Chem. Anal. (Warsaw), **52**, 715 (2007)

Non-linear Calibration Leads to Improved Correspondence Between Uncertainties

by Jens E.T. Andersen

Department of Chemistry, Technical University of Denmark, Kemitorvet building 207, DK-2800 Kgs. Lyngby, Denmark

Keywords: Quality assurance; Precision; Accuracy; Uncertainty; Non-linear calibration

Although calibrations are routine procedures of instrumental analysis and quality assurance, the working curve is rarely applied to the determination of the uncertainty budget, most likely owing to the difficulties associated with the calculation of uncertainties. The present work provides an investigation of an uncomplicated expression of the non-linear working curve that is well suited for an assessment of predicted uncertainties. At small concentrations, the working curve reduces to a straight line that corresponds to the conventional calibration line. If no interferences were disturbing the analysis, the calculation of uncertainties of calibrations must correspond to the uncertainties of unknowns that were determined by many repetitions. Thus, by introducing an average value of the law-of-propagation of errors (LPE) and observing the conditions of the central limit theorem, an excellent correspondence was obtained between predicted uncertainties and measured uncertainties. In order to validate the method, experiments were applied of flame atomic absorption spectrometry (FAAS) for the analysis of Co and Pt, and experiments of electrothermal atomic absorption spectrometry (ETAAS) for the analysis of Fe. A ten fold extension of the calibration range was identified, as represented by a lower limit of analysis (LLA) and an upper limit of analysis (ULA), which were defined by the properties of the detection system of the apparatus. It was thus found that the uncertainty of the detector dominated the contributions to the uncertainty budget, and it was proposed that a full analysis of the instrument ought to be performed before determination of every single analyte. Following this investigation, the homoscedasticy or heteroscedasticy may be identified by residuals of calibration.

Chociaż kalibracja jest powszechną procedurą w analizie instrumentalnej rzadko się jej używa do wyznaczania budżetu niepewności ze względu na trudności napotykane przy wyznaczaniu niepewnosci z krzywej kalibrowania. W pracy zaproponowano badania prostego opisu nieliniowej krzywej roboczej, który mógłby być użyteczny do oszacowania niepewności. W przypadku małych stężeń krzywe robocze przybierają zwykle kształt klasycznej krzywej kalibrowania. Jeśli w analizie nie występują interferencje, obliczanie niepewności kalibracji musi odpowiadać niepewności zmiennych wyznaczonych metodą wielu powtórzeń. Rzeczywiście, wprowadzając wartość średnią z prawa propagacji błędów i obserwując warunki wynikające z centralnego twierdzenia granicznego otrzymano doskonałą relację między przewidywanymi i mierzonymi niepewnościami. Do zwalidowania metody wykorzystano pomiary stężeń Co i Pt wykonane płomieniową, atomową spektrometrią absorpcyjną oraz pomiary stężeń Fe wykonane elektrotermiczną, atomową spektrometrią absorpcyjną. Stwierdzono 10-krotne rozszerzenie zakresu kalibracji, który jest wyznaczony przez punkty dolnej i górnej granicy analizy. Punkty te są określone przez system detekcji aparatury. Stwierdzono więc, że niepewność detektora dominowała w budżecie niepewności i zaproponowano, że pełna analiza działania instrumentu powinna poprzedzić oznaczanie próbek analitów. W wyniku przeprowadzonych badań krzywych kalibrowania można ustalić czy mamy do czynienia z więcej niż jednym rozkładem i czy wariancje tych rozkładów różnią się istotnie czy też nie.

The calibration of instruments may be considered as stable and allows consistent determinations at high precision, when the measurements are performed within a limited time frame of analysis. However, the apparently straightforward procedure of calibration and determination of an unknown, amplifies in complexity by the introduction of uncertainties and methods of deriving uncertainties [1–6]. In quality assurance, much effort has recently been devoted to evaluation of calibration protocols and derivation of uncertainties [7], comprising linear calibration [1, 3–4, 8], multivariate calibrations [2, 5–6], Monte Carlo Simulations [9], chemometrics [10], artifical-neural networks [11] and non-linear calibration [12–13]. Well established methods of calibration and proficiency testing are available in the form of ISO standards [14–17] and IUPAC recommendations [4] but they are mainly concerned with the mathematical and statistical procedures and only to a minor degree they focus on details of the instrument. Barnett [18], Winefordner *et al.* [19] and Kristensen *et al.* [20] identified physical parameters of the instrument, which each contributed to the total uncertainty of the measurement, where, *e.g.*, the uncertainty on the sensitivity was the main contributor to the uncertainty in the determination of lead in blood by ET–AAS [20]. The alleged difference between estimation of uncertainties in chemistry [21–22], as opposed to the methods applied to physics [23] may thus be related to a, so far, incomplete picture of the functionality of instruments, and the associated contributions to the overall uncertainty.

In fact, any working curve exhibits a non-linear relation between response and concentration [12–13, 24–25]. Thus, the linear range of calibration may be described as a Taylor expansion to first order of the response function, which approaches well the data

of the working curve at low concentrations (Fig. 1). The non-linearity of the response function [13] originates from physical sources such as self absorption [9], as the detectors relaxation time [19, 28] or as the saturation of the detectors signal [18–19]. Barnett [18] observed that photometric detectors in *e.g.* AAS exhibited a characteristic cut-off value of the response, which results in a constant response that becomes independent of concentration at high levels (Fig. 1). The uncertainties of calibration must be included in the uncertainty budget [26–27], and it was recently advised that analysts should accumulate and examine their data in a more global view that includes both instrument and procedure [13].

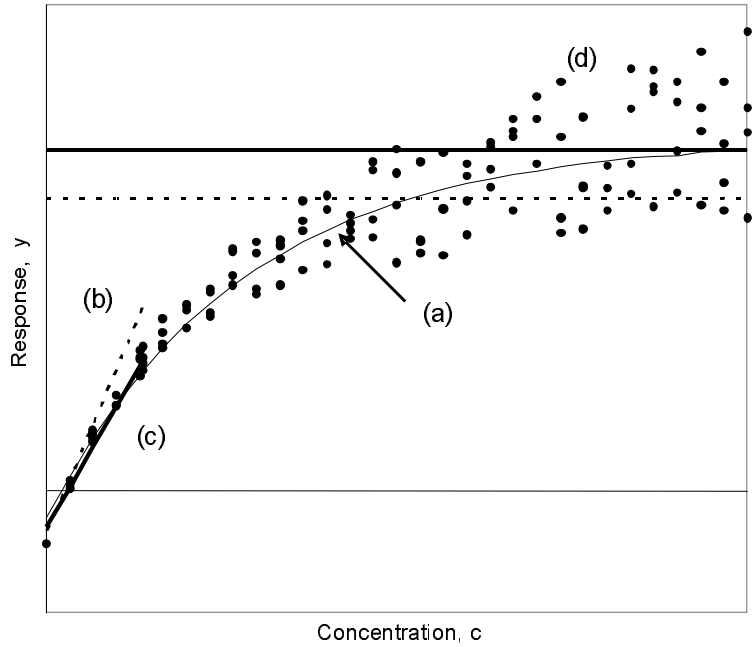

Figure 1. Schematic diagram of tentative calibration results. The general calibration curve (a) may be approximated by a tangential straight line (b) that prevails within the concentration range of eq. 4. In calibration experiments, the straight line (c) is most frequently applied by least-squares fitting within the stipulated limit of concentrations. At higher concentrations, the data (d) approaches a limit (Horizontal thick line) where concentrations renders undetermined. Some instruments are equipped with a cut-off functionality (Horizontal broken line) that results in lower concentrations of maximum response

Although methods are available for the calculations of uncertainties of non-linear responses [29], it is generally complicated to obtain an algebraic expression for the desired uncertainty that exhibits the dependence on concentration. It has been proposed that handling of uncertainties of physical measurements [27] differ from those of chemical

measurements, owing to a number of practical aspects [24, 30]. Once the uncertainty of calibration were estimated, henseforth denoted as the predicted uncertainty, it may be compared to the uncertainty obtained by repetitive determinations of an unknown [14, 31].

