Anal Bioanal Chem
DOI 10.1007/s00216-008-1912-4

ORIGINAL PAPER

Precision limits and interval estimation in the calibration of 1-hydroxypyrene in urine and hexachlorbenzene in water, applying the regression triplet procedure on chromatographic data

Milan Meloun · Zdeňka Dluhošová

Received: 26 September 2007 / Revised: 15 January 2008 / Accepted: 23 January 2008

Abstract A method for the determination of 1-hydroxypyrene in urine and hexachlorbenzene in water applying the regression triplet in the calibration procedure of chromatographic data has been applied. The detection limit and quantification limit are currently calculated on the basis of the standard deviation of replicate analyses at a single concentration. However, since the standard deviation depends on concentration, these single-concentration techniques result in limits that are directly dependent on spiking concentration. A more rigorous approach requires first careful attention to the three components of the regression triplet (data, model, method), examining (1) the data quality of the proposed model, (2) the model quality and (3) the least-squares method to be used for fulfilment of all least-squares assumptions. For high-performance liquid chromatography determination of 1-hydroxypyrene in urine and gas chromatography analysis of hexachlorbenzene in water, this paper describes the effects of deviations from five basic assumptions The paper considers the correction of deviations: identifying influential points, namely, outliers, the calibration task depends on the regression model used, and the least-squares method is based on the assumptions of the normality of the errors, homoscedasticity and the independence of errors. Results show that the approach developed provides improved estimates of analytical limits and that the single-concentration approaches currently in wide use are seriously flawed.

Keywords Calibration precision · Outliers
Influential points · Iterative reweighted least squares ·
Detection limit · 1-hydroxypyrene · Hexachlorbenzene

M. Meloun (✉)
Department of Analytical Chemistry,
Faculty of Chemical Technology, University of Pardubice,
532 10 Pardubice, Czech Republic
e-mail: milan.meloun@upce.cz

Z. Dluhošová
Institute of Public Health in Ostrava,
Partyzánske nám. 7,
702 00 Ostrava, Czech Republic

Introduction

In the analytical laboratory, the physicochemical relationship between the objective, e.g., the concentration of a certain component, and the instrumental response must be determined by a calibration procedure using "samples" of a known concentration. Linear regression is perhaps the most used and abused statistical method in calibration—a survey has been provided by the present author [1]. A common mistake is to blindly force a classical regression fit onto any set of calibration data with a presumed linear relationship. 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 [2–7]. Everyone who has worked with contaminated data sets realizes how many problems can be caused by even a small group of outlying observations, that is, observations for which y_i deviates from the relationship followed by the majority of the data.

Application of ordinary least squares (OLS) in calibration is based on the assumptions of normality, homoscedasticity and independence of the measurements [8–12]. 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 of the 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 (1) the calibration data quality of a proposed calibration model, (2) the calibration model for a given set of data, and (3) fulfilment of all least-squares assumptions.

OLS usually gives equal weight to every point of calibration data. However, every point does not have an 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 usually 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 [13, 14].

This paper summarizes a procedure of powerful general diagnostics for detecting observations that differ from the bulk of the data. These may be individual observations that do not belong to the general model, i.e., influential points or outliers. In the first calibration case the high-performance liquid chromatography (HPLC) determination of 1-hydroxypyrene in the urine from cokery workers was performed. Cokery workers are exposed to a complex mixture of gaseous and particulate contaminants. There is sufficient evidence that occupational exposure causes an excess of skin and lung cancers among coke oven workers. The most extensively studied compounds that have adverse health effects in the cokery environment are the polycyclic aromatic hydrocarbons (PAHs). Most of the mutagenic and carcinogenic PAHs are known to exist in the particulate phase. PAH metabolites in human urine can be used as biomarkers of internal dose to assess recent exposure to PAHs. PAH metabolites that have been detected in human urine include free and bound 1-hydroxypyrene as well as a number of other hydroxylated PAHs. An acid hydrolysis method was used instead of enzyme extraction, equipped with a column-switching system for the pretreatment of samples, in the gas-chromatographic determination of 1-hydroxypyrene in the urine from workers. The second calibration case considers the determination of hexachlorbenzene in water. Proposed environmental regulations require analyses of a wide variety of organic compounds in water by gas chromatography/mass spectrometry. Since the required detection limits are at the parts-per-billion level, an extraction/concentration step in the sample preparation is a necessity. Extraction of water by shaking with solvents is the simplest and the most rapid method, and the extract is then chromatographed without preconcentration on a glass capillary column.

Experimental

Calibration standards and chromatographic data

1-Hydroxypyrene in urine

The calibration data of 1-hydroxypyrene in urine were obtained by the HPLC method, which allows the determination of free and conjugated 1-hydroxypyrene in urine. After enzymatic hydrolysis to release the conjugated part of the 1-hydroxypyrene, the analyte was separated from the matrix and enriched by liquid/solid extraction in a reversed-phase column. The enzyme β-glucuronidase/aryl sulfatase was used for enzymatic hydrolysis and no optimization was performed. The components of the eluate were separated by means of HPLC and 1-hydroxypyrene was determined with a fluorescence detector. A fluorescence detector capable of measurement at the excitation wavelength of 242 nm or 336 nm and at the emission wavelength of 388 nm was used to monitor 1-hydroxypyrene. For the set of various concentration levels the peak area y in luminescence units (LU) proportional to a content of 1-hydroxypyrene x (micrograms per liter) in urine was monitored. The operational parameters for the Hewlett-Packard 1100 HPLC instrument (Agilent) were as follows: a LiChroCART 250 mm×4 mm column with LiChrospher 100 RP-18, 5 μm and 40 °C, mobile phase for solvent A—40% methanol/60% water (v/v), mobile phase for solvent B—100% methanol. The gradient program was used with the following values [time (minutes), solvent A (percent), solvent B (percent)]: 0, 90, 10; 35, 10, 90; 42, 10, 90; 45, 90, 10; 50, 90, 10. A flow rate of 0.8 ml/min was used and the sample volume was 50 μl. The interval estimate of one unknown sample with four replicate values of y^*=6010.0, 6020.0, 6000.0, 5990.0 LU requires calculation. The calibration graph contains data for 23 points with x (micrograms per liter), y (peak area in LU) values 0.05, 41; 0.05, 157; 0.1, 189; 0.1, 231; 0.2, 431; 0.2, 392; 0.5, 1,079; 0.5, 881; 1, 1,866; 1, 1,900; 2, 3,042; 2, 2,962; 3, 4,353; 4, 5,522; 5, 7,562; 5, 6,685; 6, 8,283; 7, 9,637; 8, 10,357; 8, 11,850; 10, 13,619; 12, 16,300; and 12, 16,250.

