Co-funded by the European Union

EUROPEAN UNION

COUNCIL OF EUROPE

CONSEIL DE L'EUROPE

Co-funded and implemented by the Council of Europe

General European OMCL Network (GEON) QUALITY MANAGEMENT DOCUMENT

PA/PH/OMCL (18) 153 R3

EVALUATION OF MEASUREMENT UNCERTAINTY

ANNEX 2.5

Full document title and reference	Evaluation of Measurement Uncertainty ANNEX 2: ESTIMATION OF MEASUREMENT UNCERTAINTY USING TOP-DOWN APPROACH Annex 2.5: Use of Data from Proficiency Testing Studies for the Estimation of Measurement Uncertainty *PA/PH/OMCL (18) 153 R3*
Document type	Guideline
Legislative basis	Council Directive 2001/83/EC and 2001/82/EC, as amended
Date of first adoption	19 July 2019
Date of original entry into force	1 February 2020
Date of entry into force of revised document	1 September 2022
Previous titles/other references / last valid version	Replaces *PA/PH/OMCL (18) 153 R1CORR*
Custodian Organisation	The present document was elaborated by the OMCL Network / EDQM of the Council of Europe
Concerned Network	GEON

Annex 2 to Guideline "Evaluation of Measurement Uncertainty"
PA/PH/OMCL (18) 145 (in its current version)

Estimation of measurement uncertainty using Top-down approach

Annex 2.5 Use of data from Proficiency Testing Studies for the estimation of measurement uncertainty

Contents

1. Introduction

Proficiency testing (PT), as defined in ISO/IEC 17025, is the evaluation of participant performance against pre-established criteria by means of inter-laboratory comparisons (organization, performance and evaluation of measurements or tests on the same or similar items by two or more laboratories in accordance with predetermined conditions, ISO/IEC 17043). Therefore, as defined by Eurachem, a PT scheme (PTS) is a system for objectively evaluating a laboratory's performance, helping the participant to assess the accuracy of its measurements [1].

This annex provides examples of use of PTS data to estimate the measurement uncertainty in routine testing, assuming a normal distribution of the results. Therefore, standard deviations are used to reflect the expected variability between laboratory results, as the Target Standard Deviation (TSD) defined in the PT scheme organised by the EDQM, as well as the observed variability between measurements (within-laboratory SD or repeatability).

In order to estimate the uncertainty of measurement, the laboratory can rely on:

- **Its own PT results**

  The measurement uncertainty is evaluated using the within-laboratory SD (precision component) and the difference between the laboratory mean and assigned value (bias component). These two components can be calculated using data of the last PT round or by combining data of several rounds, as long as the performance of the laboratory remains consistent.

- **PT results of all participating laboratories**

  The precision and bias components are reflected by the within- and inter-laboratory standard deviations (combined into reproducibility) based on data from all participants. The two components can be calculated using data of the last PT round or by combining data of several PT rounds, as long as the performance of the laboratory remains consistent.

The standard uncertainty (u_c) is a combination of random and systematic errors, which affect the precision and the trueness (bias) of the method/assay, respectively [4]:

$$u_c = \sqrt{u^2_{precision} + u^2_{bias}} .$$

Random errors consist in differences between individual results and their mean. Therefore, $u_{precision}$ can be estimated using the replicated results of:

- One laboratory in one PT round: S_{round} may lack of representativeness due to the limited number of independent replicates (e.g. at different days, different sample preparations) usually performed in a PT round,
- One laboratory in several PT rounds: there are as many S_{round} estimates as PT rounds. These estimates can be pooled (S_{pool}) to be more representative of the laboratory precision.
- Several laboratories in one PT round: there are as many S_{round} estimates as participating laboratories. These estimates can be pooled (S_{pool*}) to be representative of the method/assay precision.

Note. Standard deviations can be pooled if homogeneous enough (e.g. Cochran's test). Pooling means calculating a weighted average, which is done using variances (squares of standard deviations).

