INTERNATIONAL UNION OF PURE
AND APPLIED CHEMISTRY

ANALYTICAL CHEMISTRY DIVISION
COMMISSION ON GENERAL ASPECTS OF ANALYTICAL CHEMISTRY*

GUIDELINES FOR CALIBRATION
IN ANALYTICAL CHEMISTRY

PART 1. FUNDAMENTALS AND SINGLE COMPONENT CALIBRATION

(IUPAC Recommendations 1998)

Prepared for publication by

KLAUS DANZER

Institute of Inorganic and Analytical Chemistry
Friedrich Schiller University of Jena, Lessingstr. 8, D-07743 Jena, Germany

LLOYD A. CURRIE†

Chemical Science and Technology Laboratory
National Institute of Standards and Technology, Gaithersburg, MD 20899, USA

Guideline for calibration in analytical chemistry— Part 1. Fundamentals and single component calibration

Synopsis

This IUPAC nomenclature document has been prepared to establish a uniform and meaningful approach to terminology, notation, and formulation for calibration in analytical chemistry. In this first part, general fundamentals of calibration are presented, namely for both relationships of qualitative and quantitative variables (relations between variables characterizing certain types of analytes and measured signals in certain positions of a measured function on the one hand and between variables characterizing the amount or concentration of the chemical species and the intensities of the measured signals, on the other hand). On this basis, the fundamentals of the common single component calibration which models the relationship $y = f(x)$ between the signal intensities y and the amounts or concentrations x of the analyte under given conditions are represented. Additional papers will be prepared dealing with extensive relationships between several signal intensities and analyte contents, namely with multivariate calibration and with optimization and experimental design.

CONTENTS

1. INTRODUCTION

In general, calibration is an operation that relates an output quantity to an input quantity for a measuring system under given conditions.

In the chemical measurement process (CMP [1]), input quantities are given by analytical quantities characterizing certain types of analytes (chemical species) q_i and their amounts or concentrations x. Output quantities are represented by measured values, i.e., certain signals in positions z_i with intensities y_{zj}. The common case, in which a relationship has to be determined by calibration between the amount (content) of a certain analyte x_{qi} and a signal intensity y_{zj}, is only a special case of calibration.

In a more general sense, calibration in analytical chemistry refers to the relation between an analytical function $x = f(q)$ representing a pattern of analytes and their amounts or concentrations in a given test sample on the one hand (see Fig. 1, left-hand side) and a measured function $y = g(z)$ that may be represented by a spectrum, chromatogram etc. (Fig. 1, right-hand side) [2].
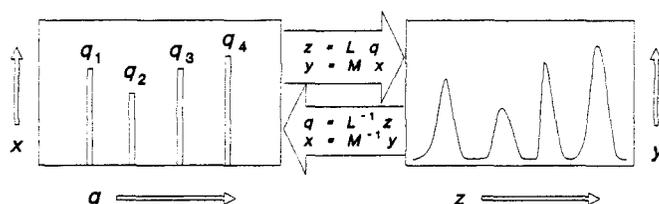
Fig. 1 Relation between the analytical function $x = f(q)$ and the measured function $y = g(z)$.

Therefore, there are four quantities q, x, z, and y that have to be related with one another. The situation is characterized in Fig. 2 by a quasi-four-dimensional representation. The foreground of the representation depicts the relationship between the species and its characteristic signal, while behind that, the relationship between signal and concentration is established. Taken together, these relationships establish the composition of the sample. These general connections are reflected in three analytically relevant applications:

(1) Calibration of quantities that characterize typical signal positions z (q-z-calibration) for identification of components and qualitative analysis

$$z = f(q) + e_z \approx L \cdot q + e_z \tag{1}$$

where f represents the underlying functional relationship, e_z the measurement error of z, and L the (approximately) linear operator that transforms q (component-specific quantity like atomic number, mass number, or typical energy values) into z.

(2) Calibration of quantities which characterize the intensity of an observed response y (y-x-calibration) for a given analyte q_i in quantitative single component analysis

$$y = F(x) + e_y \approx M \cdot x + e_y \tag{2}$$

where F represents the underlying functional relationship (the calibration function in a more

Fig. 2. Quasi-four-dimensional representation of the connection between qualitative and quantitative calibration. The q-z-relationship corresponds to a deterministic function, see (i). Frequently there exist empirical relationships like (ii) and (iii)
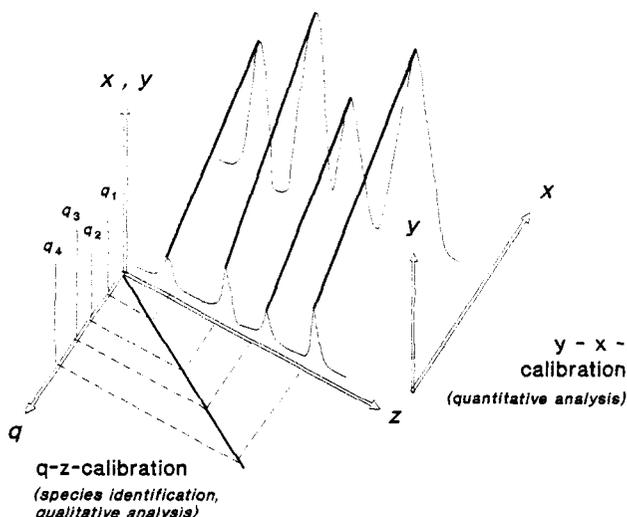
narrow sense), e_y the measurement error of y, and M is the (approximately) linear operator that transforms x into y.

(3) Quantitative multicomponent calibration

$$Y = A \cdot X + E \tag{3}$$

where Y represents the matrix of measured values, X the matrix of analyte amounts (contents), A the sensitivity matrix that transforms X into Y, and E is an error matrix. Quantitative multicomponent calibration is carried out by means of multiple or multivariate regression techniques, respectively, and will be the subject of a second paper.

2. FUNDAMENTALS

Calibration in Analytical Chemistry is the operation that determines the functional relationship between measured values (signal intensities y at certain signal positions z_j) and analytical quantities characterizing types of analytes q_i and their amount (content, concentration) x. Calibration includes the selection of the model (its functional form), the estimation of the model parameters as well as the errors, and their validation.

