WGFF Guidelines for CMC Uncertainty and Calibration Report Uncertainty

October 21, 2013

Summary

The Working Group for Fluid Flow (WGFF) defines Calibration and Measurement Capability (CMC) uncertainty (U_{CMC}) as the root-sum-of-squares (RSS) of:

1) a type B base uncertainty of the reference standard obtained by using the law of propagation of uncertainty as described in the GUM[1] and

2) a type A repeatability of *n* calibration results measured using the Best Existing Device (BED), i.e.,

$$U_{\text{CMC}} = k_{95} u_{\text{CMC}} = k_{95} \sqrt{u_{\text{base}}^2 + u_{\text{repeat, BED}}^2} \quad . \tag{1}$$

U_{CMC} represents the 95 % confidence level uncertainty of the average of *n* calibration results for the Best Existing Device using the reference standard. The number of measurements *n* should match the normal procedures recommended by the lab during calibration of a customer's device or that may be used when providing services to clients. The quantity $u_{\text{repeat, BED}}$ is the standard deviation of the mean[2] of *n* repeated measurements performed on the best existing device under test. The quantity $u_{\text{repeat, BED}}$ should be evaluated at various set points within the range of the lab's capability.

Two methods are acceptable for calculating a 95 % confidence level result from the finite number of repeatability measurements:

1) using the Welch-Satterthwaite method to find the effective degrees of freedom and k_{95} as explained in Annex G of the GUM, or
2) using the 95 % confidence level *t*-value for *n*-1 degrees of freedom, divided by 2 and assuming k_{95} = 2, i.e. $U_{\text{CMC}} = 2\sqrt{u_{\text{base}}^2 + \left(\frac{t_{95}}{2} \frac{s_{\text{repeat, BED}}}{\sqrt{n}}\right)^2}$ where *s* is the sample standard deviation of *n* repeatability measurements.[3]

These Guidelines suggest how the GUM, CIPM MRA-D-04[4], and the ILAC Policy for Uncertainty in Calibration[5] should be applied to the estimation of uncertainty in a CMC. The uncertainty

---

[1] *Guide to the Expression of Uncertainty in Measurement*, JCGM 100:2008.

[2] $s/\sqrt{n}$ where *s* is the sample standard deviation of the *n* measurements.

[3] For *n* > 5, and $\sqrt{n}\, u_{\text{base}}/s_{\text{repeat,BED}}$ between 5 and 0.5, this approach gives larger U_{CMC} values than Welch-Satterthwaite by as much as 13 %. This statement assumes that the degrees of freedom for u_{base} is large and can be considered effectively infinite. In cases where u_{base} is not large, the Welch-Satterthwaite method should be used.

contributions due to instrumentation associated with the device under test (e.g. for pressure or temperature) and the effect of property uncertainties on the DUT are assumed to be zero for U_{CMC} but must be included in calibration report uncertainties. Guidance on reporting the uncertainty of the mean performance indicator[6] for a customer's device is given in the latter part of this document.

1. CMC Uncertainty

Base Uncertainty

It is current practice for laboratories to establish a detailed uncertainty budget accounting for all the recognised measurements and sources of uncertainty to establish the uncertainty of the quantity of fluid contained within, delivered by, or passed through a device under test. The base uncertainty of the reference standard (u_{base}) is determined by a propagation of uncertainty analysis as described in the GUM. It is expected that each identified measurement, instrument, and influence factor is assessed for repeatability, stability, reproducibility, and historical performance as part of this exercise. The magnitudes of some uncertainty components (e.g., the uncertainty from density changes of fluid in connecting volumes, or diverter errors) are often based on experiments dedicated for that purpose and must be included in the uncertainty analysis. Sample uncertainty analyses are available in the literature for the commonly used types of reference standards. [7, 8, 9, 10, 11, 12, 13] These valuable examples help labs to identify uncertainty contributors and give benchmarks for reasonable values of their magnitude.

NMIs have performed uncertainty analyses on their reference standards, but not all have fully accounted for the performance of the complete, assembled calibration system over the full range of use. In some cases, important uncertainties introduced by the dynamic performance of the calibration system, which are best quantified by experiment, have not been included. Means for

---

[4] Calibration and Measurement Capabilities in the Context of the CIPM MRA, CIPM MRA-D-04, Version 2, September 2010.

[5] *ILAC Policy for Uncertainty in Calibration*, ILAC-P14:12/2010.

[6] Performance indicator may be error, meter factor, *K*-factor, discharge coefficient, etc.

[7] Engel, R., Baade, H.-J., *Water Density Determination in High-Accuracy Flowmeter Calibration- Measurement Uncertainties and Practical Aspects*, Flow Measurement and Instrumentation, **25**, pp. 40 to 53, 2012.

[8] Shimada, T., Mahadeva, D. V., and Baker, R. C., *Further Investigation into a Water Flow Rig Related to Calibration*, Flow Measurement and Instrumentation, **21**, pp. 462 to 475, 2012.