Systematic errors (bias) consist in differences between mean values and expected/assigned values. Depending on the data used to estimate the bias, a mean difference or a standard deviation can be calculated:

- One laboratory in one PT round: the difference between the mean value and the expected/assigned value may lack of representativeness, due to the limited number of independent replicates (e.g. performed at different days, using different sample preparations),
- One laboratory in several PT rounds: the difference between the mean value and the expected/assigned value averaged over the different rounds is representative of the laboratory bias,
- Several laboratories in one PT round: the difference between the mean value and the expected/assigned value averaged over the different laboratories tends toward 0 (assuming that the assigned value is the robust mean of the round results). Therefore, the bias component

should be calculated as the inter-laboratory standard deviation, S_{inter}, which increases with the differences between the laboratory means and is representative of the method/assay bias.

According to some guidelines, the bias component may be excluded from the calculation of the measurement uncertainty, when it is non-significant. A one-sample mean comparison test can be performed, referred as to z-test when the standard deviation is known (t-test, otherwise). The calculated value is simply the |z-score|:

$$z = \frac{|Mean - Assigned|}{TSD}.$$

The significance threshold is equal to 1.96 (often rounded to 2) for a confidence level of P = 95% and a two-sided test. That is, the bias may be excluded following a 'satisfactory' performance. However, it is recommended to take the bias component into account when estimating the measurement uncertainty, regardless of the z-score achieved by the laboratory.

The potential disadvantage of using PT samples is the lack of traceable reference values similar to those of certified reference materials (CRM). Consensus values, in particular, are prone to occasional errors. This demands due care in their use for uncertainty estimation. When the uncertainty ($u_{assigned}$) around the consensus value is not negligible, it should be taken into account in the calculation of the measurement uncertainty [1].

Note. In the next sections, intermediate calculated results are rounded to a limited number of decimal places, for the sake of clarity and to make it easier to reproduce subsequent results. However, in practice, it is recommended to round final calculated results only (e.g. expanded uncertainty estimates).

2. How to calculate S_{round} and S_{pool} (several PT rounds by one laboratory)

As previously mentioned, the precision component ($u_{precision}$) can be estimated using the results of a single PT round. However, S_{round} may not be fully representative of laboratory variation, as based on a limited number of replicates. A more robust estimate can be obtained by pooling the standard deviations of several PT rounds (S_{pool}).

Table 1 shows the replicated results (n = 3) of 6 rounds of a same PT scheme performed by one laboratory. The within-round standard deviations, calculated according to Eq. 1, range from 13.1 mg to 23.5 mg (estimated with DF = n – 1 = 2 degrees of freedom).

Table 1. Data (mg) of 6 PT rounds (same PT scheme).

Round	Rep. 1	Rep. 2	Rep. 3	S_{round}	DF
1	986	941	975	23.5	2
2	765	791	780	13.1	2
3	958	987	970	14.6	2
4	913	917	945	17.4	2
5	883	857	894	19.0	2
6	837	821	808	14.5	2

The within-round standard deviation is equal to:

$$S_{round} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - mean)^2}{n-1}}. \qquad (1)$$

The 6 estimates of the within-round standard deviation are homogeneous (e.g. Cochran test p-value = 0.98) and can be combined into a single value (S_{pool}). A weighted mean of the 6 variances (S^2_{round}) is carried out according to Eq. 2, with weights equal to the degrees of freedom:

$$S_{pool} = \sqrt{\frac{\sum DF_j \cdot S^2_{round(j)}}{\sum DF_j}}. \quad (2)$$

$$S_{pool} = \sqrt{\frac{2 \cdot 23.5^2 + 2 \cdot 13.1^2 + 2 \cdot 14.6^2 + 2 \cdot 17.4^2 + 2 \cdot 19.0^2 + 2 \cdot 14.5^2}{(2+2+2+2+2+2)}} = 17.4 \text{ mg.}$$

Operations on variations should always be performed using variances (S^2). When the degrees of freedom are all equal, the weighted mean calculation becomes a simple (arithmetic) mean:

$$S_{pool} = \sqrt{\sum S^2_{round(j)} / n_j}. \qquad (3)$$

n_j = number of PT rounds.

3. How to calculate S_{inter}

The bias estimate can be reported as a mean difference or a standard deviation, S_{inter} reflecting the aggregation of z-scores in PTS terms. However, S_{inter} can not be calculated directly. The standard deviation of reproducibility (S_R) should be calculated first. When there is an equal number of replicates (n) per round for the laboratories participating in the PT round, S_R is equal to the standard deviation of the PT means.

