kjemi, Inst. kjemi 7491 Norway	+4792286768 +4773591676
Dr. Christiane Jaeckle	Wacker Polymer Systems	Wacker Polymer Systems, P.O. Box 1560 D-84483 Burghausen Germany	+498677836275 +498677834606
Mr. Kjell Janné	Uppsala University	Center for Biotechnology Uppsala Sweden	018-4716252 0
Ms. Åsa Jansson	IVL - Swedish Environmental Research institute	Box 210 60 100 31 Sweden	+46859856346 +46859856390
Ms. Ingela Jedvert	Astrazeneca, Mölndal Pharmaceutical and Analytical R&D PAC	Box S-431 83 Mölndal Sweden	+46317761256
Mr. Thomas Würtz Jensen	Nycomed Pharma Pharmaceutical Development Analytical Development	Langebjerg 1 DK-4000 Denmark	+4546771236 +4546756640
Ms. Inger Jönebring	AstraZeneca Tablet Production Sweden	 Sweden	
Mr. Pär Jonsson	Umeå University	Organic chemistry SE-901 87 Sweden	+46907867102 +4690138885
Dr. Bo Jørgensen	Danish Institute for Fisheries Research Department of Seafood Research Raw-material and product technology	DTU build. 221 DK-2800 Denmark	+4545252547 +4545884774
Ms. Pia Jørgensen	Biotechnological Institute	Holbergsvej 10 DK-6000 Denmark	+4575520433 +4575529989
Dr. Mats Josefson	AstraZeneca R&D, Mölndal	AstraZeneca R&D, Mölndal Mölndal Sweden	+46-31-7761643 0
Mr. Lefteris Kaskavelis	Unilever	Olivier van Noortlaan 120 P.O.BOX 114, 3130 AC The Netherlands	(31)104605547 (31)104605671
Mr. Hector Keun	Imperial College	Imperial College London United Kingdom	02075943142 0
Dr. Rudolf Kiralj	Instítuto de Química	UNICAMP 13083-970 Brazil	551937883102 551937883023

Ms. Anita Knudsen	BMA ApS	Risingevej 1 Vallensbaek Strand Denmark	43567420 43567403
Ms. Maaret Korpelainen	Lappeenranta university of Technology	 Lappeenranta Finland	
Dr. Dora Kourti	McMaster University	1280 Main Street West Hamilton Canada	19055259140 19055211350
Dr. Gerhard Krennrich	BASF AG	BASF AG 67056 Germany	
Prof. Olav Kvalheim	University of Bergen Department of Chemistry	Allégaten 41 N-5007 Bergen Norway	
Mr. Mattias Landin	SCA Graphic Sundsvall AB Development Department Process Technology	Ortviken Paper Mill 851 23 Sweden	+4660194205 +4660574328
Mr. Rasmus Larsen	DTU Informatics and Mathematical Modelling	IMM, building 321 DK-2800 Denmark	+4545253415 +4545881397
Mr. Jukka Laukkanen	VTT Mineral Processing	Tutkijankatu 1 83500 Finland	+358135571 +35813557557
Dr. Riccardo Leardi	University of Genova Department of Pharmaceutical and Food Chemistry and Technology Section of Analytical Chemistry	via Brigata Salerno (ponte) I 16147 Italy	+390103532636 +390103532684
Ms. Valérie Lengard	CAMO ASA Consulting & Training Services	Nedre Vollgt. 8 N-0158 Oslo Norway	+4722396300 +4722396322
Mr. Casper Leuenhagen	H. Lundbeck A/S Analycal Controll Production Raw Material & Process Technology	P. Freuchensvej 21B, 1, 3 DK-4800 Nykøbing F. Denmark	+4536301311 +4536309951
Mr. Johan Lindberg	Biovitrum AB	Rapsg 751 82 Sweden	+46(0)86973825 +46(0)86973912
Mr. Carsten Lindemann	Delta	Hjortekaersvej 99 Denmark	
Ms. Anna Linusson	AstraZeneca	AstraZeneca R&D Molndal 431 43 Sweden	
Prof. Torbjörn Lundstedt	Melacure Therapeutics	Ulleråkersvägen 38 SE 756 43 Sweden	+4618530072 +461818530070
Prof. John MacGregor	McMaster University Department of Chemical Engineering	Hamilton, ON Canada L8S 4L7 Canada	
Mr. Thomas Magnussen	Novo Nordisk	Novo Nordisk Park, G8.1.59 Måløv Denmark	
Mr. Staffan Malmberg	StoraEnso Hylte AB Development	Box 300 31481 Sweden	+4634519271
Prof. Magni Martens	The Royal Veterinary and Agricultural University Chemometrics Group	The Royal Veterinary and Agricultural University 1958 Denmark	35283197 +45-35283210
Prof. Harald Martens	DTU/NTNU/KVL	Teglgaardstr 12A DK-1452 Denmark	+4521468766 +4533327240
Ms. Elaine Martin	University of Newcastle	Merz Court NE1 7RU United Kingdom	+441912226231 +441912225748

