

The effect of influential data, model and method on the precision of univariate calibration

Milan Meloun ^{a,*}, Jiří Militký ^b, Karel Kupka ^c, Richard G. Brereton ^d

^a Department of Analytical Chemistry, Faculty of Chemical Technology, Pardubice University, CZ-532 10 Pardubice, Czech Republic

^b Department of Textile Materials, Technical University, CZ-461 17 Liberec, Czech Republic

^c Trilobyte Statistical Software Ltd, CZ-530 02 Pardubice, Czech Republic

^d School of Chemistry, University of Bristol, Cantock's Close, Bristol BS8 1TS, UK

Received 29 November 2001; received in revised form 7 February 2002; accepted 28 February 2002

Abstract

Building a calibration model with detection and quantification capabilities is identical to the task of building a regression model. Although commonly used by analysts, an application of the calibration model requires at first careful attention to the three components of the regression triplet (data, model, method), examining (a) the data quality of the proposed model; (b) the model quality; (c) the LS method to be used or a fulfillment of all least-squares assumptions. This paper summarizes these components, describes the effects of deviations from assumptions and considers the correction of such deviations: identifying influential points is the first step in least-squares model building, the calibration task depends on the regression model used, and finally the least squares LS method is based on assumptions of normality of errors, homoscedasticity, independence of errors, overly influential data points and independent variables being subject to error. When some assumptions are violated, the ordinary LS is inconvenient and robust *M*-estimates with the iterative method of reweighted least-squares must be used. The effects of influential points, heteroscedasticity and non-normality on the calibration precision limits are also elucidated. This paper also considers the proper construction of the statistical uncertainty expressed as confidence limits predicting an unknown concentration (or amount) value, and its dependence on the regression triplet. The authors' objectives were to provide a thorough treatment that includes pertinent references, consistent nomenclature, and related mathematical formulae to show by theory and illustrative examples those approaches best suited to typical problems in analytical chemistry. Two new algorithms, calibration and linear regression written in S-PLUS and enabling regression triplet analysis, the estimation of calibration precision limits, critical levels, detection limits and quantification limits with the statistical uncertainty of unknown concentrations, form the goal of this paper. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Calibration precision; Outliers; Influential points; Regression triplet; Iterative reweighted least-squares

* Corresponding author. Tel.: +420-40-603-7026; fax: +420-40-603-7068.

E-mail addresses: milan.meloun@upce.cz (M. Meloun), jiri.militky@vslib.cz (J. Militký), kupka@trilobyte.cz (K. Kupka), r.g.brereton@bristol.ac.uk (R.G. Brereton).

¹ Algorithm in S-PLUS www.trilobyte.cz/calibration

1. Introduction

Linear regression is perhaps the most used and abused statistical method in calibration. Even though there is a huge literature on univariate calibration application [1–24], textbooks and reviews [25–45], a common mistake is to blindly force a ‘classical’ regression fit onto any set of calibration data with a presumed linear relationship. Most analytical procedures involve linear regression either in the calibration straight line or in the calibration curve mode. Little, if any, attention is paid to the selection of suitable calibration points or to the examination of influential points, outliers and leverages, and heteroscedasticity in the regression analysis [19,24,30,34,38,44]. Application of the ordinary least squares (OLS) in calibration is based on the assumptions of normality, homoscedasticity and independence of the measurements [46–50,53]. A source of problems may be found in the components of the regression triplet (data, model and method of estimation). OLS provides statistically accurate estimates only when all assumptions about calibration data and about a calibration model are fulfilled. When some assumptions are not fulfilled, OLS is inconvenient. Regression diagnostics represent procedures for the identification of (a) the calibration data quality of a proposed calibration model, (b) the calibration model for a given set of data, and (c) fulfillment of all least-squares assumptions. In reality, the least-squares assumptions hold only approximately, and one can expect the calibration data to contain either errors or observations that are somewhat unusual compared with the rest of the data.

OLS gives equal weight to every point of calibration data. However, every point does not have equal impact on the various least-squares results. For example, the slope in a simple calibration straight line is influenced most by the points having values of the independent variable farthest from the mean. A single point far removed from the other data points can have almost as much influence on the calibration results as all the other points combined. Such points are called high-leverage points. In designed experiments these points are not present. The term outlier refers to a

calibration point, which is in some sense inconsistent with the rest of the points in the calibration data set.

Heterogenous variances in calibration points imply that some observations are more precise than others. Rational use of the data would require that more weight be given to those that contain the most information. Two approaches to handling heterogeneous variance are the transformation of the dependent variable (signal) and use of weighted least-squares. Weighted least-squares uses the original metric of the dependent variable, but gives each observation weight according to the relative precision.

The effect of influential points, heteroscedasticity and non-normality on calibration precision limits are discussed. The authors’ objectives were to survey the relevant literatures on linear regression, to provide a thorough treatment that includes pertinent references, consistent nomenclature, and related mathematical formulae, and to show by theory and practical examples those approaches best suited to typical problems in analytical chemistry. The new algorithms Calibration and Linear Regression in S-PLUS, which enable regression triplet analysis and estimation of alternative approaches to calibration precision limits, are used.

2. Theoretical

2.1. Consistent nomenclature

2.1.1. Calibration models

Constructing calibration model $g(x)$, a relationship between the measurable quantity y , called a signal (potential, electric current etc.) or a output and the quantity x called a measurand or input, a property (content, composition, concentration, temperature etc.) being more difficult to monitor, is created. In the first stage of the calibration experiment for n samples with known (or precisely adjusted) values x_i the corresponding signal values y_i are measured. As a model of experimental measurements in the chemical measurement process CMP, the additive model of measurement errors is usually assumed $y_i = g(x_i, \beta) + \varepsilon_i$. Here β

is the set of adjustable parameters and ε_i is error term containing errors of measurement, error due to blank, etc. It is often assumed that ε_i are normally distributed and have constant variance σ^2 . In calibration model building, the suitable model $g(x)$ is selected and the parameters β are estimated from n pairs of data (x_i, y_i) by regression analysis. In many experiments, there are some restrictions on the measured variable y_i , $i = 1, \dots, n$. For example, y_i may take only positive values, with non-constant variance, $\sigma^2(y_i)$, but with constant relative error, $\sigma^2(y)/y$. All such conditions are valid in the multiplicative model of measurement errors. The combined model of measurement errors

$$y_i = g(x_i, \beta) \exp(v_i) + \varepsilon_i$$

is also used, where the errors v_i and ε_i are assumed to be independent.

For linear regression type models, according to model parameters, this task leads to the solution of the equation system. For nonlinear regression type models minimization algorithms must be used. Building a calibration model is identical with the task of building a regression model.

The second step of calibration involves inversion of the model $g(\cdot)$, i.e. for a measured response y^* for unknown concentration the corresponding value x^* and its statistical characteristics are calculated.

Calibration tasks have been classified according to criteria by Rosenblatt and Spiegelman [51] into (a) absolute calibration and (b) comparative calibration.

With reference to the application of the calibration model, the following cases may be distinguished; (a) single application of the calibration model, the calibration model is constructed from n measured points $\{x_i, y_i\}$, $i = 1, \dots, n$, and then one estimate \hat{x}^* with its confidence interval is calculated from one y^* value; (b) multiple application of the calibration model, from the calibration model, several estimates \hat{x}^* are determined from values of the analytical signal; (c) single or multiple application in combination with other measurements, the results of the second phase of calibration are used together with other variables and constants for determination of a quantity

which is a function of more variables. Here, any bias in the estimates \hat{x}^* will be included in the final systematic error of the result.

The difficulty of the calibration task depends on the model used. For nonlinear regression models the solution is sought in the form $\hat{x}^* = g^{-1}(y^*)$. On the basis of the Taylor series for this function, the approximate formula for the variance $D(\hat{x}^*)$ may be found in the form [52]

$$D(\hat{x}^*) \approx \left[\frac{\delta g(x, b)}{\delta x} \right]^{-2} \left[\frac{D(y^*)}{M} + D(g(x, b)) \right] \quad (1)$$

where $D(y^*)$ is the variance of y^* values, usually equal to σ^2 and $D(g(x, b)) = D(\hat{y})$ is the variance of prediction, estimated from the Taylor series of function $g(x, b)$. For the linear regression model the variance of prediction is given by

$$\begin{aligned} D(\hat{y}) &= \sigma^2 \left[\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right] \\ &= \sigma^2 \left[\frac{1}{n} + \frac{(y^* - \bar{y})^2}{b_1^2 \sum_{i=1}^n (x_i - \bar{x})^2} \right] \end{aligned} \quad (2)$$

where b_1 is the estimate of the slope of the regression line. On substituting into Eq. (1) the relationship results:

$$D(\hat{x}^*) \approx \frac{\sigma^2}{b_1^2} \left[\frac{1}{M} + \frac{1}{n} + \frac{(y^* - \bar{y})^2}{b_1^2 \sum_{i=1}^n (x_i - \bar{x})^2} \right] \quad (3)$$

Difficulties are caused by the generally non-symmetric distribution of quantity x^* . Only in the case of a calibration straight line and small residual variance can the distribution of x^* be assumed to be symmetric and approximately normal [10].