Table 2 shows the replicated results (n = 3) of 6 participating laboratories to a PT round. The within-round standard deviations of the laboratories, calculated according to Eq. 1, range from 12.1 mg (Lab. 2) to 21.4 mg (Lab. 1), estimated with DF = n – 1 = 2 degrees of freedom.

Table 2. Data (mg) of 6 participants/laboratories to a PT round.

Lab.	Rep. 1	Rep. 2	Rep. 3	Mean	S_{round}	DF
1	878	840	876	865	21.4	2
2	864	887	882	878	12.1	2
3	859	894	873	875	17.6	2
4	812	827	843	827	15.5	2
5	880	849	886	872	19.9	2
6	831	826	806	821	13.2	2

S_R is equal to the standard deviation of the PT means and is calculated using the usual formula of the standard deviation (Eq. 1):

S_R = SD (865, 878, 875, 827, 872, 821) = 25.5 mg.

As S_R combines the bias (S^2_{inter}) and the precision for a given number of replicates (S^2_{pool*}/n):

$$S_R = \sqrt{S^2_{inter} + S^2_{pool*}/n}, \qquad (4)$$

the bias estimate can be calculated by rearranging Eq. 4:

$$S_{inter} = \sqrt{S_R^2 - S_{pool*}^2/n}. \qquad (5)$$

For the PT results in Table 2,

$$S_{inter} = \sqrt{25.5^2 - 17.0^2/3} = 23.5 \text{ mg}.$$

Notes.

- S_{pool*} is calculated using Eq. 2. However, it represents the assay variability of any laboratory (pooled across/averaged over different laboratories), while S_{pool} in Section 2 represents the assay variability of one particular laboratory (pooled across its own results to different rounds).

- In case of unequal numbers of replicates, the use of a statistical software is recommended (one-way random anova model) as the above formulae cannot be used anymore.

4. The laboratory relies on its own PT results

Table 3 shows the results of a participant to 6 PT rounds where the melting point (measurand) in °C is measured according to Ph. Eur. 2.2.14. In the first round, as an example, the laboratory reported a mean of 115.5 °C with a standard deviation between replicates equal to S_{round} = 0.26 °C. The assigned value and target standard deviation are 115.1 and 1.2 °C, respectively. Therefore, the z-score is equal to z = (115.5 – 115.1) / 1.2 = 0.33. The assigned value was calculated using the robust mean of the 45 participants' data and reported with a standard uncertainty of 0.08 °C.

Table 3. Results of a laboratory participating to 6 PT rounds on melting point (°C, same PT scheme).

	PT statistics			**Laboratory results**			
PT round	**Nb. Labs**	**Assigned value**	**$u_{assigned}$**	**Mean result**	**S_{round}**	**Bias**	**z-score**
1	45	115.1	0.08	115.5	0.26	0.4	0.33
2	51	160.0	0.10	160.1	0.31	0.1	0.08
3	44	228.8	0.04	229.0	0.15	0.2	0.17
4	47	184.2	0.07	183.9	0.24	-0.3	-0.25
5	50	210.7	0.05	211.2	0.18	0.5	0.42
6	44	177.6	0.07	177.1	0.27	-0.5	-0.42

n = 3 replicates per round. $u_{assigned}$: standard uncertainty of the assigned value.

PT scheme: overall management of PT rounds. A PT scheme is made of several rounds of the same type, e.g. melting points, loss on drying.

The decision to pool the results generated by same type of technique or method is based on the judgment and experience of the laboratory. For example, the results obtained on PTSs on HPLC assay and HPLC related substances can be pooled providing that the major uncertainty contributors are known and the results obtained are statistically homogenous. However, a similar approach is not necessarily applicable to biological PTSs. The nature of the samples, the purpose of the test, not only the technique or method in itself have to be considered.

