Comparing methods for evaluating measurement uncertainty given in the JCGM ‘Evaluation of Measurement Data’ documents

L.T. Stant\textsuperscript{a,b,c}, P.H. Aaen\textsuperscript{a,b}, N.M. Ridler\textsuperscript{a,c}

\textsuperscript{a}n3m-labs, University of Surrey and National Physical Laboratory, UK
\textsuperscript{b}Advanced Technology Institute, University of Surrey, UK
\textsuperscript{c}Time, Quantum and Electromagnetics Division, National Physical Laboratory, UK

Abstract

The Joint Committee for Guides in Metrology (JCGM) publishes and maintains reference documents relating to general aspects in metrology. Working Group 1 of the JCGM is responsible for the Evaluation of Measurement Data series of documents that gives information for evaluating and expressing uncertainty in measurement. This paper compares several methods for evaluating measurement uncertainty that are described in these documents. Emphasis is given to situations where more than one input quantity is measured simultaneously. This leads to an investigation into how these methods perform when these quantities are high-frequency electromagnetic scattering parameters. It is shown that for measurements involving a large number of input quantities, such as those involving microwave scattering parameters, the required number of observations for the approach given in the GUM Supplements to work can be prohibitively large.

Keywords: measurement uncertainty, GUM, GUM supplements, microwave scattering parameters, standard uncertainty

PACS: 06.20.Dk, 06.30.Ka

1. Introduction

The Guide to the Expression of Uncertainty in Measurement (GUM) \cite{1} was first published in 1993 by the International Organisation for Standardization (ISO) on behalf of seven international organisations: the International Bureau of Weights and Measures (BIPM), the International Organization of Legal Metrology (OIML), the International Electrotechnical Commission (IEC), the International Organization for Standardization (ISO), the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC), the International Union of Pure and Applied Chemistry (IUPAC), and the International Union...
of Pure and Applied Physics (IUPAP). The publication of this document was a landmark event in the field of measurement uncertainty because the document presented, arguably for the first time, a comprehensive and standardised approach for evaluating uncertainty, irrespective of the type of measurement and the level of accuracy being sought. In addition, the endorsement of the document by the seven international organisations (including the two global standardisation bodies - IEC and ISO) gave the document significant international authority in many areas of the physical sciences.

In 1997, the Joint Committee for Guides in Metrology (JCGM) was established by these seven international organisations. The JCGM was tasked with maintaining both the GUM [1] and the International Vocabulary of Metrology - Basic and General Concepts and Associated Terms (VIM) [2]. The JCGM consists of two Working Groups (WGs): WG1 has responsibility for the GUM; WG2 has responsibility for the VIM. Membership of the JCGM comprises representatives of the above international organisations (referred to as the JCGM Member Organisations). In 1998, the International Laboratory Accreditation Cooperation (ILAC) became the eighth JCGM Member Organisation. In 2008, JCGM WG1 re-published the GUM [3]. Around the same time, JCGM WG1 began publishing additional documents relating to measurement uncertainty. This included supplements to the GUM [4, 5] covering topics that are not dealt with, in detail, in the GUM. The first supplement [4] introduced the use of a Monte Carlo method [6] to propagate distributions representing the uncertainty contributions. The second supplement [5] presented a detailed treatment of situations where more than one measurand is determined simultaneously during the measurement process.

With regard to methodologies for evaluating measurement uncertainty, the GUM presents methods that include the use of both Bayesian probabilistic methods [7] and classical probabilistic methods [8] to evaluate the uncertainty in the individual components of a measurement’s overall uncertainty. An informative discussion on these types of method can be found in [9]. In the GUM, classical methods are used for the treatment of Category A uncertainty components and Bayesian methods are used for the treatment of Category B uncertainty components. Category A uncertainty components are those that are evaluated using statistical techniques; Category B components are those that are evaluated using other means. Since the publication of the GUM, some authors have stated (for example, in [10, 11, 12, 13]) that this combination of different probabilistic methods (i.e., Bayesian and classical) represents an inconsistency in the GUM methodology for evaluating measurement uncertainty.

The supplements to the GUM [4, 5] resolve the above-mentioned inconsistency by introducing a method for treating the Category A uncertainties that follows a Bayesian approach [14]. Therefore, the two supplements no longer contain the inconsistency found in the original GUM. However, as a consequence of this change, there is now inconsistency between the method used to evaluate uncertainty described in the GUM and that described in the two supplements. In many situations, these different methods do not have a significant impact on the overall uncertainty that is evaluated. For situations where a consider-
able number of input quantities are observed simultaneously, the two different approaches can produce significantly different values of uncertainty. Such situations often occur in the area of high-frequency electromagnetic metrology. This paper examines in detail these procedural differences and, for the first time, illustrates their impact on measurement situations that occur in high-frequency electromagnetic metrology.