[9] Wright, J. D. and Mattingly, G. E., *NIST Calibration Services for Gas Flow Meters: Piston Prover and Bell Prover Gas Flow Facilities,* NIST SP 250-49, National Institute of Standards and Technology, Gaithersburg, Maryland, 1998.

[10] Johnson, A. N., Crowley, C. J., and Yeh, T. T., *Uncertainty Analysis of NIST's 20 L Hydrocarbon Liquid Flow Standard*, Journal of the Metrology Society of India, **26**, 3, pp. 187 to 202, 2011.

[11] Wright, J. D., Moldover, M. R., Johnson, A. N., Mizuno, A., *Volumetric Gas Flow Standard with Uncertainty of 0.02 % to 0.05 %*, ASME Journal of Fluid Engineering, **125**, pp. 1058 to 1066, 2003.

[12] Carter, M., Johansen, W., Britton, C., *Performance of a Gas Flow Meter Calibration System Utilizing Critical Flow Venturi Standards*, Journal of the Metrology Society of India, **26**, 3, pp. 247 to 254, 2011.

[13] Bean, V. E., Espina, P. I., Wright, J. D., Houser, J. F., Sheckels, S. D., and Johnson, A. N., *NIST Calibration Services for Liquid Volume*, NIST Special Publication 250-72, National Institute of Standards and Technology, Gaithersburg, MD, November 24, 2009.

assessing and including this additional uncertainty in a consistent approach which can be adopted by the NMIs and accredited flow laboratories worldwide are described next.

Repeatability of the BED, $u_{\text{repeat, BED}}$

The repeatability component measures the ability of the facility and a DUT to produce the same results over a short time period at points spread across its operating envelope. It is included in the CMC uncertainty in order to include uncertainty from sources that are presently unknown and therefore not captured in the propagation of uncertainty analysis.

To measure this component, a commercially available DUT with inherently good repeatability (the BED) is tested in the calibration system. Equipment and data to quantify this component are often already available from check standards used for a laboratory's quality system. By the nature of the test, the result will be a combination of the repeatability of the chosen BED and that of the facility.

Procedure for Assessing Repeatability

The procedure should be designed considering the financial budget available and the normal working practices of the laboratory. The procedure for a flow calibration is:

1) Establish the range of flows, temperatures, and other conditions listed in the scope of the facility to be covered.
2) Define a number of test flows and conditions which within the normal working practices, should demonstrate the least stable conditions and one or two which should demonstrate the most stable conditions. (The least stable results are generally found at the highest or lowest ends of the reference standard range.)
3) Identify the BED to use for each test.
4) Working within the operating procedures of the laboratory to ensure each test can proceed without interruption within a working day, perform multiple calibrations under stable flows and conditions to produce performance indicator values for the BED. The number of repeats should match the normal practices used when calibrating a customer's device or may be used when providing services to clients. If procedures dictate less than twenty points, the effect of the smaller sample should be fully recognised in the analysis by using the Welch-Satterthwaite or *t*-value approaches described below.
5) For each set of points, examine the data to determine the distribution to establish if a 'normal' or Gaussian distribution can be assumed.
6) If the distribution is approximately normal, calculate the sample standard deviation of the mean, $s/\sqrt{n}$. If the sample size is twenty or greater, the standard deviation of the mean can be combined by RSS with the base uncertainty and expanded using a coverage factor of 2. If the sample is smaller than twenty then, 1) use the Welch-Satterthwaite method to determine the degrees of freedom and the coverage factor necessary to achieve a 95 % confidence level, or 2) multiply the standard deviation by the 95 % confidence level *t*-value for *n*-1, divide by 2 to give the standard uncertainty for RSS with other components, and use a coverage factor of 2 on the RSS to obtain the expanded uncertainty. If the distribution of

the results is not normal or Gaussian, then further consideration and assessment is necessary to determine the correct statistical approach.

The procedure for a volume calibration is:

1) Establish the volume range and type of volumetric equipment covered in the scope of the calibration system.
2) Select the instruments (BED) to be tested.
3) Perform *n* tests of each instrument at the chosen volume, where *n* is the number of repeated measurements normally performed on a customer's device.
4) Work within the operating procedures of the laboratory to ensure each test can proceed without interruption within a working day.
5) Take the sample standard deviation of the mean of the repeated measurements and

    a) use the Welch-Satterthwaite method to determine the degrees of freedom and the coverage factor necessary to achieve a 95 % confidence level, or

    b) multiply the standard deviation by the 95 % confidence level *t*-value for *n*-1, divide by 2 to give the standard uncertainty for RSS with other components, and use a coverage factor of 2 for the expanded uncertainty.

Note that sample standard deviation of the mean ($s/\sqrt{n}$) is used here to quantify repeatability, ***not*** sample standard deviation because the goal is to capture the uncertainty for the average of *n* measurements from the reference standard.

The CMC uncertainty can be stated as:

1) the largest value over the operating range of the facility or calibration system or
2) a function describing the uncertainty across the operating range or
3) a table showing how uncertainty varies over the operating range or volume range.

