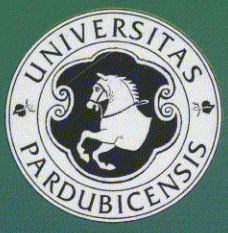


CHEMOMETRICS '95

the 4th International Chemometrics Conference of the Czech Chemical Society

July 3-7, 1995

ABSTRACTS OF PAPERS

Pardubice, Czech Republic

CHEMOMETRICS '95

the 4th International Chemometrics Conference of the Czech Chemical Society

July 3-7, 1995

ABSTRACTS OF PAPERS

Pardubice, Czech Republic

ORGANIZING COMMITTEE

Milan Meloun
Jiří Militký
Karel Kupka

PROGRAM COMMITTEE

Milan Meloun
Jiří Militký

CONFERENCE SECRETARIAT

CHEMOMETRICS'95
Department of Analytical Chemistry,
University of Pardubice,
nám. Čs. Legií 565,
532 10 Pardubice
Telephone: 004240-582 288
Fax: 004240-514 530
Email: meloun@hlb.upce.cz

Plenary Lectures

ACCEPTANCE SAMPLING BY VARIABLES IN NON – STANDARD SITUATIONS

JAROMÍR ANTOCH & JANA JUREČKOVÁ

Charles University, Dept. of Statistics
Sokolovská 83, CZ–186 00 Praha 8, Czech Republic
antoch@karlin.mff.cuni.cz, jureckova@karlin.mff.cuni.cz

ABSTRACT. Several procedures concerning the acceptance sampling by variables in the case of non–normally distributed quality parameter, resp. in the case when several normally distributed parameters are controlled simultaneously, will be described in the lecture

Research and practice in the fields of industry, chemistry, commerce etc. since many years utilize sampling procedures in evaluating product quality. Many of the methods suggested in the literature has found in the meantime its way to the standards. Unfortunately, when regarding the situation concerning the acceptance sampling by variables we can see that most of the standards are based on the assumption of normality (log-normality) of relevant parameter. Unfortunately, in the real life this is not always the case. On the contrary, we can often meet situations when:

- the distribution of the parameter of interest can be described by highly non – symmetrical, resp. truncated, distribution;
- acceptance is based on the simultaneous control of several parameters etc.

Let us assume that the requirements (AQL, LQ; α, β) on the effectiveness of the control between producer and consumer were agreed on. In the first part of the lecture we shall suppose that the control is based on the measurement of one quality parameter (characteristic) X and either lower tolerance limit L or upper tolerance limit U for controlled parameter is prescribed; the value of L (U respectively) being of course set by some standard. Instead of considering the classical assumption about normally (lognormally) distributed items we shall concentrate on the case when the parameter is either gamma distributed or follows truncated normal distribution.

More precisely, let us suppose that lower tolerance limit L for parameter X is prescribed such that the item is considered acceptable if and only if $x > L$, where x is the value of X obtained during the control. For gamma distributed items we shall propose

1991 *Mathematics Subject Classification.* 62N10.
Key words and phrases. Acceptance sampling by variables, gamma distribution, truncated normal distribution.
Partially supported by grant GAUK 365

sampling inspection procedure (n, k), where n is the number of items to be controlled and k is a constant such that the lot will not be accepted if

$$\overline{x}_n \leq kL,$$

$\overline{x}_n$ being the sample mean of measurements $x_1, \ldots, x_n$ of the respective parameter on n controlled items. On the contrary, in the case of truncated normal distributions several possibilities will be demonstrated enabling the use of the classical procedure suggested for the normal case.

In the second part of the lecture we shall concentrate on the sampling inspection procedure by m independent and normally distributed parameters which are controlled simultaneously. More precisely, we shall suppose that the parameters $X^{(1)}, \ldots, X^{(m)}$ under the control are independent, normally distributed and lower bounds $L^{(1)}, \ldots, L^{(m)}$ are given such that the item is considered acceptable if and only if $x^{(i)} > L^{(i)}$, $i = 1, \ldots, m$, where $x^{(i)}$ is the value of $X^{(i)}$. The main aim will be again to suggest sampling inspection procedure (n, k), where n is the number of items to be controlled and k is a constant such that the lot will not be accepted if

$$\overline{x}_n^{(i)} \leq L^{(i)} + k s_n^{(i)} \qquad \text{for at least one } i,\ 1 \leq i \leq m,$$

$\overline{x}_n^{(i)}$ and $s_n^{(i)}$ being the sample mean and sample standard deviation of measurements of i^{th} parameter on n controlled items. The situation when not all parameters are normally distributed will be also commented on.

REFERENCES

[1] Cohen A. C., *Estimating the mean and variance of normal populations from singly truncated and doubly truncated samples*, The Annals of Mathematical Statistics 21 (1950 a), 557 – 569.

[2] ČSN 010256, *Special types of Sampling Inspection by Variables*, Czechoslovak Office of Standards, Praha, 1969.

[3] Halperin M., *Estimation in the truncated normal distribution*, JASA 47 (1952), 457 – 465.

[4] Johnson N. L, Kotz S., *Distributions in Statistics: Continuous Univariate Distributions 1.*, Houghton & Mifflin, Boston, 1970.

[5] Jurečková J., Antoch J. et al., *Judgement of the system of evaluation of fertilizers and the proposal of the novelization of acceptance sampling by variables*, Technical Report, MFF UK, Praha, 1982 – 1984.

[6] Rusá L., *Acceptance Sampling by Variables*, Diploma Thesis, MFF UK, Praha, 1987.

[7] Sarkadi K., Vincze I., *Mathematical Methods of Statistical Quality Control*, Akadémiai KIADÓ, Budapest, 1974.

A Novel Approach to Structure-Spectra Correlation

J. T. Clerc, Ch. Affolter, K. Baumann
Pharmazeutisches Institut der Universität Bern
Baltzerstr. 5
CH-3012 Bern (Switzerland)

Summary: A new approach for the numerical representation of chemical structures of organic compounds is described and the strategy of its application to the prediction of the fingerprint region of IR spectra is presented. Preliminary results will be given.

Structure elucidation with spectroscopic methods as well as spectra prediction and estimation in effect link chemical structures with spectral data. This is based on the reasonable assumption that similar structures will give rise to similar spectra. The most common and simple representation of this link is realized in correlation tables and spectra libraries. Efficient use of the relation between structures and spectra requires the use of computers. However, computers cannot work directly on the real problem. They rather have to use a symbolic representation of the problem and of its components, and their ability to manipulate the problem is largely limited to syntactic rearrangements. Thus, the way how spectra and in particular, how structures are represented for use within a computer program is of paramount importance. The highest degree of sophistication in processing symbols with a computer is certainly realized for numbers. Thus, representations for spectra and for structures should preferably be numbers. The problem is then reduced to finding an algorithm to convert numeric structure codes into numeric spectra codes.

The relation between spectra and structures is extremely complicated, presently not known with sufficient precision, and includes many still unknown parameters. With the present state of knowledge we may at best find very crude and brutally simplified approximations, hardly ever useful for practical applications in the real world. This is particularly true for global solutions which attempt to cover a large part of the full domaine of chemistry.

Analytical problems, however, are always local problems. At a given instance the analyst is concerned with the current sample only. Thus, he may use local models, centered around the current sample and valid only for the current problem. The application range for such local approximations is strictly limited, but they tend to be simpler and often have more predictive power than global models. Construction of such ad hoc local models, however, is feasible only if the model building process is fast, cheap and easy.

In this lecture a new approach to the numerical representation of chemical structures is presented, and strategies for localization (in the sense given above) of IR spectra prediction are described. Preliminary results will be shown.

MULTIVARIATE CALIBRATION ALSO FOR SINGLE-COMPONENT ANALYSIS ?

Danzer, Klaus
Friedrich Schiller University of Jena, Department of Inorganic & Analytical Chemistry
Lessingstr. 8, D-07743 Jena

Summary: Multivariate calibration by PLS, CCA and ANN are used to improve the reliability of single-component analysis. Examples of NIR, spark-, GD-, ICP-OES and IPC-MS are given.

Nowadays, chemometrics get through all the steps of the analytical process. Sampling is strongly determined by statistical methods and analytical measurements are distinctly improved by chemometrical procedures like accumulation, filtering, deconvolution etc. Analytical evaluation achieved new dimensions by application of factor analysis and pattern recognition.

One of the most important field of application of chemometrics is analytical calibration. Methods like principal component regression (PCR), partial least squares (PLS) modelling, and artificial neural networks (ANN) make it possible to take into account entire signal functions for calibration and evaluation, e.g. whole spectra, instead of single signal values. Since the late seventies as S.Wold creates the photometric multicomponent analysis by means of PLS, multivariate calibration found many applications, particularly in the NIR spectrometry.

Noninvasive glucose determination in blood by means of NIR reflectance spectrometry is represented as an example of chemometrical methods evaluating entire spectra. NIR spectra being a priori only low shaped by glucose are additionally influenced by some patient-specific factors. Such subjective effects were eliminated by application of chemometric methods. Spectra of unsufficient quality are recognizable by means of cluster analysis and factor analysis and can be sorted out after that. On the other hand, all the spectra of good quality can be selected by means of a genetic algorithm (GA).

About 250 patients have been measured and more than 1000 spectra were recorded from them. In each case reference determinations were carried out by

means of a classical glucose analyzer (Beckman). The NIR measurements were carried out on the middle finger in reflectance by means of a measuring head consisting of two concentric glass fibres for irradiating and outgoing radiation, repectively. This equipment ist connected with (i) a grating spectrometer (Spectro 320; 800 to 1300 nm) and (ii) a diode array spectrometer (Polytec X-dap IR, 900 to 1700 nm).

The data were evaluated by means of multivariate methods using about 400 absorbance values from each spectrum. Calibration models were developed on the basis of (a) PLS regression where partly a wavelenghts selection by GA was used and (b) an artificial neural networks (radial basis function, RBF).

In general, the results of RBF are superior to those of PLS. In the case that PLS is combined with a pre-selection of spectra by clustering, factor analysis or GA the results are comparable. In almost all the cases the comparison functions (control functions) show no significant bias. By means of the grating spectrometer correlation coefficients $r > 0.8$ have been estimated, in case of the diode array spectrometer correlation coefficients > 0.9 have been found.

Other spectroscopic disciplines like optical atomic emission spectroscopy and mass spectroscopy produce a lot of signals per element from which the most sensitive (as a rule) is choosen for quantitative analysis. Chemometric methods make it possible to include several signals per element in the evaluation. The multivariate calibration can be carried out by PCR, by the established PLS modelling or by canonical correlation analysis (CCA). Examples given from spark OES, glow discharge OES, and ICP-OES show that sensitivity and detection limit can be improved significantly by this way.

It is of particular importance that a great part of disturbed signals may be included in the evaluation. Serious disturbances appear in ICP-MS caused by isotopic and molecular overlappings. By means of multivariate calibration such disturbances can be overcome as will be shown by examples.

TRANSFER OF SPECTRA AND OF CALIBRATION FUNCTION IN N.I.R.S.

Giuliana DRAVA, Michele FORINA

Institute of Food and Pharmaceutical Analysis and Technology, University of Genova - Via Brigata Salerno (ponte), 16147 Genova (Italy).

Summary: This lecture concerns the transfer of NIR spectra by means of PLS regression, in order to predict the signal of an instrument from the signal measured with another instrument. The results are compared with PDS method (Piecewise Direct Standardization). The results of the transfer of calibration function between instruments are also reported.

NIRS is widely used in quality control, in particular in food industry, since it is a fast method, which requires a very reduced sample pretreatment.

Mulitvariate calibration is applied in order to compute the calibration function relating a block of spectral variables (e.g. absorbances or reflectances at different wavelenghts) with the chemical quantity of interest.

For N samples, the chemical quantity is estimated by:

$$\mathbf{y} = \mathbf{Xb} \qquad [1]$$

$\mathbf{y}$ = column vector (N elements) of the chemical quantity;

$\mathbf{X}$ = matrix of spectral data, having N rows and as many columns as the number of variables + a column of ones;

$\mathbf{b}$ = column vector of the coefficients (V slopes + intercept).

Usually, $\mathbf{b}$ is computed by using a *biased* regression method, as PCR or PLS.

At present NIR instruments having different characteristics and performances are available, but even two identical instruments do not give identical spectra for the same sample. For this reason, calibration procedure must be performed on each instrument: moreover, since the instruments are subjected to variations during the time, calibration must be performed repeatedly on the same instrument.

Therefore, transfer of spectra and/or calibration function from an instrument to another is required in order to handle the information obtained from different instruments (e.g. a master and a server) or from the same instrument after different times, as obtained from only one instrument.

The transfer is possible by means of:

$$\mathbf{X}_{master} = \mathbf{X}_{server}\mathbf{C} \qquad [2]$$

where C is the matrix of the regression coefficients and $\mathbf{X}_{master}$ and $\mathbf{X}_{server}$ are the data matrices of the two instruments.

The transfer of spectra requires a suitable number of standards, properly selected and a suitable regression method for computing C. Shenk-Westerhaus method (1), PDS (Piecewise Direct Standardization) (2), PLS (3), PC-PLS (3), OPA-UPA (Orthogonal Procrustes Analysis, Unconstrained Procrustes Analysis) (4) can be applied.

In a recent paper, the performances of PDS were shown to be better than the method of Shenk (5).

In this lecture the results obtained with PLS are compared with PDS: the influence of the number of standards and of the criterium for their selection were also studied.

The results obtained show that:

a) the quality of the transfer procedure depends on the number of standards. For the same number of standards, experimental design techniques allow the selection of the most representative set of transfer samples, with a relevant improvement of the transfer.
b) there is a threshold number of standards for obtaining a transfer of good quality;
c) PLS gives a transfer better than PDS: only when PDS uses all the wavelenghts the performances of the two methods are similar;
d) PDS gives a smaller error than PLS only when the number of standards is below the threshold, but actually in this case both methods give poor results.

Another important advantage of PLS is that it allows the transfer of spectra also between instruments measuring a different number of variables, and a specific software is not required. A drawback is the higher computing time.

The use of PLS in two steps is also possible: first, to compute the relationship [1] between the block of spectral variables and the chemical variable (e.g. for instrument B), then to compute the relationship between two instruments [2]. By combining these equations, the calibration function of instrument A can be predicted.

The study with real data confirms that this transfer of calibration function can be performed with excellent results: in fact, the use of PLS in two steps allows a relevant reduction of the noise and, as a consequence, the predictive ability obtained with the transferred calibration function is better than the predictive ability of the coefficients computed directly.

REFERENCES

(1) J.S. Shenk, M.O. Westerhaus, *Crop Science*, **31**, 1694-1696 (1991).
(2) Y. Wang, D.J. Veltkamp, B.R. Kowalski, *Anal. Chem.*, **63**, 2750-2756 (1991)
(3) M. Forina, G. Drava, C. Armanino, R. Boggia, S. Lanteri, R. Leardi, P. Corti, P. Conti, R. Giangiacomo, C. Galliena, R. Bigoni, I. Quartari, C. Serra, D. Ferri, O. Leoni, L. Lazzeri, *Chem. Int. Lab.Sys.*, **27**, 189-203 (1995)
(4) E. Vigneau, M.F. Devaux, M. Safar, *J.Chemometrics*, **9**, 125-135 (1995)
(5) Y. Wang, B.R. Kowalski, *Appl. Spectr.*, **46**, 764-771 (1992).

BRISC (Biased Regression Imposing Smooth Coefficients)

Ildiko E. Frank
JerIl, Inc.
790 Esplanada, Stanford, CA 94305 USA

Summary: BRISC is a new biased regression method especially developed for spectral data analysis, where the predictors are not only highly correlated but also are in a predefined order. BRISC results in a regression coefficient curve that is much smoother than those from PLS, PCR, or ridge regression, and allows to pinpoint which part of the spectra is important in predicting the response.

Most regression methods are invariant to the order of the predictor variables, i.e. the predictors are regarded as a collection of measurements with no special order among them. The information that spectral data is a digitized form of a function (of wavelength), where the order of the predictor values is fixed, is totally lost in methods like PLS, PCR or ridge regression. Quite often the regression coefficient curve from PLS or PCR (scatter plot of regression coefficients against predictor ids or wavelengths) is quite rough with neighbouring values being very different. It is hard to interpret such a coefficient vector and identify important regions of the spectra.

BRISC is based on the assumption that spectral predictors close to each other on the wavelength scale are highly correlated, therefore they should have similar coefficients. This is a linear biased regression method, just like PLS, PCR or ridge regression, that results in a non-least squares solution penalizing lack of smoothness in the regression coefficient curve. Similar to PLS and PCR, BRISC calculates "composite variables" as linear combination of the original predictor variables. These composite variables, however, are quite simple. The loadings in each composite variable are either 0 or 1, i.e. the composite variables are sums of a subset of predictors. Such predictors must be contiguous neighboring predictors. Each predictor x_j, $j = 1, p$ loads into one and only one composite variable t_k, $k = 1, r$

$$t_k = \sum_{j=jst(k),jend(k)} x_j$$

with

$$jst(1) = 1 \qquad jend(r) = p \qquad jst(k+1) = jend(k) + 1$$

For example for $p = 20$ and $r = 3$ a possible set of loadings are:

```
1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1
```

The model is

$$y = a_o + \Sigma_k a_k t_k$$

where a_k the regression coefficients of the composite variables are estimated by least squares. The regression coefficients b_j, $j = 1, p$ for the original predictors can easily be calculated from a_k, $k = 1, r$ for the final model:

$$y = b_o + \Sigma_j b_j x_j$$

The key point of BRISC is to optimize the partition of the predictors. The algorithm starts with a single composite variable that is the sum of all predictors. In each step the number of composite variables is increased by one through making an additional optimal partition of the predictors. This means that one must examine p-1 potential partitions, i.e. calculate the corresponding regression coefficients and choose the partition that results in the best fit. Such exhaustive search seems to be expensive, but in reality, due to updating formulas, is quite feasible. Similar to PLS and PCR, the number of composite variables in the final model is optimized via cross-validation. After p-1 steps, p composite variables would be obtained, each consisting of a single predictor variable and the BRISC model would converge to the OLS solution.

Compared to PLS and PCR, the BRISC model is very simple and easy to interpret. Yet, its modeling (r^2) and predictive (cross-validated r^2) abilities are comparable and sometimes superior to those of PLS, PCR and ridge regression. The performance of BRISC is demonstrated and compared to other regression methods in several spectral examples.

REMOCOP The Regression Model Comparison Plot An Application to Spectral Data

Paul Geladi & Harald Martens

Dept. Organic Chemistry, Umeå University, S901 87 Umeå,
Sweden
and IDT, Ismaning, München, Germany

Summary : The REMOCOP plot is a visual aid for comparing regression models. An application to spectral data (many variables) is presented.

Regression and calibration play an important role in chemometrics. All analytical instrumentation is dependent on calibration by using some regression model. The Ordinary Least Squares method of building a Multivariate Linear Regression model has strict limitations. Therefore, biased or regularized regression models have been introduced. Important ones are Ridge Regression, Principal Component Regression and Partial Least Squares Regression. All the biased regression models are dependent on a parameter such as rank or ridge constant. This makes it possible to construct quite many models.

Comparing regression models has always been a desirable thing to do. This is perfectly possible when there is an experimental design in coded variables and with orthogonal (or almost orthogonal) factors. That is because the regression coefficients obtained are uncorrelated. When the regression coefficients become correlated, not much interpretation can be done. The classical thing to do when comparing regression models is to compare their predictive qualities by means of PRESS (Prediction Residual Sum of Squares) or SEP (Standard Error of Prediction) or one of the many variants of these. This is a univariate comparison of sizes and therefore it may not be satisfactory.

The method of REMOCOP was introduced in (1) and (2) and also presented at Chemometrics III in Brno, 1993. It is based on constructing a matrix of regression vectors and doing a principal component analysis of that matrix. The study of score plots then becomes a multivariate comparison of regression vectors. The principle was first tested on examples from multivariate image regression with many objects and few variables. It gives very good results. It allows the partition of the multivariate space of the regression vectors in volumes of excellent, good, satisfactory and unsatisfactory quality. It also allows the construction of totally new regression vectors of high quality that the biased regresion methods would not be able to obtain (1,2). Among those are models with a noninteger number of PCR or PLS components.

The presented example is from the book "Multivariate Calibration" by Martens & Næs (3). The raw data are unpublished. It is the example of the spectral analysis of Litmus when interferents and disturbances are present. It is the first use of REMOCOP for a case with many variables and fewer objects. It is also an example where more than one response variable is present. This gives regression coefficient matrices instead of vectors and makes the REMOCOP matrix a 3-way array of data. The example is presented and some conventional and newer ways of comparing the regression coefficient vectors from different models are introduced.

LITERATURE

1. Geladi P., Swerts J. & Lindgren F., "Multiwavelength microscopic image analysis of a piece of painted chinaware", Chemometrics and Intelligent Laboratory Systems, 24, 145-167, 1994.

2. Geladi P., "The regression model comparison plot (REMOCOP)", Proc. "Spectroscopy Across the Spectrum IV", Norwich, July 11-14, 1994, in press.

3. Martens H. & Næs T., "Multivariate Calibration", Wiley, Chichester, 1989.

Solving fundamental and applied problems in physical chemistry by means of chemometrics

Olav M. Kvalheim, Fred O. Libnau, Geir.M. Førland, H. Førdesdal, Egil Nodland, Alfred A. Christy, H. Høiland and Johan Sjöblom

Department of Chemistry, University of Bergen, N-5007 Bergen, Norway.

Abstract

Chemometric and statistic methodology have not been used much for problemsolving in physical chemistry. The reason for this is the strong tradition of trying to solve physical problems from first principles. If confronted with a complex problem, the physical chemist will typically try to reduce the complexity by constructing simple (often unrealistic simple) model systems and try to extrapolate from such model systems to the complex real-world system. This is opposite to the chemometric way of approaching a complex problem. The chemometrician would typically start out with the full complexity, and use experimental design and latent variable methodology to reduce the complexity and solve the problem.

In this lecture, the chemometric approach to problemsolving is illustrated on three different physical problems: i) stabilisation of water-in-oil emulsions, ii) determination of equilibrium constants in H_2O/D_2O mixtures and resolution of the HOD spectrum, and, iii) self-association of medium chain alcohols in n-decane solutions. The three examples illustrate the use of design and latent-variable methodology at different level of complexity.

FUEL SPILL IDENTIFICATION BY GAS CHROMATOGRAPHY/PATTERN RECOGNITION TECHNIQUES

Barry K. LAVINE

Clarkson University, 121 Science Center Potsdam NY 13699-5810, USA

Discriminants were developed by parametric and nonparametric pattern recognition procedures that correctly classified the gas chromatograms of neat jet fuels according to fuel type, and these discriminants were successfully used to classify gas chromatograms of jet fuels which had undergone weathering in a subsurface environment.

The potential of gas chromatography for correlating hydrocarbon spills to suspected fuel sources is recognized by many workers in the field of environmental chemistry. Typically, gas chromatograms of the fuel spill and a number of suspected sources are compared visually in order to obtain a best match. However, visual analysis of gas chromatograms can be subjective and usually cannot take into account the effects of weathering on the overall GC profile of the fuel. Therefore, evidence based on visual analysis of gas chromatograms is not always persuasive in a court of law, especially in cases involving an unweathered fuel identified as the source of a fuel spill because of the marked differences between gas chromatograms of weathered and unweathered jet fuels.

Due to the complexity of the mixture which constitutes a processed fuel, a systematic comparison of gas chromatograms is often necessary to ensure that differences in composition between various types of fuels are consistent, which is why pattern recognition methods offer a better approach to the problem of matching gas chromatograms of jet fuels than visual analysis. Pattern recognition methods can identify fingerprint patterns in the gas chromatographic (GC) data characteristic of fuel type even though the fuel samples in the training set have been subjected to a variety of conditions. Hence, classifiers can be developed from the GC data that are relatively insensitive to changes in the overall GC profile of the original fuel due to contamination, analytical error, or weathering. Furthermore, the discriminatory information that is sought in the chromatographic data often consists of subtle variations in relative peak intensities distributed across several peaks in the gas chromatograms. Pattern recognition methods are especially well suited for extracting this type of information from the large amounts of qualitative and quantitative data present in the gas chromatograms.

In this study pattern recognition methods have been used to classify the gas chromatograms of weathered and unweathered jet fuels. A data base of 302 gas chromatograms of neat jet fuel samples representing common jet fuels found in the United States was developed. Employing pattern recognition methods, the gas chromatograms of jet fuels that had undergone weathering in a subsurface environment have been correctly classified as to type using discriminants developed from the gas chromatograms of neat jet fuels. This approach has been taken because the physical

and chemical interaction of jet fuel components with the subsurface environment is not yet fully understood. The study described here is a logical extension of an earlier effort which emphasized the development of graphical and statistical pattern recognition methods for interpretation of GC profile data.

CHOICE OF OPTIMUM ANALYTICAL INSTRUMENTAL TECHNIQUE ON THE BASIS OF THE INFORMATION THEORY PARAMETERS

Mikuláš Matherny

Chair of Chemistry, Metallurgical Faculty, Technical University of Košice, SK-042 00 KOŠICE, Letná 9

Summary: The choice of optimized analytical technique was directed towards the whole set of atomic spectroscopy methods for direct analysis of natural powder materials

The analytical task was defined in the present case as follows: to determine optimum analytical methods for Ca- and Mg-carbonates and their sintered products with regard to the performance fulfillment parameters i.e. information content and efficiency as well as to the profitability as to the number of daily samples. The following methods could be taken into the consideration based on the scheme in Fig.1 : direct atomic emission spectrochemistry with arc excitation (AES), X-ray fluorescence spectroscopy (FXS), ICP-excitation and AAS spectroscopy with flame atomisation.

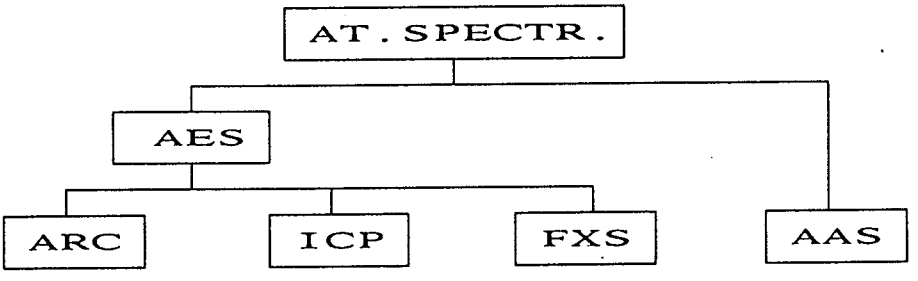


Figure 1. Scheme of the compared analytical methods

At the very beginning of evaluation of the analytical methods performance it is necessary to define concentration ranges with given guaranteed concentration determination precision for each analytical element in question. These values should be regarded as input tolerance values. The following tolerance values are discussed: $c(X_{min,T})$, $c(X_{max,T})$, and the required standard deviation of concentration determination $s(c_{X,T})$. These values should have identical dimensions. Similarly, it is necessary to determine the same parameters $c(X_{min})$, $c(X_{max})$ and $s(c_X)$ extracted from experimental data for all tested analytical methods and analytical elements (X). The most suitable way of determining these performance fulfillment parameters is to derive them directly from the calibration process [1]. In order to characterize the multielement analytical method using the information theory parameters [2], it is necessary to determine in the following sequence [3] the tolerance performance parameters and the

performance fulfillment parameters derived from experiments [4]. Both set of parameters are of identical character and therefore their calculation is identical. The information theory parameters are the following: the information content values of individual analytical elements ($I(p,p_0)$) and from these derived are the information content measure values ($MI(p,p_0)$), further, it is usually necessary to determine the element-specific efficiency values ($E(p,p_0)$) and the measure of information efficiency values ($ME(p,p_0)$) related to the experiments. The tolerance values of the precision of concentration determination and of the limit of detection have to be defined in advance for element-specific efficiency values determination since these values are limiting for the criteria of the analytical order fulfillment.

Further chemometric evaluation was directed towards the relevancy and/or redundancy of the analytical results evaluation. as a special case in chemometric evaluation, the consideration of the measure of accuracy should be regarded [4]. These sets of parameters make it possible to choose the optimum technique in agreement with the analytical order by mutual comparison of identical parameters.

REFERENCES:

1. Matherny M., Ondáš J.: Anal. Chim. Acta 1981, 133, 137
2. Matherny M., Eckschlager K.: Chem. Zvesti 1984, 38, 479
3. Matherny M.: Chem. Anal. (Warszawa) 1992, 37, 79
4. Matherny M., Eckschalger K.: Talanta 1994, 41, 113

A PERSONAL REMINISCENCE ON THE PREHISTORY OF CHEMOMETRICS

Eduard Plško
Donská 97, SK 841 06 Bratislava, Slovakia

Summary: Some practically unknown approaches to the statistical treament of analytical results performed before the formation of chemometrics are presented

Mathematical and statistical methods were widely applied to chemical measurements long before the formation of chemometrics enabled first in mid seventies by the exploitation of computers. The aim of this contribution is to show some, at the present time practically unknown examples of my approaches and results obtained in the fiell covered nowadays by chemometrics and published before its official formation as scientific branch. These attempts can so be in general understood as the prehistory of chemometrics.
Following topics will be dealt with:

1. Distribution of experimental data

Many observed deviations from the normal distribution of experimental results, if disregarded and evaluated according to the laws of Gaussian distribution can cause serious systematic errors. The majority of transformations used for the normalization of such populations has till now been of pure formal character without any attempt to elucidate and to take in account the real causes of the observed behaviour.
The determined significant skewness in the distribution of spectral lines intensities with the mode higher than the arithmetical mean was explained by similar influence of different parameters as e. g. the radial wandering of the arc or spark discharge (1, 2), random errors in repeated setting the wavelenght of the spectral line maximum in the optical axis of the measuring instrument at microphotometric measurements of photographically taken spectra (3) or at sequentially measured spectral line intensities using spectrometers with photoelectric detection (4).
The described assymetric distribution of spectral lines intensities was together with the application of internal reference element supposed as responsible for a lognormal-like distribution of spectroanalytical results (2). Later experiments (5) proved the indicated assumption. A detailed study of lognormal distribution enabled to estimate a parameter corresponding to standard deviation of the Gaussian distribution. The parameter is calculated using results gained from only twice repeated analyses performed on a set of real samples (6) and so it characterizes easier and better the precision and/or some other statistical features of a routine analytical procedure (7-9).

2. Non linear calibration

Analytical calibration fuctions in spectrochemical analysis are in general not linear and they can be formally described by higher parabole. This fact caused at the beginning of sixties several problems at the application of standard addition technique and determination of the unknown concentration by extrapolation. We contributed to the solution of the problem by the proposition of two different approaches (10). In the first case (11) more additions were used and the unknown concentration was calculated using graphically determined data. In the second case

(12) a convenient transformation of input data was proposed and the concentration was determined using a nomogram or a table. For to optimize the planning of corresponding experiments a statistical analysis enabling the determination of the relations of repeated measurements number ensuring the best precision was described (13). The last work represents so one of the first attempts performed with the aim to optimize an analytical procedure using statistical methods not only in Czechoslovakia but even in a larger scale.

3. Use of scatter diagrams in spectrochemical analysis

The evaluation of regression and correlation conditions presented graphically in scatter diagrams as applied to numerical expression of the homology of analytical and reference spectral lines by Strasheim and Holdt (1960) awoke a deep interest of many spectroscopists.

Taking in account the non linear dependence of spectral line intensity on concentration I performed a comprehensive study of regression and correlation conditions in scatter diagrams and proposed a more general formulation of spectral lines homology (14). Beside the known condition for homology of spectral lines expressed by the equality of their excitation energies, the equality of exponents in the description of their concentration dependence has to be considered too. In some cases the equality of excitation energy and the mentioned exponent relations ensures a sufficient homology.

A further analysis of scatter diagrams based on the study of the influence of experimental conditions on their parameters showed serious restrictions in their applicability (15), as well as conditions for a usefull application of reference element (16).

