

# Application of consistency checking to evaluation of uncertainty in multiple replicate measurements

Rüdiger Kessel · Michael Berglund ·  
Roger Wellum

Received: 25 October 2007 / Accepted: 13 February 2008 / Published online: 18 March 2008  
© Springer-Verlag 2008

**Abstract** Use of repeated measurements in quantitative chemical analysis is common but leads to the problem of how to combine the measurement values and produce a result with an uncertainty following the GUM. There is often confusion between repeated indications or observations of an input quantity, for whose uncertainty the GUM prescribes a type A evaluation, and complete measurements repeated on multiple sub-samples, as considered here. A solution for combining repeated measurement results and their individual uncertainties based on simple interval logic is proposed here. The individual measurement values and their uncertainties are compared with the calculated average value to see if this implies that another, possibly unknown, source of uncertainty is present. The model of the individual results is modified for this possible between-replicate effect so that the repeated measurements are consistent. Lack of consistency is a strong indication that the measurement is not fully under control and needs further development or investigation. This is not always possible, however and the method given here is proposed to ensure that the values of the repeated measurements agree with each other. A simple numerical example is given showing how the method can be implemented in practice.

**Keywords** Uncertainty of measurement · Multiple replicates · Consistency test · Repeated measurements

## Introduction

Chemical measurements are often repeated several times to improve the trust in the result. Analysts know from experience that it is difficult to control every aspect of a chemical measurement. Analysing multiple replicates helps them to have more confidence in the results.

The difference between multiple replicates and type A evaluation of an observed data series according to the GUM [1–3] is that multiple replicates are individual and complete determinations of the result obtained by measuring separate sub-samples while a type A evaluation is used to obtain the standard uncertainty for a single contributing input quantity in the evaluation of one result. In this paper we discuss the case of multiple measurements where the complete measurement procedure is carried out multiple times.

From the knowledge point of view the question is: what have we learned by redoing the measurement several times? The answer is: not much. If we have a well-described evaluation, we can expect nearly the same uncertainty for the result every time. Doing it once is then as good as doing it many times. As a consequence the uncertainty should not change significantly if a well-controlled measurement is repeated.

What is then the motivation to do multiple replicates? One good motivation for multiple replicates can be to test the assumptions on which the evaluation is based. This does not improve the knowledge about the value of the result but it improves the trust we have in the result. The assumption is that a repeated measurement of the same

---

R. Kessel (✉)  
Metrodata GmbH, Im Winkel 15-1,  
79576 Weil am Rhein, Germany  
e-mail: r.kessel@metrodata.de

M. Berglund · R. Wellum  
Institute for Reference Materials and Measurements,  
EC-JRC, Retieseweg 111, 2440 Geel, Belgium

measurand with the same procedure leads to results that have a high degree of equivalence.

### Consistency test for multiple replicates

Chemical measurements are inconsistent if the results of multiple replicates of the same measurement are not equivalent. Possible instabilities and problems can be found by testing this. A full uncertainty budget has to be calculated for every replicate. Usually the arithmetic mean is used to combine the results of the replicates.

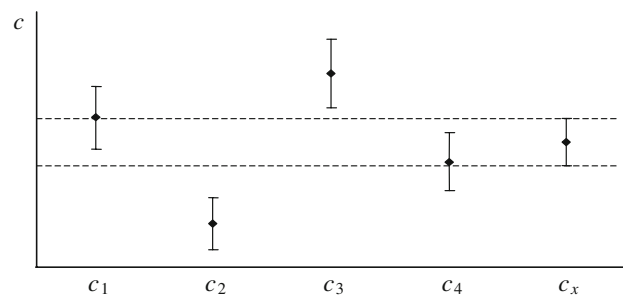
Because the same reference and calibration materials are used for all replicates, the results of the replicates are correlated [4]. This correlation is usually not taken into account. But it means that simple averaging would lead to an underestimation of the uncertainty. One possible solution is to apply a typical uncertainty for one replicate to the mean value of all replicates. The disadvantage of this method is that it is up to the analyst performing the measurement to decide which replicate was representative for the measurement and which uncertainty should be used for the mean value. It also violates the basic rule of the GUM that the uncertainty should be propagated from the input to the result quantity following the result value evaluation. An assignment of the uncertainty of the result is not foreseen in the GUM. This problem does not appear to have been tackled yet (a literature search on the topic was negative) although it can be a source of much confusion, especially for those new to applying the GUM.

