

Modelling uncertainty in a concentration range

J. Jiménez-Chacón · M. Alvarez-Prieto

Received: 11 January 2008 / Accepted: 21 July 2008 / Published online: 23 August 2008
© Springer-Verlag 2008

Abstract Combined uncertainty modelling in a concentration range is an important task for laboratories. Despite several models and regression methods reported, unsolved problems remain, including variability and type of distribution of combined uncertainty estimations and the influence of these on modelling. Intralaboratory data of eight trace elements in natural waters by flame atomic absorption spectrometry and interlaboratory data from some ISO standards were used as an experimental basis. Starting from these and applying the bootstrap technique, high relative variability of combined uncertainty estimations (second-order uncertainty) was found, but normal distributions or distributions with small deviations from normality were encountered. Linear and/or variance models are appropriate for modelling the analyzed data when ordinary weighted least-squares or repeat median robust regression methods are applied. The influences of combined uncertainty variability on modelling and the effect of points that do not follow the general trend of the remaining points are discussed. Some other guidelines are offered.

Keywords Uncertainty modelling · Standard deviation modelling · Second-order uncertainty · Linear and variance models · Ordinary, weighted, and robust regression methods

Electronic supplementary material The online version of this article (doi:10.1007/s00769-008-0440-4) contains supplementary material, which is available to authorized users.

J. Jiménez-Chacón · M. Alvarez-Prieto (✉)
Institute of Science and Technology of Materials,
University of Havana, Zapata y G, s/n, Havana, Cuba
e-mail: malvarez@imre.oc.uh.cu

Introduction

The concept of uncertainty already has a profound influence on many aspects of analytical chemistry, both practically and theoretically. Nowadays it is involved in validation [1], sampling [2], results reporting [3], specification limits [4], and many other concepts and usual laboratory activities. Several normative documents concerning uncertainty and ways to estimate it have been published [5–8]. Once the methodology is selected, it is usual to estimate the uncertainty at different concentration levels. These concentration levels are generally those selected to study precision and bias during the validation of the analytical procedure.

However, when an analytical procedure is used in routine tasks, it is required to estimate uncertainties at concentrations not studied during validation. To solve this, a mathematical model is used in order to make predictions at any concentration included in the application range of the procedure. This mathematical relationship should be as simple as possible, taking into account the supposed causal nexus between concentration and combined uncertainty (u_c). In addition, it is indispensable to apply an adequate method to estimate the regression coefficients of this mathematical relationship. Usually, this is performed by means of a numerical regression method, which should be as simple as possible, for practical reasons.

The number of selected concentrations that are to be used during the validation should be related to the valid concentration range of the analytical procedure. Generally this number is low, for economic and practical reasons. This imposes additional limitations on the modelling process because it is necessary to take decisions on the basis of limited information. Another problem faced by the analyst during the modelling process is the presence of

experimental values of combined uncertainty that do not follow the general tendency given by the causal nexus previously assumed between combined uncertainty and concentration.

Models for precision in a concentration range

Because of the practical importance of modelling precision estimators as a function of concentration, several models have been discussed in the literature. The models are presented in Table 1.

Zitter and God [9] proposed the linear model (Eq. 1) and the variance model (Eq. 6) where σ_0 is interpreted as the standard deviation at null concentration and Θ is the rate of change of standard deviation, σ , with concentration. Thompson [10] used these models to investigate the relationship of the standard deviation, within batch determination, with concentration for 25 elements in about 700 geochemical materials. He studied ordinary and weighted linear squares as regression methods for Eq. 1. As a basis for weighting, he found that the dispersion of the standard deviation increases as standard deviation increases. He pointed out that the weighted regression represents very well data for nearly all the analytes. The variance model was processed by means of an iterative non-linear weighted numerical method. No additional information is offered concerning this method, but the author established its complexity. However, similar results were obtained in comparison with the weighted linear regression method of Eq. 1.

In addition to the linear with intercept model, the ISO 5725 standard [11] also pointed out other alternatives. One of these is the linear without intercept (Eq. 2) and the other is the linear transformation of the exponential relationships expressed by Eq. 3. It is difficult to ascribe analytical meaning to the lack of intercepts of these models, because it is very well known that standard deviation is not null at

zero concentration. However, they can be used far from the detection limit. The same characteristics should be ascribed to the models from the Horwitz's Curve [12] (Eq. 4), and to Hughes and Hurley's proposal [13] (Eq. 5). The lack of intercept makes these models non-valid for use in a concentration range that starts near the detection limit.

Rocke and Lorenzato [14] proposed a model (Eq. 7) based on the physical characteristics of the measurement process. They assumed that measurement variability is approximately constant at low concentrations, but changes in a continuous mathematical manner to a relationship in which measurement variability increases as concentration increases. The two-component model of Rocke and Lorenzato is supported on the basis of combination of additive and multiplicative errors, an old idea presented in the literature. They used the maximum likelihood principle for regression, something that is not justified on the basis of differences in the variabilities of standard deviation and concentration. Additionally, it involves some numerical complexity which is not appropriate for routine practice. So, this model is not going to be considered in this work.

Models for uncertainty in a concentration range

In principle, all the functional relationships presented in Table 1 can be used to predict uncertainties in a concentration range, because generally the main component of combined uncertainty is the imprecision of measurements. In that sense, the Eurachem/CITAC guide on uncertainty has established a model for documenting u_c which depends on analytical level [6].

$$u_c = \sqrt{S_0^2 + S_1^2 w_B^2} \quad (8)$$

In this expression w_B is the mass fraction or another quantity used to express chemical measurement results of a component B. Three possible cases are pointed out:

1. uncertainty not dependent on analytical level, (S_0 dominant);
2. uncertainty entirely dependent on analytical level, (S_1 dominant); and
3. intermediate dependence (neither S_0 nor S_1 dominant).

The general form of the graphical representation of Eq. 8 shows curvature towards the combined uncertainty axis at very low concentrations and straight behavior at higher concentrations, in a linear scale on both axes. This equation is essentially the variance model. The values of S_0^2 and S_1^2 can be estimated by ordinary least-squares regression (OLS) of u_c^2 versus w_B^2 .

Another way to estimate the coefficients of Eq. 8 is the use of an iterative non-linear regression starting from appropriate initial values of both coefficients. In that sense

Table 1 Proposed models for standard deviation as a function of concentration

Model	Function	Equation
Linear with intercept	$\sigma = \sigma_0 + \Theta w_B$	1
Linear without intercept	$\sigma = \Theta w_B$	2
Exponential 1	$\sigma = q w_B^p (p < 1)$	3
Exponential 2	$\sigma = 0.02 w_B^{0.8495}$	4
Exponential 3	$\sigma = K w_B^{0.5}$	5
Variance	$\sigma^2 = \sigma_0^2 + \Theta^2 w_B^2$	6
Two-component	$\sigma = a + b w_B e^\eta$	7

w_B mass fraction or another quantity used to express chemical measurement results; $\sigma_0, \Theta, p, q, K, a, b, \eta$ are regression coefficients

Heydorn and Anglov [15] developed a novel approach to calculate coefficients of Eq. 8 to model calibration uncertainty. They transformed Eq. 8 by means of logarithms and applied non-linear regression using the Solver utility of Microsoft Excel spreadsheet to the expression:

$$\ln(u_c) = 0.5 \ln(S_0^2 + S_1^2 w_B^2). \quad (9)$$

The approach was tested for the determination of Pb by ICP-AES, demonstrating statistical control of analytical results and absence of bias. They took into account uncertainties of the measurements. The authors pointed out that the logarithmic transformation makes the variances of u_c homoscedastic.