The paper is arranged in the following sections. Section 2 reviews the procedures given in the GUM for the treatment of Category A uncertainty components. Section 3 reviews the related procedures in the GUM supplement documents for Category A uncertainty components. Section 4 shows how these two procedures impact an example that is given in both the GUM and the supplements - namely, the simultaneous measurement of resistance, reactance and impedance. Section 5 shows how the two procedures, in general, affect measurements involving significant numbers of input quantities. This is illustrated with an example using measurements of microwave scattering parameters. Section 6 follows with a discussion concerning the significance of the differences observed using these two methods. Finally, the summary and conclusions relating to the work are presented in Section 7.

2. JCGM 100:2008 - GUM

The classical statistical technique [8] applied to Category A uncertainties in the current GUM [3] is based on a series of observations of a randomly varying input quantity. After $n$ observations $x_1, x_2, \ldots, x_n$, the arithmetic mean of measured values, $\bar{x}$, and standard deviation, $s$, of a randomly varying input quantity, $X$, is written as

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i, \quad s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2}$$

(respectively, where $x_i$ is the result of the $i$th observation. Importantly, a minimum of two observations must be made ($n \geq 2$) in order for $s$ to be defined.

The standard uncertainty of the best estimate of $X$, $u(\bar{x})_{GUM}$, can be found by dividing $s$ by the square root of the number of observations:

$$u(\bar{x})_{GUM} = \frac{s}{\sqrt{n}}$$

If there are correlated (mutually dependent) input quantities present in the measurement model, the covariances of each pair of input quantities must also be calculated before the propagation stage of the uncertainty evaluation. Both the standard uncertainties and the covariances for $N$ input quantities can be represented in a symmetric ($N \times N$) matrix containing the variance ($s^2$) of each quantity along the diagonal and the covariance between $x_i$ and $x_j$ in the $i,j$th element of the matrix. This is called the “uncertainty matrix” in [4] and the “measurement covariance matrix” in [5]. An example given in the GUM, Example H.2, described later in Section 4 of this paper, demonstrates this scenario
using the example of a simultaneous measurement of resistance and reactance with voltage, current and phase as correlated input quantities.

Once the uncertainties of the input quantities have been evaluated, they are propagated through the measurement model. This requires the sensitivities of the measurand to each input quantity to be calculated to at least a first order approximation. The estimates of the input quantities are used in the measurement model to obtain the estimate of the measurand. The variances and covariances of the input quantities are combined with the sensitivity coefficients in order to obtain the variance of the measurand. The combined standard uncertainty of the measurand is equal to the positive square root of this value. The result of the measurement is then presented as the measurand estimate and combined standard uncertainty. Alternatively, the uncertainty is expressed in terms of an expanded uncertainty which is derived directly from the combined standard uncertainty.


Both GUM supplements (GUM-S1/S2) [4, 5] use a Bayesian approach [15] to assign a probability density function (PDF) to describe all input quantities. This approach results in the choice of a $t$-distribution to characterise Category A input quantities, in contrast to the approach used in the GUM. Of particular relevance to this paper is the inclusion of the degrees-of-freedom parameter, $\nu$, in the definition of the standard uncertainty and covariances of a $t$-distribution. Whereas for the Gaussian distribution $\nu$ is used as a measure of reliability of the standard uncertainty, it is explicitly required when using the $t$-distribution in order to obtain the standard uncertainty, $u(\bar{x})_{SUPP}$:

$$u(\bar{x})_{SUPP} = \frac{s}{\sqrt{n}} \times \sqrt{\frac{\nu}{\nu - 2}}$$

(3)

where $\nu = n - N$, with $n$ being the number of repeat observations used to obtain a Category A evaluation of uncertainty and $N$ being the number of input quantities. In the GUM-S1 only a univariate $t$-distribution is offered, which represents $N = 1$ input quantities. For this case (3) can be rewritten as:

$$u(\bar{x})_{SUPP} = \frac{s}{\sqrt{n}} \times \sqrt{\frac{n - 1}{n - 3}}$$

(4)