2.1.2. Calibration straight line

A straight line is the usual calibration model in a chemical laboratory. In some cases, however, the straight-line model is valid only in a limited interval, and above a limiting point $\{x_A, y_A\}$ there is a significant departure from linearity. For ex-

ample, the Lambert–Beer law is valid only up to some limiting concentrations, above which curvature occurs. The model being in the form

$$y_i = \beta_0 + \beta_1 x + \varepsilon_i, \quad i = 1, \dots, n,$$

the signal for the unknown concentration is

$$y_j^* = \beta_0 + \beta_1 \kappa + \varepsilon_j^*, \quad j = 1, \dots, M.$$

The task of calibration is to find an estimate \hat{x}^* of parameter κ , the parameter of primary interest, and of parameters β_1 and β_2 , the supplementary parameters. The estimation assumes normality of the errors ε_i and ε_j^* . The estimate \hat{x}^* and its confidence interval may be calculated by several procedures, the straight estimate of parameter κ is obtained in the form $\hat{x}^* = \bar{x} + (y^* - \bar{y})/b_1$, where y^* is the measured signal (or the average \bar{y}^* for $M > 1$ repeated measurements, respectively) and b_1 is the estimate of the slope β_1 . This estimate is generally biased and a correction can be made with Naszodi's estimates [54]

$$\hat{x}_B^* = \bar{x} + \frac{(y^* - \bar{y})b_1}{b_1^2 + \sigma^2 / \sum_{i=1}^n (x_i - \bar{x})^2} \quad (4)$$

In the construction of confidence intervals of the estimates \hat{x}^* and \hat{x}_B^* for more scattered data, the simplest is the determination of $D(\hat{x}^*)$ with an assumption of normality. The limits of the 95% confidence interval are calculated by $L_L = \hat{x}^* - 1.96\sqrt{D(\hat{x}^*)}$ and $L_U = \hat{x}^* + 1.96\sqrt{D(\hat{x}^*)}$. When the slope estimate is sufficiently precise for the approximate confidence interval for unknown parameter κ to be used in the form

$$L_{L,U} = x^* \mp t_{1-\alpha/2}(n-2) \frac{\hat{\sigma}}{|b_1|} \sqrt{\frac{1}{n} + \frac{(y^* - \bar{y})^2}{b_1^2 \sum_{i=1}^n (x_i - \bar{x})^2}} \quad (5)$$

where the variance of measurement, σ^2 , is known. When there are replicate values of y , and mean \bar{y}^* has been determined, the corresponding confidence straight lines U_L and U_U should be calculated. The intersection of straight line U_U with the lower confidence parabola P_L of the calibration straight line leads to point L_U and the intersection of straight line U_L with the upper confidence parabola P_U leads to point L_L , Fig. 1.

If the variance of measurement, σ^2 , is known it is easy to define the $100(1-\alpha)\%$ confidence interval of signal y^* in the form $U_{L,U} = \bar{y}^* \mp u_{1-\alpha/2} \sigma$, where $u_{1-\alpha/2}$ is the quantile of the normalized normal distribution. If σ^2 is unknown, the inequality $\sigma^2 < ((n-2)\hat{\sigma}^2)/\chi_{\alpha/2}^2(n-2)M$ may be used, where $\chi_{\alpha/2}^2$ is the lower $100\alpha/2\%$ quantile of the χ^2 distribution. The confidence interval of signal $U_{L,U}$ is then calculated from

$$U_{L,U} = \bar{y}^* \mp u_{1-\alpha/2} \frac{\hat{\sigma}}{\sqrt{M}} \sqrt{\frac{n-2}{\chi_{\alpha/2}^2(n-2)}} \quad (6)$$

Instead of the quantile $u_{1-\alpha/2}$ in this equation, for $M=1$ the more convenient quantile of the Student distribution $t_{1-\alpha/2}(n-2)$ is used, and the variance σ^2 is replaced by its estimate s^2 .

For the limiting $100(1-\alpha)\%$ confidence parabola of linear calibration model the Working–Hotelling relationship (cf. page 22 in [50])

$$P_{L,U} = b_1 x + b_0$$

$$\mp \hat{\sigma} \sqrt{2F_{1-\alpha}(2, n-2) \left[\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right]} \quad (7)$$

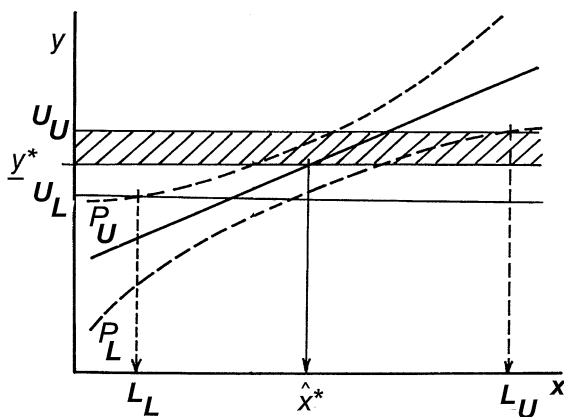


Fig. 1. Determination of the confidence interval of the concentration x for a calibration straight line: the absolute calibration and a procedure for determination x^* for the mean value of analytical signal response being referred with the horizontal line and the confidence intervals of the signal of concentration, and $f(x, b)$ is the predicted calibration straight line with its confidence intervals for a predicted observation. The horizontal x axis refers to the concentration and the vertical y axis to the signal response.

is valid. The limiting value L_U represents the solution of the equation $U_U = P_U$ with respect to variable x . The limiting value L_L is the solution of the equation $U_L = P_U$. Both equations are quadratic with respect to variable x . From Fig. 1 it can be seen that in some cases the intersection of a straight line with the parabola does not exist, but in other cases the straight line of the signal may intersect too large when the slope of the calibration straight line is small, and such a calibration straight line is a poor model. The quality of the confidence interval around the parameter x can be improved by:

1. repeating the signal measurement y^* , i.e. increasing the number of measurements M . For a sufficient number of replicates, M , the estimate $U_{L,U}$ can be calculated from the relationship $U_{L,U} = \bar{y}^* \pm u_{1-\alpha/2} \hat{\sigma} / \sqrt{M}$ with σ^2 replaced by the variance $s_{y^*}^2$ and the quantile $u_{1-\alpha/2}$ replaced by the quantile of the Student distribution $t_{1-\alpha/2}$.
2. Narrowing the confidence parabola by elimination of influential points, as will be shown in this paper: the examination of calibration data quality involves the detection of influential points, outliers and high-leverages, which cause many problems in regression analysis by shifting the parameter estimates or increasing the variance of parameters. In polynomial calibration models the confidence bands may be narrowed by the use of biased estimates calculated by the method of principal component regression [56].
3. Decreasing the residual variance s^2 and so increasing the precision of measurement, or by the use of a correct calibration model for a given data set. The quality of a proposed model can be considered in case of one controllable variable x directly from the scatter plot y versus x .

2.1.3. Application of the spline function for a calibration curve

For some physical dependencies which are not of an associative nature, polynomials are quite unsuitable and piecewise regression models (polynomials) are more convenient. Unless experimental data (x_i, y_i) , $i = 1, \dots, n$, is available another

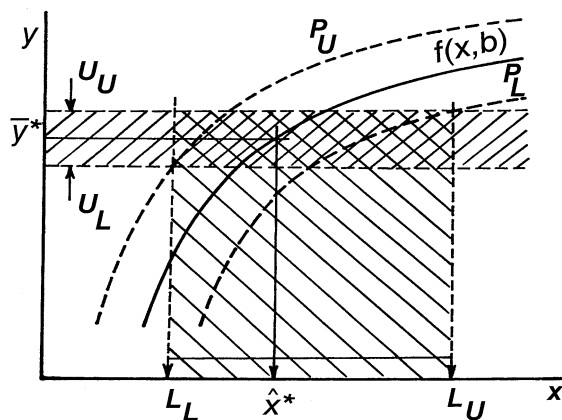


Fig. 2. Determination of the confidence interval of the concentration x for a nonlinear calibration curve: the absolute calibration and a procedure for determination x^* for the mean value of analytical signal response being referred with the horizontal line and the confidence intervals of the signal indicated by the hatched area. L_L and L_U are the lower and upper limits of the confidence interval of concentration, and $f(x, b)$ is the predicted calibration curve with its confidence intervals for a predicted signal response. The horizontal x axis refers to the concentration and the vertical y axis to the signal response.

set of knots are determined t_j , $j = 1, \dots, k$. Knots form the boundaries of intervals in which individual piecewise functions are defined. In each interval I_j bounded by knots $[t_{j-1}, t_j]$ the calibration function is expressed by the model $g_j(x)$. The quality of approximation here is dependent on the number and location of individual knots t_j , a form of the function $g_j(x)$ and on the class C^m from which the calibration model $g(x)$ comes. Note that function $g(x)$ of class C^m is continuous up to the first m derivatives. A special type of piecewise polynomial functions are called splines, Fig. 2.

Splines $S_{m+1}(x)$ are functions of class C^m which are defined as local polynomials of maximum degree $(m+1)$. Quadratic splines $S_1(x)$ which are continuous and smooth, (continuous in the first derivative), are suitable for calibration purposes. Splines $S_1(x)$ can be simply defined as truncated polynomials

$$S_1(x) = \beta_1 + \beta_2 x + \beta_3 x^2 + \sum \beta_{j+3} (x - t_j)_+^2 \quad (8)$$

here with $(x)_+^2 = x^2$ for $x > 0$ and $(x)_+^2 = 0$ for $x \leq 0$. For known t_j the $S_1(x)$ is the linear regres-

sion model. Flexibility of regression splines may be achieved by selection of knots t_i . In the proposed s-PLUS program Calibration choice from among three alternatives is possible, (a) a constant placement of knots, (b) location of knots so that in each interval I_j the same number of data points is located, (c) location of knots by the user.