2.1 Calibration function for species identification and qualitative analysis (q-z calibration, more specifically: calibration of such analytical parameters which characterize types of chemical species) is the establishment of a model (its parameter estimation and validation) of the relation between z and q for the purpose of **identification** and **qualitative analysis** on the basis of Eq.(1).

$$z = f(q) + e_z \approx L \cdot q + e_z \tag{1}$$

In analytical practice, q-z calibration refers to the position of signals on energy or energy-proportional scales, such as wavelength-, frequency- or mass/charge coordinates of spectrometers or time coordinates of chromatograms, respectively, as a characteristic signal pattern caused by the species present.

The linear operator L in identification and qualitative analysis may be [2]:

(i)   a deterministic function on the basis of natural laws, $z = f_{det}(q)$, e.g., Moseley's law of the dependence of X-ray frequencies on the atomic number [3], or

(ii)   an empirical function, $z = f_{emp}(q)$, like Kovats' indices of homologous compounds in their dependence on retention data in gas chromatography [4], or

(iii)   an empirical connection, $z = emp(q)$, represented by tables and atlases, e.g., by Colthup's table of characteristic vibrations [5], iron-atlases in atomic emission spectroscopy [6], and other types of wavelength tables [7].

While the relations $z = f_{det}(q)$ are known on the basis of natural laws, the estimation of an empirical function $z = f_{emp}(q)$ for the purpose of identification and qualititative analysis is mostly carried out by (linear) least squares to fit the observed z-values for a set of pure component standards or a multicomponent standard (frequently mixed in an "intensity-normalization" relation). On the other hand, calibration on the basis of empirical relationships $z = emp(q)$ in the form of tables, atlases and graphs are developed by classification of experimental results.

2.2 <u>Calibration function for quantitative analysis</u> is the determination of the functional relationship between y and x in the form

$$y = F(x) + e_y \tag{2}$$

where F is the calibration function. In most cases, the calibration function has to take into account the response relations for all relevant constituents and interferences. Then y depends on a vector $\mathbf{x} = (x_a, x_b \ldots x_m, x_n \ldots x_q)$ consisting of the amounts of the analyte under direct interest, x_a, of accompanying components $(x_b \ldots x_m)$ and of influencing factors $(x_n \ldots x_q)$

$$y = F(\mathbf{x}) + e_y \tag{4}$$

Under the best circumstances, Eq.(4) is a linear vector equation. The estimation of models according to Eq.(4) is subject of experimental design and optimization which will be addressed in a third report.

In the simplest case of calibration according to Eq.(2), for a given amount of $x = x_a$ where no other components and factors have to be considered, y is a scalar quantity. More generally, y may be a function of the characterizing variable z. The basic relation then takes the form

$$y(z) = M(z) \cdot x + e_y(z) \tag{4a}$$

and represents a characteristic pattern, e.g. a spectrum of a pure component.

2.3 Evaluation Function [1, 8]

In general, the evaluation function (analytical function) is the inverse of the calibration function, Eq.(2)

$$x = F^{-1}(y) \tag{5}$$

provided that the relationship between measured value y and analyte amount x has been created by calibration which is mainly the case in analytical chemistry. However, there are also other types of evaluation procedures, e.g. on the basis of natural laws, depending on the nature of the analytical method. The determination of the amounts of analytes can be based on absolute, relative or reference measurements [9].

Absolute, definitive and *reference measurements* are based on equations of the general type

$$y = A \cdot x \tag{6}$$

where the sensitivity A in analytical chemistry generally is defined as the differential coefficient dy/dx. In the case of linear models, A is given by $\Delta y/\Delta x$ [1]. For the three mentioned types of analytical measurements the sensitivity is given by mathematically well-defined relations, namely in the case of:

(a)   *Absolute measurements* by fundamental quantities like Faraday constant and quotients of atomic and molar masses, respectively;

(b)   *Definitive measurements* by fundamental quantities in combination with well-known empirical (transferable) constants (e.g. molar absorption coefficient, conductivity at definite dilution, diffusion coefficients for given media) sometimes complemented by an empirical factor (e.g. titre); and

(c)   *Direct reference measurements* by the relation of measured value to concentration (content) of a reference material (R)

$$A = y_R / x_R . \tag{7}$$

Accordingly, absolute measurements do not need (permanent) calibration[1] and definitive and direct reference measurements need only a comparison measurement (e.g. titrimetric standard) or reference measurement (reference material or spiked sample).

On the other hand, *indirect reference measurements* are based on empirical calibration functions, frequently based on linear models

$$y = B + A x + e_y \tag{8}$$

where the intercept B corresponds to the experimental *blank* and the slope A to the experi-

---

[1] Apart from the fact that both the sensitivity constants A and the conditions under which they are valid (e.g. under which a reaction proceeds quantitatively) sometime were found theoretically or experimentally

mental *sensitivity* . The parameters A and B are usually estimated by least squares fitting[2], e_y is the error of the *y*-measurement. In analytical practice, also some methods using definitive measurements, in principle, are calibrated by least squares estimating (e.g. spectrophotometry, polarography) to provide reliable estimates of A.

Calibration functions corresponding to Eq.(8) are not generally transferable over long times and not from one laboratory to another. However in the case of blank-free or blank-corrected relations

$$y = A x + e_y \qquad (9)$$

methods can be robustly calibrated under fixed experimental conditions. The experimental sensitivity coefficients (sensitivity factors) are transferable over time and between laboratories under standardized operating conditions. Because of this transferability, such methods are occasionally called *"standard-free"*. Such standard-free methods have been developed, e.g., in the field of optical emission spectrography [10], in spark source mass spectroscopy [11], and in X-ray fluorescence spectroscopy [12] for semi-quantitative multielement analysis. Standard-free methods are to be distinguished from the *"calibration-free"* methods using absolute measurements as mentioned above [9].

3. LEAST SQUARES CALIBRATION

3.1 Linear calibration model

On the condition that the errors of the measurement have a zero mean and are uncorrelated, a linear function (8) can be fitted to the measured values by means of least squares estimation (LS, or ordinary least squares estimation, OLS, respectively).

