Lawrence,

I spoke with Nick about the formatting and process that you are using going from word → Latex.

There is no value in taking this approach and it will take you more time this way to complete.

Going forward we only want the final pdf version in Latex. Nick will make his edits on paper and scan and send, or meet with you in person

For all inputs provided please annotate the pdf files that I send (or print a copy) and check off each and every comment I make.

Sure, latex from now on.

Measurement Uncertainty

3.1 Introduction

~~Foreword: Parts of this chapter are based on excerpts from the authors paper [1].~~ *add in a reference within the text Done.*

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 *(?? imagine?)* 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).

→ select one – you have repeatedly noted "physical effect or quantity" to many time. ✓

A solution widely thought to have been first established in the 3rd or 4th Millennium BC (see Figure 3.1), 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.
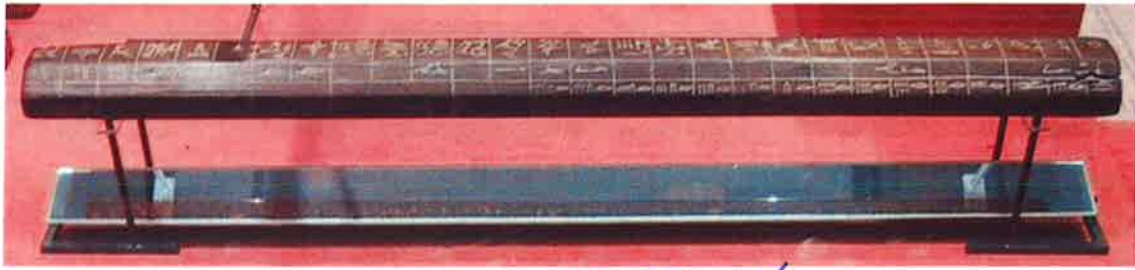
Figure 3.1: Egyptian royal cubit rod of Maya (treasurer of King Tutankhamun) 1336 – 1327 BC. The cubit is thought to have been first used in the 3rd or 4th Millennium BC.

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 Figure 3.2) and it is a fundamental tenet of metrology. Measurements with a shorter traceability chain are considered more traceable than those with longer chains.
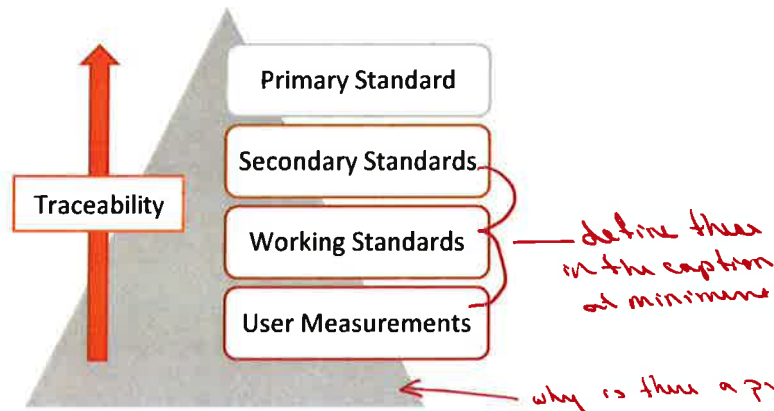

Figure 3.2: The traceability chain.

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.

define this and the need for it. ✓

because.... ✓

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.

primary ≠ important. Don't want to confuse with primary!

Working standards are used by ~~manufacturers~~ *for example,* in ~~different ways~~. Instrumentation manufacturers *who* may use the standards to ~~directly~~ calibrate their products before ~~sending~~ *shipping* it to the customer, ~~while~~ *and* more generally the standards can be used to calibrate test equipment ~~which~~ *that to* 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~~ *that* 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~~ *is scheduled for* into effect by May 2019.

Ω Ohms vs (ohms) please check No. Variable not unit. So "R".

confusing ✓ *the base*

define ✓ *(e.g...)*

Try not to ask questions → just write out the statement you wish to say. ✓

re-write

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.