2.1.4. The precision limits of calibration

One of the important goals of analytical chemistry is to report the detection and quantification capabilities of the smallest concentration or, sometimes, amount of the analyte that may be achieved with reasonable certainty when using a given procedure. To express the precision of a calibration, limiting values of the concentration for which the measurement signal is still statistically significantly different from the noise are defined:

1. The critical value, L_C , of the appropriate chemical variable (estimated net signal y , concentration x , or amount x) is the alternate to the critical level, y_C , which represents the upper limit of the $100(1 - \alpha)\%$ confidence interval of the predicted signal from the calibration model for a concentration equal to zero, i.e. the blank measurement. Currie [55] defined the critical level as ‘a decision limit at which one may decide whether or not the result of an analysis indicates detection’, Fig. 3. This level is concerned with the signal or measured concentration that corresponds to unreliable detection. For $x = 0$, an expression for the critical level y_C may be obtained

$$y_C = \bar{y} - b_1 \bar{x} + \hat{\sigma} t_{1-\alpha}(n-2) \sqrt{1 + \frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} \quad (9)$$

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 the calibration model from

$$x_C = \frac{y_C - \bar{y}}{b_1} + \bar{x} \quad (10)$$

For decisions regarding the investigation of actual states only the critical value of the net state variable or of the response variable is to be applied. These values derived from a calibration of the measurement process are decision limits to be used to assess the unknown states of systems included in this series. Looking at consecutive calibrations of the same measurement process, the critical values may vary from one calibration to another. However, since each of the critical values is a decision limit belonging to a particular measurement series, it is meaningless to calculate overall critical value across calibrations, and logically inappropriate to use these as critical values.

2. The measure of the inherent detection capability of a CMP is the minimum detectable (true) value, L_D , of the appropriate chemical variable [45,55]. The alternate is the detection limit y_D which 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 . The signal value y_D and its corresponding concentration x_D are illustrated on Fig. 3. The detection limit is the true concentration at which a given analytical procedure may be relied upon to lead to detection [45]. For the linear model the relationship

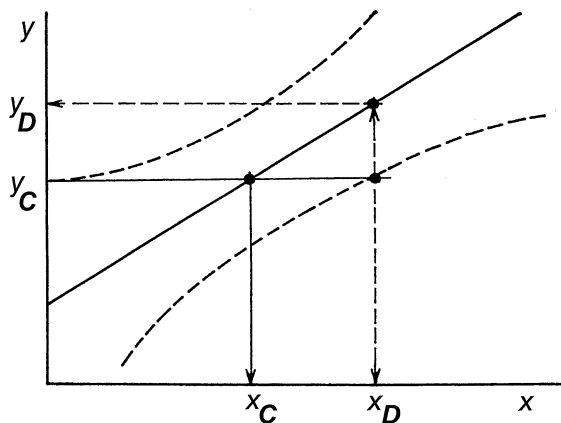


Fig. 3. Calibration design of the critical level in the response y_C and the concentration x_C units and the detection limit y_D and corresponding concentration x_D units. It includes the calibration straight line and Working-Hotelling confidence bands, cf. page 22 in [50].

$$y_D = y_C + \hat{\sigma} t_{1-\alpha}(n-2) \sqrt{1 + \frac{1}{n} + \frac{(x_D - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} \quad (11)$$

may be used. Oppenheimer [16] proposed instead of $(x_D - \bar{x})^2$ in Eq. (11) the approximation \bar{x}^2 , which is often used only. The corresponding concentration x_D is calculated from

$$x_D = \frac{y_D - \bar{y}}{b_1} + \bar{x} \quad (12)$$

The detection limit L_D gives the lowest true signal level y_D which still permits detection. The corresponding quantity x_D gives the minimum concentration, which can be distinguished from zero with probability $(1 - \alpha)$. The minimum detectable value x_D derived from a particular calibration shows whether the capability of detection of the actual measurement process is sufficient for the intended purpose. The x_D derived from a set of calibrations may serve for the comparison, the choice or the judgement of different laboratories or methods, respectively.

- As the measure of the inherent quantification capability of a CMP, the minimum quantifiable (true) value, L_Q , of the appropriate chemical variable is used. The alternate is the quantification limit y_Q , also known as the determination limit, which is the smallest signal level for which the relative standard deviation of prediction from the calibration model is sufficiently small and equal to the number C , where $C = 0.1$, usually. The quantification limit is used to decide whether or not the concentration of an analyte can be reliably determined. If the predicted value at point x_Q is given by $y(x_Q) = \bar{y} + b_1(x_Q - \bar{x})$ and the condition of quantification y_Q is then equal to $\sqrt{D(y(x_Q))}/\bar{y}(x_Q) = C$. Substitution and rearrangement leads to the expression

$$y_Q = \frac{\hat{\sigma}}{C} \sqrt{1 + \frac{1}{n} + \frac{(x_Q - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} \quad (13)$$

In practice, in the chemical laboratory, instead of $(x_Q - \bar{x})^2$ in Eq. (13) the approximation is often used only, and the corresponding concentration x_Q is given by

$$x_Q = \frac{y_Q - \bar{y}}{b_1} + \bar{x} \quad (14)$$

For nonlinear calibration models, Schwartz [17] recommends that the upper L_U and lower L_L limits of the confidence interval of concentration which correspond to different signal levels y^* be determined. Instead of the relative standard deviation of prediction from the calibration model, Schwartz uses the effective relative standard deviation

$$C(x^*) = \frac{L_U - L_L}{2x^* t_{1-\alpha/2}(n-2)} \quad (15)$$

- Alternative approaches to calibration limits: the modified quantification limit y_Q is the value of x for which $C(x) = C$. This y_Q limit is found graphically by plotting $C(x)$ against x and substituting in the calibration model. Eq. (15) may be used for linear models as well as nonlinear ones.

All four definitions may be simply used to calculate the detection limit y_D and the quantification limit y_Q for nonlinear calibration models, and also for data for which the variance of measurement is not constant (heteroscedasticity), [16]. Generally, it is valid that $y_C \leq y_D \leq y_Q$.

Note: for presentation of the defining relations, the authors use L as the generic symbol for the quantity of interest. This is replaced by y when treating net analyte signals, and x when treating analyte concentrations or amounts. Thus, L_C , L_D , and L_Q may represent y_C , y_D and y_Q , or x_C , x_D and x_Q , as appropriate.

Ebel and Kamm [18] have described an alternative procedure of determination of the detection limit y_D , and this is illustrated in Fig. 4. Even in this case for linear calibration models it is easy to determine x_P by the use of L_L . Substitution of x_P into expression for L_U leads to $L_U = y_D$.

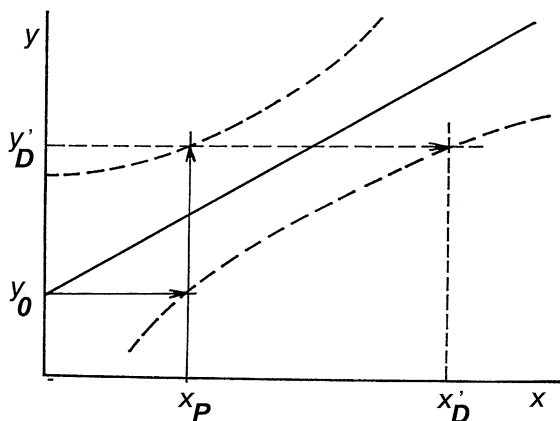


Fig. 4. The graphical procedure for determination of the detection limit y_D and x_D according to Ebel and Kamm [18].

2.2. Regression triplet

A source of problems in an OLS application may be found in components of the regression triplet, i.e. the data quality for a proposed model, the model quality for a given data set, and the regression method quality when all assumptions used in OLS are not fulfilled; regression diagnostics are used because there is no necessity for an alternative hypothesis but all types of deviations from an ideal regression triplet are discovered [50,57]. There are some basic assumptions necessary for OLS to be valid.

1. Restricted parameters: the regression parameters β are not bounded. In chemometric practice, however, there are some restrictions on the parameters, based on their physical meaning.
2. Linearity: the regression model is linear in the parameters, and an additive model of the measurement errors is valid, $y = X\beta + \varepsilon$.
3. Multicollinearity: the matrix of non-random controllable values of the regressors X has a column rank equal to m . This means that the all pairs x_j, x_k are not collinear vectors. This is the same as saying that the matrix $X^T X$ is a symmetric regular invertible matrix with a non-zero determinant.
4. Random errors: the mean value of the ran-

dom errors ε_i is zero; $E(\varepsilon_i) = 0$. This is automatically valid for all regression type models containing intercept. For models without intercept the zero mean of errors has to be tested.

5. Homoscedasticity: the random errors ε_i have constant and finite variance, $E(\varepsilon_i^2) = \sigma^2$. The conditional variance σ^2 is also constant and, therefore, the data are said to be homoscedastic.
6. Uncorrelated errors: the random errors ε_i are uncorrelated, i.e. $\text{cov}(\varepsilon_i, \varepsilon_j) = E(\varepsilon_i, \varepsilon_j) = 0$. When the errors follow the normal distribution they are also independent. This corresponds to independence of the measured quantities y .
7. Normally distributed errors: the random errors ε_i have a normal distribution $N(0, \sigma^2)$. The vector y then has a multivariate normal distribution with mean $X\beta$ and covariance matrix $\sigma^2 E$.

When the first six conditions are met, the parameter estimates b found by minimization of a least-squares are the best linear unbiased estimate (BLUE) of the regression parameters β .

2.2.1. Examination of the data quality

Examination of data quality involves detection of the influential point IP (outliers and high-leverages) [58], which cause many problems in regression analysis by shifting the parameter estimates or increasing the variance of the parameters. A calibration point may be an outlier or a potentially influential point because of errors in the conduct of the study (instrument malfunction, recording, or data entry errors) or because the data point is from a different population. On the other hand, a valid data point may appear to be an outlier, have an outlier in the residual, because the model being used is not adequately representing the calibration. Outliers are detected by analysis of the various types of residuals, hat matrix elements and related statistics. Diagnostic plots for detecting influential points are sometimes also able to detect non-normality and heteroscedasticity; a survey is provided in [57]. The most reliable tools seem to be

the graph of predicted residuals [58], the Williams graph [58] and the $L-R$ graph [59].