In order to validate the method and increase reliability, a large number of experiments were evaluated. This approach ensures that most conclusive results were endorsed by the law-of-great numbers. Average values of concentrations ($\bar{c}$) were derived on the basis of the working curve and they were determined at a standard deviation of common practice [26–27]:

$$s_c = \sqrt{\frac{\sum_i \left(c_i - \bar{c}\right)^2}{N-1}} \tag{1}$$

In this equation (eq. 1), the number of degrees of freedom (N–1) could be derived from different positions at the working curve, that is, the averages concentration as well as the standard deviation were determined by using different dilutions [26–27].

A comprehensible value of the STDEV of an algebraic function may be obtained by using the LPE. However, the LPE frequently provides STDEV's that are vastly overestimated, which may have had a diminishing effect on the dissemination of the method to analytical chemistry. A slight reduction of the magnitude of the uncertainty was obtained by subtraction of terms of covariance [32–34] in the expression of the LPE but the numerical calculations involved are cumbersome and they are not yet standard methods of spreadsheets. The reason for the overestimate is related to the 'adding-up' property of the LPE, which inevitably provides low uncertainties for uncomplicated function and high uncertainties for complicated working curves of multiple parameters. Because the event is unlikely where all contributions to the LPE add up simultaneously, a more comprehensible edition of the LPE may be proposed by comparing to the expression of average deviations (eq. 1). As an example, with a STDEV of four terms, each term will either provide a positive or a negative contribution to the final uncertainty, which yields a total of 16 combinations. Since all terms contribute either positively or negatively by the same probability, it is reasonable also to consider the average value by the LPE. Thus, it is proposed that the STDEV estimated by the LPE is averaged by the number of terms (N_t). Thus, the STDEV of a single determination of an unknown may be estimated by:

$$s_c = \sqrt{\frac{\sum_{i=1}^{N_t} \left(\frac{\partial c}{\partial x_i} \cdot s_{x_i} \right)^2}{N_t - 1}} \tag{2}$$

The application of this novel equation provides more realistic STDEV's of measurements, and it allows determination of a lower limit of analysis (LLA), an upper limit of analysis (ULA). The aim is thus to investigate the correspondence between the predicted STDEV of calibrations, as represented by eq. 2, and the STDEV of repetitive determinations of unknowns (eq. 1).

EXPERIMENTAL

Chemicals

The standards and samples of iron and of cobalt were prepared by dilution of stock solutions by 0.1 mol L^{-1} nitric acid (Merck, p.a.) in Millipore water (Resistance 18 MΩ). As samples, were applied certified reference materials (CRM's) Natural water NIST SRM 1640, Waste water EU–H–1 (1:50, EnviroMAT) and Drinking water EP–H–1 (1:100, EnviroMAT). NB! The solutions of EU–H–1 and EP–H–1 contains highly poisonous chemicals and should be handled with great caution.

Since none of the CRM's contained Pt, the solutions of EU–H–1 (diluted 1:50) were spiked (Merck, CertiPUR, 1.70341.0100) by this element to a concentration of 0.5 mg L^{-1}. In order to suppress the influence of interferences, the standards and samples of platinum were dissolved in a solution containing 0.2% (w/w) La (La(III) chloride, Sigma, p.a.) and 1% (w/w) HCl (Merck, p.a.). The standards of iron and of cobalt were prepared by dilution of standard solutions of iron(III) nitrate (Merck, CertiPUR, 1.19781.0500) and cobalt(II) nitrate (BDH Chemicals), respectively. The influence of the number of standards and the concentration of standards on the uncertainty was investigated by preparing two series of standards of iron for the determinations by ETAAS. High-precision determinations: 0, 10, 20, 30, 40 and 50 mg L^{-1}, and low-precision determinations: 0, 50, 100, 125, 150 and 200 mg L^{-1}. The determinations of platinum by FAAS were performed by using standards of 0, 10, 20, 30, 40, 50, 60 and 100 mg L^{-1}. Cobalt was determined by FAAS using a series of standards comprising 0, 0.5, 1.0, 2.0, 3.0, 3.5 and 7.0 mg L^{-1}.

Apparatus

The determinations of iron were performed by ETAAS (Perkin Elmer, AAS2100) using a wavelength of 248.3 nm, a 0.2 nm slit width, and a pre-treatment temperature and an atomization temperature of 1400°C and 2400°C, respectively.

In order to demonstrate the difference in uncertainties of two different elements, cobalt and platinum was investigated by FAAS using the same instrument (Perkin Elmer Analyst100). Both elements were analysed in an air-acetylene flame adjusted to a ratio in gas flow of 2:1. Cobalt was analysed at a wavelength of 240.7 nm (S&J Juniper & Co., England) and a slit width of 0.2 nm while Pt was analysed at a wavelength of 265.9 nm (S&J Juniper & Co., England) and a slit width of 0.7 nm. The instruments were operated according to standard conditions of analysis by setting the instrument at zero absorbance for the blanks before measurement.

Theory

In absorbance spectrophotometry, the signal is proportional to concentration, as described by the Beer-Lambert equation, and the linear range may be found by a lack-of-fit test [8]. At high concentra-

tions, deviations from the linear behaviour of the response were observed, however, and the corresponding working curve was frequently extended by the application of polynomials [12–13, 25]. A polynomial exhibits extremes that do not express correctly the physical reality. *e.g.*, the polynomial approaches in finity at high concentrations, which is not observed experimentally. A more realistic description of the experiments is obtained by applying the response function, y, based on the physical operation of a detector, as given by [19, 28]:

$$y = A \cdot (1 - \exp[B \cdot (c - c_0)]) \qquad B \leq 0 \tag{3}$$

This function (eq. 3) includes a maximum response (A) that corresponds to the signal at high concentrations where the signal becomes virtually independent of concentrations, as confirmed by experiment. The decay parameter (B) of equation 3 has the units of reciprocal concentration. The concentration (c_0) is a value that is obtained by fitting the function of equation 3 to the data of experiments and it provides a value for the intercept. At low concentration (c_{lin}), the working curve of equation 3 may be expanded to first order, which provides a straight line:

$$y \cong -A \cdot B \cdot c_{lin} + A \cdot B \cdot c_0 \qquad -\frac{B \cdot c_0 - 1 + e^{-1}}{B} < c_{lin} < \frac{B \cdot c_0 - 1 + e^{-1}}{B} \tag{4}$$

where the intercept concentration (c_0) may attain both negative and positive values. The slope and intercept are represented by $\alpha = -A \cdot B$ and $\beta = A \cdot B \cdot c_0$, respectively. Further, by subtraction of blanks or by applying the autozero functionality of the instrument, where the intercept of equation 4 vanishes, *i.e.* $A \cdot B \cdot c_0 \approx 0$, the characteristic mass (m) of ETAAS may be estimated as:

$$m \cong -\frac{0.0044 \cdot M \cdot V}{A \cdot B} \tag{5}$$

where 'M' is the molar mass of the analyte and 'V' is the sample volume. This equation (eq. 5) shows that the molar absorptivity in the present description may be interpreted as a property of the detector.