With the fundamental relations

* Model: $\quad y_i = B + A x_i + e_{yi} = E(y_i) + e_{yi}$ $\qquad$ (10a)

* Estimate: $\quad \hat{y}_i = \hat{B} + \hat{A} x_i$ $\qquad$ (10b)

* Residual: $\quad d_{yi} = y_i - \hat{y}_i = y_i - \hat{B} - \hat{A} x_i$ $\qquad$ (10c)

(where $E(y_i)$ is the expectation of y_i) the general least square criterion expressed by the sum of squares of deviations, SSD, reads [13]

$$SSD = \sum_{i=1}^{m} [(y_i - \hat{y}_i) / \sigma_i]^2 = \sum_{i=1}^{m} (d_{yi} / \sigma_i)^2 \qquad (11)^{[3]}$$

where σ_i is the standard deviation at the given point i and m the number of calibration

---

[2] Graphical methods are still applied occasionally; in recent time, neural networks also have been used to build experimental calibration models, especially in the case of nonlinear relationships

[3] The sum of squares comes from the Likelihood function
$L = (2\pi)^{-1} \sigma_1^{-1} \sigma_2^{-1} ... \sigma_n^{-1} \exp(-\frac{1}{2} SSD)$
as the product of probabilities that all the measured values y_i correspond with the estimates $\hat{y}_i$ as exact as possible. *SSD* becomes a minimum if L becomes a maximum

measurements (see Eq.(20)).

> Note 1: A corresponding criterion can be formulated for the determination of x from (error-free) y by means of the estimate $\hat{x} = \hat{B}_y + \hat{A}_y\ y$. This model, however, usually has no relevance in analytical calibration.

The *SSD* has to be minimized according to Eq.(12)

$$\sum_{i=1}^{m} (d_{yi} \ / \ \sigma_i)^2 \overset{!}{=} \min \tag{12}$$

The symbol $\overset{!}{=}$ means that the expression on the left-hand side has to be a minimum. Depending on the fulfillment of the following conditions the least squares criterion has to be modified as follows:

(1) The errors are only or essentially in the measured values y as the dependent variable:

$$A\ \sigma_x \ll \ \sigma_y \tag{13}$$

and in addition, the errors σ_y are constant in the several calibration points *(Homoscedasticity)*:

$$\sigma_{y1}^2 = \sigma_{y2}^2 = \ldots = \sigma_{yn}^2 = \sigma_y^2 \tag{14a}$$

or, expressed by the estimates of σ

$$s_{y1}^2 \overset{\alpha}{=} s_{y2}^2 \overset{\alpha}{=} \ldots \overset{\alpha}{=} s_{yn}^2 \overset{\alpha}{=} s_y^2 \tag{14b}$$

where $\overset{\alpha}{=}$ means equality for a given statistical risk of error α. Only in this homoscedastic case and if errors in x can be neglected according to Eq.(13) is the LS criterion reduced to

$$\sum_{i=1}^{m} d_{yi}^2 \overset{!}{=} \min \tag{15}$$

and the classical Gaussian LS criterion (normal or ordinary least squares, LS, NLS, or OLS) can be applied.

(2) In case that the measuring errors σ_y vary and *heteroscedasticity* must be assumed (Eqs.(14) are not valid) the least squares (LS) criterion (12) turns into

$$\sum_{i=1}^{m} (d_{yi} / \sigma_{yi})^2 \overset{!}{=} \min \ . \tag{16}$$

The model of weighted least squares (WLS) results from this criterion as will be shown in paragraph 3.3.

(3) In the most general case, if both variables are subject to error and, therefore, Eq.(13) is not fulfilled, we have

$$\sigma_i^2 \;=\; \sigma_{yi}^2 \;+\; A^2\, \sigma_{xi}^2 \tag{17}$$

In this case, in which there are errors in both variables, viz. the measured value and the analytical quantity (concentration), the sum of the d_{x+y}^2, see Fig.3, has to be minimized and orthogonal least squares fitting must be carried out, e.g. according to paragraph 3.4. The different least squares models that can be calculated are schematically shown in Fig.3.

Which model has to be used in analytical calibration depends on the fulfilment of the conditions mentioned above and on the procedure in calibration.

According to Eq.(2) experimental calibrations are mainly carried out by measurement of a set of calibration samples containing the analyte under investigation in suitably graduated amounts. If possible, analysts use materials whose concentrations are known with maximum reliability, i.e. with both high precision and trueness. In analytical practice, certified reference materials, single- and multi-component standards, and synthetic standard materials are used as calibration samples.
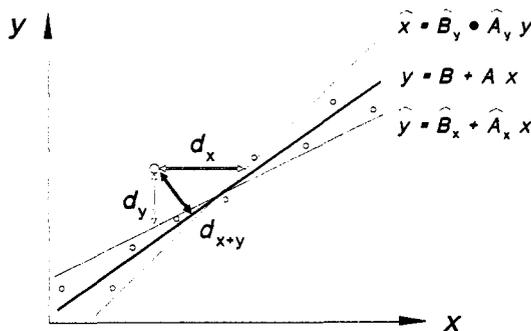

Fig. 3 Different linear least squares models. $\hat{B}_y$, $\hat{B}_x$, $\hat{A}_y$, and $\hat{A}_x$ are the estimates of B_y, B_x, A_y, and A_x

The concentrations (contents) of the calibration samples can be regarded as "true" and error-free or it can be assumed that, according to Eq.(13), the random errors of x are negligible compared with that of y. Under this condition, the specified calibration function (18) has to be used

$$y \;=\; B_x \;+\; A_x\, x \;+\; e_y \tag{18}$$

and the parameters B_x and A_x are estimated by the Gaussian (normal) least squares algorithm in case of homoscedasticity, see [14, 15],

$$\hat{A}_x \;=\; Q_{xy}/Q_{xx} \tag{19}$$

$$\hat{B}_x \;=\; (\Sigma y \,-\, \hat{A}_x \Sigma x)\,/\,m \tag{20}$$

where m is the total number of calibration experiments (index j) when obtaining the calibration function and with the following sums:

$$Q_{xx} = \Sigma (x_j - \bar{x})^2 \qquad = \Sigma x_j^2 - (\Sigma x_j)^2 / m \qquad (21a)$$

$$Q_{yy} = \Sigma (y_j - \bar{y})^2 \qquad = \Sigma y_j^2 - (\Sigma y_j)^2 \ m \qquad (21b)$$

$$Q_{xy} = \Sigma (x_j - \bar{x})(y_j - \bar{y}) = \Sigma (x_j\ y_j) - m\ \bar{x}\ \bar{y} \qquad (21c)$$

The correlation coefficient

$$r_{xy} = Q_{xy} / \sqrt{Q_{xx}\ Q_{yy}} \qquad (22)$$

which is a measure of relationship of _two random variables_, has no meaning in calibration under the conditions mentioned above because the values x are not random quantities in the calibration experiment.

