Journal of Measurement Science & Applications, JMSA. Vol (5) Issue (1) Feb. (2025)

Monte Carlo Validation Approach for Evaluating the Uncertainty of Identical Force Intervals

Ahmed A. Hawam[1], Ibrahim S. EL-Deeb[1], K. M. Khaled[2]*, Mohamed ElWakil[1]
[1] Production Engineering and Mechanical Design Department, Faculty of Engineering, Tanta University, Tanta, Egypt, PO Box 31733
[2] Force and Material Metrology Laboratory, National Institute of Standards (NIS), Egypt
*Corresponding author's email: khaled.abdelaziz@nis.sci.eg

Article Type: Research article
Received: 18/12/2024 **Accepted: 3/2/2025**

Abstract

Type A method is the GUM-adopted technique for evaluating the uncertainty of repeated observations. Unfortunately, the Type A method cannot provide any information about the dispersion of identical observations. This article introduces a reasonable approximation for the uncertainty of identical observations depending on the resolution of the limited-resolution measuring device and the number of identical observations. The proposed approach reduces the estimated uncertainty by at least 13.7% of the instrument's resolution compared to the GUM method without sacrificing reliability. Moreover, the proposed approach is simple, straightforward, and easily implemented in daily routine work.
Monte-Carlo simulation was performed to compare the proposed uncertainty statement of identical observations against the conventional Type A uncertainty of the observations indicated by finer-resolution devices. The simulation revealed that the percentage of agreement does not drop below 70% at worst according to the strictest criterion; for round-robins' criterion, the agreement remained above 99%.

Keywords: **GUM; Resolution; Uncertainty; Type A; Monte Carlo.**

1 Introduction

Measurement uncertainty is the key to reflecting the reliability of the measurement result and the confidence in decisions taken upon it. Moreover, uncertainty evaluation becomes indispensable to most measuring, testing, and calibration standards. In metrological uncertainty evaluation, input quantities are quantified either statistically using the Type A evaluation method or by scientific judgment using the Type B evaluation method. As a robust technique, the Type A uncertainty evaluation method is the Guide to the expression of uncertainty in measurement (GUM) [1] adopted methodology for quantifying the uncertainty from repeated observations. As a drawback, the result of the Type A method is affected by the limited resolution of the used measuring device as the small variations become more and more insensible by the increase of the resolution value until the different independent observations become identical. In other cases, the resolution may be adequate for the variations of the measurand's signal but, filtering and averaging the acquired samples make the results artificially identical. In these situations, the Type A method cannot provide any information about the uncertainty of the captured values. This ignorance can be a concern, especially in high-accuracy measurements and calibrations in which uncertainty can affect a decision or the classification of an under-test instrument. In such a case, the metrologist should resort to an alternative approach to the Type A method to perform its same role. Till now, there has been no agreement on a specific one, and the problem is still an open question in metrology.

To show the problem from a mathematical perspective, consider a measured quantity q and the knowledge about it is limited to a set of n repeated observations q_1, q_2, …, q_n that are independently drawn from its respective distribution (with mean μ_q and variance σ_q^2) under the same measurement conditions. According to the Type A method, the arithmetic average $\bar{q}_n$ is taken as the best available estimate of the mean μ_q and is calculated as:

$$\bar{q}_n = \frac{1}{n}\sum_{i=1}^{n} q_i \tag{1}$$

This value is considered the best estimate of q and taken as its input estimate in subsequent uncertainty models that depend on q.

Also, the variance σ_q^2 is best estimated by the *sample variance* $s^2(q_i)$ [2] which is calculated as:

$$s^2(q_i) = \frac{1}{n-1}\sum_{i=1}^{n}(q_i - \bar{q}_n)^2 \tag{2}$$

Finally, the Type A uncertainty of the quantity q, which is denoted by $u(q)$ is taken as the positive square root of the experimental variance of the mean $s^2(\bar{q}_n)$ as in Eq. (3); $s(\bar{q}_n)$ is also referred to as the experimental standard deviation of the mean.

$$u(q) = \sqrt{s^2(\bar{q}_n)} = \sqrt{\frac{s^2(q_i)}{n}} = \frac{s(q_i)}{\sqrt{n}} \tag{3}$$

When the repeated observations $\{q_1, q_2, ..., q_n\}$ are identical, Type A standard uncertainty calculated using eq. (3) becomes zero. However, this does not mean that the measurand is perfectly known; on the contrary, this indicates that the limited resolution of the measuring device cannot distinguish the variations in the measured signal. In other words, a limited resolution can result in rounding observations, making them appear artificially identical.

In subclause F.2.2.1, the GUM addressed the case of identical observations and recommended that the uncertainty attributable to repeatability should not be zero (as would result from Eq. (3)) but rather equal to 0.29δ where δ is the resolution of the measuring device (this will be referred to as the GUM method). Accordingly, the measured signal is expected to lie in the interval $[X - \delta/2, X + \delta/2]$ with equal probabilities, where X is the value of the identical observations, see Figure 1. In theory, this interpretation has a significant drawback as it assigns the same likelihood to the limits of the interval as its center. In practice, a measured quantity varies randomly around its mean, if its mean value lies on one of the interval limits, the neighborhood indications of *X* i.e., $X+\delta$ or $X-\delta$ would appear in the observations set. A noteworthy here is that the GUM did not classify this approach as a Type A evaluation method or a Type B estimation that relies on scientific judgement.