If the correlation between the replicates is taken into account in calculating the uncertainty of the result (average), then all replicates can contribute to the uncertainty. No selection of a typical replicate is necessary and the law of propagation is used.

Sometimes the complicated nature of chemical measurements gives rise to an unexpected spread of the replicates. This spread is not normally included in the propagated uncertainty. It is essential that the results of the replicates are checked for any unexpected variation. An often used method is to plot the results with an uncertainty bar and to check graphically if the expanded uncertainty intervals overlap.

Figure 1 shows the plot of four replicates  $c_1$ – $c_4$  with expanded uncertainty bars and the arithmetic mean,  $c_x$ . The intervals of  $c_1$  and  $c_4$  overlap with the interval of  $c_x$  and the others do not. This graphical method has two disadvantages. First, it is difficult to incorporate any correlation and, second, the probability that two results are equivalent is rather small in the case that the intervals only slightly overlap.

A better solution is to perform a numerical equivalence check on the results of the replicates. This is similar to the well-known statistical  $t$  test [5]:



**Fig. 1** Plot of the results  $c_1$ – $c_4$  with expanded uncertainty bars ( $k = 2$ ) and average  $c_x$

$$|Y_i - \bar{Y}| \leq k \cdot u(Y_i - \bar{Y}) \quad (\text{usually } k = 2) \quad (1)$$

All correlations between the replicates should be included in the calculation of  $u(Y_i - \bar{Y})$  and it should be noted that  $Y_i$  and  $\bar{Y}$  are always correlated. The test shows whether the variation between the results of the replicates is smaller than, or equal to, the uncertainty introduced by all statistically independent input quantities. This can be understood if the difference between the average and the replicates is introduced as a new measurand  $\varepsilon_i$ :

$$\varepsilon_i = Y_i - \bar{Y} \quad (2)$$

Although the test looks similar to the  $t$  test, the interpretation is different. Instead of a confidence interval an expanded uncertainty interval is constructed based on the standard uncertainty of  $\varepsilon_i$ . The coverage factor  $k = 2$  represents a probability of about 95% if  $\varepsilon_i$  is normally distributed. Using the constructed interval, simple interval logic can be used to compare the interval with a fixed value without uncertainty. A difference is significant if the interval does not include the value zero. As a consequence, if the assumption of no difference is not compatible with this interval then the check fails.

If this test fails there is a strong implication (about 95% probability) of an unknown effect between the replicates. The measurement should be investigated in order to find and to correct the effect. Any result found by a proper execution of the measurement procedure is as valid as any other. Therefore it is not possible to exclude results because their value may not agree with others.

If resources are not available to study this effect in detail, an additional quantity should be added to every replicate in the evaluation. It should be indicated that the quantities are included because of an unknown between-replicate effect. The value of all these quantities is zero. The analyst evaluates the standard uncertainty of these added quantities (type B evaluation according to the GUM method). In the absence of additional knowledge of the reliability of individual replicates, the influence of the between-replicate effect should be considered the same for all replicates.

There are different ways of including a between-replicate effect in the final uncertainty statement. They share in common an additional quantity that has to be introduced in the model equation to represent the effect. The uncertainty can be evaluated in various ways. It is the responsibility of the analyst to choose the method of estimating the size of the effect. The basic requirement is that the method is clearly defined, transparent, and consistent. In some fields additional requirements arise from best practice or written standards.

The uncertainty of the between-replicate effect is often evaluated from the experimental standard deviation of the result values  $Y_i$ . This evaluation is usually based on a small series of values and does not separate the between-replicate and the within-replicate variation. Another possible method is calculation of a variance analysis (ANOVA [6]), but analysis of variance performed with only a small number of values is not very useful.