> Note 2: The correlation coefficient r_{xy} can be meaningful for the relationship between random variables, but it should not be used in calibration [1].

For the evaluation of analytical measurements usually the inverse function of the calibration function (Eq.(18) is applied

$$\hat{x} = (y - \hat{B}_x) / \hat{A}_x \qquad (23)$$

provided that the requirements (1) and (2) mentioned above are correct.

> Note 3: In reality, the relationship between the measured values y and analyte amounts (concentrations) x has to be characterized by means of a three-dimensional calibration model [2]:
>
> $$y = f(x_{true}, x_{estm}) \qquad (i)$$
>
> where x_{true} is the concentration of the (certified) reference materials used for calibration and considered to be "true" ("error-free"). On the other hand, $x_{estm} = \hat{x}$ is a random variable affected by errors. Whether the condition (13) is fulfilled for x_{estm} or not, cannot be fixed _a priori_.
>
> From the three-dimensional model (i) the following three two-dimensional relations result:
> the _calibration function_:
>
> $$y = f_C(x_{true}) + e_y \qquad (ii)$$
>
> e.g. according to Eq.(18),
>
> the _analytical evaluation function_:
>
> $$x_{estm} = f_A(y) + e_x \qquad (iii)$$
>
> e.g. following Eq.(23), and
>
> the _validation function_ (bias function, recovery function):
>
> $$x_{estm} = f_V(x_{true}) + e_x \qquad (iv)$$
>
> which characterizes the accuracy of analytical results. For validation functions $x_{estm} = x_{true}$ - and only in this case - the three-dimensional relationship (i) becomes two-dimensional and the common least squares calibration is justified.

3.2 Errors in linear calibration and evaluation by ordinary LS estimation

Fundamentally, the uncertainties of measured values y estimated by calibration, e.g. according to Eq.(18), on the one hand and of analytical results (analyte amounts, concentrations) estimated by means of calibration models, e.g. according to Eq.(23), on the other hand differ from one another. The uncertainty of y values in calibration is characterized by the *confidence interval* cnf(y) = Δy_c whereas the uncertainty of estimated x values is characterized by the *prediction interval* prd(x) = Δx_p. The *prediction interval* of the measured value y, prd(y) = Δy_p, also plays a role, namely for the definition of the critical value (limit of decision), limit of detection, and limit of quantification [1,16-19].

The precision of calibration is characterized by the following special errors

3.2.1 *Residual standard deviation*

$$s_{y.x} = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{m - 2}} = \sqrt{\frac{\sum (y_i - \hat{B}_x - \hat{A}_x x_i)^2}{m - 2}} \tag{24}$$

Note that the number of degrees of freedom is $f = m - 2$ in this case of a two-parametric model according to Eq.(8). In the case of linear calibration through the coordinate origin according to Eq.(9) $f = m - 1$.

3.2.2 *Estimated standard deviation of the estimated intercept (blank) B*

$$s_B = s_{y.x} \sqrt{1/m + \bar{x}^2 / Q_{xx}} \tag{25}$$

3.2.3 *Estimated standard deviation of the estimated slope A*

$$s_A = s_{y.x} / \sqrt{Q_{xx}} \tag{26}$$

3.2.4 *Estimated standard deviation of an estimated mean $\hat{y}_e$ at position x_i*

$$s_{ye} = s_{y.x} \sqrt{1/m + (x_i - \bar{x})^2 / Q_{xx}} \tag{27}$$

3.2.5 *Estimated standard deviation of a predicted single value $\hat{y}_p$ at position x_i*

$$s_{yp} = s_{y.x} \sqrt{1 + 1/m + (x_i - \bar{x})^2 / Q_{xx}} \tag{28}$$

3.2.6 *Estimated standard deviation of a predicted mean $\tilde{y}_p$ from n repetitions in position x_i*

$$s_{\tilde{y}p} = s_{y.x} \sqrt{1/n + 1/m + (x_i - \bar{x})^2 / Q_{xx}} \tag{29}$$

3.2.7 *Estimated standard deviation of a predicted mean $\tilde{x}_p$ from n repetitions in position x_i*

$$s_{\tilde{x}p} = s_{y.x}/A \sqrt{1/n + 1/m + (y_i - \bar{y})^2 / (A^2 Q_{xx})} \tag{30}$$
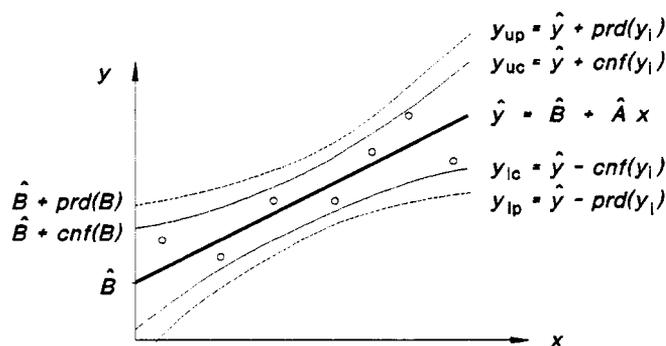
Fig. 4 Calibration straight line with relevant confidence and prediction bands; y_{up} upper prediction limit, y_{uc} upper confidence limit, y_{lc} lower confidence limit, y_{lp} lower prediction limit

The *confidence band, CB,* of the *entire calibration straight line* as shown in Fig.4 is given by

$$CB = \hat{y} \pm s_{ye} \sqrt{2 F_{\alpha;\, f1=2;\, f2=m-2}} \tag{31}$$

The following uncertainty intervals resulting from (24) to (30) are of practical interest:

3.2.8 *Confidence interval of the intercept (blank) B*

$$cnf(B) = B \pm s_B \, t_{\alpha,\, f=m-2} \tag{32}$$

3.2.9 *Prediction interval* $prd(B)$ *of a single B value* by Eq.(33) and that of an average $\bar{B}$ from *n* repetition measurements according to Eq.(34). The latter one is important with regard to the estimation of the detection limit from blanks

$$prd(B) = B \pm s_{y.x} \, t_{\alpha,\, f=m-2} \sqrt{1 + 1/m + \bar{x}^2 / Q_{xx}} \tag{33}$$

$$prd(\bar{B}) = \bar{B} \pm s_{y.x} \, t_{\alpha,\, f=m-2} \sqrt{1/n + 1/m + \bar{x}^2 / Q_{xx}} \tag{34}$$