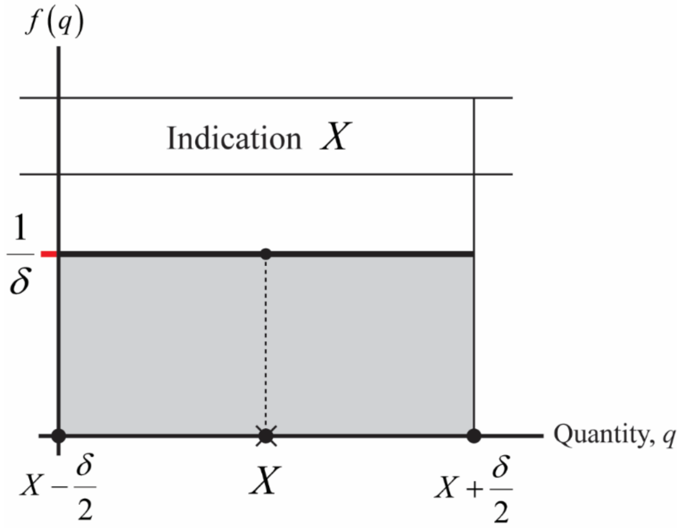

Figure 1: The probability density function $f(q)$ of the quantity q over the resolution limits of the indication X.

The literature on evaluating uncertainty in the case of identical observations is minimal. A straightforward solution for considering the effect of limited resolution on the uncertainty was proposed by Lira and Wöger, 1997 [3]. The idea of this work is to add the term $\delta^2/12$ to the variance $s^2(q_i)/n$ in eq. (3). Thus combining the conventional Type A standard uncertainty and the resolution standard uncertainty $\delta/\sqrt{12}$ that is taken as the standard deviation of the uniform distribution of span equal to the value of δ as shown in Figure 1. Accordingly, the uncertainty in the case of identical observations becomes equal to $\delta/\sqrt{12}$. This approach has widely resonated, as it is simple, to the extent that the GUM adopted this solution in its subclause F.2.2.1 as illustrated. Moreover, Other researchers, such as *Elster*, 2000 [4], adopted this combination to study the uncertainty in case the measured signal suffers from random errors combined with analogue-to-digital conversion errors.

Also, Frenkel and Kirkup, 2005 [5] conducted a Monte Carlo simulation by generating a large set of observations and rounded it to mimic the behaviour of measuring devices with limited resolution. In that study, and when the observed variance computed from the rounded data was equal to *zero* (i.e., identical observations case), the actual mean value was found to have the potential to lie within the resolution limits $[-\delta/2, \delta/2]$ while the observed mean value was equal to *zero*. By assuming a uniform distribution over the expected range $[-\delta/2, \delta/2]$, the standard uncertainty of identical observations proposed by *Lira* and *Wöger* $\delta/\sqrt{12}$ was supported, despite the theoretical drawbacks of using the uniform distribution for describing identical observations as discussed earlier and as it will be clear later in this article. The solution proposed by *Lira* and *Wöger* attempts to guess the lost piece of knowledge about the dispersion of the observations due to limited resolution. Generally, it may represent a reasonable solution despite the evidence that it mixes Type A and Type B evaluation methods that are different in their principles [6], which may result in interpretation problems [7, 8]. But, in case of identical observations with a value *X*, as Type A uncertainty vanishes, we implicitly assign the uniform distribution over the range $[X - \delta/2, X + \delta/2]$ and calculate Type A uncertainty upon it. Thus, the unsuitability of the uniform distribution to Type A uncertainty – specifically in the identical case – appears again.

In practice, there are three inevitable fundamental input quantities in any uncertainty budget: Type A uncertainty, i.e., the repeatability, the resolution uncertainty of the measuring device, and the calibration uncertainty of the measuring device. Other input quantities may be added depending on the nature of the measuring system, Eq. (4).

$$u_c^2(q) = \frac{s^2(q_i)}{n} + \frac{\delta^2}{12} + u_{\text{cal.}}^2 + \dots \tag{4}$$

, where, $u_c(q)$ is the combined standard uncertainty of the quantity *q*, and $u_{\text{cal.}}$ is the calibration standard uncertainty of the used measuring device. As apparent, the resolution uncertainty $\delta^2/12$ is indispensable in any uncertainty budget and should exist in the GUM uncertainty formula, Eq. (4).

In practice, there is a particular paradox if we considered the above solution as Eq. (4) would be rewritten in that case as:

$$u_c^2(q) = \left(\frac{s^2(q_i)}{n} + \frac{\delta^2}{12}\right) + \frac{\delta^2}{12} + u_{\text{cal.}}^2 + \dots = \frac{s^2(q_i)}{n} + 2\left(\frac{\delta^2}{12}\right) + u_{\text{cal.}}^2 + \dots \tag{5}$$

In case of identical observations, Eq. (5) can be simplified as:

$$u_c^2(q) = 2\left(\frac{\delta^2}{12}\right) + u_{\text{cal.}}^2 + \dots \tag{6}$$

Accordingly, and in either case, the resolution uncertainty is double counted. On the other side, if the resolution uncertainty is omitted, Eq. (5) would return to be identical to Eq. (4), i.e., the classical formula of the GUM uncertainty framework!

A question that is often asked is: why take more observations if the measurand is stable and the limited resolution measuring device indicates the same value. First, this may be a requirement by the standard or the protocol upon which the measurements are taken. For example, a series of observations must be taken when calibrating a force-proving instrument according to ISO 376:2011 [9], whatever the indicator's resolution, to assess the system characteristics. Another example is when calibrating a dial proving ring according to ASTM E74-18e1 [10], three observations must be taken to evaluate its repeatability. Repeated measurements may also be taken as confirmation for the identical case; in other words, if the first two observations were identical, another one or two can be taken to figure out if different values would appear, then the conventional Type A method can be employed, or the identical case is much more confirmed.

