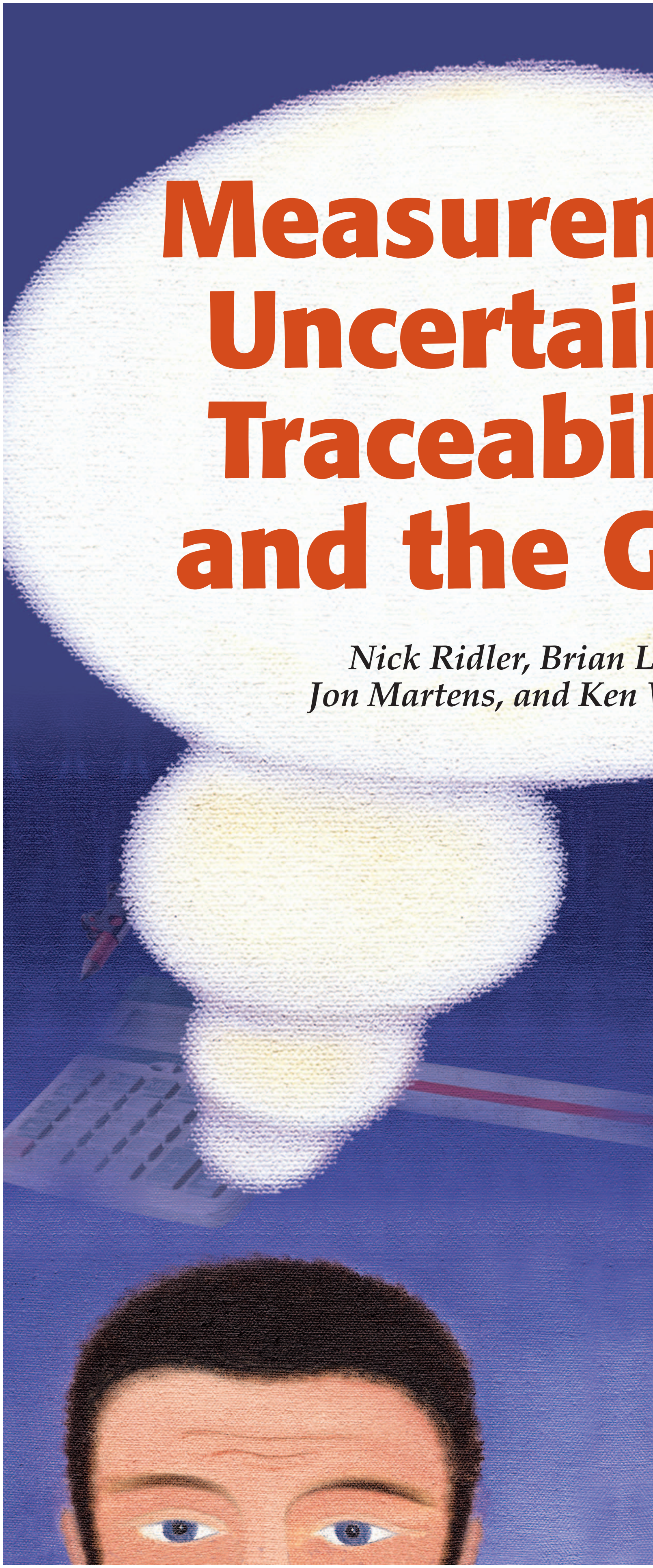
Measurement Uncertainty, Traceability, and the GUM

Nick Ridler, Brian Lee, Jon Martens, and Ken Wong



"I know what it is, because it's been measured accurately." That's a phrase that's often heard in many different walks of life. But what do we mean by "accurate"? And how *accurate* is "accurate"? Indeed, one person's idea of accuracy can often be another person's idea of inaccuracy. Just consider taking a journey from the measurement environment associated with precision coaxial connectors on the front panel of measuring instruments such as vector network analyzers (VNAs) to that associated with an open-area electromagnetic compatibility (EMC) test facility. Such a journey can often result in a reduction in accuracy of several orders of magnitude along the way.

Recognizing that accuracy describes the closeness of agreement between a measured value and its notional "true" value brings into play the concept of "error"—this being the difference between the measured value and the underlying true value. Now, we all generally agree that achieving high accuracy is usually a good thing or, equivalently, having small errors is similarly a good thing (as long as this can be accomplished at the right price). These work fine as qualitative concepts

Nick Ridler (nick.ridler@npl.co.uk) is with National Physical Laboratory, Teddington, U.K. Brian Lee and Jon Martens are with Anritsu Company, Morgan Hills, CA. Ken Wong is with Agilent Technologies, Santa Rosa, CA.

Digital Object Identifier 10.1109/MMM.2007.900114

 1527-3342/07/$25.00©2007 IEEE

that we can use in our everyday lives, e.g., "my watch is very accurate" is a useful phrase as a quality indicator. However, when we want to move from these qualitative concepts to quantitative ones, the use of terms like "accuracy" and "error" starts to cause problems. Namely, to quantify accuracy and/or error, we first need to know the underlying true value of the thing we are trying to measure. Now, under most circumstances, we don't know what the true value is—if we did, why would we be bothering to measure it?!

Therefore, as we move from everyday walks of life, where qualitative information is sufficient for our needs, to situations where the information must be quantitative, the qualitative concepts of accuracy and error become less useful. Such a transition usually occurs when we move from these everyday situations to that which we encounter when we enter our place of work to embark on scientific and engineering investigations (where reliable quantitative knowledge becomes a requirement).

It is as we make this transition that the concept of "uncertainty of measurement" comes to assist us by bringing us into a framework of quantitative measures of reliability. We will, therefore, spend the rest of this article concentrating on the concept of uncertainty in measurements so that we can establish methods for obtaining a quantitative understanding of the reliability of our measurements, whether these measurements are made in the lab, on the shop floor, in the field, or anywhere else for that matter.