References:

1. Plško, E.: Appl. Spectroscopy, **16**, 55 (1962)
2. Plško, E.: Rozpravy Nár. techn. Musea, Praha, 1962, p. 23
3. Plško, E.: Acta Chim. Hung. **32**, 419 (1962)
4. Plško, E.: Revue Roumaine de Chimie, **10**, 695 (1965)
5. Plško, E., Kubová, J.: Chem. papers, **32**, 624 (1978)
6. Plško, E.: Kémiai Közlemények, **39**, 351 (1971)
7. Medveď, J., Plško, E., Cubínek, J.: Acta Geol. Geographica Univ. Comenianae, Bratislava, Geologica Nr. 27, 183 (1974)
8. Medveď, J., Kubová, J., Plško, E.: Geol. Carpatica **30**, 253 (1979)
9. Medveď, J., Plško, E.: Proc. Symposium on Petrogenesis and Geochem. of Geol. Processes, Veda, Bratislava, 1979, p. 297
10. Plško, E.: IX. Colloquium Spectr. Internationale, Lyon, 1961, Resumés, p. 127
11. Proks, I., Plško, E., Obert, T.: Chem. zvesti, **17**, 830 (1963)
12. Plško, E., Proks, I.: Acta Chim. Hung., **30**, 267 (1962)
13. Plško, E.: Acta Chim. Hung., **41**, 373 (1964)
14. Plško, E.: Collection Czechoslov. Chem. Commun., **30**, 1246 (1965)
15. Plško, E.: XVII. Colloquium Spectr. Internationale, Firenze, 1973, Preprints I, p. 38
16. Plško, E.: Conference of Chemists in Banská Štiavnica, Dom techn. SVTS, Žilina, 1971, p. 86.

Some Recent Experiences with Quality Matters in Analytical Chemistry

Wolfhard Wegscheider
General and Analytical Chemistry, University of Mining and Metallurgy
A-8700-Leoben/Austria

Strive for better quality as signified by the introduction of various quality managements systems is currently widespread among analytical chemistry laboratories and, for that matter, among testing laboratories at large. This is frequently motivated by changing trading patterns that consistently favor the global involvement and demand an improved acceptability of analytical results worldwide.

There is a certain degree of competition and parallelism of various QA systems, but it is generally felt that the one based on ISO Guide 25 is the most suitable for testing laboratories. As these systems are better and better exploited and developed there surfaced a critical shortage of internationally accepted standard reference materials that are needed to achieve traceability. It is therefore only consistent that ISO-REMCO is working to harmonized the requirements for reference material producers. Also, the potential substitute for traceability, comparability, needs some refinement that is defined through a recently issued WELAC-(EAL) Guide.

There is still serious concern about the proliferation of methods for estimating measurement uncertainty, partly because the analytical experts tend to refrain from making any such estimations and any quality manager may lack the technical background required for this estimation. Alternative strategies for the validation of data will also be required as CITAC moves towards the definition of guidelines for quality assurance of non-routine samples particularly in research laboratories. Some chemometric ideas supporting these directions will be discussed.

Lectures and Posters

APPLICATION OF PRINCIPAL COMPONENT ANALYSIS TO THE STUDY OF MAJOR CATIONS AND TRACE METALS IN FISH FROM TENERIFE (Canary Islands)

Astorga España M.S, Peña Méndez E.M, García Montelongo F.J.
Dpt. of Analytical Chemistry, Food Science and Toxicology.
University of La Laguna, 38071 - La Laguna, Tenerife, Spain.

The role of heavy metals play as pollutants is widely recognized. Specific studies on the distribution of metal pollution have shown a possible correlation between trace metals and the major cations (Na, K, Ca and Mg) concentrations in organisms. Accordingly, these studies on the concentration distribution of major cations in relation to heavy metals in specific fish species may ilustrate a possible use of fish as a source for the assessment of the extent of pollution in the marine environment, especially in coastal waters that are presently under threat with regard to industrial contamination of varied types.

In the present study we have been examined the application principal component analysis (PCA) with the aim to study the correlations between major cations and trace metals, the relationship between the metal contents and body weight and body length of fish, and intermetallic correlations.

Fish *Chromis limbatus* were collected in the south-western coast of the Tenerife during the period November 1993 - May 1994. Twenty four samples of muscle of this fish were analysed for Na, K, Fe, Mg, Cu, Cd, Fe and Zn using Atomic Absorption Spectrophotometry.

The statistic software package, Statgraphics System Plus 5.2 was used to the chemometric study of the data. After studying the correlation matrix for the original variables, PCA was used to polish the data set to remove outliers and to reveal any clustering among the data.

IDAS, A WINDOWS BASED SOFTWARE FOR CHEMOMETRICIANS.

I. Bondarenko, B. Treiger, R. Van Grieken, P. Van Espen
Micro- and Trace Analysis Center (MiTAC), Department of Chemistry, University of Antwerpen (UIA), Universiteitsplein 1, B-2610 Antwerpen, Belgium

New Windows-based software, encompassing cluster and factor analyses, is presented.

IDAS (Integrated Data Analysis System) is a chemometrics software that encompasses cluster (CA) and factor analyses (FA), two multivariate statistical techniques frequently used in applied sciences.

CA is implemented as a three-step procedure. Hierarchical CA is the first step. Its results serve as an initial partition for nonhierarchical CA. The appropriate number of clusters can be chosen with the help of three independent clustering criteria (consistent Akaike information criterion, Davies-Bouldin criterion and total error sum of squares criterion). Finally, the internal structure of the obtained clusters and relationships between them can be revealed with fuzzy clustering. New approaches to visualise the results of fuzzy clustering are realized. Such way of clustering allows to eliminate the shortcomings inherent to each of these three clustering techniques considered independently. There is also a possibility to use each of these clusterings separately.

The FA part of **IDAS** realizes both R- and Q-modes of principal FA, abstract factor rotation (varimax algorithm) and target transformation. Two modes of FA allow to reveal relationships not only between variables that characterise objects, but between objects themselves. Therefore Q-mode FA can serve as a complimentary means to CA.

IDAS contains also data editor which allows the user to read data file in ASCII format, to enter data from keyboard, to save data matrix in binary file for subsequent use in CA and/or FA programs and to make simple editing of data matrix.

The results of multivariate data analysis can be saved in the form suitable for any text processor.

Context-sensitive as well as on-line help is available.

DETECTION OF INHOMOGENEITIES IN NIR DATA SETS

V. Centner, D. L. Massart
ChemoAC, Vrije Universiteit Brussel,
Laarbeeklaan 103, B-1090 Brussel, Belgium

Summary: *A methods to detect outliers and clusters in NIR data sets are proposed. Replicates and samples as such have been investigated before modelling and have been verified after. A tendency to cluster was evaluated by Hopkins statistic.*

The quality inspection of a data set should be the first step preceding any multivariate calibration. Since the outliers and the presence of subgroups or clusters can have a profound influence on the model it is useful to detect such inhomogeneities. Outlying replicates are usually removed. The decision about samples outlying for instance dué to different chemical characteristics is less evident. One can verify detected objects during the modelling stage and reconsider reasons for removing them. When subgroups or clusters occur one should decide whether to model all objects together or rather to develop models for each subgroup separately. The decision depends on the goal and the context of the work. It should be based more on chemical than on statistical grounds.

ISO recommends a Cochran test in order to detect outlying replicate measurements. This test is based on comparing the variance of the sample that shows the highest variance among replicates s_{max} and the sum of variances of all samples

$$C = s_{max}^2 / \sum_{i=1}^{n} s_i^2 \qquad \text{for i=1, ...n, where } \underline{n} \text{ is number of samples} \qquad (1)$$

If $C > C_{crit}$, the sample contains replicate outlier(s).
In univariate data the outlying samples are detected by Dixon and Grubbs' tests. The Dixon test measures the differences between the extreme (y_{test}) and the measurement nearest ($y_{nearest}$) to it with the range of all measurements ($y_{max} - y_{min}$)

$$Q = |y_{test} - y_{nearest}| / (y_{max} - y_{min}) \qquad (2)$$

The single outlier Grubbs test (G-ISO) compares the largest absolute value of the normalized deviation z with the critical value T

$$z = (y_{test} - \bar{y}) / (\sum_{i=1}^{n} (y_i - \bar{y})^2 / (n - 1))^{1/2} \qquad \text{for i=1, ...n} \qquad (3)$$

Grubbs sum of squares criterion (G-SS) compares the ratio

$$G = SS_1 / SS \qquad (4)$$

between the sums of squares without (SS_1) and with suspected outlier (SS) to the critical value F_{crit} for a modified F-test. NIR data are multivariate therefore the detection was performed on new latent variables (PCs) obtained by the Singular value decomposition procedure (**X = S*V*D**) and on Rao's statistic $D^2_{(k)}$

$$D^2_{(k)}(y_i) = \sum_{j=k+1}^{p} u_{ij}^2 \qquad (5)$$

The un-normed scores (u) are elements of the product matrix **S*V.** They are summed from the (k+1)-th to the p-th PC for each object (i). Since the data distribution of Rao statistic is skewed the applied G-ISO test was modified.

The index for clustering tendency, called the Hopkins statistic, examines whether objects in a data set differ significantly from the assumption that they are uniformly distributed in multidimensional PC space (null hypothesis, H_0). The Euclidean distance from an object to its nearest neighbour (W) and the distance from an artificial object, spread uniformly over the data space to the nearest real object (U) are measured. This procedure is iterated several times (e).

$$H = \sum_{e=1}^{N} U_e / \left(\sum_{e=1}^{N} U_e + \sum_{e=1}^{N} W_e\right) \qquad (6)$$

If H is greater than 0.75, there is more than 90% confidence in rejecting the null hypothesis.

The methods described above were applied to two data sets of NIR spectra of polyether polyols. The object of the calibration was to determine the hydroxyl number of the samples.

* The Cochran test applied to replicate corresponding sum of absorbances A_{sum}, PCs and A(PC) identified three outlying replicates. The recommendation to use A_{sum} could be pointed out because it covers all attributes of the multivariate data.

* On the individual latent variables (PCs) and their sums of squares (Rao statistic) were detected mean spectra outliers. The analysis supported by visual inspection of the graphs of the PC_j (D-value) plotted against the y-value is proposed.

Two typical outliers were detected and eliminated because of the chemical evidence. Other possible outliers, flagged after the data inspection, were verified by the modelling. This allowed us to distinguished one "good" leverage object and one real outlier.

* Another source of heterogeneity in the data, the tendency to create clusters, was quantified by the Hopkins statistic (H). The results obtained agree with the visual analysis. One data set was considered heterogeneous, another one homogeneous.

USE OF STECTRAL MAPPING TECHNIQUE FOR THE EVALUATION OF THE FORMATION OF INCLUSION COMPLEXES BETWEEN ANTICANCER DRUGS AND CYCLODEXTRIN DERIVATIVES

Tibor Cserháti and Esther Forgács
Central Research Institute for Chemistry, Hungarian Academy of Sciences, P.O.Box 17, 1525 Budapest, Hungary

Summary: spectral mapping technique has been used for the separation of the streng and selectivity of the relative strength of interaction between 18 commercial anticancer drugs and 7 cyclodextrin derivatives. It has been established that both the diameter of the cyclodextrin cavity and the number and type of substituents on outer sphere of the cyclodextrin molecules influence the complex forming capacity.

The relative strength of interaction between 18 commercial anticancer drugs and 7 cyclodextrin (CD) derivatives was determined by charge-transfer reversed-phase thin-layer chromatography. The anticancer drugs were Ftorafur = N-(2-furanidyl)-5-fluorouracil; Bicnu = N,N-bis(2-chloroethyl)-N-nitrosourea; Leukeran = 4-[bis (2- chloroethyl)amino]benzenebutanic acid; Vincristin 22-oxo-(3α,14ß,16α)-14, 15-dihydro-14-hydroxyeburnamenine-14-carbocyclic acid methyl ester; Vinblastine = (3α,14ß,16α)-14,15-dihydro-14-hydroxyeburnamenine-14-carbocyclic acid methyl ester; Vumon = 4'-O-demethyl-1-O(4,6-O-2-thenylidene-ß-D-glucopyranosyl)epipodophyllotoxin; Provera = 17-α-acetoxy-6-α-(methyl) progesterone; Bleogin = N^1-[3-dimethyl(sulfonio)propyl]bleomycin amide; Zitazonium = 2-[4-(2-chloro-1,2-difenilethynile)fenoxi]-N,N-diethyl-ethamine-citrate; Farmorubicin = (8S-cis)-10-[(3-amino-2,3,6-trideoxy-α-L-arabino-hexopyranosyl)oxy]-7,8,9,10-tetrahydro-6,8,11-trihydroxy-8-(hydroxyacetyl)-1-methoxy-5,12-naphthacenedione; Adriblastine = 10-[3-(amino-2,3,6-trideoxy-α-L-hexapyranonesyl)oxy]-7,8,9-tetrahydro-6,8,11-trihydroxy-8-(hydroxyacetyl)-1-methoxy-5-12-naphthacenedione; Natulan = N-(1-methylethyl)4-[(2-methylhydrazino)methyl]-benzamide; Mitomycin C = [1-aR]-6-amino-8-[(aminocarbonyl)oxymethyl]-1,1a,2,8,8a,8b-hexahydro-8a-methoxy-5-methyl-azirino-[2',3':3,4]pyrrolo[1,1a]indole-4,7-dione; Cytoxan = 2-[bis(2-chloroethyl)amino]tetrahydro-2H-1,3,2-oxazaphosphorine 2-oxide monohydrate; Estracyt = Estra-1,3,5-(10)-triene-3,17-diol-3-[bis-chloroethyl)carbamate; Deticene = 5-(3,3 -dimethyl-1-triazenyl)-1-H-imidazole-4-carboxamide; Metothrexate = 2,4,-diamino-10-methyl-pteroylglutamic acid, and Taxol =[2aR-[2aα,4ß,4aß,6ß,9α(aR*,ßS*),11α,-12α,12aα,12bα]]-ß-(Benzoylamino)-α-hydroxybenzenepropanoic acid 6,12b-bis(acetyoxy)-12-(benzoyloxy)-2a,3,4,4a, 5,6,9,10,11,12,12a,12b-dodecahydro-4,11-dihydroxy-4a,8,13,-13 tetramethyl-5-oxo-7,11-methano-14-cyclodeca[3,4]benz[1,2-b]oxet-9-yl ester. The cyclodextrins and cyclodextrin derivatives were α- and τ-cyclodextrin (α-CD and τ-CD), hydroxypropyl ß-CD (2.7 hydroxyproyl group per glucose unit), acetyl ß-CD (1.2 acetyl group per unit), carboxymethyl-τ-CD (3.2 carboxymethyl group), carboxymethyl α-CD (2.5 carboxymethyl group), and carboxymethyl ß-CD (3.6 carboxymethyl group). The original data matrix consisted of the values of the relative strength of interaction determined by reveresed-phase chromatography, the cyclodextrins and anticancer drugs being the variables and the observations, respectively. The multidimensional spectral map was visualized by the two-dimen-

sional nonlinear mapping technique. The potency values of the spectral map was considered as indicators of the capacity of anticancer drugs and cyclodextrins to inetract with each other taking into consideration simultaneously each data. In order to determine the capacity of CDs to form inclusion complexes with the anticancer drugs the same calculations were carried out on the transverse of the original data matrix. The calculation of the two-dimensional nonlinear map of spectral characterisitcs was carried out to the point when the difference between the last two iterations was lower than 10^{-8}. Anticancer drugs showed high differences in their capacity to interact with cyclodextrins, the order of the overall interaction strength was: Estracyt (6.43); Leukeran (6.23); Zitazonium (5.90); Bleogin (5.16); Vincristin (4.91); Deticene (4.57); Natulan (4.35); Methothrexate (4.16); Cytoxan (4.06); Mitomycin C (3.45); Vumon (2.72); Farmorubicin (2.27); Ftorafur (2.20); Adriblastine (2.15); Taxol (2.15); Bicnu (2.11); Vinblastine (0.67), and Provera (0.57). No significant correlation was found between the capacity of anticancer drugs to form inclusion complexes with cyclodextrin derivatives and their common physicochemical parameters indicating that complexity of forces involved in the formation of inclusion complexes. The anticancer drugs did not form separate clusters on the two-dimensional nonlinear map of spectral characteristics suggesting again that more than one substructure of the drugs participate in the formation of drug-cyclodextrin complexes. The capacity of cyclodextrins and cclodextrin derivatives to form inclusion complexes with the anticancer drugs also differed considerably: acetyl-ß-CD (8.92); carboxymethyl-ß-CD (8.67); hydroxypropyl-ß-CD (6.30); carboxymethyl-τ-CD (5.73); carboxymethyl-α-CD (4.08); τ-CD (4.05) and α-CD (2.21). The data clearly show that ß-CDs are the most appropriate complexing agents for these nonhomologous series of anticancer drugs, followed by τ- and α-CDs. The low capacity of α-CDs to form inclusion complexes can be tentatively explained by the supposition that the bulky drugs are sterically inhibited to enter the small cavity of α-CD. In each instance substituted derivatives formed stronger complexes with the drugs than the non-substituted CDs did. This phenomenon may be due to the formation of additional interactive forces between the substituents on the outer sphere of the CD ring and the polar substructures of drugs pointing outward from the hydrophobic CD cavity. The higher complex forming capacity of acetyl-ß-CD lends support to our previous qualitative conclusions. CDs and CD derivatives formed distinct clusters on the two-dimensional nonlinear map of spectral characteristics. The clusters of α-, ß- and τ-CDs are well separated from each other on the map, the cluster of τ-CDs being in the center. This result indicates that the selectivity of the interaction of cyclodextrin derivatives with anticancer drugs mainly depends on the sterical dimensions of the CD cavity and the various substituents only modify the character of the drug-CD interaction. It can be concluded from the results of calculations that anticancer drugs can interact with CDs and with CD derivatives. The interaction involves the inclusion of the drug molecule or its hydrophobic substructures into the CD cavity. This phenomenon is governed by the steric correspondence of the interacting molecules and the main interactive forces are hydrophobic ones. The polar interactions (probably hydrogen bond formation) between the various substituents on the CD ring and the substructures of drugs not included in the CD cavity may only modify the character of the inclusion complexes.

COMPARISON OF RIDGE AND PRINCIPAL COMPONENTS REGRESSION FOR NIR SPECTROSCOPY.

M. F. Devaux[1] E. Vigneau[2] E. M. Qannari[2] P. Robert[1]

1. INRA - L.T.A.N.- BP 1627 - 44316 Nantes Cedex 03 - France
2. ENITIAA - Laboratoire de Statistique - Rue de la Géraudière 44072 Nantes Cedex 03- France

Summary : Ridge and Principal components regression were compared for NIR calibration on the basis of 50 random calibration sets taken from 99 samples. A combination of the two methods was investigated.

Introduction.

Near-infrared (NIR) analysis for rapid determination of the chemical composition of food and feed products first requires to build calibration equations between the spectral data and the chemical information. Spectral data are characterised by a high collinearity among the wavelengths. Consequently, Ordinary Least Squares (OLS) is not relevant to assess the prediction equation, since the estimations of the regression coefficients are known to be very unstable. One of the most usual alternative in chemometrics is Principal Components Regression (PCR). Ridge Regression (RR) is another possible method introduced to deal with multicollinearity problem.

The principle used in PCR is to substitute the original wavelengths by their principal components and to select only the first of them associated with the largest eigenvalues. In this method, some principal components which could be useful for the prediction of unknown samples are discarded on the basis of the calibration set. In RR, all the principal components are taken into account and a constant k is added to all eigenvalues in order to reduce the unstability engendered by the smallest ones.

In the present work, 50 repetitions of NIR calibrations are achieved in order to study the stability and efficiency of RR and PCR.

Material and methods.

The methods were applied to predict the durability of pellets in feedstuff industry from NIR spectra of mixtures of raw materials to be pelleted. $n=99$ samples were available and separated into calibration (50 samples) and validation sets (49 samples). The choice of the samples of the calibration set was realised at random 50 times.

PCR and RR were applied on each of the 50 calibration sets. In PCR, the number c of principal components included in the prediction equation was chosen according to the cross-validation error of prediction (PRESS criterion). The parameter k in the case of RR was determined from a procedure suggested by Hoerl, Kennard and Baldwin (1975). For $c=1$ to $(n-3)$, k_c values were computed :

$$k_c = \frac{c\,\hat{\sigma}_c^2}{\hat{\beta}_c'\,\hat{\beta}_c}$$

where c is the number of principal components introduced in the calibration model, $\hat{\beta}_c$ is the estimation of the vector of coefficients and $\hat{\sigma}_c$ the estimation of the residual error. The minimum of the k_c values was chosen as parameter k in order to introduce a bias as small as possible.

For each method and for the 50 sets, the quality of prediction was evaluated by two criteria. The PRESS criterion was used to estimate the prediction error from the calibration set and the Squared Error of Prediction (SEP) was measured for the validation samples.

Results and discussion.

The number c of principal components selected in the model varied significantly for the 50 repetitions of calibration. $c=13$ and $c=10$ were chosen in 24% and 18% of the cases, respectively. The minimum was observed for $c=7$ and the maximum for $c=43$. The choice of the calibration set therefore appeared of prime importance. In the case of RR, the parameter k varied from 1.2E-7 to 1.7E-5 depending on the calibration set.

The PRESS values for RR were, on average, 32% smaller than the values for the OLS method. This improvement was also observed for the validation sets for which the SEP values were 37% smaller. The comparison of RR and PCR indicated that the PRESS values, evaluated for the calibration sets, were slightly larger for RR in most cases. On the contrary, the SEP values of the validation sets were 9% smaller, on average, for RR than for PCR and the prediction was improved 39 times out of 50.

Consequently, RR appeared to be an interesting alternative to PCR. Firstly, the RR predictions were at least as accurate as the PCR ones. Secondly, no principal components were discarded and the parameter k was automatically determined by calculation rather than by a cross-validation procedure.

A combination of RR and PCR was tested in order to evaluate the effect on the RR model of the last principal components which are usually interpreted as noise. In this case, the regression model was built by using to the two parameters c and k. It was observed that the PRESS values did not depend on c when c was sufficiently large.

An application of RR can be suggested by taking a larger number of principal components than in PCR and introducing the parameter k in order to reduce the unstability generated by the last components included.

Reference :

Frank I. E. and Friedman J. H. (1993). A Statistical View of Some Chemometrics Regression Tools. *Technometrics*, 35 (2), 109-148.

Hoerl A. E., Kennard R. W. and Baldwin K. F. (1975). Ridge regression : some simulations. *Comm. Statist.*, 4 (2), 105-123.

INTERACTION OF FACTORS INFLUENCING CELL PROLIFERATION: COMPARISON OF POLYNOMIAL REGRESSION AND ISOBOLOGRAPHIC METHODS.

Ladislav Dušek[1], Alois Kozubík[2], Jiřina Hofmanová[2]

[1] Department of Environmental Studies, Faculty of Science MU
Kotlářská 2, 611 37 Brno, Czech Republic

[2] Institute of Biophysics, Academy of Sciences of the Czech Republic
Královopolská 135, 612 65 Brno, Czech Republic

Summary: Both polynomial regression and isobolographic analysis proved that inhibitors of arachidonic acid metabolism enhance the effect of TGF-β1 on cell proliferation synergistically. Furthermore, regression analysis allowed to model quantitative effect of each combination of factors.

The evaluated experiments were designed as complete factorial combination of various levels of following factors: TGF-β1 (transforming growth factor), Esculetin (ESCUL;) and ETYA (competitive inhibitor of arachidonic acid). Influence of regulation factors on cell proliferation (measured as level of ^{3}H-thymidine incorporation) was studied using mink lung epithelial cell line CCL 64 that represents prototype of cells with extremely high sensitivity to TGF-β1.

The results of experiments were analyzed using standard two - way ANOVA (fixed model) after preliminary tests of normality and homogeneity of variance. The ANOVA models were used to assess the significance of fixed effects of single factors and their interaction as well. The null hypothesis for an interaction between two combined factors was that the response of ^{3}H-thymidine incorporation was not differing among specific levels of one factor depending upon the particular level of the second factor.

Experiment I tested the combinations of TGF-β1 with ESCUL and experiment II realized the combinations of TGF-β1 with ETYA. The dose response curve of TGF-β1 was very steep in comparison with ETYA and ESCUL. Wide range of concentrations of inhibitors of arachidonic acid metabolism decreasing the cell response in the range from 10 to 50 % (ESCUL) and from 30 % to 70 % (ETYA) were combined with relatively less efficient concentrations of TGF-β1: 0.125 pM - threshold dose, 0.25 pM - 5 % of inhibition and 0.5 pM - 20 % of inhibition of ^{3}H-thymidine incorporation.

For both experiments, stepwise multiple regression and regression on all possible subsets of independent parameters identified equations including the best variables for the cubic polynomial model of the data. Both combination of TGF-β1 with ESCUL (Exper. I) and TGF-β1 with ETYA (Exper. II) caused highly significant ($P = 0.001$) overadditive inhibition of ^{3}H-thymidine incorporation. This interactive effects led to the incorporation of interactive components in the regression models:

$$Y = a_0 + a_1(TGF) + a_2(TGF^2) + a_3(ESCUL) + a_4(ESCUL^2) + a_3(TGF)(ESCUL) + a_4(TGF)(ESCUL^2) + a_5(TGF)(ESCUL^3)$$

$$Y = a_0 + a_1(TGF) + a_2(TGF^2) + a_3(ETYA) + a_4(ETYA^2) + a_3(ETYA^3) + a_4(TGF)(ETYA) + a_5(TGF^2)(ETYA)$$

Response surfaces were modelled by fitting untransformed data to polynomials. The evaluation of regression models was based on the correlation between observed and predicted ^{3}H-thymidine incorporation levels, the lack-of-fit statistic (F-test), and the significance of estimated parameters.

Contribution of additive and interactive part of the constructed regression models to the total inhibition effect was analyzed separately with respect to concentration levels of the combined factors. This approach allowed to quantify synergistic potentiation of TGF-β1 by ESCUL and ETYA as well.

The effect of ESCUL significantly increased the steepness of curve describing effect of TGF-β1 on ^{3}H-thymidine incorporation even in the range of relatively low concentrations. Derived from tested range of concentrations, it appeared that quantitative portion of interactive effect in the total inhibition effect decreased with increasing dose of ESCUL.

The application of each tested dose of ETYA had significant enhancing effect on inhibition efficiency of each dose of TGF-β1. The quantitative portion of interactive effect in the total inhibition effect increased directly with increasing dose of ETYA within single dose of TGF-β1 and there was no significant difference between various doses of TGF-β1 within single dose of ETYA. Overadditive inhibition effect of TGF-β1 x ETYA combinations was less quantitatively apparent than in combinations with ESCUL, mainly due to relatively high total inhibition of cell response caused by higher doses of ETYA alone (more than 60 % of inhibition).

Isobolographic analyses proved the results from response surface models and provided very useful tool for graphical presentation of the data.

APPLICATION OF MULTIVARIATE STATISTICAL METHODS FOR THE EVALUATION OF THE HPLC RETENTION BEHAVIOUR OF SOME PHENOL DERIVATIVES

Esther Forgács and Tibor Cserháti
Central Research Institute for Chemistry, Hungarian Academy of Sciences, P.O.Box 17, 1525 Budapest, Hungary

Summary: The retention characteristics of polyethylene(PEE)- and octadecylsilica(ODS)-coated HPLC columns were compared using ring substituted phenol derivatives as solutes. The retention data matrix was evaluated by stepwise regression analysis and by principal component analysis followed by two-dimensional nonlinear mapping. Calculations proved that the retention behaviour of PEE column sligthly differs from that ODS column and the discrepancy may due to the steric and electronic interactions occurring between the polar substructures of phenol derivatives and the free silanol groups not-covered by polyethylene.

The retention of 19 ring-substituted phenol derivatives was determined on polyethylene-coated (PEE) and octadecylsilica (ODS) HPLC columns. The eluents were mixtures of methanol - 25mM K_2HPO_4 (methanol concentration 35 - 70 vol%) and methanol - water (methanol concentration 55 - 80 vol%). Linear correlations were applied to describe the dependence of logk' value on the concentration of methanol in the eluent (C vol%):

$$\log k' = \log k'_0 + b.C$$

where $\log k'_0$ and b are retention parameters characterising the retention capacity and the specific hydrophobic surface area of the solutes, respectively.

and the retention data were evaluated by multivariate mathematical statistical methods to find the relationship between the retention characteristics and physico-chemical parameters of phenol derivatives. Stepwise regression analysis was applied to find the molecular characteristics of phenol derivatives influencing significantly their retention on PEE column. The molecular characteristics included in the calculation were: π = Hansch - Fujita's substituent constant characterizing hydrophobicity; H - Ac and H - Do = indicator variables for proton acceptor and proton donor properties, respectively; M - RE = molar refractivity; F and R = Swain - Lupton's electronic parameters characterizing the inductive and resonance effect, respectively; σ = Hammett's constant, characterizing the electron-withdrawing power of the substituent; Es = Taft's constant, characterizing steric effects of the substituent; B_1 and B_4 = Sterimol width parameters determined by the distance of substituents at their maximum point perpendicular to attachment. Principal component analysis (PCA) was used to find the similarities and dissimilarities between the retention characteristics of PEE and ODS columns as well as the physicochemical parameters of phenol derivatives. The parameters of eqn. 1 both for PEE and ODS columns, the various physicochemical parameters of phenol derivatives listed above were the variables and the phenol derivatives were the observations. In each instance the relationship between log k' value and concentration of methanol in the eluent was significantly linear indicating that the retention behaviour of phenol derivatives follows the general rule also on PEE column. Stepwise regression analysis found significant linear correlation between the retention parameters (slope and intercept values of

eqn. 1. in both eluent systems) and physicochemical parameters of phenol derivatives.

Methanol-25mM K_2HPO_4 eluent systems:

$\log k'_0(PEE) = 0.23 - 0.61(\pm 0.15).B_1$

$n = 15 \qquad r = 0.7386 \qquad r_{99\%} = 0.6614$

$b(PEE) = -2.52.10^{-2} + 4.48.10^{-2}(\pm 1.90.10^{-3}).Es$

$n = 15 \qquad r = 0.5288 \qquad r_{95\%} = 0.5139$

Methanol-water eluent systems:

$\log k'_0(PEE) = 2.46 - 2.25(\pm 9.10^{-3}).\sigma$

$n = 15 \qquad r = 0.7116 \qquad r_{99\%} = 0.6614$

$b(PEE) = -4.17.10^{-2} + 2.52.10^{-2}(\pm 9.10^{-3}).\sigma$

$n = 15 \qquad r = 0.6108 \qquad r_{95\%} = 0.5923$

The results of stepwise regression analysis indicates that the retention of ring-substituted phenol derivatives on PEE column significantly depends on the steric and electronic parameters of phenol derivatives while the influence of molecular lipophilicity is negligible. These results suggest that polar interactions between the hydrophilic substructures of solutes and the silanol groups of silica support not covered by the polyethylene play a considerable role in the retention. The results of PCA proved that four principal components (background variables) contain the overwhelming majority of the information (89.09%) of the 15 chromatographic and physicochemical parameters. Each chromatographic parameter has great loadings in the same (first and second) principal components indicating the similarity between the retention behaviour of ODS and PEE columns. The specific contact surface areas of phenols determined on PEE column form a cluster with the electronic parameters on the two-dimensional nonlinear map of PC loadings. This results indicates again the occurrence of polar interactions between the dissociable substructures of phenol derivatives and the surface of polyethylene-coated silica support. A distinct cluster contains the intercept values determined on ODS and PEE columns, the lipophilicity and steric parameters of phenol derivatives. This result suggests that the lipophilicity and steric parameters of phenol derivatives has an marked effect on the retention capacity of both columns. Phenol derivatives do not form clusters neither according to the nature of the substituents nor according to the substituent position on the non-linear map of PCA variables. This finding indicates that retention behaviour of phenols is influenced in similar degree by the quality and position of substituents.