The quantifying of uncertainty in metrology should not be treated as pure math. For each uncertainty source, *scientific judgment* is vital in determining the limits of the uncertainty interval and specifying the probabilities of its values. While the requirement of scientific judgment is evident in the Type B evaluation method, its role in the Type A evaluation method is excluded implicitly in GUM and GUM-based uncertainty guides in which the observations are considered the only available knowledge about the measurand. This position is arguably unscientific [11] as the metrologist can provide other relevant information that may amend the uncertainty value. In this article, when the conventional Type A method cannot give any information about the dispersion of a set of identical observations, a reasonable guess of its standard uncertainty is proposed based on a new interpretation for the case of the identical observation. The proposed solution significantly reduces the estimated uncertainty compared to the GUM method without sacrificing the reliability of the proposed uncertainty statement. Moreover, it is simple, straightforward, and easily implemented in laboratories' daily routine work.

2 Type A Uncertainty in Light of the Central Limit Theorem

The *Central Limit Theorem* (CLT) forms an essential basis for inferential statistics. According to CLT, if samples of size *n* are taken from a population of a random quantity q that has a *mean* μ_q and *variance* σ_q^2, the sampling distribution of sample means $\hat{q}_n$ (i.e., the probability distribution of sample means) would follow a normal distribution with expectation $E[\hat{q}_n]=\mu_q$ and variance $\mathrm{Var}[\hat{q}_n]=\sigma_q^2/n$, i.e., $\hat{q}_n \sim N(\mu_q, \sigma_q^2/n)$.This theorem is applied either when the population has any form, but the sample size *n* is relatively large ($n \geq 30$) or the population is normally distributed regardless of *n*.

When interpreting the philosophy of the Type A uncertainty evaluation method, it becomes apparent that it aims to reconstruct the sampling distribution of sample means of the quantity q by the sample variance available in hand $\bar{q}_n$. Figure 2 illustrates this idea; Figure 2.a shows the *unknown* probability distribution of the quantity *q* that follows the normal form with an unknown expectation μ_q and unknown variance σ_q^2; Figure 2.b also shows its respective *unknown* sampling distribution that also follows the normal form with an expectation μ_q and variance σ_q^2/n, this sampling distribution gives the probability of the random variable $\hat{q}_n$ that is the *sample mean* of sample size *n*.

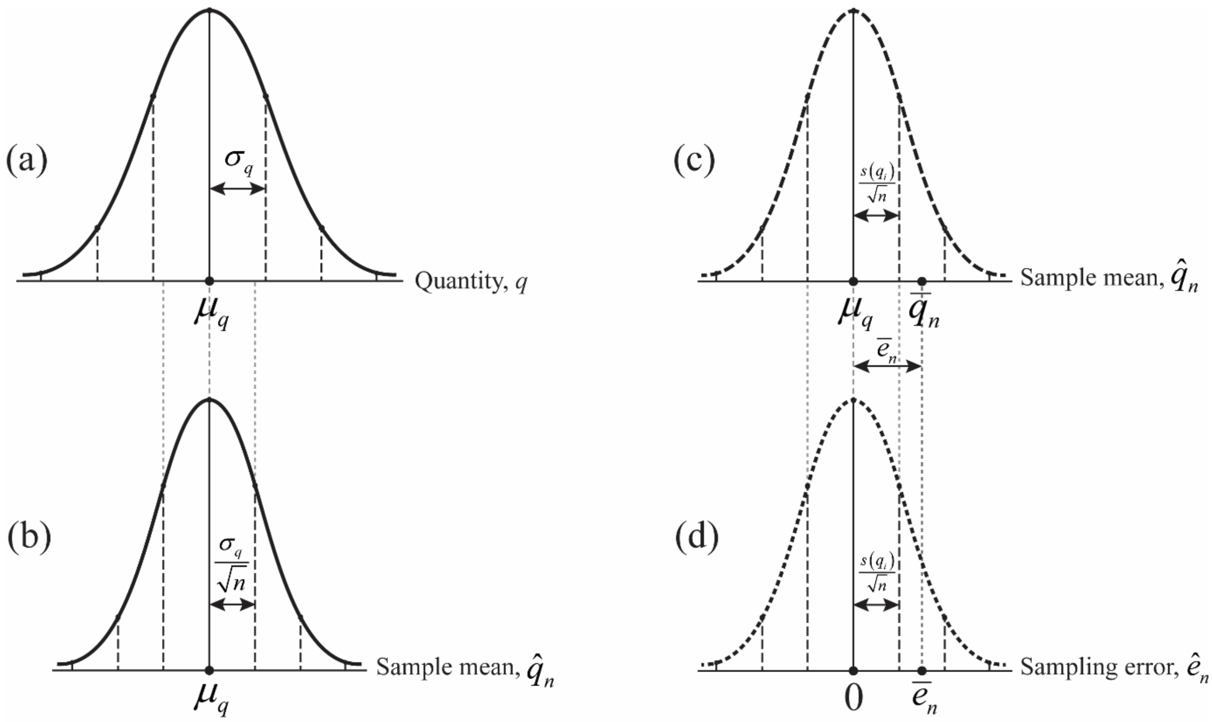

Figure 2: The steps of reconstructing the sampling distribution of sample means of the quantity q. (a) The distribution of q, (b) The sampling distribution of q, (c) The reconstructed sampling distribution of q, (d) The distribution of the sampling error of sample means.

For reconstructing the sampling distribution, the sample variance $s^2(q_i)$, as calculated in Eq. (2) is taken as the best estimator of the population parameter σ_q^2 [12]. The question now is: to what extent does the value of the sample mean $\bar{q}_n$ estimate the parameter μ_q? In other words, how much is the error $\bar{e}_n$ (Figure 2.c) between $\bar{q}_n$ and μ_q? This error, which is known as the sampling error, when estimated, would be considered as the uncertainty associated with $\bar{q}_n$; this is consistent with the fundamental concept of uncertainty that states that errors with unknown values (because here, μ_q is unknown) are uncorrectable and considered as the sources of uncertainty [13, 14].

