Measurement Uncertainty

# 3.1 Introduction

Foreword: Parts of this chapter are based on excerpts or taken verbatim from the authors paper [1].

A measurement is an observation of a physical effect or quantity which provides useful information. This information, through the ages, has been used to facilitate advancement of both scientific knowledge and industrial development – from the production of standardised stone blocks to build the pyramids of ancient Egypt, to the production of standardised car parts to build Henry Ford’s Model T. In the scientific realm, advanced measurement techniques at laboratories such as CERN are used to convince the world that new subatomic particles exist.

To successfully communicate information about a measurement, the recipient needs to be able to either make or mentally construct a similar observation to that of the original measurer (or *metrologist*). The simplest way of doing this is to provide the recipient with the same physical effect or quantity for which to make their own observation (if you require a new nut for a bolt from a hardware shop, you might intuitively take the bolt with you), however, this can be inconvenient or impractical with larger objects, or if the recipient is located far away. Instead, you might substitute the physical effect or quantity with a more portable representation. For example, if you were to measure the size of a doorway to see if a new piece of furniture may fit through it, you might cut a piece of string to the same length and use this as the representation of the width of the item. However, this approach is very wasteful and also impractical for many physical effects (temperature, flow, pressure).

A solution widely thought to have been first established in the 3rd or 4th Millennium BC, is a system of units. In such a system, a discretised value of a physical effect or quantity is standardised and knowledge of its value is disseminated to all people who wish to use it. Typically, a range of discrete values are chosen, such that the system of units can be conveniently used to represent all measurements. Knowledge of the discretised values is obtained from a primary standard which becomes the definition of the unit and is used to create copies of the standard which can be given to users of the unit system to perform measurements with. The most common method of performing measurements with a unit system is to use a standard to calibrate a measuring instrument, which can then be used to measure an arbitrary value of a physical effect or quantity in the units defined by the standard.



Figure 1: Egyptian royal cubit rod of Maya (treasurer of King Tutankhamun) 1336 – 1327 BC. The cubit is thought to have been first used in the 3rd or 4th Millennium BC.

The introduction of a regulated system of units enables commerce, as traded goods can be reliably valued between merchants across cities. This application is encountered by all citizens, and so there is a high demand for standards to be produced from the primary standard and distributed far and wide. It becomes impractical to create all standards by copying the primary standard directly (in some cases because the value of the primary standard is perturbed each time it is measured), and so a tiered structure of standards is used. In this structure, there is a tier consisting of a small number of standards which are created directly from measurements of the primary standard, followed by subsequent tiers of larger numbers of standards which are derived from measurements of those in the tier above. For any produced standard, it should be possible to trace the lineage back to a measurement of the golden standard. This is referred to as a traceability chain (see Fig. 2) and it is a fundamental tenet of metrology. Measurements with a shorter traceability chain are considered more traceable than those with longer chains.

Traceability

Figure 2: The traceability chain.

Today, the primary standards are maintained in most countries by a National Measurement Institute (NMI) and co-ordinated by the Bureau of International Weights and Measures (BIPM). To accommodate international trade and compatibility, a routine process of inter-comparisons is undertaken to ensure that the primary standards between countries are agreeable.

Secondary standards are also kept by the NMIs and are calibrated against the primary standard as infrequently as possible. These standards are then used by the NMI to calibrate working standards which are sent to them by manufacturers and research institutes. Another task of each NMI is to investigate new and improved methods of measurement. Secondary standards are also typically used at NMIs during these investigations.

Working standards are used by manufacturers in different ways. Instrumentation manufacturers may use the standards to directly calibrate their products before sending it to the customer, while more generally the standards can be used to calibrate test equipment which identifies faulty products. Larger research institutes typically use working standards to recalibrate instrumentation prior to performing very sensitive measurements. To ensure that product specifications and scientific measurements are traceable and of high quality, accreditation services such as the United Kingdom Accreditation Service (UKAS) [2] exist to certify manufacturers and laboratories that demonstrate good measurement practice and use traceable measurements.