Feeling Uncertain About Uncertainty?

So, what do we mean by "uncertainty" in the context of a measurement result? Historically speaking, a phrase such as "my measurement uncertainty is . . ." was really only ever heard in the realms of calibration and measurement standards laboratories. Even there, the process of evaluating uncertainty was often considered a "black art," the skills for which were possessed by only a chosen few, such as laboratory managers and perhaps the occasional quality assurance auditor.

Indeed, under these circumstances, the very name "uncertainty" turned out to be aptly chosen, as people were generally *uncertain* as to what it meant, *uncertain* about what to do with it, and *uncertain* about how to go about finding it out. This perhaps coincidental choice of terminology (after all, the choice could have been made to talk about measurement *certainty* rather than the more pessimistically sounding *uncertainty*) has nevertheless stuck, and it is now a part of the language of those desiring a quantitative indication of the reliability of measurement results.

What Is Uncertainty?

According to international sources [1], the definition of uncertainty of measurement is as follows:

> Parameter, associated with the result of a measurement, that characterizes the dispersion of the values that could reasonably be attributed to the measurand.

This definition, albeit both rigorous and precise, is perhaps, to the uninitiated, about as clear as a foggy day! (Note that the definition also contains three footnotes to explain the term "parameter;" the concept of uncertainty components; and that a result is a best estimate of its value!) So, an attempt at a more digestible form of wording might be something along the lines of:

> Uncertainty expresses the doubt about the result of a measurement.

or

> Uncertainty is an interval that is likely to contain the "true value" of the quantity being measured.

Although this latter description has reintroduced the term "true value," the good news is that it turns out that we don't actually need to know the true value in order to determine the uncertainty! This means that we can now quantify the reliability of a measured value in terms of its uncertainty without needing to know the true value of the thing we are trying to measure. This is the goal we should be striving for, and this is why when we need to demonstrate the reliability of a measurement, we should look to a determination of the uncertainty of the measurement.

Before finishing the discussion of the meaning of the term "uncertainty," it is perhaps worthwhile to consider another related term that is often used to describe measurement reliability—that is, measurement repeatability. Repeatability is often used somewhat loosely to imply high measurement accuracy or, alternatively, small measurement error. However, measurement repeatability only really indicates the closeness of agreement between successive results of measurements that are usually made under essentially the same conditions. As such, repeatability only provides an indication of the size of random errors that may be present in a measurement process but says nothing about any systematic errors that may also be present (such as biases, offsets, etc.). Repeatability is therefore often of little use when a measurement must be understood outside the context of its immediate determination (e.g., when others want to use or reproduce a measurement result in a different context).

Why Is Uncertainty Important?

If the uncertainty of a measurement has been determined, then we have a clear indication about the reliability of that measurement. As such, the measurement can then be compared with other values (measured or otherwise) in a meaningful way, enabling statements of equivalence or compliance to be made. To illustrate this, consider two measurements of the same physical quantity made by two separate measurement systems. This might be, for example, a measurement made by a supplier of a device followed by a measurement made by the customer for the device. Almost certainly, the

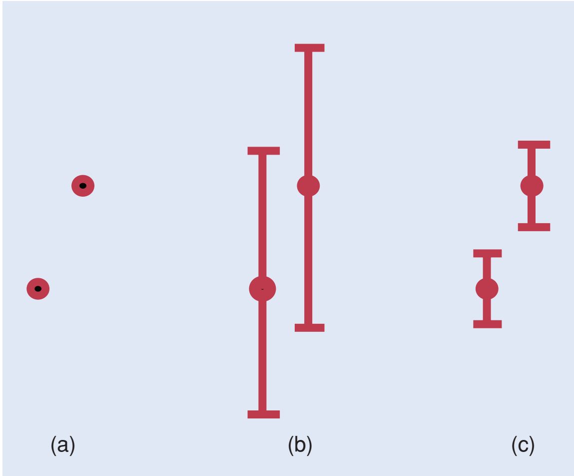


Figure 1. *Graph of two values: (a) without stated uncertainties, (b) with large uncertainties, and (c) with small uncertainties.*

two measurements will not agree exactly, and so there will be a difference between the two values. This is illustrated in Figure 1(a), where the vertical axis indicates the size of the parameter being measured (using arbitrary units and scale sensitivity). Since the dots representing the measurements are not exactly in line, the figure shows a difference between the two values. The question then becomes "Is this difference significant?" In other words, should there be concern about the difference, for whatever reason, or not?

Now, if we assume that the uncertainty of both the measurements has been evaluated, we can then reexamine these measurements with "uncertainty bars" attached. One possible scenario is presented in Figure 1(b), which shows that the uncertainty bars from each measurement overlap. In other words, although the values for the measurement results do not agree exactly, the uncertainty bars for the results accommodate the amount of disagreement. This means that the results are actually showing agreement to within these stated uncertainties.

Another possible scenario for the same two measurements is shown in Figure 1(c). On this occasion, the uncertainty bars for these two measurements do not overlap. This shows that the amount of disagreement between the measurements is greater than the uncertainty bars and, thus, these measurements are not in agreement. The results shown in Figure 1 demonstrate the importance of having uncertainty information about measured results since it is only really possible to demonstrate measurement equivalence (i.e., agreement or otherwise) after the uncertainty of measurement has been established.

Another use of measurement uncertainty is demonstrated in Figure 2, which shows a series of measurements made on a product with respect to a specification limit. It is clear that for result a), the value, along with its uncertainty, is within the specification limit; i.e., the product meets the specification and is therefore accepted. Similarly, it is clear that result d), along with its uncertainty, is outside the specification limit; i.e., the product fails to meet the specification and is therefore rejected.