As the sample mean $\hat{q}_n$ is a random variable, the sampling error $\hat{e}_n$ is also a random variable that can be calculated as:

$$\hat{e}_n = \hat{q}_n - \mu_q \tag{7}$$

Thus, the distribution of the sampling error (Figure 2.d) would follow the normal form with an expectation equal to zero and with the same variance of $\hat{q}_n$, i.e., $s^2(q_i)/n$. Accordingly, the standard uncertainty associated with $\bar{q}_n$, i.e., the standard uncertainty of Type A can be considered as the standard deviation $s(q_i)/\sqrt{n}$ of the reconstructed sampling distribution, and this value can be expanded using the coverage factors of the normal distribution [14].

The following are some notes about the philosophy of the Type A uncertainty method in light of the above explanation and the CLT:

- The sampling distribution of sample means (Figure 2.b) is always considered Gaussian, necessitating that the measurand q must have the Gaussian form. The reason is that the number of repeated observations upon which the sampling distribution is reconstructed is always relatively small, say three or five, in most metrological applications [15]. Accordingly, and as long as the number of repeated observations is less than thirty, there is no guarantee for the sampling distribution's normality except the normality of the quantity q. This is also supported from a physical perspective as the noise responsible for repeated measurements dispersion is usually considered Gaussian, even the unavoidable electrical noise coming from analogue-to-digital conversion circuits [16, 17].
- The expectation of the distribution of the sampling error $E[\hat{e}_n] = 0$ (Figure 2.d), this means that there is no bias in the measurement result. Accordingly, no correction should be applied to the measurement result, which can be considered a merit of the Type A method.
- $\overline{q}_n$ as calculated per Eq. (1) is not itself a random variable; it is an outcome of the random variable $\hat{q}_n$, so the term $s^2\left(\overline{q}_n\right)$ in Eq.(3) becomes more precise if written as $s^2\left(\hat{q}_n\right)$.

The span of the standard uncertainty interval $2\left[s\left(q_i\right)/\sqrt{n}\right]$ is symmetrical about μ_q (Figure 2.c), not about $\overline{q}_n$. Nevertheless, as $\overline{q}_n$ is the available estimator of μ_q, it is considered as the result of the measurement as well as the center point of the span $2\left[s\left(q_i\right)/\sqrt{n}\right]$.

3 A new interpretation for identical observations case

The idea of the current paper is to guess the term $s(q_i)$ in case of identical observations depending on the available knowledge implied in that case and keep using Eq. (4) of the GUM method without modifications. The proposed point of view depends on the following reasonable assumptions:

- *The* identical observations are drawn from a normal distribution; this is a fundamental assumption of the Type A evaluation method, as illustrated in the previous section.
- As the domain of the normal distribution ranges from $-\infty$ to $+\infty$, we must consider a truncated normal distribution that entirely lies in the X-indication interval. Accordingly, it is assumed that the truncated domain of the observations' normal distribution is entirely within the interval $\left[X-\delta/2, X+\delta/2\right]$, where $X = \overline{q}_n$ is the value of the identical observations and their mean, and δ is the resolution of the measuring device. This assumption is necessary for the identical case; otherwise, neighborhood indications of X would exist in the set of observations.
- The expectation of the observations' distribution is not necessarily equal to X, while the maximum expected span width of the distribution's domain (truncated) depends on

the location of that expectation concerning X. As obvious in Figure 3, when the expectation of distribution B becomes $\bar{q}_n^*$ instead of $X = \bar{q}_n$, its maximum expected span width w_{B} is shortened concerning the maximum expected span width of distribution A, w_{A}, which has an expectation $\bar{q}_n$. As $\bar{q}_n^*$ gets farther away from $X = \bar{q}_n$ as its maximum expected span width becomes narrower to keep the identical condition and prevent neighbourhood indications of X from appearing.

As evident in Figure 3, the maximum expected span width is achieved in case the expected value of the observations' distribution $\bar{q}_n$ is equal X as apparent in distribution A; thus, it can be considered as the measurement result; once again, this is also a merit as there is no need for a correction in the measurement result. In that case, the maximum expected span width $w_{\mathrm{A}} = (X + \delta/2) - (X - \delta/2) = \delta$. For calculating the standard deviation of the observations' distribution, w_{A} is assumed to be equal to $6.6\, s_n(q_i)$, where $s_n(q_i)$ is the standard deviation of distribution A in Figure 3; this means that the percentage of the area of the distribution encompassed between the limits $X - \delta/2$ and $X + \delta/2$ equals 99.9%; thus,

$$w_{\mathrm{A}} = \delta = 6.6\, s_n(q_i) \rightarrow s_n(q_i) = \frac{\delta}{6.6} \tag{8}$$

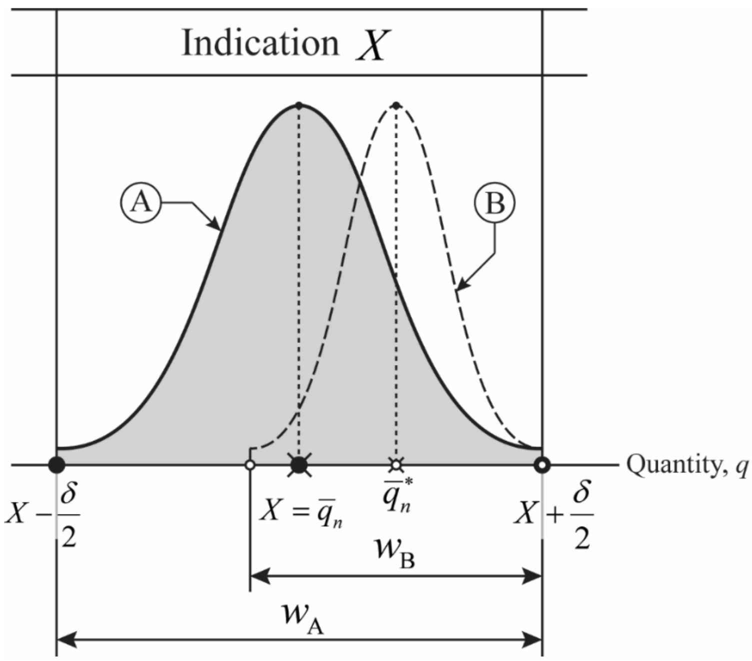

Figure 3: The expected probability distribution of identical observations.