The selection of physical effects and quantities for which primary standards are kept is only a subset of those for which recognised units exist. This is because many units are derived quantities, where their value can be obtained by calculation using definitions of other units. For example, the definition of the unit of resistance (R, ohms) can be derived from that of voltage (V, volts) and current (I, amperes), because R=V/I. The eight fundamental “base” units which make up the International System of Units (SI), are the metre, kilogram, second, ampere, kelvin, candela and mole. From these unit definitions, it is possible to define any other derived unit in use. NMIs will usually keep secondary standards of most derived quantities which users may wish to calibrate against, which are themselves traceable to one or more primary standards of different units. Although traditionally all primary standards were defined by physical artefacts, these are being gradually replaced by definitions involving physical constants (e.g. Plank, Boltzmann), which do not degrade over time or use. The “Ninth SI Units” [3], a proposition currently under review by the BIPM, is the redefinition of four of the SI units (the ampere, the kilogram, the kelvin and the mole) which may come into effect by May 2019.

What is the difference between a measurement with good traceability compared to one with poor traceability? If they are traceable to primary standards, then are they not both a correct representation of some value in a unit system? The answer is that confidence in the result of each measurement may differ, which is caused by measurement uncertainty.

It is impossible to know the true value of a quantity being measured. Many physical effects are involved during the measurement process, of which only a few are useful. All other processes contribute error to the measurement, causing a reduction in accuracy (the observed value deviating from the true value). Typical sources of error in measurement are thermal noise (electronic measurements), parallax (bifocal meter reading), and drift of environmental conditions from those at which a measuring instrument was calibrated. In some cases, it is possible to quantify and correct for these errors, but there are often many sources (some of which contribute very small errors) which cannot be corrected for. This is due either because the error cannot be quantified, or because the value of the error will change over the duration of the measurement process (random errors). Any source of error which cannot be removed from a measurement becomes a source of uncertainty, because the deviation of the measured value from the true value due to this source of error is uncertain. If it is possible to quantify the amount of uncertainty in a measurement, then a degree of confidence can be formed about its value.

If every measurement has an associated uncertainty in its value, then any measurement involving the results of previous measurements will include uncertainty contributions from both measurements. The earlier question can now be answered – measurements with good traceability involve fewer sources of uncertainty than those with poor traceability, leading to a higher degree of confidence in the former. It is because of this fact that NMIs strive to reduce the uncertainties in their primary standard definitions, which in turn reduces the uncertainty in all traceable measurements.

Because the amount of error in a source of uncertainty is unknown, probability and statistical theories are used to describe the amount of uncertainty associated with it. By the nature of these theories there are often several methods which can be used to obtain a result, providing different values. To ensure consistency and portability of uncertainty definitions, measurement guides were created in each industry and area of science, which specialised in processing the results of typical measurements. In addition, different guides were produced depending on the level of accuracy required – as more accurate measurements often require more effort to complete. Although this practice allowed suitable trade and measurement comparisons within each field, ambiguities existed in uncertainty definitions between fields. To address this, a landmark document was published in 1993 by the International Organisation for Standardisation (ISO), the Guide to the Expression of Uncertainty in Measurement (GUM) [4]. This document was the work of representatives from seven international organisations: the BIPM, the International Organisation of Legal Metrology (OIML), the International Electrotechnical Commission (IEC), the 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 GUM, updated in 2008 [5], is still used today as a reference for the evaluation of measurement uncertainty in many laboratories and industries across the world. The seven original organisations which wrote the GUM, together with the International Laboratory Accreditation Cooperation (ILAC, of which UKAS is a member), form the Joint Committee for Guides in Metrology (JCGM), who maintain the GUM and subsequent additional documents. These additional documents consist of the International Vocabulary of Metrology (VIM) [6] and two supplements to the GUM [7,8]: Supplement 1 covers the use of a Monte Carlo method [9] in uncertainty evaluation; Supplement 2 covers the situation where more than one quantity is measured at the same time (multivariate).

Throughout this thesis, the methodologies presented in the GUM will be used. The international authority of the guide, developed by the endorsement of the seven international organisations behind it (including the two global standardisation bodies IEC and ISO), gives strong motivation to use it as a basis for a framework to evaluate uncertainty in measurement.