4.1. Use of the results of one PT round

Table 4 shows the calculation steps of the combined standard uncertainty u_c considering the results of the first PT round.

$$u_c = \sqrt{u^2{}_{precision} + u^2{}_{bias}},$$

$u_{precision}$: variation between replicated results (within-round variation),

u_{bias}: combination of the mean bias and related errors.

u_c is estimated with and without the mean bias component, given the fact that it is not significant (z-score ≤ 2). It is up to the laboratory to choose one of the two calculated values, although the one including the mean bias is recommended in this guideline.

Table 4. Use of the results of the first PT round.

Component (°C)	First PT round
Precision (S_{round})	0.26
Mean bias (b)	0.4
Error of mean bias (u_b)	0.15
Combined standard uncertainty, u_c (b included)	0.50
Combined standard uncertainty, u_c (b excluded)	0.30

The mean bias (b) is equal to the difference between the mean and assigned value. The mean bias comes with some error (u_b), which combines the standard error of the laboratory mean (SE_m) and error of the assigned value (when applicable):

$$u_b = \sqrt{SE_m^2 + u_{assigned}^2} = \sqrt{S_{round}^2/n + u_{assigned}^2}\,. \qquad (6)$$

The error of the assigned value is negligible when it is lower than 0.3 × TSD [1]. For the first PT round, $u_{assigned}$ (0.08 °C) is lower than 0.3 × 1.2 = 0.36 °C and can be omitted in Eq. 6:

$$u_b = \sqrt{0.26^2/3} = 0.15\,°C.$$

The combined standard uncertainty, including the mean bias, is equal to:

$$u_c = \sqrt{S_{round}^2 + b^2 + u_b^2}, \qquad (7)$$

$$u_c = \sqrt{0.26^2 + 0.4^2 + 0.15^2} = 0.50\,°C.$$

The combined standard uncertainty is u_c = 0.30 °C following the exclusion of the mean bias (the error of the mean bias should be kept). The corresponding decrease is 40 % ((0.30 - 0.50) / 0.50), which may have some practical consequences. For this reason, it is recommended not to exclude the mean bias whatever the z-score of the laboratory. Note that u_c applies to individual results (n = 1). For the mean of n = 3 results, the precision component is S_{round} / SQRT(3) = 0.15 °C. Table 5 shows the expanded uncertainty (k = 2, P ≈ 95% confidence level) calculated for an individual result (n = 1) or the mean of n = 3 results.

Table 5. Expanded uncertainty for 2 testing formats.

Testing format	Expanded uncertainty (U)	
(e.g. in routine testing)	Mean bias included	Mean bias excluded
Single result (n = 1)	**± 1.0 °C**	**± 0.60 °C**
Mean of n = 3 independent results	**± 0.91 °C**	**± 0.42 °C**

Expanded uncertainty estimates are rounded to 2 significant digits (common practice).

4.2. Use of the results of several PT rounds

Table 6 shows the calculation steps of the combined standard uncertainty (u_c) considering the results of the 6 PT rounds. The precision component is the average of the within-round variations, calculated according to Eq. 2 (or Eq. 3 as there is an equal number of degrees of freedom per round).

$$S_{pool} = \sqrt{(2 \cdot 0.26^2 + \cdots + 2 \cdot 0.27^2)/(2 \cdot 6)} = 0.24\,°C.$$

The bias component should summarise the 6 individual biases (0.4, 0.1, 0.2 °C, etc.) as well as their errors. This is achieved by the formula:

$$u_b = \sqrt{RMS_b^2 + u_{assigned}^2}.$$

RMS_b is the root mean square of the individual biases calculated in Table 3:

$$RMS_b = \sqrt{\sum bias^2/n_{bias}},$$

$$RMS_b = \sqrt{(0.4^2 + 0.1^2 + 0.2^2 + 0.3^2 + 0.5^2 + 0.5^2)/6} = 0.37\,°C.$$

Note. The above formula is used for a same number of results per round. A weighted mean calculation should be used otherwise.

Moreover, $u_{assigned}$ is taken as the median of the standard uncertainties of the assigned values:

$u_{assigned\text{-}median}$ = median (0.08, 0.10, 0.04, 0.07, 0.05, 0.07) = 0.07 °C.

The median is less sensitive to potential 'outlying' results. However, the mean can also be used as suggested in some references [5].