2.2.2. Examination of model quality

Examination of model quality can be considered directly from the scatter plot of y versus x . Individual parameters are tested for significance using the Student's t -test. The Fisher–Snedecor F -test of significance of the regression model proposed is based on the testing criterion $F_R = \hat{R}^2(n-m)/[(1-\hat{R}^2)(m-1)]$ which has the Fisher–Snedecor distribution with $(m-1)$ and $(n-m)$ degrees of freedom, means an estimate of the determination coefficient, n is a number of data points and m is the number of parameters, for a straight line $m=2$. With the use of F_R the null hypothesis $H_0: R^2=0$ may be tested and concerns a test of significance of all regression parameters β . Other statistical characteristics calculated are the determination coefficient R^2 , the mean quadratic error of prediction MEP, the Akaike information criterion AIC, and the predicted coefficient of determination \hat{R}_p^2 while the standard deviation of ordinary residual $s(e)$ examining the linearity of the proposed model can be used as a resolution criterion among various models; a definition may be found in [50,57]. To check a proposed regression model with reference to the data, a simple test based on the residuals is the sign test. Incorrectness of a proposed model causes non-randomness of residuals, and this non-randomness may be tested by a sign test. The number of sequences n_U of the same sign of residuals is estimated, for example, for residuals $-1, -1, 1, -1, 1, 2, 1$ the number of sequences is equal to 4, $n_U=4$. Then, the number of residuals with a positive sign (n_+) and a negative sign (n_-) is determined. For medium sample sizes the theoretical number of sequences n_t and its variance D_t are defined by

$$n_t = 1 + \frac{2n_+n_-}{n_+ + n_-} \approx 1 + \frac{n}{2} \quad (16)$$

$$D_t = \frac{2n_+n_-(2n_+n_- - n_+ + n_-)}{(n_+ + n_-)^2(n_+ + n_- - 1)} \approx \frac{n}{4} \quad (17)$$

When $n_U = n_t - 2\sqrt{D_t}$, there is a trend in the residuals and the model is incorrect.

2.2.3. Examination of the regression method used

Several tests for the fulfilment of three important assumptions for the least-squares method are performed, namely, homoscedasticity, absence of autocorrelation, and normality of random errors.

The Cook–Weisberg test of the homoscedasticity of residuals; identification of heteroscedasticity in data is based on the idea that the variance of a measured quantity at the i th point is an exponential function of the variable $x_i\beta$ of the type $\sigma_i^2 = \sigma \exp(\lambda x_i \beta)$ where x_i is the i th concentration. The test for homoscedasticity is carried out by checking the null hypothesis $H_0: \lambda = 0$. Cook and Weisberg [60] introduced the test criterion

$$S_f = \frac{\left[\sum_{i=1}^n (\hat{y}_i - \bar{y}) \hat{e}_i^2 \right]^2}{2\hat{\sigma}^4 \sum_{i=1}^n (\hat{y}_i - \bar{y})^2} \quad (18)$$

where $\bar{y} = (\sum_{i=1}^n \hat{y}_i)/n$. When the null hypothesis is valid the test statistics has approximately the $\chi^2(1)$ distribution with one degree of freedom.

The test of the normality of errors: at first, the normality of errors may be simply examined by a $Q-Q$ plot containing the order statistics of classical residuals $\hat{e}_{(i)}$ in dependence on the quantile of the normalized normal distribution u_{P_i} for $P_i = i/(n+1)$. Since small samples exhibit a supernormality effect, independent recursive residuals are used instead of classical, because this effect then does not exist [50]. At second, the most convenient test for linear models seems to be the Jarque–Berra test [61] which is based on the criterion

$$L(\hat{e}) = n \left[\frac{\hat{g}_1}{6} + \frac{(\hat{g}_2 - 3)^2}{24} \right] \quad (19)$$

where \hat{g}_1 is the residual skewness, \hat{g}_2 is the residual kurtosis. When $L(\hat{e}) > \chi_{0.95}^2 = 5.99$, the null hypothesis H_0 about the error normality is rejected. In this test, the supernormality effect of small samples may again disturb statistical testing.

Violation of some assumptions for OLS method: according to the test for fulfilment of the assumptions for the least-squares method LS, and the result of regression diagnostics, a more accurate regression model is constructed as follows:

- (a) When heteroscedasticity is found in the data, the weighted least-squares method WLS is used, cf. page 102–104 in [50].
- (b) When autocorrelation is found in the data, the generalized least-squares method GLS is used, cf. page 110 in [50].
- (c) When some restrictions apply to the parameters, the conditioned least-squares method CondLS is used, cf. page 98 in [50].
- (d) When multicollinearity is found in the data or when polynomial calibration models are used, the principal component method PCR is used, cf. page 116 in [50,56].
- (e) When all variables are subject to random errors, the extended least-squares method ELS is used, cf. page 121 in [50].
- (f) When the data have an error distribution other than normal, heteroscedasticity or the data contain outliers or high-leverage points, some robust methods or the iterative method of reweighted least-squares IRWLS are used, cf. page 126 in [50]: when the distribution of the errors in the dependent variable y is not normal (violation of assumption 7 for OLS method, cf. [57]) the parameter estimates obtained by OLS are not the best possible estimates. In such a case, instead of the least squares criterion some other robust criterion can be used, that is not so sensitive to violation of the assumption about the error distribution, and also not sensitive to influential points. The most convenient criteria seem to be the group of M -estimates: all M -estimates are related to the minimization criterion

$$U_M = \sum_{i=1}^n \rho \left[\frac{e_i}{\sigma} \right] = \sum_{i=1}^n \rho \left[\frac{y_i - \mathbf{x}_i^T \mathbf{b}_M}{\sigma} \right] \quad (20)$$

where x_i is the i th value of concentration, σ is the parameter of spread and $\rho(\cdot)$ is a convenient function determined from the probability density $p(\varepsilon)$. By analytical minimization of U_M , a set of normal equations is obtained:

$$\sum_{i=1}^n \psi \left[\frac{e_i}{\sigma} \right] x_{ij} = 0, \quad j = 1, \dots, m \quad (21)$$

where the function $\psi(x) = \delta \rho(x) / \delta x$ represents

the derivative of function $\rho(x)$ with respect to x . Then, if $r_i = e_i / \sigma$, this equation may be expressed in a form which corresponds to the weighted least-squares method WLS

$$\sum_{i=1}^n w_i(r) y_i x_{ij} = \sum_{i=1}^n \sum_{k=1}^m w_i(r) x_{ij} x_{ik} b_k, \quad j = 1, \dots, m \quad (22)$$

where $w_i(r) = \psi(r_i) / r_i$. The parameters estimated by the iterative method of reweighted least-squares IRWLS by using the following procedure:

1. Select $w_i(r) = 1$, $i = 1, \dots, n$ and set $j = 1$.
2. Estimate the residuals by OLS method. In order to reach convergence, corrected least-squares estimates are used [63].
3. Calculate the weights $w_i(r_j)$ from Eq. (22), for $j = j + 1$.
4. Use the reweighted least-squares to estimate \mathbf{b}_j and the residuals r_j .
5. If the estimates \mathbf{b}_j and \mathbf{b}_{j+1} are not close enough, go to step 3, otherwise $\mathbf{b}_j = \mathbf{b}_M$.

It should be noted that in the j th iteration the weights used have been calculated from residuals \hat{e}_{j-1} in the $(j-1)$ th iteration. By applying this method, the robust estimate of parameter and can be evaluated. An independent estimate $\hat{\sigma}$ from the residuals \hat{e}_{j-1} determined in the previous iteration seems to be most convenient. A useful expression is

$$\hat{\sigma} = \frac{\text{med}(|\hat{e}_i - \text{med}(\hat{e}_i)|)}{0.6745} \quad (23)$$

where $\text{med}(\hat{e}_i)$ is the median calculated from all residuals and for sake of simplicity, the indices $(j-1)$ denoting the actual iteration used for residual estimation, are omitted. The constant 0.6745 for large sample size fixes the value $\hat{\sigma}$ to be equal to the residual standard deviation $\hat{\sigma}$ but forms a normal error distribution. A simpler option is $\hat{\sigma} = 2.1 \text{med}(|\hat{e}_i|)$. The advantage of the IRWLS method is the fact that after termination of iterative refinement of parameter estimate the covariance matrix of OLS method is already the estimate $D(\mathbf{b}_M)$. The regression model identification and parameters estimation also may be found in the book by Walter and Pronzato [64].

3. Methodology

3.1. Procedure of calibration model building

The quality of the calibration model is evaluated with the use of classical tests, regression diagnostics and some supplementary information about the ‘data + model + method’. The procedure for construction of a calibration regression model consists of following steps:

1. Proposal of a calibration model: the procedure should always start from the simplest linear model of the straight line.
2. Regression triplet analysis.
 - (a) Examination of the data quality: if influential points are found, it is necessary to decide whether these points should be eliminated from the data. If points are eliminated, the whole data treatment must be repeated.
 - (b) Examination of the model quality: if some parameters are statistically insignificant, they are omitted in the new model.
 - (c) Examination of the regression method used: according to the test for fulfilment of assumptions for the least-squares method, and the result of regression diagnostics, the more accurate regression model is constructed.
3. Construction of the more accurate calibration model: on base of results of the regression triplet the new and more accurate calibration model is proposed.
4. Precision limits of calibration and the point and interval estimates of unknown concentration: the precision of a calibration is expressed with three limiting values of the concentration for which the measurement signal is still statistically significantly different from the noise: the critical value, L_C (Eqs. (9) and (10)), the minimum detectable (true) value, L_D (Eqs. (11) and (12)), and the minimum quantifiable (true) value, L_Q (Eqs. (13)–(15)).

3.2. Software used

For creation of regression diagnostic graphs and the computation of regression based charac-

teristics the in-house constructed algorithms Calibration and Linear Regression in S-PLUS were written, and are freely available on the Internet at the address URL: <http://www.trilobyte.cz/calibration>. On basis of the authors’ own experiences with both algorithms, the commercial software ADSTAT 2.5 (Advanced Statistics) (Trilobyte Statistical Software Ltd., Pardubice, Czech Republic) was constructed cf. [62].