An additional criterion for estimation of the unknown between-replicate effect can be derived from the equivalence check in Eq. (1).

The uncertainty of the unknown effect can be chosen so that the equivalence check will just be passed. This value will be dependent on how the results were combined. If a non-weighted arithmetic mean is used, the minimum additional uncertainty  $u(\delta_{\text{between}})$  that should be added to the  $n$  replicates can be calculated from the equation:

$$u(\delta_{\text{between}}) \geq \sqrt{\frac{n}{n-1} \cdot \left[ \left( \frac{Y_{i, \max} - \bar{Y}}{k} \right)^2 - u^2(Y_{i, \max} - \bar{Y}) \right]} \quad (\text{usually } k = 2) \quad (3)$$

For  $Y_{i, \max}$  the result of the replicate with the maximum deviation from the mean value is used (see also Appendix 2). Alternatively, Eq. (3) can be calculated for all results which did not pass the equivalence check and the largest value for  $u(\delta_{\text{between}})$  is the lower limit for the uncertainty to pass the check.

The equation does not give any information about the between-replicate effect. It only gives a minimum value for the uncertainty that needs to be added so that all replicates will pass this particular check. It is the responsibility of the analyst to decide which value is reasonable to cover the possible between-replicate effect.

### Interpretation of the consistency test

It is important to interpret the result of this test carefully. If the difference is always smaller than the uncertainty-based interval and the test passes then there is no indication of an uncontrolled effect. It does not mean that the result is correct. It only means that it was not possible to show the presence of a problem by repeating the measurement. If at

least one difference is greater than the interval and the test fails then it is not possible to judge from the results where the test fails. The only conclusion that can be drawn is that the evaluation is not consistent. It is then necessary to investigate the measurement in detail to find the reason for the inconsistency. A result can be excluded if, for instance, information in the measurement protocol indicates a problem during the measurement. Possible reasons in chemical measurements are contamination, misreading of balances or other instruments, losses because of procedure failure, and unidentified uncertainty contributors. Inhomogeneity of the sample can also be a cause. If the reason cannot be identified and the effect cannot be corrected, the uncertainty of the result of every replicate should be increased.

It is important to be clear that the test does not allow us to judge the result of individual replicates. The uncertainty of all results should therefore be increased. We know that the unknown effect is a between-replicate effect. It has a different value for every replicate, because if it always had the same value we could not recognize it by repeating the measurement. The estimate of the unknown effect has to be independent for every replicate. As a consequence separate quantities should be introduced for every result to model this independence.

It would not be correct to simply add the same quantity (same name) to the model of all the replicates. This would model a common effect that is only known within the uncertainty but is always constant and has the same value. Even if we do not know the value exactly we would need to know that it is always the same from one replicate to the next. This is not the case, however, for the between-replicate effect.

For the value and uncertainty of the individual quantity which is added to the model of all replicates, the following reasoning should be used: As long as we have no more information, the best estimate for this quantity or effect is zero. The uncertainty of all quantities should be the same, because we do not have the information needed to know where the effect is possibly higher and where lower.

### Calculation scheme

The general model for evaluation of  $m$  multiple replicates depending on  $n$  input quantities can be written in the form:

$$\begin{aligned} Y_1 &= f_1(X_1, \dots, X_n) \\ &\vdots \\ Y_m &= f_m(X_1, \dots, X_n) \end{aligned} \quad (4)$$

Often the same model function is used for all replicates.

$$f_i(\cdot) \equiv f(\cdot) \quad (i = 1, \dots, m) \quad (5)$$

The arithmetic mean is the best estimate of the combined result if the uncertainties of the replicates are similar, which is usually the case.

$$\bar{Y} = \frac{1}{m} \cdot \sum_{i=1}^m Y_i \tag{6}$$

The consistency test can be applied by calculating the differences between the average and each replicate result  $\varepsilon_i$ .

$$\begin{aligned} \varepsilon_1 &= Y_1 - \bar{Y} \\ &\vdots \\ \varepsilon_m &= Y_m - \bar{Y} \end{aligned} \tag{7}$$