In order to extend the working curve as much as possible with due respect to the level of uncertainties, responses of equation 3 is investigated in terms of the LPE. The concentration of an unknown is calculated by isolating the concentrations of equation 3:

$$c = c_0 + \frac{1}{B} \cdot \ln\left(1 - \frac{y}{A}\right) \tag{6}$$

In this 'inverse representation' [4, 9], the response must be smaller than the maximum response value ($y < A$). However, experiments frequently provide responses that are larger than the maximum response value (A), owing to the STDEV and general spread of data. In such a case, the concentration remains 'undetermined' because the STDEV on concentration becomes too high. The investigation of equation 6 in terms of the LPE (eq. 2) results in a formula for the standard deviation on concentration, which depends on the response of an unknown and on the standard deviations of parameters, as given by $\overline{y}$, s_{c_0}, s_A, s_B and N, respectively:

$$s_c = \sqrt{\frac{s_{c_0}^2 + \left[\frac{s_B}{B^2} \cdot \ln\left(1 - \frac{\bar{y}}{A}\right)\right]^2 + \left[\frac{s_y}{B} \cdot \frac{1}{A - \bar{y}}\right]^2 + \left[\frac{s_A}{A \cdot B} \cdot \frac{\bar{y}}{A - \bar{y}}\right]^2}{N_t - 1}} \qquad \bar{y} < A \tag{7}$$

Since the number of independent parameters equals four, the STDEV was averaged by three, according to the proposition of equation 2. The response of the unknown is determined by a number of repetitions, which provides an average value of the response ($\bar{y}$) and the concomitant standard deviation (s_y). Thus, equation 7 provides a value for the STDEV on the corresponding concentration only at this single concentration. Insertion of equation 3 into the expression of equation 7, yields a STDEV that depends on concentration, as required:

$$s_c = \sqrt{\frac{s_{c_0}^2 + \left[\left(\bar{c} - c_0\right) \cdot \frac{s_B}{B}\right]^2 + \left[\frac{s_y}{A \cdot B} \cdot \exp\left(B \cdot \left(c_0 - \bar{c}\right)\right)\right]^2 + \left[\frac{s_A}{A \cdot B} \cdot \left(\exp\left(B \cdot \left(c_0 - \bar{c}\right)\right) - 1\right)\right]^2}{3}} \tag{8}$$

The STDEV of the unknown (s_y) also depends on concentration, and the equation (eq. 8) provides only a value at a single concentration. In principle, the STDEV of an unknown might resemble that of a standard at a given concentration but it may be expected that it is always larger, owing to matrix effects or to sample-introduction accessories. The difference in uncertainty structure of standards and of unknowns is therefore an important issue of consideration, as proposed by Tellinghuisen [12]. However, by assuming different values of s_y, equation 8 may be applied to investigate extremes of the STDEV on concentration. A minimum value of s_c is obtained by assuming a value of zero ($s_y \cong 0$), while the maximum value of the STDEV is found when $s_y = s_A$. Accordingly, the STDEV(1s) on concentration may be depicted as a function of concentration at these two extremes, which provides a general overview of the precision of the method. By assuming that the determination of the concentration of a single unknown has the same uncertainty structure as that of the calibration data, the expression of equation 8 corresponds, at least qualitatively, to the uncertainty function of polynomials [12].

Concentration range of calibration

A calibration range wider than the one defined by the linear range (eq. 4) prevails by applying the non-linear function of equation 3, and the extension may be calculated by considering the concentrations at the extremes. The LLA is an important figure of merit of the analytical protocol but in the present method the upper limit of analysis (ULA) should also be considered, owing to the limitations imposed by equation 6. The ULA could arbitrarily be determined by the concentration at which the response equals the value of the maximum response (A) less double the value of the STDEV proper. According to equation 6, the ULA may then be defined as:

$$c_{max} = \text{ULA} = c_0 + \frac{1}{B} \cdot \ln\left(\frac{2 \cdot s_A}{A}\right) \tag{9}$$

At the lower extreme of concentrations, the LLA may be defined by the STDEV at the intercept concentration (c_0, eq. 6):

$$\bar{c} = c_0 \qquad \Rightarrow \qquad s_c = LLA = \sqrt{\frac{1}{3} \cdot \left(s_{c_0}^2 + \left[\frac{s_{y_0}}{A \cdot B} \right]^2 \right)} \tag{10}$$

Accordingly, the LLA may be expressed in terms of the STDEV of the intercept (s_{c_0}) and the STDEV of the signal blank (s_{y_0}), and it becomes independent of s_A and s_B. However, the LLA depends on the reciprocal response ($1/A$) and on the characteristic concentration ($c_m = -1/B$). This result implies that an instrument of a high maximum response and a high sensitivity most likely produces the minimum LLA. As an example, the difference in LLA's of AAS instruments and ICP–MS instruments may be explained in terms of differences in A and B.

Experimental determination of the fitting parameters

An attempt to apply the method of least-squares fitting of the response function (eq. 3) to the data of experiments may frequently be performed in vain because it yields an infinite number of values for the parameters (Fig. 1). The problem of determining the fitting parameters can be solved by first considering the response value at high concentration, that is, the A-value is determined at concentrations where the signal of the detector saturates. However, as in the case of Co determined by FAAS [18], the signal may saturate at a response value that resides significantly below that of the cut-off response. Many instruments of analytical chemistry are manufactured with a built-in cut-off functionality, which may be difficult to identify, owing to the statistical spread of data at high concentrations. However, if the extrapolated value corresponds to the measured value of A at high concentrations, then the maximum response value were situated below the characteristic cut-off value (Fig. 1). Thus, once the A-value is determined, the two independent parameters (c_0 and B) of the response function (eq. 3) may be calculated by MSExcel Solver [35].