4. Results and discussion

4.1. Case study 1. The effect of influential points on precision limits and interval estimate in the additive model of measurement errors

The relationship between absorbance y and the concentration of nitrate x in solution is described by the Lambert–Beer law, $y = \varepsilon dx + \text{const}$, where d is the cuvette length in cm (here $d = 1$ cm), ε is the molar absorptivity and const is the absorbance of the blank. It is to estimate the parameters of the Lambert–Beer law and three precision limits, the critical level L_C , the detection limit L_D and the quantification limit L_Q with the interval estimate of one unknown sample which with replicate absorbances of $y^* = 0.601, 0.602, 0.600, 0.599, M = 4$.

4.1.1. Data

Calibration graph for $n = 16$, $\{1000 \cdot x \text{ [mg of } \text{NO}_3^-], y \text{ [mAU]}\}$: 5 110, 16.1 272, 16.5 224, 21.3 274, 27.5 338, 32.4 389, 38.2 449, 45.3 522, 52.3 595, 57.5 649, 63.2 708, 71.2 791, 80.3 885, 86.2 946, 91.8 1005, 98.2 1067.

4.1.2. Solution

4.1.2.1. Proposal of a calibration model. OLS method and original data gives the calibration straight line model $y = 64.78(\pm 6.35) + 10.20(\pm 0.11) x$, where the standard deviations of the parameter estimates are given in brackets. Presence of influential points causes the interval estimates $[L_L, L_U]$ for b_0 and also for b_1 to be rather broad.

4.1.2.2. Regression triplet

Examination of the data quality. In step 1 three diagnostic plots, i.e. the graph of predicted residuals (Fig. 5a), the Williams graph (Fig. 5b) and Gray's $L-R$ graph (Fig. 5c) indicate that point number 2 is an outlier. When this masking point is removed and influential points detection is repeated, points numbers 1, 16 and 15 are detected as outliers with the graph of predicted residuals (Fig. 5d), points number 1 and 16 with the Williams graph (Fig. 5e) and points numbers 1 and 16 with the Gray's $L-R$ graph. It is evident that, apart from point number 2 which is strongly masking the influence of other points, that there are two other influential points, numbers 1 and 16, and to some extent number 15. Point number 2 also causes heteroscedasticity in the data and the non-normality of random errors being detected in residuals (Table 1). The mean quadratic error of prediction MEP is 190.43, $AIC = 82.86$, and the statistics describing fitness are: the residual standard deviation $s(e) = 12.57$ mAU and the mean of the absolute values of residuals $\bar{e} =$

6.53 mAU. The residual skewness is 8.92 and the residual kurtosis 11.30 so that the residual distribution is strongly asymmetrical with a sharp peak, and, therefore, non-normal. In step 2 three outliers (number 1, 2, 16) were removed from the data and OLS was applied again. Better fit was proven by all regression diagnostics: lower value $MEP = 0.11$ and $AIC = -30.73$, much lower values of $s(e) = 0.29$ mAU and $\bar{e} = 0.19$ mAU. The resulting residuals distribution is symmetric and exhibits homoscedasticity. Due to the better fit in step 2, the calibration results are also more reliable (Fig. 6): the point estimate of the unknown concentration x^* is more accurate in step 2 and its interval estimate $[L_L, L_U]$ is also nearer, the interval limits having changed from $[51.59, 53.40]$ in step 1 to $[52.75, 52.80]$ in step 2.

Examination of the model quality. Even though both parameter estimates are statistically significant at the significance level $\alpha = 0.05$ in step 1, in step 2 more reliable estimates of intercept and slope with their standard deviations are achieved. Therefore, the interval estimates of the intercept

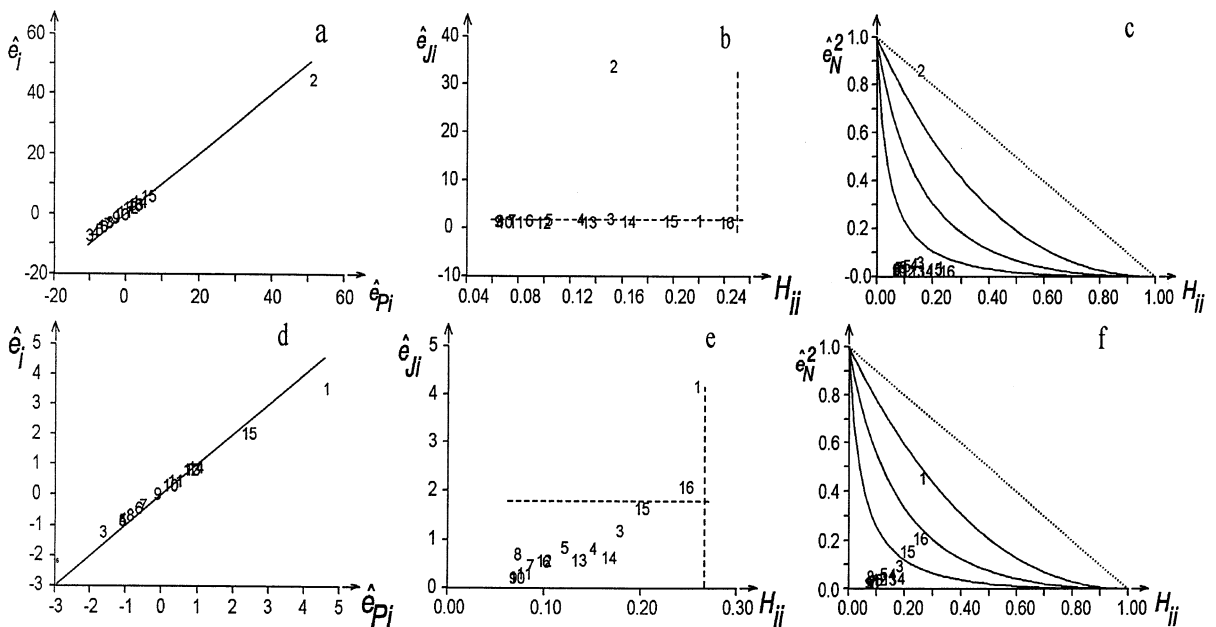


Fig. 5. Diagnostic plots indicating influential points based on residual and hat matrix elements for the original data set of Example 1: (a) graph of predicted residuals, (b) Williams graph, (c) Gray's $L-R$ graph, and these plots after excluding point number 2 from data: (d) graph of predicted residuals, (e) Williams graph, (f) Gray's $L-R$ graph.

Table 1

The effect of influential points on calibration precision limits and interval estimates of unknown concentration of nitrate for 16 points of calibration data $\{x [\mu\text{g of NO}_3^-], y [\text{mAU}]\}$ in the additive model of measurement errors

Characteristic	Step 1	Step 2
Parameters of calibration model $y = \beta_0 + \beta_1 x$, (in brackets the S.D.)		
Intercept $b_0(s_0)$, [mAU]	64.78 (6.35)	53.03 (0.19)
$[L_L, L_U]$ for b_0	[51.16, 78.39]	[52.61, 53.45]
Slope $b_1(s_1)$	10.20 (0.11)	10.36 (0.00)
$[L_L, L_U]$ for b_1	[9.96, 10.43]	[10.36, 10.37]
<i>Critical level</i>		
y_C [mAU], x_C [μg]	83.38, 1.8	53.60, 0.05
<i>Detection limit</i>		
y_D [mAU], x_D [μg]	101.47, 3.60	54.17, 0.11
y_D [mAU], x_D [μg] (Ebel, Kamm)	100.99, 3.55	54.17, 0.11
<i>Quantification limit</i>		
y_Q [mAU], x_Q [μg]	119.07, 5.33	54.74, 0.16
<i>Unknown concentration</i>		
x^* [μg] for $y^* = 600.0$ mAU, $M = 4$	52.49	52.78
$[L_L, L_U]$ for x^* [μg]	[51.59, 53.40]	[52.75, 52.80]
<i>Regression diagnostics for a fitness test of the calibration straight line</i>		
Mean error of prediction, MEP	190.43	0.111
Akaike information criterion, AIC	82.86	-30.73
Residual standard deviation $s(e)$ [mAU]	12.57	0.29
Mean of absolute values of residuals [mAU]	6.53	0.19
Residual skewness, g_1	8.92	0.02
Residual kurtosis, g_2	11.30	3.42
Homoscedasticity of errors is	Rejected	Accepted
Normality of random errors is	Rejected	Accepted
Conclusion: calibration results are	False	True

The repeated signal ($M = 4$) of unknown sample y^* [mAU] = 601.0, 602.0, 600.0, 599.0 leads to the mean \bar{y}^* [mAU] = 600.0. Step 1, straight line fitted data with outliers using OLS. Step 2, straight line fitted data without three outliers using OLS.

and slope also are nearer (Fig. 6), from [51.16, 78.39] for b_0 in step 1 to [52.61, 53.45] in step 2 and from [9.96, 10.43] for b_1 in step 1 to [10.36, 10.37] in step 2. Fig. 6 shows that the precision limits and all three regions are lower in step 2 and, therefore, calibration is more precise and reliable than for the original data with all outliers. It can be concluded that influential points have a strong effect on the values of precision limits in calibration and also on the interval estimate of unknown concentration.

Examination of the regression method used. Of the seven assumptions of OLS, only homoscedasticity and normality of random errors need be examined here. Removing outliers from data, heteroscedasticity (tested with the Cook–Weisberg test) is now rejected and residuals normality (tested with the Jarque–Berra test) is accepted. Therefore, the crucial assumptions are valid and OLS can be applied to give final results.