ELIMINATION OF SPECTRAL AND/OR OTHER INTERFERENCES IN ICP-AES USING MULTIVARIATE CALIBRATION APPROACH

J. Havel, D. Kalný

Department of Analytical Chemistry, Masaryk University, Kotlářská 2, 611 37 Brno, CZECH REPUBLIC

Summary: *All difficulties and tedious procedures to eliminate spectral and/or nonspectral interferences can be easily overcome applying multivariate calibration approach with Partial Least Squares evaluation method.*

Classical approach for correction of spectral and other interferences in ICP-AES and that one of multivariate calibration method using Partial Least Squares (PLS) evaluation method have been once more compared and studied in details. Application of PLS method has been already proposed some time ago[1].
The classical correction method is based on computation of correction factors a_{ij} from line intensities according to the expression:

$$a_{ij} = (I_{ij}/c_{ij})(I_j/c_j)^{-1}$$

where I_{ij} is the intensity of the interfering line of the i-th interferent at the wavelength corresponding to the maximum of the analytical line of the j-th analyte, I_j is the intensity of the j-th analyte, and c_{ij}, c_j are corresponding concentrations of interferents and analytes in calibration solutions. Such calibration matrix obtained is then used for the prediction of concentrations of determined elements in unknown samples from measured raw intensities. This approach does not require any selection of interference-free lines, any computation of correction factors, and no background corrections.The concentration of the each analyte is computed from the equation:

$$c_A = c_M - \sum_{i=1}^{n} a_{ij} c_i$$

where c_A is the concentration of the analyte, c_M the concentration of the analyte computed from the raw intensity measured at the analytical line, and a_{ij}, c_i are corresponding correction factors and the concentrations of the interferents in

solution to be analyzed, respectively. The set of such equations (one equation per element) is solved by the Gauss's elimination method.

On the other hand, multivariate calibration method (PLS) requires only that two-dimensional matrix of the intensities is measured for arbitrary lines of determined elements, for a set of calibration solutions. The calibration matrix is constructed using appropriate experimental design method applying full or reduced factorial design.

All measurements have been made on UNICAM PU 7000 instrument with a free-running 40.68 MHz generator and spectrometer with an echelle grating. Model mixtures of rare earth elements have been analyzed, because these elements have rich spectra and interferences are frequent. Nonspectral interferences have been modelled adding calcium to analyzed solutions. The demandingness of both correction approaches has been compared to each other and results obtained have been statistically evaluated.

References:

1. Havel J., Moreno C., Hrdlička A., Valiente M., 2nd Scand. Symp. on Chemometrics, 28-31 May, Bergen, Norway, 1991.

Enhanced Additivity Model of Parameter $log(L^{16})$

Pavel Havelec and Jiří G.K. Ševčík*

Department of Analytical Chemistry, Charles University, Albertov 2030, CZ-128 40 Prague 2, The Czech Republic
*ALS - Analytical Laboratory Systems GmbH, P.O.Box 1225, D-92602 Weiden, Germany

Summary: *An additivity model of apolar solute-solvent parameter $log(L^{16})$ was enhanced and tested for large sets of nonaromatic (more than 900) and aromatic (more than 800) compounds.*

The solute-solvent interactions in gas chromatography have been discussed in many papers over the last thirty years. The most recent works have been published by Abraham, Carr and Poole and their coworkers; the autors assume the linear solvation energy relationships. The principal solvation equation can be written in the form,

$$log(SP) = l*log(L^{16}) + r*R_2 + s*\pi_2^H + a*\alpha_2^H + b*\beta_2^H + SP_0 \quad (1)$$

The dependent variable SP can be any retention characteristic (e.g., gas-liquid partition coefficient K, specific volume V_g, adjusted retention time t_r'or volume V_r', etc.). The independent variables $log(L^{16})$, R_2, π_2^H, α_2^H and β_2^H are parameters representing the properties of the individual solutes.

The regression coefficients l, r, s, a, b and SP_0 in this equation characterize the stationary phases, in other words, they specify the ability of a phase to interact with a solute in a certain way. In equation 1, Carr replaces the product $r*R_2$ with $d*\delta_2$.

The solute descriptor $log(L^{16})$ is the decadic logarithm of the partition coefficient for transfer of the solute from the gas phase to n-hexadecane at 298 K. The product $l*log(L^{16})$ is a combination of an endoergic cavity term with an exoergic general dispersion interaction term. The regression coefficient l reflects the stationary phase ability to separate adjacent members of a homologous series and exhibits a sensible correlation with the partial Gibbs free energy of solvation of a methylene group into a stationary phase.

The parameter $log(L^{16})$ can be obtained experimentally, or can be back-calculated from Abraham's regression equation and/or using the additivity method.

In our previous work, we proposed an additivity model for prediction of this parameter. For nonaromatic compounds it holds that

$$\log\left(L^{16}(X)\right) = \sum_{i=1} l_i \times FG_i + \sum_{j=1} m_j \times SC_j + \sum_{k=1} n_k \times IC_k \qquad (2)$$

where

FG, SC, IC ....the number of particular groups (*FG*) forming compound *X*, structural contributions (*SC*), interactional contributions (*IC*),

i, j, k ....the identification number of a group (*i*), of a structural contribution (*j*), of an interactional contribution (*k*),

l, m, n ..the regression coefficients for the contributions of particular groups (*l*), structural contribut-ions (*m*),interactional contributions (*n*).

In this work, equation 2 was extended to involve additional functional contributions and tested by various statistical methods. The methods applied confirm statistical significance of equation 2 and of the given contributions.

The additivity model enables a fast and reliable estimation of $log(L^{16})$ for any compound and will find use in biochemical, clinical, pharmaceutical applications.

New Software Developments for Titrations

The Programs MINI_T 5.0, ITERAX and T_ALPHA

Prof. Dr. G. Hägele, Dipl.-Chem. B. Drenker

HEINRICH-HEINE-Universität Düsseldorf
Institut für Anorganische Chemie und Strukturchemie I
Universitätsstr. 1, D-40225 Düsseldorf, Germany

Summary: We describe some of our new PC-programs to control and interprete titrations. MINI_T 5.0 is a computer titration software for most applications. ITERAX simulates and iterates titration curves. It allows the determination of electrode parameters (Blank-titration), pK-values and complex stability constants. T_ALPHA communicates with the new TitroLine alpha computer titrator (SCHOTT-GERÄTE GmbH) and allows PC-control under MS-WINDOWS 3.1.

MINI_T 5.0

This computer titration software (MS-DOS) controls several equipment from SCHOTT-GERÄTE GmbH, Hofheim, Germany. We use one piston burette T 110 or T 200 and a microprocessor pH-meter CG 841. This combination enables a low budget system for most applications, f. e. mV- or pH-titrations with linear or dynamic volume increments. In our research group it is used to measure the experimental data for potentiometric determination of pK-values and complex stability constants.

In this new version (MINI_T 5.0 M) an additional photo sensor was implemented by using a METTLER TOLEDO two channel pH/ionmeter Delta 355. It enables simultaneous transmission measurements in potentiometric titrations to determine precipitations. This is important for later simulations of the titration curve, because the simulation only works in a homogeneous aquatic system.

A new conductometric titration system is realized in another version (MINI_T 5.0 R). Here a RADIOMETER ABU 93 / 2 piston triburette with 2 electrode channels is combined with a RADIOMETER CDM 92 conductivity meter and a PC. It enables potentiometric and conductometric titrations with up to 30 equivalence points which are automatically calculated.

ITERAX

This PC-program (MS-DOS) was created for simulation and iteration of potentiometric titration curves. It is equipped with a comfortable menu system which enables a very simple use of the program. Up to 8 components or 20 species can be calculated. Every data can be saved, f. e. calculated titration curves, molar

fraction data sets (x - pH, x - ml). The program is also useful for exact electrode calibration by „Blank-titrations". Therefore the 4 parameters of the extended Nernst equation (1) are solved by iteration:

$$\Delta E = \Delta E^0 + G \log [H^+] + j_{II} [H^+] + j_{OII} K_w/[H^+] \qquad (1)$$

This method allows an exact concentration calibration - a buffer calibration uses activities - and so the dissociation constants can be calculated as pK_c.

The most important application is the determination of complex stability constants. Some examples for biorelevant organo-phosphonic acids are described on the poster.

T_ALPHA

This WINDOWS 3.1 program is used to control the new „TitroLine *alpha*"-Titrator (SCHOTT-GERÄTE GmbH) by a RS 232 C port (COM1 to COM4). All basic functions are accessed by intuitive buttons. Any other functions can be chosen from a listbox.

Every data output can be viewed and edited in a window. The data lists needed for other programs like ITERAX can be saved as text files on disk or can be transferred to a LIMS.

The use of this program with a TitroLine *alpha* offers one of the lowest budget systems for use in student laboratories, research groups or in industry. The TitroLine *alpha* contains 100 predefined methods f.e. H_2SO_4 vs. NaOH or Karl-Fischer-Titrations and so on. Measurements of pH, mV, temperature or µA (for dead stop titrations) are possible. The TitroLine *alpha* software runs in the background, so the PC can be used for other activities like iterations of previous titrations.

Iteration and Simulation of Microscopic Titration

Determination of Macroscopic and Microscopic Stability Constants by the PHOTO_T concept

Prof. Dr. G. Hägele, Cand.-Chem. C. Verwey and Dipl.-Chem. C. Arendt

HEINRICH-HEINE-Universität Düsseldorf
Institut für Anorganische Chemie und Strukturchemie I
Universitätsstr. 1, D-40225 Düsseldorf, Germany

Summary: We describe the hard- and software combination used in the PHOTO_T concept. The six programs are demonstrated on several organo phosphorus compounds. New version are shown. Iteration and simulation algorithms are explained.

PHOTO_T-Concept

Nowadays, macroscopic and microscopic dissociation constants are of increasing importance in matters of research and product development. The determination of those parameters is both economical and convenient in using the PHOTO_T concept developed in Düsseldorf for compounds with titration dependent UV-Vis-spectra.

Six PC-programs are combined in the concept. Five of them, PHOTO_T, ITERAX, GENTIT, MICRO_IT and MICROSIM, are written by our research group, one, TPC°2000 (SCHOTT-Geräte, Hofheim, Germany), is a commercial titration program.

The PHOTO_T concept is connecting potentiometric pH-titration with the registration of UV-Vis-spectra. So called spectrophotometric titration curves are gained from pH-dependent spectra. These curves can be used to determine microscopic dissociation constants as well as the formation of dimeres.

GENTIT and MICROSIM allow to simulate (macroscopic or microscopic) titration curves as well as the corresponding molar fraction diagrams for all components.

ITERAX performs graphically and user controlled iteration of macroscopic stability constants. Impurities and additional strong acids or bases are considered as well.

MICRO_IT iterates microscopic pk_a-values using spectrophotometric and potentiometric data.

MICRO_IT

MICRO_IT requires 2 experimental functions: "absorption vs. volume of titrator" and "pH vs. volume of titrator", as well as the macroscopic pK_a-values. The latter are gained from potentiometric data by iteration with ITERAX.

The function "absorption ($\sim\alpha_i$) vs. volume of titrator" is generated by PHOTO_T from pH-dependent absorption spectra. α_i values are the sums of molar fractions of microscopic species containing the observed group in the observed deprotonation status.

MICRO_IT minimises the difference (rms) between experimental and simulated function „α vs. pH" to find the microscopic pk-values

MICROSIM

MICROSIM simulates titration from microscopic dissociation constants. It is designed to be used in teaching and titration planning.

The pH-dependent properties of protolyte solutions can be predicted by simulation with ionspecific parameters. Chemical shift δ, energy eigen values or molar decadic absorption coefficient ε_λ can be chosen.

Input masks prove the input and reject invalid values. Context sensitive help is available on every topic. A brief introduction to microscopic dissociation concept is included.

All simulation results can be viewed, saved and printed as ASCII text and HPGL graphic. Scale, stile and labels in graphics can be changed.

NMR- and UV- controlled Titrations:

Macroscopic and microscopic stability constants.

Prof. Dr. G. Hägele, Dipl.-Chem. J. Ollig and Dipl.-Chem. H.W. Kropp

HEINRICH-HEINE-Universität Düsseldorf
Institut für Anorganische Chemie und Strukturchemie I
Universitätsstr. 1, D-40225 Düsseldorf, Germany

Summary: We describe automated combination techniques - NMR- and UV-controlled titrations. Macroscopic and microsopic dissociation concepts of several organo-phosphorus compunds are demonstrated.

Protolytic- and metalcomplex formation equilibria are monitored by automated combination techniques. Theory, experimental setup and results from appplications are given. Examples are chosen from the following classes of compounds:

Phosphonic acids, Posphinic acids, Aminophosphonic acids, Aminophosphinic acids, Phosphonocarboxylic acids, Phosphinocarboxylic acids and combinations with some biorelevant cations. In addition aminoacids and peptides are studied. The simultaneous determination of stability constants in mixtures by NMR-controlled titrations is shown.

EVALUATION OF STATIONARY PHASE POLARITY BY PRINCIPAL COMPONENT ANALYSIS IN GAS-LIQUID CHROMATOGRAPHY

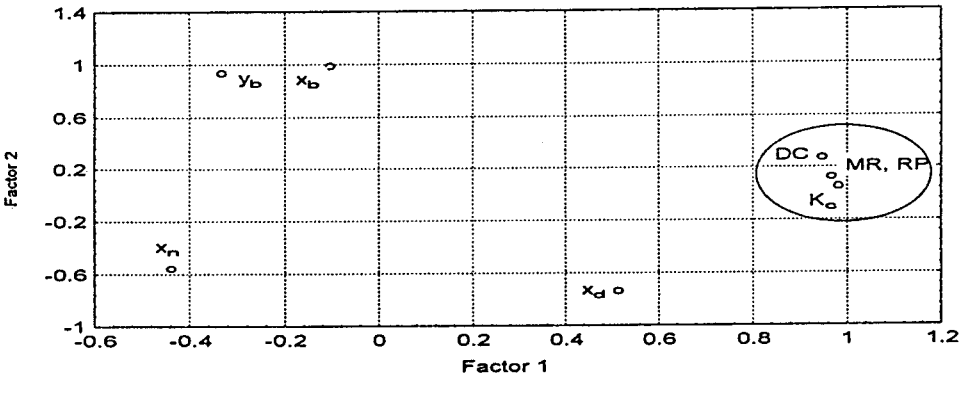
K. Héberger
Central Research Institute for Chemistry of the Hungarian Academy of Sciences
POB. 17. H-1525 Budapest, HUNGARY

Summary: The presentation classifies and characterizes (i) the different polarity indicators and (ii) the various stationary phases.

Recently, principal component analysis (and related techniques) proved to be a useful tool in characterizing the polarity of stationary phases[1-3]. However, there is no comprehensive study available as to which polarity scale can be applied successfully and under which conditions. Therefore our aim was to evaluate the different polarity scales: the first five McReynolds constants (MR), the Kováts coefficients (K_C), Snyder's selectivity parameters (x_b, x_n, x_d), the retention polarity (RP) and Castello's DC, y_b values using 30 different stationary phases. As input matrix Castello's data[4] were used.

The first principal component explains 54%, the second 35% and the third 10% of the total variance, indicating that no single polarity parameter is applicable alone. The most informative plot (Fig.1) identifies a small cluster: MR, DC, K_C and RP carry very similar information. Any of them can be used for the characterization of polarity. Moreover, MR and RP are identical, *i.e.* linearly not independent. Likewise, y_b and x_b are similar. Therefore, the usage of RP and y_b is not justified. The score plots determine the similarity among stationary phases.

Fig. 1: Unrotated Factor Loadings



1. Z. Juvancz, T. Cserháti, K. E. Markides, J. S. Bradshaw and M. L. Lee: *Chromatographia*, **38**, 227 (1994)
2. E. Forgács: *Anal. Chim. Acta*, **296**, 235 (1994)
3. T. Cserháti: *Anal. Lett.*, **27**, 2615 (1994)
4. G. Castello, G. D'Amato and S. Vezzani: *J. Chromatogr.*, **646**, 361 (1993)

SIMULATION OF BAND PROFILES IN CHROMATOGRAPHY BY NUMERICAL CALCULATIONS

Pavel Jandera

University of Pardubice, nám. Legií 565, 532 10 Pardubice

Summary: Simulation of band profiles in chromatography can be used as a very useful tool to describe, predict and optimize separation under conditions where nonlinear adsorption isotherms apply, such as in preparative liquid chromatography.

The chromatographic theory is well established and can be used in straightforward manner to predict and optimize the retention and the quality of separation under usual analytical conditions, where the distribution of the analytes between the stationary and the mobile phases is controlled by a linear isotherm. However, under certain circumstances, the linear isotherm does not apply. The most important case is preparative liquid chromatography, where the requirement of a high production rate makes it necessary to use high sample concentrations, so that the chromatographic separation occurs under "overloaded" conditions, where significant deviations from linearities of the distribution isotherms usually are observed. The attempts at extrapolating the results of linear chromatography to nonlinear conditions usually are not successful because of the convolution between the band broadening resulting from the nonlinearity of the isotherm (thermodynamic origin) and the band broadening of kinetic origin. The rule of the additivity of variances does not apply here and it is not possible to separate the two effects in the analysis of the band profiles.

Such systems can be described by the equilibrium-dispersive model of non-ideal, nonlinear chromatography. To solve the basic mass-balance equation of chromatography postulated by this model, numerical calculations are necessary to predict the sample band profiles. From among various approaches, we selected the finite difference method and Rouchon's algorithm of solution. Here, the axial dispersion coefficient is replaced by an estimate of the apparent diffusion coefficient and the kinetic equation is replaced by the isotherm equation, the parameters of which are more easy to determine in practice using an independent method. Numerical solution of the basic differential equation is possible by substituting differentials by finite differences, using the concept of the theoretical plate for the column chromatography.

Successful application of this approach requires accurate determination of the distribution isotherms of the sample components. The dynamic methods, such as frontal analysis, save time and are straightforward when single-component isotherms are measured. These methods can be used in principle also to measure multicomponent (competitive) equilibrium isotherms, necessary for calculations of the band profiles of the sample compounds in mixture,

however, the optimum form of the equation to be used to describe the competitive isotherms has not been established unequivocally yet.

The possibilities of the application of the numerical band simulation approach are illustrated by examples of calculation of the elution profiles of high-concentration sample bands under overloaded column conditions used in preparative liquid chromatography with either isocratic or gradient elution modes. The effects of the sample solvent and of the limited solubility of the sample in the mobile phase on the isotherms and on the elution profiles can be also predicted. In addition to preparative chromatography, the numerical calculation approach is useful also to model some more complex situations occasionally met in analytical high performance liquid chromatography such as the effects of preferential adsorption of the stronger solvent on a polar adsorbent used as column packing in normal-phase chromatography during gradient elution experiments. Another practical example is modelling of the effects of using sample solvent with the elution strength different from that of the mobile phase, when higher sample volumes are injected to improve the possibilities of detection and quantitation of the components of the sample mixture.

INVESTIGATION OF ENANTIOMERS IN AROMA SYSTEMS USING MULTIVARIATE ANALYSIS OF GC/FTIR/MS DATA

L.Jirovetz[1], A.Nikiforov[2] and A.Woidich[3]

[1]Institute of Pharmaceutical Chemistry, University of Vienna, A-1090 Vienna, Althanstreet 14, Austria

[2]Institute of Organic Chemistry, University of Vienna, A-1090 Vienna, Waehringerstreet 38, Austria

[3]Food Research Institute, A-1190 Vienna, Blaasstreet 29, Austria

Summary: The enantiomers in complex aroma mixtures were investigated by enantioselective gas chromatography (cyclodextrin and peptide phases). Beside classic methods of MS and IR data interpretation also Multivariate Data Analysis (MVDA) was tested.

The degree of racemisation of aroma compounds is of basic importance in complex aroma mixtures (e.g. lactones in strawberry extracts or monoterpenes in essential oils). These enantiomers were investigated by enantioselective gas chromatography. Different column types of chiral phases on basis of peptides (D- and L-valine) and cyclodextrins (alkyl-, hydroxyl- or carboxyl-groups) were used. Some achiral columns (polar and non-polar phases) were also necessary for correlation experiments. The cyclodextrin-columns showed highest separation quality in all cases. As detectors the flame ionisation detector, mass spectrometer (MSD from Hewlett-Packard) and infrared detector (IRD from Hewlett-Packard) were taken. Classis methods of MS and IR interpretation could be combined with Multivariate Data Analysis (MVDA). The MVDA gives valuable information on the degree of racemisation of aroma compounds in such complex aroma systems. These basic methology and the results of MVDA will be presented as well as mentioned GC separations with both achiral and chiral phases.

REMARK ON THE DYNAMICAL CALIBRATION

Stanislav KEPRTA

Charles University, Prague

Summary: *This contribution describes the calibration of a sensor when the records of signals of a dynamical process from the sensor and a standard were available.*

Suppose we register continuous signals in equally spaced time moments. As an example consider the measurements on a forging press of velocity v and distance s measured by a standard, and velocity v_c and distance s_c measured by a sensor which will be manufactured on a large scale and which will be used in everyday practice. Assume we have available several records of the signals under the same conditions. The main objective of the analysis is to calibrate the sensor and to estimate the repeatability and the precision of the measurement.

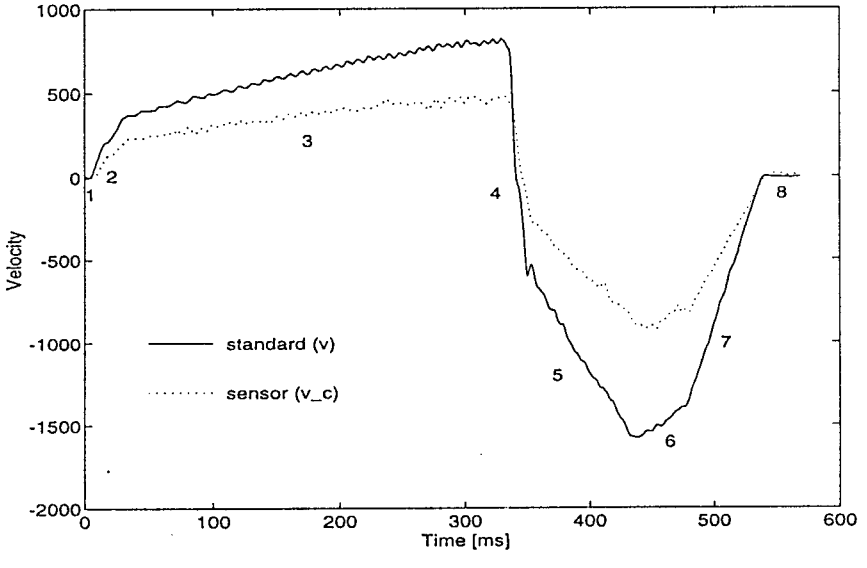


Figure 1. The signals corresponding to the velocities v and v_c consisting of several phases (i.e. no action denoted by 1, acceleration denoted by 2, free fall denoted by 3, etc.).

The examination of the data revealed that the signal v_c from the sensor was delayed in comparison with the signal v from the standard. To predict the actual velocity from v_c we had to estimate the delay δ as precisely as possible. To do that we applied two different approaches:

- maximization of the correlation between $v(t)$ and $v_c(t+\delta)$ over δ;
- detection of change-points between phases.

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX

Shift (in ms) estimated by	Number of experiment								
	1	2	3	4	5	6	7	8	9
maximum correlation	4	4	4	4	4	4	4	4	4
diference of change-points between phases 3 and 4	4	3	3	4	4	3	4	3	3
diference of change-points between phases 5 and 6	7	8	7	9	7	8	8	8	7

Table 1. Estimated shift between v and v_c.

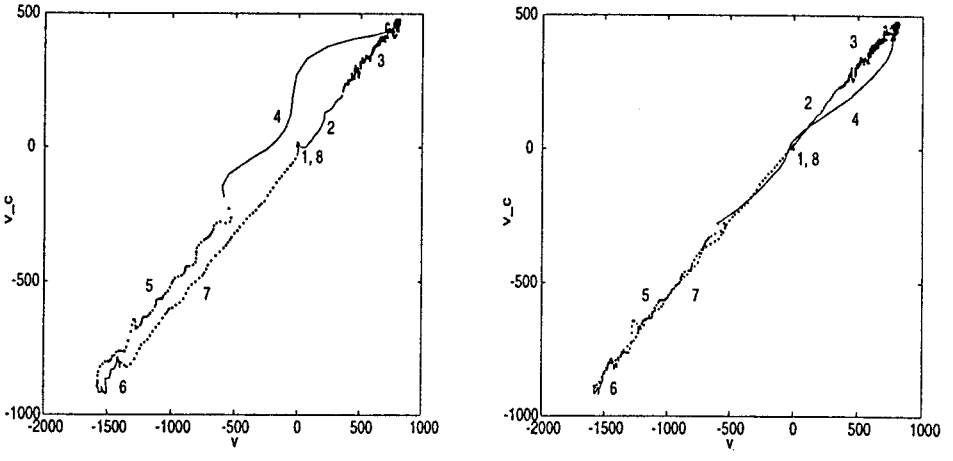


Figure 2. $v_c(t)$ versus $v(t)$ and $v_c(t+\hat{\delta})$ versus $v(t)$, $\hat{\delta} = 4$ ms.

Another problem we had to deal with was that the precision of the sensor was considerably lower than that of the standard. The sensor forms an integral part of the forging press and therefore reflects all its vibrations and movements while the standard is connected with the press only indirectly. This is sometimes reflected in unexpected outliers in observed values.

The main target was to calibrate the sensor and to estimate its repeatability. During the lecture the main statistical problems will be summarized, appropriate models will be described and an illustration based on the real data mentioned above will be presented.

ACKNOWLEDGEMENT: This research was partially supported by the grant GA ČR – 201/94/0472.

ADDRESS: Dept. of Probability and Mathematical Statistics, Faculty of Mathematics and Physics of Charles University, Sokolovská 83, 186 00 PRAHA 8, Czech Republic
E-mail address: keprta@karlin.mff.cuni.cz

NEW METHOD OF DETERMINATION OF NUMBER OF COMPONENTS IN THE MIXTURE

Karel Komers, University of Pardubice, Faculty of Chemical Technology, Department of Physical Chemistry, Nám. Čs. legií 565, 532 10 Pardubice, Czech Republic

Summary: Determination of the number of species in a mixture, based on a new type of mathematical analysis of the matrix of suitable physico-chemical properties (e.g. absorbances) which depend on the composition of the mixture and on another experimental condition (e.g. wavelength of the used light), is described.

Many quantitatively measurable physico-chemical properties of mixtures (e.g. absorbance, electrolytic conductivity, polarographic current, area of peaks in GC or HPLC spectrum, molar polarisation and refraction) have additive character. Such property of the mixture A is connected with the properties of its components A(s) through the equation $A = \sum A(s)$ for the number of components s = 1 to n. A(s) is linearly dependent on the concentration of the component s in the mixture c(s) according to the relation $A(s) = e(s).c(s)$; e(s) depends on the quality of the component and on the experimental conditions (e.g. length of the measuring cell and the wavelenght of the light in the case of absorbance, temperature in conductivity, electric potencial in polarographic current, composition of the mobile phase and the type of the detection in GC or HPLC). If the ratio of the concentrations of n compounds in the given mixture is altered (e.g. because of the chemical reaction between these components or of changing the equilibrium constant of this reaction by temperature, pressure, pH, etc.), the following equation can be written for the total property A(i,j) for every combination of concentrations j (j=1 to k) and given experimental condition i (i=1 to r)

$$A(i,j) = \sum e(s,i).c(s,j) , \qquad s=1 \text{ to } n \qquad (1)$$

If A(i,j) is measured at r experimental conditions and k combinations of concentrations, a matrix of the type **A(r,k)** is obtained, whose rank (for $r,k \geq n$) gives the number of undependent components in the mixture[1]. The main difficulty in the determination of this rank lies in the experimental errors in the A(i,j) values. This problem was discussed by several authors[2-6]. The solutions publicated do not lead allways, especially for multicomponent mixtures, to unambiguous results.

The proposed method makes use of the fact, that in the matrix **A(r,k)**, which must be of at least (n + 1) order for n components , holds <u>in all rows</u> that the property from any column A(i,n+1) is a linear combination of the properties from the other arbitrary chosen n columns A(i,j) in the same row

$$A(i,n+1) = \alpha_0 + \sum \alpha_j.A(i,j) , \quad j=1 \text{ to } n \qquad (2)$$

The coefficients α only are functions of c(s,j) of the particular components in the chosen columns. <u>In all columns</u> a similar equation is valid

$$A(n+1,j) = \beta_0 + \sum \beta_i.A(i,j) , \quad i=1 \text{ to } n \qquad (3)$$

The coefficients β only are functions of e(s,i) of the particular components in the chosen rows. For all n it is $\alpha_0 = \beta_0 = 0$. For a one-component system it is e.g. $\alpha_1 = c(1,2)/c(1,1)$, $\beta_1 = e(1,2)/e(1,1)$. For a two-component system it is $\alpha_1 = [c(1,3).c(2,2) - c(2,3).c(1,2)]/c$, $\alpha_2 = [c(2,3).c(1,1) - c(1,2).c(2,1)]/c$, $c = [c(1,1).c(2,2) - c(1,2).c(2,1)]$. For β_1, β_2 the same equations are valid with e(s,i) instead of c(s,j).

Equations (2) and (3) are not valid for a number of A(i,j) smaller than m = n. They hold for m > n, but with other dependences $\alpha \sim c(s,j)$ and $\beta \sim e(s,i)$.

For the verification of (2) and (3) the model matrixes of absorbances of the type **A(12,12)** were calculated for n = 1 to 10 on the basis of 120 e(s,i) and 120 c(s,j) values, obtained from the generator of random numbers. With the original computation program in Quick Basic all A(i,j) values were calculated exactly and then (again by random number generator) with imitated errors ± 0 to 1, 0 to 5 and 0 to 10% of the exact A(i,j). In every matrix the values of α and β for m = 1 to 10 at various combinations of rows and columns were calculated with the method of linear regression using the program included in Qattro Pro. The validity of (2) and (3) was checked by the values of the correlation coefficient R, the coefficients α or β and their standard deviations (s.d.).

Results

The theoretical assumptions were confirmed using the exact model matrixes. Equations (2) and (3) for m < n have R < 1, α_0 and $\beta_0 > 0$ and great s.d. for individual α and β. For m = n it is R = 1 , s.d. and $\alpha_0, \beta_0 = 10^{-14}$ to 10^{-15}. For m > n equations (2) and (3) are also fulfilled; R stays 1.

The model matrixes with artificial errors have also the growing dependence $R \sim m$, but not to R = 1 for $m \geq n$. The growth of R is smaller at greater errors in A(i,j). R has for $m \geq n$ the greatest and practically constant value.

The suggested method reacts sensitively also on the dependences between the concentrations of the mixed components. For example for the direct proportionality c(s,j) = a.c(1,j) the determinated n shows only the number of linearly independent species in the mixture. The linear dependence between the concentrations of the mixed components c(s,j) = a.c(1,j) + b (a,b = constants) is expressed in $\beta_0 = f[a,b,e(s,i)] > 0$. A nonlinear dependence (resulting e.g. from the kinetic equation or equilibrium constant) do not influence the real number of mixed species determined with this method as n.

The proposed method has been checked on several experimental matrixes of absorbances taken from literature[2,5,6] and based on two own measurements. The next checking of this method will be continued.

Literature

1) Kankare J.J.: Anal.Chem. 42 (1970) 1322
2) Ainsworth S.: J.Phys.Chem. 65 (1961) 1968
3) Ainsworth s.: J.Phys.Chem 67 (1963) 1613
4) Sternberg J.C.,Stillo H.S.,Schwendeman R.H.: Anal.Chem. 32 (1960) 94
5) Wallace R.M.: Anal.Chem. 64 (1960) 899
6) Katakis D.: Anal Chem 37 (1965) 876

NONLINEAR TWO POINT CALIBRATION OF BIOSENSORS

J.Krejčí[1], M.Kraus[2]

(1) Krejčí Engineering, Riegrova 318, 666 01 Tišnov, Czech Republic

(2) COMWORK, Severovýchod 3, Zábřeh, Czech Republic

Summary: *The enzymatic biosensors show the nonlinear calibration curve. The contribution proposes two parametric nonlinear aproximation of calibration curve.*

KEY WORDS: Calibration curve, biosensor, glucose electrode

The calibration curve of the enzymatic biosensors shoves the typical nonlinarity which has a saturation character. For low concentrations of the analyte the calibration curve is linear. The calibration curve converges hyperbolically to its asymptotical value for high analyte concentrations.