The standard deviation $s_n(q_i)$ of distribution A is guessed as it is calculated from the n identical observations. For estimating the standard deviation of the observations' population, the value of the variance $s_n^2(q_i)$, the square of the standard deviation $s_n(q_i)$, calculated per eq. (8), is factorized by the quantity $n/(n-1)$; this can be considered as the sample variance $s_{n-1}^2(q_i)$ of the observations and can be calculated as:

$$s_{n-1}^2(q_i) = \frac{n}{n-1} \times \left(\frac{\delta}{6.6}\right)^2 = \frac{n \cdot \delta^2}{43.56(n-1)}; n > 1 \tag{9}$$

Finally, in case of identical observations, eq. (3) can be rewritten as:

$$u(q) = \frac{\left(\sqrt{\frac{n}{n-1}} \times \frac{\delta}{6.6}\right)}{\sqrt{n}} = \frac{\delta}{6.6\sqrt{n-1}}; n > 1 \tag{10}$$

Accordingly, in case of identical observations, eq. (4) of the GUM uncertainty framework can be rewritten as:

$$u_c^2(q) = \left(\frac{\delta}{6.6\sqrt{n-1}}\right)^2 + \frac{\delta^2}{12} + u_{\text{cal.}}^2 + \ldots = \frac{\delta^2}{43.56 \times (n-1)} + \frac{\delta^2}{12} + u_{\text{cal.}}^2 + \ldots; n > 1 \tag{11}$$

In Figure 4, the standard uncertainty as per eq. (10) is plotted versus the number of identical observations at different resolution values. As apparent, the uncertainty decreases with the decrease in resolution, and this is because the observations' distribution span width becomes narrower with the reduction of the resolution. Additionally, at a certain resolution value, the uncertainty decreases by the increase of the number of identical observations n, that is because the standard deviation of the reconstructed sampling distribution – evaluated via dividing by $\sqrt{n}$ in Eq. (10) – upon which the uncertainty is evaluated decreases by the increase of n. In other words, this can be interpreted as by the increase of n, the sampling error decreases, and the confidence in X as an estimator of μ_q increases. The same question can be reformulated from a different perspective: what knowledge is gained by a further identical observation that decreases uncertainty. The answer is simple: it is a memory feature of the proposed approximation; as the new observation confirms the identical case, the uncertainty associated with the measurement result decreases.

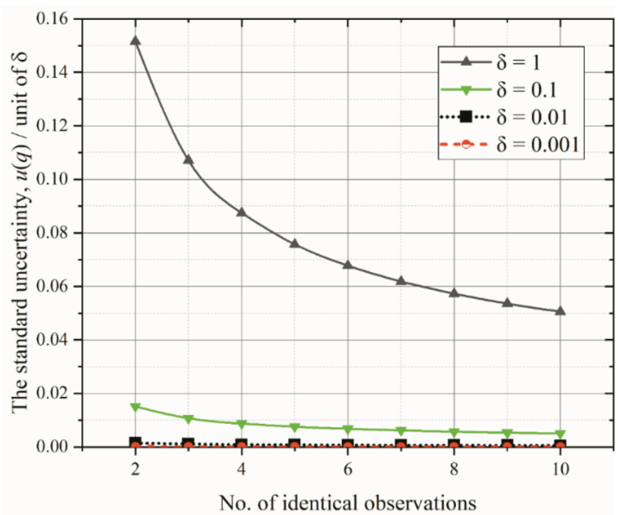

Figure 4: The standard uncertainty of identical observations as per Eq. (10), plotted at different resolution values.

If the uncertainty values in Figure 4 are compared to their counterparts calculated as specified in GUM-Subclause F.2.2.1, it would be found that the difference between the values $\delta/\sqrt{12}$ and $\delta/6.6\sqrt{n-1}$ is always positive. As shown in Figure 5, the difference is minimum at n=2 and equals 13.7% of the value of δ and increases steadily until it reaches its maximum value at n=10 with 23.8% of the value of δ . Thus, a reasonable significant reduction in the uncertainty compared to the GUM method is achieved.

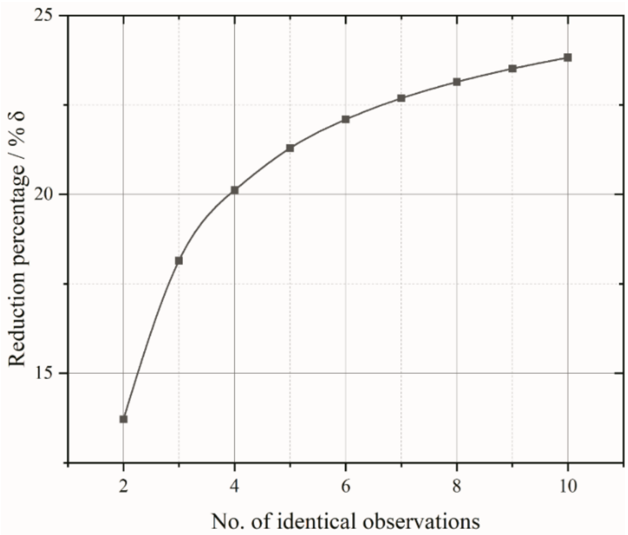

Figure 5: The reduction percentage in the resolution uncertainty value compared to the GUM method at different values of n.