3.2.10 *Confidence interval of an estimated mean* $\hat{\bar{y}}_{ie}$ *at position* x_i

$$cnf(\bar{y}_i) = \bar{y}_{ie} \pm s_{y.x} \, t_{\alpha,\, f=m-2} \sqrt{1/m + (x_i - \bar{x})^2 / Q_{xx}} \tag{35}$$

3.2.11 *Prediction interval of a single value* $\hat{y}_{ip}$ *at position* x_i

$$prd(y_i) = y_{ip} \pm s_{y.x} \, t_{\alpha;\, f=m-2} \sqrt{1 + 1/m + (x_i - \bar{x})^2 / Q_{xx}} \tag{36}$$

3.2.12 *Prediction interval of a mean* $\hat{\bar{y}}_{ip}$ *from n repetitions in position* x_i

$$prd(\bar{y}_i) = \bar{y}_{ip} \pm s_{y.x} \, t_{\alpha;\, f=m-2} \sqrt{1/n + 1/m + (x_i - \bar{x})^2 / Q_{xx}} \tag{37}$$

3.2.13 *Prediction interval of a mean* $\bar{x}_{ip}$ *from n repetitions for a measured value* y_i

$$prd(\bar{x}_i) = \bar{x}_{ip} \pm s_{y.x} / A \, t_{\alpha;\, f=m-2} \sqrt{1/n + 1/m + (y_i - \bar{y})^2 / (A^2 Q_{xx})} \tag{38}$$

3.3 Weighted linear least squares estimation (WLS)

In cases in that homoscedasticity according to Eq.(15) is not given, the estimated standard deviation s_y is frequently a function of the measured quantity, $s_y = f(y)$, strictly speaking $\sigma_y = f[E(y)]$. That means that the standard deviation can be a function of the expected y value. In this case the calibration system is heteroscedastic and weighted least squares fitting has to be applied [20, 21]. By means of weighting, the different variances of the certain calibration points are considered. In variance weighting, the weights are

$$w_{yi} = \frac{1/s_{yi}^2}{(\sum^p 1/s_{yi}^2)/p} \tag{39}$$

for p calibration points (index i) and the minimization criterion (16) becomes

$$\sum w_{yi}\, d_{yi}^2 \overset{!}{=} \min \tag{40}$$

The calibration coefficients are calculated analogous to Eqs. (19) and (20) by means of the weights $w_i = w_{yi}$:

$$A_{x,w} = \frac{m \sum w_i\, x_i\, y_i - \sum w_i\, x_i \sum w_i\, y_i}{m \sum w_i\, x_i^2 - (\sum w_i\, x_i)^2} \tag{41}$$

$$B_{x,w} = (\sum w_i\, y_i - A_{x,w} \sum w_i\, x_i)/m \tag{42}$$

The estimate of the residual standard deviation is

$$s_{y.x.w} = \sqrt{\sum w_i\, (y - \hat{y})^2/(m-2)} \tag{43}$$

Other quantities characterizing uncertainties can be estimated in analogy to Eqs. (25) to (38). Some software packages for regression analysis allow one to enter an estimate of the functional dependence $\sigma_y = f[E(y)]$ and to carry out a suitable weighting with this function.

The decision on weighted or unweighted least squares can be reached on the basis of a statistical test or on the basis of a theoretical model.

3.4 Linear least squares fitting for errors in both variables (orthogonal LS)

In Fig. 3, three calibration lines are given. First, the model to estimate y from (practically) error-free x values. This relationship is commonly used for calibration in form of ordinary least squares fitting:

$$\hat{y} = \hat{B}_x + \hat{A}_x\, x \tag{44}$$

Another model can be formulated to estimate x from y values under the condition that $s_y \ll A \cdot s_x$:

$$\hat{x} = \hat{B}_y + \hat{A}_y \; y \tag{45}$$

Note 4: It should be explicitly noted that Eq.(45) is not the inverse function of (44) and, therefore, $\hat{B}_y \neq -\hat{B}_x / \hat{A}_x$ and $\hat{A}_y \neq 1 / \hat{A}_x$.

Note 5: In analytical chemistry, Eq.(45) has no practical relevance, as a rule. It serves only as a means to estimate the orthogonal calibration line according to Eq.(46).

The calibration function in case where errors in both variables exist

$$\hat{y} = \hat{B} + \hat{A} \; x \tag{46}$$

has to be determined by orthogonal least squares minimizing, that means the errors in both the dependent and the independent variable are minimized simultaneously. This orthogonal least squares calibration has to be applied if both the measured values y and the analytical values x (concentrations) are error-affected quantities. The model (46) cannot be determined directly but only by approximations, e.g. the slope A can be estimated as the geometric mean (GM) of the straight lines (44) and (45) [13] by

$$\hat{A} = \tan[\; \tfrac{1}{2} \; (\tan^{-1} \hat{A}_x + \tan^{-1} \hat{A}_y)] \tag{47}$$

with $\hat{A}_x$ according to Eq.(19) and $\hat{A}_y = Q_{xy} / Q_{yy}$. The estimate of B is obtained from (46) analogously to (20). Another procedure for the estimation of A was proposed by Wald [22]

$$A = \frac{\sum\limits_{i=1}^{g} y_i - \sum\limits_{j=h}^{m} y_j}{\sum\limits_{i=1}^{g} x_i - \sum\limits_{j=h}^{m} x_j} \tag{48}$$

where m is the number of calibration measurements and $g = m/2 = h - 1$ for even m and $g = (m + 1)/2 = h$ when m is uneven. Furthermore it should be mentioned that the first principal component p_1 of a principal component analysis (PCA) [23-25] gives a good approximation of the orthogonal calibration line.

Several approximation models for orthogonal calibration procedures have been compared [28], the problem is also dealt with in other studies [29-31].

4. STATISTICAL TESTS

For the application of the proper calibration model, it is important to test whether the conditions given in section 3.1 are fulfilled. For a first information, by means of commercial software packages one can examine visually the residual errors of a given calibration model and get pre-information on the character of the residual errors. Typical plots as shown in Fig.5 also give information on tests that have to be carried out, such as randomness, normality, linearity, homoscedasticity, etc.