Hexachlorbenzene in water

The calibration data of hexachlorbenzene content (micrograms per liter) in water were measured with a Fisons Instruments MEGA II HRGC chromatograph equipped with an electron capture detector. The detector was maintained at 300 °C, with the flow rate of makeup gas (N_2) at 45–50 ml/min. The flow rate of the carrier gas (N_2) was 1 ml/min. A 1-μl injection was performed in the splitless mode with an injection port temperature of 240 °C. The column was a DB-5 capillary column (30 mm× 0.32 mm, 0.25-μm film thickness). The oven temperature

program was as follows: initial temperature 50 °C, 40 °C/min to 210 °C, 1.5 °C/min to 240 °C, 10 °C/min to 270 °C, retained for 3 min at 270 °C. The signal of the unknown sample y^* was 250.0 mV. The calibration graph contains data for 24 points with x (micrograms per liter), y (millivolts) values 0.2, 6.69; 0.2, 13.3; 0.5, 16.74; 0.5, 17.16; 1, 35.41; 1, 34.88; 2, 75.53; 2, 77.53; 2.5, 84.89; 2.5, 90.95; 3, 96.71; 3, 107.38; 4, 127.27; 4, 134.89; 5, 158.22; 5, 160.65; 7.5, 208.17; 7.5, 215.28; 10, 261.04; 10, 252.25; 15, 313.7; 15, 318.2; 20, 377.8; and 20, 367.4.

Calibration model building procedure

The quality of the calibration model was evaluated with the use of regression diagnostics and some supplementary information about the "data + model + method." Consistent with the above reasoning, the numerical calibration procedure should exploit the ability of the regression triplet procedure to detect the group of aligned experimental data and the optimal precision and accuracy of the estimates provided by the OLS regression. Consequently, the following computational sequence was developed:

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, a more accurate regression model is constructed.
3. *Construction of a more accurate calibration model*. On the basis of the results of the regression triplet a 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, y_C, the minimum detectable (true) value, y_D, and the minimum quantifiable (true) value, y_Q.

Supporting information available

Complete experimental and computational procedures, input data specimens and corresponding output in numerical and graphical form for the program Calibration in S-Plus are available free of charge online at http://meloun.upce.cz in the block DATA [32].

Results and discussion

The determination of 1-hydroxypyrene in the urine from cokery workers

Urine samples were first treated with acid hydrolysis, followed by solvent extraction, prior to being injected into the separation system for the determination with HPLC and fluorescence detection. The HPLC method allows the determination of free and conjugated 1-hydroxypyrene in urine. After enzymatic hydrolysis to release the conjugated part of the 1-hydroxypyrene the analyte was separated from the matrix and enriched by liquid/solid extraction in a reversed-phase column. The components of the eluate were separated by means of HPLC and 1-hydroxypyrene was determined with a fluorescence detector. Calibration was carried out using standards prepared with urine from unexposed persons. The urine was spiked with 1-hydroxypyrene, then processed and analyzed in the same way as the assay samples. For the set of various concentration levels the peak area y in LU proportional to a content of 1-hydroxypyrene x (micrograms per liter) in urine was monitored. The resulting calibration graph described by the linear calibration model $y=\beta_0+\beta_1 x$ had very good correlation coefficients, $r>0.999$. The parameters of the linear calibration model and three precision limits, the critical level L_C, the detection limit L_D and the quantification limit L_Q, were estimated using the following steps:

1. *Proposal of a calibration model*. A calibration straight line is calculated using the OLS with the resulting area under the peak as a function of the standard solution concentration, and gives the calibration model $y=244.9\ (89.9)+1345.0\ (16.5)x$, where the standard deviations of the parameter estimates are given in parentheses. The presence of influential points causes the interval estimates $[L_L, L_U]$ for b_0, i.e., [57.9, 432.0], and for b_1, i.e., [1,310.8, 1,379.2], to be rather broad (Table 1).
2. *Regression triplet* (a) *Examination of the data quality.* In step 1 five diagnostic plots were made of influential points behind the calibration graph (Fig. 1a), i.e., the graph of the jackknife residuals indicating suspicious points only (Fig. 1b), the Williams graph (Fig. 1c), the graph of predicted residuals (Fig. 1d), the Pregibon plot (Fig. 1e) and Gray's L-R graph (Fig. 1f); all these graphs indicate that points 15, 19 and 20 are strong outliers. When these masking outliers are removed and the detection of influential points is repeated, points 9 and 10 are also indicated as outliers. It is evident that, apart from points 15, 19 and 20, which strongly mask

Table 1 The effect of influential points (outliers) on calibration precision limits and interval estimates of unknown concentration of 1-hydroxypyrene in urine for 23 points of calibration data [x (micrograms per liter), y (peak area in luminescence units, LU] in the additive model of measurement errors

Characteristic	Step 1	Step 2
Parameters of calibration model $y=\beta_0+\beta_1 x$ (in parentheses, the standard deviation)		
Intercept $b_0(s_0)$ (LU)	244.9 (89.9)	170.7 (38.6)
L_L, L_U for b_0, (LU)	57.9, 432.0	88.8, 252.6
Slope $b_1(s_1)$	1,345.0 (16.5)	1,344.4 (7.1)
L_L, L_U for b_1	1,310.8, 1,379.2	1,329.3, 1,359.5
Critical level		
y_C (LU), x_C (μg/l)	503.4, 0.192	283.0, 0.084
Detection limit		
y_D (LU), x_D (μg/l)	749.7, 0.375	393.2, 0.165
Quantification limit		
y_Q (LU), x_Q (μg/l)	990.6, 0.554	502.2, 0.247
Unknown concentration		
x^* (μg/l) for y^*=6000.0 LU, M=4	4.279	4.336
$[L_L, L_U]$ for x^* (μg/l)	[4.141, 4.418]	[4.272, 4.400]
Regression diagnostics for a fitness test of the calibration straight line		
Fisher–Snedecor F test, F_R vs. $F_{1-\alpha}(m-1)(n-m)$	6,676.5 vs. 4.32	35,764.2 vs. 4.49
Determination coefficient R^2	0.9967	0.9998
Mean error of prediction	103,476.0	15,957.0
Akaike information criterion	265.6	175.1
Residual standard deviation $s(e)$ (LU)	308.9	122.9
Mean of absolute values of residuals $\overline{e}$ (LU)	208.0	90.5
Homoscedasticity of errors	Rejected	Accepted
Normality of random errors	Rejected	Accepted
Conclusion: calibration results	False	True