Dr. Arthur Mateos	FMC Corporation	Box 8, US1 Zip Code: 08543 New Jersey, USA	609-951-3149 609-951-3372
Mr. Rune Mathisen	Borealis as	Rønningen 3960 Norway	+4735577665 +4735577055
Mr. Inger Hedvig Matveyev	EMT	Bergsbygda N-3914 Norway	+4735518541 +4735575250
Ms. Elisabeth Micklander	The Royal Veterinary and Agricultural University Food technology	Rolighedsvej 30 1958 Denmark	+4535283500 +4535283505
Mr. Pentti Minkkinen	Lappeenranta university of Technology	Lappeenranta University of Technology. FIN-53851 Finland	+35856212102 +35856212199
Ms. Tina Moe	Novo Nordisk Engineering A/S Manufacturing Information Systems Team Optimizer	Krogshoejvej 55 2880 Bagsvaerd Denmark	+4544447777 +4544443777
Dr. Trine Møgelberg	Novo Nordisk	Laurentsvej 50, bygn 8L 1.36 2800 Denmark	44421192 44421260
Mr. Peter Paasch Mortensen	Novozymes A/S	Hallas Alle 1 axs36 DK-4400 Denmark	+4544435110 +4544435740
Prof. Lars Munck	Royal Veterinary and Agricultural University Chemometrics Group, Food Technology Department of Dairy and Food Science	Rolighedsvej 30 DK-1958 Frederiksberg C Denmark	
Dr. Allan Aasbjerg Nielsen	Technical University of Denmark Informatics and Mathematical Modelling Image analysis	IMM/DTU building 321 DK-2800 Kgs. Lyngby Denmark	+4545253425 +4545881397
Mr. Jesper Pram Nielsen	The Royal Veterinary and Agricultural University	Rolighedsvej 30 1958 Frederiksberg C. Denmark	35283500 35283505
Mr. Henrik Aalbrog Nielsen	Technical University of Denmark Informatics and Mathetical Modelling Statistics	Richard Petersens Pl., Bygn. 321 2800 Lyngby Denmark	45253418 45881397
Mr. Martin Høigaard Nielsen	University of Copenhagen Department of Chemistry	Tagensvej 175, 4. th 2400 København NV Danmark	35857584
Dr. Andreas Niemöller	Bruker Optik GmbH FT-NIR Application	Rudolf-Plank-Str. 23 Ettlingen Germany	+497243504679 +497243504673
Ms. Rikke Nikolajsen	The Royal Veterinary and Agricultural University Department of Dairy and Food Science Chemometrics Group	Lersø Parkalle 105 2100 Ø Denmark	(+45)39165261 (+45)39165201
Ms. Bjørg Narum Nilsen	MATFORSK	Osloveien 1 N-1430 Norway	++64970100 ++64970333
Dr. Jonas Nilsson	Biovitrum AB	Rapsgatan 7 Sweden	
Dr. Lars Nord	Carlsberg Laboratory Department of Chemistry	Gamle Carlsberg Vej 10 DK-2500 Valby Denmark	+4533275221 +4533274708
Mr. Lars Nørgaard	The Royal Veterinary and Agricultural University Chemometrics Group Department of Dairy and Food Science	Rolighedsvej 30 DK-1958 Frederiksberg C Denmark	+4535283267 +4535283245
Mr. Maurice O´Sullivan	The Royal Veterinary and Agricultural University	Sensory Science, Rolighedsvej 30 DK-1958 Denmark	0045-3528-3174 +45-3528-3210