4.1.2.3. Construction of a more accurate calibration model. When point 2 is omitted, the classical least-squares method gives the residual regression model $y = 54.97(\pm 0.78) + 10.33(\pm 0.01)x$, with determination coefficient $\hat{R}^2 = 0.9999$, the quadratic error of prediction $\text{MEP} = 2.93$, $\text{AIC} = 82.86$ and the residual standard deviation $s(e) = 1.40$ mAU. All these statistics demonstrate a significant improvement in the statistical regression characteristics, too. Despite the good degree of fit of the regression straight line to the experimental points, the residual diagnostics indicate the presence of some other influential points, i.e. points 1 and 16 and to some extent 15. Since calibration requires the highest precision, points 1 and 16 were removed from the original data set. The regression model by OLS now has the form $y = 53.03(\pm 0.19) + 10.36(\pm 0.00)x$ with determination coefficient $\hat{R}^2 = 0.9999$, the mean quadratic error of prediction $\text{MEP} = 0.11$, $\text{AIC} = -30.73$, the residual standard deviation $s = 0.29$ mAU, and the mean of absolute values of residuals $\bar{e} = 0.19$. Descriptive statistics of residuals prove better fitness and, therefore, more reliable calibration results to have been achieved.

4.1.2.4. Precision limits of calibration. In Table 1 and on Fig. 6 it is shown that in step 1 region I of unreliable detection is from 0 up to $x_C = 1.80 \mu\text{g NO}_3^-$, region II of qualitative estimation is from $x_C = 1.80 \mu\text{g NO}_3^-$ up to $x_D = 3.60 \mu\text{g NO}_3^-$, and region III of quantitative estimation of unknown concentration is above $x_Q = 5.33 \mu\text{g NO}_3^-$. In step 2 region I is $[0, 0.05 \mu\text{g NO}_3^-]$, region II is $[0.05, 0.11 \mu\text{g NO}_3^-]$ and region III above $0.16 \mu\text{g NO}_3^-$. Therefore, it may be concluded that the estimates for the critical level (y_C, x_C), the detection limits (y_D, x_D) and the quantification limits (y_Q, x_Q) strongly depend on the influential points.

4.2. Case study 2. The effect of heteroscedasticity on precision limits and interval estimate in the multiplicative model of measurement errors

The calibration model for the determination of the unknown concentration of silver content [ng dm^{-3}] in $[\text{Bi(m)Sb(n)}](2)\text{Ag(z)Te(3)}$ alloy for 40 points of calibration data $\{x [\text{ng ml}^{-1}], y [\text{mAU}]\}$ in the multiplicative model of measurement errors.

The signal of unknown sample was $[\text{mAU}] = 300.0$.

4.2.1. Data

Calibration graph, $n = 40, \{x [\text{ppm}], y [\text{mAU}]\}$:
 0.00 27.88, 0.00 21.46, 0.00 20.69, 0.00 22.17, 0.00 19.10, 0.50 72.59, 0.50 78.04, 0.50 84.45, 0.50 73.89, 0.50 72.50, 1.01 109.16, 1.01 111.66, 1.01 126.21, 1.01 114.88, 1.01 114.08, 1.52 179.86, 1.52 178.19, 1.52 172.80, 1.52 164.17, 1.52 173.14, 2.02 211.92, 2.02 221.80, 2.02 220.57, 2.02 209.45, 2.02 220.93, 2.52 266.71, 2.52 291.74, 2.52 268.55, 2.52 277.55, 2.52 261.50, 3.53 355.50, 3.53 359.75, 3.53 375.93, 3.53 357.84, 3.53 361.24, 4.54 433.34, 4.54 475.18, 4.54 436.88, 4.54 454.61, 4.54 525.00.

4.2.2. Solution

4.2.2.1. Proposal of a calibration model. In step 1 the calibration straight line $y = 23.3(\pm 3.7) + 97.1(\pm 1.6)x$ was proposed (where the standard deviation of parameter estimates are in brackets) and fitted through original data with the use of OLS, Table 2.

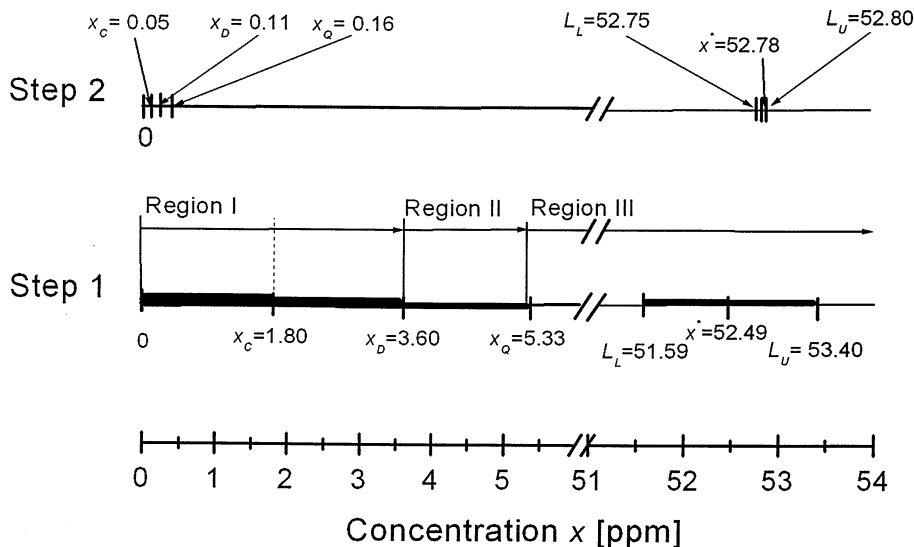


Fig. 6. The three principal analytical regions of calibration precision limits and the point and interval estimates $[L_L, L_U]$ of the unknown concentration in dependence on regression triplet analysis for Example 1 and Table 1 where Region I means 'the region of unreliable detection', Region II means 'the region of qualitative estimation', and Region III means 'the region of quantitative estimation of unknown concentration'. Calibration precision limits x_C, x_D and x_Q , and with $[L_L, L_U]$ are calculated. In Step 1 the original data with all outliers are fitted with the straight line using OLS while in Step 2 the data without the three outliers numbers 1, 2, and 16 and using OLS.

Table 2

The effect of influential points and heteroscedasticity on calibration precision limits and interval estimates of unknown concentration of silver in [Bi(m)Sb(n)](2)Ag(z)Te(3) alloy for 40 points of calibration data $\{x, [\text{ng ml}^{-1}], y, [\text{mAU}]\}$ in the multiplicative model of measurement errors

Characteristic	Step 1	Step 2	Step 3
Parameters of calibration model $y = \beta_0 + \beta_1 x$, (in brackets the S.D.)			
Intercept $b_0(s_0)$, [mAU]	23.32 (3.77)	23.04 (1.99)	24.34 (1.42)
$[L_L, L_U]$ for b_0	[15.68, 30.96]	[19.02, 27.07]	[21.44, 27.25]
Slope $b_1(s_1)$	97.11 (1.56)	97.42 (1.36)	95.16 (0.60)
$[L_L, L_U]$ for b_1	[93.96, 100.27]	[94.66, 100.18]	[93.94, 96.39]
<i>Critical level</i>			
y_C [mAU], x_C [μg]	33.99, 0.11	28.66, 0.06	28.36, 0.04
<i>Detection limit</i>			
y_D [mAU], x_D [μg]	44.27, 0.22	34.11, 0.11	32.32, 0.08
y_D [mAU], x_D [μg] (Ebel, Kamm)	43.91, 0.21	33.95, 0.11	32.27, 0.08
<i>Quantification limit</i>			
y_Q [mAU], x_Q [μg]	54.19, 0.32	39.40, 0.17	36.22, 0.13
<i>Unknown concentration</i>			
\bar{x}^* [μg] for $\bar{y}^* = 300.0$ mAU	2.85	2.84	2.90
$[L_L, L_U]$ for \bar{x}^* [μg]	[2.77, 2.93]	[2.77, 2.93]	[2.86, 2.93]
<i>Regression diagnostics for a fitness test of the calibration straight line</i>			
Mean error of prediction, MEP	228.9	250.6	31.8
Akaike information criterion, AIC	213.8	140.6	82.7
Residual standard deviation, $s(e)$ [mAU]	14.13	5.66	5.53
Mean of absolute values of residuals, \bar{e} [mAU]	8.65	8.49	4.52
Residual skewness, g_1	3.34	8.10	0.17
Residual kurtosis, g_2	10.95	13.95	2.45
Homoscedasticity of errors is	Rejected	Rejected	Accepted
Trend in residuals is	Rejected	Accepted	Rejected
Normality of random errors is	Rejected	Rejected	Rejected
Conclusion: calibration results are	False	False	True

The signal of unknown sample was y^* [mAU] = 300.0. Step 1, straight line fitted data with outliers using OLS. Step 2, straight line fitted data with outliers using IRWLS. Step 3, straight line fitted data without eight outliers using IRWLS.

4.2.2.2. Regression triplet

Examination of the data quality. The straight line was fitted through points with the use of OLS and 8 outliers were detected (point numbers 8, 11, 27, 33, 36, 37, 38, 40). Heteroscedasticity in the data was detected.

Examination of the model quality. Original data were fitted with the straight line and using the IRWLS method. Regression diagnostics prove that better fit was achieved. The IRWLS method found the model $y = 23.0 (\pm 2.0) + 97.4 (\pm 1.4)x$.

Examination of the regression method quality. As

heteroscedasticity and non-normality of random errors in signal values were detected, the IRWLS method was applied again on the data without the eight outliers. The resulting model $y = 23.3 (\pm 1.4) + 95.2 (\pm 0.6)x$ seems to be more reliable as the regression diagnostics achieved better values than in previous steps.

4.2.2.3. Precision limits of calibration. Table 2 and Fig. 7 show that from step to step precision limits decrease and the confidence interval for unknown concentration is nearer.

4.3. Case study 3. The effect of the calibration model proposed on precision limits and interval estimate in the multiplicative model of measurement errors

The calibration model of chromium content [ppm] measured spectrophotometrically leads to the calibration curve. It is to examine a regression triplet and its effect on the precision limits and point and interval estimates of the unknown concentration for straight line and IRWLS and for the spline function. The signal of the unknown sample for $M=4$, y^* [mAU] is 500.3, 501.4, 499.5, 499.8 leading to $\bar{y}^* = 500.0$.