On the basis of a vectorial combination of independent uncertainties at low and high concentration levels, Thomson and Wood [16] proposed a function with the form:

$$u_c = \sqrt{(w_{DL}/2)^2 + A^2 w_B^2}. \quad (10)$$

In this expression, w_{DL} is the concentration at the detection limit and A is a constant of proportionality, the main factor which determines u_c at high concentrations. The authors called such a function an “uncertainty function”, the algebraic relationship describing how uncertainty of measurement varies with the concentration of the analyte in respect of a specific analytical procedure applied to a certain class of material. The same authors defined “characteristic function” as an uncertainty function that describes performance (referring to the results obtained by a specific analytical procedure in particular circumstances). Evidently, this latter model is essentially Eq. 6, also. The coefficients of Eq. 10 can be estimated from intralaboratory validation results (precision estimators, bias estimation, and detection limit). In that sense, w_{DL} is obtained from the estimated detection limit and coefficient A is estimated from an empirical relationship between the repeatability and reproducibility standard deviations ($\sigma_r/\sigma_R \approx 0.5$), the last of which derived from the Horwitz’s function [12]. When results of reproducibility at different concentrations from collaborative trials are available the coefficients can be estimated by regression.

Some of the discussed models are inappropriate for concentration ranges starting near the detection limit because of the lack of an intercept (Eq. 2–5). The regression with others is mathematically complex and not so appropriate for routine tasks (Eq. 6 with the mentioned iterative method and Eq. 7 applying the maximum likelihood principle for regression).

On the basis of professional judgment and from previous experience [15], it is known that the variability of u_c increases with its value. It is, then, attractive to provide a different experimental support of this assertion in order to

quantify how large in relative terms is this variability and to determine if the statistical distribution of combined uncertainty is normal or not. These characteristics have an important influence on the modelling of u_c by means of regression along a concentration range. Additionally, it could be useful to employ not previously reported tools and resources for regression of u_c on w_B . Furthermore, the analyst needs guidelines to handle cases that could be presented in practical situations during modelling, such as a low number of experimental values of combined uncertainty and cases where there are points that do not follow the general tendency shown by the other points.

The main goals of this work are:

1. to estimate, starting from an experimental basis, the relative variability of combined uncertainty at different levels of concentration, i.e., second-order uncertainty;
2. to draw conclusions about the type of distribution of u_c in the studied data;
3. to give an idea about the influence of these considerations on modelling; and
4. to establish some guidelines in order to find appropriate regression methods and models to describe the behaviour of u_c with concentration, making use of some additional tools not employed previously.

For such purposes, inter and intralaboratory data starting near the detection limit are used.

Materials and methods

Intralaboratory results

Validation data from determinations of traces of Cd, Co, Cr, Cu, Fe, Mn, Pb, and Zn in natural waters by flame atomic absorption spectrometry (FAAS) were used. The procedures were written as standard operating procedures starting from a series of internationally recognized standards [17]. Precision and bias were simultaneously characterized by means of the same experiment. A fully nested experimental design with three factors (days, analyst, and repeatability in the order of nesting) and ANOVA model II were applied to decompose the total variance in its components. So, standard deviations of intermediate precision time different (s_T), analyst different (s_A) and standard deviation of repeatability (s_r) were obtained for three levels of concentration. From these, the standard deviation of intermediate precision time–analyst different was obtained ($s_{I(TA)}$) [18]. The principal uncertainty sources of $s_{I(TA)}$ are the uncertainty due to the analytical signal and the calibration process. But some other sources, for example aliquots and dilutions of synthetic “samples” to a fixed volume are also included. For example, uncertainties

due to these sources represent up to 8.8% of $s_{I(TA)}$, calculated according to Ref. [6]. So, the analytical procedure is simple, but the uncertainty budget includes some sources of uncertainty, not only uncertainty due to calibration. The concentration levels of synthetic “samples” were 0.05, 0.5, and 1.0 mg l⁻¹ for Cd and Zn and 0.25, 1.5, and 4.0 mg l⁻¹ for the rest of the elements. These concentrations were prepared from laboratory primary standards of high purity metals (Johnson–Matthey, UK), using calibrated glassware and a calibrated analytical balance. Uncertainties of standard concentrations were estimated using the Eurachem/CITAC approach [6] and a Microsoft Excel template routinely used in the laboratory for such purposes. Calibrations were performed starting from primary standards different from those used for synthetic “samples”.

Calibration curves and blank tests were prepared and measured independently by each analyst on each day. A Philips PU 9100 FAAS was used with instrumental conditions optimized for highest sensitivity. Calibrations were performed with the Barnett’s algorithm to correct curvature [19]. The concentrations of the blank assays were calculated for each element and the intermediate precision time-analyst different was also obtained by means of ANOVA. This standard deviation was considered as the combined uncertainty for the blank concentration.

The bias was tested from recoveries according to Barwick and Ellison [20]. Finally, an estimate of uncertainty was obtained for each concentration level, by combining the uncertainty due to the experiment to evaluate the bias and the intermediate precision time-analyst different according to Moroto and co-workers [21]. This additional uncertainty source is included in the uncertainty budget. All the above calculations were performed with Microsoft Excel. These combined uncertainties should be considered as more complete at the highest organizational level within our laboratory. Additionally, this is the starting point to estimate the combined uncertainty for the interlaboratory organizational level [8].

In order to estimate the confidence intervals for uncertainty associated with the concentration level it is necessary to perform repetitions of the fully nested design. This is impractical and to overcome this difficulty the bootstrap technique was used [22, 23]. The replicate of the design was generated taking samples at random with replacement from the experimental absorbances of the reference solutions obtained by each analyst on each day. This sampling ensures that repeatability is the unique source of variability included in each sample. The concentrations were calculated using the selected absorbances and the corresponding calibration curve. An estimate of u_c for each replicate of the design was obtained in a similar way as experimental data. The 95% confidence interval for combined uncertainty was calculated according to:

$$2u_c - P_{0.975} \leq u_c \leq 2u_c + P_{0.025} \quad (11)$$

where u_c is the experimental combined uncertainty associated with the level of concentration, $P_{0.975}$ and $P_{0.025}$ are the 97.5 and 2.5% percentiles, respectively, of the 2000 simulated values of u_c . This resource can be considered as a type B evaluation of an uncertainty of a estimated combined uncertainty.

In the case of the concentration of the blank assay the above procedure is applied but only the confidence interval for the precision time-analyst different is considered. All the calculations for simulations were performed with a program expressly written in MatLab [24].