However, for results b) and c), the accompanying uncertainty interval straddles the specification limit line and, therefore, demonstrating compliance to this specification is more difficult. For case b), it is tempting to say that the result is within specification. But if this were a safety-critical application, would we feel comfortable accepting this product (recognizing that the uncertainty interval suggests that the underlying true value might actually be outside the specification limit)? In a similar manner, it is tempting to reject case c) as failing to meet the specification. However, there may be instances where the true value contained within the uncertainty interval is inside the specification limit, so that one would be rejecting a product that should actually pass the test.

In general, for result b), it is not possible to state compliance (unless a lower level of confidence is acceptable, thus reducing the size of the uncertainty interval). Similarly, for result c), it is not possible to state noncompliance (again, unless a lower level of confidence is acceptable for the uncertainty interval) [2].

How to Evaluate Uncertainty

These days, there are clear guidelines on how to evaluate uncertainty for just about all types of measurements. These can be found in the *Guide to the Expression of Uncertainty in Measurement* [3] (commonly known as the GUM, as used in the title of this article), which was first published in 1993 by the International Organization for Standardization (ISO).

In the introduction to the GUM, some clear goals are set out, including:

> …a worldwide consensus on the evaluation and expression of uncertainty in measurement would permit the significance of a vast spectrum of measurement results in science, engineering,

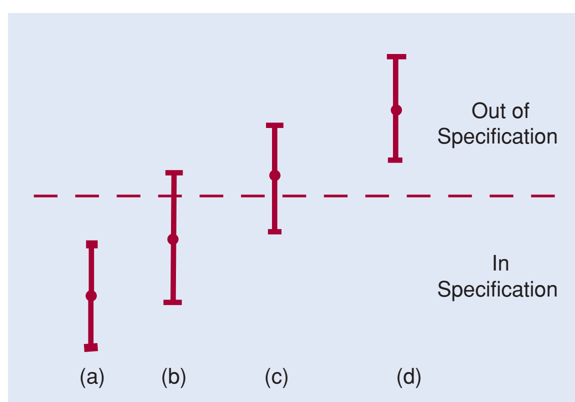


Figure 2. *In or out of specification? This graph shows four compliance scenarios.*

commerce, industry and regulation to be readily understood and properly interpreted.

…it is imperative that the method for evaluating and expressing uncertainty be uniform throughout the world so that measurements performed in different countries can be easily compared.

In general, the material presented in the GUM has met these and other goals. Any serious student or practitioner of the art and science of evaluating measurement uncertainty should undertake some study of the GUM. However, bearing in mind that the GUM runs to around 100 pages in length and contains some fairly advanced mathematics and statistics, some more digestible texts might be more appropriate for the beginner or casual observer in this field. Examples of such texts include [4] and [5].

The basic approach advocated in the GUM is that of describing a measurement using a measurement model in the form of a functional relationship between input and output quantities. The term "input quantities" is used to describe the things that are actually determined during a measurement process and the term "output quantity (or quantities)" is used to describe that which we want to determine. This approach recognizes that, in most cases, the thing we want to determine is not measured directly, but is determined from a series of other quantities.

Let's use an example to illustrate this. We want to determine the dc resistance of a resistor, and the two measuring instruments available to us are a voltmeter and an ammeter. Well, by connecting a battery to the resistor we can measure the voltage drop across the resistor using our voltmeter and the current flowing through the resistor using our ammeter. In this case, our (two) input quantities are voltage and current and our single output quantity is resistance. The functional relationship between the input quantities and the output quantity is simply Ohm's law,

$$R = f(V, I) = \frac{V}{I}, \tag{1}$$

where V and I are the measured voltage and current, respectively; R is the resistance; and f represents the functional relationship between the input quantities (V and I) and the output quantity, R.

Extending this approach to any number of input and output quantities gives the following "general" measurement model:

$$\boldsymbol{y} = f(\boldsymbol{x}), \tag{2}$$

where $\boldsymbol{x}$ is a vector representing the input quantities, $\boldsymbol{y}$ is a vector representing the output quantities, and f is the function describing their interrelationship.

Having established the measurement model, the next step is to actually do the measurement. In the above Ohm's law example, this involves getting values of voltage and current using the voltmeter and ammeter. We can then "calculate" a value for the resistance using the measurement model.

We then need to estimate the uncertainty in the input quantities (in our example, these are the voltage and current readings). This can involve a variety of approaches depending on the situation. For example, if the measurement is prone to significant random variation (e.g., if an interconnect mechanism causes a significant variation in the observed measurement result), then repeating the measurement a number of times gives an indication of this variation. Statistical measures of variability (e.g., the experimental standard deviation) can then be used to quantify this variation. Alternatively, a measurement may be prone to systematic errors (e.g., due to an offset, or bias, present in a measuring instrument) that are described in the instrument's specification. In this case, the instrument specification can sometimes be used to quantify these errors. In practice, the uncertainty is often caused by a combination of several different effects that all need to be taken into account.

The uncertainty in the final result is determined from the uncertainties in the input quantities. To do this, an experimenter will often undertake a "sensitivity analysis" of the measurement model to determine which parameters have the most significant effect on the uncertainty of a final measurement result. Such an analysis can involve using

- partial differentiation of the terms in the measurement model [3]
- practical experiments to observe the effects of variations in the terms in the model [6]
- computer-intensive methods using random numbers to simulate the variations in the terms in the model [7] (note that an example of this approach is the so-called Monte Carlo method where random numbers are used to simulate distributions for the measured quantities)
- experience gained from previous measurement situations.

A generalization of the approach using partial derivatives, following from (2), results in the following succinct matrix form for the expression of uncertainty [8]:

$$V_y = J_x V_x J_x^T,$$

where J_x is the Jacobian matrix containing the partial derivatives for the input quantities, V_x is the uncertainty matrix (or covariance matrix) describing the uncertainties in the input quantities, and V_y is the uncertainty matrix for the output quantities. The superscript symbol T is used to denote matrix transposition.