The repeated signal (M=4) of unknown sample y^*=6010.0, 6020.0, 6000.0, 5990.0 (LU) leads to the mean $\overline{y}^* = 6000.0$ LU. Step 1—straight line fitted data with outliers using ordinary least squares (OLS); step 2—straight line fitted data without five outliers (points 9, 10, 15, 19, 20) using OLS (Calibration in S-Plus)

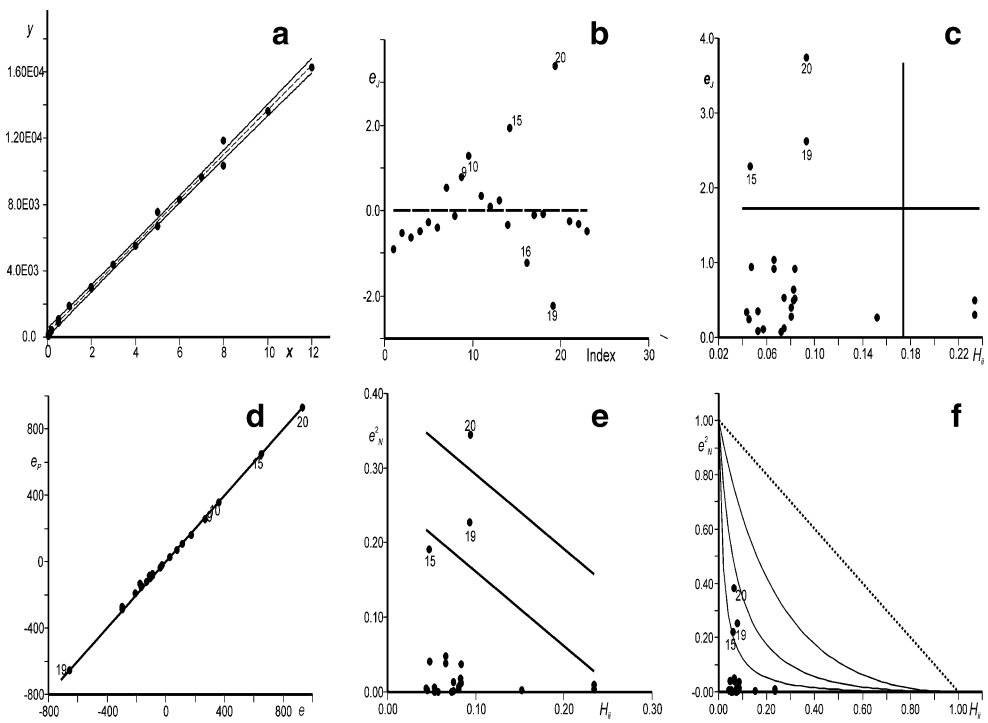

Fig. 1 Diagnostic plots indicating five influential points, five outliers including two leverages, based on residuals and hat matrix elements for the original data set of 1-hydroxypyrene in urine: **a** calibration straight line; **b** scatter plot of jackknife residuals indicating six suspicious points; **c** Williams graph indicating three outliers (*15*, *19*, *20*) and two leverages (22, 23); **d** Graph of predicted residuals indicating five outliers (*19*, *9*, *10*, *15*, *20*); **e** Pregibon plot indicating one strong influential point (*20*) and one medium influential point (*19*); **f** Gray's L-R graph indicating three outliers (*15*, *19*, *20*) and two leverages (22, 23)

the influence of other points, that there are two other influential points, points 9 and 10. Points 19 and 20 could cause slight heteroscedasticity only in the data and the nonnormality of random errors being detected in residuals (Table 1). The mean quadratic error of prediction (MEP) is 103,476.0, the Akaike information criterion (AIC) is 265.6 and the statistics describing goodness of fit are the residual standard deviation $s(e)$= 308.9 LU and the mean of the absolute values of residuals $|\overline{e}| = 208.0$ LU. The residual distribution is asymmetrical with a sharp peak, and is therefore nonnormal. The calibration results prove that the model is false and should be corrected in step 2.

In step 2 five outliers (nos. 9, 10, 15, 19, 20) were removed from the data and the OLS procedure was applied again. A better fit was proven by all of the regression diagnostics: lower values for both criteria, MEP=15,957.0 and AIC=175.1, were achieved, and much lower values of $s(e)$=122.9 LU and $|\overline{e}| =$ 90.5 LU were attained, expressing better fitness. The resulting residuals distribution is symmetric and Gaussian. and exhibits homoscedasticity. Because of the better fit achieved in step 2, the calibration results are also more reliable (Fig. 2, Table 1): the point estimate of the unknown concentration x^*=4.336 μg/l is more accurate in step 2 than in step 1, x^*=4.279 μg/l, and its interval estimate $[L_L, L_U]$ is also nearer, the interval limits having changed from [4.141, 4.418] in step 1 to [4.272, 4.400] in step 2. (b) *Examination of the model quality.* Even though both parameter estimates are statistically significant at the α=0.05 significance level

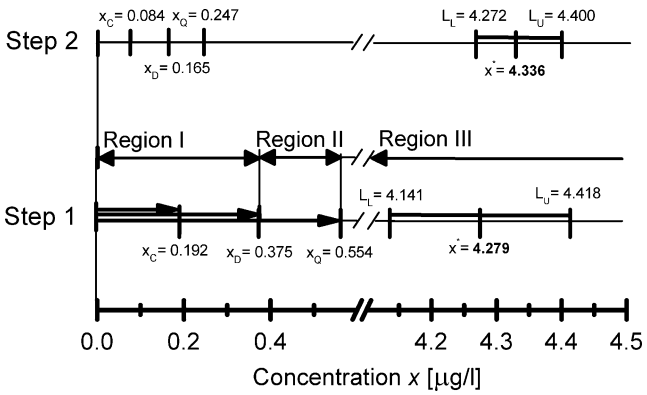

Fig. 2 The three principal analytical regions of calibration precision limits and the resulting point estimate $\widehat{x}^*$ and interval estimates $[L_L, L_U]$ of the unknown concentration with dependence on regression triplet analysis for 1-hydroxypyrene in urine and the data in Table 1, where *region I* is the region of unreliable detection, *region II* is the region of qualitative estimation and *region III* is *the* region of quantitative estimation of unknown concentration. Calibration precision limits x_C, x_D and x_Q, and $\widehat{x}^*$ with $[L_L, L_U]$ are estimated. In step 1 the original data with all outliers are fitted with the straight line model using ordinary least squares (OLS), while in step 2 the data without the five outliers 9, 10, 15, 19 and 20 are fitted using OLS