Rewritten.

It is impossible to know the true value of a quantity being measured. ~~Many~~ *as* Many physical effects are involved during the measurement process, of which only a few are useful. All other processes

→ this is not clear –

→ why "other"?

Rewritten

imperfect calibration

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) *relevance to EE measurement?* 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).

No, "uncertain" because I am trying to show the definition of the word "uncertainty".

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 *unknown(?)*. If it is possible to quantify the amount of uncertainty in a measurement, then a degree of confidence can be formed about its value.

rephrase so you are not responding to your own question

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.

Please define: error, accuracy, uncertainty. already done why.

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. *some lines different* 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 *which fields?*. To address this, a landmark document was published in *given examples* 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

I have rewritten this.
rewritten.

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.

This chapter describes the evaluation of uncertainty prescribed in the GUM and highlights an inconsistency in the current version of the GUM and associated documents (which can have a profound effect on electromagnetic measurements). This work was published by the author in 2016 in Measurement (Elsevier) [1].

3.2 The Measurement Process

In contrast to basic evaluations of uncertainty, where only repeat measurements of the quantity of interest are analysed, 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 with Monte Carlo methods.

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. ✓

[handwritten, top right: for clarity: ~~the~~ the f(.) is the measurement itself. This is not a curve fit to data as a modelling engineer will do. — Also how is the modelling done. Is it just tabulated data. Cross reference with creation of a definition of a model.]

[handwritten, right margin: Added as footnote. Using GUM/VIM vocab.]

3.2.1 Modelling the Measurement

[handwritten, left margin: Added explicit reference to VIM to clarify]

We can define a set of measurands *Y* as a functional relationship depending on *N* ~~other~~ input quantities $X_1, X_2, ..., X_N$:

[handwritten: f(.) ↑ why bold?]

$$Y = f(X_1, X_2, ..., X_N) \tag{3.1}$$

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

[handwritten: where ȳ is ... and $\bar{x}_{1...N}$ is ...]

$$\bar{y} = f(\bar{x}_1, \bar{x}_2, ..., \bar{x}_N) \tag{3.2}$$

Each input quantity ~~could either~~ can be ~~obtained~~ observed during ~~this~~ the present measurement, ~~from a~~ a result from previous *[handwritten: why not the present measurement!?]* measurement, or from another source of information such as a datasheet or specification. An

[handwritten: meaning?]

example of a measurement model could be for a temperature measurement, where the ~~functional relationship~~ *[handwritten: world, input quanti.]* 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.

3.2.2 Input Quantities

[handwritten: use same title as your numbered list.]

[handwritten, left margin: define Sources, of]

~~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, *[handwritten: need to define / example... provided by datasheet.]*

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 *[handwritten: Added sentence.]* 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.

3.2.2.1 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), $\bar{x}$, and standard deviation, s, of a randomly varying input quantity, X, is written as

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i , \qquad s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad (3.3)$$

respectively, where x_i is the result of the ith observation. Importantly, a minimum of two observations must be made (n = 2) in order for $\bar{x}$ and 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:

$$u(\bar{x})_{GUM} = \frac{s}{\sqrt{n}} \qquad (3.4)$$

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, $u(x)_{SUPP}$:

$$u(\bar{x})_{SUPP} = \frac{s}{\sqrt{n}} \times \sqrt{\frac{\nu}{\nu-2}} \tag{3.5}$$

where $\nu = n-N$, 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:

$$u(\bar{x})_{SUPP} = \frac{s}{\sqrt{n}} \times \sqrt{\frac{n-1}{n-3}} \tag{3.6}$$