For example, in our above dc resistance example (1), the (1×2) Jacobian matrix is

$$J_x = \left(\frac{\partial f}{\partial V}, \frac{\partial f}{\partial I} \right),$$

the (2 × 2) uncertainty matrix describing the uncertainties in the input quantities is

$$V_x = \begin{pmatrix} u^2(V) & u(V, I) \\ u(I, V) & u^2(I) \end{pmatrix},$$

and so the (1 × 1) uncertainty matrix describing the output quantity is simply

$$V_y = u^2(R).$$

The symbols $u(.)$ in both V_x and V_y above represent the uncertainties in the quantities V, I, and R.

An advantage of this matrix formulation is that any correlation between the input quantities and/or the output quantities can be accommodated by the off-diagonal elements in the uncertainty matrices (e.g., the $u(V, I)$ and $u(I, V)$ terms in V_x, used in the above dc resistance example). The need to consider correlation between terms in a measurement model is particularly important when there may be some form of interdependence (either physical or otherwise) between such terms. This can often be the case for measurements made at microwave frequencies involving complex-valued quantities [9].

The above process results in an uncertainty interval that is effectively one standard deviation (i.e., corresponding to a level of confidence of approximately 68%). However, for most practical circumstances, it is desirable to enlarge the overall uncertainty interval so that it becomes very likely that it contains the notional true value. One often then speaks of a certain level of confidence (e.g., 95%) that the uncertainty interval contains this true value. To achieve a given level of confidence (sometimes called "coverage probability"), a multiplying factor (termed "coverage factor" and represented by the symbol k) is applied to the overall uncertainty interval. Therefore, the coverage factor is closely related to the level of confidence described by the uncertainty interval. For example, for many situations, a coverage factor of $k = 2$ provides an uncertainty interval with a level of confidence that approaches 95%.

Some Recent Developments

Since the publication of the GUM in 1993, work has continued on applying the GUM approach to various measurement situations. Work has also been undertaken to adapt the methods of the GUM to cover other areas that were not originally included within the scope of the GUM and circumstances where the assumptions made by the GUM do not strictly apply. One such area of particular relevance to microwave engineers is the evaluation of uncertainty for complex-valued quantities. This is of particular relevance to the microwave community since many of the quantities that are encountered in microwave engineering are

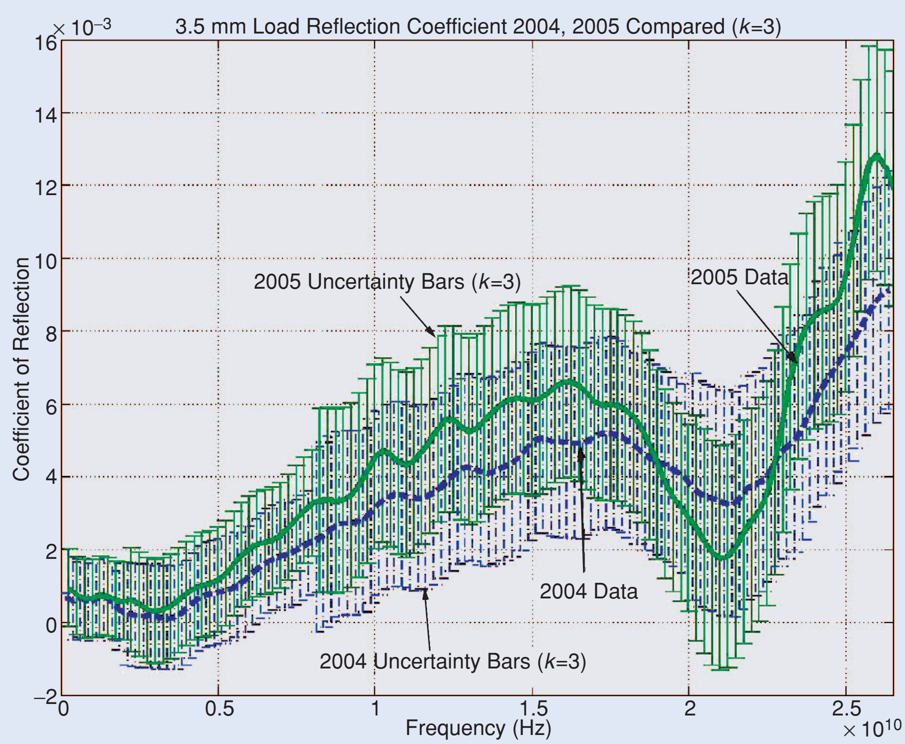


Figure 3. *Two sets of results with uncertainty bars representing the uncertainty, using a coverage factor $k = 3$.*

complex-valued, i.e., quantities having both magnitude and phase, such as *S*-parameters, etc. Several papers have now been published on this subject (see, for example, [9] and [10]).

In addition, several recent editions of *Metrologia* (www.iop.org/EJ/journal/met)—see, for example, the August 2006 and December 2006 issues—have contained many interesting articles on extending uncertainty practices beyond the scope originally covered by the GUM.

Finally, the various parties responsible for producing the GUM have now reformed a working group to put in place documents to supplement the information given in the GUM. Supplementary documents on the following topics are being planned [11]:

- a general introduction
- concepts and basic principles
- using a Monte Carlo method to propagate distributions
- using models with multiple output quantities
- modeling
- the role of measurement uncertainty in deciding conformance to specified requirements
- applications of the least-squares method.

Some Case Studies

We now present three case studies to illustrate some of the points covered in this article. The first case study demonstrates the importance of choosing an appropriate coverage factor when assessing equivalence between two sets of measurements. The second case study shows how a Monte Carlo method can be used to establish uncertainty estimates for complex-valued quantities, and the third case study shows how assumptions about probability density functions can have an effect on the size of the evaluated uncertainty of measurement.