Equation (4) is undefined if $n$ is less than four. This effectively prevents the standard uncertainty from being calculated for a single input quantity according to the guidance given in the GUM-S1 (and the GUM-S2). The commercial ramifications of this condition are significant and are discussed further in Section 6. Figure 1 illustrates the ratio between the standard uncertainty values calculated for different numbers of observations of a single Category A input quantity using the GUM and the GUM-S1/S2 approaches. It can be seen that when $n = 4$, $u(\bar{x})_{SUPP} = \sqrt{3} \times u(\bar{x})_{GUM}$, and as the number of observations increases the results from both approaches converge: If $n$ tends to infinity, the
\[ V(X) = \frac{\nu}{\nu - 2} \frac{S}{n} = \frac{1}{n(n - N - 2)} \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})^\top \]  
\[ S = \frac{1}{\nu} \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})^\top \]  
\[ V(X) = \begin{bmatrix} u(x_1)^2 & u(x_1, x_2) & \ldots & u(x_1, x_n) \\ u(x_2, x_1) & u(x_2)^2 & \ldots & u(x_2, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ u(x_n, x_1) & u(x_n, x_2) & \ldots & u(x_n)^2 \end{bmatrix} \]  

Figure 1: Scaling factor to convert from a GUM standard uncertainty to a GUM Supplement standard uncertainty, as a function of number of observations.

The \( t \)-distribution tends towards a Gaussian distribution. However, most commercial laboratories would avoid making large numbers of measurements as this reduces the efficiency of the process.

For measurements involving multiple input quantities, such as the measurement of a vector quantity, a multivariate/joint distribution should be used as suggested in the GUM-S2. The variances and covariances between all pairs of input quantities are obtained using a matrix form of (3) (section 5.3.2 of [5]):

\[ V(X) = \frac{\nu}{\nu - 2} \frac{S}{n} = \frac{1}{n(n - N - 2)} \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})^\top \]  
\[ S = \frac{1}{\nu} \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})^\top \]  
\[ V(X) = \begin{bmatrix} u(x_1)^2 & u(x_1, x_2) & \ldots & u(x_1, x_n) \\ u(x_2, x_1) & u(x_2)^2 & \ldots & u(x_2, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ u(x_n, x_1) & u(x_n, x_2) & \ldots & u(x_n)^2 \end{bmatrix} \]  

where \( V(X) \) is the uncertainty matrix, \( x_i \) is a sample from the array of vectors containing input quantity indications and \( \bar{x} \) is the arithmetic mean of that array. For this multivariate case, the minimum value of \( n \) will increase linearly with \( N \), such that the standard uncertainty is undefined unless \( n > N + 2 \). The consequences of this condition are demonstrated in Section 5.

4. Example - H.2 of the GUM/9.4 of the GUM-S2

Both the GUM and the GUM-S2 provide an identical example which can be used to demonstrate the different standard uncertainties obtained when applying the method suggested in each document. The example is a simultaneous
Table 1: The indication values from the example “Simultaneous Resistance and Reactance Measurement” and their statistical properties as evaluated by the approaches given in [3] (example H.2) and [5] (example 9.4).

<table>
<thead>
<tr>
<th>Value</th>
<th>V/V</th>
<th>I/A</th>
<th>Φ/rad</th>
</tr>
</thead>
<tbody>
<tr>
<td>x₁</td>
<td>5.007</td>
<td>19.663</td>
<td>1.0456</td>
</tr>
<tr>
<td>x₂</td>
<td>4.994</td>
<td>19.639</td>
<td>1.0438</td>
</tr>
<tr>
<td>x₃</td>
<td>5.005</td>
<td>19.640</td>
<td>1.0468</td>
</tr>
<tr>
<td>x₄</td>
<td>4.990</td>
<td>19.685</td>
<td>1.0428</td>
</tr>
<tr>
<td>x₅</td>
<td>4.999</td>
<td>19.678</td>
<td>1.0433</td>
</tr>
<tr>
<td>x₆</td>
<td>4.999</td>
<td>19.661</td>
<td>1.0445</td>
</tr>
<tr>
<td>¯x</td>
<td>4.9990</td>
<td>19.6610</td>
<td>1.04446</td>
</tr>
</tbody>
</table>

\[ u(\bar{x})_{GUM} = 0.0026, \quad u(\bar{x})_{SUPP} = 0.0045 \]

\[ u(\bar{x})_{GUM} / u(\bar{x})_{SUPP} = 1.732 \]