in step 1, in step 2 more reliable estimates of the intercept and the slope with their standard deviations were achieved. Therefore, the interval estimates of the intercept and slope also are nearer (Fig. 2): from [57.9, 432.0] for b_0 in step 1 to [88.8, 252.6] in step 2 and from [1,310.8, 1,379.2] for b_1 in step 1 to [1,329.3, 1,359.5] in step 2. Figure 2 shows that the precision limits and all three regions are lower in step 2 and that therefore calibration is more precise and reliable than for the original data with all outliers. It can be concluded that influential points have a very strong effect on the values of the precision limits in calibration and also on the interval estimate of unknown concentration. (c) *Examination of the regression method used.* Of the seven assumptions of OLS, only homoscedasticity and the normality of random errors need be examined here. When outliers are removed from the data, homoscedasticity (tested with the Cook–Weisberg test) is accepted and the normality of residuals (tested with the Jarque–Berra test) is also accepted. Therefore, the crucial assumptions are valid and OLS can be applied to give the final results.

3. *Stepwise construction of a more accurate calibration model*. When points 15, 19 and 20 are omitted, the classical least-squares method OLS gives the residual regression model y=225.9 (47.6)+1338.9 (9.2) x, with the determination coefficient $\widehat{R}^2 = 0.9991$, MEP= 27696.1, AIC=205.6 and $s(e)$=162.7 LU. 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 9 and 10. Since calibration requires the highest precision, points 9 and 10 were removed from the original data set. The regression model by OLS now has the form y=170.7 (38.6)+1344.4 (7.1) x, with $\widehat{R}^2 = 0.9998$, MEP=15957.6, AIC=175.1, $s(e)$= 122.9 LU and $|\overline{e}| = 90.5$ LU. Descriptive statistics of residuals prove that better fitness and therefore more reliable calibration results have been achieved.
4. *Precision limits of calibration.* In Table 1 and Fig. 2 it is shown that in step 1 region I of unreliable detection is from 0 up to x_C=0.192 μg/l 1-hydroxypyrene, region II of qualitative estimation is from x_C= 0.192 μg/l 1-hydroxypyrene up to x_D=0.375 μg/l 1-hydroxypyrene and region III of quantitative estimation of unknown concentration is above x_Q=0.554 μg/l 1-hydroxypyrene. In step 2 region I is from 0 to 0.084 μg/l 1-hydroxypyrene, region II is from 0.084 to 0.165 μg/l 1-hydroxypyrene and region III is above 0.247 μg/l 1-hydroxypyrene. 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, namely, on the outliers.

The effect of the calibration model proposed for hexachlorbenzene in water on precision limits

An extraction/concentration step in the determination of hexachlorbenzene in water in the sample preparation is a necessity. One liter of cold water was extracted with 1 ml isooctane and shaken for 10 min. The extract was chromatographed without preconcentration on a glass capillary column according to the method of Hrivnák [33] and determination was performed in the following steps:

1. *Proposal of a calibration model.* In step 1 the calibration straight line y=38.4 (7.4)+18.6 (0.9)x was proposed (where the standard deviations of the parameter estimates are in parentheses) and fitted through original data with the use of the OLS method.
2. *Regression triplet* (a) *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 (Fig. 3, Table 2) and means that the model must be changed. (b) *Examination of the model quality.* In the second step the original data were fitted with the quadratic spline function with the use of the OLS method. Regression diagnostics for a fitness test prove that the quadratic spline fits the data better and therefore the calibration results are more reliable in step 2 than in step 1 (Table 2). (c) *Examination of the regression method quality.* As heteroscedasticity and nonnormality 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 iterative method of reweighted least squares (IRWLS) must be used. Figure 4 shows calibration precision limits in steps 1, 2 and 3. It is obvious that the more convenient calibration model with an application of the IRWLS leads to a more optimistic calibration precision and nearer confidence interval of the unknown concentration in steps 2 and 3.

Conclusions

The major goal of this study was to propose a procedure that allows the use of regression triplet examination in calibration, namely, when some basic assumptions for classical OLS are violated. For testing assumptions for OLS, regression diagnostics are recommended as they do not require knowledge of an alternative hypothesis or the fulfilment 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 for giving a 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., detection of influential points.

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, the model proposed and the heteroscedasticity in the data. The procedure used for calibration and determination of the detection limit can be routinely applied to other chemical systems and endows analytical methods with adequate sensitivity and detectability. The detection limit thus obtained takes account of the sensitivity of the analytical method, the nature of the analyte, and the risk of false positives and falses negatives that the analyst is willing to accept.

Appendix

Calibration models

Previous work on the CALIBRATION algorithm in S-Plus [1] has assumed the *additive model of measurement errors* $y_i = g(x_i, \beta) + \varepsilon_i$ where β is the set of adjustable parameters and ε_i is the error term containing errors of measurement, the error due to the blank, etc. It is often assumed that the ε_i are normally distributed and have a constant variance σ^2. 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 the model in the form

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

the signal for the unknown concentration is

$$y_i^* = \beta_0 + \beta_1 \kappa + \varepsilon_i^*, \quad j = 1, ..., M.$$

The task of calibration is to find an estimate $\widehat{x}^*$ of unknown parameter κ, the parameter of our primary interest, and of parameters β_0 and β_1, the supplementary parameters. The estimation assumes normality of the errors ε_i and ε_j^*. The estimate $\widehat{x}^*$ and its confidence interval may be calculated by several procedures: *the straight estimate* of κ is obtained in the form $\widehat{x}^* = \overline{x} + (y^* - \overline{y})/b_1$, where y^* is the measured signal (or the average $\overline{y}^*$ for $M>1$ repeated measurements, respectively) and b_1 is the estimate of the slope β_1. This estimate is