Finally, the combined standard uncertainty reported in Table 6 is calculated as:

$$u_c = \sqrt{S_{pool}^2 + RMS_b^2 + u_{assigned}^2}.$$

As in the previous section, $u_{assigned}$ (0.07 °C) is lower than 0.3 × TSD (0.3 × 1.2 = 0.36 °C) and can be omitted.

Table 6. Use of the results of the 6 PT rounds.

Component (°C)	n = 1	n = 3
Precision (S_{pool})	0.24	0.14
Bias (RMS_b)	0.37	0.37
$u_{assigned}$	Negligible $u_{assigned-median}$	Negligible $u_{assigned-median}$
Combined standard uncertainty (u_c)	**0.44**	**0.40**

n = 1: for individual results, n = 3: for the mean of 3 individual results (precision = S_{pool} / SQRT(3))

Note. A relative standard uncertainty should be reported when the magnitude of the bias and/or TSD depend on (e.g. increase with) the assigned value. This may require a transformation (e.g. log) of the replicated results. All the calculation steps may be performed on log-transformed data and the final standard and expanded uncertainty estimates back transformed (e.g. exponentiation in case of a natural log transformation).

4.3. Reporting of measurement uncertainty

The expanded uncertainty of measurement of the laboratory (k = 2, P ≈ 95% confidence level), calculated for the various scenarios, rounded to 2 significance digits, is equal to:

- Last PT round, mean bias included: U = 1.0 ºC,
- Last PT round, mean bias excluded: U = 0.60 ºC,
- Pool of 6 PT rounds: U = 0.88 ºC.

5. The laboratory relies on all participant's results

Table 7 shows the results of a participant to 10 PT rounds where the density of a liquid (measurand) was measured according to Ph. Eur. 2.2.25. In this monograph, density is defined as the mass of a unit volume of the substance at 20 °C, expressed in grams per cubic centimetre. However, the results in Table 7 are reported in mg/cm^3 instead of g/cm^3 for convenience (less decimal points and leading zeros).

In the first round, as an example, the laboratory reported a mean of 1120 mg/cm^3. The assigned value and target standard deviation are 1119 and 2 mg/cm^3, respectively. Therefore, the z-score is equal to $z = (1120 – 1119) / 2 = 0.5$. The assigned value was calculated using the robust mean of the participants' data and was reported with a standard uncertainty of 0.15 mg/cm^3.

The standard deviation between replicates results (S_{round}) pooled across the different laboratories is equal to S_{pool^*} = 1.5 mg/cm^3. In addition, the standard deviation of the participants' means, excluding questionable and unsatisfactory results, i.e. |z-scores| > 2, is equal to S_R = 2.4 mg/cm^3.

Table 7. Results of a participating laboratory to 10 PT rounds.

	PT statistics					Lab. results	
PT round	Nb. Labs	Assigned Value	$u_{assigned}$	S_R	S_{pool*}	Mean Result	z-score
1	61	1119	0.15	2.4	1.5	1120	0.5
2	57	846	0.15	1.9	2.1	845	-0.5
3	47	865	0.15	2.2	1.9	862	-1.5
4	44	1261	0.15	3.1	2.8	1261	0.0
5	42	960	0.20	2.0	2.2	958	-1.0
6	57	912	0.20	2.7	1.6	911	-0.5
7	44	845	0.20	2.1	1.8	842	-1.5
8	50	918	0.20	3.1	2.5	915	-1.5
9	43	911	0.20	1.8	1.5	911	0.0
10	57	864	0.15	1.9	2.2	863	-0.5

n = 3 replicates per round; density results reported in mg/cm^3.

The 10 PT rounds are part of the same PT scheme.

TSD = 2 mg/cm^3 for the 10 PT rounds.