4.1 Linearity: Whether the chosen linear model is adequate can be seen from the residual deviation distribution over the x values. In Fig. 5a the errors scatter randomly around the zero line indicating that the model is suitable. On the other hand,

Fig. 5 Typical plots of residual deviations

in Fig. 5b it can be seen that the errors show systematic deviations and even in the given case where the deviations alternate in the real way, it is indicated that the linear model is inadequate and a nonlinear model must be chosen. The hypothesis of linearity can be tested

(a) *a priori* (no actual nonlinear model is considered) by comparison of the deviations of the means from the calibration line (the residual standard deviations $s_{y.x}$ (24)) with that of the y values from their means (s_y)

$$\hat{F} = \frac{s_{y.x}^2}{s_y^2} = \frac{\sum_{i=1}^{p} m_i\,(\overline{y_i} - \hat{y_i})^2 / (p - 2)}{\sum_{i=1}^{p} \sum_{j=h}^{m_i} (y_{ij} - \overline{y_i})^2 / (m - p)} \qquad (49)$$

(m_i number of measurements in the p calibration points; $\sum_{i=1}^{p} m_i = m$, usually $m_1 = m_2 = \ldots = m_p$, and $p \cdot m_i = m$). The test is carried out by comparison of the quotient (49) with $F_{\alpha;\,f1=p-2;\,f2=m-p}$.

(b) *a posteriori* (comparison with a certain nonlinear model) by comparison of the residual standard deviations of the linear model with that of the nonlinear model:

$$\hat{F} = \frac{s_{y.x.lin}^2}{s_{y.x.non}^2} = \frac{\sum_{i=1}^{m} (y_i - \hat{y_i})^2 / (m - 2)}{\sum_{i=1}^{m} (y_i - \hat{y_i}^2) / f_{non}} \qquad (50)$$

The number of the degrees of freedom in case of linear models is $f_{lin} = m - 2$ or $m - 1$, respectively, depending on whether two parameters are estimated according to Eq.(8) or one parameter according to Eq.(9). In the nonlinear case f_{non} results from the actual model (e.g. for a quadratic equation $y = a + b x + c x^2$, $f_{non} = m - 3$). A suitable test for (50) can also be carried out according to Mandel [32]

$$\hat{F} = \frac{s_{y.x.lin}^2 - s_{y.x.non}^2}{s_{y.x.non}^2} \tag{51}$$

by comparison with $F_{\alpha; f1=1; f2=f_{non}}$. In each case where $\hat{F} \geq F_{\alpha; f1; f2}$ the linear model cannot be applied.

4.2 <u>Homoscedasticity</u>: Unequal variances are recognizable from residual plots as in Fig. 5c where frequently d_y is a function of x in the given "trumpet"-like form. In such a case, the test of homoscedasticity can be carried out in a simple way by means of the Hartley-Test [33] (for equal m_i in the p calibration points)

$$\hat{F}_{max} = s_{max}^2 / s_{min}^2 \tag{52}$$

In cases in which the situation is more obscure as represented in Fig. 5c, the Bartlett-Test of homogeneity of variances [34] has to be applied

$$\hat{\chi}^2 = 2.303 / c \left(f \lg s^2 - \sum_{i=1}^{p} f_i \lg s_i^2 \right) \tag{53}$$

where $f = m - p = \sum f_i$ is the total number of degrees of freedom (p is again the number of calibration points in each of which m_i repeated measurements are carried out), $s^2 = \sum (f_i s_i^2 / f)$ the weighted variance, s_i^2 the variances of the i-th group (point) with the degrees of freedom f_i and c is a correction constant that should be calculated according to $c = \{\sum (1/f_i - 1/f) [3(m - 1)] + 1\}$ when the number of degrees of freedom is low.

$\hat{\chi}^2$ has to be compared with the critical value $\chi^2_{\alpha; f}$ and the null hypothesis $s_1 = s_2 = ... = s_p$ must be rejected if $\hat{\chi}^2 \geq \chi^2_{\alpha; f}$.

4.3 <u>Test of calibration parameters</u>: In some cases, it may be useful to compare experimentally found calibration parameters A and B, respectively, with theoretically expected values α and β. The comparison is carried out by means of Student's t-Test

$$\hat{t} = |A - \alpha| / s_A \tag{54}$$

$$\hat{t} = |B - \beta| / s_B \tag{55}$$

With regard to validation, especially the null hypotheses $\alpha = 1$ and $\beta = 0$ are of interest. The respective hypothesis must be rejected if $\hat{t} \geq t_{\alpha; f}$.

5. VALIDATION OF CALIBRATION

As a rule, the trueness of analytical results is guaranteed by validation experiments. Validation of a calibration procedure is based on the validation function (recovery funcion) $x_{estm} = f(x_{true})$, see Eq.(iv) section 3.1, note 3. Two practical ways are used to investigate the trueness of analytical results:

5.1 *Analysis of certified reference materials* (CRM's) with "true" contents. The validation function is estimated by normal LS regression

$$E(x_{estm}) = b + a\, x_{true} \tag{56}$$

where a and b are validation coefficients with the analytical meaning of a constant bias (b) and a proportional bias (a). The estimates of a and b can be determined by Eqs. (19) and (20). By testing the null hypotheses $a = 1$ and $b = 0$ according to Eqs. (54) and (55) the absence of biases of the analytical results can be verified.

Systematic deviations are also detected if the corresponding confidence intervals of the validation coefficients do not include 0 or 1, respectively, namely

(i) an additive bias if $b + \Delta b < 0$ or $b - \Delta b > 0$ ($b > |\Delta b|$)
(ii) a proportional bias if $a + \Delta a < 1$ or $a - \Delta a > 1$ ($a > |\Delta a|$)

5.2 *Analysis* of a given set of test samples with graduated concentrations *by two independent methods*, the one, I, for which accuracy is to be verified, in direct comparison with another method, II, which is known to be accurate. The special recovery function in this case holds

$$x_{I,estm} = b + a\, x_{II,true} \tag{57}$$

Because both quantities, $x_{I,estm}$ and $x_{II,true}$, are subject to error in this processing, orthogonal least squares fitting according to Eqs. (47), (48), principal component analysis or robust fitting must be applied. The tests on significant deviations from $a = 1$ and $b = 0$ are carried out as above.