Let us suppose that the calibration curve has a functional form $s = C/\alpha$ for $C < C_0$ and $s = a + b/(C + d)$ for $C > C_0$. Very good aproximation of the calibration curve may be constructed if the calibration curve is "sewn up together" from its initial and asymptotical parts. To obtain the minimal number of parameters it is possible to add additional conditions of smoothness. And in addition to it, it is possible to complete the calibration curve from parts which correspond physically to the solved problem. This "sew up together" must have better behavior then formally designed aproximation by for instance a polynom, because it is composed from physically reasonable parts.

Foregoing idea may be solved mathematically and the final expresion is

$$C = \alpha * s \qquad \text{for } s < s_0$$

$$C = 2*C_0 + \alpha * \frac{s_{oo}\,(s - s_o) - s_o\,(s_{oo} - s_o)}{s_{oo} - s} \qquad \text{for } s > s_0$$

where C - Analyte concentration

C_0 - Limit analyte concentration up to which the shape of calibration curve may be considered to be linear.

(limit of linearity)

s - Ouput signal of biosensor (Faradaic current, time derivative of current, conductivity,...)

s_0 - Output signal of biosensor for concentration of limit of linearity ($s_o=s(C_o)$).

s_{oo} - Saturated output signal of biosensor ($s_{oo}= \lim_{C\to\infty} s(C)$)

α - Calibration constant of linear per of the calibration curve.

The parameter s_o is defined by the type of enzymatic reaction. The parameters α a s_{oo} may be determined by the calibration in two points C_1 and C_2 one of which lies below of limit of linearity C_o and the second one lies above of C_o by

$$\alpha = \frac{C_1}{s_1} \qquad \text{for } C_1 < C_0$$

$$s_{oo} = s_2 + \alpha * \frac{(s_2 - s_0)^2}{(C_2 - \alpha * s_2)} \qquad \text{for } C_2 > C_0$$

where s_1(C_1) and s_2(C_2) are the signals in the points C_1 and C_2 respectivelly.

The proposed approximation of the calibration curve involves only the operations of addition, subtraction, multiplication and division. That is why it is very suitable for making programmes in one chip computers.

The results were verified experimentally while using a device with nonlinear calibrating curve implemented into a microprocessor, which served for glucose measurement in vitro and clinically.

Conclusions

The nonlinear calibrating curve allowes a simple and quick calibration of biosensor and comprises also a saturation part of calibration curve. Only two calibrating points are needed for the determination of the calibration curve parameters. From the theoretical analysis it results that this calibration method may be used effectively for majority of biosensors' calibration types.

A Hybrid of Simplex Method and Simulated Annealing

Vladimír Kvasnička
Department of Mathematics, Slovak Technical University
812 37 Bratislava, Slovakia
email: kvasnic@theochem.chtf.stuba.sk

Summary: *A hybrid of simplex optimization method and a stochastic simulated annealing approach is discussed. It provides efficient optimization technique for finding global minima of highly multimodal functions.*

Standard simplex method [1,2] belongs to well-known nongradient optimization techniques often used in chemistry. Its generalization towards its randomization was done by Price [3] (cf. also ref. [4]). His the so-called control random search (CRS) method is able to look for, to some limited extent, global minima, and may be now considered as one of forerunners of modern stochastic optimization techniques (genetic algorithms [5], simulated annealing [6], evolution strategy [7], etc.). The purpose of this communication is to give a brief outline of our recent efforts to elaborate a hybrid of the simplex method and the simulated annealing approach, which would be more effective and robust than its single predecessors.

Let f be an objective function

$$f: D \subset R^n \to R \tag{1}$$

where $D=\prod[a_i,b_i]$ is a domain of f. The following optimization problem is considered

$$x_{opt} = \arg\min_{x \in D} f(x) \tag{2}$$

Let $P = \{x_1, x_2, \ldots, x_p\} \subset D$ be a population of p points from the domain D. A subset (simplex) composed of n+1 points is randomly selected from points in P

$$S = \{x_{\alpha_1}, x_{\alpha_2}, \ldots, x_{\alpha_{n+1}}\} \subset P \tag{3}$$

One point of S with highest functional value is especially distinguished

$$x_H = \arg\max_{x \in S} f(x) \tag{4}$$

A center of gravity of all simplex points but the point x_H is determined by

$$\bar{x} = \frac{1}{n}\left(\sum_{x \in S} x - x_H\right) \tag{5}$$

A reflection of x_H with respect to the center of gravity $\bar{x}$ is (cf. ref. [4])

$$x^* = x_H + r(\alpha_o, \sigma)(\bar{x} - x_H) \tag{6}$$

where $r(\alpha_o, \sigma)$ is a random (logistic or Gaussian) number with a mean α_o and a "standard deviation" σ. In our model calculations these parameters are specified as follows: logistic distribution, α_o=2, and σ=0.5.

The created reflection point x^* is included into the population P for further simulated annealing process together with elimination of the point x_H from the population P by a probability determined by the Metropolis criterion [8]

$$Pr = min\left(1, exp\left(-\left(f(x^*) - f(x_H)\right)/T\right)\right) \tag{7}$$

where T is called the 'temperature" and plays an important role of the control parameter of the simulated annealing approach. It is decreased from T_{max} to T_{min} by successive multiplication by α , $T \leftarrow \alpha T$, where $0 << \alpha < 1$. Simple pseudo-Pascal implementation of the proposed method is outlined by the following algorithm

```
procedure Simplex_Simulated_Annealing;
begin P:=set of p randomly selected points of D;
      T:=Tmax;
      while T>Tmin do
      begin for k:=1 to kmax do
            begin S:=set of n+1 randomly
                     selected points of P;
                  x*:=reflection(S);
                  Pr:=min(1,exp(-(f(x*)-f(xH))/T));
                  if random<Pr then xH:=x*;
            end;
            T:=α*T;
      end;
end;
```

The algorithm is initialized by a random generation of p points of population P. The initial (maximal) temperature is set equal to T_{max}. Outer cycle is repeated while $T>T_{min}$, where the temperature T is subsequently multiplicatively decreased by $T:=\alpha*T$. The inner cycle is repeated k_{max} times, where k_{max} is sufficiently great number (usually from 10^3 to 10^5).

The proposed method has been also extended to the so-called *parallel* version, where the population P is divided into disjoint subpopulations P_1 , P_2 , ... , P_q , where the above described method is applied independently for these subpopulations. The simulated annealing is synchronized for all subpopulations so that all have the same temperature T, which is simultaneously decreased for all of them. In order to introduce an interaction between subpopulations, randomly (with very low level of probability, e.g. $Pr=10^{-3}$) selected subpopulations mutually exchange best solutions so that the worst ones are eliminated. We have carried out trial computations with this approach in the framework of UNIX parallel environment and it is observed that the efficiency and robustness of the method is substantially increased with respect to the nonparallel simplex simulated annealing described above.

References

[1] R. Hooke and T.A. Jeeves: *Direct Search Solution of Numerical and Statistical Problems*. J. Assoc. Comp. mach. 8(1961), 212.

[2] J.A. Nelder and R. Mead: *A Simplex Method for Function Minimization*. Comp. J. 7(1965), 308.

[3] W.L. Price: *A Controlled Random Search Procedure for Global Optimization*. Comp. J. 20(1976), 376.

[4] I. Křivý and J. Tvrdík: *The Controlled Random Search Algorithm in Optimizing Regression Model*. Comp. Stat. Data Anal. (to appear).

[5] G. Goldberg: *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley, Reading (1989).

[6] P.J.M. van Laarhoven and E.H.L. Aarts: *Simulated Annealing: Theory and Applications*. D. Reidel Publishing Co., Dordrecht (1987).

[7] H.-P. Schwefel: *Numerical Optimization of Computer Models*. Willey, Chichester (1981).

[8] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller: *Equation of State Calculations by fast Computing Machines*. J. Chem. Phys. 21(1953), 1087.

PROGRAM FOR THE DETERMINATION OF EQUILIBRIUM CONSTANTS FROM POTENTIOMETRIC, SPECTROPHOTOMETRIC, NMR AND OTHER DATA

Mojmír Kývala, Ivan Lukeš,
Department of Inorganic Chemistry, Charles University,
Albertov 6, 128 43 Praha 2, Czech Republic

Summary: A new computer program for the analysis of solution equilibria by the generalized least squares method is presented. Original data taken from our coordination chemistry research are treated to illustrate the versatility of the program.

The normal approach in the determination of equilibrium constants in solution consists of the measurement of suitable physical properties of the solution in dependence on the composition of the solution. These physical properties are usually the electromotive force of a cell, light absorbance, chemical shift or relaxation time in NMR and the *g*-factor in EPR spectra, the integral intensity or peak height in optical spectra, the height of a polarographic wave, etc. It holds in general that the measured quantity y is a function of the equilibrium concentrations of the components $c = \{c_i\}$, i = 1, .., n+m, (n is the number of reactants, m is the number of complexes) and of vector χ of certain parameters $y = \phi(\chi;c)$. The program described here now includes almost twenty different calibration functions, the commonest of which are, e.g.

$$y = E_0 + S \cdot log\, c_1 + j_1 c_1 + j_2 \frac{\beta_1}{c_1} = \phi(E_0, S, j_1, j_2; c_1)$$

for potentiometric determinations using a glass electrode in a cell with non-negligible liquid junction potential (c_1 is the equilibrium concentration of H^+ ions, $\beta_1 \equiv K$ is the dissociation constant of water) or

$$y = \sum_{i=1}^{n+m} a_i c_i = \mathbf{a}^T c = \phi(\mathbf{a};c)$$

for spectrophotometric or NMR determinations (**a** can be the vector of molar absorption coefficients for the given wavelength, the vector of real numbers proportional to the chemical shifts or relaxation rates of the pure components, etc.).

The dependence of the equilibrium concentrations on the analytical concentrations is contained in the mass balance equations (solutions prepared by the titration)

$$c_k + \sum_{i=1}^{m} q_{ik} \beta_i \prod_{j=1}^{n} c_j^{q_{ij}} = \frac{c_k V + c'_k x}{V + x} = \frac{n_k + c'_k x}{V + x} \quad , \quad k = 1, .., n$$

where q_{ij} is the element of the matrix of stoichiometric coefficients **q** of type (m,n), β_i is the stability constant of the i-th complex, c_i, n_i is the concentration or mass amount of the i-th reactant in the titrated solution, c_i' is the concentration of the i-th reactant in the titration solution and x is the volume of titration solution added. Thus, formally, $y = f(\chi, \mathbf{c}, \mathbf{n}, V, \mathbf{c}', \beta; x)$, where function f can be expressed analytically only in the simplest cases. If we now have k independent pairs of real numbers $\{x,y\}$, then this functional relationship can be used to calculate $l \leq k$ unknown values of any of parameters χ, **c**, **n**, V, **c'**, β.

Provided $l < k$, then each measured value Y_i of the test quantity y is considered to be a realization of a random variable $Y_i = y_i + \varepsilon_i$, where ε is a random vector with multi-dimensional normal distribution, zero mean value and variance-covariance matrix **V**. Under these assumptions, the maximum reliable estimate of the unknown parameters is the solution of a overdetermined set of k equations (for l unknown) according to the generalized least squares method.

The formulation of matrix **V** is based on the assumption that not only variables y and x, but any of the known parameters χ, **c**, **n**, **V**, **c'**, β can be accompanied by a random measuring error with the zero mean value and constatnt variance σ_y^2, σ_x^2, etc. Let us assume that these errors are the only source of random variable ε, that they act additively on value y (when multiplied by corresponding partial derivative) and that they are independent, thus e.g.

$$\varepsilon_i = \sigma_y + \left(\frac{\partial y}{\partial x_i}\right)\sigma_x + \left(\frac{\partial y}{\partial c_{k_i}}\right)\sigma_{c_k} + \left(\frac{\partial y}{\partial c_{l_i}}\right)\sigma_{c_l} + \left(\frac{\partial y}{\partial V_i}\right)\sigma_V + \ldots$$

then the elements of matrix **V** can be readily calculated from a definition such as $V_{ii} = D(\varepsilon_i) = E(\varepsilon_i^2)$ and $V_{ij} = V_{ji} = cov(\varepsilon_i, \varepsilon_j) = E(\varepsilon_i \varepsilon_j)$.

The choice of vector χ has a dominant effect on the correctness of the estimates of the equilibrium constants. Where possible, preference is given to external calibration over internal calibration. The conditionality of parameters E_0, S, j_1, j_2, see above, can, for example, be significantly increased if each actual titration is preceded by a titration of a strong acid by a strong hydroxide of known concentrations (at a given temperature and ionic strength). The program then assigns each of these pairs of titrations a joint calibration function (a joint subvector of vector χ).

All the elements of vector χ in all the calibration functions implemented so far are linear, and thus the program does not require their initial estimates. If the selected elements of vector χ are the only unknown parameters, then a calculation is made of their unbiased estimates by the iterative linear least squares method (matrix **V** is in general a function of unknown parameters), where biased estimates are also available, obtained by the ridge regression method. If initial estimates of the unknown stability constants are not given, then the program simply generates them.

In the whole program, the first derivatives are calculated analytically, while the second derivatives (required in calculation of the maximum and average internal and parametric curvature of the model, Box bias of the estimates and residuals etc.) are calculated numerically from the analytical first derivatives.

In adition to a point estimate of the mean value of the estimate of the unknown parameters, the program also offers classical analysis of this estimate in a linearization approximation (variance, *t*- and *F*-interval estimate of the mean value, a significance test, correlation matrix, etc.) as well as likelihood ratio analysis (interval estimate of the mean, sigificance test). At each point, several types of residuals are calculated along with a number of scalar and vector quantities for indication of golden points, outliers and high leverage points. The program also tests several hypotheses on homogeneity, independence and normality of the standardized residuals.

The percentage points of the tested distributions are calculated analytically directly from their definitions (where neccesary, using the value of the transcendental function $2/\pi^{1/2} erf$). The corresponding critical values are calculated by the Newton iteration method with analytical first derivatives.

Investigation of Cancerogenesis in Surroundings of Nuclear Power Plants

Maria Lętkovičová [1], Ján Korec [1], Vladimír Prikazský [2], Viktor Príkazský [3], Zdeněk Smejkal [4]

[1] Nuclear Power Plants Research Institute, Trnava, Slovak Republic
[2] State Medical Institute, Centre of Epidemiology, Prague, Czech Republic
[3] State Medical Institute, Poprad, Slovak Republic
[4] University of Pardubice, Czech Republic

Summary: This work analyses the long-term incidence of cancers and mortality in surroundings of the Nuclear Power Plant Jaslovské Bohunice.

The purpose of this work is to summarize data about incidence of cancers and about the chosen indicator - the life reducing in Trnava region and to analyze these data in respect to operation of nuclear-energetic plant in one of investigated territory.

Data and methods: The data about cancer incidence were summarized according to regions and years, separately for men and women. The main investigated territory is the Trnava region, the comparative territories is the Nitra and Levice regions.

These regions were chosen from following reasons:
Approximately 60% of Trnava region inhabitants live in the distance shorter than 20 km from the NPP which first reactor was put into operation in 1979. In Levice region there are the NPP Mochovce, which aren't under operation now and this analysis serves as a starting point for the next period. Nitra region is a similar one, neutral, without any big industrial plant. The real data about cancer incidence were summarized for years 1968 - 90, the trend was calculated till 2000. There exists one assumption that tumours caused due to long-term activity of small doses of injurants need a long period to become evident. The harmful ecologic influences function just this way, that means permanent small doses. The best way for this work was to investigate summary of all cancers, because the irradiation of the body was inhomogeneous and the analysis of doses in different tissues can't be performed due to broad spectrum of contaminating radionuclides (besides of the kind of radionuclide there has a significant meaning also the chemical form of radionuclides), different entries into body and the time distribution of irradiation from environment.

The mortality was investigated as from the point of view of percentage of premature deaths and potential lost years of live (PLYL). This analysis was done for each municipality in Trnava region for men and women separately in years 1986 - 90 in the distance shorter than 30 km from NPP Jaslovské Bohunice. 114 municipalities were investigated. All the data about deaths in 1986 - 90 were considered and the investigated mortality parameter was the PLYL. The basis for calculation was an average men age (69) and women age (74) in Slovakia. Each death which comes sooner than at average age is calculated as premature one and the "not living years" are calculated into the whole summary. The division of the whole sum by the number of premature deaths makes an average for one premature deaths. All the not living years were calculated in each municipality in years 1986 - 90 to exclude an accidental influence of one year.

The mortality was calculated by the accumulation analysis. It was assumed if there exists a harmful influence of NPP to inhabitants in its surroundings, it will become evident after some time even with a different type of mortality in related municipalities. Some reducing of life length was considered. The causes of death weren't investigated because this information can easy undergo a

subjective error. It was assumed that there could become evident different types of mortality in municipalities which are settled in a direction of dominant winds from the NPP, where could be an influence of radioactive deposit, or in river basins of rivers containing water from the NPP.

Results: Table of incidence of cancer tumours (relative counts according to 100.000 inhabitants)

Region	*Situation*	*FORECAST*					*Equation*
	1990	*1992*	*1994*	*1996*	*1998*	*2000*	
MEN							
Trnava	365	372	375	378	381	385	334.6 + 1.574 t
Levice	437	449	457	466	474	482	355.1 + 3.827 t
Nitra	415	442	455	469	482	495	292.6 + 5.923 t
SLOVAKIA	383	393	403	412	421	430	280.1 + 4.453 t
WOMEN							
Trnava	270	299	304	308	312	316	248.4 + 2.139 t
Levice	330	324	327	330	333	336	288.2 + 1.510 t
Nitra	351	327	335	343	351	360	235.8 + 3.613 t
SLOVAKIA	299	289	293	297	302	307	240.8 + 1.962 t

In the table there is in the column "Situation" the last real value from an Oncologic Register for given region (before there were given values in 1968-90); in the section "Forecast" there are extrapolated values till the year 2000 according to added equations. We can see that in 2000 the number of indicated tumours in Trnava regions will be at the same level as in Nitra and Levice in about 1977. The same can be said with some caution about women cancer. The trend equation has an independent variable time (t). The direction gives a rate of increasing number of cancer tumours in given region. This coefficient makes us possible to easy compare differences among regions. So we can say that if in Trnava region the cancer cases grow with a factor 1.6, that factor is in Nitra equal to 5.9 and 3.8 in Levice.

The mortality is very similar in the whole Trnava region. The average percentage of premature deaths was 44% for men and 37% for women, the average number of PLYL was 15.5 years for men and 11.5 years for women. There are no significant differences in the whole Trnava region. There were 4 from 114 municipalities as for men as for women out of the central accumulation. These municipalities are either very small and these extremes are caused due to small numbers or there doesn't exist any relationship to NPPs which are very far and aren't in direction of wind or water ways which could affect these results.

Conclusion: The incidence of cancer tumours in Trnava region is one of the smallest in Slovakia and shows a very low growth in comparison as to the compared regions. The mortality in surroundings of NPP is similar one as in broad region. Results of this study do not prove any harmful influence of NPP to inhabitants even in the closest surroundings.

NONLINEAR WEIGHTED LEAST SQUARES CALIBRATION

Ján Mikuláš Lisý
Katedra fyzikálnej chémie CHTF STU, Radlinského 9
812 37 Bratislava, Slovakia

Summary: A simple and easy programmable interative algorithm for the nonlinear weighted least squares method suitable for nonlinear calibration has been developed.

Polynomial models are very often used in a curvilinear calibration. In many cases the constraints internally involved in the polynomial model may be in contradiction with the calibrating experimental data. For instance the quadratic polynomial has one extreme and its second derivative is constant. When treated data are in such contradiction with the model the estimates of the parameters may have spurious physical meaning. In spite of that the polynomials are used because they are linear in their parameters, i.e. their parameters can be easily estimated. Using presented algorithm the nonlinear (in parameters) models become as easy to handle as linear models without any kind of linearization [1,2,3].

According to the least squares method the model $f(x_i,a)$ is fitted using N experimental data (X_i, Y_i) with standard deviations σ_{Xi} and σ_{Yi} by minimization of the sum S,

$$S = \sum_{i=1}^{N} \left(R_i / \sigma_{Ri}\right)^2 , \qquad (1)$$

where σ_{Ri} is the standard deviation of the residual R_i: $R_i = Y_i - f(X_i,a)$. Minimization with respect to the M parameters a_j (j=1, 2, ... , M) of the model gives M nonlinear equations [3]:

$$\sum_{i=1}^{N} \frac{f(X_i,a)}{\sigma_{Ri}^2} \frac{\partial f(X_i,a)}{\partial a_j} = \sum_{i=1}^{N} \left\{ \frac{Y_i}{\sigma_{Ri}^2} \frac{\partial f(X_i,a)}{\partial a_j} + \frac{1}{2}\left(\frac{R_i}{\sigma_{Ri}^2}\right)^2 \frac{\partial \sigma_{Ri}^2}{\partial a_j} \right\}, \quad j=1, 2, \ldots , M \qquad (2)$$

For any model, which is linear with respect to each of its parameters, the last equations may be rewritten in matrix form:

$$\mathbf{B}.\mathbf{a} = \mathbf{g} . \qquad (3)$$

The matrix $\mathbf{B}$ is a symmetric matrix $M \times M$ with components

$$B_{jk} = \sum_{i=1}^{N} \frac{\partial f(X_i,a)}{\partial a_j} \frac{\partial f(X_i,a)}{\partial a_k} \frac{1}{\sigma_{Ri}^2} . \qquad (4)$$

The components of the vector g are:

$$g_j = \sum_{i=1}^{N} \left\{ \frac{Y_i}{\sigma_{Ri}^2} \frac{\partial f(X_i,a)}{\partial a_j} + \frac{1}{2} \left(\frac{R_i}{\sigma_{Ri}^2} \right)^2 \frac{\partial \sigma_{Ri}^2}{\partial a_j} \right\}. \quad (5)$$

The parameters a_j are the components of the vector **a**:

$$\mathbf{a} = (a_1, a_2, \ldots, a_M)^T \quad (6)$$

For the nonlinear model the matrix **B** is not symmetric. The number of its rows is equal to the number of its parameters. The number of its columns is equal only to the number of its linear parameters. By simple mathematical manipulations equations (2) can be rearranged in such a form for which matrix **B** is symmetric. Equations (2) are rewritten in the form:

$$0 = \sum_{i=1}^{N} \left\{ \left(\frac{Y_i - f(X_i,a)}{\sigma_{Ri}^2} \right) \frac{\partial f(X_i,a)}{\partial a_j} + \frac{1}{2} \left(\frac{R_i}{\sigma_{Ri}^2} \right)^2 \frac{\partial \sigma_{Ri}^2}{\partial a_j} \right\}, j=1, 2, \ldots, M. \quad (7)$$

To the left side and the right side of each equation of the set (7) the expression:

$$\sum_{k=1}^{M} a_k \left\{ \sum_{i=1}^{N} \frac{\partial f(X_i,a)}{\partial a_j} \frac{\partial f(X_i,a)}{\partial a_k} \frac{1}{\sigma_{Ri}^2} \right\} \quad (8)$$

is added. After that the right side of each equation denoted as G_j is:

$$G_j = g_j' + \sum_{k=1}^{M} a_k \left\{ \sum_{i=1}^{N} \frac{\partial f(X_i,a)}{\partial a_j} \frac{\partial f(X_i,a)}{\partial a_k} \frac{1}{\sigma_{Ri}^2} \right\}. \quad (9)$$

where by g_j' is denoted the right side of each equation of the set (7).

Then the solution for nonlinea model is in matix fom

$$\mathbf{B}.\mathbf{a} = \mathbf{G} \quad (10)$$

where matrices **B** and **a** are defined exactly in the same way as for linear models by equations (4) and (6). The components of the vector **G** are given by (9). Damping of the iterative solution in matrix form (10) and the other features of the method are the same as in the case of the polynomial models. The solution satisfies Deming's criteria and provides the variance-covariance matrix of the parameters.

References:

1. W. E. Deming, Statistical Adjustment of Data, Dover, N. York 1964.
2. P. Gans, Data Fitting in the Chemical Sciences, Wiley, N. York 1992.
3. J. M. Lisý et al., Comput. Chem., 15, 135 (1991).

COMPUTATIONAL PROBLEMS IN EVALUATION OF VAPOUR-PRESSURE OSMOMETRY AND OTHER COLLIGATIVE PROPERTY EQUILIBRIUM DATA

Milada LUBALOVÁ, Josef HAVEL

Department of Analytical Chemistry,
Faculty of Science, Masaryk University, Kotlářská 2,
611 37 Brno, Moravia, CZECH REPUBLIC
and
Maria MARTÍNEZ, Nuria MIRALLES, Ana SASTRE*

Department d'Enginyeria Química,
Universitat Politécnica de Catalunya
E.T.S.E.I.B., Diagonal, 647, 08028 Barcelona, SPAIN

Summary: *Computational problems in evaluation of colligative properties equilibrium data were studied by simulation on several noise levels and experimental data evaluated.*

Aggregation equilibria of solutes A, B, C, and etc.:

$$p\,A + q\,B + r\,C + ... \quad <\!\!=\!\!=\!\!> \quad A_pB_qC_r....$$

are often determined by colligative methods.

Colligative property data (vapour-pressure lowering, freezing point depression, etc.) are all based on the relations

$$Y = K_x.S$$

or

$$Y = K_{x1}.S + K_{x2}.S^2 + K_{x3}.S^3,$$

where S is sum of the concentrations of all solute species and Kx' s are calibration constants of a given instrumental set-up.

The computational problems and difficulties of vapour pressure osmometry equilibrium data evaluation have been studied.
Using both, simulated and experimental data with different noise levels, conditions for reliable computing of the equilibrium constants and the calibration parameters K_x were studied. For the computation general least squares optimizing program CPLET [1] was used. Three dimensional graphs of supersurfaces of the response function (the sum of squares of residuals) as a function of parameters were demonstrated. The limits of the method were studied and conditions for reliable computation of parameters were found.

Aggregation constants of di-(2-ethylhexyl)monothiophosphoric acid and trioctylmethylammonium chloride in toluene were determined.

References:

1. Havel J., Miralles N., Sastre A., Aguilar M., Computers Chem. **16**, 319-323 (1992).

QUANTITATIVE ANALYSIS OF TUNNELLING SPECTRA

Pavel Mach

Czech Technical University in Prague, Faculty of Electrical Engineering,
Department of Electrotechnology, Technická 2, 166 27 Prague 6, Czech Republic

Summary: Inelastic Electron Tunnelling Spectroscopy (IETS) is a method of vibrational spectroscopy that makes the measurement of spectra with the IR, Raman and tunnelling bands possible. This method has been used to the measurement of spectra of tetrametyltetrafenyltrisiloxan irradiated by energetic electrons. A form of spectral bands and their changes have been investigated by the use of a nonlinear regression completed with the Levenbergh dumping. An approximation by the Gaussian form of a band has been found the best. Changes of spectral bands after irradiation of material by different doses of electrons have been found, too.

The inelastic electron tunnelling spectroscopy has been developed in 1964. This method is based on the measurement of a nonlinearity of the tunnelling sample current versus voltage characteristics. The nonlinearity of the sample is very low and is strongly joined with the creating of an inelastic tunnelling channel through the junction. This channel is caused by vibrational states of molecules presented inside the junction. The measurement of the spectrum is usually carried out by a modulating technique. The sample is immersed directly into liquid He. Therefore the measurement is carried out at the temperature of 4.2 K. There is possible to find three types of bands in the tunnelling spectrum: the IR bands, the Raman bands and the tunnelling bands. The tunnelling bands do not correspond with bands measured by other types of vibrational spectroscopies. An IET spectrum analysis is usually realized by the same way as the analysis of the IR spectrum, mostly by the use of an IR spectra handbook or an expert system.
The IET spectroscopy is more sensitive in comparison with the IR one. There is possible to obtain spectra of high quality by the measurement of very thin films of material (thinner than monolayer).
The spectra are measured using a modulating technique. The tunnelling junction is powered by sinusoidal current of a low amplitude superposed on DC voltage which is slowly increased from 30 to 500 mV. A nonlinearity of junction causes a distortion of ac voltage across the junction. It is possible to derive that the second harmonics of this distorted voltage is proportional to the second derivative of the ac signal across the junction and represents the vibrational spectrum. The spectral band of such the spectrum can be described by the equation:

$$\frac{d^2 I_{in}}{dU^2} = n_0.(h\omega_0).[D * \chi * \psi] \qquad (1)$$

Here I_{in} .... inelastic current, U .... DC voltage across the junction, $n_0(h\omega_0)$ .... number of molecules taking part in the vibrational mode having the energy $h\omega_0$, D .... spectral weight function, χ .... modulating function that represents the broadening of the spectral band caused by the amplitude of the modulating signal, ψ .... temperature function that describes the thermal broadening of the band. The width of the band decreases with the decrease of the temperature and amplitude of the modulating signal. The general theory that describes the amplitude of the spectral band is not known till this time and therefore the band form is also not known.

If the IET spectroscopy is often used to the study of structural changes in material caused by different influences (e.g. by irradiation, ageing or any other influence). The

changes are investigated by the comparison of spectra after different doses of irradiation. The spectra are first decomposed in separate bands and then both the amplitude and area of corresponding bands are compared. This process is usually realized by a computer.

The measured spectrum is described by the equation:

$$S(\nu) = P_0(\nu) + \sum_1^i P_i(\nu) \qquad (2)$$

Here $P_0(\upsilon)$ .... background of the spectrum (is given by the elastic tunnelling),

$P_i(\upsilon)$ .... profile function that describes the form of the band indexed by i.

Three profile functions have been used to fitting of spectral bands: the Cauchy function, the Gaussian function and function given by the first two elements of mathematical series that represents the Cauchy function. Every band has been described by three parameters - by the position, the amplitude, and the FWHM (Full Width in the Half of the Maximum). The FWHM differs for different forms of the bands.

The spectral band is given by the convolution of three functions (see eq. 1) and one of them (spectral weight function) is not known. Therefore we have calculated the convolution of both the remained functions and compared the result with the profile functions. The best result has been achieved by the Gaussian function. It is obviously that the measured form of the bands will different in comparison with the calculated one because the spectral weight function has not been included into our calculation. However, it has been found by the analysis of measured spectra that the form of the IET bands is really very similar to the Gaussian function.

The modelling of spectra has been carried out by a nonlinear regression. The Newton-Raps method has been used. However, the iteration process is convergent in the case of a good primary estimation only. Therefore the nonlinear regression has been completed by the Levenbergh dumping. This combination creates a useful tool to the spectrum modelling. The program has been written in Turbo Vision language.

The spectra have been analyzed in two steps. At first the background of the spectrum has been found and subtracted. This step is very simple because the background of the IET spectrum is caused by elastic current and the current vs. voltage characteristics of the tunnelling junction is linear. Then the spectrum has been processed by the method mentioned above. The accuracy of the modelling depends on the number of iteration steps and has been evaluated according to the standard deviation between measured spectrum and model. There are three ways of modelling possible: to prescribe the number of iterations or to prescribe the maximum admissible value of the standard deviation, however, in our opinion the best way is to prescribe the maximum admissible value of the standard deviation and together the number of the iteration steps. When one of these conditions is realized the program is stopped.

We have investigated tetrametyltetrafenyltrisiloxan irradiated by energetic electrons. The modelling has been provided with the aim to analyze both the spectral bands form and area of corresponding spectral bands after the irradiation. The best results have been achieved when the bands have been fitted by the Gaussian form. The changes of bands area have also been analyzed. The most significant changes have been found in the bands corresponding with the CH_3 group.

t-TEST FOR SPREADSHEETS

Ivan Medonos

Department of Chemistry of Natural Compounds
Institute of Chemical Technology, Technická 5, 166 28 Prague 6, Czech Republic
(Present address: WATREX, Hošťálkova 42, 169 00 Prague 6, Czech Republic)

Summary: The Student's t-test algorithm is used for evaluation of up to ten series of data. Data editation and evaluation proceeds in spreadsheet.