5.1. Use of the results of one PT round

The laboratory decides to estimate the uncertainty using the results of a single round, e.g. the last one (number 10). The reproducibility standard deviation (S_R) can be a suitable source of measurement uncertainty (following the exclusion of questionable and unsatisfactory results, i.e. |z-scores| > 2) as it includes the:

- Precision component (S_{pool*}: assay variability pooled across different laboratories) and,
- The mean bias component (S_{inter}: inter-laboratory variability, which comprises systematic effects due to differences in laboratory operations [6]).

$$S_R = \sqrt{S_{inter}^2 + S_{pool*}^2/n}.$$

According to the last PT round, S_R = 1.9 mg/cm^3. However, this value applies to the means of n = 3 replicates, i.e. the testing format of the PT round. In order to estimate the standard uncertainty of a measurement (u_c), S_{inter} should be calculated as an intermediate step using Eq. 5. For the last PT round:

$$S_{inter} = \sqrt{1.9^2 - 2.2^2/3} = 1.4 \text{ mg/cm}^3,$$

And

$$u_c = \sqrt{1.4^2 + 2.2^2} = 2.6 \text{ mg/cm}^3.$$

Therefore, the expanded uncertainty of measurement is equal to U = 5.2 mg/cm^3 (k = 2, P ≈ 95% confidence level).

In a PT round, the uncertainty of the assigned value ($u_{assigned}$) is usually considered as negligible if it is lower than 0.3 × TSD (0.3 × 2 mg/cm³ = 0.6 mg/cm³). For the 10th round, $u_{assigned}$ = 0.15 mg/cm³ is lower than the calculated threshold and thus negligible. However, should it be greater than the threshold, it should be added to the calculation of the standard uncertainty using Eq. 8:

$$u_c = \sqrt{S_{inter}^2 + S_{pool*}^2 + u_{assigned}^2}. \qquad (8)$$

Note.

The standard uncertainty (u_c) is applicable to an individual result. For the mean of n results, the precision component, S_{pool*}^2, should be divided by n.

When the uncertainty of the assigned value is not reported, it can be set to:

$$u_{assigned} = \frac{S_R}{\sqrt{N_P}}.$$

For the 10th round, the calculated value is 1.9 / SQRT(57) = 0.25 mg/cm³.

5.2. Use of the results of several PT rounds

The same approach as described in section 5.1 can be used, except that S_R is averaged over several PT rounds (Eq. 9). A weighted mean should be calculated with the number of participants of the ith round (P_i) – 1 as weight. At the end, the laboratory can expect a more robust estimate of the measurement uncertainty.

$$S_{R+} = \sqrt{\frac{\sum(P_i-1)\cdot S_{R(i)}^2}{\sum(P_i-1)}}, \qquad (9)$$

$$S_{R+} = \sqrt{\frac{60\cdot 2.4^2+56\cdot 1.9^2+\cdots+42\cdot 1.8^2+56\cdot 1.9^2}{60+56+\cdots+42+56}},$$

$$S_{R+} = 2.4 \text{ mg/cm}^3.$$

Similarly, S_{pool*} can be averaged over the PT rounds using Eq. 10, which is applicable to a same number of replicates (n) across the laboratories and PT rounds. The weight of each round is equal to $w_i = (n-1) \times P_i$. For the first round, as an example, $w_1 = (3-1) \times 61 = 122$.

$$S_{pool+} = \sqrt{\frac{\sum(n-1)P_i\cdot S_{pool*(i)}^2}{\sum(n-1)P_i}}. \qquad (10)$$

$$S_{pool+} = 2.0 \text{ mg/cm}^3 \text{ for } n = 3.$$

Next, in order to estimate the standard uncertainty of a measurement (u_c), S_{inter+} should be calculated as an intermediate step using Eq. 5:

$$S_{inter+} = \sqrt{2.4^2 - 2.0^2/3} = 2.1 \text{ mg/cm}^3,$$

And,

$$u_c = \sqrt{2.1^2 + 2.0^2} = 2.9 \text{ mg/cm}^3.$$

Therefore, the expanded uncertainty of measurement is equal to U = 5.8 mg/cm³ (k = 2, P ≈ 95% confidence level).

As in section 5.1, the uncertainty of the assigned value ($u_{assigned}$) can be added to the measurement uncertainty using Eq. 6, if it exceeds 0.3 × TSD. As in Section 4.2, $u_{assigned}$ can be set to the median of the uncertainty values, i.e. 0.175 mg/cm³ for the 10 PT rounds (negligible as below 0.3 × TSD = 0.6 mg/cm³).