Interpretation of the parameter 's_y'

The uncertainty structure of standards may be different from that of the samples [13] but in the present analysis, it was assumed that the uncertainty structure deviated only as a result of the presence of interferences or as a result of the application of sample introduction accessory, such as flow-injection analysis (FIA). Matrix interferences could be identified by either a concentration that differs significantly from the certified value or by an elevated STDEV of the certified value. It was recognised (above) that the LLA depends on the value of s_y at the intercept concentration (c_0) where the STDEV's on A and B are unimportant. All standard deviations (eq. 8), except for the value of s_y, are provided as a result of the fitting of the response function (eq. 3) to the calibration data. The s_y – value is determined in terms of the appropriate response of the method, such as absorbance, counts per second or current density. It refers to the determination of the unknown where at least two repetitions were required, as to estimate s_y by equation 1. However, application of eq. 8 shows that a minimum value for the standard deviation of the unknown is obtained when s_y equals zero. This means that the STDEV of a single determination of an unknown may be provided by assigning to s_y a zero value. Thus, a STDEV may be assigned to a single unknown, owing to the uncertainties related to the uncertainties of calibration. Of course, a more realistic and correct value of s_y may be obtained by performing a large number of repetitions ($N > 30$) of the same unknown, which is generally recommended because it ensures repeatability and it diminishes the risk of capital errors. Application of standard-addition methods in matrices is thus expected to increase the value of s_y, owing to the presence of potential interferences. Thus, multiple repetitions ensures a reliable estimate of s_y, and the minimum value equals that of a corresponding standard, that is, s_y (unknown) $\geq s_y$ (standard).

RESULTS AND DISCUSSION

The investigation of functions and data by statistical methods were performed by observing the conditions of the law-of-great numbers, which imposes at least 30 data to the analysis [21]. A schematic diagram of a tentative calibration experiment is shown in Figure 1, where the threshold value of the instrument defines an upper limit to the response. If the maximum response of an analyte were positioned below the threshold value, then the response would follow the curve of equation 3 (Fig. 1, solid line), also at high concentrations. Moreover, as in UV–VIS spectrometry and in AAS, the response is limited by the cut-off at the maximum response value (Fig. 1, broken line). A calibration line may be constructed by using the data within the linear range of responses (Figs. 1(b)–1(c)) that was estimated by the tangential approximation (eq. 4, Fig. 1(b)) to the working curve (eq. 3, Fig. 1(a)). This approach provides a rapid and reliable estimate of the calibration range of linear response, where the maximum concentration of linear response is given by the limits of eq. 4. That is, the calibration range defined by the limits of the tangential approximation (eq. 4) may also be applied as the calibration range of least-squares regression (eq. 4, fig. 1(c)). Data above the limiting response of calibration (Fig. 1(d) and thick horizontal line, respectively) may be considered as 'undetermined' because the relative uncertainty of concentrations exceeds 50%. Occasionally, the data above the limiting response is hidden behind the instrument cut-off value [18] (Fig. 1, horizontal broken line).

The distribution of data portrayed in Fig. 1 may be obtained by depicting a large number of measurements in the same diagram. However, such an extensive investigation of the uncertainties of the working curve is rarely affordable in experiments of analytical chemistry [21]. The application of data too low in number to the analysis may result in an incomplete description of the uncertainty that belongs to the uncertainty budget [26–27]. That is, if too low a number of data were applied to the calibration procedure, the uncertainty might accidentally be underestimated in comparison to the true uncertainty of an infinite number of measurements. Further, an underestimated uncertainty poses a risk of erroneously identifying outliers. Accordingly, in order to assess the combined performance of the method and analyte, a careful and extensive analysis of uncertainties is needed [8].

In order to sought out the details of uncertainties of a given method using different analytes, the working curves of Co and Pt were investigated by AAS using the same instrument, as shown in Figures 2(a)–2(b). The determinations of Pt (N = 48) clearly exhibited a large spread of data around the working curve, as compared to those of the Co results (N = 325). These results (Figs. 2(a)–2(b)) demonstrate that the same instrument produced different levels of uncertainties, and they show that Co was determined at high precision while Pt was determined at low precision. In comparison to these measurements, the working curve of iron was established in experiments of ETAAS,

where 110 measurements were applied to the analysis (Fig. 2(c), N = 110). The general trend of an increasing uncertainty as a function of concentration was confirmed by all three experiments (Figs. 2(a)–2(c)), which is a common feature of working curves [3, 23, 26].

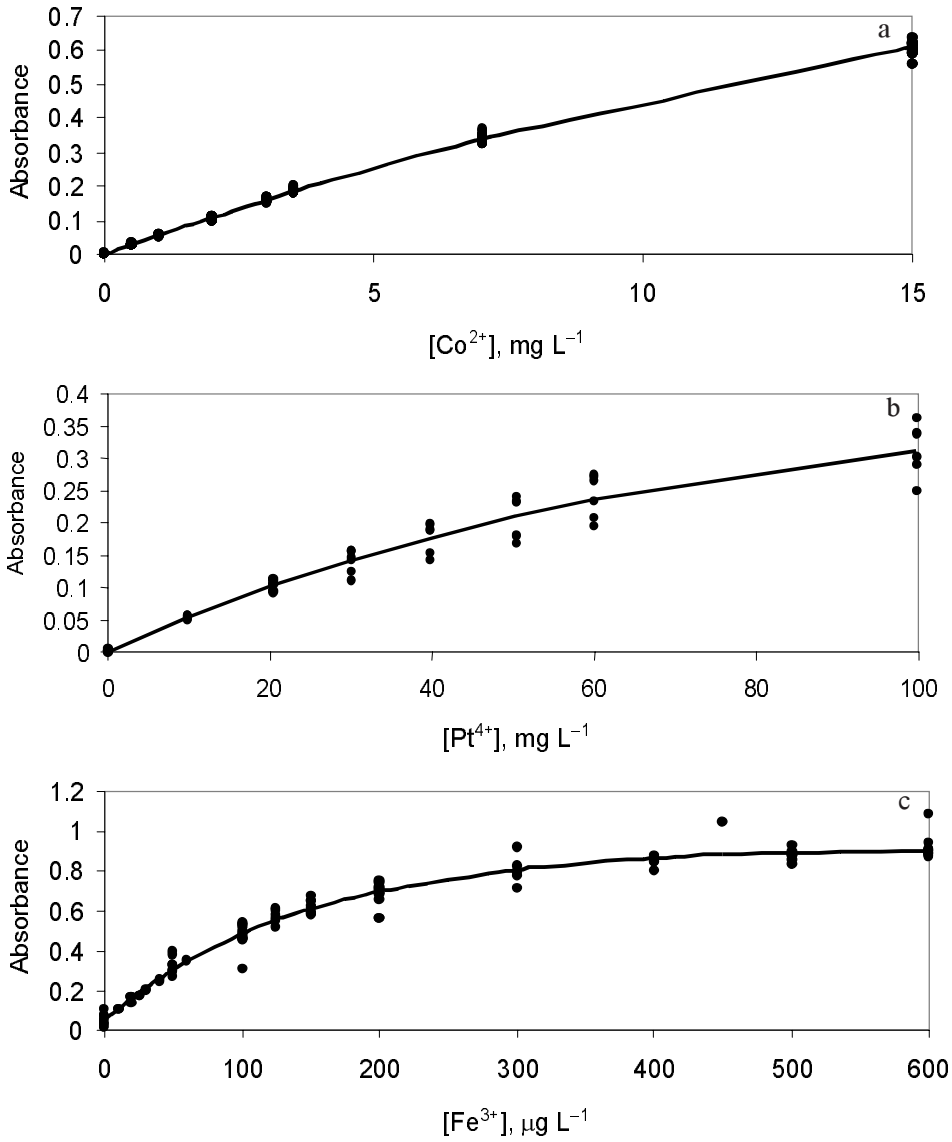

Figure 2. The pooled calibration data and corresponding working curves (eq. 3) of the FAAS determinations of Co (a), Pt (b) and ETAAS determinations of Fe (c). The data were obtained during many series of experiments with N(Co) = 324, N(Pt) = 47 and N(Fe) = 109