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 3.3 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 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.
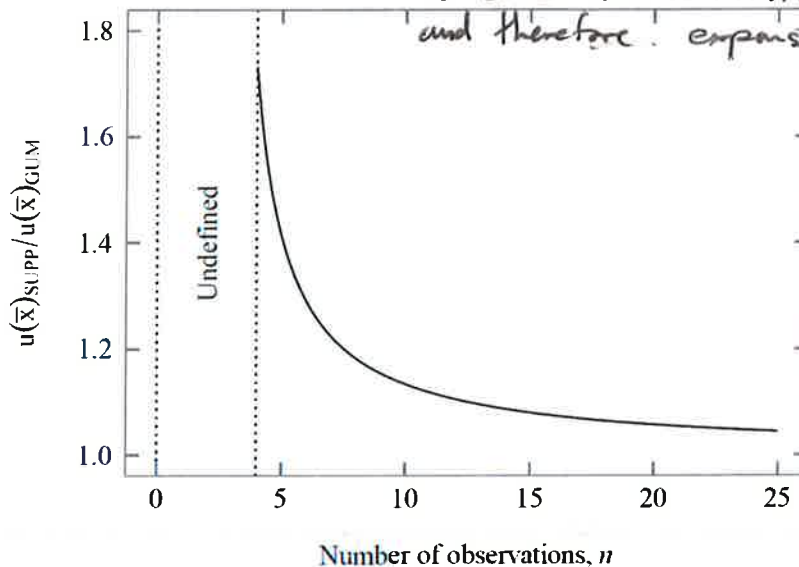

Figure 3.3: Scaling factor to convert from a GUM standard uncertainty to a GUM Supplement.

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]):

$$U_X = \frac{v}{v-2}\frac{S}{n} = \frac{1}{n(n-N-2)}\sum_{i=1}^{n}(x_i - \overline{x})(x_i - \overline{x})^{\mathsf{T}} \tag{3.7}$$

$$S = \frac{1}{v}\sum_{i=1}^{n}(x_i - \overline{x})(x_i - \overline{x})^{\mathsf{T}} \tag{3.8}$$

$$U_X = \begin{bmatrix} u(x_1)^2 & u(x_1,x_2) & \cdots & u(x_1,x_n) \\ u(x_2,x_1) & u(x_2)^2 & \cdots & u(x_2,x_n) \\ \vdots & \vdots & \ddots & \vdots \\ u(x_n,x_1) & u(x_n,x_2) & \cdots & u(x_n)^2 \end{bmatrix} \tag{3.9}$$

where U_X is the uncertainty matrix, x_i is a sample from the array of vectors containing input quantity indications and $\overline{x}$ 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 shortly.

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:

$$R = \frac{V}{I}\cos\theta \ , \quad X = \frac{V}{I}\sin\theta \ , \quad Z = \frac{V}{I} \qquad (3.10)$$

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 the previous section. 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 3.1. 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 $N = 3$, $\sqrt{v/(v-2)} = \sqrt{(n-N)/(n-N-2)} = \sqrt{3}$. 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.

Table 3.1: 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).

Value	V/V	I/A	Φ/rad
x_1	5.007	19.663	1.0456
x_2	4.994	19.639	1.0438
x_3	5.005	19.640	1.0468
x_4	4.990	19.685	1.0428
x_5	4.999	19.678	1.0433
x_6	4.999	19.661	1.0445
$\bar{x}$	4.9990	19.6610	1.04446
$u(\bar{x})_{GUM}$	0.0026	0.0077	0.00061
$u(\bar{x})_{SUPP}$	0.0045	0.0134	0.0011
$\frac{u(\bar{x})_{SUPP}}{u(\bar{x})_{GUM}}$	1.732	1.732	1.732

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 3.2 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 3.2: 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).

Method	$u(R)/\Omega$	$u(X)/\Omega$	$u(Z)/\Omega$
GUM	0.058	0.241	0.193
GUM-S2	0.130	0.540	0.431
$\frac{GUM-S2}{GUM}$	2.241	2.241	2.233

[handwritten: improve resolution ✓]

[handwritten: Section # 7. ✗ Too deep.]