# 3.2 The Measurement Process

In contrast to basic evaluations of uncertainty, where only repeat measurements of the particular quantity are used, the GUM prescribes a more rigorous approach which models the measurement process and propagates uncertainty through the model to the result (measurands). This allows any uncertainties from previous measurements, including those involving standards in the traceability chain, to be correctly included in the result. The measurement model can be simple, such as measuring resistance using input quantities of voltage and current, or complicated and multivariate, requiring many input quantities and producing many output quantities. In some cases, the measurement model may not be known and can be defined as a black box, but this has certain limitations discussed later in Section 3.5.

The GUM defines a process that is to be followed when evaluating uncertainty in measurement. It consists of the following steps:

1. Modelling the measurement.
2. Evaluating standard uncertainty of input quantities.
3. Determining combined standard uncertainty of the measurands.
4. Determining expanded uncertainty of the measurands.

Each of these steps will now be explained.

# 3.3 Modelling the Measurement

We can define a set of measurands ***Y*** as a functional relationship depending on *N* other input quantities *X1,**X2,**…,**XN*:

The estimate of the measurands can therefore be found by evaluating the model using the estimates of each input quantity:

Each input quantity could either be obtained during this measurement, from a previous measurement, or from another source of information such as a datasheet or specification. An example of a measurement model could be for a temperature measurement, where the functional relationship includes the value observed from the meter, the measured values of two calibration temperatures, and the assumed values of those calibration temperatures. Using this method, uncertainty from the calibration can be correctly included in the evaluation. This is especially true for uncertainties caused by systematic errors, which do not vary during the measurement process, and so cannot be evaluated purely by performing repeat measurements.

# 3.4 Input Quantities

Components of uncertainty in measurement can be divided into two categories: Category A uncertainty components are those that are evaluated using statistical analysis of a series of observations (i.e. repeats); Category B components are those that are evaluated using other means.

The GUM presents methods that include the use of both Bayesian and classical probabilistic methods to evaluate the uncertainty in the input quantities for a measurement model. In particular, classical methods [10] are used for the treatment of Category A uncertainty components and Bayesian methods [11] are used for the treatment of Category B uncertainty components. An informative discussion on these types of method can be found in [12]. Since the publication of the GUM, some authors have stated (for example, in [13-16]) 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 [7, 8] resolve the above-mentioned inconsistency by introducing a method for treating the Category A uncertainties that follows a Bayesian approach [17]. 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 considerable 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, which is the topic of this thesis.

## 3.4.1 Category A Evaluation

### GUM Method

The classical statistical technique [10] applied to Category A uncertainties in the current GUM assigns a Gaussian probability distribution to a series of observations of a randomly varying input quantity, itself represented by a Gaussian distribution. Therefore, after *n* observations *x*1, *x*2, …, *x*n, the best available estimate (arithmetic mean of measured values), *x*, and standard deviation, s, of a randomly varying input quantity, *X*, is written as

respectively, where *x*i is the result of the *i*th observation. Importantly, a minimum of two observations must be made (n = 2) in order for and *s* to be defined. The standard uncertainty of the best estimate of *X*, u()GUM, can be found by dividing *s* by the square root of the number of observations:

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 conveniently represented in a symmetric (*N* x *N*) matrix containing the variance of each quantity (*s*2) along the diagonal and the covariance between *x*i and *x*j in the *i*, *j*th element. This is called the “uncertainty matrix” in the GUM and the “measurement covariance matrix” in the GUM Supplement 2. An example given in the GUM, Example H.2, described later in this chapter, 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.

### GUM Supplements Method

Both GUM supplements (GUM-S1/S2) [7, 8] use a Bayesian approach [11] to assign a probability density function (PDF) to describe all input quantities. This approach results in the choice of a *t*-distribution to characterize Category A input quantities, in contrast to the Gaussian distribution used in the GUM. Of particular relevance to this paper is the inclusion of the degrees-of-freedom parameter, ν, in the definition of the standard uncertainty and covariances of a *t*-distribution. Whereas for the Gaussian distribution ν 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(x)SUPP:

where ν=*n-N*, with *n* being the number of observations 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 the previous equation can be rewritten as:

Equation (3.6) 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 later in this chapter. 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()SUPP = √3 x u()GUM, and as the number of observations increases the results from both approaches converge: If *n* tends to infinity, 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.