The uncertainty of s_A was estimated by application of the common formula of standard deviation (eq. 1) using responses at high concentrations. Thus, the two additional parameters c_0 and B, and the corresponding uncertainties s_{c_0} and s_B may be determined by fitting the curve of equation 3 to the data, as shown by the solid lines of Figures 2(a)–2(c).

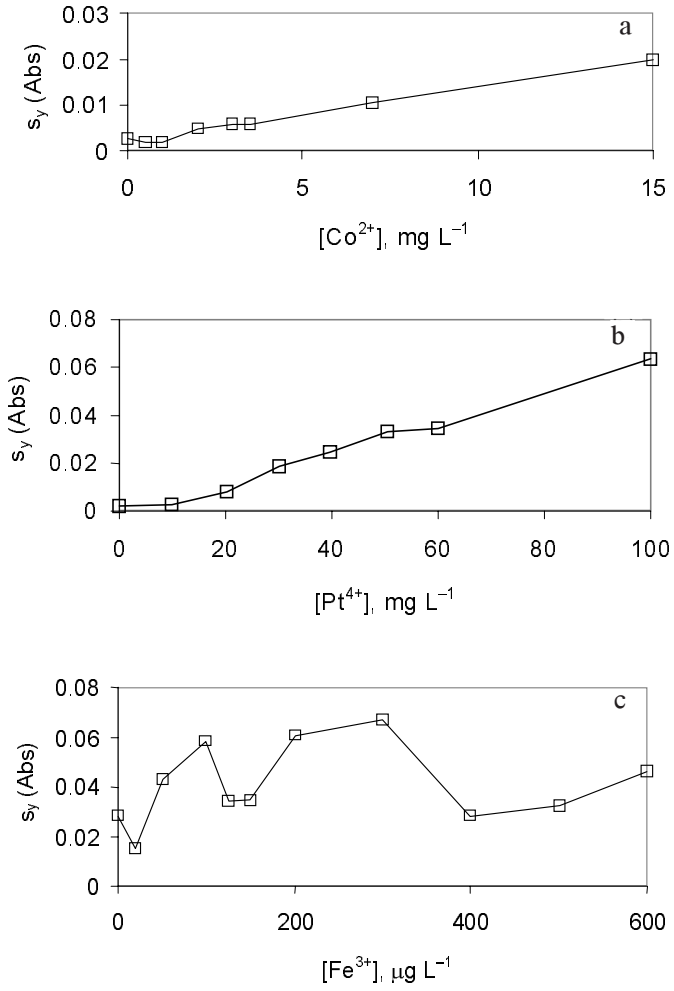

Figure 3. The absolute values of the standard deviations of unknowns (s_y) that exhibit the same uncertainty structures as those of the standards depicted as a function of concentrations. The uncertainties of Co (a) and the uncertainties of Pt (b) exhibits heteroscedasticy and the uncertainties of Fe (c) displays homoscedasticy, which demonstrates that these concepts are related to the method of determination. Note that the magnitude of s_y influences the total uncertainty mainly at high concentrations (eq. 8)

By using MSExcel Solver, correlations [1, 32–34] need not be considered because the parameters of slope and intercept are thus truly undependent [35]. The total uncertainty is given by equation 8 for the determination of an unknown but the value of s_y (Figs. 3(a)–3(c)) need be determined independently because it does not correspond to the uncertainty of Figures 2(a)-2(c). The value of s_y must equal, at least, the magnitude of the calibration data but most likely it becomes larger, owing to interferences and sample introduction accessories. A minimum uncertainty of an unknown may be estimated, however, by calculating s_y using the uncertainties of Figures 2(a)–2(c), as portrayed in Figures 4(a)–4(c). The calculation of s_y shows that the results of FAAS (Figs. 3(a)–3(b)) were approx. proportional to concentration while the results of ETAAS (Fig. 3(c)) were independent of concentration. The results of Figure 3(c) demonstrate the principle of uncertainties homoscedasticy [4] but this principle was only valid for measurements of ETAAS. The contribution of s_y to the total uncertainty (eq. 8) is rather limited at low and medium concentrations, and reside at the level of centi-absorbances, as demonstrated by the depiction of Figures 3(a)–3(c) and Figures 4(a)–4(c) where the RSD's are shown in the latter figures. The wide distribution of data at high concentrations observed in the experiments of Pt-analysis (Fig. 2b) results in a large contribution of s_y to the total RSD (Fig. 4(b)). A general linear dependence of the STDEV on concentration [16] could only be recognised at high concentrations. As expected, the RSD increased when the concentration approached the limit of detection and it approached infinity when the concentration approached zero. The figures of merits based on the lowest attainable uncertainties were calculated by equations 7 and 8 and they are shown in Table 1 together with corresponding values of straight lines. The extension of the straight line follows that of the manufacturer's specifications only for the determinations of Co but it narrows for Pt and for Fe. A particularly narrow concentration range of linear calibration was predicted for the iron results, and it widened only slightly by applying a significant curvature to the working curve, which is a result of the relatively large spread of data. However, the calibration range was extended considerably by applying a curve to the determinations of Co and Pt, as compared with the extension of the straight line. The concentration range of the working curve may also be extended by the application of polynomial regressions [12–13, 25] or by the application of artificial neural networks [11] but it leads to cumbersome and less comprehensible estimates of the figure of merits. The intercepts were determined as zero, within limits of uncertainty (Tab. 1), thus confirming the setting of the auto-zero function of the instrument that was executed before experiment. The characteristic concentrations were relatively large, and approached the value of the ULA (eq. 9).