4.3.1. Data

Calibration graph, $n = 14$, $\{x$ [ppm], y [mAU] $\}$: 0.99 59.0, 2.01 111.0, 2.98 162.0, 5.02 257.0, 6.99 343.0, 9.03 422.0, 11.10 486.0, 13.70 558.0, 17.40 642.0, 20.02 711.0, 24.90 807.0, 29.70 865.0, 35.30 899.0, 38.10 919.0.

4.3.2. Solution

4.3.2.1. Proposal of a calibration model. In step 1 the calibration straight line $y = 40.4(\pm 14.7) + 39.5(\pm 2.7)x$ was proposed (where the standard deviation of parameter estimates are in brackets) and fitted through original data with the use of the IRWLS method.

4.3.2.2. Regression triplet

Examination of the data quality. As the straight line does not fit the data well, influential points of a straight line model were not indicated. Poor fitness proves false calibration results (Table 3) and the model must be changed.

Examination of the model quality. Original data were fitted with the quadratic spline function with the use of the IRWLS method. Regression diagnostics for a fitness test prove that the quadratic spline fits the data better and,

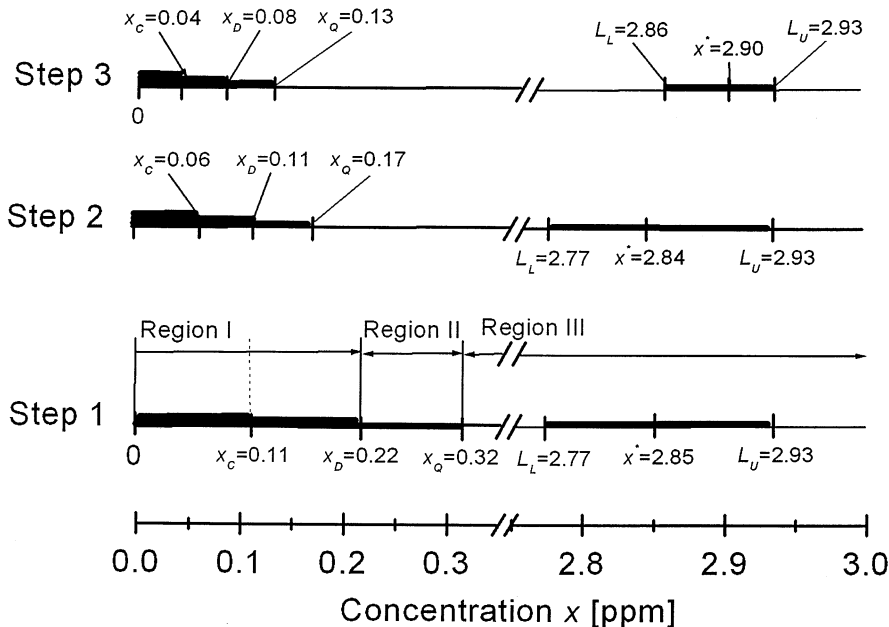


Fig. 7. The three principal analytical regions of calibration precision limits and the point and interval estimates $[L_L, L_U]$ estimates of the unknown concentration in dependence on regression triplet analysis for Example 2 and Table 2. Description of regions is the same as in Fig. 6: in Step 1 the original data with all outliers are fitted with the straight line using OLS, in Step 2 the same as in Step 1 but using the IRWLS method, and in Step 3 the original data without eight outliers and using the IRWLS method were fitted with the straight line.

Table 3

The effect of the model proposed and heteroscedasticity on calibration precision limits and interval estimates of the unknown concentration of chromium for 14 points of calibration data $\{x$ [ppm], y [mAU] $\}$ in the multiplicative model of measurement error

Characteristic	Step 1	Step 2
Parameters of calibration model $y = \beta_0 + \beta_1x (+\beta_2x^2)$, where in brackets the S.D.		
Intercept $b_0(s_0)$, [mAU]	40.4 (14.7)	26.6 (4.5)
$[L_L, L_U]$ for b_0	[8.3, 72.4]	[16.8, 36.4]
Slope $b_1(s_1)$	39.5 (2.7)	47.4 (0.6)
$[L_L, L_U]$ for b_1	[33.8, 45.3]	[46.0, 48.8]
Quadratic term $b_2(s_2)$	----	-0.6 (0.0)
$[L_L, L_U]$ for b_2	[--, --]	[-0.7, -0.6]
<i>Critical level</i>		
y_C [mAU], x_C [μg]	74.9, 0.87	54.4, 0.54
<i>Detection limit</i>		
y_D [mAU], x_D [μg]	105.3, 1.64	78.0, 1.05
y_D [mAU], x_D [μg] (Ebel, Kamm)	102.0, 1.56	76.7, 1.02
<i>Quantification limit</i>		
y_Q [mAU], x_Q [μg]	132.6, 2.33	100.3, 1.54
<i>Unknown concentration</i>		
\bar{x}^* [μg] for $\bar{y}^* = 500.0$ mAU, $M = 4$	11.62	12.03
$[L_L, L_U]$ for \bar{x}^* [μg]	[10.37, 13.31]	[11.54, 12.54]
<i>Regression diagnostics for a fitness test of the calibration graph</i>		
Mean error of prediction, MEP	82772.8	267.9
Akaike information criterion, AIC	112.0	59.6
Residual standard deviation, $s(e)$ [mAU]	51.09	7.66
Mean of absolute values of residuals, \bar{e} [mAU]	75.50	9.41
Residual skewness, $g_1(e)$	4.21	0.53
Residual kurtosis, $g_2(e)$	4.90	2.04
Homoscedasticity of errors is	Rejected	Accepted
Trend in residuals is	Accepted	Rejected
Normality of random errors distribution is	Rejected	Accepted
Conclusion: calibration results are	False	True

The repeated signal ($M = 4$) of unknown sample y^* [mAU] = 500.0, 501.0, 499.0, 500.0 mAU leads to the mean \bar{y}^* [mAU] = 500.0. Step 1, calibration straight line fitted data using IRWLS. Step 2, calibration curve fitted data using the quadratic spline and IRWLS.

therefore, the calibration results are more reliable in step 2 than in step 1.

Examination of the regression method quality. As heteroscedasticity and non-normality of random errors in signal y were proven in the data (the multiplicative model of measurement errors), OLS does not seem to be a convenient regression

method and the IRWLS must be used. Fig. 8 shows calibration precision limits in step 1 and in step 2. It is obvious that the more convenient calibration model with an application of the weighted least squares IRWLS leads to a more optimistic calibration precision and nearer confidence interval of the unknown concentration.

5. Conclusions

The major goal of this study was to propose a procedure that allows the use of regression triplet examination, namely when some assumptions for OLS are violated. For testing assumptions for OLS regression diagnostics are recommended as they do not require a knowledge of an alternative hypothesis or the fulfillment of some conditions of the classical test, but all kinds of deviations from the ideal state are discovered. Selected diagnostic plots were chosen as suitable to give reliable indication of influential points. The spread of points around the calibration straight line is related to the precision of the instrument. This has a significant effect on the critical level, detection and quantification limit, and also on the confidence interval for the unknown concentration. In evaluating calibration experiments, attention should be paid to the model and to the data quality, i.e. the influential points detection.

Another major objective of this study was to provide a comprehensive guide to the regression triplet effect on the statistical uncertainty of the unknown concentration (amount) and on precision limits. It was proven that all three precision limits strongly depend on the influential points, model proposed and the heteroscedasticity in data.

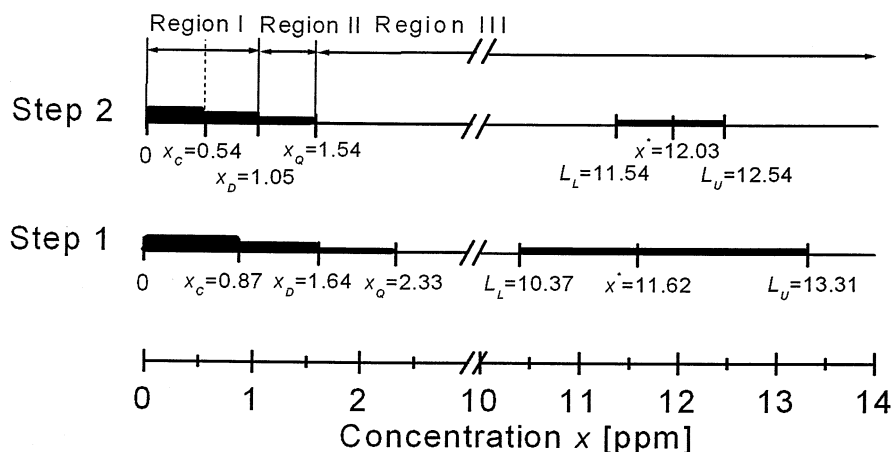


Fig. 8. The three principal analytical regions of calibration precision limits and the point and interval estimates $[L_L, L_U]$ of the unknown concentration in dependence on regression triplet analysis for Example 3 and Table 3. Description of regions is the same as in Fig. 6: in Step 1 the original data with all outliers are fitted with the straight line using IRWLS, and in Step 2 the original data are fitted with the quadratic spline and using IRWLS.

Acknowledgements

The financial support of the Grant Agency of the Czech Republic (Grant No. 303/00/1559) and of the Ministry of Education (Grant No. MSM253100002) is gratefully acknowledged.