Mr. Henrik Öjelund	IMM	IMM, DTU DK-2800 Denmark	45253372 45882673
Ms. Ing-Marie Olsson	Umeå University Kemi institutionen Organisk kemi	Umeå University 90187 Sweden	+46907867102
Prof. Mathias Otto	Freiberg University	Leipziger Str. 29 D-09599 Germany	+4937313468 +493731393666
Ms. Joan Grønkjær Pedersen	Biotechnological Institute Integrated ProcessOptimisation	Holbergsvej 10 DK-6000 Denmark	+4575520433 +4575529989
Ms. Dorthe Kjær Pedersen	The Royal Veterinary and Agricultural University Department of Dairy and Food Science Food Technology	Rolighedsvej 30 1958 Frederiksberg C Denmark	4535283564 4535283245
Mr. Poul Erik Petersen	FOSS A/S Chemometric Department	Slangerupgade 69 3400 Denmark	+4548208509 +4548208070
Mr. Lars Petersen	Aalborg University Esbjerg	Amagervej 16, 3 sal (Lejl. 2) Esbjerg Ø Denmark	
Dr. Alexey Pomerantsev	Semenov Institute of Chemical Physics Polycert	4, Kosygin Str Moscow Russia	+7(095)9397483 +7(095)9397483
Ms. Vibeke T. Povlsen	The Royal Veterinary and Agricultural University Department of Dairy and Food Science Food Technology	Rolighedsvej 30 1958 Denmark	35283501 35283505
Ms. Tarja Rajalahti	Umeå University Organic Chemistry Research Group for Chemometrics	SE-90187 Umeå Sweden	+46907867102 +4690138885
Dr. Róbert Rajkó	University of Szeged, College Faculty of Food Engineering Department of Unit Operations and Environmental Engineering	Mars tér 7. H-6725 Hungary	
Mr. Ari Rantala	Outokumpu Mintec Oy Automation Product development	PO Box 84 Espoo Finland	+358408212944 +35894213373
Dr. Satu-Pia Reinikainen	Lappeenranta university of Technology Department of Chemical Technology Chemometrics Group	PO Box 20 Lappeenranta Finland	+358-5-6212112 +358-5-6212199
Mr. Marlon M. Reis	Chemistry Institute - UNICAMP	Cidade Universitária Zeferino Vaz, s/n Campinas Brazil	
Dr. Carsten Ridder	Foss Electric A/S Chemometric Centre	Slangerupgade 69 DK-3400 Hillerød Denmark	+4570103370 +4570103371
Mr. Åsmund Rinnan	The Royal Veterinary and Agricultural University Department of Dairy and Food Science Chemometrics Group	Rolighedsvej 25 1958 Denmark	4535283501 4535283505
Dr. Oxana Rodionova	Polycert Semenov Institute of Chemical Physics	4, Kosygin Str. Moscow Russia	+7(095)9397483 +7(095)9397483
Mr. Christian Rosen	Lund University Industrial Electrical Engineering and Automation	Box 118, SE-211 00, Lund Sweden	+46(0)462224582 +46(0)46142114
Mr. Jonas Röttorp	IVL - Swedish Environmental Research institute	Box 210 60 Stockholm Sweden	
Ms. Bente Sæten	Weifa AS R&D	P.O. Box 9113 Grønland NO-0133 Norway	+4722998600 +4722998601

Mr. Andy Scott	GlaxoSmithKline R&D, Pharmacy Division	Office 5F010, R&D, Park Road SG12 0DP United Kingdom	441920882942
Dr. Terhi Siimes	Nokia Mobile Phones Research and Technology Access	P.O.Box 407 FIN-00045 NOKIA GROUP Finland	+358407706758 +358718037290
Dr. Rolf Singer	Pronosco A/S	Kohavevej 5 DK-2950 Denmark	45650600 45650620
Mr. Jukka Sinisalo	Kemira	P.O. Box 44 02771 Espoo Finland	+358108622440 +358108622466
Mr. Jonas Sjöblom	Solvina AB	Gruvgatan 37 Göteborg Sweden	+46(0)317096316 +46(0)317096316
Mr. Michael Sjöström	Umeå University	Umeå University SE-901 87 Sweden	+46907865119 +4690138885
Mr. Erik Skibsted	Department Of Chemical Engineering Process Analysis and Chemometrics	Nieuwe Achtergracht 166 NL 1018 The Netherlands	0031205256991 0031613212321
Mr. Erik Skov	Radiometer Medical A/S Research and Sensor Development	Åkandevej 21 2100 Danmark	38273585
Mr. Erik Slinde	Institute of Marine Research Department of Aquaculture	Postboks 1879 Nordnes N-5098 Bergen Norway	+47-99538480 t+47-55236379
Prof. Age Smilde	University of Amsterdam Department of Chemical Engineering Process Analysis and Chemometrics	Nieuwe Achtergracht 166 NL-1018 WV Amsterdam Netherlands	+31205255062 +31205255604
Dr. Ib Søndergaard	Technical University of Denmark BioCentrum-DTU Biochemistry and Nutrition	Søltofts Plads, building 224 DK-2800 Denmark	+4545252733 +4545886307
Dr. Anders Sparén	AstraZeneca Tablet Production Sweden TQP	S-15185 Södertälje Sweden	+46855251430 +46855259083
Mr. Halfdan Stenholm		Denmark	
Mr. Bo Stenlöf	The Swedish Institute for Food and Biotechnology Flavour and Sensory Evaluation	Box 5401 S - 402 29 Göteborg Sweden	+46313355665 +4631833782
Ms. Laila Stordrange	University of Berge	Allégt. 41 5007 Norway	+4755589851 +4755589490
Mr. Rolf Sundberg	Stockholm University	Mathem. statistics SE-10691 Sweden	
Dr. Olof Svensson	AstraZeneca Pharmaceutical Analytical R&D Analytical and Technical Development	Pepparedsleden 1 S-431 83 Sweden	+46317761883 +46317763813
Mr. Ketil Svinning	NORCEM A/S	Setrevn. 2 N-3950 Norway	4735572314 4735570400
Mr. Kent Tano	LKAB	Malmberget S-983 81 Sweden	
Mr. Paul Taylor	McMaster University	1280 Main Street West Hamilton Canada	19055259140 19055211350
Dr. Pekka Teppola	BASF AG Scientific Computing	ZDP/C-C13 67056 Ludwigshafen Germany	+496216094523 +496216049463