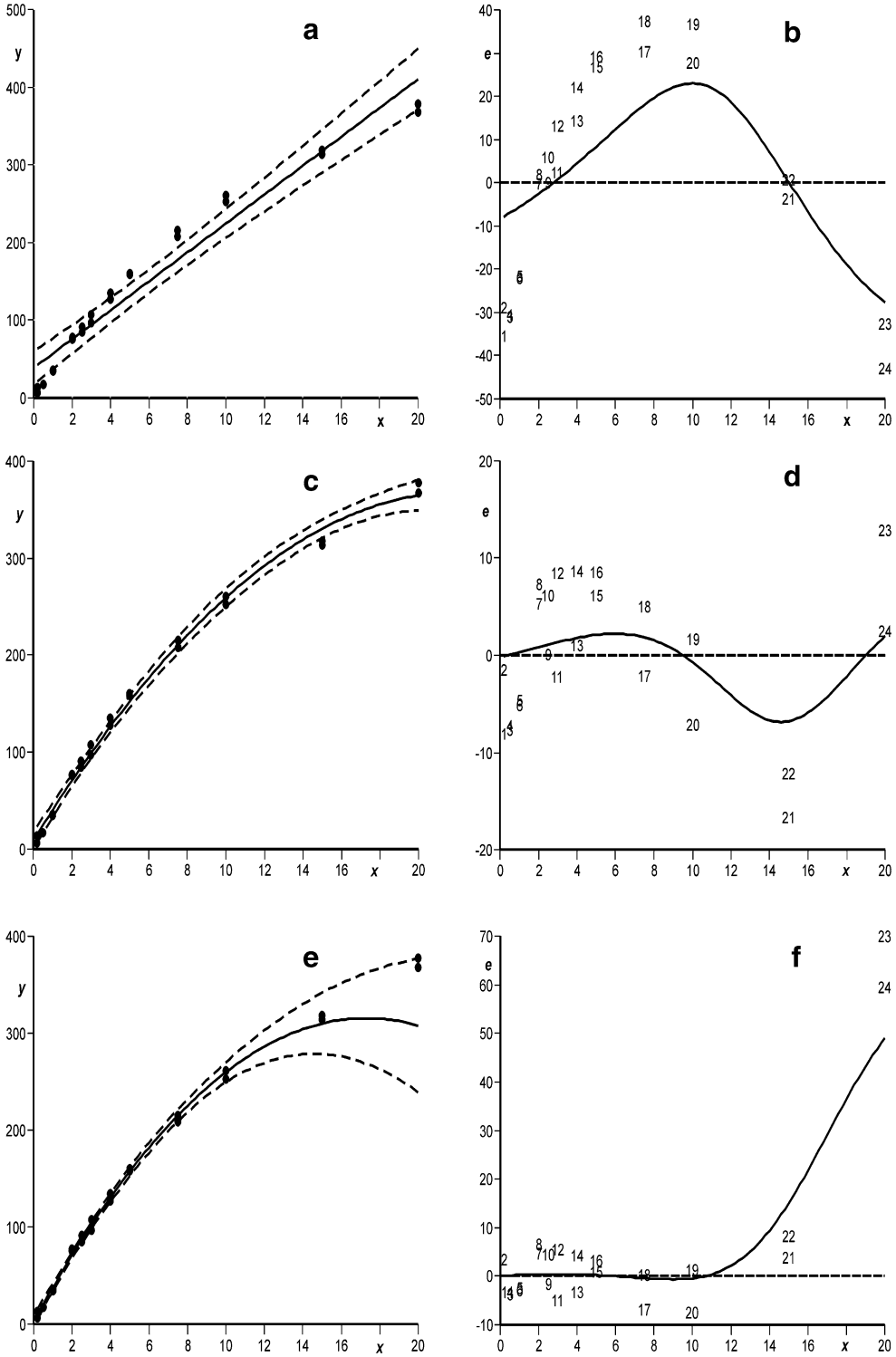

Fig. 3 Calibration model building and testing with the use of the regression triplet for original chromatographic data of hexachlorbenzene in water: **a** the calibration straight-line model used; **b** scatter plot of classical residuals for the straight-line model ad a used, **c** the calibration spline function model with OLS used; **d** scatter plot of classical residuals for spline model ad c used; **e** the calibration weighted spline function model with the iterative method of reweighted least squares (IRWLS) used; **f** scatter plot of classical residuals for the spline model ad e used

generally biased and a correction can be made with *Naszodi's estimates* [15]:

$$\hat{x}^{*}_{\mathrm{B}} = \bar{x} + \frac{(y^{*} - \bar{y})b_1}{b_1^2 + \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}.$$

In the construction of confidence intervals of the estimates $\widehat{x}^{*}$ and $\widehat{x}^{*}_{B}$ for more scattered data, the simplest is the determination of the variance of unknown concentration $\sigma^2(\widehat{x}^{*})$ with an assumption of normality. The limits of the 95% confidence interval are then calculated by $L_{\mathrm{L}} = \widehat{x}^{*} - 1.96\sqrt{\sigma^2(\widehat{x}^{*})}$ and $L_{\mathrm{U}} = \widehat{x}^{*} + 1.96\sqrt{\sigma^2(\widehat{x}^{*})}$.

Table 2 The effect of the model proposed and heteroscedasticity on calibration precision limits and interval estimates of the unknown concentration of hexachlorbenzene in water for 24 points of calibration data [x (micrograms per liter), y (peak height in millivolts)] in the multiplicative model of measurement error

Characteristic	Step 1	Step 2	Step 3
Parameters of the calibration model used $y=\beta_0+\beta_1 x$ and $(+\beta_2 x^2)$ (in parentheses, the standard deviation)			
	Straight line, OLS	Spline, OLS	Spline, IRWLS
Intercept $b_0(s_0$, (mV)	38.4 (7.4)	8.3 (3.0)	2.7 (2.2)
$[L_L, L_U]$ for b_0 (mV)	[23.0, 53.8]	[2.0, 14.6]	[-1.8, 7.3]
Slope $b_1(s_1)$	18.6 (0.9)	32.4 (1.0)	36.1 (1.1)
$[L_L, L_U]$ for b_1	[16.8, 20.4]	[30.4, 34.4]	[33.9, 38.4]
Quadratic term $b_2(s_2)$	–	-0.73 (0.05)	-1.04 (0.10)
$[L_L, L_U]$ for b_2	[–, –]	[-0.83, -0.63]	[-1.26, -0.83]
Critical level			
y_C (mV), x_C (μg/l)	59.8, 1.15	17.7, 0.29	8.73, 0.17
Detection limit			
y_D (mV), x_D (μg/l)	77.7, 2.11	25.8, 0.55	14.0, 0.32
Quantification limit			
y_Q (mV), x_Q (μg/l)	94.6, 3.02	33.5, 0.79	19.6, 0.46
Unknown concentration			
$\bar{x}^*$ (μg/l) for $\bar{y}^* = 250.0$ mV	11.37	9.49	9.39
$[L_L, L_U]$ for $\bar{x}^*$ (μg/l)	[10.36, 12.60]	[9.00, 10.00]	[8.92, 10.04]
Regression diagnostics for a fitness test of the calibration graph			
Fisher–Snedecor F–test, F_R vs. $F_{1-\alpha}(m-1)(n-m)$	438.9 vs. 4.30	2511.2 vs. 3.47	7929.8 vs.3.47
Determination coefficient R^2	0.9523	0.9958	0.9987
Mean error of prediction	792.2	81.6	100.9
Akaike information criterion	158.2	101.6	74.1
Residual standard deviation $s(e)$ (mV)	25.9	7.8	4.4
Mean of absolute values of residuals e (mV)	20.8	6.2	6.1
Homoscedasticity of errors	Accepted	Accepted	Rejected
Trend in residuals	Accepted	Rejected	Rejected
Normality of random errors distribution	Accepted	Rejected	Rejected
Conclusion: calibration results	False	True	True