The data table was designed and the related macros was developed for the evaluation of the data series, which are to be compared by Student's *t*-test. The original design developed for Quattro Pro 2.0 (DOS) was transcribed for environment of Microsoft Excel 5.0 (Windows). In this case the macros are written in Visual Basic.

The **input table** (= the part of the spreadsheet - up to 10 data columns, up to 100 data rows) allows common editation functions of spreadsheet in data columns (adding; deleting, copying, moving... of data), in the lower part the most popular statistical values are shown (average, deviation...).

The macro is written for the visualization of the **result table**. In this part of the spreadsheet each data series is compared with each other and this information is displayed: *t*-parameter computed from the data series, critical *t*-parameter from table of critical data and the result YES/NO, which means, that these two data series (=samples) are(/are not) the subsets of the same population.

For the possibility of comparing each of real data series with the data imitation (where only the number, mean and deviation are known) the special row is added in this part.

The table of critical Student's *t*-values for 95% probability is added. In this case the program computes with such probability.

EXPLORATORY AND CONFIRMATORY DATA ANALYSIS

Milan Meloun and Jiří Militký [1)]
Department of Analytical Chemistry, University Pardubice,
532 10 Pardubice, Czech Republic
and
[1)]Department of Textile Materials, Textile Faculty, Technical University,
461 17 Liberec, Czech Republic,

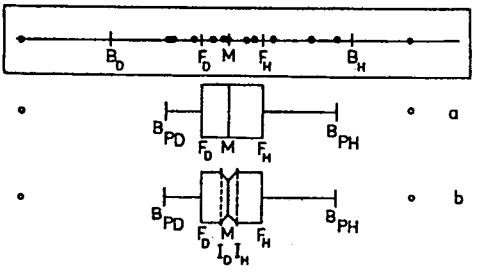
Summary: Exploratory data analysis (EDA) isolates certain statistical features and patterns of data with the use of the quantile plot, the dot and jittered dot diagrams, the (notched) box-and-whisker plot, the midsum plot, the symmetry plot, the curtosis plot, the differential quantile plot and the quantile-box plot, the kernel estimation of probability density function, the histogram, the frequency polygon, the bar chart, the rootgram, the quantile-quantile plot, the rankit plot and the conditioned rankit plot. The power and Box-Cox transformation is often used.

The first step of univariate data analysis called an exploratory data analysis (EDA) isolates certain basic statistical features and patterns of data while the second step, a confirmatory data analysis (CDA) stresses evaluation of an available evidence by tests of probability models. According to Tukey the EDA is a "detective work". Its tools are various descriptive graphically oriented techniques which are typically free of strict statistical assumptions about data. The EDA-techniques are often called "distribution-free" and are based on a continuity and differentiability of underlying density only. The EDA techniques are quite effective for an investigation of statistical behavior of data from new or non-standard analytical procedures.

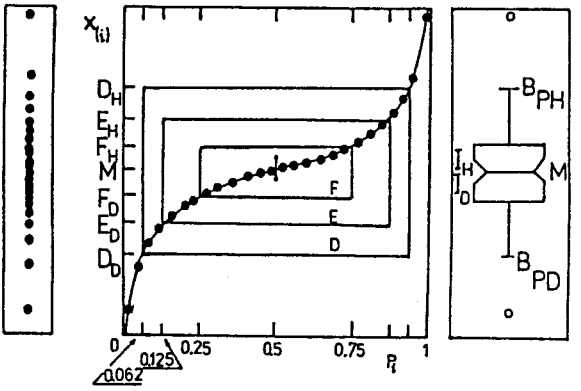
For graphical visualization of data the EDA uses *the quantile plot, the dot and jittered dot diagrams, the (notched) box-and-whisker plot* while the sample distribution is investigated by *the midsum plot, the symmetry plot, the curtosis plot, the differential quantile plot and the quantile-box plot.* The construction of sample distribution i.e. the estimation of probability density function is done by *the stem-and-leaf display, the kernel estimation of probability density function, the histogram, the frequency polygon, the bar chart, the rootgram, the quantile-quantile plot, the rankit plot and the conditioned rankit plot.*

When an exploratory data analysis (EDA) finds that the sample distribution differs from a normal one or when a confirmatory data analysis (CDA) does not prove a sample independence and a sample homogeneity, the original data should be transformed. The power transformation and the Box-Cox transformation improves a sample symmetry and stabilizes a sample variance. *The Hines-Hines selection graph* and *the plot of logarithm of maximal likelihood function* enables to find an optimum power transformation. Diagnostics of the EDA and CDA about assumptions of the actual data set are also examined.

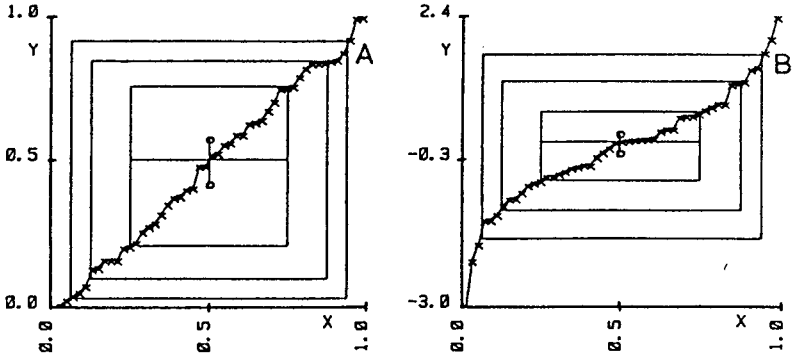
According to results of an examination about sample assumptions the classical, robust or adaptive estimates of location and spread are calculated.



Construction of (a) the box-and-whisker plot and (b) the notched box-and-whisker plot from the dot diagram. Empty circles indicate outliers



An example of a quantile-box plot. The dot diagram (left) and the notched box-and-whisker plot (right) are given for comparison



The rankit plot for samples from (A) rectangular and (B) normal distributions

REGRESSION DIAGNOSTICS IN NONLINEAR CALIBRATION

Milan Meloun and Jiří Militký[1)]
Department of Analytical Chemistry, University Pardubice,
532 10 Pardubice, Czech Republic
and
[1)]Department of Textile Materials, Textile Faculty, Technical University,
461 17 Liberec, Czech Republic,

Summary: Linear and nonlinear calibration contains steps: model proposition, exploratory data analysis, parameter estimation, regression diagnostics, construction of improved model, quality of calibration model and determination of unknown concentration. Detection limit, determination limit and critical level (blank) are estimated together with interval estimate of unknown concentration.

Calibration in analytical chemistry consists from two steps, a construction of calibration model and an inversion of calibration model when from measured signal y^* (e. g. absorbance) the unknown concentration x^* including its confidence interval is estimated:

(i) In case of linear models both steps use the calibration straight line. The straight estimate of unknown concentration x^* is estimated by

$$\hat{x}^* = \bar{x} + \frac{(y^* - \bar{y})}{b_1}$$

where y^* is the measured signal (or the average $\bar{y}^*$ for $M > 1$ repeated measuremets, respectively) and b_1 is the estimate of the slope. This estimate is generally biased and a correction is made by Naszodi modified estimates

$$\hat{x}_B^* = \bar{x} + \frac{(y^* - \bar{y})\, b_1}{b_1^2 + \dfrac{\sigma^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}}$$

Kruchthof proposed the inversion estimate while Schwartz the nonlinear estimate, details *cf. ref.* [2].

In the construction of confidence intervals of the estimate $\hat{x}^*$ or $\hat{x}_B^*$ for more scattered data besides the simplest confidence limits there are several approaches of calculation:

(ii) In case of nonlinear calibration models the linear spline, quadratic spline and cubic spline are used. There are three possibilities to choose knots in calibration curve.

(iii) To express the precision of a calibration, limiting values of the concentration for which the measurement signal is still significantly different from the noise are usually defined. For the precision and sensitivity of calibration methods, three levels of signal are identified:

(a) The *critical level* y_C represents the upper limit of the $100(1 - \alpha)\%$

confidence interval of the predicted signal from the calibration model for the concentration equal to zero i.e. the *blank measurement.* Signals above this critical level y_C are significantly different from the noise. The concentration x_C corresponding to this critical level y_C is determined from $x_C = (y_C - \bar{y})/b_1 + \bar{x}$.

(b) The ***detection limit*** y_D corresponds to the concentration for which the lower $100(1 - \alpha)\%$ confidence interval of signal prediction from the calibration model is equal to y_C,

$$y_D = y_C + \sigma . t_{1-\alpha/2}(n - 2) \sqrt{1 + \frac{1}{n} + \frac{(x_D - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}}$$

(c) The ***determination limit*** y_s is the smallest signal level for which the relative standard deviation of prediction from the calibration model is sufficiently small and equal to the constant number C, where C is 0.1 usually.

Procedure of the (linear and nonlinear) calibration usually contains steps:

(1) **Proposed model:** starting from the simplest model, models of higher power are build.

(2) **Exploratory data analysis:** using diagnostic graphs the multicollinearity, heteroscedasticity, autocorrelation, normality of errors, influential points i.e. outliers and extremes, are indicated.

(3) **Parameter estimation:** using the methods of the least squares or the rational ranks, the best estimates of unknown parameters are estimated. Statistical test of significance of each parameter follows. Quality of the regression model suggested is examined using regression characteristics: the mean error of prediction, the Akaike information criterion, the determination coefficient, the predicted determination coefficient and the standard deviation of prediction.

(4) **Regression diagnostics:** an identification of influential points and an examination of seven assumptions of the least squares method is carried out by the examination of the regression triplet (i. e. *data, model, method*).

(5) **Construction of improved model:** the estimation of parameters of improved model using various modifications of the least-squares is performed.

(6) **Quality of calibration model:** the estimates of the model parameters and precision limits of calibration are calculated.

(7) **Determination of unknown concentration:** the point and interval estimates of unknown x^* are calculated.

References:

1. J. Milităký and M. Meloun, Anal. Chim. Acta 277, 267 - 271 (1993);
2. M. Meloun, J. Milităký and M. Forina, *CHEMOMETRICS FOR ANALYTICAL CHEMISTRY, Volume 1. PC-Aided Statistical Data Analysis,* and *Volume 2. PC-Aided Regression and Related Methods,* Ellis Horwood, Chichester, 1992 and 1994;
3. ADSTAT, TriloByte Statistical Software Ltd, Pardubice.

NONLINEAR REGRESSION ON PERSONAL COMPUTERS

Jiří Militký, Milan Meloun[1)] and Karel Kupka[2)]
Department of Textile Materials, Textile Faculty, Technical University,
461 17 Liberec, Czech Republic,
[1)]Department of Analytical Chemistry, University Pardubice,
532 10 -Pardubice, Czech Republic
and
[2)]Trilobyte Statistical Software Ltd,
530 02 Pardubice, Czech Republic

Summary: The paper describes basic numerical techniques for the nonlinear model building by using of a regression method. Some procedures for the nonlinear least squares minimization are discussed. Comparison of numerical suitability of leading statistical packages for personal computers is presented.

The main task in construction of regression models is to estimate a parameter vector a $(m \times 1)$ in known nonlinear model $f(x, a)$. The estimation process is based on experimental data $\{y_i, x_i\}$, $i = 1, ..., N$. The values x_i (without detriment to generality x_i is supposed to be a scalar) creates nonstochastic set. The measured values y_i are usually supposed to be expressed in frame of additive measurement model

$$y_i = f(x_i, a) + \epsilon_i, \qquad i = 1, ..., N \qquad (1)$$

Here ϵ_i represents the so-called experimental random error, presumably independent identically distributed random variables having normal distribution $N(0, \sigma^2)$. Considering these assumptions the estimates a^* of the parameter vector a may be found by minimizing the least squares (LS) criterion

$$S(a) = \| y - f(a) \| \qquad (2)$$

where y is $(N \times 1)$ the vector of measured quantities, $f(a)$ is $(N \times 1)$ the vector of corresponding model values (the i-th component is $f(x_i, a)$) and $\| . \|$ is the Euclidean norm.

Generally, the nonlinear regression problem may be divided into two parts:
- minimization of LS criterion $S(a)$,
- statistical analysis of estimates, model and data quality.

In most of the programs the linearization of nonlinear model $f(x, a)$ is used for both, the minimization and the statistical analysis. This approach may be observed even in cases where it is not acceptable due to high nonlinearity of model $f(x, a)$. In sequel the problem of minimization of LS criterion is discussed.

For estimation of parameters by minimizing of the LS criterion $S(a)$ the derivative algorithms are profitably being used. General scheme of minimization is expressed by the following sequence of steps:

1. Input of first estimates.
2. Finding a suitable directional vector v_j.
3. Determination of the scalar α_j such that the increment $d_j = \alpha_j \; v_j$ is acceptable.
4. Testing whether the minimum has been reached. If no minimum has been found, the new value $a_{j+1} = a_j + d_j$ is computed and new iteration step (step 2) is started.

To find an optimum scalar α Taylor expansion up to the quadratic terms is substituted

for $S(a)$ in v direction. After analytical minimization with respect to α the well-known Raleigh factor results.

The notorious Gauss-Newton method is based on the same expansion where the second derivative terms are omitted. The following relationship results

$$v_{GN} = (J^T J)^{-1} J^T e \qquad (3)$$

Here e is the vector of residuals having the components $e_i = y_i - f(x, a)$ and J $(N \times m)$ is the Jacobian matrix with the components

$$J_{jk} = \frac{\mathrm{d}\, f(x_j, a)}{\mathrm{d}\, a_k}$$

Numerical experiments have shown that the Gauss-Newton method solves the nonlinear LS problem efficiently when the residuals are sufficiently small at the solution, but it can be less efficient if the residuals are significantly nonlinear and their optimal values are large.

In the MNOPT program the modification of the so-called *double dogleg strategy* of Dennis and Mei is applied. Optimum directional vector v_D is here sought on the line segment TB of a triangle OTB defined by its vertices

$$O \equiv a_j; \qquad T \equiv a_j + \alpha_1 v_{GN}; \qquad B \equiv a_j - \alpha g$$

Here g is the gradient vector of $S(a)$. The vector v_D can be then expressed in the form

$$v_D = \mu \alpha_1 v_{GN} + (1 - \mu) \alpha v_G \qquad (4)$$

Parameter μ lies in the interval (0, 1) and defines the position on the line segment TB. Parameters α and α_1 are computed as a Raleigh factors.

This procedure requires in each iteration only one inverse (in computing $(J^T J)^{-1}$) to be done. Moreover, maximum step length is here limited with respect to admissible region where the quadratic expansion of $S(a)$ is acceptable.

The MNOPT procedure is compiled allowing to do besides effective numerical estimation of regression parameters also their statistical analysis dependent upon the size of bias (model nonlinearity). For the minimization of LS criterion $S(a)$ the *double dogleg strategy* is carried out.

Instead of inverting the matrix J^TJ the special pseudoinverse technique is employed (a variant of the "rational rank" method). Owing to this minimization is accomplished even in cases where J^TJ is singular or nearly singular. For evaluation of the maximum acceptable step length the heuristic procedure based on the "trust region" is used. The program is completed by a *grid search* routine applied if no acceptable directional vector v can be found.

Numerical accuracy of the MNOPT has been compared with some well known statistical packages on the base of extensive published and simulated test problems. It will be shown that the most of the tested packages are not able to solve these test problems. Results of test study clearly show that the MNOPT is one of the best.

References:

1. M. Meloun, J. Militký and M. Forina: *CHEMOMETRICS FOR ANALYTICAL CHEMISTRY, Volume 1. PC-Aided Statistical Data Analysis,* and *Volume 2. PC-Aided Regression and Related Methods,* Ellis Horwood, Chichester, 1992 and 1994;
2. ADSTAT, TriloByte Statistical Software Ltd. Pardubice

SOLUTION OF INVERSE NONISOTHERMAL KINETIC PROBLEM

Jiří Militký
Dept. of Textile Materials, Faculty of Textiles,
Technical University, Liberec, Czech Republic

Summary: In this contribution the selected features of inverse noniso-thermal kinetic problem solving are presented. The corresponding statistical problems connected with formulation of regression criterion for parameters estimation are discussed.

There exists a lot of methods for solving of the inverse kinetic problems. For correct statement of optimization target function (regression criterion) the various statistical question (i. e. model of measurements, statistical nature of errors etc.) must be answered.

Description of chemical kinetics is based on equations defining the relations between rate of conversion $\dot{\alpha}$, rate of temperature changes $\dot{T}$ and state of investigated system represented by variables (α, T).

For creation of the nonisothermal kinetic models the factorized rate equation is usually used

$$\dot{\alpha} = \frac{d\alpha}{dt} = k(T,\underline{\beta})\, f(\alpha,\underline{\delta}) \qquad (1)$$

In this eqn the α denotes degree of conversion ($0 \leq \alpha \leq 1$) and t is corresponding time of reaction.

Temperature term k(T, $\underline{\beta}$) is dependent on temperature only. Model parameters $\underline{\beta}$ are thermodynamical characteristics of investigated reaction (activation energy, entropy etc.).

Kinetic term f(α, $\underline{\delta}$) is dependent on degree of conversion only. Model parameters $\underline{\delta}$ are connected with kinetic model of investigated reaction.

The temperature is usually linear function of time of reaction

$$T(t) = T_0 + \Phi t \qquad (2)$$

where T_0 is starting temperature and Φ is rate of heating. Often the integral data (α_i, T_i), i = 1, ..., N are only available. The formal integration of eqn (1) is then adopted. Resulting relation has the form

$$g(\alpha,\underline{\delta}) = F(t,\underline{\beta}) \qquad (3)$$

where

$$g(\alpha,\underline{\delta}) = \int_0^{\alpha} \frac{d\alpha}{f(\alpha,\underline{\delta})}; \quad F(t,\underline{\beta}) = \int_0^{t} k[T(t),\underline{\beta}]\,dt$$

degree of conversion α can be simply determined from the relation

$$\alpha = g^{-1}[\underline{\delta},F(t,\underline{\beta})] \qquad (4)$$

In model (4) the parameters $\underline{\beta}$ and $\underline{\delta}$ can be estimated on the base of experimental data. This task is denoted as an inverse kinetic problem (IKP).

In sequel we assume that the aim is estimation of the parameter vector $\underline{p}=(\underline{\beta},\underline{\delta})$ for known integral model $g^{-1}(t)$.

Commonly used least squares (LS) criterion has the form

$$S(\underline{p}) = \sum_{i=1}^{n} [\alpha_i - g^{-1}(F(t_i),\underline{p})]^2 \qquad (5)$$

Parameter estimates $\underline{\hat{p}}$ can be obtained by minimizing of S(p). The LS criterion is not generally applicable and its effectiveness is dependent on the very strong assumptions about measured variable α and errors of measurements.

In the modelling of nonisothermal kinetics these assumptions are not often acceptable.

1. Model of measurements is not additive but often multiplicative. The statistical technique for discrimination between additive and multiplicative measurement models wil will be proposed.

2. The relative measurements errors are commonly constant and error variances $D(\epsilon_i)$ are increasing function of measured quantity α.

3. Independence of measurements in nonisothermal kinetic experiments cannot be generally accepted. Due to special experimental arrangements (measuring on the one system only) the errors caused by process conditions variation (e.q. thermal fluctuations) are cumulative.

In the case of precise measurement device (d_i are small) the least squares criterion must be replaced by the first difference form

$$S_D = \sum_{i=2}^{N} \{(\alpha_i-\alpha_{i-1})-[g^{-1}(F(t_i),\underline{p})-g^{-1}(F(t_{i-1},\underline{p})]\}^2 \quad (6)$$

Full derivation of S_D and other special variants of regression criteria will be presented. It can be concluded that the more realistic assumptions about data and their origin lead to the more general estimation problem. For effective solution of these tasks the special nonlinear optimization methods will be proposed.

The practical applicability of proposed techniques will be demonstrated on the simulated and real data (crystallization kinetics of special glass).

NONISOTHERMAL THERMOANALYTICAL KINETIC MODELS BUILDING

Jiří Militký and Jaroslav Šesták[1)]
Dept. of Textile Materials, Faculty of Textiles,
Technical University, Liberec, Czech Republic
and
[1)]Institute of Physics, Czech Academy of Science,
Prague, Czech Republic

Summary: Main features of nonisothermal kinetic models building are discussed. The aplication of the partial regression graph concept for kinetic and temperature terms evaluation are described.

Simplified description of chemical kinetics is based on equations defining the relations between rate of conversion $\dot{\alpha}$, rate of temperature changes $\dot{T}$ and state of investigated system represented by variables (α, T).

For nonisothermal thermoanalytical data treatment the factorized mdel is often assumed

$$\dot{\alpha} = \frac{d\alpha}{dt} = k(T,\underline{\beta})f(\alpha,\underline{\delta}) \qquad (1)$$

In this retation the α denotes degree of conversion ($0 \leq \alpha \leq 1$) and t is corresponding time of reaction.

Temperature term k(T, $\underline{\beta}$) is dependent on temperature only. Model parameters $\underline{\beta}$ are thermodynamical characteristics of investigated reaction (activation energy, entropy etc.).

Kinetic term f(α, $\underline{\delta}$) is dependent on degree of conversion only. Model parameters $\underline{\delta}$ are connected with kinetic model of investigated reaction.

Relation (1) is based on a formal assumption that the temperature dependent in nonisothermal kinetic model is rate constant only.

The temperature is usually linear function of time of reaction

$$T(t) = T_0 + \Phi t \qquad (2)$$

where T_0 is starting temperature and Φ has meaning of rate of heating or cooling.
From results of thermoanalytical measurements it is possible to obtain information about α, $\dot{\alpha}$ and T simultaneously, i.e. the data

$$(\dot{\alpha}_i, \alpha_i, T_i) \quad i=1,..N$$

are determined. For nonisothermal models building is then possible to use eqn (1) directly.

In thermal analysis a number of candidate nonisothermal kinetic models obviously exists and the selection of suitable one is very difficult. The main problem lies in fact, that

computed thermodynamical parameters (activation energy) are dependent on used kinetic model, too.

For specification of kinetic $f(\alpha, \underline{\delta})$ and temperature $k(T, \underline{\beta})$ term is suitable to use partial regression graphs combined with broad spectrum of possible models. We propose generalized kinetic model including a lot of kinetic and temperature terms in the form

$$\frac{d\alpha}{dt} = \beta_1 T^{\beta_2} \exp(-\frac{\beta_3}{RT})\, \alpha^{\delta_1}(1-\alpha)^{\delta_2}[-\ln(1-\alpha)]^{\delta_3} \qquad (3)$$

Let the error of $\dot{\alpha}_i$ or α_i measurements are multiplicative. Then the eqn (3) can be correctly logaritmically transformed into linear regression model. In matrix notation has the form

$$\underline{y} = X\,\underline{A} \qquad (4)$$

where $\underline{y}$ is (N x 1) vector, X is (N x 6) matrix and $\underline{A}$ is (6 x 1) unknown vector.

For investigation of partial linearity between $\underline{y}$ and j-th column $\underline{X}_j$ of matrix X the projection into space L orthogonal to space defined by columns of matrix $X_{[j]}$ is used. Matrix $X_{[j]}$ is created by dropping of j-th column $\underline{x}_j$ from matrix X.

Corresponding projection matrix into space L has the form

Let

$$P_{[j]} = E - X_{[j]} (X_{[j]}^T X_{[j]})^{-1} X_{[j]}^T \qquad (5)$$

$$\underline{u} = P_{[j]}\,\underline{y}; \quad \underline{v} = P_{[j]}\,\underline{x}_j$$

Partial regression graph is then dependence of vector $\underline{u}$ on vector $\underline{v}$. If the term $\underline{x}_j$ is correctly specified the partial regression graph forms straight line. Systematic nonlinearity is indication of incorrect specification of $\underline{x}_j$ and random pattern shows unimportance of $\underline{x}_j$ for explaining the variability of $\underline{y}$.

Practically it is possible by selecting of β_2, δ_1, δ_2 and δ_3 to specify some kinetic model. From partial regression graph it is then suitable to investigate his correctness.

This approach has been applied for simulated data i.e. combination of the Arrhenius term with $\ln \beta_1 = 23.03$; $\beta_2 = 0$; $\beta_3 = 200$ kJ mol^{-1} and JMA term with $\delta_1 = 0$; $\delta_2 = 1$ and $\delta_3 = 0.5$). These simulated data have been used for investigation of incorrect specification of temperature term on shape of corresponding partial regression graphs. In the correct case, the combination of JMA kinetic term and Arrhenius one was used for creation of partial regression graph. In the incorrect case the Arrhenius term was replaced by simple exponential.

It will be show that incorrect specification of temperature term leads also to the strong nonlinearity in partial regression graphs.

LIMITS OF DETECTION AND QUANTIFICATION - - A STATISTICAL OVERVIEW AND NEW APPROACHES

Jan Mocak[1], Alan M. Bond[2]

[1]Department of Analytical Chemistry, Slovak Technical University, SK-81237 Bratislava, Slovak Republic
[2]Department of Chemistry, Monash University, Clayton, Victoria 3168, Australia

Summary: Statistical analysis of the problems in evaluating the limits of detection, LOD, and the limit od quantification, LOQ, is presented in agreement with new approaches utilising the upper confidence limit of the individual blank signal observation and taking into account uncertainty in the position of the calibration line.

An important task in characterising the performance of any trace analysis procedure is the specification of its limit of detection, LOD, and limit of quantification (determination), LOQ. Despite of the long time existing LOD and LOQ definitions, approved by IUPAC and the American Chemical Society, their correct statistical interpretation under real experimental conditions is still rather problematic. Therefore in this work an assessment of the LOD and LOQ will be presented from the standpoint of a correctly and unambiguously calculating LOD and LOQ, and relevantly, new methodologies will be given.

Although some authors consider the LOD exclusively in the signal domain, the real goal in trace analysis is to obtain the LOD and LOQ *concentration* values (c_D and c_Q). The conversion from the signal to the concentration domain is commonly made by projection of the signals y_D and y_Q, related to the LOD and LOQ, and *corrected for blank*, through the *calibration plot* (obtained by regression), onto the concentration axis. Usually a linear calibration model $y = q_0 + q_1 c$ is assumed and in this case the concentrations c_D and c_Q, calculated by projection, are influenced by the errors of the intercept q_0 and slope q_1 (*analytical sensitivity*). There are two reasons for a non-zero intercept q_0: (*a*) the use of the mean blank signal $\bar{y}_b$ instead of the "true" blank signal value μ_b, and (*b*) the uncertainty of the regression line position. The problem (*a*) can be overcome by using the Student's $t(\nu, \alpha)$ instead of the commonly used coefficient $k_D=3$ (ν being the number of degrees of freedom, e.g. ν_b in the blank measurements) and the term $(1+1/n_b^{1/2})$ in the definition of the y_D:

$$y_D = \bar{y}_b + t(\nu_b, \alpha)(1+1/n_b^{1/2})\, s_b \qquad (1)$$

The problem (*b*) is associated with the fact that the "true" position of any regression line is given by generally unknown population regression coefficients.

However, a *confidence band*, which expresses a set of confidence intervals of the signal around the regression line, can be computed by means of the t- statistic. With this approach, the intersection of the appropriate (i.e. LOD or LOQ) signal value with the upper bound of the confidence band determines the relevant concentration LOD or LOQ value. Calculation for the given numbers of freedom ν (given by the number of observations minus number of regression parameters) and the confidence level α is possible not only for both existing straight-line calibration models but also for a non-linear calibration model.

Differences between the classical IUPAC approach and our new Upper Limit Approach can be briefly demonstrated as follows:

A. Classical IUPAC approach. The most common application based on the IUPAC and ACS definitions employs the mean blank signal, $\bar{y}_b$, as the reference point value for the calculation of the signal LOD and LOQ values, y_D and y_Q, regardless of the intercept q_0 position of the calibration plot. Then the concentration LOD and LOQ values are:

$$\mathbf{LOD} = (y_D - \bar{y}_b) / q_1 ; \qquad \mathbf{LOQ} = (y_Q - \bar{y}_b) / q_1 \qquad (2a,b)$$

If the condition $q_0 > 0$ is valid (or $q_0 > \bar{y}_b$ for the gross signals plot) then the found LOD and LOQ concentration values are overestimated, i.e. they may be larger than they should be. If $q_0 < 0$, (or $q_0 < \bar{y}_b$ for the gross signals) then the LOD and LOQ values may be smaller than the "true" values. The conditionality of these statements follow from the fact that the incorrect position of the calibration curve may either compensate for the above mentioned effect or, alternatively, cause the error to be enhanced. Moreover, the error described in the problem (*a*) has to be concerned.

B. Upper limit approach. The upper limit approach, ULA, utilises the upper bound of the signal vs. concentration confidence band for obtaining the concentration counterparts of the signal LOD and LOQ values, thus the uncertainty of the regression line is taken into account. *One-sided* upper confidence limit at $100(1 - \alpha)$ % probability level is consistent with the IUPAC definition of LOD and LOQ. Estimation of the *concentration* LOD and LOQ values is calibration model dependent. For the general straight line the results are:

$$\mathbf{LOD} = \{ t(\nu, \alpha)\, s_y / q_1 \} [1 + 1/n + \bar{c}^2 / \Sigma(c_i - \bar{c})^2]^{1/2} \qquad (3)$$

$$\mathbf{LOQ} = 3 \{ t(\nu, \alpha)\, s_y / q_1 \} [1 + 1/n + \bar{c}^2 / \Sigma(c_i - \bar{c})^2]^{1/2} \qquad (4)$$

where s_y is the residual standard deviation, c_i - the concentration coordinates of the calibration points, and $\bar{c}$ denotes their mean. The way of the LOQ calculation, consistent to the LOD calculation, will be demonstrated in detail in the lecture.

NON-TRADITIONAL WAYS OF SIMULATION OF VOLTAMMETRIC SIGNALS

Jan Mocak[1], Alan M. Bond[2]

[1]Department of Analytical Chemistry, Slovak Technical University,
SK-81237 Bratislava, Slovak Republic
[2]Department of Chemistry, Monash University, Clayton,
Victoria 3168, Australia

Summary: The possibilities of the mathematical solution of physical processes by infinite series are extended when using the algorithms transforming diverging or very slowly converging series into well behaved ones. The application in a broad scale of voltammetric conditions is presented for single sweep voltammetry and partly also for cyclic voltammetry.

The theory of linear sweep and cyclic voltammetry is based on solving Fick's law boundary value problems, involving various cases of simultaneous diffusion, charge transfer, and, sometimes, a coupled chemical reaction. Except the most simple cases, a numerical solution (mainly digital simulation) has been used for the signal calculation under different conditions. Such a solution is sufficiently flexible, but its precision is limited, even though it is generally better than that of a real experiment.

An alternative way, used in this paper, is based on the infinite series solution. Such an approach, considered as an analytical solution, is straightforward, independent on the previous points of the current - potential dependence, and very accurate in the potential region where the series is convergent; usually before the peak. Unfortunately, entire convergence of the respective power series is exceptional and attempts to use this approach in the region of a considerable divergence failed or produced erratic results. Therefore we have succesfully applied several linear and nonlinear lozenge algorithms of the series transformation (generalised Euler transform, Eta and Epsilon algorithms), elaborated for changing divergent series into convergent ones and accelerating their convergence. This is important especially in cyclic voltammetry where the calculated response is composed of two terms, one of them being the dimensionless current at the potential far beyond the peak, necessary for constructing the base-line of the backward process. Another important case of this kind is the treatment of two-step redox processes.