Case Study 1—Choosing Coverage Factors

An expression of the uncertainty of a measurement should always give the coverage factor, k, chosen to achieve the stated uncertainty. Without the coverage factor, it is not possible to determine if two measurements of the same device are significantly different. The following plots illustrate this point.

The first plot (Figure 3) shows two measurements of the linear reflection coefficient as a function of frequency for a coaxial termination. The uncertainty bars represent the uncertainty using a coverage factor of $k = 3$. If the coverage factor is not stated, the differences between the two measurements appear to be acceptable.

Figure 4 shows the same measurements with the uncertainty bars representing the uncertainty using a coverage factor $k = 1$. If the coverage factor is not given, the measurements appear to show significant differences.

The reason for the apparent agreement in Figure 3 and disagreement in Figure 4 is that the uncertainty intervals are representing two very different levels of confidence, since the coverage factor directly relates to

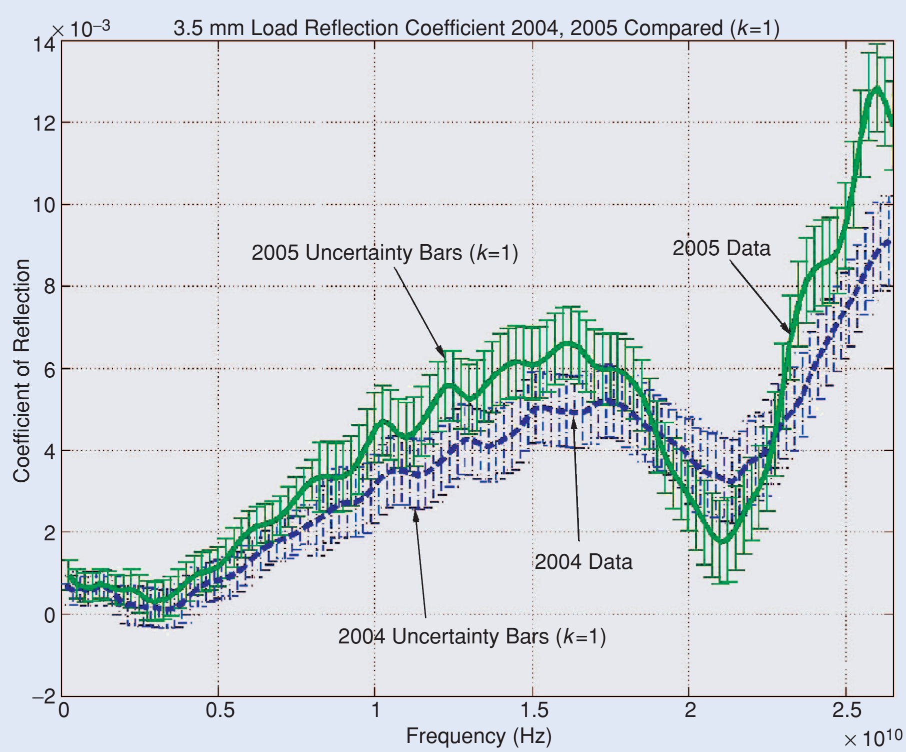


Figure 4. *Two sets of results with uncertainty bars representing the uncertainty using a coverage factor* $k = 1$.

the level of confidence. If a sufficiently large coverage factor is chosen, then any two sets of measurement results can be made to appear consistent (i.e., to show agreement). Similarly, if a sufficiently small coverage factor is used, any two sets of results can be made to appear to disagree. Therefore, the choice of coverage factor is very important, and it is equally important to state the chosen coverage factor used to report measurement results.

As a rule of thumb, a coverage factor of two (or thereabouts) will often provide a reasonably high level of confidence (approaching 95%) suitable for most applications, including the assessment of consistency between two sets of data. This is also the value of coverage factor that is usually used by calibration and test laboratories when reporting results to their customers.

Case Study 2—Complex-Valued Quantities and the Monte Carlo Method

A problem that affects most areas of microwave measurement is the presence of complex-valued measurement quantities (such as *S*-parameters and noise parameters) that have both a magnitude and phase component. The presence of such quantities can introduce further difficulty in what is often already a challenging area in which to operate.

For example, consider a simple one-port measurement on a calibrated VNA. There are many potential sources of error, including repeatability errors associated with the raw data for the measurement device (from connector repeatability, noise effects, etc.), repeatability errors associated with the calibration process, and systematic errors resulting from imprecise knowledge about the calibration standards. The complex-valued nature of terms in the measurement model also requires careful consideration. For example, each of the error terms used to model the VNA (e.g., directivity, test port match, etc.) will be a complex-valued quantity. In addition, the observed readings of the VNA for both measurement devices and calibration standards will also be complex-valued.

Many of these sources of uncertainty (e.g., the error terms) may also be correlated to some extent. In addition, the real and imaginary parts within a given error term may also be correlated (how this can happen will be explained shortly). Just adding up the magnitudes of the effects ignores all correlations and can overestimate the resulting uncertainty.

To show this effect, we will run through the case of a one-port VNA measurement with some simplifications. The error model uses typical values for the error coefficients for a commercial VNA operating at microwave frequencies. Connector repeatability and noise effects are assessed together as repeatability errors for both the measurement of the device and the calibration standards and are treated as uncorrelated. A typical distribution (assumed uncorrelated) for load reflection coefficients is also employed. Typical distributions for the open and short will also be used, but here correlation does play a role. Among a collection of opens and shorts, taking into account manufacturing and characterization tolerances, the magnitudes of the reflection coefficients are generally very tightly confined. There is somewhat more variability in the phase of the reflection coefficient (due to the effective offset length changing or the characterization of that offset length). This establishes a correlation between the real and imaginary parts (since the magnitude is roughly constant and the phase changing, something like a small part of a circle is traced out on the real-imaginary plane).