Each  $\varepsilon_i$  should fulfil the condition:

$$|\varepsilon_i| \leq k \cdot u(\varepsilon_i) \quad (\text{usually } k = 2) \tag{8}$$

If the test fails, the model of evaluation for the multiple replicates should be modified:

$$\begin{aligned} Y_1 &= f_1(X_1, \dots, X_n) + \delta X_{1,\text{rep}} \\ &\vdots \\ Y_m &= f_m(X_1, \dots, X_n) + \delta X_{m,\text{rep}} \end{aligned} \tag{9}$$

The additional term  $\delta X_{i,\text{rep}}$  represents the between-replicate effect. The expectation value of this quantity is zero and the standard uncertainty is the same for all replicate results. The minimum value for the uncertainty can be calculated by use of Eq. (3).

**Example of a multiple replicate evaluation**

The following example will illustrate the evaluation of a determination with four replicates. From a sample solution four aliquots have been taken. They have been measured with a linear instrument having a known relationship between the concentration of interest and the indication of the instrument. The instrument has been calibrated and the calibration factor is known from previous measurements. A typical example of this situation would be the determination of concentration using an optical absorption measurement. To keep the example as simple as possible all other influences are assumed to be negligible. The results for the replicates  $c_i$  can be evaluated with the model:

$$\begin{aligned} c_1 &= K_{\text{cal}} \cdot I_1 + \delta c_{1,\text{rep}} \\ c_2 &= K_{\text{cal}} \cdot I_2 + \delta c_{2,\text{rep}} \\ c_3 &= K_{\text{cal}} \cdot I_3 + \delta c_{3,\text{rep}} \\ c_4 &= K_{\text{cal}} \cdot I_4 + \delta c_{4,\text{rep}} \end{aligned} \tag{10}$$

$K_{\text{cal}}$  is the common calibration factor measured before and is evaluated with the type B method in this calculation.  $I_1 \dots I_4$  are the observations for the replicates evaluated with

the method type A. The  $\delta c_{i,\text{rep}}$  terms are introduced to model a possible between-replicate effect. In the beginning this effect is assumed to be zero with zero uncertainty (constant).

The best estimate for the sample is the average of the replicate results  $c_x$ .

$$c_x = \frac{c_1 + c_2 + c_3 + c_4}{4} \tag{11}$$

For the consistency test the differences are calculated:

$$\begin{aligned} \varepsilon_1 &= c_1 - c_x \\ \varepsilon_2 &= c_2 - c_x \\ \varepsilon_3 &= c_3 - c_x \\ \varepsilon_4 &= c_4 - c_x \end{aligned} \tag{12}$$

For this example, the following data values are used:

Calibration factor	$K_{\text{cal}} = 1.176 \cdot 10^{-6} \pm 0.047 \cdot 10^{-6}$ ( $k_{p95} = 2$ )
Intensity of replicate 1	$I_1 = 5.35, u(I_1) = 0.11$
Intensity of replicate 2	$I_2 = 4.352, u(I_2) = 0.087$
Intensity of replicate 3	$I_3 = 5.81, u(I_3) = 0.12$
Intensity of replicate 4	$I_4 = 4.953, u(I_4) = 0.099$

The following results and their correlation coefficients were calculated by use of a software tool [7] based on the approach presented in Ref. [8], assuming the value of the between-replicate term is zero and has no uncertainty:

$$\begin{aligned} c_1 &= 6.29 \cdot 10^{-6} \pm 0.36 \cdot 10^{-6} \quad (k_{p95} = 2) \\ c_2 &= 5.12 \cdot 10^{-6} \pm 0.29 \cdot 10^{-6} \quad (k_{p95} = 2) \\ c_3 &= 6.84 \cdot 10^{-6} \pm 0.39 \cdot 10^{-6} \quad (k_{p95} = 2) \\ c_4 &= 5.83 \cdot 10^{-6} \pm 0.33 \cdot 10^{-6} \quad (k_{p95} = 2) \\ c_x &= 6.02 \cdot 10^{-6} \pm 0.27 \cdot 10^{-6} \quad (k_{p95} = 2) \end{aligned}$$