We have found the application of infinite series in calculating voltammograms describing diffusion in finite space as hitherto most successful. It is substantial for thin-layer electrochemistry, voltammetry of intercalation processes, polymer- or adsorbed electrode films. Under limited diffusion field conditions there exist four limiting cases in voltammetry, dependent on the values of dimensionless parameters L and λ. These express the film thickness and extent of reversibility, resp.:

$$L = l\,[nFv/(RTD)]^{1/2}\,; \qquad \lambda = k^{0}\,[RT/(nFvD)]^{1/2} \qquad (1a,b)$$

where **D** denotes diffusion coefficient in limited diffusion space, k^0 denotes heterogeneous rate constant at formal potential, **l** - real thickness of diffusion space, and **v** - scan rate. For large values of **L**, calculated current - potential curves have the same form as prescribed for conventional linear sweep voltammetry with reversible or irreversible electrode reaction. On the other hand, for low **L** values, a thin-layer electrochemistry behaviour can been found as the second extreme. The shape, magnitude, and position of a voltammogram depend again not only on parameter **L**, but in the case of irreversible reaction also on heterogeneous rate constant and transfer coefficient α. Dimensionless currents (current functions) will be referred as I^*_{11}, I^*_{10}, I^*_{22}, and I^*_{20} for the cases (a) thick film, reversible electrode reaction, (b) thick film, irreversible electrode reaction, (c) thin film, reversible electrode reaction, (d) thin film, irreversible electrode reaction, resp.:

$$I^*_{11} = -\sum j^{1/2} (-z)^j \qquad (2)$$

$$I^*_{10} = -\sum (-z)^j / [(j-1)!\,]^{1/2} \qquad (3)$$

$$I^*_{22} = -\sum j\,(-z)^j \qquad (4)$$

$$I^*_{20} = -\sum (-z)^j / [(j-1)!\,] \qquad (5)$$

where $z = \exp(E^*)$, E^* denotes dimensionless potential $(E - E^0)(nF/RT)$ for reversible processes, and $(E - E^0)(\alpha n_a F/RT) + \ln[(\pi D \alpha n_a F v/RT)^{1/2}/ k^o]$ for irreversible ones; the symbol $\sum$ denotes summation from $j = 1$ to $j = \infty$.

Unlike the currents I^*_{10} and I^*_{20}, all real voltammetric currents for irreversible processes are smaller compared to reversible ones since they also depend on transfer coefficient α (from the interval 0 to 1). Interestingly, regardless the electrochemical reversibility, the real currents depend on $v^{1/2}$ in a thick film situation (common voltammetry), whilst in the thin film case depend on v^1.

For description of a possible multi-dimensional diffusion (e.g. in the intercalation processes) it is also important to consider the dimensionality factor *d* describing the one-, two-, or three- dimensional diffusion, resp., in the host lattice. Calculated dependences of peak current and peak potential on the parameters **L** and *d* (*d* = 1, 2 or 3) reveal their influences upon the shape and maximum current of the voltammogram. When *d* is higher, the current changes with **L** are slower and the peak is less sharp. The infinite series with an appropriately incorporated "transfer" function enable the treatment of intermediate cases, i.e. the films of the intermediate thicknesses. The transfer function, multiplicating each series term, is $\tanh(j^{1/2}L)$, $\coth(j^{1/2}L) - 1/(j^{1/2}L)$, and $[I_1(j^{1/2}L)/I_0(j^{1/2}L)]$ for *d* = 1, 2, and 3, resp., with I_1 and I_0 denoting hyperbolic Bessel functions of the order 1 and 0, resp.

Several further non-traditional voltammetric simulations will also be discussed.

DQLIMITS - A COMPUTER PROGRAMME FOR CALCULATION OF DETECTION AND QUANTIFICATION LIMITS

Jan Mocak

Department of Analytical Chemistry, Slovak Technical University,
SK-81237 Bratislava, Slovak Republic

Summary: The limit of detection (LOD) and limit of quantification (LOQ) are calculated using traditional methods, based on the IUPAC and ACS definitions, as well as a new approach, in which the uncertainty of the calibration plot position is concerned. Comparison of electrochemical trace analysis methods will be used for demonstration.

Our computer programme **DQLIMITS** enables calculation of the limit of detection, LOD, and the limit of quantification (determination), LOQ, using four ways: (A) Classical way based on the most common interpretation of the IUPAC [1] and the American Chemical Society definitions [2]. (B) Practical approach by Miller & Miller [3] following out the regression analysis of the calibration plot. (C) Our new upper limit aproach (ULA2) utilising a calibration line with a non-zero intercept [4]. (D) Our new upper limit aproach (ULA1) utilising a calibration line forced to pass through the origin [4]. The calculations are not cumbersome, as it ought to be since they are planned to be used by a large number of users. However, the use of a computer makes the calculations more comfortable, precise and errorless. The main task is, therefore, to explain the way how the calculations are performed.

Although it is not the ultimate goal, the *signal values* of the LOD and LOQ are computed first. The signal value, y_D, corresponding to the *limit of detection*, must reflect the value of the true signal (related to some non-zero analyte concentration) which is significantly different from the blank signal value. The IUPAC convention [1] defines this y_D signal value in terms of the population mean, μ_b, the population standard deviation of the blank signal, σ_b, and using the factor $k_D = 3$, as:

$$y_D = \mu_b + k_D \sigma_b \qquad (1)$$

Similarly, the *limit of quantification*, LOQ, referring to the smallest concentration or the quantity which can be quantitatively analysed [2], is related to the signal

$$y_Q = \mu_b + k_Q \sigma_b \qquad (2)$$

The recommended use of the factor $k_Q = 10$ has no statistical inference, therefore our way of deriving this signal (explained in detail later) is coupled with the factor $k_D = 9$, which creates a logical link to the signals y_D and y_I, the signal defined by the factor $k_I = 6$ at which the analyte is surely detected.

In analytical chemistry practice only a very limited number of observations of the blank signal, n_b, are made. The same is true for the number of the signal observations, n_a, used for construction of the calibration plot. This plot is inevitably required in determining the LOD and LOQ by the IUPAC recommended method

because it enables to find the concentration counterparts to the signals y_D and y_Q. Therefore for *real* analytical measurements, which are based on a limited data set, the following changes can be considered: (a) Replacement of the population characteristics μ_b and σ_b by the sample characteristics $\bar{y}_b$ and s_b. (b) Use of the appropriate Student t- distribution, chosen according to the number of degrees of freedom, ν, and the confidence level α (e.g. 0.05). The extent, how these changes are incorporated into the y_D signal calculation is different for the ways (A) and (B) - eqs. (3) and (4), resp., and the ways (C) and (D) - eq. (5):

$$y_D = \bar{y}_b + 3\, s_b \tag{3}$$

$$y_D = q_0 + 3\, s_y \tag{4}$$

$$y_D = \bar{y}_b + t(\nu, \alpha)(1 + 1/n_b^{1/2})\, s_b \tag{5}$$

The term $(1 + 1/n_b^{1/2})$ expresses the correction for the uncertainty of the $\bar{y}_b - \mu_b$ determination and is maximal for the smallest n_b and approaches 1 for the n_b sufficiently large. Then the k_D value is $k_D = t(\nu, \alpha)(1 + 1/n_b^{1/2})$.

The calibration plot $y = q_0 + q_1 c$ is used for conversion from the signal to the concentration domain. The reference point for measuring y_D on the signal axis is (1) $\bar{y}_b$, (2) q_0, (3) μ_b, (4) μ_b, for the ways (A) to (D), resp. Then, using the *net signal* calibration plot, the concentration LOD and LOQ for (A) to (D) are, resp.:

$$LOD = 3\, s_b / q_1\ ; \qquad LOQ = 10\, s_b / q_1 \tag{6a,b}$$

$$LOD = 3\, s_y / q_1\ ; \qquad LOQ = 10\, s_y / q_1 \tag{7a,b}$$

$$LOD = \{t(\nu, \alpha)\, s_y / q_1\}[1 + 1/n + \bar{c}^2 / \Sigma(c_i - \bar{c})^2]^{1/2}\ ; \quad LOQ = 3\, LOD \tag{8a,b}$$

$$LOD = 3\, \{t(\nu, \alpha)\, s_y / q_1\}\ ; \qquad LOQ = 3\, \{t(\nu, \alpha)\, s_y / q_1\} \tag{9a,b}$$

where s_y is the residual standard deviation (expressing variability around the regression line), c_i - the concentration coordinates of the calibration points, and $\bar{c}$ denotes their mean.

REFERENCES:

[1] IUPAC: Spectrochim. Acta, 1978, 33 B, 242; 248.

[2] ACS Commitee on Environmental Improvement: Anal. Chem. 1980, 52, 2242.

[3] MILLER, J.C., MILLER, J.N: Statistics for Analytical Chemistry. E. Horwood, Chichester, 1984. Chapter 4.7.

[4] MOCAK, J., BOND, A.M., MITCHELL, S., SCOLLARY, G.: Book of Abstracts 12AC/3EC. Perth, W. Australia, 1993, p. 74; idem: Pure & Appl. Chem., in press.

ROBUST METHODS IN CALIBRATION

Ivo Müller

Charles University, Prague

Summary: *M-estimators are introduced into controlled calibration procedure. Asymptotic behavior of the resulting inverse estimator is studied. A preliminary test of regression parameters is also discussed.*

Calibration can be viewed as a two-step procedure. During the first step, often referred to as a calibration phase, or a calibration experiment, a "response" variable Y is related to an "explanatory" variable X through a calibration curve. The joint observations of (X, Y) are used for estimating parameters of the calibration curve. In the second step, called a prediction phase or inverse regression, inferences about an unknown X corresponding to a new observation Y, are made on the basis of previously estimated parameters.

We consider the following model of a calibration experiment

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{E}$$

where $\boldsymbol{Y}$ is a $n \times q$ matrix of responses, $\boldsymbol{X}$ is a known $n \times p$ matrix of explanatory variables, $\boldsymbol{B}$ is a $p \times q$ matrix of regression parameters and $\boldsymbol{E}$ is a $n \times q$ matrix of experimental errors. n is a number of observations in a calibration experiment, p is a number of explanatory variables and q a number of responses.

On the error matrix $\boldsymbol{E} = (e_1, ..., e_n)^T$ we impose following assumptions:

$$\mathrm{E}(e_i) = 0, \quad i = 1, ..., n$$
$$\mathrm{E}(e_i e_i^T) = \Sigma = (\sigma_{jk}^2)_{j,k=1,...,q}, \quad i = 1, ..., n$$

Errors are assumed to be independent from observation to observation, with a distribution function F.

We assume the prediction experiment to follow the same model. Variables associated with the prediction stage are denoted by a subscript $_0$.

$$\boldsymbol{Y}_0 = \mathbf{1}_m \boldsymbol{x}_0^T \boldsymbol{B} + \boldsymbol{E}_0,$$

where $\boldsymbol{Y}_0$ is a $m \times q$ matrix, $\mathbf{1}_m$ is a m-vector of 1's, $\boldsymbol{x}_0$ is a $p \times 1$ vector and $\boldsymbol{B}$ is the same as above. m is a number of observations in a prediction experiment. All the m observations are made at the same unknown p-dimensional value of $\boldsymbol{x}_0$. Errors are assumed independent of errors at the calibration stage.

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX

In order to improve on robustness in the linear model described above, we sought for alternative estimates of the matrix B. We chose M-estimates particularly because they are well theoretically explored, widely used and implemented in software packages.

In the one-dimensional, linear regression case (q=1), M-estimator $\hat{\beta}$ of a parameter β is defined through minimizing the expression

$$\sum_{i=1}^{n} \rho(Y_i - \sum_{k=1}^{p} b_k x_{ki}),$$

where the minimization is done over all $b \in \Re^p$ and where $\rho : \Re^1 \to \Re^1$ is a suitable loss function.

Let's denote

$$Q_n := n^{-1} \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^T$$

Q_n is called a design matrix and we impose the following assumptions:

(A) $n^{-1}Q_n \to Q$ positive definite
and
(B) $n^{-1} \sum_{i=1}^{n} ||\boldsymbol{x}_i||^3 = O(1)$, or (B') $n^{-1} \sum_{i=1}^{n} ||\boldsymbol{x}_i||^4 = O(1)$

Under these assumptions and depending on the smoothness properties of ρ (and its derivative ψ) we gain strong asymptotic approximations of M-estimates. These representations, in particular the asymptotic linearity result

$$n^{1/2}(\hat{\beta} - \beta) = \gamma(\psi, F)^{-1} Q_n^{-1} n^{-1/2} \sum_{i=1}^{n} \psi(e_i)\boldsymbol{x}_i + o_p(1),$$

are then exploited to formulate asymptotic properties of $\hat{x}_0$, the estimator in the prediction step.

Prior to estimating x_0, a preliminary test is performed on the hypothesis $H_0 : B = 0$. A meaningful estimator of x_0 is available only conditionally, i.e. when H_0 is rejected. Several types of tests are proposed, both "classical" (multidimensional normal) and robust (R- and M-type).

Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics of Charles University, Sokolovská 83, 186 00 PRAHA 8, The Czech Republic
E-mail address: muller@karlin.mff.cuni.cz

Wavelet Filters for Processing of Noisy Analytical Data

S. G. Nikolov, H. Hutter, M. Grasserbauer

Institute of Analytical Chemistry, Vienna University of Technology
Getreidemarkt 9/151, Vienna 1060, Austria

Summary: We describe wavelet methods for recovery of noisy analytical data. A comparison with polynomial smoothing filters is provided. Results from de-noising of chromatograms, SIMS images and Raman spectra are presented.

Wavelets

A multiscale representation that has recently attracted great interest in the signal processing community is the *wavelet transform*, which dates back to Strömberg (1983) and Meyer (1989). The two-parameter family of translated and dilated functions

$$\psi_{a,b}(x) = |a|^{-1/2} \psi\left(\frac{x-b}{a}\right) \quad a,b \in R, \quad a \neq 0$$

defined from a single mother function $\psi : R \to R$ is called a wavelet basis. The function ψ is called a *wavelet* (a short wave) which comes from the requirements that it should integrate to zero *'waving'* above and below the x-axis. The basic idea of the wavelet transform is to represent any arbitrary function f as a superposition of wavelets. The Continuous Wavelet Transform (CWT) of f is given by

$$(Wf)(a,b) = \langle f, \psi_{a,b} \rangle = |a|^{-1/2} \int_{x \in R} f(x) \psi\left(\frac{x-b}{a}\right) dx.$$

A particularly well studied problem is the construction of *orthogonal wavelets* for discrete signals, which permit a compact non-redundant multiscale representation of the signal data. A large number of signals are better approximated by substantially fewer wavelet basis functions than by sine and cosine functions for instance. The development of a fast implementation of the Discrete Wavelet Transform (DWT) by Mallat (1989) has made wavelet methods a particularly valuable tool for processing of analytical data.

De-noising via Thresholding

All wavelet coefficients $w_{j,k}$ in the final decomposition of signals correspond to details. If the absolute value of a wavelet coefficient is small and if it is omitted the signal would not change very much. Therefore, thresholding of the wavelet coefficients is a very efficient way of removing *unimportant* or *undesired* details from a signal, since in most cases a small number of wavelet coefficients with large amplitudes preserve most of the energy of the original data set.

Different thresholding methods like

$$\hat{w}_{jk}^{hard} = \begin{cases} 0, w_{jk} < t \\ w_{jk}, w_{jk} \geq t \end{cases} \qquad \textit{hard thresholding}$$

$$\hat{w}_{jk}^{soft} = sign(w_{jk})(|w_{jk}| - t)_+ \qquad \textit{soft thresholding}$$

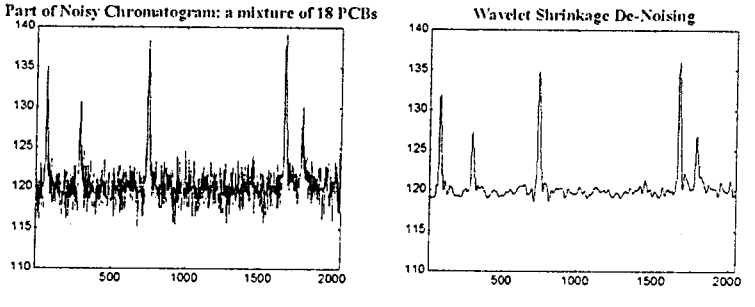
universal thresholding (hard or soft thresholding with $t = \sigma\sqrt{2\log(n)}/\sqrt{n}$, where σ is the variance of the Gaussian noise and n is the size of the data set) and others have been used to solve problems ranging from Gaussian noise removal to density estimation and inverse problems.

De-noising of Chromatograms

In our investigations we have used the signal de-noising algorithm proposed by Donoho and Johnstone (1992). If we have a chromatogram $y_i = f(t_i) + \sigma z_i$, $i = 1,..,n$, where z_i is white Gaussian noise and $n = 2^{J+1}$, we can use a 1-d pyramidal filtering by making the following three steps:

- Perform the forward pyramid wavelet transform of Cohen, Daubechies, Jawerth, and Vial (1992) to the normalized data $y_k/\sqrt{n}$, yielding noisy wavelet coefficients $w_{j,k}$, $j = j_0,...,J$, $k = 0,..,2^j - 1$;
- Apply universal soft thresholding to the noisy wavelet coefficients, yielding estimates $\hat{w}_{j,k}$;
- Make the inverse wavelet transform, producing the chromatogram reconstruction $\hat{y}_i$;

This wavelet domain filter shrinks the wavelet coefficients to zero. Thus, because the few large wavelet coefficients preserve almost the whole l^2 energy of the signal, the shrinkage *kills* the noise without distorting the chromatogram peaks.



Other Investigations

Our comparative study of algorithms for Gaussian noise removal clearly shows that wavelet methods outperform classical polynomial smoothers like the Savitzky-Golay filters in terms of MSE, SNR improvement and peak area preservation. We have investigated different wavelet filters for removal of Poisson distributed noise in Secondary Ion Mass Spectrometry images and shot noise in Raman spectra. In all reconstructions the noise is significantly suppressed, while sharp edges in the original signal remain sharp after the processing.

COMBINING THERMODYNAMIC AND SPECTROSCOPIC INFORMATION IN SPECTRAL ANALYSIS
Characterization of a single sample

Jan Nygren, José Manuel Andrade and Mikael Kubista
Molecular Biotechnology Group, Department of Biochemistry and Biophysics,
Chalmers University of Technology, S-413 90 Gothenburg, Sweden.

Summary: We demonstrate the effectiveness of chemometric methods to characterize thermodynamic equilibria by studying the effect of temperature on the spectral response. The temperature dependence of the equilibrium constant as well as the spectral responses of the components are determined. The approach works well even when the component spectra overlap extensively, and is applicable to all techniques providing a linear response.

The information content in a series of spectra recorded on a number of samples is insufficient to determine the spectral responses and the concentrations of the components.[1] Recently we showed that if the components are in chemical equilibrium thermodynamic relations can be exploited to determine their spectral responses and concentrations as well as the thermodynamic constants describing the equilibrium.[2] In this work we show that also a single sample can be characterized this way by studying the effect of temperature on the equilibrium.

Theory

Spectra recorded at different temperatures are arranged as rows in a matrix **A**, which is decomposed into an orthonormal basis set:

$$\mathbf{A} = \mathbf{TP}' + \mathbf{E} \approx \mathbf{TP}' = \sum_{i=1}^{r} \mathbf{t}_i \mathbf{p}_i'$$

where r is the number of components. Since the spectra are linear combinations of the concentrations, **C**, and spectral responses, **V**, of the components,

$$\mathbf{A} = \mathbf{CV}' + \mathbf{E} \approx \mathbf{CV}' = \sum_{i=1}^{r} \mathbf{c}_i \mathbf{v}_i'$$

$\mathbf{V} = \mathbf{RP}'; \mathbf{C} = \mathbf{TR}^{-1}$, where **R** is an r×r rotation matrix. It can be determined by requiring that the equilibrium constant, K, relating the component concentrations obeys the expected temperature dependence as given by the van't Hoff equation:

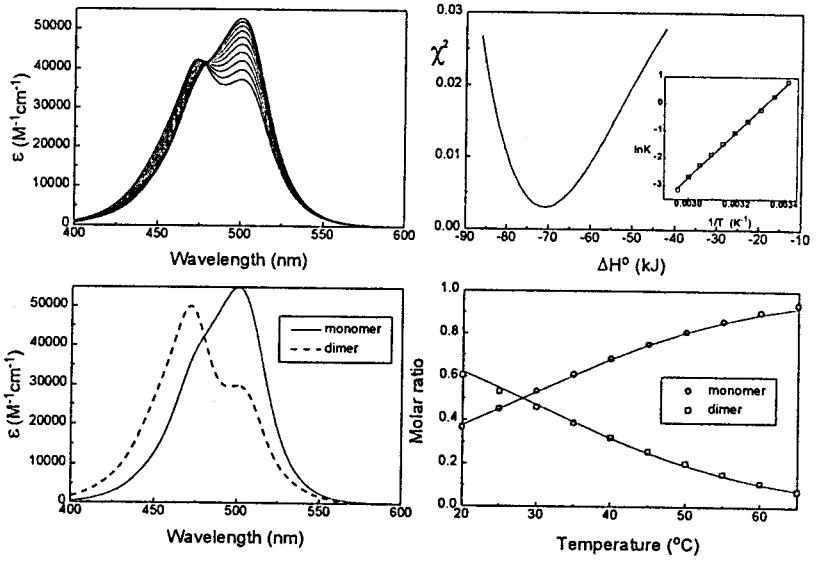
$$\frac{d\ln K}{dT} = \frac{\Delta H^{\varnothing}}{RT^2} \quad \text{or} \quad \ln K(T_1) - \ln K(T_2) = -\frac{\Delta H^{\varnothing}}{R}\left(\frac{1}{T_2} - \frac{1}{T_1}\right)$$

where $\Delta H^{\varnothing}$ is the molar enthalpy change, $R = 8.31\ \text{Jmol}^{-1}\text{K}^{-1}$ is the universal gas constant and T is temperature in Kelvin. Concentrations and equilibrium constants are calculated for various trial matrices $\hat{\mathbf{R}}$, and the one that produces best linear

regression of lnK(T) with respect to 1/T determines **C**, **V** and $\Delta H^{\varnothing}$. When varying $\hat{\mathbf{R}}$ it is important to note that its elements are not independent owing to constrains. For example, for a two-component equilibrium when the spectral response of one component is known, all elements are linearly dependent and the whole matrix is defined by anyone of its elements.[3]

Results and Discussion

To illustrate the approach we determine the dimerization constant of thiazole orange (TO) in water. Figure top, left shows absorption spectra of TO (36 μM) recorded at various temperatures in the range 20-65 °C. The distinct isosbestic point evidences the presence of exactly two components, which are the TO monomer and dimer. The TO dimer could not be obtained in pure form, but the spectrum of the monomer was recorded separately using an extensively diluted sample. This allowed us to describe matrix **R** with a single element, and calculate solutions for various trial matrices. The chi-squares of the linear regressions of lnK with respect to 1/T for these solutions is shown as a function of calculated enthalpy change in top right. Best solution produced an almost perfectly linear van't Hoff plot (insert) and gave $\Delta H^{\varnothing}$ =-71 kJmol^{-1}. The spectral profiles and concentrations corresponding to this solution are shown in the bottom panels.



[1] W. Lawton and E. Sylvestre, Technometrics 13 (1971) 617.
[2] M. Kubista, R. Sjöback and B. Albinsson, Anal. Chem. 65 (1993) 409; M. Kubista, R. Sjöback and J. Nygren, Anal. Chim. Acta 302 (1995) 121.
[3] S. Eriksson, S. Kim, M. Kubista and S. Nordén, Biochem. 32 (1993) 2987.

QUANTITATIVE ANALYSIS OF EPR SPECTRA OF POWDERED SAMPLES CONTAINING A MIXTURE OF VARIOUS PARAMAGNETIC CENTRES

P. Pelikán[a] and M. Liška[b]

[a] Department of Physical Chemistry, Slovak Technical University, 812 37 Bratislava, Slovakia
[b] Institute of Inorganic Chemistry, Slovak Academy of Sciences, 842 36 Bratislava, Slovakia

Summary: The presented paper deals with non-linear optimisation of EPR spectra of randomly oriented paramagnetic centres in solids. The proposed model of calculation enables to obtain the relative quantitative content of various paramagnetic centres in the measured sample.

The dependence of the experimentally obtained EPR spectrum Y(B) on the external magnetic field induction B can be expressed by linear combination of derivation bands (Lorentz, Gauss, and so on)

$$Y(B) = x_0 + \sum_{i}^{n} x_i y_i(B) \tag{1}$$

where x_0 is constant background of the measured spectrum;

x_i (i = 1,n) are coefficients of linear combination which represent the relative quantities of individual paramagnetic particles;

n is the number of various paramagnetic particles in the measured sample.

In the case of randomly oriented radicals in the solid state the individual derivation bands $y_i(B)$ must be averaged through the elementary angle volume [1]

$$y_i(B) = N_i \int_0^{\pi} \int_0^{\pi/2} P_i(\vartheta,\varphi) G_i(\vartheta,\varphi) \sin\vartheta \, d\vartheta \, d\varphi \tag{2}$$

where the angles ϑ and φ are defined according to the principal axes of the g-tensor [2].

N_i is normalisation constant which keeps the condition [3]

$$\int_0^{\infty} \int_0^{B} y_i(B) dB' dB = 1 \tag{3}$$

$P_i(\vartheta,\varphi)$ is the transition probability of the given transition and $G_i(\vartheta,\varphi)$ is the shape of the given spectral band.

The best estimate of the n_p unknown parameters (the spin-hamiltonian parameters and relative concentrations of paramagnetic particles) can be obtained by minimisation of the sum of squared deviations

$$F(\vec{p}) = \sum_{1}^{n_B} \left[Y^{calc}(B_l) - Y^{exp}(B_l) \right]^2 = \min \tag{5}$$

where $\vec{p}$ is the vector of n_p unknown parameters, $Y^{exp}(B_l)$ are experimentally measured spectral points and $Y^{calc}(B_l)$ are theoretically calculated spectral points according to Eq. (1).

If the experimental spectrum is normalised, i.e.

$$\int_0^{\infty} \int_0^{B} \left[Y^{exp}(B') - x_0 \right] dB' dB = 1 \tag{6}$$

the following condition is fulfilled

$$\sum_{i}^{n} x_i = 1 \tag{7}$$

The minimisation task can be divided into two interconnected parts:

i) non linear minimisation problem with regard to the spin-hamiltonian parameters, which are determined by physical character of the solved problem;

ii) linear minimisation problem with regard to the vector

$$\vec{x} = \{x_i\} \quad \text{for } i = 0,n \tag{8}$$

including binding condition (7) and, simultaneously

$$x_i \geq 0 \quad \text{for } i = 1,n \tag{9}$$

The linear minimisation is performed for each calculation of the F criterion.

Great number of the interconnected optimised parameters requires a special modification of the minimisation algorithm. It consists of gradual non-linear optimisations of the approximately defined small sets of optimised parameters that is cyclically repeated. At the end of minimisation procedure all of the optimised parameters are released.

The proposed model of calculation was used for investigation of various glasses and ceramics materials. The obtained results are in good agreement with chemical expectations.

References

[1] P. Pelikán, M. Čeppan, M. Liška, Applications of Numerical Methods in Molecular Spectroscopy, CRC Press, New York, 1994.

[2] W. Gordy, Theory and Applications of Electron Spin Resonance, Wiley, New York, 1980.

[3] P. Pelikán, M. Liška (to be published).

Chemometric Study of Organic Pollution in Organisms of the Coast of Canary Islands (Spain)

Peña Méndez E.M, Astorga España M.S, García Montelongo F.J.
Dpto. of Analytical Chemistry, Food Science and Toxicology.
University of La Laguna, 38071 - La Laguna, Tenerife, Spain.

Aquatic plants and animals can accumulate petroleum hydrocarbons from the environment even when the pollutants are present at very low concentrations. In the last time, there has been growing concern about the concentration of organic compounds in the sea environment. Some of these compounds, such as the Polynuclear aromatic hydrocarbons (PAHs), have well known mutagenic and carcinogenic activity.

Geographical and seasonal variations in the organisms and environments, however, makes it difficult to use the levels of pollutants found in theirs tissues as a measure of the degree of pollution of the environment.

In this work we have been used the **Multivariate Data Treatment** on data obtained by measurement of petroleum-derived hydrocarbons (*n*-alkanes, hydrocarbons, PAHs) in the tissue of endemic limpets (*Patella crenata, Patella ullisiponensis aspera*) , during the period 1992-1993. Samples were collected from differents stations in the coast of Tenerife Island.

The results obtained by GC/FID with SPB-5 capillary collumn, were subjeted to statistical processing by Statistical Software Package, Statgraphs Plus 5.2. We studied the correlations among variables directly connected by the pollution. PCA and FA followed by Varimax rotation, provide significant information about the relations among variables related to pollution.

PREDICTION OF GASOLINE OCTANE NUMBERS USING FT-MIR AND PLS. SAMPLES FROM CATALYTIC REFORMING PROCESSES

Prad[redacted] D.; Andrade, J.M.; Muniategui, S.; Fernández, E.

Dept. Analytical Chemistry. University of A Coruña. Campus da Zapateira s/n E-15071, A Coruña, Spain

Summary: using Fourier Transform Medium Infrared Spectroscopy and Partial Least Squares, improvements have been obtained in sample throughput, precision, lower costs and safer environmental working conditions when compared with current standard methods. Large routine -sample-sets were considered to assess the models performance.

INTRODUCTION

The catalytic reforming process is a generaly implemented unit in refineries over the world. The conceptual function of this process is to upgrade low octane number straight-run naphthas to higher octane motor fuel blending components by catalytically promoting specific groups of chemical reactions, typically aromatic groups, cycloalkanes and branched chains. These can be called reformed naphthas or gasolines.

The two most critical quality tests of gasolines are the prediction of the knock characteristics: the Research Octane Number (RON) and the Motor Octane Number (MON). Both are normalized by ASTM; using two standardized motors (ca. US$ 155000 per motor) and 500 mL of standard blends mL. Motors are noisy and produce exhaust gases and, what is worst, have large precission intervals. Even the standard deviation is strongly dependent on the RON and MON values.

Spectroscopy methods are well known, have robust equipments, do not requiere skilled staffs nor special conditions in industrial laboratories and they are quite fast. Surprisingly, medium-IR has not been extensively used although it presents, *a priori*, good perspectives.

EXPERIMENTAL PART

Samples

310 samples collected from the daily output of catalytic reforming processes from Spanish refineries were used. Aliquots of 50 mL were placed in chromatographic vials, closed and sealed. Vials were stored in a refrigerator.

Instrumental set-up

The FT-MIR-PLS method was implemented on a Perkin-Elmer 1610 FT-IR

spectrometer operating in the middle-infrared (MIR) region and furnished with an NaCl traditional cell of constant thickness. The spectrophotometer performance was checked by means of one polystyrene film. All the studied bands fell within the acceptable error ranges. Spectra were scaned with a 4 cm^{-1} resolution using backgroud substraction.

Procedure

PLS was used in its two variants (1-block and 2-block) for statistical modelling. Data were mean-centered prior to all analyses. The optimal number of latent variables for each model was determined by cross-validation. The optimal model chosen in each case was that resulting in the minimum standard error of prediction (SEP).

The FT-MIR-PLS method has been checked for repeatability and reproducibility. Repeatability was found to be < 0.1 O.N.; reproducibility, was found to be as low as < 0.3 O.N.

Samples were classified as "type A" or "type B" according with their aromatic contents. Different models were developed for either type A or type B gasolines.

RESULTS AND DISCUSSION

RON and MON prediction by FT-IR-PLS can be resumed by their SEP values:

	Type A gasolines, SEP	Type B gasolines, SEP
MON	(n=61) 0.33 O.N.	(n=112) 0.31 O.N.
RON	(n=15) 0.37 O.N.	(n= 36) 0.42 O.N.

Using PLS-2 block, these values were not improved; probably due to the different number of latent variables needed to model the MON and RON parameters.

COMPARISON OF DIFFERENT SAMPLING SEASONS IN SOILS USING PROCRUSTES ROTATION

Prada, D.; Carlosena, A.; Andrade, J.M.; Kubista, M.(*); Fernández, E.; López, P

Dept. Analytical Chemistry. University of A Coruña. Campus da Zapateira s/n E-15071, A Coruña, Spain
(*) Dept. of Biophysics and Biochemistry. Lundberg Laboratory. Göteborg University. Medicinaregatan 9C. Göteborg. Sweden

Summary: Procrustes rotation is used to select the minimum number of original variables that are important to describe a system and to compare different sampling seasons. Four samplings were carried out to develop a monitoring scheme to control metal pollution in a medium-city area. Future efforts to monitor metal pollution will be based on analysing only two or (at most) three metals, as Procrustes rotation has revealed.