Interlaboratory results

Numerical examples of results of two interlaboratory studies reported in different parts of ISO 5725 [18, 25] were selected. The fully nested [25] or the staggered-nested [18] designs were used to obtain estimates of precision at several concentration levels. In one of the examples [25] the combined uncertainty was the combination of the reproducibility standard deviation and uncertainty of bias estimation, but this latter component was negligible due to the experimental design used and the number of laboratories involved. In the other example [18], only the reproducibility standard deviation was considered for u_c . In both cases the confidence intervals for u_c were calculated with MatLab [24] taking samples at random with replacement from the concentration reported by each laboratory by means of the bootstrap technique. Equation 11 was used for 2000 replicates of the design.

Others results of interlaboratory studies were also used [26]. Unfortunately, only the final estimates of the standard deviation of the reproducibility were reported and the confidence interval could not be calculated because of the impossibility of obtaining replicates of the design. This example is exactly what the analyst usually has at hand, when no confidence intervals of u_c are available.

Results and discussion

Results of simulations

For the intralaboratory results, the 95% confidence intervals calculated from Eq. 11 show that the estimates of combined uncertainty obtained have a high relative variability. This variability, expressed in percent as confident interval over experimental u_c , resulted about 22–65% for the blank assay with a mean and median values of 43 and 48%, respectively. For the concentration levels the relative variability was about 11–47% with a mean and median

values of 27 and 26%, respectively. The difference between the mean of the 2000 simulations and the experimental value of u_c associated with the level of concentration or blank assay is the bias of the bootstrap. This bias was less than 2% relative in all cases.

The distributions of simulated combined uncertainties by means of the bootstrap were studied using Statgraphics Plus for Windows [27]. A confidence level of $\alpha = 0.05$ was selected for the statistical tests. The frequency histogram, quartile plot, normal probability plot, Kolmogorov-Smirnov, Kuiper V , Anderson-Darling A^2 and some other goodness-of-fit tests suggest that the results of the simulations of the intralaboratory estimated values of u_c are normally distributed in about 2/3 of the cases. In the remaining 1/3 the data are not normal. However, the symmetry plots show that the majority of distributions are only slightly asymmetric (negative skewed). For blank assays, the results of simulated combined uncertainty follow a clear non-normal distribution in 2/3 of the cases (of Pb, Zn, Cd, Co, and Cr, the last three showed great deviation from normality).

For interlaboratory results, the 95% confidence intervals for u_c are also wide. The bias of the bootstrap was less than 1% relative in all cases. The results of combined uncertainty showed a high variability, also. This variability, expressed in the same way as in the intralaboratory cases, was from 16 to 49% with a mean and median values of 27 and 23%, respectively. About a half of the groups of simulated combined uncertainties can be assumed to be normally distributed, but the majority of them were slightly asymmetric (negative skewed). Only one resulted somewhat positive skewed. The results of the bootstrap can be consulted in the Electronic Supplemental Material.

Intralaboratory results

Statistical evidence of bias for all the elements during the validation process was not found. A linear tendency of the estimated value of u_c with concentration is observed for all the elements considered in the intralaboratory validation. The plots of u_c at the respective concentration levels of Cr and Cu shown in Fig. 1 are typical examples. The plots also contain the confidence intervals represented by a short bar below and above the particular value of u_c . Because, in practice, it is frequent that samples have analyte concentration between detection limit and the bottom end of the mass concentration range considered in the validation, the blanks were included in the plot in order to give an idea of the behaviour of u_c below this bottom end.

As starting point, the OLS regression was used to estimate the coefficients of Eqs. 1 and 8 to model the behaviour of u_c with concentration. The values of the combined uncertainty for blanks are also included in the fit.

In Fig. 1a, c the solid (Eq. 1) and broken (Eq. 8) lines are the fitting results of both models to the data.

Equation 1 with the OLS fit represents well the data for the mass concentration range considered in the validation. The regression line passes inside all the confidence intervals of u_c which represents a criterion on statistical basis. The value of the intercept σ_0 is a good estimation of the experimental value of combined uncertainty for the blank assay. Extrapolations at concentrations lower than the concentration range considered in the validation can be made because the behaviour of u_c with concentration can be assumed to be linear. The same was observed for the rest of the elements, which are detailed in the Electronic Supplemental Material.

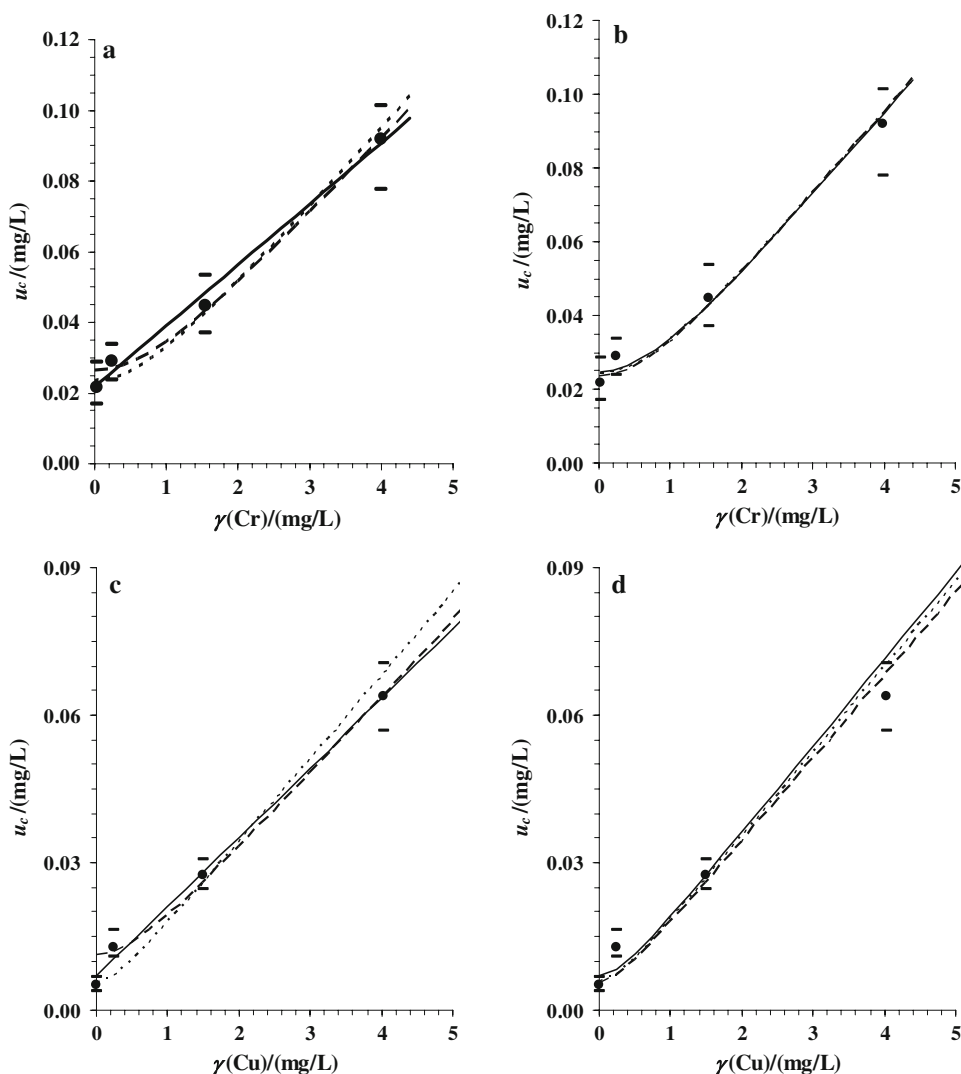
Equation 8 with the OLS fit also represents well the data for both elements in the mass concentration range of validation. The regression curves pass through all the confidence intervals of u_c . For Cr, the value of S_0 can be considered a good estimation of the experimental combined uncertainty for the blank assay. In the case of Cu, the predicted value of u_c for the blank is greater than expected taking into account the variability of the experimental value. Extrapolations with Eq. 8 at concentrations lower than the first level of concentration could be made for Cr, but is not recommended for Cu because there is a tendency to overestimate u_c .