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

[handwritten: ✗ no – this is the GUM example title as per the previous section.]

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.

[handwritten: ✗ I like this ——— part # ?]

Because each S-parameter is a complex-valued quantity ~~(S = (S_Re, S_Im))~~, there are $2m^2$ input quantities

[handwritten: for ✓]   *[handwritten: of an m port device.]*

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 3.3 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 3.3: 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.

Ports, m	Input quantities, N	Required minimum number of repeat observations, n, for $u(\bar{x})_{SUPP}$ to be defined.	$\frac{u(\bar{x})_{SUPP}}{u(\bar{x})_{GUM}}$
1	2	5	1.732
2	8	11	1.732
3	18	21	1.732
4	32	35	1.732
⋮	⋮	⋮	⋮
8	128	131	1.732

[margin annotation: improve resolution ✓]

[margin annotation: no brackets. ea.† "Equation" only appears at sentence start as per IEEE.]

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 3.4 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.

[margin annotation: often the case ✓]
[margin annotation: capture these numbers ✗]
[margin annotation: engineer performing the measurements]
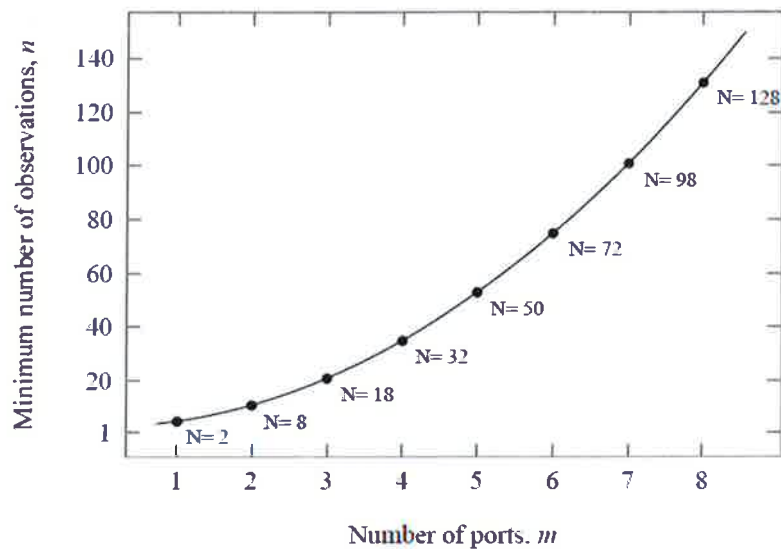*[margin annotation: No, I like these phrases.]*

Figure 3.4: 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.~~

please clarify as the caption should be readable by itself ✓

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.~~ *It is not straight-forward to select the approach because...* 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. *(what about time savings?) Mentioned here*

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.

3.2.2.2 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 could be [ ]:

- 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;

Try to avoid bulleted lists just write them in the text

now explained earlier

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. Category B uncertainty components are not restricted to Gaussian or t-distributions, and could for example be normal (rectangular), beta, or Cauchy distributions. Unless the combined standard uncertainty is determined via a Monte Carlo method, as explained in the following section, the standard uncertainty must be known for the value to be used as an input quantity.

3.2.3 Evaluating Combined Standard Uncertainty

In order to determine the standard uncertainty of the measurand – the combined standard uncertainty – the uncertainties of the input quantities must be propagated through the measurement model. The GUM offers several methods to achieve this, which will be described in this section.

3.2.3.1 Monte Carlo Methods

the? ✗ There are lots of techniques! e.g. MCMC...

Supplement 1 of the GUM [7] covers the use of a Monte Carlo technique to determine combined standard uncertainty in the measurand. The Monte Carlo technique has three important benefits for the propagation of uncertainty:

1. The measurement model does not need to be known explicitly. In some cases, the algorithm used to obtain a measurement result is proprietary and cannot be made available to the metrologist. Alternatively, the measurement model may be very complicated or involve numerical solutions which cannot be differentiated as required by other propagation methods.