2. Reported Uncertainty for the Performance Indicator in a Calibration Report

The result of a calibration of a customer's device is reported in some form of "performance indicator". Examples of performance indicators are error, meter factor, *K*-factor, and discharge coefficient. When the uncertainty of the mean performance indicator for a customer's device U_{PI} is reported, it will include additional components over the base uncertainty due to:

1) instrumentation and characteristics associated with the DUT u_{AI},

2) fluid properties (if applicable) u_{prop}, and

3) repeatability or short-term reproducibility for the customer's DUT $u_{\text{repeat or reprod, DUT}}$, and the uncertainty of the performance indicator can be expressed as:

$$U_{\mathrm{PI}} = 2u_{\mathrm{PI}} = 2\sqrt{u_{\mathrm{base}}^2 + u_{\mathrm{AI}}^2 + u_{\mathrm{prop}}^2 + u_{\text{repeat or reprod, DUT}}^2} \quad . \qquad (2)$$

Other uncertainties can be included at the discretion of the lab. Note that the reported uncertainty of the performance indicator must not be less than the uncertainty stated in the CMC at that point in the operating range, i.e. U_{PI} must be $\geq U_{\mathrm{CMC}}$. [5]

Associated Instrumentation and Property Uncertainties

Laboratory instrumentation is often used to measure output quantities from a flow device under test, like pressure, temperature, frequency, voltage, or current. These instruments associated with the DUT contribute to the uncertainty of the performance indicator. For some DUT's the uncertainties of fluid properties like density, viscosity, or thermal conductivity affect the uncertainty of the performance indicator reported by the laboratory. Using the example of a critical flow venturi, the contributions of the pressure and temperature instrumentation to the uncertainty of the discharge coefficient are,

$$u_{\mathrm{AI}} = \sqrt{u_{\mathrm{P}}^2 + \frac{1}{4}u_{\mathrm{T}}^2} \ . \qquad (3)$$

The uncertainties contributed to the performance indicator by the uncertainty in fluid properties for a critical flow Venturi are,

$$u_{\mathrm{prop}} = \sqrt{u_{\mathrm{C*}}^2 + \frac{1}{4}u_{\mathrm{M}}^2} \, , \qquad (4)$$

where *C** and *M* represent the critical flow factor and the molar mass for the gas.

For volumetric instruments u_{AI} and u_{prop} can be related to the characteristics of the customer's volumetric instrument under calibration that may be insignificant for the BED and not included in the base uncertainty, e.g. the scale resolution, the expansion coefficient of the material, the meniscus reading, evaporation rate, etc.

The uncertainty components that apply in the associated instrumentation and properties categories must be considered on a case by case DUT basis and quantified by a thorough uncertainty analysis just as for the reference standard, following the GUM.

Repeatability or Short-Term Reproducibility of the Customer's DUT

The type A uncertainty contributed by the customer's device must be included in the uncertainty of the performance indicator. For flow calibrations, most laboratories make multiple measurements, at multiple set points, on more than one occasion, and often approach the set point from different directions on different occasions to assess hysteresis. A variety of methods are acceptable for quantifying $u_{\text{repeat or reprod, DUT}}$, from these various sets of data, but in all cases, the method must be clearly stated in the report. Acceptable methods are:

1) The Welch-Satterthwaite method and effective degrees of freedom applied to multiple measurements at each set point,
2) The sample standard deviation of the mean of multiple measurements at each set point ($s/\sqrt{n}$) with statistical corrections using the *t*-value for the finite sample size,

3) The sample standard deviation of multiple measurements at each set point, with weighting based on the *t*-value at 95 % confidence for the number of points and then divided by 2 to give the standard uncertainty for RSS with other components.

Note: where two or more sets of data are taken at one set point, the standard deviations may be pooled.

Long-Term Reproducibility and Inter-Laboratory Comparison Results

Long-term reproducibility measures the calibration system stability over years. By means of trend-analyses, (based on results of frequent cross checks between facilities, recalibration campaigns, and lab to lab comparisons) calibration system's drift behaviour can be derived and used to underpin the estimated uncertainty budget for drift. Such analyses can be based on e.g. ISO 7066-1 (linear calibration relationships) or ISO 7066-2 (non-linear relationships) where a model can be established to indicate the shift over time of the reference value disseminated by the facility or calibration system. Obviously, models can only be made on the basis of enough results over time. Therefore, due to a lack of information, a new facility's trend (stability) figures will initially be based on the experience of the expert.

Comparison results (proficiency tests) are required to validate these models or estimates of NMIs and accredited laboratories. Failing to show consistent long-term reproducibility and comparison results that support a laboratory's CMC can result in suspension of a CMC entry or accreditation if appropriate corrective actions are not taken.

There is no established approach for including comparison results in a laboratory's CMCs. A laboratory that obtains unsuccessful comparison results (i.e. standardized degree of equivalence $|E_n| > 1$) must conduct diagnostic tests and re-examine their uncertainty analysis and revise their CMCs or improve their measurement system.