6. ROBUST CALIBRATION

If the basic conditions for the use of least squares fitting are not fulfilled or if strongly deviating calibration points appear (*"outliers"* or, more exactly, *leverage points*), the ordinary least squares method fails, i.e., the estimated parameters are biased and, therefore, are not representative of the relation between x and y. Whereas normality of the measured values can be frequently obtained by a suitable transformation, especially in the case of outlying calibration points [35], robust calibration has to be applied. In the simplest case the robust parameter estimation can be carried out by means of median statistics. Between all the calibration points all possible slopes $A_{ij} = (y_j - y_i) / (x_j - x_i)$ for $j > i$ are calculated. After arranging the A_{ij} according to increasing values, the average slope can be estimated as the median by

$$\tilde{A} = \text{med}\,\{A_{ij}\} \tag{58}$$

and the intercept $\tilde{B}$ will then be obtained by

$$\tilde{B} \;=\; \text{med } \{y_i - \tilde{A} \; x_i \} \tag{59}$$

Estimates of the variances and uncertainty intervals in robust calibration can be taken from the literatur [35, 36].

Note 6: Robust calibration corresponds in most cases to the problem of outlying calibration points (leverage points). In consideration of that, attention must be directed to the linearity of the relationship in general and the (residual) randomness.

The relationship between the most important calibration models depending on the fulfilment of certain conditions are represented schematically in Figure 6. As outlined above, the ordinary least squares estimation can only be applied if the measured values are independent and normal-distributed, free from outliers or leverage points and characterized by homoscedastic errors. Additionally, the values in the analytical quantity (concentration) must be practically error-free.
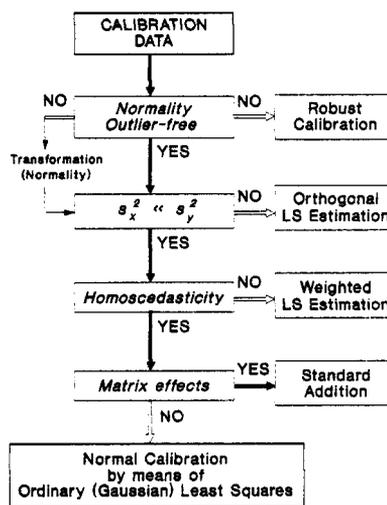
Fig. 6 Different calibration models in dependence on the fulfillment of certain statistical and chemical conditions

From the chemical point of view, in cases where matrix effects appear and no suitable certified reference materials (CRM's) are available, the calibration can be performed in the sample matrix itself by means of standard addition.

7. CALIBRATION BY STANDARD ADDITIONS

When matrix effects appear or are to be expected and matrix-matched calibration samples are not available, the standard addition method (SAM) can be the calibration method of choice. Especially in the case of environmental and biochemical systems and generally in ultra trace analysis, SAM is frequently used. By addition of standard solutions to the sample a similar behaviour of the calibration set and the sample is created provided that the analyte is added in the form of same species.

The model of standard addition is based on the prerequisite that blanks do not appear or can be eliminated, see Eq.(9). For the initial concentration x_0 of the analyte in the sample under investigation it holds

$$y_0 = A_x x_0 \qquad (60)$$

where y_0 is the measured value of the unspiked test sample. Known amounts x_i of the analyte are added to the sample; in doing so it is recommendable to use equimolar amounts of x_i in the range of $x_1 \approx x_0$, $x_2 = 2 x_1$, ..., $x_p = p x_1$ (frequently $p = 3$ or 4 is used). Therefore, some ideas about the initial concentration x_0 should exist.
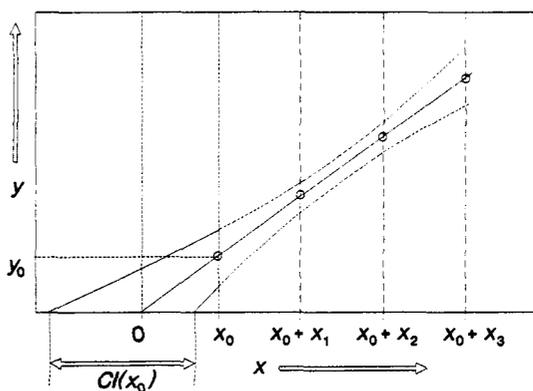
Fig. 7   Calibration by standard addition

The standard addition (SA) calibration function is estimated by least squares fitting. The slope is obtained by

$$A_x = (\overline{y}_p - \overline{y}_0) / x_p. \qquad (61)$$

This procedure is justified if the sensitivity of the determination of the species in the sample is the same as of the species added: $A_x = \Delta y / \Delta x = y_0 / x_0 = (\overline{y}_p - \overline{y}_0) / x_p$ (p is the number of samples with standard additions). The solution for x is given by the (SA) analytical function whose extrapolation for $y = 0$ yields

$$x_0 = \overline{y}_0 / A_x = x_p \cdot \overline{y}_0 / (\overline{y}_p - \overline{y}_0) \qquad (62)$$

with the confidence interval

$$\text{cnf}(x_0) = x_0 \pm s_{y.x}\, t_{\alpha;\, f=m-2} / A_x \sqrt{1/m + [(-x_0 - x_p/2)^2 / Q_{xx}]} \qquad (63)$$

which is wider as compared to normal calibration (see Eq.35 divided by A_x) because of the extrapolation to $x_i = -x_0$. The number of calibration measurements m results from $m_0 + p \cdot m_i$ or $m_0 + \Sigma^p m_i$, respectively, where m_0 is the number of measurements of y_0.

When a blank appears, it must be estimated from a sufficiently large number of blank measurements and the measured values must be corrected in this respect. To ensure the adequateness of the SA calibration model, $p \geq 2$ additions should be carried out. Only in the case when it is definitely known that the linear model holds true, then one single addition (m_i times repeated) may be carried out. In general, linearity can be tested according to Eqs.(49) to (51).

Although standard addition calibration is an unsafe method if linearity in the range $x < x_0$ is not experimentally verified but only supposed, there is scarcely an alternative when matrix effects are seriously suspected.

SUMMARY

The fundamentals of calibration in analytical chemistry are represented. For several analytical methods the calibration data may have different statistical characteristics. Depending on whether the data are normally distributed, error-free in the independent variable (the content of the calibration standards), and homoscedastic, the calibration has to be carried out by means of normal or ordinary least squares (LS, NLS, or OLS), weighted least squares (WLS), orthogonal least squares (treatment of errors of both variables), or by robust calibration as shown in the flow chart, Fig. 6.

In most practical cases, ordinary (Gaussian) least squares calibration may be applied. Errors in both variables have to be considered and, therefore, orthogonal least squares fitting has to be applied especially when the results of two methods have to be compared in the course of validation procedures.