Practically, by considering the proposed approach, a rational estimation of the standard uncertainty of identical observations can be achieved with a significant reduction compared to the GUM method. Someone might see this approach as an extension to the Type A method as it mimics its respective steps in reconstructing the sampling distribution of sample means of the measured quantity; someone else might see it as a Type B method as it depends on scientific judgment. Either way, the GUM considered the same tactic and introduced a similar approach – depending on scientific judgement – to solve the problem of Type A identical observations in subclause F.2.2.1, as discussed in the introduction. In theory, the proposed approach might be considered as a “Type B estimation for a Type A indeterminate case.” It is noteworthy here that the proposed uncertainty statement in Eq. (10) is just a rational approximation for the uncertainty of identical observations. It should not be treated as a result of an analytical derivation for solving a statistical problem. Accordingly, an assessment of its results from a metrological point of view should be performed before its implementation, as discussed in the next section.

4 Metrological assessments of the proposed standard uncertainty statement of identical observations

For assessing the proposed statement of the uncertainty of identical observations from a metrological perspective, a hypothetical error-free measuring device with limited resolution

$\delta = 1$ measuring a Gaussian signal in the interval $[X - 0.5, X + 0.5]$ with expectation $X = \overline{q}_n$ (taken arbitrarily equal to 15.00) and standard deviation $1/6.6$, is compared against a set of other hypothetical error-free measuring devices measuring the same signal but with finer resolution α. The agreement between the results is assessed as follows [18]:

Consider the measurement result of the measuring device with limited resolution is $X = \overline{q}_n$ and is $\tilde{q}_n$ for the device with finer resolution; accordingly, the absolute difference between the two values $\Delta = |\overline{q}_n - \tilde{q}_n|$. Also, consider the expanded uncertainty of the measuring device with limited resolution $U_{\text{L}} = 2\left(\delta / \left(6.6\sqrt{n-1}\right)\right)$ while the Type A expanded uncertainty for the device with finer resolution $U_{\text{F}} = 2\left(s(q_i)/\sqrt{n}\right)$.

Agreement Criterion 1 (AC1): The first criterion confirms the agreement between two measuring systems if the absolute difference between their measurement results is less than or equal to the minimum of their expanded uncertainties, i.e., $\Delta \le \min\left(U_{\text{L}}, U_{\text{F}}\right)$.

Agreement Criterion 2 (AC2): The second criterion confirms the agreement between two measuring systems if the absolute difference between their measurement results is less than or equal to the *root sum of squares* (RSS) of their expanded uncertainties, i.e., $\Delta \le \sqrt{U_{\text{L}}^2 + U_{\text{F}}^2}$.

On the one hand, AC1 represents the ultimate agreement grade according to the *American Society of Mechanical Engineers* ASME B89.7.3.3 standard [18], which is devoted to assessing the reliability of uncertainty statements. On the other hand, AC2 is commonly considered as the agreement criterion between laboratories in round robins. It is noteworthy here that AC1 is stricter if compared to AC2.

Monte-Carlo computer simulation software was developed to mimic the behaviour of the hypothetical measuring devices used for the assessment. The simulation software generates a Gaussian random sample of controllable size n (n = 2, 3, …, 10) with expectation $\overline{q}_n = 15.00$ and standard deviation equal to 1/6.6. The elements of the generated sample are rounded to the nearest digit allowed by the resolution α of the hypothetical finer-resolution device; in the simulation, six hypothetical finer-resolution measuring devices with different resolutions $\alpha = 10^{\ell}$ where $\ell = -1, -2, ..., -6$ were considered. Then, the mean value and Type A expanded uncertainty of the sample i.e., $\tilde{q}_n$ and U_{F} respectively, are calculated. After that, the agreement between the results of the limited-resolution device $(\overline{q}_n, U_{\text{L}})$ and the results of the finer-resolution device $(\tilde{q}_n, U_{\text{F}})$ is checked using the agreement criteria AC1 and AC2. After running the simulation for 10^6 trials, the number of positive agreements recorded for both AC1 and AC2 is obtained, and its percentage relative to the total number of trials is calculated. Pilot tests were also conducted to ensure that the obtained results are not limited to the tested resolutions, and depend mainly on the ratio γ where $\gamma = \alpha/\delta$ and $\delta = 10^{\lambda}$ as long as $\ell \le \lambda - 1$, while ℓ and $\lambda \in \{0\} \cup \mathbb{Z}^-$; in the current assessment $\gamma = \alpha/10^0 = \alpha$. The simulation results are plotted as the percentage of positive agreements at different α values versus the number of identical observations for AC1 in Figure 6 and for AC2 in Figure 7.

For AC1 (Figure 6), as clear, the general trend of the set of curves is the increase of the agreement percentage by the increase of the number of identical observations. This can be explained as follows: by the increase in the number of observations, the mean value $\tilde{q}_n$ gets closer to $\bar{q}_n$; consequently, the value of Δ decreases, which results in more agreements when Δ compared to the minimum of U_{L} and U_{F}. As evident, the minimum percentage of agreement of AC1 does not drop below 70% at worst. For most metrological practices, when three observations are recommended, the agreement percentage gets above 80%; if the number of observations is raised to five, the agreement percentage approaches 87%.

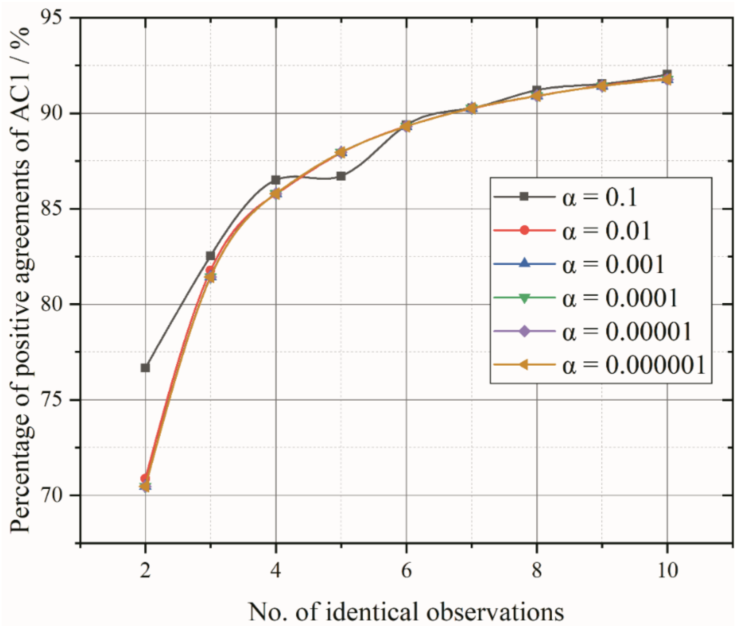

Figure 6: The percentage of positive agreement of AC1, $U_{\mathrm{L}} = 2\left(\delta/\left(6.6\sqrt{n-1}\right)\right)$.