The measurement of resistance and reactance, which uses a measurement model with multiple input quantities and multiple output quantities (measurands). The input quantities are voltage \( V \), current, \( I \), and phase, \( Φ \), and the measurands are resistance \( R \), reactance, \( X \), and impedance, \( Z \). The measurement model is defined as:

\[ R = \frac{V}{I} \cos θ, \quad X = \frac{V}{I} \sin θ, \quad Z = \frac{V}{I} \quad (8) \]

Six sets of indication values [2] \( (n = 6) \) of \( V, I, Φ \) are obtained independently by measurement. The version of this example given in the GUM uses only \( n = 5 \) sets, but one additional set of values of \( V, I, Φ \) has been added for the GUM-S2 example to allow (3) to be defined for \( N = 3 \) input quantities, a condition which was explained at the end of Section 3. These values, together with their arithmetic means and standard uncertainties as calculated from the two approaches using (2) and the matrix form of (3) (which is applicable to measurements involving multiple input quantities), are presented in Table 1. The ratios of the standard uncertainties from each approach is also included in the table, which are identical for all these input quantities due to their dependence only on \( n \) and \( N \), which are also equal for all these input quantities (e.g., when \( n = 6 \) and \( N = 3 \), \( \sqrt{n/(n-2)} = \sqrt{(n-N)/(n-N-2)} = \sqrt{3} \)). This explains why standard uncertainties evaluated with Category A methods using the minimum number of observations following the GUM-S1/S2 approach are always 1.732 times larger than the standard uncertainties calculated following the GUM approach.

This difference in the input quantity uncertainties calculated from the two approaches propagates through the measurement model and therefore significantly affects the combined standard uncertainties of the measurands. Table 2
Table 2: A comparison of the results obtained for the example “Simultaneous Resistance and Reactance Measurement” using the approaches given in [3] (example H.2) and [5] (example 9.4).

<table>
<thead>
<tr>
<th>Method</th>
<th>( u(R)/\Omega )</th>
<th>( u(X)/\Omega )</th>
<th>( u(Z)/\Omega )</th>
</tr>
</thead>
<tbody>
<tr>
<td>GUM</td>
<td>0.058</td>
<td>0.241</td>
<td>0.193</td>
</tr>
<tr>
<td>GUM-S2</td>
<td>0.130</td>
<td>0.540</td>
<td>0.431</td>
</tr>
<tr>
<td>( GUM-S2 )</td>
<td>2.241</td>
<td>2.241</td>
<td>2.233</td>
</tr>
</tbody>
</table>

presents the combined standard uncertainties of the measurands for the described example as evaluated by both approaches, together with a ratio of the uncertainty values. For all three measurands the combined standard uncertainty calculated using the GUM-S1/S2 method is more than double the equivalent values calculated using the GUM method. For other measurement models with higher sensitivities to the input quantities, this difference could be even greater.

5. Application to microwave scattering parameters

High-frequency electromagnetic metrology often involves using multiple complex-valued quantities. Common input quantities for this type of measurement, measured using instruments such as vector network analysers (VNA), are scattering parameters (S-parameters). These measurements characterise the linear response of a microwave device to a stimulus on a specified connection (port), and can be used to calculate important device parameters such as gains and reflection coefficients. A single S-parameter represents either a transmission or reflection coefficient for a pair of ports, and therefore a device with \( m \) ports requires \( m^2 \) S-parameters to fully characterise the response of the device:

\[
S = \begin{bmatrix}
S_{11} & S_{12} & \cdots & S_{1m} \\
S_{21} & S_{22} & \cdots & S_{2m} \\
\vdots & \vdots & \ddots & \vdots \\
S_{m1} & S_{m2} & \cdots & S_{mm}
\end{bmatrix}
\] (9)

Because each S-parameter is a complex-valued quantity \((S = (S_{Re}, S_{Im}))\), there are \( 2m^2 \) input quantities required in a measurement model which uses the complete device response. All of these quantities are correlated, so a multivariate distribution should be used to represent them. It has been shown in Section 4 that for a Category A evaluation of uncertainty, both the number of repeat observations and the number of input quantities have a significant effect on the difference in uncertainty as calculated from the two approaches presented in the GUM and the GUM-S1/S2. Table 3 shows the ratio of uncertainties calculated from both approaches when applied to a measurement using scattering parameters obtained from the minimum number of repeat observations, \( n \), for devices with \( m \) ports.
Table 3: The difference in standard uncertainties obtained using the GUM \( u(\bar{x})_{GUM} \) and the GUM-S1/S2 \( u(\bar{x})_{SUPP} \) approaches to measure a full set of scattering parameters for microwave devices with various numbers of ports, \( m \). Each device has \( 2m^2 \) input quantities, \( N \), and requires a minimum of \( N + 3 \) repeat observations, \( n \), in order for \( u(\bar{x})_{SUPP} \) to be defined.