The signal of unknown sample is y^*=250 mV. Step 1—calibration straight line fitted data using OLS; step 2—calibration curve fitted data using the quadratic spline and OLS; step 3—calibration curve fitted data using the weighted quadratic spline and the iterative method of reweighted least squares (*IRWLS*) (Calibration in S-Plus)

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, ..., n, are available another set of knots are determined t_j, j=1, ..., 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 regression model $g_j(x)$. The quality of the approximation here is dependent on the number and the location of the individual knots t_j, a form of the function $g_j(x)$ and on the class C_m from which the calibration model $g_j(x)$ comes. A special type of piecewise polynomial functions is known as *splines* [1, 12].

The precision limits of calibration

Chemists are concerned with two types of limits when evaluating data quality. The first is a detection limit, used to decide whether or not an analyte is present; the second is a quantification limit, used to decide whether or not the concentration of an analyte can be reliably determined. Three different assessment criteria introduced by Curie [16–24] are used in this paper however, i.e., (1) the critical level, (2) the detection level and (3) the quantification level (previously called the determination level). The notation, assumptions and derivation have been given previously [1, 12]. As shown in Fig. 5, the *critical level* y_C (or the blank measurement) is the assay response above which an observed response is reliably recognized as detected, i.e., the response at which one may decide whether or not the result of an analysis indicates the presence of residue. The *detection limit* y_D is the actual net response which may a priori be expected to lead to detection. The *quantification limit* y_Q is the level at which measurement precision will be satisfactory for quantitative determination, i.e., a result which is satisfactorily close to the limiting expected value. Also defined are the analyte quantities x_C, x_D and x_Q, corresponding to y_C, y_D and y_Q, respectively, through the

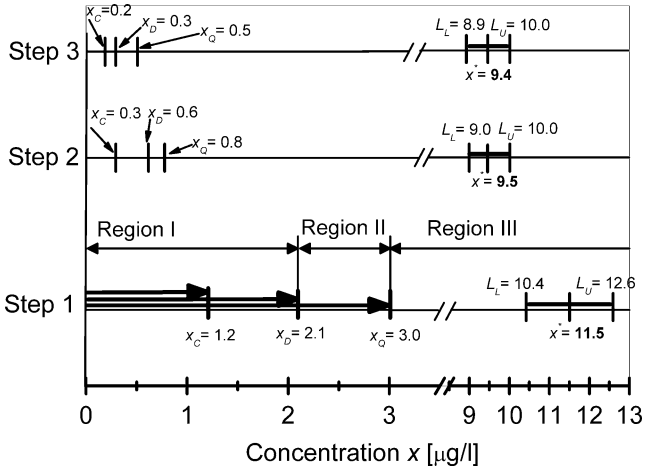

Fig. 4 The three principal analytical regions of calibration precision limits and the resulting point estimate $\hat{x}^*$ and interval estimates [L_L, L_U] of the unknown concentration with dependence on regression triplet analysis for hexachlorbenzene in water and the data in Table 2, where *region I* is the region of unreliable detection, *region II* is the region of qualitative estimation and *region III* is the region of quantitative estimation of unknown concentration. Calibration precision limits x_C, x_D and x_Q, and $\hat{x}^*$ with [L_L, L_U] are estimated. In step 1 the original data with all outliers are fitted with the straight-line model using OLS, in step 2 the same data are fitted but with the use of the quadratic spline function model with OLS used, and in step 3 the weighted quadratic spline function model and the IRWLS method are used

calibration curve. All these definitions may be simply used to calculate y_D and y_Q for nonlinear calibration models, and also for data for which the variance of measurement is not constant (heteroscedasticity) [1, 25]. Generally, it is valid that $y_C \leq y_D \leq y_Q$.

Basic assumptions of classical OLS regression

A source of problems in an OLS application may be found in the 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 "classical" OLS regression 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 [12, 26, 27]. 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 physicochemical meaning.
2. *Linearity*. The regression model is linear in the parameters, and an additive model of the measurement errors is valid, $y=X\beta+\varepsilon$. If a linear relation does not exist initially, data can often be transformed to obtain linearity. Linearity is often checked by using either the product–moment correlation coefficient, r, or the coefficient of determination, R^2. R^2 gives a measure of the portion of the total variability in a data set that is explained by a particular calibration model.
3. *Multicollinearity*. The matrix of nonrandom controllable values of the regressors X has a column rank equal to m. This means that the all pairs $\boldsymbol{x}_j$, $\boldsymbol{x}_k$ are not collinear vectors.
4. *Random errors*. The independent variable x either must be free of error or its level of error must be insignificant compared with the level of error in the dependent variable y. The mean value of the random errors in y denoted here as ε_i should be zero; $E(\varepsilon_i)=0$. This is automatically valid for all regression-type models containing an intercept. For models without an intercept the zero mean of errors has to be tested.
5. *Homoscedasticity*. The random errors ε_i in the dependent variable y have constant and finite variance, $E\left(\varepsilon_i^2\right) = \sigma^2$. 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 in the dependent variable y are uncorrelated, i.e., $\mathrm{cov}(\varepsilon_i, \varepsilon_i)= E(\varepsilon_i, \varepsilon_i)=0$. This corresponds to independence of the measured quantities $\boldsymbol{y}$.
7. *Normally distributed errors*. The random errors ε_i have a normal distribution $N(0, \sigma^2)$ with a mean of 0 and a variance of σ^2. The vector $\boldsymbol{y}$ 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.