$r(c_i, c_j)$	$c_1$	$c_2$	$c_3$	$c_4$	$c_x$
$c_1$	1	0.49	0.49	0.49	0.79
$c_2$	0.49	1	0.49	0.50	0.76
$c_3$	0.49	0.49	1	0.49	0.81
$c_4$	0.49	0.50	0.49	1	0.78
$c_x$	0.79	0.76	0.81	0.78	1

$$\begin{aligned} |\varepsilon_1| &= 270 \cdot 10^{-9}, U(\varepsilon_1) = 220 \cdot 10^{-9} \\ &\quad (k_{p95} = 2) \text{ (consistency test failed)} \\ |\varepsilon_2| &= 900 \cdot 10^{-9}, U(\varepsilon_2) = 190 \cdot 10^{-9} \\ &\quad (k_{p95} = 2) \text{ (consistency test failed)} \\ |\varepsilon_3| &= 820 \cdot 10^{-9}, U(\varepsilon_3) = 240 \cdot 10^{-9} \\ &\quad (k_{p95} = 2) \text{ (consistency test failed)} \\ |\varepsilon_4| &= 190 \cdot 10^{-9}, U(\varepsilon_4) = 210 \cdot 10^{-9} \\ &\quad (k_{p95} = 2) \text{ (consistency test passed)} \end{aligned}$$

The consistency test failed for three replicates. From the measurement results we have good reason to doubt the assumption that no between-replicate effect is present. After assigning a standard uncertainty of  $u(\delta c_{i,\text{rep}}) = 0.51 \cdot 10^{-6}$  (calculated by applying Eq. 3) to the quantities representing the between-sample effect, the calculated results are:

$$\begin{aligned} c_1 &= 6.3 \cdot 10^{-6} \pm 1.1 \cdot 10^{-6} \quad (k_{p95} = 2) \\ c_2 &= 5.1 \cdot 10^{-6} \pm 1.1 \cdot 10^{-6} \quad (k_{p95} = 2) \\ c_3 &= 6.8 \cdot 10^{-6} \pm 1.1 \cdot 10^{-6} \quad (k_{p95} = 2) \\ c_4 &= 5.8 \cdot 10^{-6} \pm 1.1 \cdot 10^{-6} \quad (k_{p95} = 2) \\ c_x &= 6.02 \cdot 10^{-6} \pm 0.58 \cdot 10^{-6} \quad (k_{p95} = 2) \end{aligned}$$

$r(c_i, c_j)$	$c_1$	$c_2$	$c_3$	$c_4$	$c_x$
$c_1$	1	0.04	0.06	0.05	0.54
$c_2$	0.04	1	0.05	0.04	0.52
$c_3$	0.06	0.05	1	0.05	0.55
$c_4$	0.05	0.04	0.05	1	0.53
$c_x$	0.54	0.52	0.55	0.53	1

$$\begin{aligned} |\varepsilon_1| &= 270 \cdot 10^{-9}, U(\varepsilon_1) = 910 \cdot 10^{-9} \\ &\quad (k_{p95} = 2) \text{ (consistency test passed)} \\ |\varepsilon_2| &= 900 \cdot 10^{-9}, U(\varepsilon_2) = 900 \cdot 10^{-9} \\ &\quad (k_{p95} = 2) \text{ (consistency test passed)} \\ |\varepsilon_3| &= 820 \cdot 10^{-9}, U(\varepsilon_3) = 910 \cdot 10^{-9} \\ &\quad (k_{p95} = 2) \text{ (consistency test passed)} \\ |\varepsilon_4| &= 190 \cdot 10^{-9}, U(\varepsilon_4) = 910 \cdot 10^{-9} \\ &\quad (k_{p95} = 2) \text{ (consistency test passed)} \end{aligned}$$

Comparing the results it is obvious that the best estimates do not change. The reason is that the best estimate of the between-replicate effect  $\delta c_{i,\text{rep}}$  is zero. The uncertainty of the unknown concentration  $c_x$  has more than doubled. This uncertainty includes the assumption that there is a possible between-replicate effect. Since the effect was not investigated in detail, it could not be corrected for. Only the minimum uncertainty to ensure that the calculation is consistent was added. If more knowledge about the effect were available, it should have been used. This approach follows the principle that a reasonable solution should be used. In this case, as a “side-effect”, the correlation of the results  $c_1$  to  $c_4$  is reduced to a level which could be ignored. By adding independent quantities to all results the statistical dependants or correlation will always be reduced.