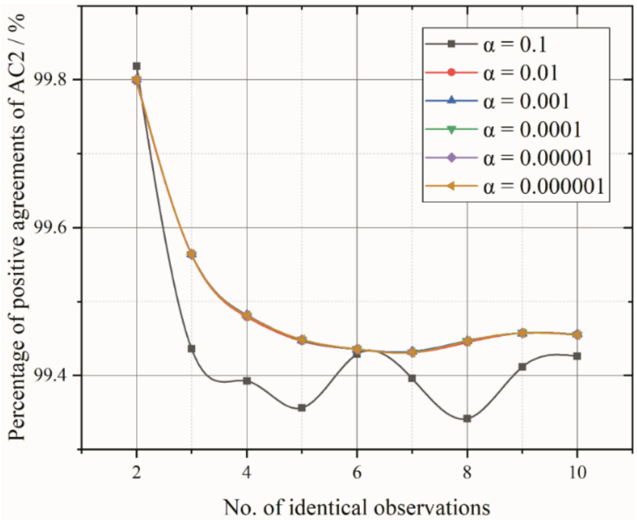

Figure 7: The percentage of positive agreement of AC2, $U_{\mathrm{L}} = 2\left(\delta/\left(6.6\sqrt{n-1}\right)\right)$.

For AC2 (Figure 7), the agreement percentages for the set of curves are relatively stable, and its variations are within a margin of width of less than 0.6%. At the same time, the agreement percentage remains 99% at worst.

Finally, it is apparent from Figure 6 and Figure 7 that the curves with $\alpha \leq 0.01$ are quite identical. So, it can be concluded that using a relatively very fine resolution measuring device with $\gamma < 0.01$ would increase the measurement cost without positively affecting the agreement percentage between them.

The simulation was extended to determine the agreement percentage of AC1 and AC2 if the GUM method was used instead of the proposed statement. Accordingly, all simulation settings were left as they were except for the value of U_{L} was set equal to $2\delta/\sqrt{12}$. The agreement percentages of AC1 are plotted in Figure 8, while the agreement percentages of AC2 were 100% at all simulated points.

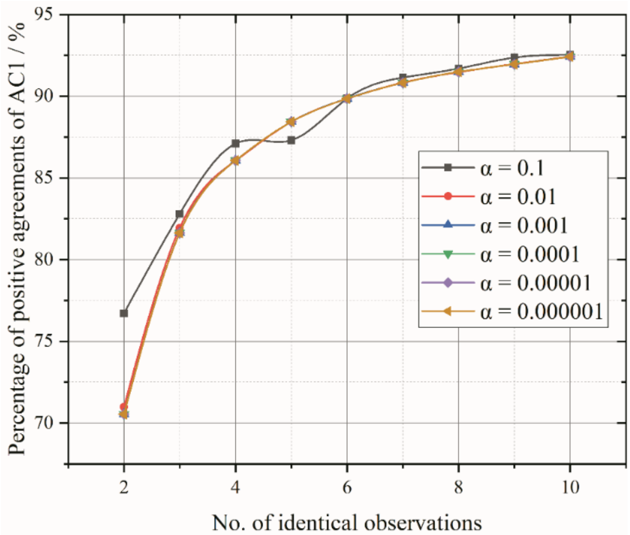

Figure 8: The percentage of positive agreement of AC1, $U_{\mathrm{L}} = 2\delta/\sqrt{12}$.

When comparing Figure 6 with Figure 8, it would be clear that the results are nearly the same. This means that the critical parameter in determining the agreement percentage of AC1 is U_{F}, i.e., U_{F} is usually the minimum when compared to U_{L}.This indicates that the Type A uncertainty calculated in the traditional way is actually less than the inferred value $\delta/\sqrt{12}$. So, it would be recommended to reduce the inferred value – based on rational assumptions – to get close to the traditional Type A value. This emphasizes the reasonability of the proposed uncertainty statement that could decrease the inferred value of identical observations without sacrificing the agreement percentage. This indicates that the inferred value did not drop below the traditional Type A uncertainty limit, which confirms the reliability of the proposed uncertainty statement.

For AC2, although the differences are trivial, they are expected. As setting U_{L} equal to $2\delta/\sqrt{12}$ which is always greater than $2\left(\delta/\left(6.6\sqrt{n-1}\right)\right)$, would result in larger values of their root sum of squares. This means more agreements when compared to Δ.

5 Case study: Calibration of force transducers according to ISO 376

According to ISO 376:2011 [9, 19], the relative repeatability uncertainty w_3 contributes to the classification of force transducers while it is calculated as:

$$w_3 = \frac{\left|\frac{X_2 - X_1}{(X_1 + X_2)/2}\right|}{\sqrt{3}} \tag{12}$$

while X_1 and X_2 are the obtained deflections in series 1 and 2 respectively, i.e., without rotating the transducer. From Eq. (12), it is obvious that the uncertainty evaluation philosophy considers the absolute difference $|X_2 - X_1|$ as the half range above which a uniform distribution is assigned.