2. Full knowledge of the probability distributions of the input quantities are used and preserved through the uncertainty propagation. Because the input quantity distributions are sampled directly, the complete probability distribution of the measurand can be obtained (see Figure 3.5). This can be very useful when more exotic distributions such as u-shaped distributions are used for input quantities, or if the measurement model is strongly nonlinear, when one cannot make assumptions about the probability distribution of the measurand.

3. The uncertainty propagation preserves nonlinearities in the measurement model. Alternative propagation methods presented in the GUM cause the measurement model to be linearised around the estimate. In most cases where a nonlinear measurement model is used, however, the uncertainty values are sufficient small that a linear approximation is valid [5, 5.1.5]. Often, an initial Monte Carlo propagation is used to validate this assumption.

4. All correlations between input quantities are preserved. For many measurements involving multiple input quantities (especially in electromagnetic measurements), the uncertainties of one or more input quantities may be correlated. This means that when the value of one quantity changes, it affects the values of others. The effect can both increase or decrease the combined standard uncertainty in the measurand significantly. Chapter 4 will discuss the impact of correlations on VNA measurements.

The primary disadvantage of Monte Carlo methods is the time required to process them. For an accurate evaluation of uncertainty, the number of samples must be sufficiently large. Generally, the GUM recommends 10^6 samples for a 95% coverage interval accurate to one or two significant digits [7, 7.2.1]. The number of samples increases with the size of the desired coverage interval of similar accuracy. For many measurements today, the processing power of modern computers is sufficient for the duration of uncertainty propagations using Monte Carlo methods to be acceptable. However, in situations where the measurement model is very time-consuming to process, or where the uncertainty evaluation must be very fast, linear propagation techniques may be preferred.

A detailed explanation of the steps involved in performing a Monte Carlo propagation can be found in Section 7 of [7].
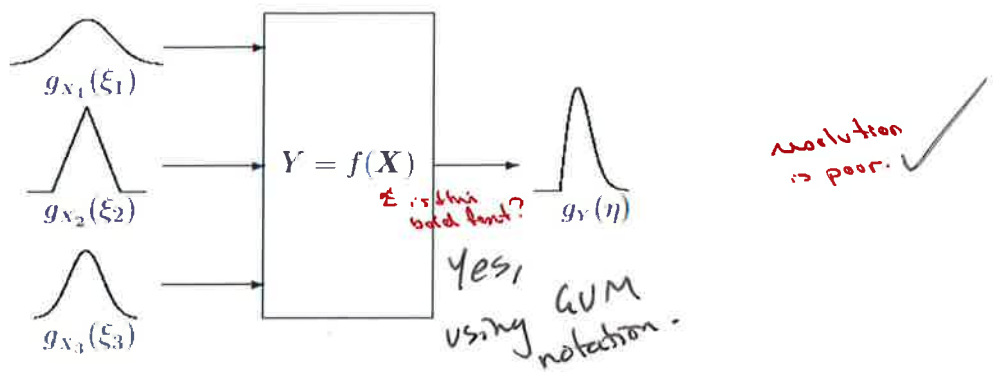
Figure 3.5: An illustration of the propagation of distributions from 3 input quantities g_{x1}, g_{x2}, g_{x3}, through the measurement model, Y, to the measurand, g_Y [7].

3.2.3.2 Law of Propagation of Uncertainty

The primary propagation method presented in the GUM is the Law of Propagation of Uncertainty (LPU). This method uses first-order derivatives of the measurement model, together with the variances (and co-variances) of the input quantities, to determine a value for the combined standard uncertainty. The use of first-order derivates specifies a linearised measurement model, which in many applications is a valid assumption as discussed previously. The LPU provides different equations for combining independent (uncorrelated) and correlated input quantities. For independent input quantities,

$$u_c^2(\bar{y}) = \sum_{i=1}^{N} \left(\frac{\partial f}{\partial x_i}\right)^2 u^2(\bar{x}_i), \tag{3.11}$$

where $u_c^2(\bar{y})$, the combined variance, is the square of the combined standard uncertainty, f is the function describing the measurement model, and x_i is an input quantity with variance $u^2(\bar{x}_i)$. For correlated input quantities,

$$u_c^2(\bar{y}) = \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} u(\bar{x}_i, \bar{x}_j). \tag{3.12}$$