Another layer of correlation occurs within some of the error coefficients. While directivity is largely established by the load measurement, the open and short measurements jointly affect a pair of error terms (source match and reflection tracking). Thus, a deviation in, for example, the short measurement (or its characterization) affects multiple error terms and these contributions can sometimes partially cancel in the overall uncertainty.

A Monte Carlo method [12] is used to undertake this analysis, assuming normal distributions for all input parameters (for simplicity). On each pass of the process (over 1,000 passes are normally used), values are selected from the appropriate distributions for the calibration measurements, characterizations, and the device-under-test (DUT) measurement. The calibration is then applied to calculate a corrected DUT reflection coefficient for that pass. The output "distribution" of calculated values is then compared with the actual measurement result. The device in this example was chosen to have a return loss of approximately 14 dB.

First, we will suppose that the measurement repeatability terms are dominant [e.g., if a very wide intermediate frequency (IF) bandwidth is used in the VNA, or if the connectors on the devices are not particularly good]. Since the repeatability error terms are assumed uncorrelated, one would expect the real and imaginary parts of the resulting measurements to be uncorrelated. Figure 5 shows that this is indeed the case, since the error distribution is fairly symmetric in both the real and imaginary directions. It turns out, in this case, that a simple addition of magnitude effects would not lead to a noticeably different answer.

Now consider the case where repeatability terms no longer dominate (e.g., if a lower IF bandwidth is used or if high-quality connectors are used) and imprecise knowledge of the calibration standards becomes an important consideration. Because of the correlation between some of the error terms, one would expect there to be some correlation in the resulting measurements. This can be seen in Figure 6, where the variation in the imaginary component is considerably larger than the variation in the real component. In this particular case, the simple addition of magnitudes of effects would lead to an uncertainty estimate that is about a factor of two higher.

Figures 5 and 6 show the resultant effects of errors interacting during a measurement process involving complex-valued quantities. The use of a Monte Carlo method has enabled the complicated, and complex-valued, error structure in the measurement results to be determined. This information can be used to provide realistic uncertainty estimates for complex-valued measurement quantities [7].

Case Study 3—Mismatch Uncertainty in Power Measurement at Microwave Frequencies

One of the more daunting tasks encountered during the evaluation of uncertainty is the choice of the probability density function (pdf) used to characterize each uncertainty component in the measurement model—see the "Assigning Distributions to Uncertainty Components and Establishing Standard Uncertainty" sidebar. The availability of different measurement techniques and instrumentation often makes that choice even more difficult. Additional confusion can also arise due to the complex-valued nature of some measurement quantities.

An example where different choices of pdf are often made is in determining the uncertainty due to impedance mismatch during a microwave power measurement. This mismatch uncertainty is due to unwanted reflection (and its subsequent impact on transmission) of a signal in the transmission path of a measurement circuit.

Consider a measurement circuit comprising a generator with a power sensor and meter. The impedance mismatch factor, M, can be expressed as

$$M = |1 - \Gamma_L \Gamma_G|^2,$$

where Γ_G and Γ_L are the reflection coefficients of the generator and the sensor, respectively. The uncertainty in the mismatch, $U(M)$, depends on how Γ_L and Γ_G are determined. We will consider three cases for calculating $U(M)$, assuming different pdfs for Γ_L and Γ_G.

First, consider the situation where we only have scalar determinations of Γ_L and Γ_G. Under these conditions, the limits of the mismatch uncertainty are given by $\pm 2|\Gamma_L||\Gamma_G|$ (see, for example, [5]). Since the scalar determinations of the reflection coefficients do not contain any phase information, we have the following two scenarios:

Case a: Using maximum values for $|\Gamma_L|$ and $|\Gamma_G|$, denoted by $\max|\Gamma_L|$ and $\max|\Gamma_G|$, respectively, assuming both have a uniform distribution for the phase components (e.g., if the values of $|\Gamma_L|$ and $|\Gamma_G|$ are obtained from a product data sheet), gives:

$$U(M) = \pm \frac{2 \max |\Gamma_L| \max |\Gamma_G|}{d_1},$$

where the pdfs for $|\Gamma_L|$ and $|\Gamma_G|$ are assumed to be uniform/rectangular distributions.

Case b: Using constant values for $|\Gamma_L|$ and $|\Gamma_G|$, again assuming both have a uniform distribution for the phase components (e.g., from measurements of $|\Gamma_L|$ and $|\Gamma_G|$ made using a scalar network analyzer),

$$U(M) = \pm \frac{2\,|\Gamma_L|\,|\Gamma_G|}{d_2},$$

where the pdf for $U(M)$ is assumed to be a U-shaped distribution.

Values for the divisors d_1 and d_2 are chosen based on the assumed pdfs—see the "Assigning Distributions to Uncertainty Components and Establishing Standard Uncertainty" sidebar.