For all the elements between the respective first and third concentration levels, there are no significant differences between the fits with Eq. 1 and 8, and both can be used without distinction to make predictions of u_c . It was found in this work that when the first concentration level is approximately greater than 3.5 times the detection limit (Cu, Co, and Mn) the extrapolations with Eq. 8 at concentrations lower than this level tend to overestimate u_c .

With respect to Eq. 9, the procedure indicated by Thompson and Wood [16] to obtain coefficient A^2 from validation results cannot be applied. In that sense, it is necessary to have precision estimates expressed as relative standard deviation (RSD) at concentrations at least 50 times the detection limit. This requirement is only fulfilled for the second and third concentration levels of Mn and Cu. For those elements, there is a little difference between the OLS estimates of the coefficients and the estimates from the procedure proposed in reference [16].

As observed from confidence intervals in Fig. 1, the absolute variability in u_c increases with the value of u_c . This suggests the use of weighted least-squares regression (WLS) instead of OLS to estimate the coefficients of Eq. 1 and 8. Considering the results mentioned above and the fact that, in practice, only an estimated point of u_c is obtained per concentration, the inverse of the square value of u_c , i.e. $1/u_c^2$, is used as weighting factor.

Fig. 1 Estimated values of u_c and their 95% confidence intervals at different levels of mass concentration (γ) of Cr and Cu. In **a** and **c** solid lines represent the OLS fit of Eq. 1, broken and dotted lines represent, respectively, the OLS and WLS fit of Eq. 8 to the data. In **b** and **d** the solid, broken, and dotted lines represent, respectively, the NLLT, WLS, and NLW fit of Eq. 8 to the data



For all the elements, the estimation of the coefficients of Eq. 1 by WLS regression does not yield considerable differences from those estimated by OLS regression. The results are very similar (Table 2) and are also very similar to the predicted values of combined uncertainty.

In the case of Eq. 8, the use of the WLS regression yields differences with respect to OLS only for Cd, Co, Cu, and Mn. A better estimation of the experimental combined uncertainty of the blank assay for these elements is obtained. An example can be seen for Cu in Fig. 1c (dotted line). However, the predictions of u_c near the first concentration level tend to be lower than expected despite the high weight assigned to this level in the regression. For Cr no noticeable difference between OLS and WLS is observed.

A possible inconvenience in estimation of the coefficients of Eq. 8 by linear WLS regression is the previous transformation of the data indicated in Ref. [6], that is, to square both members of Eq. 8 prior to regression. Taking

this into account, the coefficients S_0^2 and S_1^2 are also estimated by two different ways. One is the nonlinear weighted regression (NLW) using the same weighting factor ($1/u_c^2$) in which no previous transformation of the data is needed. The other is nonlinear regression of the logarithmic transformation (NLLT) by mean of Eq. 9 suggested by Heydorn and Anglov [15], in which no weights are assigned. In both cases the Solver utility of Microsoft Excel spreadsheet was used. As starting values for the coefficients, those obtained by the WLS regression were taken, as indicated in Ref. [15].

The results of the NLLT, NLW, and WLS are presented in Fig. 1b, d by mean of the solid, broken, and dotted lines, respectively. For Cr there are no differences between the regression methods and for Cu the differences that appear in the plot can be considered as not remarkable. For the rest of elements similar behaviour was observed. Compared with WLS, the use of the two nonlinear regressions does not yield an improvement for Cu because the problem

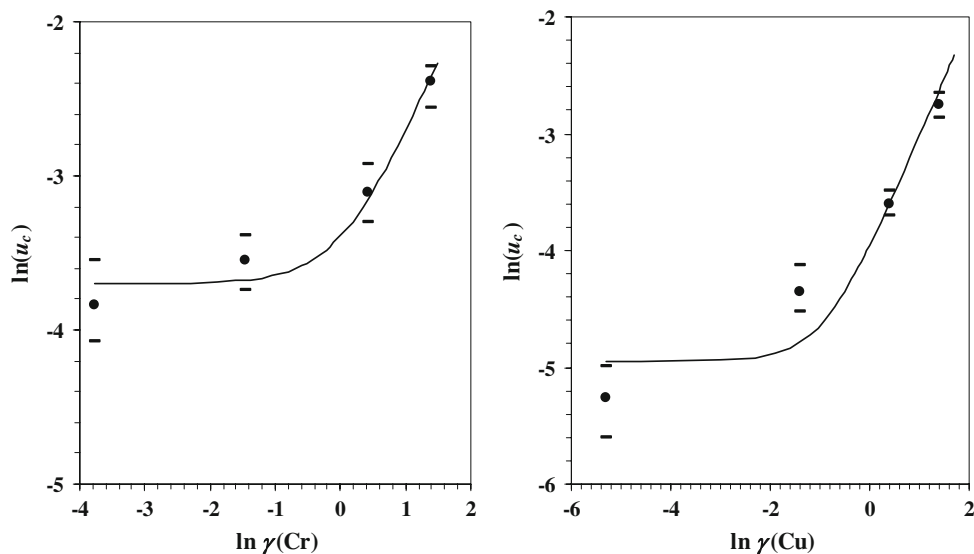
Table 2 Estimated values of σ_0 (upper) and θ (lower) for Eq. 1 obtained from OLS, WLS, and RMM

Element	OLS	WLS	RMM
Cd	0.00353	0.00351	0.00353
	0.02201	0.02178	0.02220
Co	0.01646	0.01577	0.01536
	0.02201	0.02244	0.02225
Cr	0.02181	0.02224	0.02214
	0.01723	0.01656	0.01723
Cu	0.00698	0.00564	0.00580
	0.01410	0.01523	0.01443
Fe	0.03390	0.03512	0.03546
	0.02033	0.01899	0.02036
Mn	0.01018	0.00753	0.01157
	0.02023	0.02259	0.01977
Pb	0.04771	0.04777	0.04772
	0.00726	0.00717	0.00725
Zn	0.00459	0.00482	0.00486
	0.02434	0.02354	0.02258

σ_0 is expressed in mg l^{-1} , θ has no units

concerned with the predictions of u_c near the first concentration level is still present. The same was observed for Cd, Co, and Mn. From all the above results, it is concluded that for our intralaboratory results the transformation indicated in Ref. [6] has no influence on estimation of the coefficients of Eq. 8. Moreover, Heydorn and Anglov [15] pointed out that transformation of data using Eq. 9 results in homogeneous variances and it is not necessary to use a weighting factor when nonlinear regression is applied. Figure 2 is a plot of the logarithmic transformation of u_c for Cr and Cu.