5.3. Reporting of measurement uncertainty

The expanded uncertainty of measurement of the assay (k = 2, P ≈ 95% confidence level), calculated for the various scenarios is equal to:

- Last PT round: U = 5.2 mg/cm^3,
- Pool of 10 PT rounds: U = 5.8 mg/cm^3.

For comparative purposes, the expanded uncertainty calculated using the laboratory results only is also reported.

- Last PT round: U = 5.5 mg/cm^3 (according to the approach in section 4.1.),
- Pool of 10 PT rounds: U = 5.5 mg/cm^3 (according to the approach in section 4.2.).

Note: The expanded uncertainty (U) is reported with two significance digits (common practice).

6. Monitoring the measurement uncertainty

A way to monitor the measurement uncertainty is to plot estimates of the combined standard uncertainty (u_c, y-axis) according to round numbers (x-axis). Participants should use the approach described in section 4.1 (Eq. 7) if they are interested in monitoring their own measurement uncertainty.

It is recommended to plot u_c (for n = 1 as in Table 8), unless the number of replicates is the same in each round of the PT scheme. Figure 1 shows the corresponding run chart where the median value (3.29 mg/cm^3) is depicted by the horizontal dotted line. The combine standard uncertainty has remained rather stable over time. If applicable, a horizontal line representing a validity threshold may be added on the run chart.

Table 8. Combined standard uncertainty of each PT round on density (u_c in mg/cm^3).

Round	1	2	3	4	5	6	7	8	9	10
$u_{precision}$*	1.5	2.1	1.9	2.8	2.2	1.6	1.8	2.5	1.5	2.2
Bias (b)**	1	-1	-3	0	-2	-1	-3	-3	0	-1
u(b)***	0.87	1.21	1.10	1.62	1.27	0.92	1.04	1.44	0.87	1.27
u_c	2.00	2.62	3.72	3.23	3.23	2.10	3.65	4.16	1.73	2.73

With frequent negative bias estimates, the laboratory tends to underestimate the density.

* In absence of S_{round} in Table 7, $u_{precision}$ is taken as S_{pool*}.
** The bias estimate is equal to the difference between the mean and assigned value.
*** The error of the bias is taken as S_{pool*} / SQRT(n) as $u_{assigned}$ is negligible in Eq. 6.

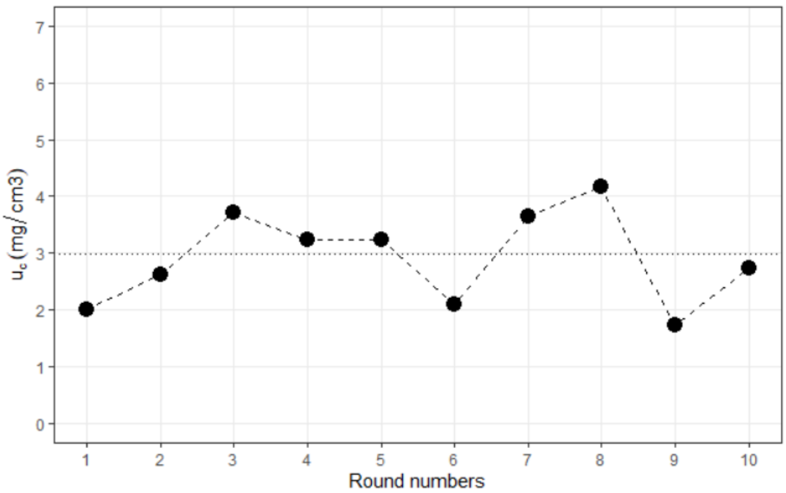

Figure 1. Combined standard uncertainty against round numbers.

7. Overview of approaches

This annex presents 4 approaches to estimate the measurement uncertainty. The first 2 approaches rely on the laboratory results only, and differ by the number of PT rounds taken into account: 1 round in Section 4.1 versus several rounds in Section 4.2. The later approach is expected to provide more robust estimates of the uncertainty contributors, and thus a combined standard uncertainty (u_c) that better reflects the performance of the laboratory on the mid-/long-term. In that perspective, a minimum of 6 different rounds over an appropriate period of time is recommended [5].