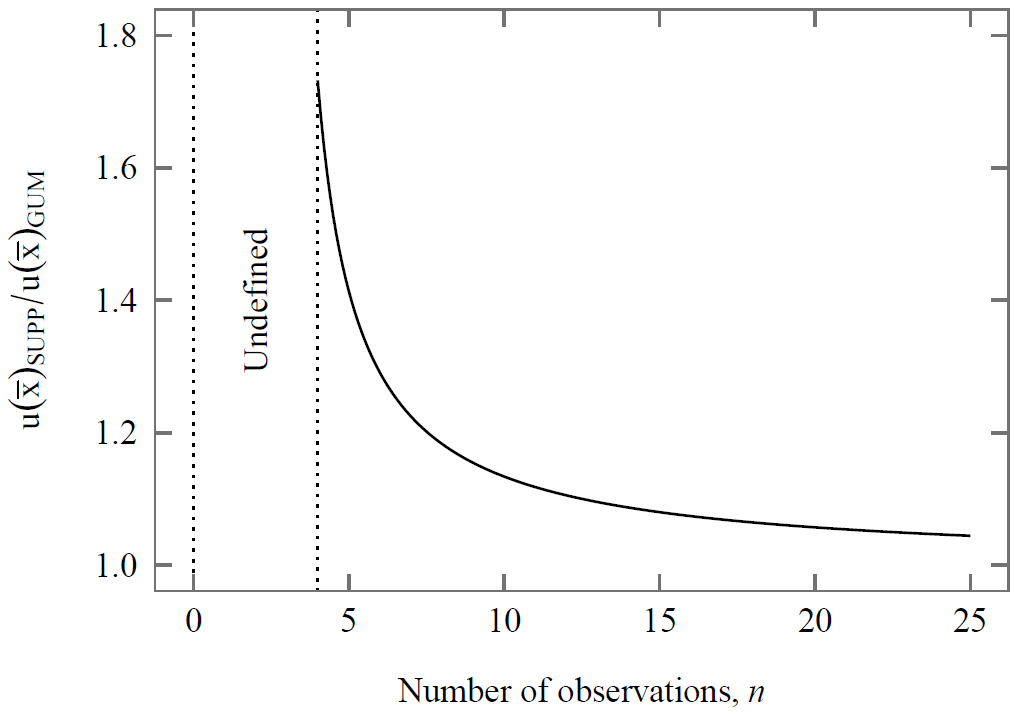


Figure 3: Scaling factor to convert from a GUM standard uncertainty to a GUM Supplement.

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 (3.5) (section 5.3.2 of [8]):

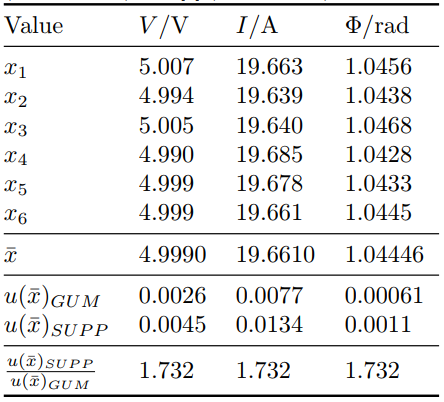
where ***V***(***X***) is the uncertainty matrix, ***x***i is a sample from the array of vectors containing input quantity indications and is the arithmetic mean of that array. For this multivariate case, the minimum value of n will increase linearly with *N*, such that the standard uncertainty is undefined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

### Comparison of GUM and GUM Supplements approach using example H.2/9.4

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 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:

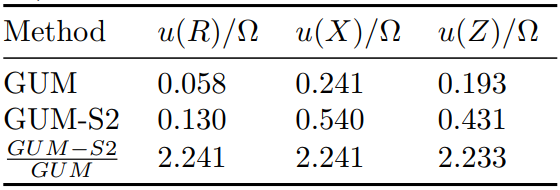
Six sets of indication values [6] (*n* = 6) of *V*; *I*; φ are obtained independently by measurement. The version of this example given in the GUM uses only *n* = 5 sets, but one additional set of values of *V*; *I*; φ has been added for the GUM-S2 example to allow (3.5) to be defined for *N* = 3 input 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 (3.4) and the matrix form of (3.5) (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, . 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.