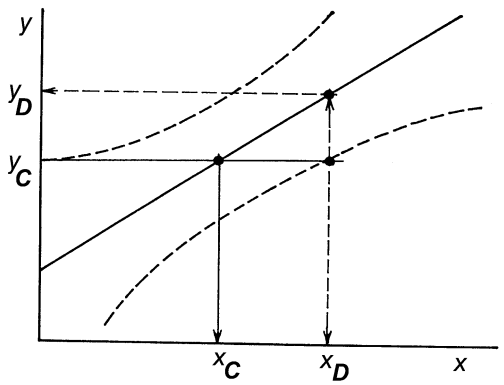

Fig. 5 Calibration design of the critical level in the response y_C and the concentration x_C units and the detection limit y_D and corresponding concentration x_D units. It includes the calibration straight line and both confidence bands

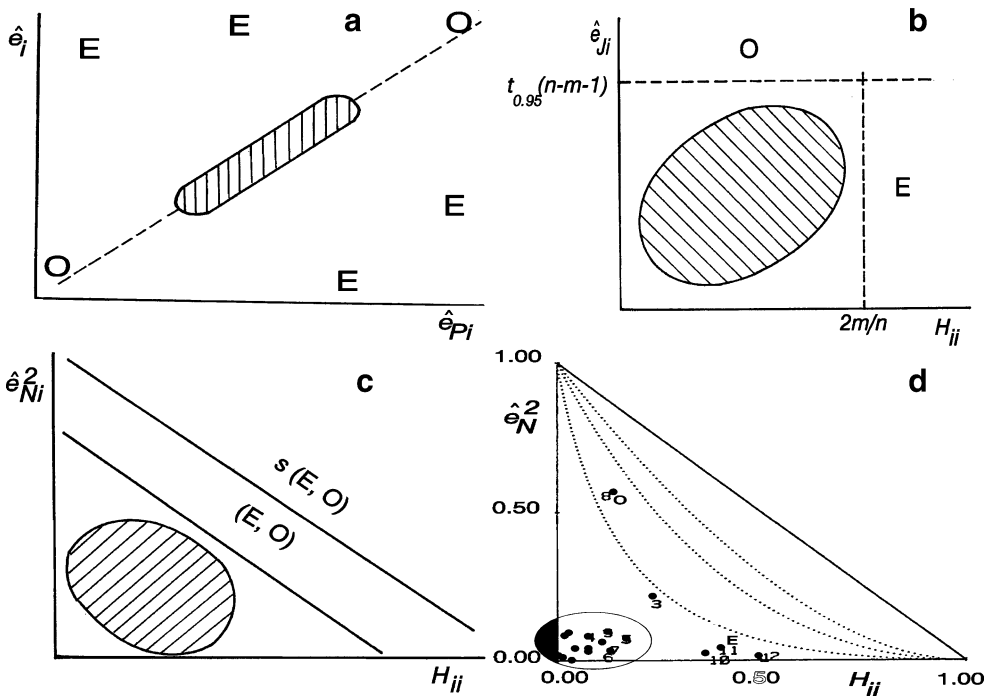

Fig. 6 Diagnostics indicating influential points are based on plots of various residuals and hat matrix elements: **a** graph of predicted residuals; **b** Williams graph; **c** Pregibon graph; **d** Gray's L-R graph. *E* a leverage point, *O* an outlier. The *filled random pattern* denotes that all assumptions of OLS are fulfilled

Examination of the data quality

Examination of data quality involves detection of the *influential point* (i.e., outliers and high-leverage points) which cause many problems in regression analysis by shifting the parameter estimates or increasing the variance of the parameters; a survey was provided in previous work [26]. A calibration point may be an outlier or a potentially influential point because of errors how the study was conducted (instrument malfunction, recording or data entry errors) or because the data point is from a different population. 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 nonnormality and heteroscedasticity [26, 27]. For the identification of influential points, i.e., outliers and high-leverage points, various types of residuals are used (Figure 6):

1. The *graph of predicted residuals* has on the *x*-axis the predicted residuals $\widehat{e}_{P,i}$ and on the *y*-axis the ordinary residuals $\widehat{e}_i$. The high-leverage points are easily detected by their locations, as they lie outside the line $y=x$, and are located quite far from this line. The outliers are located on the line $y=x$, but far from its central pattern.
2. The *Williams graph* [28] has on the *x*-axis the diagonal elements H_{ii} and on the *y*-axis the jackknife residuals $\widehat{e}_{J,i}$. Two boundary lines are drawn, the first for outliers, $y=t_{0.95}$ $(n - m - 1)$ and the second for high-leverage points, $x=2$ m/n. Note that $t_{0.95}(n - m - 1)$ is the 95% quantile of the Student distribution with $(n - m - 1)$ degrees of freedom.
3. The *Pregibon graph* has on the *x*-axis the diagonal elements H_{ii} and on the *y*-axis the square of normalized residuals $\widehat{e}^2_{N,i}$. Since the expression $E\left(H_{ii}+\widehat{e}^2_{N,i}\right) = (m+1)/n$ is valid for this graph, two different constraining lines can be drawn, $y=-x+2$ $(m+1)/n$ and $y=-x+3(m+1)/n$. To distinguish among influential points the following rules are used: (1) a point is *strongly influential* if it is located above the upper line; (2) a point is *influential* if it is located between the two lines. The influential point can be either an outlier or a high-leverage point.
4. *Gray's L-R graph* [29] has on the *x*-axis the diagonal elements H_{ii} and on the *y*-axis the squared normalized residuals $\widehat{e}^2_{N,i}$. All the points will lie under the hypotenuse of a triangle with a 90° angle in the origin of the two axes and the hypotenuse defined by the limiting equality $H_{ii}+\widehat{e}^2_{N,i}=1$. In Gray's L-R graph, contours of the same critical influence are plotted, and the locations of individual points are compared with them. It may be determined that the contours are hyperbolic as described by $y=(2x-x^2-1)/(x(1-K)-1)$, where $K=n(n-m-1)/(c^2m)$ and c is a constant. For $c=2$, the constant K corresponds to the limit $2/\sqrt{m/n}$. The constant c is usually equal to 2, 4 or 8.