The next 2 approaches rely on the results of all participating laboratories, with, again, the opportunity to take 1 or several PT rounds into account (Sections 5.1 and 5.2). These approaches provide an estimate of the method variation or reproducibility, i.e. for an 'average' laboratory, while the first 2 approaches provide an estimate of the variation of one particular laboratory.

Table 9 provides an overview of the different approaches discussed in this annex.

Table 9. Overview of approaches using PT data to estimate the measurement uncertainty.

Section	PT data	Scope*	Combined standard uncertainty (u_c) **	Remark
4.1	1 laboratory 1 round	Lab. uncertainty, short term	$\sqrt{S_{round}^2 + b^2 + SE_m^2 + u_{assigned}^2}$	Includes the precision and bias components. However, provides a onetime estimate of the laboratory uncertainty (which may not be representative enough).
4.2	1 laboratory several rounds	Lab. uncertainty, Mid-/long-term	$\sqrt{S_{pool}^2 + RMS_b^2 + u_{assigned}^2}$	Provides a robust estimate of laboratory uncertainty. PT rounds (min. of 6) should belong to the same PT scheme.
5.1	Several labs. 1 round	Method uncertainty, short term	$\sqrt{S_{pool*}^2 + S_{inter}^2 + u_{assigned}^2}$	Provides a onetime estimate of the method uncertainty (reproducibility, i.e. including inter-lab variation).
5.2	Several labs. several rounds	Method uncertainty, Mid-/long-term	$\sqrt{S_{pool+}^2 + S_{inter+}^2 + u_{assigned}^2}$	Provides a robust estimate of the method uncertainty (reproducibility, i.e. including inter-lab variation).

* Lab. uncertainty: uncertainty estimate of a given laboratory; Method uncertainty: uncertainty estimate of an 'average' laboratory (method reproducibility). ** The combined standard uncertainty (u_c) is applicable to an individual result.

Precision components

S_{round}: standard deviation of n replicated results in a PT round.

S_{pool}: S_{round} averaged over several PT rounds by one single laboratory.

S_{pool*}: S_{round} averaged over several laboratories in a single PT round.

S_{pool+}: S_{pool*} averaged over several PT rounds.

Bias components

b: mean bias i.e. difference between the laboratory mean and assigned value in a PT round.

SE_m: standard error of the laboratory mean, equal to S_{round} / SQRT(n).

RMS_b: root mean square of mean biases (b) of several PT rounds by one single laboratory.

S_{inter}: inter-laboratory standard deviation calculated from a single PT round.

S_{inter+}: S_{inter} averaged over several PT rounds.

$u_{assigned}$: standard uncertainty of the assigned value.

8. References

1. Selection, Interpretation and Use of Proficiency Testing Schemes (2011), 2nd Edition, Eurachem
   https://www.eurachem.org/images/stories/Guides/pdf/Eurachem_PT_Guide_2011.pdf

2. Eurachem/CITAG Guide CG4, Quantifying Uncertainty in Analytical Measurement, 3rd Edition (2012) https://www.eurachem.org/images/stories/Guides/pdf/QUAM2012_P1.pdf

3. Eurolab Technical Report 1/2007 – Measurement uncertainty revisited https://www.eurolab.org/pubs-techreports

4. Miller JC & Miller JN, Statistics for Analytical Chemistry, 3rd Edition (UK, 1993)

5. Eurolab Technical Report 1/2006 - Guide to the Evaluation of Measurement Uncertainty for Quantitative Test Results – Measurement uncertainty revisited https://www.eurolab.org/pubs-techreports

6. Nordtest Technical Report 537, Handbook For Calculation of Measurement Uncertainty In Environmental Laboratories, Edition 3.1, 11/2012

   http://www.nordtest.info/images/documents/nt-technical-reports/nt_tr_537_ed3_1_English_Handbook%20for%20Calculation%20of%20Measurement%20uncertainty%20in%20environmental%20laboratories.pdf

7. BARWICK, V.J., ELLISON, S.L.R.: VAM Project 3.2.1 Development and Harmonisation of Measurement Uncertainty Principles. Part (d): Protocol for uncertainty evaluation from validation data (Jan 2000)

   https://blpd.dss.go.th/knowledge_el/VAM_uncertainty-0452.pdf