Supplement 2 to the GUM offers a matrix formulation of (3.12) [8, 6.2.1.3], which handles the multivariate case where multiple measurands are encountered. If the covariance matrix of dimension $N \times N$ associated with $\bar{x}$ is

$$U_{\bar{x}} = \begin{bmatrix} u(\bar{x}_1, \bar{x}_1) & \cdots & u(\bar{x}_1, \bar{x}_N) \\ \vdots & \ddots & \vdots \\ u(\bar{x}_N, \bar{x}_1) & \cdots & u(\bar{x}_N, \bar{x}_N) \end{bmatrix}, \tag{3.13}$$

the covariance matrix of dimension $m \times m$ associated with $\bar{y}$ is

$$U_{\bar{y}} = \begin{bmatrix} u(\bar{y}_1, \bar{y}_1) & \cdots & u(\bar{y}_1, \bar{y}_m) \\ \vdots & \ddots & \vdots \\ u(\bar{y}_m, \bar{y}_1) & \cdots & u(\bar{y}_m, \bar{y}_m) \end{bmatrix}, \tag{3.13}$$

and the sensitivity matrix $C_{\bar{x}}$ of dimension $m \times N$ containing the first-order partial derivatives of the measurement model to each input quantity (the Jacobian of the measurement model) is given by evaluating

$$C_X = \begin{bmatrix} \dfrac{\partial f_1}{\partial X_1} & \cdots & \dfrac{\partial f_1}{\partial X_N} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial f_m}{\partial X_1} & \cdots & \dfrac{\partial f_m}{\partial X_N} \end{bmatrix} \qquad \text{\textcolor{red}{✓}} \tag{3.14}$$

at $X = \bar{x}$, then $U_{\bar{y}}$ is given by

$$U_{\bar{y}} = C_{\bar{x}} U_{\bar{x}} C_{\bar{x}}^{\mathsf{T}}. \tag{3.15}$$

The LPU does not provide any information about the shape of the probability distribution of $U_{\bar{y}}$ or its components. The results of the measurement are obtained from the estimates and the combined standard uncertainties – the positive square roots of the diagonal terms of the covariance matrix.

3.2.3.3 Finite Difference Methods

Included in the definition of the LPU is an alternative method of determining the sensitivity coefficients in (3.14), without the need to know the measurement model f, explicitly. This technique can be described as a finite-difference method and involves measuring the change in Y while varying a particular X_i and holding all other input quantities constant. This is often used ~~in industry~~ where there may not be a model available for a particular process, but a rudimentary uncertainty analysis is required. Typically, the sum of the estimate and the standard uncertainty of each input quantity $\bar{x}_i + u(\bar{x}_i)$ is used, although a more rigorous version also includes the standard uncertainty subtracted from the estimate $\bar{x}_i - u(\bar{x}_i)$ to check for asymmetry. Because only two points are used to solve for each sensitivity coefficient (the estimate and the estimate plus standard uncertainty), this uncertainty propagation also linearises the measurement model.