Table 1: The indication values from the example “Simultaneous Resistance and Reactance Measurement” and their statistical properties as evaluated by the approaches given in [5] (example H.2) and [8] (example 9.4).



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 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.

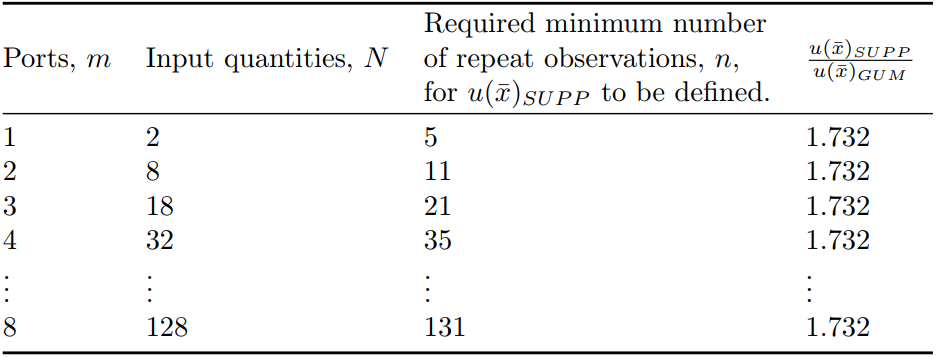
Table 2: A comparison of the results obtained for the example “Simultaneous Resistance and Reactance Measurement” using the approaches given in [5] (example H.2) and [8] (example 9.4).



### Comparison of GUM and GUM Supplements approach using microwave scattering parameters example

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), as described in Chapter 2. Because each S-parameter is a complex-valued quantity (S = (SRe, SIm)), there are 2m2 input quantities required in a measurement model which uses the complete device response. All these quantities are correlated, so a multivariate distribution should be used to represent them. It has been shown previously 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()GUM) and the GUM-S1/S2 (u()SUPP) approaches to measure a full set of scattering parameters for microwave devices with various numbers of ports, *m*. Each device has 2*m*2 input quantities, *N*, and requires a minimum of *N* + 3 repeat observations, *n*, in order for u()SUPP to be defined.



It can be seen that for devices with multiple ports, *n* can become large in order for (3.5) 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 × 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 × 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.

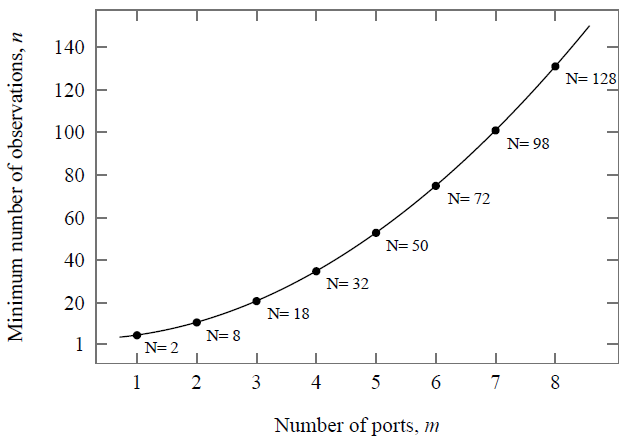


Figure 4: The minimum number of observations, *n*, required to calculate the standard uncertainty of a full set of S-parameters for a microwave device with m ports using the GUM-S2 approach. The number of input quantities, *N*, for each device is also shown. Because n is minimized, *n* = *N* + 3.

### 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 has been demonstrated in this chapter using both the example of simultaneous resistance and reactance measurement and a typical microwave measurement. 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 two 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.

For the work in this thesis, which is based on multivariate electromagnetic measurement problems, the GUM approach is used. In addition, an existing software framework, introduced later, which is included as part of the complete framework presented in this work, also uses the GUM approach for processing Category A uncertainty components.