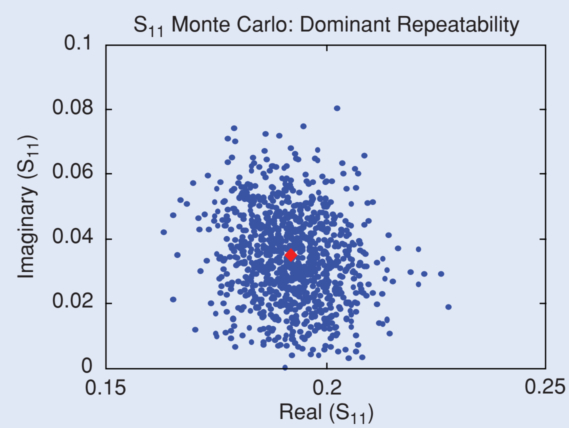


Figure 5. *A Monte Carlo estimate of uncertainty for the case where repeatability errors are dominant. The red diamond indicates the actual measured value.*

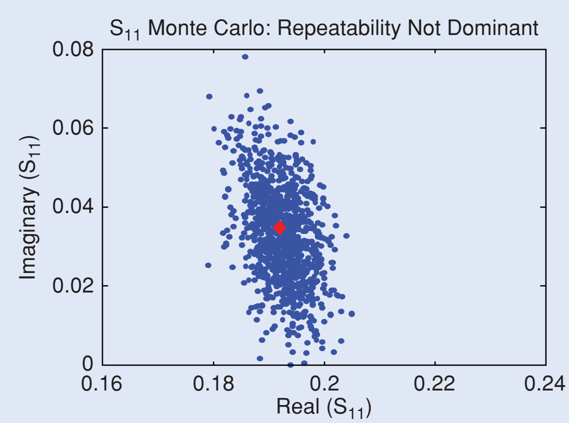


Figure 6. *A Monte Carlo estimate of uncertainty for the case where repeatability errors do not dominate. Again, the red diamond indicates the actual measured value.*

Assigning Distributions to Uncertainty Components and Establishing Standard Uncertainty

Before combining uncertainty components, an assumption about the type of pdf of each uncertainty component needs to be made. This is so that the size of each uncertainty component can be down-converted to the equivalent standard deviation for the assumed pdf. This then allows all the uncertainty components to be combined using the Law of Propagation of Uncertainty (LPU) [3]. The simplest realization of the LPU is the so-called root-sum-squares (RSS) approach where a number, n, of uncertainty components represented by standard deviations, u_i ($i = 1, \ldots, n$), are combined to give an overall uncertainty, u_c, also representing a standard deviation

$$u_c = \sqrt{u_1^2 + u_2^2 + \ldots u_n^2}\,.$$

The following pdfs are often used to specify uncertainty components in microwave measurements.

1) If the uncertainty component is characterizing a random error process (e.g., electrical noise, connector repeatability, etc.), then the pdf is often assumed to be a normal distribution (see Figure 7). Now, if the size of the uncertainty component has been established from observing a series of m repeated measurements, x_j ($j = 1, \ldots, m$), then the standard deviation in the mean, $\bar{x}$, of these repeated values is estimated as

$$s(\bar{x}) = \sqrt{\frac{\sum_{j=1}^{m}(x_j - \bar{x})^2}{m(m-1)}}.$$

On the other hand, if the uncertainty component is characterized by a normal distribution and is given at a specified level of confidence (e.g., as is often the case for results quoted on certificates by calibration laboratories), the uncertainty interval is divided by the coverage factor, k, (stated on the calibration certificate) to down-convert it to the equivalent standard deviation. For example, if the uncertainty component is specified at a level of confidence of 95%, the coverage factor is likely to be $k = 2$ or thereabouts.

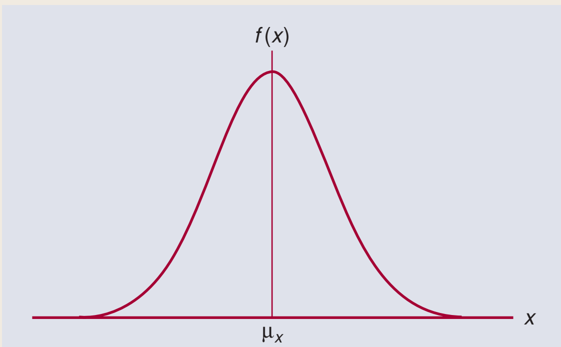


Figure 7. *A normal, or Gaussian, distribution often used to describe variations due to random error processes.*

If we have vector determinations of Γ_L and Γ_G (e.g., using a VNA), then the uncertainty becomes

$$U(M) = \pm k \sqrt{\begin{array}{l}\left(\frac{\partial M}{\partial|\Gamma_L|}\right)^2 u(|\Gamma_L|)^2 + \left(\frac{\partial M}{\partial\theta_L}\right)^2 u(\theta_L)^2 \\ + \left(\frac{\partial M}{\partial|\Gamma_G|}\right)^2 u(|\Gamma_G|)^2 + \left(\frac{\partial M}{\partial\theta_G}\right)^2 u(\theta_G)^2,\end{array}}$$

where θ_L and θ_G are the phases of the reflection coefficients of the load and the generator, respectively; the symbols $u(.)$ represent the uncertainties in the four input quantities with pdfs that are assumed to be normal distributions (However, in general, this will not be the case, particularly if $|\Gamma_G|$ and $|\Gamma_L|$ are small. More details on this can be found in [9].); and k is the coverage factor. However, the task to evaluate this uncertainty becomes rather more complicated.

This example shows the importance of determining the pdf for each different measurement technique. Whenever possible, the evaluation of uncertainty should be based on quantitative data derived from real experiments so that the conditions of a measurement set-up can be interpreted reliably.

Conclusions

This article has given an overview of some of the issues involved when considering the accuracy of measurements. The concept of uncertainty of measurement has been advocated as a reliable metric when assessing the quality of a measurement. This is because the "true value" of any measurement quantity is, at best, hard to find, whereas the uncertainty can almost always be quantified. A measurement result that includes a statement about the uncertainty and its confidence interval, or coverage factor, therefore is of far greater value to all concerned, whether this is in science, engineering, commerce, industry, or regulation.

Uncertainty of measurement, when used correctly, leads to continuous improvement and often results in improved efficiency, cost reductions, and better value for customers. With traceable measurements (a requirement implicit in a statement of uncertainty of a measurement), we can achieve accuracy via consensus. This leads to an agreed equivalence (or harmonization) of measurements regardless of where in the world these measurement are made. This provides an underpinning framework of reliability for global trade both within and across international boundaries.