<table>
<thead>
<tr>
<th>Ports, ( m )</th>
<th>Input quantities, ( N )</th>
<th>Required minimum number of repeat observations, ( n ), for ( u(\bar{x})_{SUPP} ) to be defined.</th>
<th>( \frac{u(\bar{x})<em>{SUPP}}{u(\bar{x})</em>{GUM}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>5</td>
<td>1.732</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>11</td>
<td>1.732</td>
</tr>
<tr>
<td>3</td>
<td>18</td>
<td>21</td>
<td>1.732</td>
</tr>
<tr>
<td>4</td>
<td>32</td>
<td>35</td>
<td>1.732</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>8</td>
<td>128</td>
<td>131</td>
<td>1.732</td>
</tr>
</tbody>
</table>

It can be seen that for devices with multiple ports, \( n \) can become large in order for (3) to be defined and calculate the standard uncertainty. It is likely that the user will not always have the time or resources available to perform such a quantity of measurements. In microwave measurement environments, connections are typically made by hand using coaxial connectors. A typical measurement may include a Category A evaluation of uncertainty due to connection repeatability. Considering the specific example of a 4-port device, this requirement would result in the need for a minimum of \( 35 \times 4 = 140 \) repeat coaxial connections to be made in order to perform a Category A evaluation of the standard uncertainty using the GUM-S1/S2 approach. By contrast, the classical approach used in the GUM is defined with just 2 repeat observations, which would require only \( 2 \times 4 = 8 \) repeat coaxial connections to be made. Figure 2 shows the minimum number of repeat observations required when using the GUM-S1/S2 approach, \( n \), in order to be able to calculate a Category A evaluation of the standard uncertainty of a full set of S-parameters for a microwave device with \( m \) ports. In all cases, the standard uncertainty obtained using the GUM-S1/S2 approach is approximately 1.7 times larger than that obtained using the GUM approach.

6. Discussion

The inconsistency of the approaches used in the GUM and its supplements to calculate the standard uncertainty of Category A input quantities of a measurement has two noticeable consequences for the user:

1. There can be a large difference in the standard uncertainties reported by each approach, which is demonstrated using both the example of simultaneous resistance and reactance measurement shown in Section 4 and a typical microwave measurement shown in Section 5. This leads to the
question: “Which approach should be used?” The answer is not straightforward. The GUM approach is likely to be more attractive to commercial laboratories and test engineers since this leads to achieving smaller uncertainties in their results.

2. For situations involving multiple Category A input quantities, the Bayesian approach introduced in the GUM-S1/S2 can require a large number of observations before the standard uncertainty can be defined. Although the standard uncertainty calculated using the GUM approach will become less reliable with fewer observations, it is still possible to obtain a result with only 2 observations of any number of input quantities. In a commercial laboratory the additional measurements required by the GUM-S1/S2 approach can be impractical, with many laboratories typically using only two or three measurements per device following the GUM approach. For a single input quantity this would require a potential doubling of the number of observations and therefore the test duration, which would either slow throughput or require more test stations to be added. If implemented, the additional time or financial investment would then produce uncertainties that are significantly larger than those obtained using the GUM approach.
7. Summary and Conclusion

This paper has highlighted an inconsistency between the GUM and its supplements when evaluating the standard uncertainty for Category A input quantities of a measurement model. The classical approach used in the GUM and the Bayesian approach used in the GUM supplements have both been described and their results compared using two examples. The supplement approach was shown not to work for measurements with less than four observations of the input quantities. The standard uncertainties provided by the supplement approach is always higher than those provided by the GUM approach, by a scaling factor with an upper limit of 1.732. Finally, for measurements involving a large number of input quantities, such as microwave scattering parameters, the required number of observations for the supplement approach to work can be prohibitively large. In contrast, the GUM approach only requires two observations for any number of input quantities.

Acknowledgements

This work was part-funded through the European Metrology Programme for Innovation and Research (EMPIR) Project 14IND10, MET5G. The EMPIR is jointly funded by the EMPIR participating countries within EURAMET and the European Union.

References