If all of the input quantities are considered independent and the standard uncertainty was chosen as the value with which to perturb the input quantities, then by subtracting the estimate of the measurand from each sample and adding the results in quadrature, the combined standard uncertainty in the measurand can be obtained ~~easily~~:

$$u_c(\bar{y}) = \sqrt{\sum_{i=1}^{N}((\bar{x}_i + u(\bar{x}_i)) - \bar{x}_i)} \#(3.16)$$

3.2.4 Expanded Uncertainty and Coverage Intervals

Although it is recommended to express a result with combined standard uncertainty $u_c(\bar{y})$, it is often required, especially in safety critical applications, for the uncertainty to encompass a larger fraction of the distribution of values that could reasonably be attributed to the measurand. An "expanded uncertainty" U is instead used and is related to the combined uncertainty by $U = ku_c(\bar{y})$ [5, 6.2.1]. The multiplying factor k is termed the "coverage factor" and is typically in the range 2 to 3, often either of those two integer values. Using expanded uncertainties, the result can be expressed as $\bar{y} \pm U$, which is a popular format for datasheets and specifications.

To obtain a coverage factor that provides a given level of confidence (e.g. 95%) is not straightforward, and depends on the probability distribution of the measurand. If all input quantities are Category A uncertainty components and the measurement model is linear, then the measurand distribution can be assumed to be Gaussian. In this case, the coverage interval is known as a confidence interval and can be given as a percentage by $\mathrm{erf}(z/\sqrt{2}) \times 100$ where $\mathrm{erf}(x)$ is the Gauss error function of x.

In situations where the above conditions cannot be met, a level of confidence can be obtained by calculating the effective degrees of freedom ν_{eff} of the distribution of the measurand. This process is explained in Annex G of [5]. For Monte Carlo propagations with sufficient samples, the confidence interval can be found by analysing the distribution of the measurand and obtaining the deviation from the estimated value which encompasses the desired percentage of samples (e.g. 95%).

3.3 Sensitivity Analysis

A benefit of propagating uncertainties through the measurement model is that an analysis of the sensitivity of the measurands to each input quantity can be performed. The sensitivity coefficients obtained from the measurement model can either be compared directly or multiplied by the standard uncertainty of the respective input quantity, in order to obtain an uncertainty figure for the measurand which can be compared with those calculated for other input quantities. This method is similar to the finite difference propagation technique described in 3.5.3, which can also be used to perform a sensitivity analysis. Because the input quantities are perturbed from their estimate sequentially (while all others are held at their estimate), this form of sensitivity analysis is termed "sequential perturbation".

The results of the sensitivity analysis can be very useful to the metrologist. Not only can the relative impact of different input quantity uncertainties be reviewed, but also complicated behaviour in the combined standard uncertainty may be better understood; Figure 3.6 shows an example of this feature. Sensitivity analyses are also an efficient approach to improving combined standard uncertainty. Once input quantities with dominant contributions have been identified they can be targeted for improvement – or in some cases an alternative measurement model can be used which avoids them.

[handwritten: reducing ?]
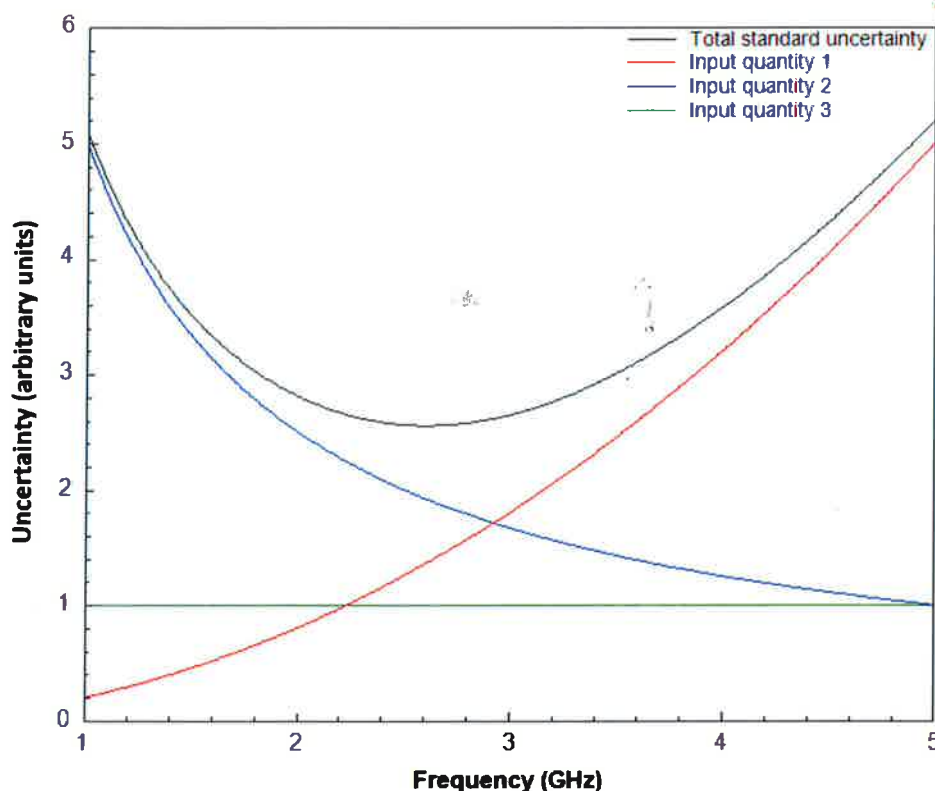
*[handwritten: * some more description of what is happening in this figure would be helpful]*