Mr. Bernt Thelin	Bioglan AB	PO Box 50310 SE-202 13 Malmö Sweden	+4640287510 +4640291955
Ms. Lisbeth G. Thygesen	The Royal Veterinary and Agricultural University Department of Food and Dairy Science Food Technology	Rolighedsvej 30 1958 Frederiksberg C Denmark	+4535283323 +4535283245
Mr. Henrik Toft Pedersen	The Royal Veterinary and Agricultural University Department of Dairy and Food Science Food Technology	Rolighedsvej 30 1958 Frederiksberg C Denmark	(+45)35283564 (+45)35283245
Mr. Magnus Tolleshaug	Nycomed Pharma Pharmaceutical Development Analytical Development	Langebjerg 1 4000 Denmark	+4546771273 +4546771295
Mr. Giorgio Tomasi	The Royal Veterinary and Agricultural University Chemometrics Group	Duevej 58 2000 Danmark	35283501 35283505
Mrs. Kristin Tøndel		 Norway	
Mr. Johan Trygg	University of Umeå Research Group for Chemometrics	901 87 Umeå Sweden	+46907865999 +4690138885
Mr. Goran Urden	Fresenius Kabi	Hagvagen 4 755 97 Sweden	
Mr. Frans van den Berg	The Royal Veterinary and Agricultural University Food Technology Chemometrics Group	Rolighedsvej 30 DK-1958 Denmark	35283502 35283505
Mr. Paul van Zomeren	University of Groningen Pharmacy Pharmaceutical Analysis	PO Box 196 NL-9700 AD The Netherlands	+31503633335 +31503637582
Mr. Serguei Verzakov	Institut fuer Chemo- und Biosensorik	Mendelstr. 7 D-48149 Germany	+4902519802892
Dr. Jürgen von Frese	Institut für Chemo- und Biosensorik Chemometrics	Mendelstr. 7 D-48149 Germany	+49-2519802871 +49-2519802890
Mr. Jesper Wagner	Novo Nordisk Engineering A/S Manufacturing Information Systems Team Optimizer	Krogshøjvej 55 2880 Denmark	+4544422779
Mr. Lars Wallbäcks		Barrgränd 15 SE-944 71 Piteå Sweden	+4691197515,31339
Dr. Peter Wentzell	Dalhousie University Department of Chemistry	Halifax, Nova Scotia B3H 4J3 Canada	902-494-3708 902-494-1310
Mr. Frank Westad	MATFORSK	Arildsvingen 12 Oslo Norway	+4795765125 +4764970333
Dr. Johan Westerhuis	Chemical Engineering, University of Amsterdam Process Analysis and Chemometrics	Nieuwe Achtergracht 166 1018 WV AMSTERDAM The Netherlands	+31205255265 +31205255604
Mr. Tony Wiklund	AstraZeneca R&D	Mölndal Sweden	+46317761000 +467763811
Ms. Malin Wikström	Umeå Universitet Organic Chemistry Chemometrics	Organisk kemi, Umeå Universitet 901 87 Umeå Sweden	+46(0)907865999 +46(0)90138885
Dr. Barry M. Wise	Eigenvector Research, Inc.	830 Wapato Lake Road Manson, WA 98831 USA	509-687-2022 509-687-7033

Ms. Bettina Weber Wismann	FOSS Electric A/S	Slangerupgade 69 3400 Denmark	+4548208599 +4548208070
Prof. Svante Wold	University of Umeå	 Sweden	
Ms. Gunilla Wormbs	Arla Foods Innovation	Torsgatan 14 S-105 46 Sweden	+4686773206 +468203329
Mr. Ransford Yeboah	University of Ghana	Accra Africa	
Mr. Christian B. Zachariassen	CP Kelco Lille Skensved Quality Control Development	Ved Banen 16 DK-4623 Lille Skensved Denmark	+4556165616 +4556169446
Ms. Deborah Zink	Merck & Co Inc. Natural Products Chemistry	PO Box 2000 R80Y335 Rahway NJ 07065 USA	1-732-594-7207 1-732-594-6880

100 years ago…

Karl Pearson

On Lines and Planes of Closest Fit to Systems of Points in Space

Phil. Mag. (6), 2, 559-572, **1901**