The uncertainty of the between-replicate effect  $u(\delta c_{i,\text{rep}})$  could also be estimated from the experimental standard deviation of the values for  $c_1$ – $c_4$ . The experimental standard deviation for this series is  $s(c_i) = 0.73 \cdot 10^{-6}$ . This value is higher, but of the same order of magnitude found previously with Eq. (3). This estimation is only based on a series of four values. It leads to an overestimation, because the within-replicate variation would be counted twice.

### Summary

The procedure discussed above can be generalised. Multiple replicate measurements should be checked for possible between-replicate effects. Possible correlations between replicates should be investigated and should be included in the check. Any check can be used as long as it is fit for purpose and clearly defined.

If the test fails then additional quantities should be added to the replicates in order to model the (unknown) between-replicate effect. The effect should, of course, be investigated and corrected if possible. If the effect cannot be corrected then an uncertainty should be estimated using scientific judgement. The minimum uncertainty that should be added can be derived from the check criterion as described above.

**Acknowledgments** We thank Raghu Kacker (NIST) for fruitful discussions and comments.

### Appendix 1: The symbols used in this paper

$Y_i$	Value of the measurands; variable representing the state of knowledge
$u(y_i)$	Standard uncertainty associated with result $Y_i$
$X_i$	Input quantity; variable representing the state of knowledge
$u(x_i)$	Standard uncertainty associated with the input quantity $X_i$
$r(x_i, x_j)$	Correlation coefficient between $X_i$ and $X_j$
$c_i$	Concentration for replicate $i$ (results of the measurements)
$c_x$	Average concentration (end result)
$\varepsilon_i$	Difference between single result $c_i$ and the average $c_x$
$I_i$	Observed intensity (detector output) for replicate $i$
$\delta c_i$	Small deviation of the concentration with expectation zero

### Appendix 2: Calculation of the minimum additional uncertainty for the equivalence check

In cases when a non-weighted arithmetic mean is used to combine measurement results, a general equation for the minimum additional uncertainty  $u(\delta_i)$  that should be added to the  $n$  replicates can be derived. One contribution  $Y_m$  can be separated from the arithmetic mean.

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i = \frac{Y_m}{n} + \frac{1}{n} \sum_{i \neq m} Y_i \quad (13)$$

The difference between the average and  $Y_m$  is:

$$\varepsilon_m = Y_m - \bar{Y} \quad (14)$$

Combining Eqs. (13) and (14) gives:

$$\varepsilon_m = Y_m - \left[ \frac{Y_m}{n} + \frac{1}{n} \sum_{i \neq m} Y_i \right] = \frac{n-1}{n} \cdot Y_m - \frac{1}{n} \sum_{i \neq m} Y_i \quad (15)$$

An independent uncertainty component is added to all replicates, which leads to a second difference  $\varepsilon_m^*$ :

$$\begin{aligned} \varepsilon_m^* &= \frac{n-1}{n} \cdot (Y_m + \delta_m) - \frac{1}{n} \sum_{i \neq m} (Y_i + \delta_i) \\ &= \left[ \frac{n-1}{n} \cdot Y_m - \frac{1}{n} \sum_{i \neq m} Y_i \right] + \left[ \frac{n-1}{n} \cdot \delta_m - \frac{1}{n} \sum_{i \neq m} \delta_i \right] \end{aligned} \quad (16)$$

The two terms in Eq. (16) are obviously independent and the left term is equal to difference without the additional terms. The uncertainty of  $\varepsilon_m^*$  can be calculated.