Fig. 2 Natural logarithm of the estimated values of u_c and their confidence intervals as a function of natural logarithm of mass concentration for Cr and Cu. The *solid line* in the figure is the result of non-linear fit [15] of Eq. 9 to the data



It is evident from Fig. 2 that the logarithmic transformation does not make the variability of u_c constant. This was statistically shown by means of the Cochran's test. The differences between confidence intervals are not so manifest as in Fig. 1 but are still present. To investigate this in a deeper way, the logarithmic transformation was applied to the result of the bootstrap technique for all elements. The variances of the new transformed data [$\ln(u_c)$] were statistically compared and in all the cases the variances were statistically different. This transformation with the aim of variance stabilization could be useful only if the relative second-order combined uncertainty is constant. However, compared with the variances of the non-transformed data, the logarithmic transformation makes the differences smaller but it does not eliminate them. In our cases, the coincidence in the results of NLW and NLLT could be explained by the fact that in the NLLT the starting values of the coefficients were taken from WLS regression.

The problem of unequal variability associated with the dependent variable u_c justified the use of WLS instead of OLS. In the case of Eq. 1 for all elements and for Eq. 8 with the exception of Cd, Co, Cu, and Mn, there are no noticeable differences between the estimated values of the coefficients obtained with both regression methods. So, which of the regression methods should be selected? If WLS and OLS are compared, in the former the confidence intervals for coefficients are smaller and the confidence intervals for prediction are more realistic [28]. In practice, the main interest is to make point predictions of u_c in a defined concentration range and not to obtain a confidence interval of the predictions (second-order uncertainty). Taking this into account, the OLS regression can be selected to estimate the coefficients of both models despite the unequal variability observed in the values of u_c used in the modelling process.

The OLS and WLS regression methods have one characteristic in common: they are very sensitive to the presence of outliers. That is, both regressions methods are not robust. This means that outlier data can have a large influence on the values of the estimated coefficients and, of course, in the predicted values of the dependent variable. The fact that in Fig. 1a both regression curves (OLS and WLS with Eq. 8) can be considered similar is because there is no an experimental value of u_c out of the general tendency described by Eq. 8.

Several robust regression methods have been developed. Gao and co-workers [29] applied the iterative least-square linear regression with the linear model (Eq. 1) to precision data from trace element determinations in soil and water deposit certified reference materials. However, the numerical technique is complex and difficult to implement in routine tasks. Other single robust regression methods have been proposed [30]. The repeat median method (RMM) proposed by Siegel [31] is very resistant to outliers and easy to implement, for example, in a Microsoft Excel spreadsheet. Compared with OLS and WLS regressions the application of a RMM to Eq. 1 does not yield appreciable changes in the estimated values of coefficients as can be observed in Table 2. This is because the experimental behaviour of u_c with concentration is in concordance with the relationship expressed by mean of Eq. 1 and none of the values of u_c can be considered far from that general tendency.

The RMM can also be applied to Eq. 8. For this, the method is applied to the linear model that results from u_c^2 versus the square of concentration. In this case, it produces noticeable changes in the estimated values of the coefficient, mainly in S_0^2 , only for Co, Cu, and Mn. These changes are appreciable only with respect to WLS regression as can be seen in Fig. 3 for Co and Cu. The behaviour of u_c with concentration between the blank and the first concentration level can not be assumed according to the functional relationship established by Eq. 8 for low concentrations. In other words, the S_0^2 coefficient cannot be assumed as dominant in the relationship. The RMM does not take into account the blank values of u_c for Cu, Co, and Mn during the respective regressions. It does not mean that such values of u_c are considered as outliers. To confirm if a suspicious point is an outlier, it is necessary to perform a specific statistical test with robust regressions [30]. However, the statistical test did not indicate these blanks as outliers. For Cd, Co, Cu, and Mn, the application of WLS in order to estimate the coefficients of Eq. 8 is not recommended because a high weight is assigned to an experimental value that does not follow the tendency described by the model. The RMM solves the problem of the predictions of u_c near the first concentration level observed for the WLS. However, the tendency to

overestimate u_c is also presented when it is necessary to estimate combined uncertainties below the first concentration level.

It is known that the RMM is not equivariant for transformations of the independent variable [32]. Linearization by means of squares of both sides of Eq. 8 produces a magnification of u_c variability, which could be problematic during the application of robust regression. With the aim of showing this fact and starting from results of the bootstrap, squaring of u_c doubles its relative variability. So, in the case of Eq. 8, this regression method should be approached with caution, but from a practical point of view.

Interlaboratory cases

The results of two collaborative studies reported as examples in ISO 5725 [18, 25] are represented graphically in Fig. 4. Figure 4a, b corresponds to the results of ISO 5725-3 (flame atomic absorption spectrometric determination of V in steel) while Fig. 4c, d correspond to the results of ISO 5725-4 (flame atomic absorption determination of Mn in iron ores). In the plots also appear the confidence intervals for u_c calculated from bootstrap technique. At first sight, it is observed that in the two examples the values of u_c increase with concentration showing a linear trend. However in both cases, the relative position of the value of u_c belonging to the higher concentration does not exactly follow the linear tendency shown by the previous values of u_c . In the case of ISO 5725-3 it is a little above and in 5725-4 a little below. So, these values should be considered out of the general tendency of the rest of the points. Moreover, this inconvenience could lead to assumption of apparent curvature in the behaviour of u_c with concentration in both plots.

In principle, the models represented by Eqs. (1) and (8) can be used to describe the experimental behaviour of u_c with concentration shown in Fig. 4. The problem is centred on selection of the regression method to estimate the coefficients. In that sense, the OLS, WLS, and RMM fits of Eq. 8 are represented in Fig. 4a, c by solid, broken and dotted lines, respectively. It is clear that the application of OLS to Eq. 8 is not suitable. There is a noticeable tendency to obtain very high or very low estimated values of u_c when concentration decreases. Moreover, unacceptable (negative) estimated values of u_c are obtained in the case of ISO 5725-3. The results obtained with WLS and RMM are much better than the obtained with OLS. Both regression methods represent well the behaviour of u_c in almost the whole concentration range considered. The predicted values of combined uncertainties are acceptable considering the high variability of experimental values of u_c . The RMM does not take into account the value of u_c belonging to the higher concentration. In both examples, the last point is

Fig. 3 Estimated values of u_c and their confidence intervals at different levels of mass concentration of Co and Cu. Solid, broken, and dotted lines represent, respectively, the OLS, WLS, and RMM fit of Eq. 8 to the data