Examination of the model quality

Examination of calibration model quality can be considered directly from the calibration scatter plot of y versus x. Individual parameters are tested for significance using the Student t test. The *Fisher–Snedecor F test of significance of the calibration model proposed* is based on the testing criterion

$$F_{\mathrm{R}} = \widehat{R}^2(n-m) \Big/ \left[\left(1-\widehat{R}^2\right)(m-1)\right]$$

which has the Fisher–Snedecor distribution with (m - 1) and (n - m) degrees of freedom, $\widehat{R}^2$ is 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 H0: R^2=0 may be tested and concerns a test of significance of all regression parameters β. Other statistical characteristics calculated are the MEP and the AIC, while $s(e)$ and $\overline{e}$ examining the linearity of the proposed model also can be used as resolution criteria among various models; definitions of these characteristics may be found in previous works [1, 12, 27].

Examination of the regression method used

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

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 ith point is an exponential function of the variable $x_i\ \beta$ of the type $\sigma_i^2 = \sigma^2 \exp(\lambda\ x_i\ \beta)$, where x_i is the ith concentration. The test for homoscedasticity is carried out by checking the null hypothesis H0: λ=0. Cook and Weisberg [30] introduced the test criterion

$$S_{\mathrm{f}} = \frac{\left(\sum_{i=1}^{n}(\widehat{y}_i-\overline{y})\widehat{e}_i^2\right)^2}{2\widehat{\sigma}^4\sum_{i=1}^{n}(\widehat{y}_i-\overline{y})^2},$$

where $\overline{y} = \left(\sum_{i=1}^{n}\widehat{y}_i\right) \Big/ n$. When the null hypothesis is valid the test statistics have approximately a $\chi^2(1)$ distribution with one degree of freedom.

The test of the normality of errors

First, the normality of errors may be simply examined by a rankit quantile–quantile plot containing the order statistics of classical residuals $\widehat{e}_{(i)}$ depending on the quantile of the normalized normal distribution u_{Pi} for P_i=i/(n+1), i=1, ..., n. Since small samples exhibit a supernormality effect, independent recursive residuals $\widehat{e}_{\mathrm{R}(i)}$ are used instead of classical ones, because this effect then does not exist [12]. Second, the most convenient test for linear models seems to be the Jarque–Berra test [31], which is based on the criteria of the residual skewness and the residual kurtosis. When $L(\widehat{e}) > \chi^2_{0.95}$= 5.99, the null hypothesis H0 about the error normality is rejected.

Violation of some assumptions for the OLS method

The effects of deviations from these basic assumptions for the OLS method, and methods that correct for these effects leading to a more accurate regression model, are noted:

1. When heteroscedasticity is found in the data, the *weighted least squares method* is used [12].
2. When autocorrelation is found in the data, the *generalized least squares method* is used [12].
3. When some restrictions apply to the parameters, the *conditioned least squares method* is used [12].
4. When multicollinearity is found in the data or when polynomial calibration models are used, the *principal component method* is used [12, 26].
5. When all variables are subject to random errors, the *extended least squares method* is used [12].
6. 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 IRWLS method is used [12]: when the distribution of the errors in the dependent variable y is not normal, the parameter estimates obtained by OLS are not the best possible estimates. In such cases, instead of the least-squares criterion some other *robust* criterion can be used, less sensitive to violation of the assumption about error distribution, and also not sensitive to influential points.

Acknowledgements The financial support of the IGA Grant Agency (grant no. NR9055-4/2006) and of the Czech Ministry of Education (grant no. MSM0021627502) is gratefully acknowledged.

References

1. Meloun M, Militký J, Kupka K, Brereton RG (2002) Talanta 57:721–740
2. Massart DL, Kaufman L (1986) Anal Chim Acta 187:171–179
3. Lee JJ (1991) Biometrics 47:1573–1580

4. Ortiz MC, Arcos J, Juarros JV, López-Palacios J, Sarabia LA (1993) Anal Chem 65:678–682
5. Walczak B (1995) Chemom Intell Lab Syst 28:259–272
6. Zorn ME, Gibbons RD, Sonzogni WC (1997) Anal Chem 69:3069–3075
7. Zorn ME, Gibbons RD, Sonzogni WC (1999) Environ Sci Technol 33:2291–2295
8. Boque R, Faber NKM, Rius FX (2000) Anal Chim Acta 423:41–49
9. International Organization for Standardization (2000) International standards: capability of detection: part 2. Methodology in the linear calibration case, ISO 11842-2 (2000) E. International Organization for Standardization, Geneva
10. Kuselman I (2001) Accredit Qual Assur 6:54
11. Kuselman I, Goldfeld I, Anisimov B (2001) Accredit Qual Assur 6:61–65
12. Meloun M, Militký J, Forina M (1994) Chemometrics for analytical chemistry, vol 2. PC-aided regression and related methods. Horwood, Chichester
13. Gort SM, Hoogerbrugge R (1995) Chemom Intell Lab Syst 28:193–199
14. Horowitz F (1982) Anal Chem 54:67A–76A
15. Naszodi LJ (1978) Technometrics 20:201
16. Currie LA, Svehla G (1994) Pure Appl Chem 66:595–608
17. Currie LA, Horwitz W (1994) Analysis 22:M24–M26
18. Currie LA (1995) Pure Appl Chem 67:1699–1723
19. Currie LA (1997) Chemom Intell Lab Syst 37:151–181
20. Danzer K, Currie LA (1998) Pure Appl Chem 70:993–1014
21. Currie LA (1988) ACS Symp Ser 361:1–62
22. Currie LA (1999) Anal Chim Acta 391:103–103
23. Currie LA (1999) Anal Chim Acta 391:105–126
24. Currie LA (1999) Anal Chim Acta 391:127–134
25. Oppenhelmer L, Capizzi TP, Weppelman RM, Mehta H (1983) Anal Chem 55:638–643
26. Meloun M, Militký J, Hill M, Brereton RG (2002) Analyst 127:433–450
27. Meloun M, Militký J (2001) Anal Chim Acta 439:169–191
28. Williams DX (1973) Appl Stat 22:407–408
29. Gray JB (1985) Am Stat Assoc 102–107
30. Cook RD, Weisberg S (1983) Biometrika 70:1
31. Jarque CM, Berra AK (1987) Int Stat Rev 55:163
32. Trilobyte Statistical Software (2001) ADSTAT. Trilobyte Statistical Software, Pardubice
33. Hrivnák J (1985) Anal Chem 57:2159–2160