2) If the uncertainty component is characterizing a systematic error (e.g., a value found in a manufacturer's data sheet), different distributions are used depending on the nature of the contributing error mechanism. The most common case for such errors is to assume a uniform (i.e., rectangular) distribution (Figure 8). This implies that there is no specific knowledge about values within the interval described by the limits of the distribution. Under these circumstances, the uncertainty interval is divided by $\sqrt{3}$ to establish the equivalent standard deviation.

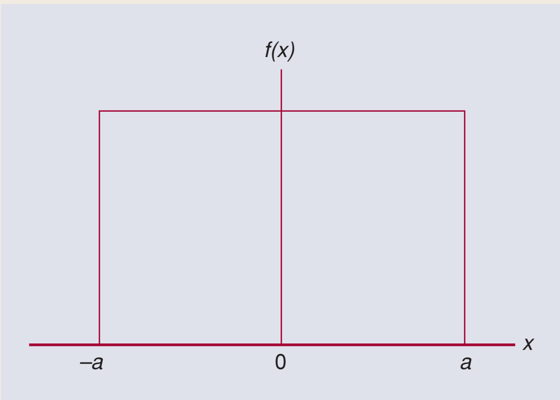


Figure 8. *A uniform (or rectangular) distribution, with limits −a and +a, used for uncertainty components when there is no specific knowledge about the whereabouts of a value within a given range.*

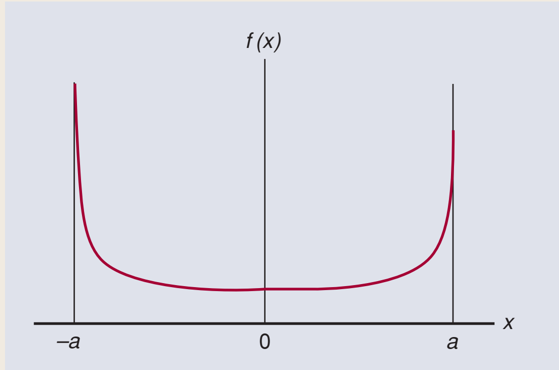


Figure 9. *A U-shaped distribution, with limits −a and +a, often associated with mismatch uncertainty where the vector representing the imperfect match has unknown phase.*

3) Since measurements at microwave frequencies often involve vector quantities (having both magnitude and phase), there is sometimes a need to consider using a U-shaped distribution (see Figure 9) to represent the uncertainty contribution (in the absence of detailed phase information for a given vector) [13]. Under these circumstances, the uncertainty interval is divided by $\sqrt{2}$ to establish the equivalent standard deviation.

In practice, all of the above three distributions (normal, rectangular, and U-shaped) may need to be considered during an evaluation of the uncertainty in a given measurement application.

With the widespread use of the GUM across the vast spectrum of measurement areas, there is now a standardized approach for expressing the uncertainty in measurement. This accelerates meaningful communication of all quantitative results (scientific or otherwise) across all stages (including research and development, manufacturing and testing, etc.) and promotes consistency of measurement in every aspect of life, including fair trade, public health, and the natural environment.

References

[1] BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, and OIML, *International Vocabulary of Basic and General Terms in Metrology*, 2nd ed. Geneva, Switzerland: International Organization for Standardization, 1993.

[2] "Guidelines on assessment and reporting of compliance with specification (based on measurements and tests in a laboratory)," LAC-G8, International Laboratory Accreditation Cooperation, 1996.

[3] BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, and OIML, *Guide to the Expression of Uncertainty in Measurement*, 1st ed. Geneva, Switzerland: International Organization for Standardization, 1993.

[4] B.N. Taylor and C.E. Kuyatt, "Guidelines for evaluating and expressing the uncertainty of NIST measurement results," NIST Tech. Rep. TN1297, Gaithersburg, MD, 1994.

[5] UKAS, "The expression of uncertainty and confidence in measurement," *M3003*, 2nd ed., 2007 [Online]. Available: www.ukas.org

[6] N.M. Ridler and J.C. Medley, "An uncertainty budget for VHF and UHF reflectometers," Teddington, U.K., NPL Rep. DES 120, May 1992 [Online]. Available: www.npl.co.uk

[7] M.J. Salter and N.M. Ridler, "A generalised approach to the propagation of uncertainty in complex S-parameter measurements," in *Proc. 64th ARFTG Conf.*, Orlando, FL, pp. 1–14, Dec. 2004.

[8] M.G. Cox, M.P. Dainton, P.M. Harris, and N.M. Ridler, "The evaluation of uncertainties in the analysis of calibration data," in *Proc. 16th IEEE Instrumentation and Measurement Technology Conf.*, Venice, Italy, pp. 1093–1098, May 24–26, 1999.

[9] N.M. Ridler and M.J. Salter, "An approach to the treatment of uncertainty in complex *S*-parameter measurements," *Metrologia*, vol. 39, no. 3, pp. 295–302, 2002.

[10] C.M. Wang and H.K. Iyer, "Uncertainty analysis for vector measurands using fiducial inference," *Metrologia*, vol. 43, no. 6, pp. 486–494, 2006.

[11] W. Bich, M.G. Cox, and P.M. Harris, "Evolution of the 'Guide to the Expression of Uncertainty in Measurement'," *Metrologia*, vol. 43, no. 4, pp. S161-S166, 2006.

[12] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML, "Evaluation of measurement data—Supplement to the 'Guide to the Expression of Uncertainty in Measurement'—Propagation of distributions using a Monte Carlo method," Joint Committee for Guides in Metrology, Bureau International des Poids et Mesures, JCGM 101, to be published.

[13] I.A. Harris and F.L. Warner, "Re-examination of mismatch uncertainty when measuring microwave power and attenuation," *IEE Proc.-H*, vol. 128, no. 1, pp. 35–41, Feb. 1981.