Figure 3.6: An example of results from a sensitivity analysis which reveal the origins of the complicated behaviour of the combined standard uncertainty with respect to a variable (in this case frequency).

3.4 Conclusions

[handwritten: presented]

This Chapter has explained how measurements underpin modern life, supporting trade and commerce and facilitating new discoveries in science and engineering. Through traceability and the unit system, the evaluation and management of uncertainty in measurements provides a backbone of confidence and trust that supports this infrastructure. In an attempt to standardise the definition and representation of measurement uncertainties, an internationally-used guidance document, the GUM, offers rigorous methods to evaluate them. The methods presented in the GUM have been used throughout the work presented in this thesis. However, the GUM continues to be developed, and recently an inconsistency was created in the evaluation of Category A uncertainty components,

[handwritten: generates]

[handwritten: these uncertainties]

[handwritten: and this has a ... impact for microwave metrology]

[handwritten: Next line says this.]

This chapter has reviewed the inconsistency from the objective of electromagnetic measurements, an area of metrology where the effects are shown to be potentially significant, ~~because~~ *as shown in this chapter!* ~~have been~~ Three methods for propagating uncertainty through a measurement model to determine the combined standard uncertainty of the measurands were described. Although the Monte Carlo method preserves the most information about both the uncertainties of the input quantities and the measurement model, the higher computational effort can be prohibitive in some cases. Instead, the LPU provides two linear alternatives, which are often much more efficient but require validation to ensure that the measurement model can be treated as linear.

The idea of expanded uncertainty and coverage intervals was introduced, these being met frequently in Category B uncertainty components defined from datasheets and specifications. ~~It was stated that~~ a confidence interval can only *be* easily calculated from either Category A uncertainty components and a linear measurement model, or if a Monte Carlo propagation is used and the probability distribution of the measurand analysed. In other cases, a coverage interval can be calculated using knowledge of the input quantities and guidance from the GUM.

fix sentence structure

Finally, this Chapter described sensitivity analysis, which can be carried out using results from the LPU procedure. The framework presented in this thesis utilises a sensitivity analysis to allow the user to examine and attempt to minimise significant sources of uncertainty, which is especially important in sensitive electromagnetic measurements such as those made on-wafer.

Bibliography ~~References~~

[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

[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

[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

[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

[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

(Handwritten annotations: "Lined up?", "— Use IEEE style please consistently. Many need to be fixed")

[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

Information for list of figures for this chapter

1. https://commons.wikimedia.org/wiki/File:Louvres-antiquites-egyptiennes-img_2748.jpg
2. Own work
3. Own work [1]
4. Own work [1]
5. Gum Supplement 1 [7]
6. Own work using NIST Microwave Uncertainty Framework