$$u^2(\varepsilon_m^*) = u^2(\varepsilon_m) + u^2\left(\frac{n-1}{n} \cdot \delta_m - \frac{1}{n} \sum_{i \neq m} \delta_i\right) \quad (17)$$

$$\begin{aligned} u^2\left(\frac{n-1}{n} \cdot \delta_m - \frac{1}{n} \sum_{i \neq m} \delta_i\right) &= \left(\frac{n-1}{n}\right)^2 \cdot u^2(\delta_m) \\ &\quad + \frac{1}{n^2} \sum_{i \neq m} u^2(\delta_i) \end{aligned} \quad (18)$$

All uncertainties of the additional terms have the same value  $u(\delta)$ .

$$\begin{aligned} u^2\left(\frac{n-1}{n} \delta_m - \frac{1}{n} \sum_{i \neq m} \delta_i\right) &= \left(\frac{n-1}{n}\right)^2 u^2(\delta) + \left(\frac{n-1}{n^2}\right) u^2(\delta) \\ &= \frac{(n-1)^2 + n-1}{n^2} u^2(\delta) \\ &= \frac{n-1}{n} u^2(\delta) \end{aligned} \quad (19)$$

Combining Eqs. (17) and (19) leads to:

$$u^2(\varepsilon_m^*) = u^2(\varepsilon_m) + \frac{n-1}{n} \cdot u^2(\delta) \quad (20)$$

Equation (20) solved for  $u(\delta)$ :

$$\begin{aligned} u^2(\delta) &= \frac{n}{n-1} \cdot [u^2(\varepsilon_m^*) - u^2(\varepsilon_m)] \\ u(\delta) &= \sqrt{\frac{n}{n-1} \cdot [u^2(\varepsilon_m^*) - u^2(\varepsilon_m)]} \end{aligned} \quad (21)$$

With Eq. (21) the uncertainty value of an additional between-replicate term can be calculated. This is added to all replicates if the uncertainty of the epsilon without the term and the uncertainty of the epsilon with the term are known. The term  $u(\varepsilon_m)$  can be calculated from the equivalence check in Eq. (1) under the assumption that the check is just passed.

$$u^2(\varepsilon_m^*) \geq \left(\frac{Y_m - \bar{Y}}{k}\right)^2 \quad (\text{usually } k = 2) \quad (22)$$

Combining Eq. (21) and (22) leads to:

$$\begin{aligned} u(\delta) &\geq \sqrt{\frac{n}{n-1} \cdot \left[ \left(\frac{Y_m - \bar{Y}}{k}\right)^2 - u^2(Y_m - \bar{Y}) \right]} \\ &(\text{usually } k = 2) \end{aligned} \quad (23)$$

For  $Y_m$  the replicate should be selected for which  $\varepsilon_m$  in respect of  $k \cdot u(\varepsilon_m)$  is maximum.

## References

1. Ellison SLR, Rösslein M, Williams A, Eurachem/Citac (eds) (2000) Quantifying uncertainty in analytical measurement, 2nd edn. <http://www.measurementuncertainty.org/mu/QUAM2000-1.pdf>
2. European co-operation for Accreditation (1999) Expression of the uncertainty of measurement in calibration, EA, EA-4/02. <http://www.european-accreditation.org>
3. International Organization for Standardization (1995) Guide to the Expression of Uncertainty in Measurement, 2nd edn. International Organization for Standardization, Geneva, ISBN 92-67-10188-9
4. Kessel R (2003) A novel approach to uncertainty evaluation of complex measurements in isotope chemistry, Dissertation. University of Antwerp, Antwerp
5. Box GEP, Hunter WG, Hunter JS (1978) Statistics for experimenters. Wiley, NY, ISBN 0-471-09315-7
6. ISO: GUIDE 35 (1989) Certification of reference materials, 2nd edn. International Organisation for Standardisation, Geneva
7. GUM Workbench Version 2.3.6 2005 Metrodata GmbH, Im Winkel 15-1, D-79576, Weil am Rhein, Germany. <http://www.metrodata.de>
8. Kessel R, Berglund M, Taylor P, Wellum R (2000) How to treat correlation in the uncertainty budget, when combining results from different measurements, paper presented at AMCTM 2000, ISBN 981-02-4494-0