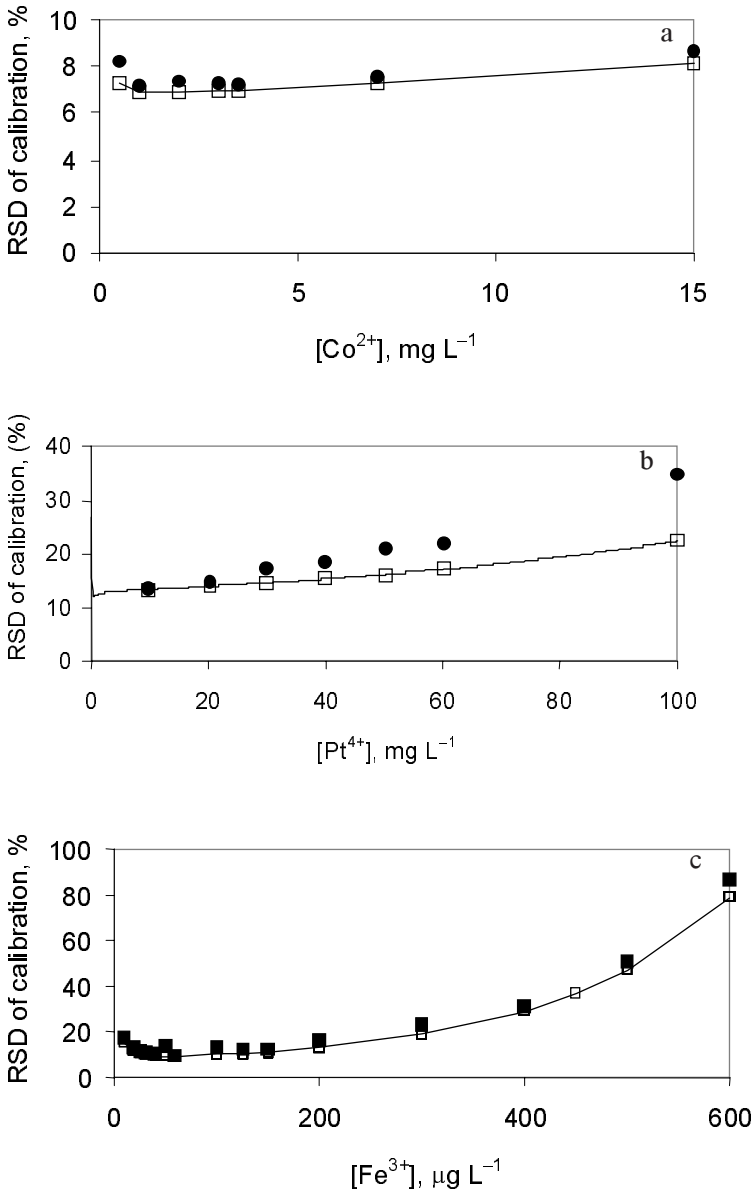

Figure 4 . The RSD of calibration (eq. 8 divided by concentration) depicted as a function of concentration for the determinations of Co by FAAS (a), for the determinations of Pt by FAAS (b) and for the determinations of Fe by ETAAS (c). The relative uncertainties are shown with (●) values of s_y (Fig. 3) and without (□) values of s_y that originate from tentative unknowns with the same uncertainty structure as that of the calibrations. The line represents a theoretical lower limit of the RSD that may be associated with a single determination of an unknown, and the line displays a characteristic concentration of minimum RSD

Table 1. The figures of merits that was obtained by least-squares regression of a working curve (eq. 3) and the calibration line using pooled data. The concentration range recommended by the manufacturer is shown in parentheses

Regression:	Curve					Line		
Parameter	LLA	ULA	A (Abs)	c_m (eq. 3)	c_0	c_{lin} (eq. 4)	LLA	ULA
Pt/ FAAS (mg L^{-1})	0.2	80	0.41 ± 0.06	68 ± 11	0.04 ± 0.02	43	1.1/(0)	20.0/(60)
Co/FAAS (mg L^{-1})	0.04	39	1.15 ± 0.08	20 ± 2	0.03 ± 0.03	13	0.15/(0)	3.5/(3.5)
Fe/ETAAS (µg L^{-1})	5	250	0.91 ± 0.09	131 ± 27	1 ± 8	84	22/(0)	150/(600)

Two independent series of determinations were performed of Fe by this particular reference material by using standards in the range 0–50 mg L^{-1}, as shown in Table 2.

Table 2. Two independent measurement of iron in NIST SRM 1640 by ETAAS using standards of 0, 10, 20, 30, 40 and 50 µg L^{-1}. Standard linear calibration; certified 34.3 ± 1.6 mg L^{-1} (95% confidence range)

#	c, µg L^{-1} Measured	s_c Predicted
1	34.7	1.2
2	34.3	1.2
$\bar{c}$	34.5	1.2
s_c	0.2	
RSD%	0.7	3.4

The results of Table 2 demonstrate that the certified value was indeed recovered to high accuracy by using a straight line for the calibration. Although the two independent series of measurements displayed almost identical concentrations; the RSD imposed by the calibration data amounted to approx. 3% (Tab. 2). However, this RSD of 3.4% (Tab. 2) was obtained by calculating the STDEV of the calibration line [26–27], which provides the value of s_c that originates from calibration data and a single determination of the unknown. In this particular series of determinations, the certified value was excellently reproduced by both determinations at an RSD of much less than one percent (Tab. 2). Since the determination of Fe was performed by only two successive experiments, the STDEV became lower than that of the certified value, simply by means of chance. If more determinations were applied to the analysis, the two STDEV's were expected to approach a mutual value. On average, the two independent determinations deviated by only 0.7%, which appears to fall short by a factor of 5 of the RSD of

calibrations. Thus, the only reliable value of the STDEV was obtained by considering the calibration data as opposed to the STDEV of repetitions of the determination of unknowns. A low number of data applied to the calibration may accidentally promote too low values for the STDEV, even when the average value approaches well the certified value (Tab. 2). Apparently, the results of Table 2 are characterised by high precision and high accuracy.

In order to investigate in full detail the differences between a calibration line of common practice and a working curve (eq. 3), a very large number of experiments were performed (Tab. 3), which ensures that the conditions of high quality data [21] (central limit theorem) was fulfilled throughout. The analysis of Co may be performed by considering a straight line within the concentration range of 0.15–3.5 mg L^{-1} but the application of a working curve (eq. 3) widens the range to 0.06–39 mg L^{-1}, according to Table 1. The certified reference materials EU and EP were analysed at three different dilutions and the RSD values thus obtained were used to assess the quality of measurements. The R-values of EU ranges from 0.4 to 1.7 while it ranges from 0.6–3.1 for the EP determinations, which indicates a satisfactory repeatability. The combined determinations of Table 3 reproduce the certified values within the limits of uncertainty and no significant differences (95% of confidence) were identified among the average values of unknowns. The RSD of calibration was determined by eq. 8 and the RSD of unknowns were found by equation 1. Some of the RSD's shows a trend towards increased values at high dilutions, which is consistent with the results of Figure 4 where the uncertainty increases at lower concentration values.

The application of individual curves or of individual lines (eqs. 3 and 4) for the determination of concentrations provides the minimum values of RSD's of calibration (Tab. 3). Since the RSD of calibration of individual lines or of individual curves were applied to the analysis, the measurements were performed within a limited time span where the instrument operated at high stability with minimum drift of the detector. The STDEV estimated by using the average values for the calculation also showed to be underestimated because the average value represents the most likely event, and the distribution of average values therefore do not convey the true spread of all data (boxes in Tab. 3). Accordingly, the STDEV calculated by using the average values cannot be designated as a useful parameter of analytical results. Conversely, the STDEV of calibration that was calculated by the pooled curves seemed to be vaguely overestimated. However, in order to assess the more realistic performance of a particular instrument, this type of full analysis ought to be performed for a number of selected species. The analyses of Figures 2–3 and of Table 3 demonstrate that it was not possible to estimate a concentration of Co at an RSD below approx. 3%. It should be kept in mind that the determination of Co by the FAAS instrument was virtually insensitive to interferences, and Co constitutes an element that may be considered as an easily analysed element that brings forward the maximum performance of the instrument.