REFERENCES

1.   Currie, L.A., IUPAC Commission on Analytical Nomenclature, "Nomenclature in Evaluation of Analytical Methods including Detection and Quantification Capablilities", *Pure & Appl. Chem.* 67 (1995) 1699

2.   Danzer, K., "Calibration - A Multidimensional Approach", *Fresenius' J. Anal. Chem.* 351 (1995) 30

3.   Moseley, H., Phil. Mag. 26 (1913) 1024

4.   Kovats, E., Helvet. Chim. Acta 41 (1958) 1915; Analyt. Chem. 36 (1964) 31A

5.   Colthup, N.B., Daly, L.H., Wiberly, S.E., *Introduction to Infrared and Raman Spectro-scopy*, Academic Press, New York, London, 1975

6.   De Gregorio, P., Savastano, G., *Spektrum des Eisens von 2206 bis 4656 Å mit Analysenlinien*, Specola Vaticana, 1972; Peter, H., Scheller H., *Atlas für Gitter-spektrographen 2250 bis 6500 Å*, Carl Zeiss, Jena, 1982

7.   Harrison, G.R., *M.I.T. Wave-Length Tables of 100.000 Spectrum Lines*, New York, 1939; Meggers, W.E., Corliss, C.H., Scribner, B.F., *Tables of Spectralline-Intensities*, Nat. Bur. Standards Monograph 32, 1961

8.   IUPAC, "Orange Book", H. Freiser, G.H. Nancollas, Eds., *Compendium of Analytical Nomenclature*, 2nd edition, Blackwell Scientific Publ., Oxford, 1987

9.   Hulanicki, A., IUPAC Commission on General Aspects of Analytical Chemistry, "Absolute Methods in Analytical Chemistry", Pure & Appl. Chem. 67 (1995) 1905

10.   Harvey, C.E., *A Method of Semi-Quantitative Spectrographic Analysis*, A.R.L., Glendale, Calif., 1947

11.   Ramendik, G.I., Elemental Analysis Without Standard Reference Samples: The General Aspect and the Realization in SSMS and LMS, *Fresenius'J. Anal. Chem.* 337 (1990) 772

12.   Sherman, J.,The Theoretical Derivation of Fluorescent X-Ray Intensities from Mixtures, *Spectrochim. Acta* 7 (1955) 283; Fei He, Van Espen, P.J., General Aspects for Quantitative Energy Dispersive X-Ray Fluorescence Analysis Based on Fundamental Parameters, *Analyt. Chem.* 63 (1991) 2237

13.   Danzer, K., "Problems of Calibration in Trace, in situ-micro and Surface Analysis", *Fresenius' J. Anal. Chem.* 337 (1990) 794

14.   Currie, L.A., Svehla, G., IUPAC Commission on Analytical Nomenclature, "Nomen-clature for the Presentation of Results in Chemical Analysis", *Pure & Appl. Chem.* 66 (1994) 595

15.   Harter, H.L., "The Method of Least Squares and Some Alternatives", Parts I-IV, *Intern. Statist. Rev.* 42 (1974) 147; 235; 43 (1975) 1; 269

16.   Currie, L.A., "Limits for Qualitative Detection and Quantitative Determination", *Analyt. Chem.* 40 (1968) 586

17.   Kaiser, H., "Zum Problem der Nachweisgrenze", *Fresenius'Z. Anal. Chem.* 209 (1965) 10; 216 (1966) 80

18. Currie, L.A., Ed., *Detection in Analytical Chemistry: Importance, Theory, and Practice*, ch.1, ACS Sympos. Ser. 361, Washington 1988
19. Currie, L.A., Horwitz, W., "IUPAC Recommendations for Defining and Measuring Detection and Quantification Limits", *Analusis* 22 (1994) 24
20. Garden, J.S., Mitchell, D.G., Mills, W.N., "Nonconstant Variance Regression Techniques for Calibration-Curve-Based Analysis", *Analyt. Chem.* 52 (1980) 2310
21. Draper, N.R., Smith, H., *Applied Regression Analysis*, Wiley, New York, 2nd ed., 1981
22. Wald, A., "The Fitting of Straight Lines if Both Variables are Subject to Error", *Ann. Math. Statist.* 11 (1940) 284
23. Malinowski, E.R., Howery, D.G., *Factor Analysis in Chemistry*, Wiley, New York, 1980
24. Flury, B., *Common Principal Components and Related Multivariate Models*, Wiley, New York, 1988
25. Martens, H., Næs, T., *Multivariate Calibration*, Wiley, New York, 1989
26. Natrella, M.G., *Experimental Statistics*, National Bureau of Standards Handbook 91, Washington, 1963
27. MacTaggart, D.L., Farwell, S.O., "Analytical Use of Linear Regression", Part I: "Regression Procedures for Calibration and Quantitation"; Part II: "Statistical Error in Both Variables", *J. Assoc. Official Anal. Chemists* 75 (1992) 594-608; 608-614
28. Danzer, K., Wagner, M., Fischbacher, C., "Calibration by Orthogonal and Common Least Squares - Theoretical and Practical Aspects", *Fresenius' J. Anal. Chem.* 352 (1995) 407
29. Fuller, W.A., *Measurement Error Models*, Wiley, New York, 1987
30. Meloun, M., Militký, J., Forina, M., *Chemometrics for Analytical Chemistry*, Vol. 2: *PC-Aided Regression and Related Methods*, Ellis Horwood, New York, 1994
31. Mandel, J., "Fitting Straight Lines When Both Variables are Subject to Error", *J. Quality Technol.* 16 (1984) 1
32. Mandel, J., *The Statistical Analysis of Experimental Data*, Wiley, New York, 1964
33. Hartley, H.O., "The Maximum F-ratio as a Short Cut Test for Heterogeneity of Variance", *Biometrika* 37 (1950) 308
34. Bartlett, M.S., "Properties of Sufficiency and Statistical Tests", *Proc. Roy. Soc. A* 160 (1937) 268
35. Rousseeuw, P.J., Leroy, A.M., *Robust Regression and Outlier Detection*, Wiley, New York, 1987
36. Huber, P.J., *Robust Statistics*, Wiley, New York, 1981
37. Sharaf M.A., Illman, D.L., Kowalski B.R., *Chemometrics*, Wiley, New York, 1986

INDEX OF TERMS