When X_1 and X_2 are identical and equal to X, the value of $w_3 = 0$. In this case, two possible solutions can be applied to guess the value of w_3. First, it can be considered that the absolute difference between X_1 and X_2 equals the resolution of the system indicator δ, and still, apply the uniform distribution according to the equation:

$$w_3 = \frac{(\delta/X)}{\sqrt{3}} \tag{13}$$

The second choice is adopting the proposed philosophy in this article and considering that the variations in readings are owed to Gaussian disturbances. Accordingly, Eq. (10) can be reformulated to suit the application as follows:

$$w_3 = \frac{(\delta/X)}{6.6} \tag{14}$$

Thus, the reduction percentage obtained using Eq. (14) instead of Eq. (13) is approximately 73.75%, which can substantially affect the classification of the under-calibration transducer.

6 Conclusion

This article proposes a new approach for estimating the standard uncertainty of identical observations. The proposed approach considers the resolution of the measuring device and the number of identical observations in the estimation process. As merits, the proposed approach reasonably reduces the estimated uncertainty compared to the GUM method by at least 13.7% of the instrument's resolution without sacrificing the reliability of the proposed uncertainty statement. Moreover, it is simple, straightforward, and suitable to implement in laboratories' daily routine work. It was observed that the proposed uncertainty estimation decreases when the resolution of the measuring device decreases, or the number of identical observations increases. Based on the metrological assessment, when comparing the results of the proposed approach for identical observations of limited-resolution devices to the conventional Type A method calculated for finer-resolution devices, it was found that the agreement percentage did not drop below 70% at worst for the strictest criterion. For the common round-robins' criterion, the agreement percentage remained above 99%, confirming the proposed approach's validity.

Furthermore, it was noted that improving the resolution of the finer-resolution measuring device by more than 10^{-2} of the resolution of the limited-resolution device does not significantly affect the agreement percentage between them.

7 Declarations

7.1 *Study Limitations*

None.

7.2 *Funding Source*

None.

7.3 *Competing Interests*

The authors have no financial or proprietary interests in any material discussed in this article.

7.4 *Ethical Approval*

Not Required

References

[1] Evaluation of measurement data — Guide to the Expression of Uncertainty in Measurement JCGM 100:2008, BIPM et al., 2008.

[2] M3003-The Expression of Uncertainty and Confidence in Measurement M3003, UKAS, United Kingdom, 2012.

[3] H. L. Ignacio and W. Wolfgang, "The evaluation of standard uncertainty in the presence of limited resolution of indicating devices," Measurement Science and Technology, vol. 8, no. 4, p. 441, 1997. [Online]. Available: http://stacks.iop.org/0957-0233/8/i=4/a=012.

[4] E. Clemens, "Evaluation of measurement uncertainty in the presence of combined random and analogue-to-digital conversion errors," Measurement Science and Technology, vol. 11, no. 9, p. 1359, 2000. [Online]. Available: http://stacks.iop.org/0957-0233/11/i=9/a=315.

[5] R. B. Frenkel and L. Kirkup, "Monte Carlo-based estimation of uncertainty owing to limited resolution of digital instruments," Metrologia, vol. 42, no. 5, p. L27, 2005. [Online]. Available: http://stacks.iop.org/0026-1394/42/i=5/a=L01.

[6] L. Ignacio, "Resolution revisited," Metrologia, vol. 43, no. 3, p. L14, 2006. [Online]. Available: http://stacks.iop.org/0026-1394/43/i=3/a=N03.

[7] R. Kacker and A. Jones, "On use of Bayesian statistics to make the Guide to the Expression of Uncertainty in Measurement consistent," Metrologia, vol. 40, no. 5, p. 235, 2003/09/02 2003, doi: 10.1088/0026-1394/40/5/305.

[8] I. Lira and W. Wöger, "Comparison between the conventional and Bayesian approaches to evaluate measurement data," Metrologia, vol. 43, no. 4, p. S249, 2006/08/04 2006, doi: 10.1088/0026-1394/43/4/S12.

[9] ISO 376:2011 Metallic materials — Calibration of force-proving instruments used for the verification of uniaxial testing machines, International Organization for Standardization, 2011.

[10] ASME B89.1.14 - 2018 - Calipers, The American Society of Mechanical Engineering, 2018.

[11] A. O'Hagan and M. Cox, "Simple informative prior distributions for Type A uncertainty evaluation in metrology," Metrologia, vol. 60, no. 2, p. 025003, 2023/02/22 2023, doi: 10.1088/1681-7575/acb93d.

[12] R. Larson and B. Farber, Elementary Statistics: Picturing The World, Six ed. Pearson Education, Inc., 2015.

[13] S. Bell, A Beginner's Guide to Uncertainty of Measurement, United Kingdom: National Physical Laboratory, 2001.

[14] A. A. Hawam, "The Quantifying of Uncertainty in Measurement," in Handbook of Metrology and Applications, D. K. Aswal, S. Yadav, T. Takatsuji, P. Rachakonda, and H. Kumar Eds. Singapore: Springer Nature Singapore, 2022, pp. 1-31.

[15] S. Mekid, Metrology and Instrumentation: Practical Applications for Engineering and Manufacturing. Wiley, 2021.

[16] M. McCarthy, "Peak-to-Peak Resolution Versus Effective Resolution," in "Application Note," Analog Devices, Inc., 2003.

[17] W. Kester, "ADC Input Noise: The Good, The Bad, and The Ugly. Is No Noise Good Noise?," in "Analog Dialogue," Analog Devices, Inc., 2006.

[18] ASME B89.7.3.3 - 2002, Guidelines for Assessing the Reliability of Dimensional Measurement Uncertainty Statements, The American Society of Mechanical Engineers, New York, 2002.

[19] EURAMET Calibration Guide No. 4: Guidelines on the Uncertainty of Force Measurements, EURAMET, 2022.