Table 3. Results of the analysis of Co in two different CRM's denoted as EU and EP. The concentrations (mg L^{-1}) were determined by the working curve at three dilutions. The contemporary method of analysis is given by the method denoted as 'Individual lines'. The same results were analysed by using the working curve of equation 4 (Individual curves). The standard deviation of the average values were obtained by calculating the concentrations on the basis of comparing the three average values, as indicated by boxes. Certified EU: 0.74 ± 0.02 mg L^{-1}, EP: 0.095 ± 0.003 mg L^{-1}

Cobalt FAAS	EU	EU	EU	EU	EP	EP	EP	EP	Total	EU	EP
Certified	0.74	0.74	0.74	0.74	0.095	0.095	0.095	0.095		0.74	0.095
Measured at dilution factor	25	50/3	10	25, 50/3, 10	10	5	10/3	10, 5, 10/3		25, 50/3, 10	10, 5, 10/3
N	105	111	60	276	105	111	42	258		276	258
Pooled curves											
c	0.72	0.71	0.73	0.72	0.093	0.094	0.087	0.091		0.72	0.091
s_c	0.05	0.05	0.05	0.01	0.007	0.007	0.006	0.004		0.09	0.012
RSD% predicted	7	7	7	0.9	8	7	7	3		7	8
RSD% measured	5	3	4		6	4	5			4	7
Individual curves											
c	0.71	0.70	0.71	0.71	0.092	0.093	0.093	0.093		0.71	0.093
s_c	0.02	0.02	0.02	0.003	0.004	0.004	0.004	0.001		0.02	0.004
RSD% predicted	2	2	1	0.2	3	2	3	0.4		2	3
RSD% measured	3	3	2		4	4	5			3	4
c	0.75	0.72	0.72	0.73	0.097	0.096	0.087	0.094		0.72	0.096
s_c	0.04	0.02	0.01	0.01	0.007	0.004	0.002	0.005		0.02	0.004
RSD% predicted	3	2	1	1	4	2	2	3		2	3
RSD% measured	3	3	2		5	4	5			3	4

Thus, it was disclosed that the distribution of concentrations of unknowns could be satisfactorily modelled by applying the pooled curves for the calibration and the estimation of STDEV's (eq. 8). The general concept of obtaining high precision by a limited number of standards in the process of calibration may need adjustments because it may be confused with the concept of accuracy. Suppose that a determination was performed by double determinations using a calibration line of some 10 standards [17]. In such an example, an RSD of much less than 0.5% was typically assigned to the result by the manufacturers software report (not shown). A standard deviation below 0.5% thus represents the software's ability to predict a concentration by the aid of one particular calibration line, and broken rational functions are very useful tools for providing such an excellent precision [18]. However, the precision of analysis of, say, 0.5% does not comply with the prediction of the uncertainty budget where the uncertainty on dilution alone most frequently exceeds this value [26]. A repetition of the determination most likely reveals that another concentration results from the analysis, which deviates from the previous result by more than 0.5%. Accordingly, the instrument precision is not a useful parameter for the estimate of uncertainty of unknowns. Since the manufacturer's software does not take into account the distribution of data of the calibration line but merely considers the spread induced by a double determination, the precision becomes underestimated and thus unrealistic. A more realistic accuracy was therefore calculated on the basis of the distribution of data around the calibration line and the uncertainty of the response , as given by equation 8. A maximum accuracy could then be obtained by analysing Co in the CRM EU diluted by a factor of 10, which provided an RSD of 1.3% (Tab. 3) that corresponds to the RSD of iron determination in Table 2. Even by this estimate of accuracy, it did not match the RSD of repeated determinations of the concentration of unknowns. It should be noted that the RSD of unknowns in Table 3 was independent of the method applied to the analysis, and it varied within the range 2.9–6.6%.

By adding 5.0 mg L^{-1} of platinum to the certified reference material EU, the contents were determined at a high recovery (not shown). The uncertainties of the determination were much higher than those of the other elements, which corresponds well to the spread of data observed in Figure 2. Also in this case, the uncertainty is minimal for the results of a straight line using independent and individual determinations. Although a low number of determinations (N = 6) were applied to the analysis of Pt, the general trend was also found with large STDEV's of the working curve of combined results (RSD = 14%) and lower STDEV's of the straight line (RSD = 9%). However, the choice of a working curve instead of a line, allows application of a wider concentration range (0.3–80 mg L^{-1}), and thus less dilution before measurement.

The results of Table 2 show that the certified value was determined at a relative uncertainty of only 3.4% while the RSD of the certified concentration was 4.7%. The RSD of 3.4% even include the contribution from repeatability but the low value

indicate that not enough repetitions were performed, as to reproduce the more realistic value attached to the SRM (Tab. 2). However, another choice of standards of calibration generates much higher uncertainties, as shown in Table 3 where the concentration range of calibration was stretched to its maximum, according to the general level of uncertainty (eq. 9). Thus, a significant curvature of the working curve contributes to a high STDEV, despite the fact that the mean value corresponded well to the certified value (Tab. 3). By modelling the data by the working curve of equation 3, the STDEV decreased significantly and displayed a higher degree of consistency (Tab. 3). The large STDEV's of Table 3 (curve) may be considered as representatives of the maximum reasonable values that model the repeatability of the method. Accordingly, this more pessimistic view on uncertainties may well reveal a method of local determination of reproducibility and accuracy that are usually only possible to estimate by a determination of the STDEV of repeated determinations of a CRM. Once the STDEV of the CRM corresponds to the STDEV of calibrations, then the state of statistical control is obtained.

The determination of the LLA corresponds to a determination of the limit of detection (LOD(3σ)) with the distinction, however, that the LLA includes information about the spread of data around the calibration curve. Thus, a higher value and a more realistic estimate may be obtained by a concentration of an RSD of less than 50%, which corresponds to a minimum concentration that is accessible to analysis by the parent method. The classical LOD considers only the noise level and the slope of the calibration line, which in many cases underestimates the minimum concentration, as recognised by Huber [23] who considered three different methods for the calculation of a realistic lower concentration of analysis.

A comparison of the results obtained by each series of determinations to the results of the certified reference values should be performed comprehensibly by observing the fundamental differences between standard deviations and confidence ranges. In principle, a standard deviation cannot be compared to a confidence range, except for the special case where they were equal, which occurs after approx. 6 repetitions (N = 6, 95% level of confidence). The uncertainty of the certified reference values of the present work (Tab. 2 and 3) are represented by confidence ranges of an unknown number of determinations where, in addition, outliers were removed by the manufacturer from the original set of data. This approach thus provides uncertainties that most definitely remain artificially low, as compared to an uncertainty that was represented by a standard deviation of the method.