INTRODUCTION

When environmental studies are made, two questions are posed: first, should all initially considered variables be maintained in future samplings?; second, are all analytical tests necessary to describe the system?. Considering the latter question, we are looking for some technique to find the minimum number of original variables that allows us to describe our system with sufficient accuracy. To answer the former question some kind of inter-sampling-seasonal is needed. Studying what happens in an inter-seasonal data sense, we should be able to describe which variables have a similar pattern/behaviour along the different sampling seasons.

PROCRUSTES ROTATION

Both topics can be solved considering that the main idea is to compare two or more multidimensional data sets (one from each sampling season). This can be achieved using the mathematically well known technique *Procrustes rotation*. In the Greek legend, Procrustes lodged travellers in his bed and during sleep either cut their legs or elongated them to fit them precisely into the bed. In analogy with Procrustes himself, in Procrustes rotations the two sets of projections are rotated to a consensus target vector so as to match as closely as possible in the least squares sense.

1st. question: SELECTION OF THE MINIMUM NUMBER OF ORIGINAL VARIABLES TO DESCRIBE ONE SAMPLING SEASON

The main objective is to select redundant variables in raw data, which is equivalent to the identification of a subset of k variables which conveys the main structure of the raw data. Each variable is deleted in turn from the data set and the two subspaces are compared by rotation, translation and streching by using the Procrustes rotation

technique. A statistic is defined, the "predicted residual error sum of squares", PRESS. Important variables will produce larger PRESS values and they should be maintained in the data set.

An important task is to select the optimum number of PCs since the subspace comparison is made in a PCA base. This is made by svd and calculating the W_m statistic. The W_m values are expressed in a cross-validation scheme and represent the increase in predictive information supplied by the *m* th component divided by the average predictive information in each of the remaining components. Important components should yield values of W_m greater than about 0.9 .

2nd question: INTER-COMPARISON OF SAMPLING SEASONS

Krzanowski proved the following two results:

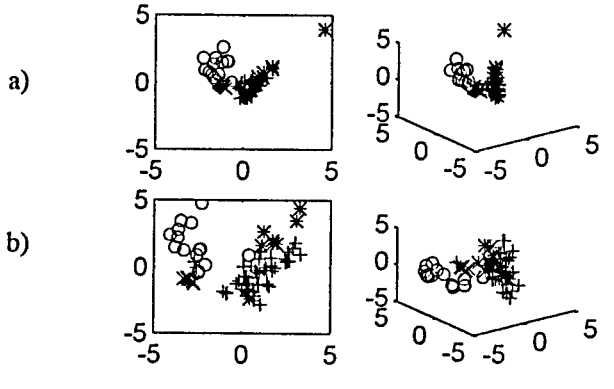
1.- the minimum angle between an arbitrary vector in the space of the first k principal components from the first sampling season and the one most nearly parallel to it in the space of the first k principal components of the second sampling season is given by $\cos^{-1}\sqrt{\lambda_i}$.

2.- a set of mutually orthogonal vectors can be defined into which the differences between the subespaces can be partitioned. The angles between these vectors are given by $\cos^{-1}\sqrt{\lambda_i}$

The r critical angles $\cos^{-1}\sqrt{\lambda_1}$, ..., $\cos^{-1}\sqrt{\lambda_r}$ can measure the extent to which two r-dimensional portions differ. If they have strong similarities there will be r angles close to zero; if not, the r values are large (close to 30 or greater). A very important advantage of this approach is that we can define a *consensus vector* directly linked with the original measured variables.

RESULTS AND DISCUSSION

The Figure shows that two variables are enough to explain the main characteristics of the soils considered in this work (in this case, the first sampling season) (a) Selected variables (Pb, Co) (b) all initial variables.



A NEW APPROACH IN BIASED REGRESSION

E. M. QANNARI E. VIGNEAU M. SEMENOU

ENITIAA LABORATOIRE DE STATISTIQUE
LA GERAUDIERE 44072 NANTES CEDEX FRANCE

We discuss an optimisation problem which provides a new characterisation to Ridge Regression (RR). A variant of this optimisation problem leads to a new family of biased estimators that include Stein estimators and Principal Components Regression (PCR) as particular cases.

Formulation of the problem

With the usual notations and conventions, let us consider the general linear model of the form :

$$Y = X\beta + \varepsilon .$$

The ordinary least squares (OLS) estimator is given by $B_0 = V^{-1} X'Y$, where $V=X'X$. It is known that B_0 may be far removed from the true parameter β, in presence of near-collinearity among the independent variables.

Several alternatives to OLS are proposed in order to handle this problem. The rationale behind these methods is to trade the high variability of the parameter estimates for some bias (hopefully negligible). The most popular biased estimation methods are PCR and Ridge Regression (RR).

Let us consider linear estimators that can be written as follows :

$$B = T X' Y = T V B_0$$

where T is a pxp matrix. Being a linear transform of B_0, B is a biased estimator if T is different from V^{-1} :

$$E(B) = T V \beta .$$

It can be shown that the Mean Squared Error associated with B, MSE(B), can be written as follows:

$$\begin{aligned} MSE(B) &= \beta'(TV-I)'(TV-I)\beta + \sigma^2 \operatorname{trace}(TVT') \\ &= D(B) + V(B) \end{aligned}$$

where V(B) is the total variance of B and D(B) is the squared bias term of B.

First Optimisation problem : Ridge estimator

From the inequality :

$$0 \leq D(B) \leq \|TV-I\|^2 \, \|\beta\|^2$$

it appears that the squared bias term D(B) vanishes when $T = V^{-1}$ and approaches 0 when $\|TV-I\|^2$ approaches 0. This suggests the following optimisation problem:

$$\min_T \|TV-I\|^2 \quad \text{under the constraint trace } (T V T') = c$$

c being a fixed positive scalar.

The rationale behind this optimisation problem is to fix the total variance of the parameter estimates at an "acceptable" level while allowing a bias as small as possible. The solution of this optimisation problem leads to RR, namely :

$$T = (V + k\,I)^{-1}\ .$$

Second Optimisation problem : Generalised Stein estimator

As an alternative to the previous optimisation problem, we consider the constraint $\det(TVT') = c$, (det denotes the determinant) instead of $\text{trace}(TVT') = c$. This new constraint takes into account correlation effects and its rationale is to control the level of the generalised variance.

The new optimisation problem is therefore as follows :

$$\min_{T} \|TV - I\|^2 \qquad \text{under the constraint } \det(TVT') = c$$

The solution of this problem leads to a biased estimation method which consists in a non uniform shrinkage as the directions corresponding to the eigenvectors associated with the smallest eigenvalues have a shrinkage coefficient smaller than those of the other directions.

The family of estimators generated from this optimisation problem covers a large range of estimators including OLS, stein estimator and PCR.

Application:

On basis of real data sets (calibration from NIR spectra), we compare the prediction ability of OLS, PCR, RR, Stein estimation method and the new approach we suggest. It appears that this new approach leads to a significant improvement over OLS, RR and Stein estimation and to a slight improvement over PCR.

REFERENCES

Hoerl A. and Kennard W. (1970). - Ridge regression : biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55-67.

Naes T. (1985). - Multivariate calibration when the error covariance matrix is structured. *Technometrics*, 27(3) 301-311.

EXAMPLES USING MULTIDIMENSIONAL IVS-PLS

S. Rännar, F. Lindgren, A. Berglund and S. Wold

Research Group for Chemometrics
Department of Organic Chemistry
University of Umeå
s-901 87 Umeå
Sweden

Summary: In this poster is presented examples where the predictability of PLS-models is improved by using multidimensional IVS-PLS.

In modern chemistry it is often easy to measure or calculate hundreds, or even thousands, of variables on one sample or compound, eg in multivariate calibration using digitalized spectra and in 3D-QSAR. It is therefor important to have proper computational tools that can extract the information from the huge amount of data available.

To find correlations between the chemical structure and some kind of activity (biological or chemical) Partial Least Squares, PLS [1], is often used and there are many documented QSAR-studies where it has proved to be useful [2,3].

One way of dealing with PLS-modelling of large data matrices with many variables is to reduce the number of variables. In IVS-PLS (Interactive Variable Selection) [4] this variable selection is done componentwise based on cross validation and it usually leads to a better predictive capability of the model. When working with the IVS-procedure we have found that there is a correlation between the cutoffvalues in the different components. To deal with this we have developed a multivariate IVS-procedure which is used in the examples presented in this poster. By reducing the variables in several dimensions at the same time we have improved the predictabilities of the models.

REFERENCES

1. S. Wold, A. Ruhe, H. Wold, W. J. Dunn III, SIAM J Sci. Stat. Comp, 5, 1984, 735-743
2. M. Cocchi and E. Johansson, Quant. Struct.-Act. Relat., 12, 1993, 1-8
3. S. Hellberg, L. Eriksson, J. Jonsson, F. Lindgren, M. Sjöström, B. Skagerberg, S. Wold and P. Andrews, Int. J. Pept. Prot. Res., 37, 1991, 414-424
4. F. Lindgren, P. Geladi, S. Rännar and S. Wold; Interactive Variable Selection (IVS) for PLS. Part one; Theory and Algorithms. J. of Chemometrics, 8, 349-363 (1994)

CALCULATION OF RESTRICTION FRAGMENT LENGTHS FROM DISTANCES MIGRATED IN AGAROSE GEL

BOHUSLAV RITTICH[1] and ALENA ŠPANOVÁ[2]
[1]Faculty of Chemistry, Technical University Brno, Veslařská 230, Brno, Czech Republic, [2]Institute of Animal Physiology and Genetics of the Academy of Sciences of the Czech Republic, Královopolská 135, Brno, Czech Republic

Summary: Regression analysis was used to expression of the relationship between DNA fragment lengths and distances migrated in agarose gels

Standard mixtures of DNA fragments were separated by electrophoresis in agarose gels. Regression analysis was used to fit curves to the data points obtained by plotting DNA fragment lengts (the number of base pairs - bp) versus distances migrated in gel (D). The equations which describe the relation between bp and D were expressed in the form:

$$bp = a + bD + c\ln D \quad (1)$$

$$bp = \Sigma\, a_n D^n \quad (2)$$

$$\ln bp = a + bD + c\ln D \quad (3)$$

$$\ln bp = \Sigma\, a_n D^n \quad (4)$$

where a,b,c are constants, n-polynominal order.

Polynomials with orders ranging from 1 to 3 were used. It was found that the third order polynomials of equations (2) and (4) are the highest order equations in which all regression coeficient are statistically significant. The equation (3) is also suitable.

tRNA sequence analysis based on principal properties

Maria Sandberg, Michael Sjöström and Svante Wold
Research Group for Chemometrics, Department of Organic Chemistry, University of Umeå, S-901 87 Umeå, Sweden.

Summary: Twenty nucleosides have been characterized with the use of 22 variables, both theoretical and experimental. The scores from a principal component analysis are used as descriptors in sequence analysis.

Modified nucleosides are naturally occurring within the cell as building blocks in transfer ribonucleic acid (tRNA, a macromolecule in the protein synthesis machinery that e.g. decodes the genetic message into functional proteins). Modified nucleosides are also important e.g. as biological markers of diseases like cancer [1] and acquired immunodeficiency syndrome (AIDS) [2].

Quantitative descriptors based on physico-chemical variables have previously been used in quantitative structure-activity relationships (QSAR) of peptides [3], epoxides [4] and dioxines [5]. The principal component scores (here referred to as principal properties, PP:s) for the nucleosides gives the basis for discussion of the chemical properties among the nucleosides, a strategy for expansion to new objects, and use in quantitative sequence analysis.

An example is given, where the nucleoside PPs have been used for the quantitative description of a number of tRNA sequences.

References

[1] Salvatore, F. , Recent Results Cancer Res, 84, (1983) 360-377.
[2] Borek, E et al. Cancer Res 46, (1986) 2557-2561
[3] Jonsson, J et al. Quant Struct Act Relat, 8, (1989) 204-209.
[4] Eriksson, L. et al. Environmental Toxicology and Chemistry, Vol 13, No 4 , (1994) 1-9.
[5] Tysklind, M. et al. Environmental Toxicology and Chemistry, Vol 12 (1993) 659-672.

Abstract

H.C. Smit and R. Kilg
Laboratory for Analytical Chemistry, University of Amsterdam
Nieuwe Achtergracht 166, 1018 WV Amsterdam, The Netherlands

Information-theoretical optimization of separations

The chromatographic process is considered as an information channel.
The information rate (bits/sec) can be calculated, using the power spectrum and power spectral density of signal and noise.
A new information-theoretical optimization criterion is the amount of information obtained during the separation.
The uncertainty in the determination of peak intensities is minimized using this criterion; the quality of the separation is maximal.
The calculated optimum separation (resolution) does not correspond with the usual numerous resolution, but results in a considerable overlap.
According to this approach, optimum chromatography means the application of very short columns, rapid detection/injection, special chemometrical signal processing software, and drastically reduced analysis time.

Structure elucidation of polyunsaturated fatty acid ethyl esters by multivariate modelling of FT-IR spectra.

Hilde Strom Norsk Hydro a.s
Frode Brakstad Norsk Hydro a.s
Olav Kvalheim University of Bergen

An approach is given for selecting the most significant absorption range for FTIR spectrum structure correlations. The method is based on principal component analysis (PCA) and partial least squares regression (PLS) on normalized spectral intensities for selection of absorption-ranges. The quantitative spectrum-structure modelling is based on partial least squares regression (PLS) on normalized spectral intensities as independent variables, and positions and numbers of carbon - carbon double bonds as structural descriptors (responses).

The procedure is demonstrated by using transmission FTIR of straight chain mono- and polyunsaturated fatty acid ethyl esters, ranging in carbon - carbon double bond position from 6 -15 and number of double bonds from 1- 6, incorporating both cis- and trans-isomers and chain length from C_{18} - C_{24}.

On the basis of relative intensities of different absorption - ranges in the spectra, quantitative models relating them to the structural descriptors was developed. The methods give estimates of

* **carbon - carbon double bond positions for cis- and trans isomers**

* **carbon - carbon double bond number for cis isomers**

CALCULATION AND ANALYSIS OF MODEL MATRIXES OF ABSORBANCES ON PC FOR VERIFYING A NEW METHOD OF DETERMINATION OF NUMBER OF COMPOUNDS IN MIXTURE

Andrea Sloučková, University of Pardubice, Faculty of Chem. Technology, Department of Phys. Chemistry, nám. Čs. legií 565, 532 10 Pardubice, Czech Republic

Summary: An original program is presented, calculating the model matrixes of absorbances of the 12th order in Quick Basic for 1 to 10 components in the mixture and imitating random errors in the absorbances with the generator of random numbers. The resulting matrixes are used to verify the new method of determination of the number of species in the mixture with the help of linear regression in Quattro Pro.

For verification of the new method of determination of number of compounds in the mixture (s. Komers K. : this book), the model matrixes of absorbances of mixtures with 1 to 10 compounds were needed. The matrixes were obtained on PC as follows:

1) Two series of 120 random natural numbers (y) from 1 to 10 were calculated with the generator of random numbers in BASIC according to the equation

$$y = \text{INT}\,(\text{RND} * 10 + 1)$$

The first series was used for 12 combinations of absorption coefficients (e) in 12 rows, the second one for 12 combinations of concentrations (c) in 12 columns of the matrix for mixtures of 1 to 10 species (s).

2) Both series of random numbers were input as DATA into the original computation program in QUICK BASIC, which calculated the individual elements of the matrix for the given number of species (n) according to the equation

$$A(i,j) = \Sigma\, e(s,i) \,.\, c(s,j) \quad [\, s=1 \text{ to } n \,]$$

in rows i=1 to 12 and columns j=1 to 12. This calculation was made, on the one hand, exactly and, on the other hand, with simulated errors 0 to ± 1, 5 and 10% of the exact A(i,j) value. These errors were calculated also with the generator of random numbers. The main parts of the computation program include the following instructions:

READ e(s,i) : reading of the absorpt. coefficient e of the component s in the row i .

READ c(s,j) : reading of the concentration c of the component s in the column j .

FOR i=1 TO 12 : FOR j=1 TO 12 : FOR s=1 TO n : LET a = e(s,i) * c(s,j) : LET A(i,j) = A(i,j)+a : NEXT s : NEXT j : NEXT i : calculation of the absorbance A(i,j) of the mixture with n components as the element in the place i,j of the matrix .

LET b = A(i,j) * x * 0.01 : calculation of the error x% of the value A(i,j) .

LET A(i,j) = INT (b(i,j) * 10) : rounding of the value to the number with 4 digits.

The complete program included the following possibilities: Option of n=1 to 10, option of random errors in A(i,j) from 0 to arbitary x% with linear distribution, checking of the selected values e(s,i) and c(s,j) and also option of the dependences between the concentrations of the components in the given mixture (not used in the described computation). The calculated matrix was automatically saved in QUICK BASIC.

The matrixes were calculated for 1 to 10 components in the mixture exactly and with random errors 0 to ±1, 0 to ±5 and 0 to ±10% in A(i,j) values. These matrixes were transfered in QUATTRO PRO. For each of them the validity of the above-mentioned method was checked for various combinations of rows and columns.

This verification was made with the help of linear regression of the given combination of rows or columns and based on the values of the correlation coefficient, coefficients at individual independent variables and their standard deviations (s. Komers K.: this book). The computing program in QUATTRO PRO makes these computations easy and quick.

APPLICATION OF NEURAL NETWORKS TO A SOLVATOCHROMIC APPROACH TO GAS CHROMATOGRAPHY

Daniel Svozil and Jiří G.K. Ševčík*
Department of Analytical Chemistry, Charles University, Albertov 2030, CZ-128 40 Prague 2
* ALS GmbH., P.O.Box 1225, D-92602 Weiden

Summary: *The Kohonen neural network was used for the selection of the solute descriptors. The feed-forward neural network was applied for the prediction of the polarity/polarizability.*

The continuous development of gas chromatographic methods is accompanied by efforts to describe solute-stationary phase interactions. Characterisation of stationary phases is a pivotal question.

The most popular characterisation system for stationary phases is the Rohrschneider-McReynolds system assuming only one type of interaction for a selected standard compound (e.g., 2-butanone should only exhibit orientation forces). The main disadvantage of this system lies in the fact that the selected standard compounds do not actually represent only one type of interaction.

A more prospective approach is the solvation concept, as by Abraham. It holds that:

$$\log SP = SP_0 + rR_2 + s\pi_2^H + a\alpha_2^H + b\beta_2^H + l\log L^{16} \qquad (1)$$

The dependent variable in this equation can be ,e.g., $\log t_r'$, where t_r' is the adjusted retention time. Parameter R_2 represents excess molar refraction, α_2^H and β_2^H are the effective acidity and basicity parameters respectively, π_2^H expresses polarity/polarizability and $\log L^{16}$ is the decadic logarithm of the solute partition coefficient between the mobile and the stationary phase at 25°C.

The regression coefficients in (1) describe the given stationary phase. The parameter *r* corresponds to the ability of the stationary phase to interact with π- a n-electron couples of the solute, the parameter *s* describes stationary phase participation in dipole-dipole interactions and the parameters *a* or *b* represent, respectively, the basicity or the acidity of the

stationary phase. The parameter $\underline{l}$ describes the disperse and cavitation forces

The applicability of this equation depends on the availability of the solute parameters. The additivity principle has been applied to some of the solute parameters (e.g. $logL^{16}$, R_2) and thus characteristic values are predictable for any compound.

In our work, attention has been paid to the parameter π, which has been assumed to be non-additive. An artificial feed-forward neural network, which can handle strongly non-linear dependencies, has been used for calculation of π. The key problem is the choice of descriptors characteristic for the given solute. The Kohonen neural network has been used for their selection. The descriptors have been selected from the following groups: physico-chemical constants, topological indices, electron and quantum-chemical indices, and indices of the molecular shape. The input vector for the feed-forward neural network has been obtained from various combinations of these descriptors.

COMPUTATION OF METAL IONS INTERACTIONS WITH POLYELECTROLYTES. ACIDOBASIC AND COMPLEXATION PROPERTIES OF HUMIC ACIDS

D. Široký[a], D. Fetsch[b*], P. Lubal[a], and J. Havel[a]

[a] Department of Analytical Chemistry, Faculty of Science, Masaryk University, Kotlářská 2, 611 37 Brno, CZECH REPUBLIC

[b] Ecole Europeenne des Hautes Etudes des Industries Chimiques de Strasbourg (EHICS), 1, rue Blaise Pascal, 6700 Strasbourg, FRANCE

Summary: *Acidity constants of humic acid (HA) and stabilities of Cu(II), Cd(II), Pb(II), Ca(II), Ba(II), and uranyl complexes were computed. Computational methodology has been proposed.*

Humic acids (HA) are polyelectrolytes with characteristic functional groups [1] occurring as natural products in nature.

In frame of the study of toxic metal ions transport in the environment, the speciation of several metal ions (Ba^{2+}, Ca^{2+}, Cd^{2+}, Cu^{2+}, Pb^{2+}, UO_2^{2+} under the presence of HA's has been investigated by spectrophotometric and potentiometric methods using preparative of humic acid extracted from northern Bohemia brown coal.

On the base of discrete model [2] of humic acid, the acidimetric titrations of its salts were explained suggesting HA as a mixture of several ligands with different structures. The protonization constants of HA fragments were determined.

During computation several models were tested. Finally, by means of SUPERQUAD and LETAGROP-ETITR (NYTIT) programs the most probable explanation of data has led to the stability constant values decreasing in the following order :

$$Cu^{2+} > Ba^{2+} > Pb^{2+} > Cd^{2+} > Ca^{2+}.$$

* *On the leave from*

This fact is in agreement with Irwing-Williams [1] results.

Both, acidobasic models for HA and for HA-metal ions interactions were proposed and computational methodology for polyelectrolytes equilibrium studies suggested.

References:

1. F. J. Stevenson, HUMUS CHEMISTRY: GENESIS, COMPOSITION, REACTIONS, J. Wiley - Interscience, New York 1982.

2. N. Paxeus, M. Wedborg, Anal. Chim. Acta **69**, 87 (1985).

COMPUTERISED POTENTIOMETRIC STRIPPING ANALYSIS OF THALLIUM IN ENVIRONMENTAL SPECIMENS

I. Švancara[1], P. Ostapczuk[2], J. Arunachalam[2,3], H. Emons[2], K. Vytřas[1]

1) Department of Analytical Chemistry, Faculty of Chemical Technology, University of Pardubice, náměstí Čs.legií 565, CZ-532 10 Pardubice, CZECH REPUBLIC
2) Institute of Applied Physical Chemistry, The Research Centre Juelich (KFA), D-524 25 Juelich, FEDERAL REPUBLIC OF GERMANY
3) Permanent address: Department of Atomic Energy, Brabha Atomic Research Centre, 500 762 AP Hyderabad, INDIA.

Summary: A method for the determination of thallium at trace and ultratrace level has been developed based on the use of computer-controlled potentiometric stripping analysis with a mercury film electrode. The method is applicable to analyse thallium in various environmental specimens.

The determination of thallium in environmental specimens still represents a serious problem because of extremely low concentration levels at which this element occurs in the nature [1]. Potentiometric stripping analysis (PSA) in combination with a computer has already been proposed for environmental analysis of Tl because, among others, this technique offers a very low detection limit [2]. Computerised PSA is also relatively insensitive towards matrix effects so that the determination can be carried out even in unpretreated samples - either directly or after their decomposition [3]. This simplifies the whole analytical procedure, and in particular, minimises the risk of contamination or loss of the analyte in consequence of the manipulation with the sample during pretreatment operations [4].

The method developed is based on potentiostatic pre-concentration of thallium at a mercury film electrode (deposited onto a glassy carbon support) with a subsequent re-oxidation of amalgamated Tl. The re-oxidation step has been interpreted to be a process where both bivalent mercury and oxygen may act as the chemical oxidant. A supporting electrolyte containing 0.02 - 0.07 M HNO_3 + 0.007 M HCl + 0.02 M CH_3COONa + 0.001 M EDTA + 8×10^{-5} M Hg^{II} has been found to be optimum. The constituents of the supporting medium and their concentrations have been chosen in order to meet the following criteria. Firstly, nitric acid was incorporated into the supporting electrolyte due to the fact that this mineral acid had been used as a digestion reagent for the decomposition of samples by means of high pressure ashing (HPA). All the experimental conditions have been optimised with respect to the presence of nitric acid in the electrolyte, thus avoiding possible difficulties connected with direct analyses of the sample digests containing about 3 M HNO_3. Hg^{II} added into the supporting medium in the form of a plating solution ($HgCl_2$ + HCl) ensures "In Situ" formation of the mercury film. EDTA present in the electrolyte is capable to suppress

interferences from Pb(II) ions up to a 500-fold excess in concentration with respect to Tl(I) and from Cd(II) and Cu(II) up to their 100-fold excess, respectively. Moreover, the use of EDTA has also resulted in a significant enhancement of the response of Tl. Similarly, a marked enhancement of the analytical signal can be achieved by deaeration with argon gas, i.e. by removing oxygen dissolved in the solution. In both cases, the improved sensitivity of the method is in direct relation with the lowering of the concentration of oxidant. Due to these factors together with other favourable effects of the constituents of the supporting electrolyte, the method is able to operate in a wide concentration range. The analytical signal versus concentration dependence is linear from 9×10^{-11} to 2.5×10^{-8} M Tl (pre-concentration time from 5 to 30 min.) and one can achieve a detection limit of about 5×10^{-11} Tl (10 ng/L) when using the pre-concentration for 30 minutes.

The applicability of the method is demonstrated on various samples such as tap water (spiked), rain water, river sediment or sewage sludge . Except water samples, all the remaining had to be decomposed by HPA. The results have shown good recovery (spiked samples) or agreed well with those declared for reference materials or those obtained by Isotope Dilution-Mass Spectrometry (ID-MS). The method has also been tested to determine Tl in digested biological samples (e.g. pine needles, poplar leaves and mussel tissues), representing the specimens with particularly low concentrations of thallium.

The results of analyses have been evaluated by conventional statistics. Possibilities and limitations of PSA for the determination of Tl in various environmental specimens, including biological materials, are commented and some conclusions given. Finally, potentional applicability of the method to fully automated routine analyses is discussed.

References:

1) E.Waidmann, K.Hilpert, M.Stoeppler, *Fresenius J.Anal.Chem. 338* (1990) 572.
2) J.K.Christensen, L.Kryger, N.Pind, *Anal.Chim.Acta 141* (1982) 131.
3) D.Jagner, *Analyst 107* (1982) 593.
4) P.Ostapczuk, *Anal.Chim.Acta 273* (1993) 35.

Modeling and prediction of molecular properties. Theory of Grid-Weighted Holistic Invariant Molecular (G-WHIM) descriptors.

R. Todeschini [1], G. Moro [2], U. Cosentino [2], L. Bonati [2], M. Lasagni [2], D. Pitea [2], R. Boggia [3]

[1] Dipartimento di Scienze dell'Ambiente e del Territorio - via Emanueli, 15 - 20126 Milano (Italy)
[2] Dipartimento di Chimica Fisica ed Elettrochimica - via Golgi, 19 - 20133 Milano (Italy)
[3] Istituto di Scienze e Tecnologie Farmaceutiche ed Alimentari - via Brigata Salerno, ponte - 16147 Genova (Italy)

Summary: A development of recently proposed molecular descriptors (WHIM descriptors) is proposed to deal with scalar fields. These new descriptors (G-WHIM) are able to synthesize chemical information obtained from several kinds of interaction energies.

Recently proposed three-dimensional molecular indices (WHIM descriptors[1,2]) have been used to searching for quantitative structure-activity and structure-property relationships for several classes of compounds. WHIM descriptors contain information about the whole molecular structure in terms of size, shape, symmetry and atom distribution. These indices are calculated from the cartesian coordinates of the atoms weighted in different manners and represent a very general approach to describe molecules in a unitary conceptual framework.
In this paper the WHIM descriptor approach is extended to treat scalar fields. The new G-WHIM (Grid-Weighted Holistic Invariant Molecular) descriptors are calculated from the coordinates of the grid points where the field has been evaluated, each weighted by the field value. The fields considered in the first examples are related to non-bonding and electrostatic interaction energies, both evaluated by classical potentials. Two different sets of descriptors are obtained, for negative and positive potential separately.
The information contained in the whole grid is synthesized in a few parameters to be used in Sructure-Activity Relationship studies, thus avoiding difficulties related to chemical information spread out over a great number of points and overcoming the problem of results' dependence from the molecule alignment. Their interpretability is quite evident and is defined by the same mathematical properties of the algorithm used for their calculation.
Applications of G-WHIM to some QSAR interesting data sets gives good models, confirming the high modeling power of these descriptors.

[1] R. Todeschini, M. Lasagni, and E. Marengo. New molecular descriptors for 2D and 3D structures. Theory. *J. of Chemometrics*, **8**, 263-272 (1994)

[2] R. Todeschini, P. Gramatica, R. Provenzani, and E. Marengo. Weighted holistic invariant molecular descriptors. Part 2. Theory development and applications on modeling physicochemical properties of polyaromatic hydrocarbons. *Chemometrics and Intelligent Laboratory Systems*, **27**, 221-229 (1995).

CLUSTER ANALYSIS OF MOLECULAR GRAPHS BY PRINCIPAL COMPONENT ANALYSIS

VARMUZA K.*, SCSIBRANY H.

Technical University of Vienna, Department of Chemometrics,
A-1060 Vienna, Lehargasse 4/152, Austria

KERBER A., LAUE R.

University of Bayreuth, Lehrstuhl II für Mathematik
D-95440 Bayreuth, Germany

Summary: A set of 165 binary molecular descriptors is used to characterize chemical structures. Projection of the multivariate data by principal component analysis allows interactive definition of clusters containing similar chemical structures. This method has been implemented in the software ToSiM, running under MS-DOS. Isomeric structures have been automatically generated by the software MOLGEN, running under MS-Windows.

Sets containing between about twenty and some hundred chemical structures are often obtained as a result of searches in databases, also by computer-assisted structure elucidation, or in investigations of structure-property relationships. In these cases clustering of chemical structures into groups of similar structures may be helpful for further interpretations or for building data models.

For a clustering of chemical structures usually each structure is characterized by a set of molecular descriptors and hierarchical methods are applied; the result is a dendrogram. Application of a projection of the multivariate data by principal component analysis (PCA) has advantages in comparison with dendrogram methods: (a) In a scatter plot two dimensions can be used for a representation of the data structure and therefore visualization is in general better than by a dendrogram. (b) Loading plots provide information about the influence of the descriptors on clustering. (c) No hierarchical construction of the clusters is necessary. (d) The human is very capable and flexible in recognizing two- or three-dimensional clusters.

Two software products have been used for this work. MOLGEN [1] allows to compute the complete set of connectivity isomers corresponding to a given brutto formula. The construction is redundancy free and complete (Table 1). Generation of isomers can be restricted by a *goodlist* (substructures that have to be contained in the isomers) and by a *badlist* (forbidden substructures). Furthermore, for instance lower and upper limits can be defined for valence and ring size. Output of the generated structures can be directed to an ASCII file in Molfile format.