## 3.4.2 Category B Evaluation

Category B uncertainty components are those which have not been obtained by repeated measurements. The GUM presents a list of possible sources:

* Previous measurement data;
* Experience or knowledge of relevant materials and instruments;
* Manufacturer’s specifications;
* Data provided by calibration and other certificates;
* Reference data from handbooks;

Values obtained from these sources will typically be an estimate accompanied by either a standard uncertainty or an expanded uncertainty. The latter can be converted to a standard uncertainty, the process of which is described in section 3.6. Once the estimate and standard uncertainty of the quantity are known, they can be used as input quantities for the measurement model.

# 3.5 Evaluating Combined Standard Uncertainty

## 3.5.1 Monte Carlo Methods

## 3.5.2 Law of Propagation of Uncertainty

## 3.5.3 Finite Difference Methods

# 3.6 Expanded Uncertainty and Confidence Intervals

# 3.7 Conclusions

# Bibliography

[1] L.T. Stant, P.H. Aaen, N.M. Ridler, Comparing methods for evaluating measurement uncertainty given in the JCGM ‘Evaluation of Measurement Data’ documents, Measurement, Volume 94, 2016, Pages 847-851, ISSN 0263-2241, <https://doi.org/10.1016/j.measurement.2016.08.015>.

[2] UKAS, Website, Accessed March 2018, <http://www.ukas.com/about/our-role/>

[3] BIPM, Draft of the ninth SI brochure, 10 November 2016, pp. 2-9. Retrieved March 2018, <http://www.bipm.org/utils/common/pdf/si-brochure-draft-2016b.pdf>

[4] International Organization for Standardization 1993 Guide to the Expression of Uncertainty in Measurement 1st ed (Geneva) ISBN 92-67-10188-9

[5] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP and OIML 2008 Guide to the Expression of Uncertainty in Measurement JCGM 100:2008 (GUM 1995 with minor corrections) URL [http://www.bipm.org/utils/common/documents/jcgm/JCGM 100 2008 E.pdf](http://www.bipm.org/utils/common/documents/jcgm/JCGM%20100%202008%20E.pdf)

[6] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP and OIML 2012 International Vocabulary of Metrology Basic and General Concepts and Associated Terms (VIM) JCGM 200:2012 URL [http://www.bipm.org/utils/common/documents/jcgm/JCGM 200 2012.pdf](http://www.bipm.org/utils/common/documents/jcgm/JCGM%20200%202012.pdf)

[7] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP and OIML 2008 Supplement 1 to the “Guide to the Expression of Uncertainty in Measurement" – Propagation of distributions using a Monte Carlo method JCGM 101:2008 URL [http://www.bipm.org/utils/common/documents/jcgm/JCGM 101 2008 E.pdf](http://www.bipm.org/utils/common/documents/jcgm/JCGM%20101%202008%20E.pdf)

[8] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP and OIML 2011 Supplement 2 to the “Guide to the Expression of Uncertainty in Measurement" - Extension to any number of output quantities JCGM 102:2011 URL [http://www.bipm.org/utils/common/documents/jcgm/JCGM 102 2011 E.pdf](http://www.bipm.org/utils/common/documents/jcgm/JCGM%20102%202011%20E.pdf)

[9] N. Metropolis, S. Ulam, The Monte Carlo method, J. Am. Stat. Assoc., 44 (1949), pp. 335-341

[10] J. Neyman, Outline of a theory of statistical estimation based on the classical theory of probability, Philos. Trans. R. Soc. A, 236 (767) (1937), pp. 333-380

[11] A. Gelman, J. Carlin, H. Stern, D. Rubin, Bayesian Data Analysis, Chapman & Hall, London (2003)

[12] D. White, In pursuit of a fit-for-purpose uncertainty guide, Metrologia 53 (4) (2016) S107–24

[13] Kacker R and Jones A 2003 Metrologia 40 235—48

[14] Kacker R 2006 Metrologia 43 1—11

[15] Kacker R, Toman B and Huang D 2006 Metrologia 43 S167—77

[16] Bich W 2014 Metrologia 51 S155—58

[17] Elster C, Wöger W and Cox M G 2007 Metrologia 44 L31—32

# List of Figures

1. <https://commons.wikimedia.org/wiki/File:Louvres-antiquites-egyptiennes-img_2748.jpg>
2. Own work
3. Own work [1]
4. Own work [1]