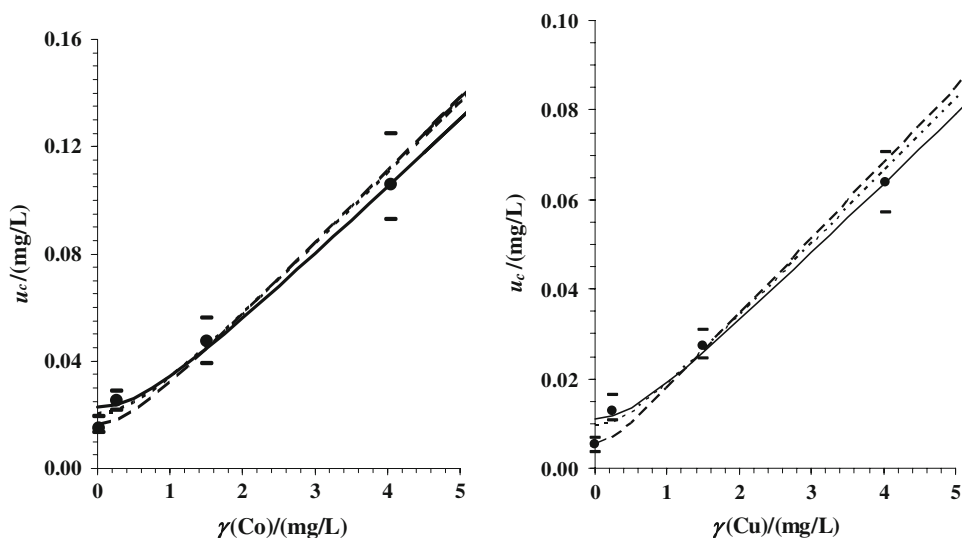
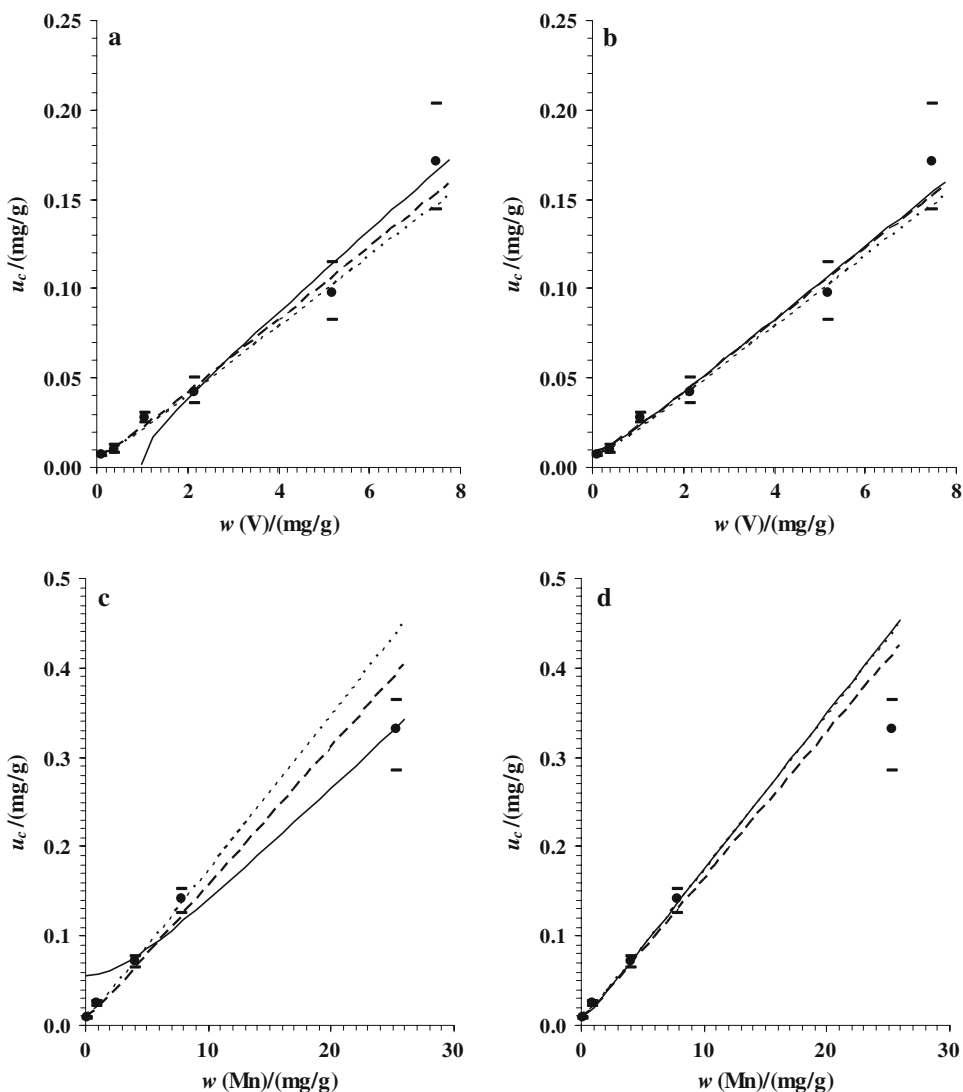


Fig. 4 Results of interlaboratory studies reported in ISO 5725-3 (a, b) and ISO 5725-4(c, d). In a and c solid, broken, and dotted lines represent, respectively, the OLS, WLS, and RMM fit of Eq. 8 to the data. In b and d solid, broken and dotted lines represent, respectively, the NLLT, NLW and RMM fit of Eq. 8 to the data



identified as an outlier according to the statistical test intended for that purpose [30]. A similar effect is produced by WLS because the assignation of a low weight to the last point for both plots. Practically, along the whole concentration range the problems of underestimation or overestimation are not presented. However, near the higher concentration, the estimated values of u_c can be considered different from the experimental ones in the case of ISO 5725-4.

In Fig. 4b, d the results of the NLLT, NLW, and RMM fit of Eq. 8 are presented by solid, broken and dotted lines, respectively. As can be observed, there are no noticeable differences between the two non-linear (NLLT and NLW) and the RMM linear method for the Eq. 8 coefficient estimation. Both, the NLLT and NLW do not seem to be so sensitive to the presence of possible outliers. However, the logarithmic transformation does not make the variability of u_c constant. As in intralaboratory examples, this transformation makes the differences among the variances smaller but it does not eliminate them. So, the RMM is recommended taking in to account its computational simplicity.

The OLS, WLS, and RMM fits of Eq. 1 to the same interlaboratory results are represented in Fig. 5a (ISOI 5725-3) and 5b (ISO 5725-4) by solid, broken, and dotted lines, respectively. With the OLS regression, considerable underestimates (ISO 5725-3) and overestimates (ISO 5725-4) of values of u_c below the respective third concentration level are obtained. However, the results with WLS and RMM are much better than those obtained with OLS; practically along the whole concentration range the problems of underestimation or overestimation are not presented. As in Eq. 8, in both examples the last point is identified as an outlier. The results are represented in the Fig. 5 on a log–log scale because this kind of representation has the advantage that nicely separates all points and

make easier the graphical representation along the whole concentration range. If a linear scale were selected the problems of lack of fit at lower concentrations with the OLS regression could not be well appreciated.

Figure 6 consists of two plots with the results of an interlaboratory study of determination of Mn in steel and iron reported in ISO 10700 by FAAS [26]. This is a situation frequently found in practice, that is, there is no information about the variability of estimated values of u_c . The intervals that appear in the plots represent 20% of the experimental respective values of u_c . Of course, this is arbitrary, but roughly approximated to the results obtained with the bootstrap technique for the former interlaboratory results. The values of u_c increase with concentration showing a linear trend. The combined uncertainty belonging to the fifth, sixth, and ninth concentration levels seem to be apart from the behaviour of the remaining points. These can be observed better in a linear scale but in this kind of representation the first four concentration levels practically overlap with the origin (see the Electronic Supplemental Material to view the results with linear scales).