References

- [1] R.G. Krutchkoff, Classical and inverse regression methods of calibration, *Technometrics* 9 (1967) 425–439.
- [2] L.A. Currie, Limits for qualitative detection and quantitative determination, *Anal. Chem.* 40 (1968) 586–593.
- [3] J. Berkson, Estimation of a linear function for a calibration line; consideration of a recent proposal, *Technometrics* 11 (1969) 649–660.
- [4] G.M. Tallis, Note on a calibration problem, *Biometrika* 56 (1969) 505–508.
- [5] E.J. Williams, A note on regression methods in calibration, *Technometrics* 11 (1969) 189–192.
- [6] B. Hoadley, A bayesian look at inverse linear regression, *J. Am. Stat. Assoc.* 65 (1970) 356–369.
- [7] H. Scheffé, A statistical theory of calibration, *Ann. Stat.* 1 (1973) 1–37.
- [8] L.T. Frazier, An analysis of a bayes inverse regression method of confidence intervals in linear calibration, *J. Stat. Comp. Simul.* 3 (1974) 99–103.
- [9] J.D. Ingle, R.L. Wilson, Difficulties with determining the detection limit with nonlinear calibration curves in spectrometry, *Anal. Chem.* 48 (1976) 1641–1642.
- [10] L.M. Schwartz, Nonlinear calibration curves, *Anal. Chem.* 48 (1976) 2287–2289.
- [11] L.M. Schwartz, Nonlinear calibration, *Anal. Chem.* 49 (1977) 2062–2068.
- [12] ACS Committee on Environmental Improvement, Guidelines for data acquisition and data quality evaluation in environmental chemistry, *Anal. Chem.* 52 (1980) 2242–2249.
- [13] T. Lwin, J.S. Maritz, A note on the problem of statistical calibration, *Appl. Stat.* 29 (1980) 135–141.
- [14] P.J. Brown, Multivariable calibration, *J. R. Stat. Soc. B* 44 (1982) 287–321.
- [15] W. Merkle, Statistical methods in regression and calibration analysis of chromosome aberration data, *Rad. Env. Biophys.* 21 (1983) 217–233.
- [16] L. Oppenheimer, T.P. Capizzi, R.M. Weppelman, H. Mehta, Determining the lowest limit of reliable assay measurement, *Anal. Chem.* 55 (1983) 638–643.
- [17] L.M. Schwartz, Lowest limit of reliable assay measurement with nonlinear calibration, *Anal. Chem.* 55 (1983) 1424–1426.
- [18] S. Ebel, U. Kamm, Statistische definition der Bestimmungsgrenze, *Fresenius Z. Anal. Chem.* 318 (1984) 293–294.
- [19] D.L. Massart, L. Kaufman, Least median of squares: a robust method for outlier and model detection in regression and calibration, *Anal. Chim. Acta* 187 (1986) 171–179.
- [20] A. Dobrigal, D.A.S. Fraser, Line calibration and conditional inference, *Commun. Stat. Theory Meth.* 16 (1987) 1037–1048.
- [21] R. Smith, M. Corbett, Measuring marathon courses: an application of statistical calibration theory, *Appl. Stat.* 36 (1987) 283–295.

- [22] R. Tibshirani, Noninformative priors for one parameter of many, *Biometrika* 76 (1989) 604–608.
- [23] B.P. Carlin, A.E. Gelfand, A sample reuse method for accurate parametric empirical bayes confidence intervals, *J. R. Stat. Soc. B* 53 (1991) 189–200.
- [24] J.J. Lee, A note on the conditional approach to interval estimation in the calibration, *Biometrics* 47 (1991) 1573–1580.
- [25] C.h. Osborne, Statistical calibration: a review, *Int. Stat. Rev.* 59 (1991) 309–336.
- [26] D.L. MacTaggart, S.O. Farwell, Analytical use of linear-regression, part. I: regression procedures for calibration and quantitation, *J. AOAC Int.* 75 (1992) 594–608.
- [27] R.L. Cooley, Exact scheffé-type confidence intervals for output from groundwater flow models, *Water Resources Res.* 29 (1993) 35–50.
- [28] R.D. Fisch, G.A. Strehlau, A simplified approach to calibration confidence sets, *Am. Stat.* 47 (1993) 168–171.
- [29] L.A. Currie, G. Svehla, Nomenclature for the presentation of results of chemical analysis, *Pure Appl. Chem.* 66 (1994) 595–608.
- [30] M.C. Ortiz, J. Arcos, J.V. Juarros, J. López-Palacios, L.A. Sarabia, Robust procedure for calibration and calculation of the detection limit of trimipramine by adsorptive stripping voltametry at a carbon paste electrode, *Anal. Chem.* 65 (1993) 678–682.
- [31] K.S. Johnston, S.S. Yee, K.S. Booksh, Calibration of surface plasmon resonance refractometers using locally, *Anal. Chem.* 69 (1997) 1844–1851.
- [32] J. Mocak, A.M. Bond, S. Mitchell, G. Scollary, A statistical overview of standard (IUPAC and ACS) and new procedures for determining the limits of detection and quantification: application to voltametric and stripping techniques, *Pure Appl. Chem.* 69 (1997) 297–328.
- [33] L.A. Currie, W. Horwitz, IUPAC recommendation for defining and measuring detection and quantification limits, *Analysis* 22 (1994) M24–M26.
- [34] B. Walczak, Outlier detection in multivariate calibration, *Chemometr. Intell. Lab. Syst.* 28 (1995) 259–272.
- [35] L.A. Currie, Nomenclature in evaluation of analytical methods including detection and quantification capabilities (IUPAC Recommendations 1995), *Pure Appl. Chem.* 67 (1995) 1699–1723.
- [36] L.A. Currie, Detection: international update, and some emerging dilemmas involving calibration, the blank, and multiple detection decisions, *Chemometr. Intell. Lab. Syst.* 37 (1997) 151–181.
- [37] L. Renman, D. Jagner, Asymmetric distribution of results in calibration curve and standard addition evaluations, *Anal. Chim. Acta* 357 (1997) 157–166.
- [38] M.E. Zorn, R.D. Gibbons, W.C. Sonzogni, Weighted least-squares approach to calculating limits of detection and quantification by modeling variability as a function of concentration, *Anal. Chem.* 69 (1997) 3069–3075.
- [39] K. Danzer, L.A. Currie, Guidelines for calibration in analytical chemistry-part 1, fundamentals and single component calibration (IUPAC Recommendation 1998), *Pure Appl. Chem.* 70 (1998) 993–1014.
- [40] L.A. Currie, Detection-overview of historical, societal, and technical issues, *ACS Symposium Series* 361 (1988) 1–62.
- [41] L.A. Currie, International recommendation offered on analytical detection and quantification concepts and nomenclature, *Anal. Chim. Acta* 391 (1999) 103–103.
- [42] L.A. Currie, Nomenclature in evaluation of analytical methods including detection and quantification capabilities (IUPAC Recommendations 1995), *Anal. Chim. Acta* 391 (1999) 105–126.
- [43] L.A. Currie, Detection and quantification limits: origins and historical overview, *Anal. Chim. Acta* 391 (1999) 127–134.
- [44] M.E. Zorn, R.D. Gibbons, W.C. Sonzogni, Evaluation of approximate methods for calculating the limit of detection and limit of quantification, *Environ. Sci. Technol.* 33 (1999) 2291–2295.
- [45] L.A. Currie, Detection and quantification and the evaluation of low-level data: some international perspectives and continuing challenges, *J. Radioanal. Nucl. Chem.* 245 (2000) 145–156.
- [46] R. Boque, N.K.M. Faber, F.X. Rius, Detection limits in classical multivariate models, *Anal. Chim. Acta* 423 (2000) 41–49.
- [47] International standards: capability of detection: part 2, methodology in the linear calibration case, *ISO 11842-2* (2000) E.
- [48] I. Kuselman, Introduction to the proceedings of the international conference on metrology-trends and applications in calibration and testing laboratories, 16–18 May 2000, Jerusalem, Israel, *Accred. Quality Assur.* 6 (2001) 54–54.
- [49] I. Kuselman, I. Goldfeld, B. Anisimov, Analysis of long-term distributions of calibration parameters and calibration intervals for an atomic absorption spectrophotometer, *Accred. Quality Assur.* 6 (2001) 61–65.
- [50] M. Meloun, J. Militký, M. Forina, Chemometrics for analytical chemistry, in: *PC-Aided Regression And Related Methods*, vol. 2, Ellis Horwood, Chichester, 1994.
- [51] J.R. Rosenblatt, C.H. Spiegelman, A bayesian analysis of the linear calibration problem-discussion, *Technometrics* 23 (1981) 329.
- [52] S. Ebel, U. Brecht, Confidence Interval of calibration and analytical measurement using ion-sensitive electrodes, *Z. Anal. Chem.* 327 (1987) 157.
- [53] S. Ebel, U. Brecht, Limit of detection and limit of determination using ion-sensitive electrodes and direct potentiometry, *Z. Anal. Chem.* 329 (1987) 7–11.
- [54] L.J. Naszodi, *Technometrics* 20 (1978) 201.
- [55] R.G. Brereton, Introduction to multivariate calibration in analytical chemistry, *Analyst* 125 (2001) 2125–2154.
- [56] M. Meloun, J. Militký, M. Hill, R.G. Brereton, Problems in Regression Modelling and Their Solutions, *The Analyst* 127 (2002) 433–450.
- [57] M. Meloun, J. Militký, Detection of single influential points in regression model building (tutorial), *Anal. Chim. Acta* 439 (2001) 169–191.

- [58] D.X. Williams, Letter to the editor, *Appl. Stat.* 22 (1973) 407–408.
- [59] J.B. Gray, Graphics for regression diagnostics, Proceedings of the statistical computing section, Amer. Statist. Assoc., (1985) 102–107.
- [60] R.D. Cook, S. Weisberg, Diagnostics for heteroscedasticity in regression, *Biometrika* 70 (1983) 1.
- [61] C.M. Jarque, A.K. Berra, A test for normality of observations and regression residuals, *Int. Stat. Rev.* 55 (1987) 163.
- [62] ADSTAT, Trilobyte Statistical Software, Pardubice, Czech Republic, 2001.
- [63] G.R. Philip, E.M. Eytng, Comparison of conventional and robust regression in analysis of chemical data, *Anal. Chem.* 55 (1983) 1134–1138.
- [64] E. Walter, L. Pronzato, Identification of Parametric Models from Experimental Data, Springer, 1997.