* correspondence author (Email: kvarmuza@email.tuwien.ac.at)

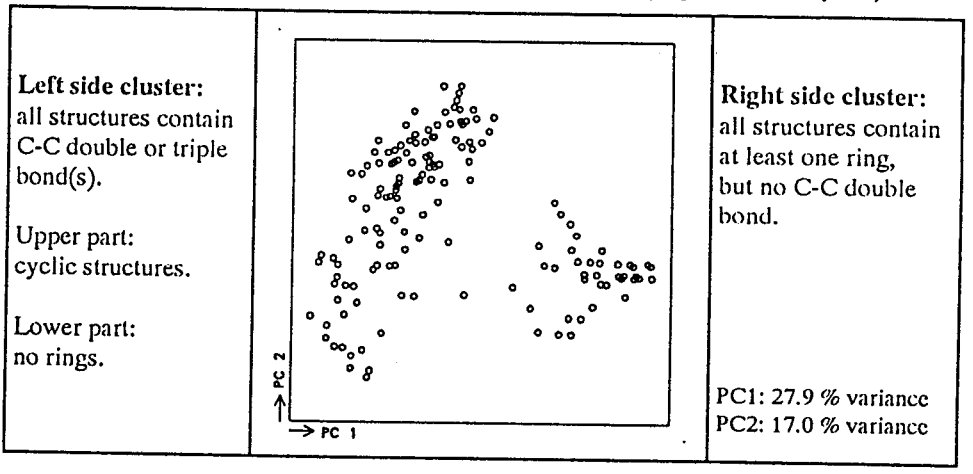
Table 1. Examples of isomer generation with MOLGEN (PC-486, 40 MHz).

formula	number of isomers	computing time (s)
C5 H8 O	205	0.2
C15 H32	4347	29
C8 H10 O3	3869189	3433

The other software used was ToSiM [2] containing a number of methods of computer chemistry, such as a fast substructure search, determination of maximum common substructures, determination of topological equivalent atoms and bonds and of asymmetric carbon atoms in a molecule. Chemical structures can be imported from files in the Molfile format. A set of 165 binary molecular descriptors is used to characterize chemical structures; a part of them has been developed for the mass spectroscopic database *MassLib* [3]. The descriptors can be divided into six groups (A to E): 26 descriptors for Aromatic compounds, 9 for Branches in chains and non-aromatic rings, 30 for non-aromatic Cycles, 41 for a characterization of the neighborhood of Double bonds between carbon atoms, 27 for particular Elements, and 32 for Functional groups.

For principal component analysis a subset (typical 10 to 30) of the descriptors having maximum variance is used. Interactive investigation of the scatter plots (Figure 1) is supported by software tools, allowing for instance the display of chemical structures or descriptor definitions. Clusters can be selected manually and investigated by a search for large and frequently occurring common substructures.

Figure 1. Cluster analysis by PCA of all 205 isomers from C_5H_8O (20 descriptors).



[1] Grund R., Kerber A., Laue R.: *MOLGEN, ein Computeralgebra-System für die Konstruktion molekularer Graphen,* MATCH **27**, 87-131 (1992).

[2] Scsibrany H., Varmuza K.: *ToSiM: PC-Software for the Investigation of Topological Similarities in Molecules,* in: Software-Development in Chemistry **8**, Gesellschaft Deutscher Chemiker: Frankfurt am Main, 235-249 (1994).

[3] Henneberg D., Weimann B., Zalfen U.: *Computer-Aided Interpretation of Mass Spectra Using Databases with Spectra and Structures. I. Structure Searches,* Organic Mass Spectrometry **28**, 198-206 (1993).

COMBINATORIC GENERATION OF REACTION MECHANISM AND ITS REDUCTION BY VARIOUS METHODS

J. Wágner†, T. Vidóczy, Zs. Kúti, K. Héberger, A. Németh and D. Gál
Central Research Institute for Chemistry of the Hungarian Academy of Sciences
POB. 17, H-1525 Budapest, HUNGARY
†Informatics Centre of Semmelweis University of Medicine
H-1089 Budapest, HUNGARY

Summary: The program generates all stoichiometrically possible reactions on the basis of the species (reactants, intermediates, products) with subsequent exclusion of reactions according to complexity considerations and thermochemical constraints.

The program written in Borland-Pascal language is built from the following blocks:

1. Interface to user. The input of formulae, structure and thermochemical properties of reactants and products is user friendly, it requires no special computer knowledge. The structure of the species including multiplicity of bonds between atoms or atomic groups has to be given. A library of species is built to speed up the input of new set of species. Thermochemical data, if available, can also be given.
2. Generation of all stoichiometrically possible reactions. Only those reactions can be included, in which all the products are involved in the list of species given in block 1.
3. Exclusion of reactions exceeding the limit of complexity. The complexity of reactions can be characterized by two approaches: (i) The pair of atoms between which there is a change of multiplicity of chemical bonds, *i.e.* the number of pairs of atoms (κ) between which "something happened"; (ii) the number of pairs of atoms (χ) between which a chemical bond has been formed, without a bond being present prior to the reaction. In both cases theoretical consideration set a limit ($\kappa < 5$, $\chi < 2$) above which the reaction is not considered as an elementary one, thus it should be excluded from the possible mechanism.
4. Exclusion of reactions which are thermochemically unfavourable. Thermochemical considerations provide another, independent way to reduce the mechanism: (i) above a certain limit of free energy the reaction may be considered as unprobable; (ii) if comparing reactions with identical reactants, a reaction can be neglected below a certain limit of exothermicity. The temperature dependence of thermochemical data (heat of reaction, free energy) is taken into account.
5. Beside these exclusions the program can filter out reactions according to the number of reactants and products involved in elementary reactions.
6. Output of the results. List of elementary reactions at any stages of filtering can be obtained on monitor, in printed form or on a floppy disk.

The program is under development, the graph representation of the reaction routes is being built in, which enables the presentation of even very complex reactions in an easily comprehensible way.

THE INTERSECTION OF TWO REGRESSION LINES

Jana Kubanova, Stanislav Kolda, Bohdan Linda
Univerzita Pardubice

Summary: The aim of this contribution is to find both analytical solution of interval estimate x-coordinate of two regression lines at various dispersions and this interval by Monte Carlo method.

The matematical statistics methods are used as well at the evaluation of the chemical experiments. The use of some methods can make dificulties and problems to express exactly the final results. Many such problems can be solved by developing a simulation.

The basic idea of the Monte Carlo methods is the relationship among the characteristics of various random variables and events, that describe the investigating process in analytical form.

Instead of calculation many complicated analytical expressions it is possible to find in experimental way the values corresponding characteristics and by the help of them to calculate further parameters of investgating process.

This method has found wide applicability with the introduction computers in various scientifics branches.

For example in chemistry when we analyse the titration curves, we can meet the problem to find the intersection of two regression lines. We have used both these procedures to solve this problem.

The analytical way of solution:
Let's consider two idependent regression models

$$y^{(1)} = x^{(1)}\beta^{(1)} + \epsilon^{(1)}$$
$$y^{(2)} = x^{(2)}\beta^{(2)} + \epsilon^{(2)}$$

where $\epsilon^{(1)}$ and $\epsilon^{(2)}$ are two independent column random vectors with distribution $N(O, KG^2I)$, $K > 1$ and $N(O, G^2I)$.
The matrix $x^{(1)}$ of the type $(n_1, 2)$ and the matrix $x^{(2)}$ of the type $(n_2,2)$ are known and their values are $h(x^{(1)}) = h(x^{(2)}) = 2$.
We will mark the vectors of unknown parameters $\beta^{(1)}$ and $\beta^{(2)}$ this way.

$$\beta^{(1)} = \begin{bmatrix} \alpha_1 \\ \beta_1 \end{bmatrix} \qquad \beta^{(2)} = \begin{bmatrix} \alpha_2 \\ \beta_2 \end{bmatrix}$$

The estimates of unknown parameters that we won by the help of the method of least squares (from two independent random samples of sizes n_1 and n_2 we'll mark

$$\hat{\alpha}_1 = a_1, \quad \hat{\alpha}_2 = a_2, \quad \beta_1 = b_1, \quad \beta_2 = b_2$$

$$k\sigma^2 = s_1^2 \cdot \frac{Se^{(1)}}{n_1 - 2}\sigma^2 = s_2^2 = \frac{Se^{(2)}}{n_2 - 2}$$

Let x-coordinate of the intersection of these two lines is ξ. If we choose the random variable Z in the form

$$Z = \bar{y}_1 + b_1(\xi - \bar{x}_1) - \bar{y}_2 - b_2(\xi - \bar{x}_2)$$

then it is simple to prove, that Z has normal distribution $N(O, (C_1 + C_2)\sigma^2)$ where

$$C_1 = k\left(\frac{1}{n_1} + \frac{(\xi - \bar{x}_1)^2}{\sum (x_j^{(1)} - \bar{x})^2}\right), \quad C_2 = \frac{1}{n_2} + \frac{(\xi - \bar{x}_2)^2}{\sum (x_j^{(2)} - \bar{x}_2)^2}$$

and that random variable

$$\frac{Se^{(1)} + kSe^{(2)}}{k\sigma^2}$$

has χ^2 (chi square) distribution for $n_1 + n_2 - 4$ degrees of freedom. We'll use the usual way of solution and show, that the random variable

$$\left(\frac{Z\sqrt{k}\sqrt{n_1 + n_2 - 4}}{\sqrt{C_1 + C_2}\sqrt{S_e^{(1)} + kS_e^{(2)}}}\right)^2$$

has F distribution for 1 and $n_1 + n_2 - 4$ degrees of freedom.

If we choose the level of significance γ, we get with the solution of the relation

$$P(F_1, n_1 + n_2 - 4 < F_1, n_1 + n_2 - 4, \gamma) = 1 - \gamma$$

100(1-γ) percentage confidence interval for value ξ.

The simulation way of solution (Monte Carlo method):
The program showing the calculation of interval estimate by Monte Carlo method is based in generalisation coefficients a_i, b_i, i = 1,2 of regression lines by the help of relevant distribution.

For every generalized foursome coefficients it is calculated x-coordinate of intersection of respective lines. (1 - α)100 percentage of these values makes then the searching interval estimate.

The program and the results of comparison will be released during the conference.

CONTENTS

Plenary Lectures

Lectures and posters

Complete Alphabetic Index of Authors of Contributed Papers

List of Addresses

Prof. Christian Affolter
Dept. of Analytical Chemistry
University of Bern
Pharmaceutical Institute
Balzerstrasse 5
CH-3012 Bern, Switzerland

Dr. Jose Manuel Andrade
Dept. of Analytical Chemistry
University of La Coruna
A Zapateira s/n
E-15071 A Coruna, Spain

Prof. Jaromir Antoch
Dept. of Probability and Statistics
Charles University
Sokolovská 83
CZ-186 00 Praha 8, Czech Republic

Dr. C. Arendt
Inst. für Anorganische Chemie
und Strukturchemie I
Heinrich Heine Universitat Duesseldorf
Universitatsstrasse 1
D-40225 Duesseldorf, Deutschland

Dr. Soledad Astorga España
Dept. of Analytical Chemistry
University of La Laguna
AV Trinidad S/N.
38071 La Laguna - Tenerife, Spain

Dr. Peter Boehm
Food Science and Toxicol WBB
SCHERING Aktiengesellschaft
Ernst Schering Str.14
D-59179 Bergkamen, Deutschland

Dr. Igor Bondarenko
Department of Chemistry
Universitaire Instelling Antwerpen
Universiteitsplein 1
B-2610 Antwerpen, Belgium

Ing. Vítězslav Centner
Dept. Biomedische en Farmac. Analyse
Vrije University Brussel
Fac. Geneskunde en Far.
Laarbecklaan 103
B-1090 Brussels, Belgium

Prof. J. T. Clerc
Department of Pharmacy
University of Bern
Baltzerstrasse 5
CH-3012 Bern, Switzerland

Dr. Tibor Cserháti
Central Research Institute for Chemistry
Hungarian Academy of Sciences
P. O. Box 17,
H-1525 Budapest, Hungary

Prof. Dr. Klaus Danzer
Friedrich-Schiller-Universität Jena
Inst. für Anorg. und Anal. Chemie
Lessingstr. 8
D-07743 Jena, Deutschland

Dr. Marie Francoise Devaux
INRA - LTAN
Centre de Recerche de Nantes
Rue de la Geraudiere, B. P. 1627
F-44316 Nantes Cedex 03, France

Dr. Giuliana Drava
Institute of Food and Pharmaceutical Analysis
University of Genova
Via Brigata Salerno (ponte)
I-16147 Genova, Italy

Dr. B. Drenker
Inst. für Anorganische Chemie
and Strukturchemie I
Heinrich Heine Universitat Duesseldorf
Universitatsstrasse 1
D-40225 Duesseldorf, Deutschland

Dr. Diana Elerle
Laboratory of Hygienic Investigation
National Environment Health Centre
Klijánu Str. 7
LV 1012 Riga, Latvia

Dr. Esther Forgács
Separation Science
Central Res. Inst. for Chemistry
Hungarian Academy of Sciences
P. O. Box 17
H-1525 Budapest, Hungary

Dr. Ildikó E. Frank
Jeril Inc.,
790 Esplanada, Stanford CA 94305
USA

Prof. Montelongo F. García
Depart. Química Analitica
Facultad de Química
Universidad de La Laguna
E-38271 La Laguna, Spain

Dr. Paul Geladi
Department of Chemistry
University of Umea
S-90187 Umea
Sweden

Dr. Werner Guenther
InCom Bureau
Sentaweg 16
D-40468 Duesseldorf, Deutschland

Prof. Dr. Gerhard Hägele
Inst. für Anorganosche Chemie
und Struktur Chemie
Heinrich Heine Universitat Duesseldorf
Universitatsstrasse 1
D-40225 Duesseldorf, Deutschland

Prof. RNDr. DrSc. Josef Havel
Katedra analytické chemie
Přírodovědecká fak. Masarykovy university
Kotlářská 2
CZ-611 37 Brno, Czech Republic

Mgr. Pavel Havelec
Dept. of Analytical Chemistry
Faculty of Sciences, Charles University
Albertov 2030
CZ-128 40 Praha 2, Czech Republic

Dr. Károly Héberger
Department of Oxidation Processes
Central Research Institute for Chemistry,
Hungarian Academy of Sciences, POB 17
Pusztaszeri út 59-67
H-1525 Budapest, Hungary

Ing. Petr Homola
odd. hygienických laboratoří
Okresní hygienická stanice
Jana Černého 361
CZ-502 41 Hradec Králové, Czech Republic

Dr. István Horváth
Depart. of Inorg. and Analytical Chemistry
A. József University, P. O. Box 440
H-6701 Szeged, Hungary

Prof. Agnar Hoskuldsson
Dept. of Applied Construction & Production
Danish Technical University
Akademivej Build 358
DK-2800 Lyngby, Denmark

Ing. Jaroslav Hovorka
Department of Environmental
Spolchemie, a. s.
Revoluční 86
CZ-400 32 Ústí nad Labem, Czech Republic

Ing. Jaromír Chládek
Oddělení kontroly jakosti 1, TřDPCH
CHEMOPETROL, a. s., divize petrochemie
436 70 Litvínov, Czech Republic

Dr. Alfred Christy
Dept. of Chemistry
University of Bergen
Allegt. 41
N-5007 Bergen, Norway

Prof. Ing. DrSc. Pavel Jandera
Department of Analytical Chemistry
University of Pardubice
Fac. of Chem. Technology
nám. Čs. legií 565
CZ-53210 Pardubice, Czech Republic

Dr. Leopold Jirovetz
Analytical Service
Institute of Pharmaceutical Chemistry
University Vienna, Pharmacy Centre
Althanstreet 14
A-1090 Vienna, Austria

Mgr. Stanislav Keprta
Department of probability and math. statistics
Faculty of Mathematics and Physics
Charles University
Sokolovská 83
CZ-186 00 Praha 8, Czech Republic

Prof. Ing. CSc. Stanislav Kolda
Katedra matematiky
Univerzita Pardubice
Studentská 84
CZ-530 09 Pardubice, Czech Republic

Doc. Ing. CSc. Karel Komers
Physical Chemistry
University of Pardubice
nám. Čs. legií 565
CZ-532 10 Pardubice, Czech Republic

Dr. Alois Kozubík
Dept. of Environmental
Faculty of Science, Masaryk University
Kotlářská 2
CZ-611 37 Brno, Czech Republic

Ing. Vladimír Kraják
Hygiene laboratory
OHS
Jana Černého 361
CZ-502 41 Hradec Králové, Czech Republic

Dr. Jan Krejčí
BIOSENSORS
KREJČÍ ENGINEERING
Riegrova 318
CZ-666 01 Tišnov, Czech Republic

Assoc. Prof. Mikael Kubista
Dept. of Biochemistry and Biophysics
Lundberg laboratory, Gothenburg University
Medicinare g. 9c
S-41390 Gothenburg, Sweden

Ing. Karel Kupka
Manager
TriloByte Statistical Software Ltd.
U sokolovny 21
CZ-530 02 Pardubice, Czech Republic

Prof. Olav V. Kvalheim
Department of Chemistry
University of Bergen
Allegt. 41
N-5007 Bergen, Norway

Prof. Ing. DrSc. Vladimír Kvasnička
Department of Mathematics
Fac. of Chem. Technology
Slovak Technical University
Radlinského 9
SK-812 37 Bratislava, Slovakia

Dr. Mojmír Kývala
Katedra anorganické chemie
Přírodovědecká fakulta UK
Albertov 6
CZ-128 43 Praha 2, Czech Republic

Ing. CSc. Lubomír Lacko
Specializovaný státný zdravotní ústav
Cesta k nemocnici 1
SK-975 56 Banská Bystrica, Slovakia

Prof. Barry Lavine
Department of Chemistry
School of Science, Clarkson University
121 Science Center Potsdam NY
13699-5810, USA

Dr. CSc. Marie Letkovičová
Department of Epidemiology
Nuclear Power Plants Institute
Okružná 5
SK-917 00 Trnava, Slovakia

Doc. Ing. Ján Mikuláš Lisý
Katedra fyzikální chemie
Chemickotechnol. fakulta
Slovenská technická univerzita
Radlinského 9
SK-812 37 Bratislava, Slovakia

Ing. CSc. Marek Liška
Institute of Inorganic Chemistry
Slovak Academy of Sciences
Martina Rásuza 10
SK-911 01 Trenčín, Slovakia

Studentka Milada Lubalová
Katedra analytické chemie
Přírodovědecká fakulta MU
Kotlářská 2
CZ-611 37 Brno, Czech Republic

Doc. Ing. CSc. Pavel Mach
Faculty of Electrical Engineering
Czech Technical University
Technická 2
CZ-166 27 Praha 6, Czech Republic

Prof. Ing. DrSc. Mikuláš Matherny
Katedra chemie
Hutnická fak., Technická universita
Letná 9
SK-04385 Košice, Slovakia

Dr. Vladimír Matzner
Státní zkušební ústav lehkého průmyslu
Čechova 59
CZ-370 65 České Budějovice, Czech Republic

Ing. CSc. František Pudil
Food Chemistry and Analysis
Institute of Chemical Technology
Technická 5
CZ-166 28 Praha 6, Czech Republic

Doc. Ing. DrSc. Oldřich Pytela
Katedra organické chemie
Fakulta chemickotechnologická
Univerzita Pardubice
nám. Čs. Legií 565
CZ-532 10 Pardubice, Czech Republic

Dr. E. Mostafa Qannari
SMAD
ENITIAA
La Geraudiere,
F-44072 Nantes Cedex, France

Dr. Stefan Rannar
Department of Organic Chemistry
Umea University
S-90187 Umea, Sweden

Ing. CSc. Bohuslav Rittich
Faculty of Chemistry
Technical University
Veslařská 230
CZ-637 00 Brno, Czech Republic

Dr. Willem G. de Ruig
Dienst Landbouwkundig Onderzoek
RIKILT-DLO, Postbus 230
6700 AE Wageningen
The Netherlands

Dr. Maria Sandberg
Department of Organic Chemistry
Umea University
S-90187 Umea, Sweden

Dr. Ing. J. Dirk Schladot
Environmental Specimen Bank
Research Centre Juelich
Inst. of Appl. Phys. Chemistry
Stettenicher Forst
52425 Juelich, Deutschland

Dr. Ir. Henri C. Smit
Laboratory for Analytical Chemistry
University of Amsterdam
Nieuwe Achtergracht 166
1018 WV Amsterdam, Netherlands

Ing. Andrea Součková
Dept. of Physical Chemistry
University of Pardubice
Fac. of Chem. Technology
nám. Čs. legií 565
CZ-532 10 Pardubice, Czech Republic

Dr. Hilde Strom
Norsk Hydro a. s.
Research Centre Porsgrunn
POB 2560,
N-3901 Porsgrunn, Norway

Mgr. Daniel Svozil
Dept. of Analytical chemistry
Faculty of Sciences, Charles University
Albertov 2030
CZ-128 40 Praha 2

Ing. Darina Szalayová
chemik-metodik
VSŽ LABORTEST s. r. o.
SK-04454 Košice, Slovakia

Prof. Pál Szepesvary
Soroksari ut. 38 - 40, V.11
H-1095 Budapest, Hungary

Doc. RNDr. Jiří Ševčík
Dept. of Analytical Chemistry
Faculty of Sciences, Charles University
Albertov 2030
CZ-128 40 Praha 2, Czech Republic

Student David Široký
Katedra analytické chemie
Přírodovědecká fak. Masarykovy university
Kotlářská 2
CZ-611 37 Brno, Czech Republic

Doc. Ing. CSc. Zdeněk Šmejkal
Katedra anorganické technologie
Univerzita Pardubice
nám. Čs. legií 565
CZ-532 10 Pardubice, Czech Republic

Ing. Ivan Švancara
Katedra analytické chemie
Univerzita Pardubice
nám. Čs. legií 565
CZ-53210 Pardubice, Czech Republic

Prof. Roberto Todeschini
Dept. of Environmental Science
University of Milan
Via Emanueli 15
I-20126 Milano, Italy

Dr. Vladislav Tomišič
Laboratory of Physical Chemistry
Faculty of Science, University of Zagreb
Maruličev trg 19/II, POB 163
HR-41001 Zagreb, Croatia

Dr. Boris Treiger
Dept. of Chemistry
Universitaire Instelling Antwerpen
Universiteitsplein 1
B-2610 Antwerpen, Belgium

Prof. Dr. Kurt Varmuza
Department for Chemometrics
Technical University of Vienna
Lehargasse 4/152,
A-1060 Vienna, Austria

Dr. C. Verwey
Inst. für Anorganische Chemie
und Strukturchemie
Heinrich Heine Universitat Duesseldorf
Universitatsstrasse 1
D-40225 Duesseldorf, Deutschland

Dr. Evelyne Vigneau
SMAD
ENITIAA
La Geraudiere,
F-44072 Nantes Cedex 03, France

Ing. František Všianský
OKB
Zavodní nemocnice NH
Pavlovova 31
CZ-704 00 Ostrava 3, Czech Republic

Prof. Dr. Wolfhard Wegscheider
General and Analytical Chemistry
University of Mining and Metallurgy
Franz-Josef Strasse 18
A-8700 Leoben, Austria

Dr. Wolfgang Werther
Institute of Chemometrics
Technical University of Vienna
Lehargasse 4I152
A-1060 Vienna, Austria

BENZINA

CHRAŇTE SVÉ ŽIVOTNÍ PROSTŘEDÍ

•

BEZOLOVNATÉ BENZINY

BA 95N NATURAL

BA 98N NATURAL SUPER PLUS

V SÍTI ČERPACÍCH STANIC A.S. BENZINA

farmakon
OLOMOUC

PRODUCTION PROGRAMME

- pharmaceutical substances

antidepressants:	*dothiepin hydrochloride* *amitriptyline hydrochloride* *nortriptyline* *selegiline*
cardiovascular drugs:	*diltiazem hydrochloride* *naftidrofuryl oxalate* *xanthinol nikotinate*
anticoagulants:	*ethyl biscoumacetate* *warfarin* *hymecromone*
ascorbic acid	

- intermediates for pharmaceutical and chemical industry

- fine chemicals

- custom manufacturing of organic compounds

- disinfectans for sanitary use

FARMAKON, Na Vlčinci 3, 771 17 OLOMOUC, CZECH REPUBLIC
tel.: +42 (0)68 / 522 3941, Fax: +42 (0)68 / 289 60

a.s. • 747 70 Opava 9 • tel.0653/443 111 • fax 0653/31 104

FARMACEUTICKÝ PODNIK

s dlouholetou tradicí

Vyrábí léčivé přípravky
humánní a veterinární,
farmaceutické substance,
zejména námelové alkaloidy,
silymarin, silybinin, ciclosporin,
morfinanové alkaloidy
a jejich polosyntetické analogy.
Mezi tradiční
výrobní sortiment
náleží rovněž
extrakty,
léčebná kosmetika
a nápojové sirupy.

- KORAMO Kolín is the largest producer of lube oils, greases and waxes in the Czech Republic with a product range of over 150 lubricants of a trade mark **MOGUL**.
- KORAMO KOLÍN is a highly competitive manufacturer with much experience of over ninety years. From 1929 to Second World War was KORAMO a subsidiary of the Vacuum Oil Company.
- KORAMO has a workforce of about 600 people.
- KORAMO supplies the 50% of the Czech lube oil market and the 80% of the Czech wax market with its premium quality products.
  Chemopetrol Litvínov Refinery is the exclusive supplier of a hydrocracked feedstock for KORAMO.
- KORAMO invests substantially to maintain the high technical standard of its technology processes. Since 1990, KORAMO has spent US $ 15 million on the production facilities modernization, new incinerating plant and on the initiation of an integral computer control system construction.
- Kolín is situated in central Bohemia about 50 kms of Prague.

MOGUL

CHEMOPETROL

Pánevní oblast severu Čech se stala jedním z nejlidnatějších území našeho státu. Stěžejním průmyslovým podnikem tohoto regionu specializovaným na těžkou chemii je akciová společnost Chemopetrol Litvínov s hlavním předmětem podnikání - výrobou a prodejem motorových paliv, plastů a výrobků souvisejících s hlubokým zpracováním ropy.

Původně německá továrna, s jejíž výstavbou se začalo v roce 1939, sloužila k výrobě motorových paliv na bázi severočeského hnědého uhlí. Od druhé poloviny 50. let byl zahájen a v roce 1972 ukončen postupný přechod na ropu. V současné době je Chemopetrol, a. s. schopen za rok zpracovat pět miliónů tun této suroviny a svým obratem se řadí na čtvrté místo v České republice.

Vytvořením akciové společnosti 1. ledna 1994 začala rozsáhlá vnitřní restrukturalizace dříve státního podniku. V rámci diverzifikace se vyčleňují neklíčové aktivity a vznikají tak dceřiné společnosti, jichž do současného období pracuje osm, šest z nich má sídlo přímo v areálu. Přechodem na uvedený systém se tak snížil počet zaměstnanců na nynějších 6 900 osob, aniž by došlo k rozsáhlému propouštění.

Negativním rysem činnosti Chemopetrolu je ekologické zatěžování oblasti exhalacemi, zejména při spalování energetického uhlí, proto přijala akciová společnost ekologický program zaměřený především na ochranu ovzduší. Jeho hlavním cílem je modernizace současných technologií a tím vyloučení negativního vlivu na zdraví obyvatel severočeské oblasti. Aktivním přístupem k ochraně životního prostředí společnost soustavně zabezpečuje celkové snížení ekologické zátěže.

Chemopetrol, akciová společnost se výrazným způsobem podílí na tvorbě sociálního prostředí severočeské oblasti. Sponzoruje kulturní a sportovní akce, zdravotnictví a vzdělávání. Usiluje tak o dobré vztahy s orgány státní správy, samosprávy a sousedními podniky.

CHEMOPETROL

LACHEMA, akciová společnost
621 33 Brno-Řečkovice, Karásek 28
Česká republika

Lachema a.s., přední farmaceutický výrobce s tradicí i budoucností

V roce 1951 byla založena Lachema, národní podnik, s cílem pokrýt trh v oblasti čistých, laboratorních a speciálních chemikálií. Rozšiřováním výrobního programu se Lachemě nejen podařilo tohoto cíle dosáhnout, ale zároveň, postupnou specializací, se stala i předním farmaceutickým výrobcem na našem trhu.

V roce 1991 vznikla akciová společnost Lachema, se sídlem správy společnosti v Brně. K brněnské aglomeraci, která je tvořena Výzkumným ústavem čistých chemikálií a třemi výrobními divizemi - farmaceutickou, diagnostickou a chemickou, patří ještě odštěpné závody v Blansku a Neratovicích.

Svým širokým polem působnosti, technickou vybaveností a vstřícným jednáním se Lachema a.s. řadí k nejlepším a nejvyhledávanějším podnikům, které jsou schopny vyhovět požadavkům svých zákazníků k jejich spokojenosti.

Lachema a.s., pharmaceutical producer with tradition and future

Lachema, state enterprise, was established in 1951 with the goal of covering the market in the field of pure, laboratory and specialty chemicals. By expanding its manufacture programme.

Lachema is not only succeeded in reaching this goal but also, due to gradual manufacture specialization, became the leading manufacturer of pharmaceuticals on the market.

In 1991, the joint stock company Lachema was established, with the seat of the company's management located in Brno. To the Brno company's agglomeration consisting of the Research Institute of Pure Chemicals and three manufacturing divisions - pharmaceuticals, diagnostic and chemical, belong also the subsidiary plants at Blansko and Neratovice.

Through its wide range of activities, technical facilities and friendly approach, Lachema a.s. ransk among the leading and most popular companies which are able to meet to satisfaction their customer's requirements.

OSTACOLOR a.s.

533 54 PARDUBICE - RYBITVÍ

phone: 040/47311
fax: 040/42107

EXPORT AND IMPORT OF ORGANIC DYES, PIGMENTS AND INTERMEDIATES

OSTACOLOR a.s.

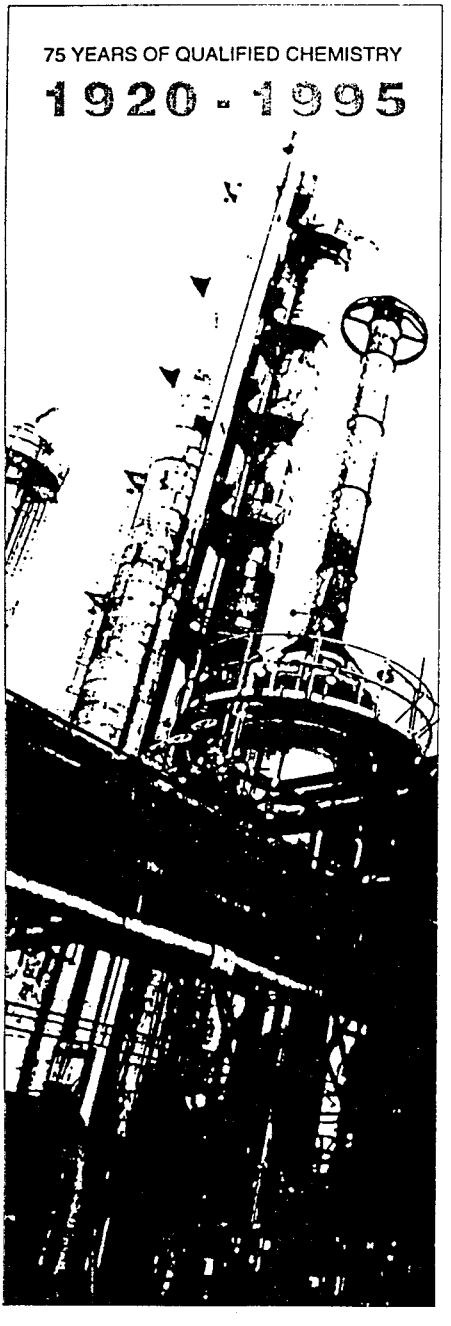
533 54 PARDUBICE - RYBITVÍ

phone: 040/47311
fax: 040/42107

PRODEJ, EXPORT, IMPORT ORGANICKÝCH BARVIV A PIGMENTŮ

EXPORT A IMPORT ORGANICKÝCH POLOTOVARŮ

Synthesia



HISTORY

The Czechoslovak joint-stock factory for explosives was established in the year 1920. Factory for production of organic and drugs was founded in the following year (1938). In the year 1942 the factory for production of plastic materials was organized. Transformation to the joint-stock company "SYNTHESIA" has been completed in the year 1994.

SYNTHESIA
THE GREATEST PRODUCER OF THE QUALIFIED CHEMISTRY IN THE CZECH REPUBLIC

LIST OF PRODUCTS
(more than 1000 products)

- Organic dyes and pigments
- Inorganic chemicals
- Fertilizers
- Pesticides
- Basic organic products and intermediates
- Pharmaceutical fine chemicals
- Special chemicals
- Explosives
- Plastics

SYNTHESIA a.s., 53217 Pardubice-Semtín, CZECH REPUBLIC
Phone: +42 (0)40 6821111 Telefax: +42 (0)40 46465 Telex: 194235 vchz c