The uncertainty structure [13] of ETAAS differed from that of the other methods, and the results indicate that differences in uncertainty structures are related to the technology, *e.g.* the detector, and not only to properties of elements. Tentatively, it may be proposed that the constant s_y value of ETAAS originates from small fragments of graphite that are released in an uncontrollable manner during the atomisation step. The uncon-

trolled release of graphite thus produces as fluctuation of the background signal, which is independent of concentration. The differences in uncertainty structure may be included in future efforts in schemes of interlaboratory testing and schemes of proficiency testing where it is important to make reliable estimates of uncertainties, as opposed to a procedure of minimising uncertainties of a low number of determinations.

CONCLUSION

A thorough examination of analytical results by observing the central-limit theorem led to a satisfactory correspondence between predicted uncertainties of calibrations and the uncertainties obtained by repetitions of unknowns. The correspondence was established by proposing an average value for the propagating uncertainties, where the sum of deviations was averaged by the number of terms in equation 2. The response of the detector was disclosed as a major contributor to the total uncertainty of analysis, and the uncertainty of the detectors response constitutes a major contributor to the uncertainty budget. It may thus be suggested that the instrument should be carefully investigated by day-to-day variations of the responses. In line of these results, a database is desired of calibrations with a search functionality of instrument performances accessible on the internet. A calibration line of 10 standards [17] is inadequate for a reliable estimate of the uncertainty of calibration. More data are required, and for the assessment of uncertainties associated with a novel method of analysis, it is proposed that in excess of a hundred data are appropriate.

The main results of the present report may be summarised as follows:

- A large number of measurements is required for the determination of the calibration line and for the determination of unknowns. In excess of one hundred determinations may be determined in accordance with the conditions of the central-limit theorem.
- A reliable estimate of uncertainties and a correspondence between predicted uncertainties (eq. 8) and measured uncertainties (eqs. 1 and 6) was obtained by averaging the LPE (eq. 2).
- Once a full investigation of the uncertainties was performed, the determination of the concentration of an unknown could be performed by using a low number of standards and a low number of unknowns. The associated uncertainty of the unknown is then assessed by the uncertainty of the full investigation.
- A bias of the method may only be recognised, if the deviation from the expected result were significantly larger than the uncertainty of the full investigation.
- The uncertainty of the method is more important to interlaboratory comparisons than the uncertainty of the mean.

- Any instrument of chemical analysis exhibits a saturation of the response when the concentration exceeds the value of the ULA. The response at very high concentrations in AAS is represented by the threshold parameter 'A' and the associated uncertainty s_A where both should be determined before determination of the working curve (eq. 3).
- The calibration range of the non-linear calibrations extends from the LLA (eq. 10) to the ULA (eq. 9). The ULA may be approximated by using the characteristic concentration: ULA $\cong -B^{-1}$.
- A difference in uncertainty structure was identified between experiments of FAAS and experiments of ETAAS.
- In all experiments of calibrations, a minimum value of the RSD was found within the calibration range.

Acknowledgements

The financial support from Direktor Ib Henriksens Foundation and P.A. Fiskers Foundation is gratefully acknowledged. Many thanks are due to Flemming Ege Hansen (Perkin-Elmer), Francisca Henriksen and André Markussen for their enthusiastic technical assistance.

REFERENCES

1. Galbács G., Gornushkin I.B. and Winefordner J.D., *Talanta*, **63**, 351 (2004).
2. Brereton R.G., *Analyst*, **125**, 2125 (2000).
3. Brüggemann L., Morgenstern P. and Wennrich R., *Accred. Qual. Assur.*, **10**, 344 (2005).
4. Danzer K. and Currie L.A., *Pure Appl. Chem.*, **70**, 993 (1998).
5. Olivieri A.C., Faber N.M., Ferré J., Boqué R., Kalivas J.H. and Mark H., *Pure Appl. Chem.*, **78**, 633 (2006).
6. Faber K. and Kowlski B.R., *Fresenius J. Anal. Chem.*, **357**, 789 (1997).
7. Paneva V.I. and Ponomareva O.B., *Accred. Qual. Assur.*, **4**, 177 (1997).
8. Baumann K., *Process Control Qual.*, **10**, 75 (1997).
9. Centner V., Massart D.L. and de Jong S., *Fresenius J. Anal. Chem.*, **361**, 2 (1998).
10. Otto M., *Fresenius J. Anal. Chem.*, **359**, 123 (1997).
11. Hernández-Caraballo E.A., Avila-Gómez R.M., Rivas F., Burguera M. and Burguera J.L, *Talanta*, **63**, 425 (2004).
12. Tellinghuisen J., *Analyst*, **125**, 1045 (2000).
13. Tellinghuisen J., *Analyst*, **130**, 370 (2004).
14. ISO11095, *Int. Org. Standard.*, Geneve, (1996).
15. ISO5725, *Int. Org. Standard.*, Genéve, (1994).
16. ISO11843, *Int. Org. Standard.*, Genéve, (2003).
17. ISO8466-1, *Int. Org. Standard*, Genéve, (1990).
18. Barnett W.B., *Spectrochim. Acta*, **39B**, 829, (1984).
19. Winefordner J.D., Petrucci G.A., Stevenson C.L. and Smith B.W., *J. Anal. Atom. Spectr.*, **9**, 131 (1994).

20. Kristiansen J., Christensen J.M. and Nielsen J.L., *Mikrochim. Acta*, **123**, 241 (1996).
21. Valcárcel M. and Ríos, A., *Tr. Anal. Chem.*, **18**, 68 (1999).
22. Horwitz W. and Albert R., *Analyst*, **122**, 615 (1997).
23. Huber W., *Accred. Qual. Assur.,* **8**, 213 (2003).
24. Cuadros-Rodríguez L., Gámiz-Gracia L., Almansa-López E. and Laso-Sánchez J., *Tr. Anal. Chem.*, **20**, 195 (2001).
25. Rubeska I. and Svoboda V., *Anal. Chim. Acta*, **32**, 253 (1965).
26. EURACHEM/CITAC Guide, 2nd ed., eds. S.L.R. Ellison, M. Rosslein and A. Williams, (2000).
27. Guide to the Expression of Uncertainty in Measurement, 1st ed., International Organization for Standardization, Genéva, Switzerland (1995)
28. Andersen J.E.T., *Talanta*, **54**, 131 (2001).
29. Salter C. and de Levie R., *J. Chem. Ed.*, **79**, 268 (2002).
30. Valcárcel M., Ríos A. and Maier E., *Accred. Qual. Assur.*, **4**, 143 (1999).
31. Cuadros-Rodríguez L., Gámiz-Gracia L., Almansa-López E. and Bosque-Sendra J.M, *Tr. Anal. Chem.*, **20**, 620 (2001).
32. Meyer E.F., *J. Chem. Ed.*, **74**, 1339 (1997).
33. Salter C., *J. Chem. Ed.*, **77**, 1239 (2000).
34. Whitehead D., *Talanta*, **20**, 193 (1973).
35. Salter C. and de Levie R., *J. Chem. Ed.,* **79**, 268 (2002).

Received March 2007
Revised July 2007
Accepted July 2007