Applying OLS fit to Eq. 1, (Fig. 6a, solid line) the ninth point affects the regression and produces a marked overestimation of u_c for the first four points. However, the application of WLS (broken line) to this model is not so affected by the position of the ninth point because of the low weight assigned during the regression. Although the assigned weights to the fifth and sixth points are high, the weights assigned to the first four points are higher. So the effects of the fifth and sixth points are superseded or compensated during the WLS fit. The RMM (dotted line) is not affected at all by the relative position of these three points that do not follow the general trend of the remaining points. In this example, the ninth point is identified as

Fig. 5 Results of interlaboratory studies reported in ISO 5725-3 (a) and ISO 5725-4 (b). Solid, broken, and dotted lines represent, respectively, the OLS, WLS, and RMM fit of Eq. 1 to the data

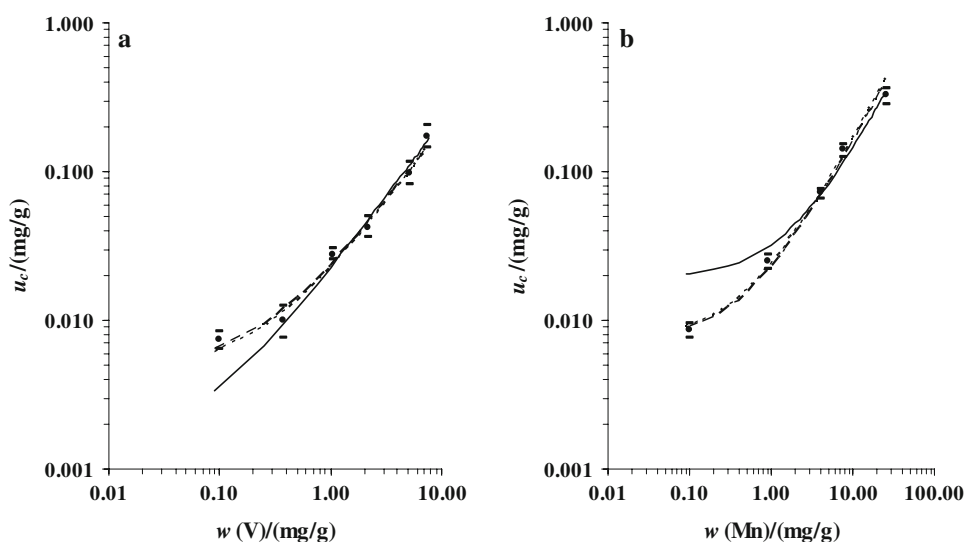
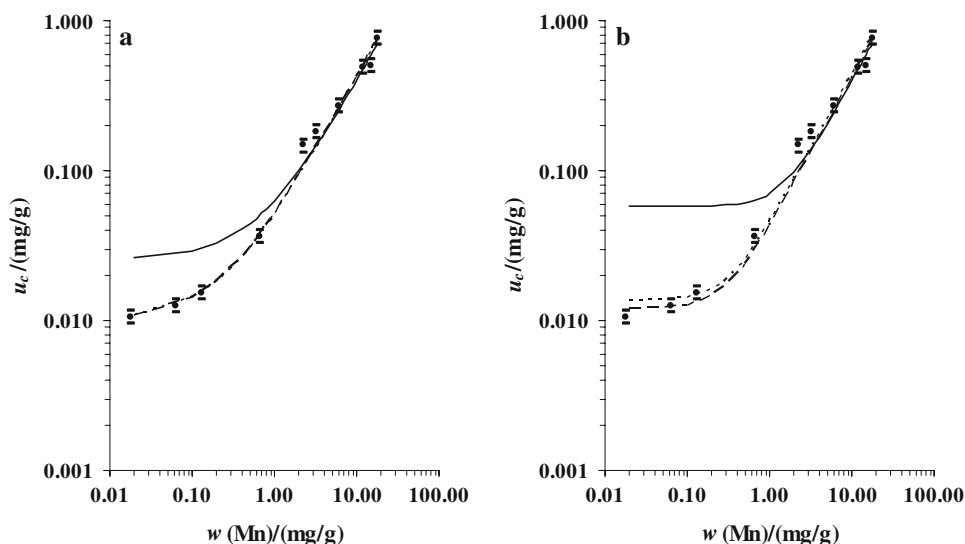


Fig. 6 Results of interlaboratory studies reported in ISO 10700. *Solid, broken, and dotted lines* represent, respectively, the OLS, WLS, and RMM fit of Eq. 1 (a) and Eq. 8 (b) to the data



outlier according to the statistical test [30]. Compared with WLS the predictions of u_c at higher concentrations are very similar, but at lower concentrations the RMM produces small overestimations of u_c . In the original written standard [26], the data were fitted through linear transformation of Eq. 3 by means of logarithms and represented in a log–log plot, something that can mask the lack of fit and be misleading.

The application of OLS (Fig. 6c, solid line) to Eq. 8 is not suitable and also predicts overestimated values of u_c near the bottom end of the concentration range. The overestimation is greater than in Eq. 1 as can be seen comparing Fig. 6a, b. The application of WLS fit (Fig. 6b, broken line) dramatically improves predictions in all the concentration range, principally in the lower part. Similar results were obtained with the RMM (Fig. 6b, dotted line). Compared with Eq. 1, the differences in predictions between WLS and RMM with Eq. 8 are greater. As in the intralaboratory and the above interlaboratory examples, the results obtained with the NLLT and NLW (not shown) are very similar to the obtained with WLS and RMM. So, RMM is also recommended because it is computational simplicity.

The fact that the use of WLS with both models is more appropriated (if compared with OLS fit) is not a consequence of taking in to account the unequal variability of experimental values of u_c in the regression. It is a consequence of the position inside the concentration range of the points that show a different tendency of the rest. That is, if such points are positioned at the upper part of the concentration range, the low assigned weight confers some “robustness” to WLS. On the other hand, if those points are positioned at low concentrations WLS is not recommended. The application of the RMM to the models represented by Eqs. (1) and (8) confirms that in the

examples the points considered as suspicious are apart from the general tendency shown by the rest of the points. This robust method does not take into account these points in the respective regressions. The differences in predictions between WLS and RMM along the concentration range are magnified with Eq. 8, if compared with Eq. 1. It is an additional argument about the sensitivity of the variance model to departures of experimental values of u_c from the general behaviour of the rest of the values.

Despite the mentioned limitations of RMM with Eq. 8, its application to the studied cases produced good results. A Microsoft Excel template for robust regression by means of the RMM is included in the Electronic Supplemental Material.

Conclusions

The process of modelling combined uncertainty in a concentration range is not always a single task. Several aspects should be considered for the selection of the appropriate model. It should be as simple as possible, taking into consideration the supposed causal nexus between concentration and combined uncertainty. Despite the diversity of available proposed models, none can be considered as a universal solution for all the possible cases. In that sense, the linear model and/or the variance model were successfully selected for the cases discussed in this work.

The selected regression method should be as simple as possible. In that sense, OLS is attractive. It is based on the fact that the dependent variable is normally distributed with a constant variability. In the examples studied, only slight departures from normality were observed. The magnitudes of these departures do not seem enough to compromise the underlying hypotheses of OLS. On the other hand, in this

work high relative variability of u_c estimates was found. This is in agreement with the statements of GUM [5] that only one or two significant figures should be used to express u_c . With OLS the hypothesis concerning constant variability of the dependent variable is clearly violated. As alternative, WLS takes into account the lack of constancy of u_c variability. But, when the points follow the relationship expressed by the previous selected model, there are no real differences between the results of predictions made with OLS and WLS. The fundamental goal during modelling u_c versus concentration is to perform point predictions of combined uncertainty. The variability of those predictions, i.e., the second-order uncertainty, something for which WLS is more realistic than OLS, does not matter.

However, both regression methods are very sensitive to outliers, something that is necessary to keep in mind in some cases. If there are points that do not follow the relationship expressed by the selected model, the predictions of u_c could deteriorate. If the suspicious point is at the high part of the concentration range, the use of WLS could be a proper way for modelling due to the low weight assigned to such a point. In contrast, if the suspicious point is at the lower end, because of the high assigned weight, the regression with WLS could not be favourable.

However, there are several additional tools that can be useful. In addition to the OLS and WLS, robust regression seems promising for solving some cases. The diagnosis of outliers can throw some light during the modelling process, a topic that has not been explored before. It is difficult to consider as an outlier a point obtained from a high number of independent experimental results, as used in the experimental designs applied in intra or interlaboratory conditions. However, in some cases it seems that there are objective bases to consider some points as outliers, taking into account the high variability of u_c estimates.

An important fact that it is necessary to consider is the sensitivity of the variance model to outliers, if compared with the linear model. Nonlinear regression of the logarithmic transformation (NLLT) could be a solution to that problem because the results of predictions of u_c with NLLT are very similar to those obtained with RMM. However, NLLT is complex computationally, does not really make the variability of u_c constant, and compared with RMM does not yield a considerable improvement in predictions.

Finally, if the above guidelines do not solve the problem at hand, it is preferable to provide overestimations of u_c , which are more realistic estimates of combined uncertainty instead of underestimations which provide a false impression about the laboratories' measurement capabilities.

Acknowledgments The authors express their gratitude to A. Boza, A. Montero, S. Alleyne, and O. Collazo for their help in part of the

experimental labour and to Professor Carlos Bouza Herrera for his comments.

References

- IUPAC (2002) Pure Appl Chem 74:835–855. doi:[10.1351/pac200274050835](https://doi.org/10.1351/pac200274050835)
- Lee JC, Ramsey MH (2001) Analyst (Lond) 126:1784–1791. doi:[10.1039/b104946c](https://doi.org/10.1039/b104946c)
- ISO/IEC 17025:2005 Requisitos generales para la competencia de los laboratorios de ensayo y calibración, ISO, Geneva, Switzerland (certified translation)
- International Laboratory Accreditation Conference (1996) Guidelines on assessment and reporting of compliance with specification, ILAC-G8
- ISO (1995) Guide to the expression of uncertainty in measurement, Geneva, Switzerland (ISBN 92-67-10188-9)
- EURACHEM/CITAC (2000) Guide quantifying uncertainty in analytical measurement, 2nd edn
- Royal Society of Chemistry, Analytical Methods Committee (1995) Analyst (Lond) 120:2303–2308. doi:[10.1039/an9952002303](https://doi.org/10.1039/an9952002303)
- ISO/TS 21748:2004 Guidance for the use of repeatability, reproducibility and trueness estimates in measurement uncertainty estimation, ISO, Geneva, Switzerland
- Zitter H, God C (1971) Fresenius Z Anal Chem 255:1–9. doi:[10.1007/BF00423819](https://doi.org/10.1007/BF00423819)
- Thompson M (1998) Analyst (Lond) 113:1579–1587. doi:[10.1039/an9881301579](https://doi.org/10.1039/an9881301579)
- ISO 5725-2:1994 Accuracy (trueness and precision) of measurements and results—Part 2, ISO, Geneva, Switzerland
- Thompson M, Lowthian PJ (1997) J AOAC Int 80:676–679
- Hughes H, Hurley PW (1987) Analyst (Lond) 112:1445–1449. doi:[10.1039/an9871201445](https://doi.org/10.1039/an9871201445)
- Rocke DM, Lorenzato S (1995) Technometrics 37:176–184. doi:[10.2307/1269619](https://doi.org/10.2307/1269619)
- Heydorn K, Anglov T (2002) Accred Qual Assur 7:153–158. doi:[10.1007/s00769-002-0440-8](https://doi.org/10.1007/s00769-002-0440-8)
- Thompson M, Wood R (2006) Accred Qual Assur 10:471–478. doi:[10.1007/s00769-005-0040-5](https://doi.org/10.1007/s00769-005-0040-5)
- American Public Health Association (1998) Standard methods for the examination of water and waste water, 18th edn, Washington DC
- ISO 5725-3:1994, Accuracy (trueness and precision) of measurements and results—Part 3, ISO, Geneva, Switzerland
- Barnett WB (1984) Spectrochim Acta [A] 39b:829–836
- Barwick BV, Ellison SL (1999) Analyst (Lond) 124:981–990. doi:[10.1039/a901845j](https://doi.org/10.1039/a901845j)
- Moroto A, Jordi R (1999) Anal Chim Acta 391:173–175. doi:[10.1016/S0003-2670\(99\)00111-7](https://doi.org/10.1016/S0003-2670(99)00111-7)
- Wehrens R, Putter H, Buydens LM (2000) Chemom Intell Lab Syst 54:35–52. doi:[10.1016/S0169-7439\(00\)00102-7](https://doi.org/10.1016/S0169-7439(00)00102-7)
- Royal Society of Chemistry, Analytical Methods Committee (2001) No 8
- The MathWorks Inc. (2002) MatLab The language of technical computing, version 6.5
- ISO 5725-4:1994, Accuracy (trueness and precision) of measurements and results—Part 4, ISO, Geneva, Switzerland
- ISO 10700:1994 Steel and iron. Determination of manganese content, Flame atomic absorption spectrometric method, ISO, Geneva, Switzerland
- Statistical Graphics Corp. (2000) Statgraphics Plus for Windows, version 5.1
- Miller JN, Miller JC (2002) Estadística y Quimiometría para Química Analítica, Prentice Hall, Madrid, 4th edn

29. Gao Z, He X, Zhang G, Li Y, Wu X (1999) *Talanta* 49:331–337. doi:[10.1016/S0039-9140\(98\)00383-X](https://doi.org/10.1016/S0039-9140(98)00383-X)
30. Massart DL, Vandeginste BGM, Buydens LMC, de Jong S, Lewi PJ (1997) Smeyers-Verbeke J “Handbook of chemometrics and qualimetrics” Part A. Elsevier, Amsterdam
31. Siegel AF (1982) *Biometrika* 69:242–244. doi:[10.1093/biomet/69.1.242](https://doi.org/10.1093/biomet/69.1.242)
32. Rousseuw PJ (1984) *J Am Stat Assoc* 79:871–880. doi:[10.2307/2288718](https://doi.org/10.2307/2288718)