The measurement uncertainty of complex quantities

Blair Hall

Measurement Standards Laboratory of New Zealand
Callaghan Innovation

blair.hall@callaghaninnovation.govt.nz

BIPM Workshop on Measurement Uncertainty, 15-16 June 2015, BIPM, Sèvres France.

Outline

- What is special about complex quantities?
- International comparisons and CMCs
- Developments in the last 15 years
- Final comments

CallaghanInnovation

Impedance is a complex-valued quantity that is fundamental in RF and MW metrology

This talk will cover some aspects of the measurement uncertainty associated with complex-valued quantities in RF metrology that have challenged us

We begin by showing the evolution of ideas in a series of international comparisons.

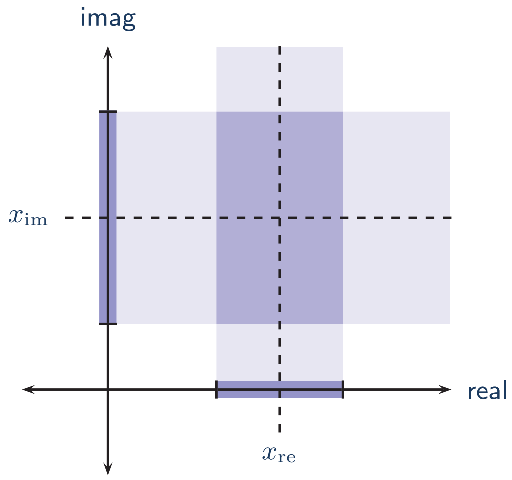
We then go back and identify key developments in our understanding over the same period of time.

Complex quantities are two dimensional ...

- Two components:

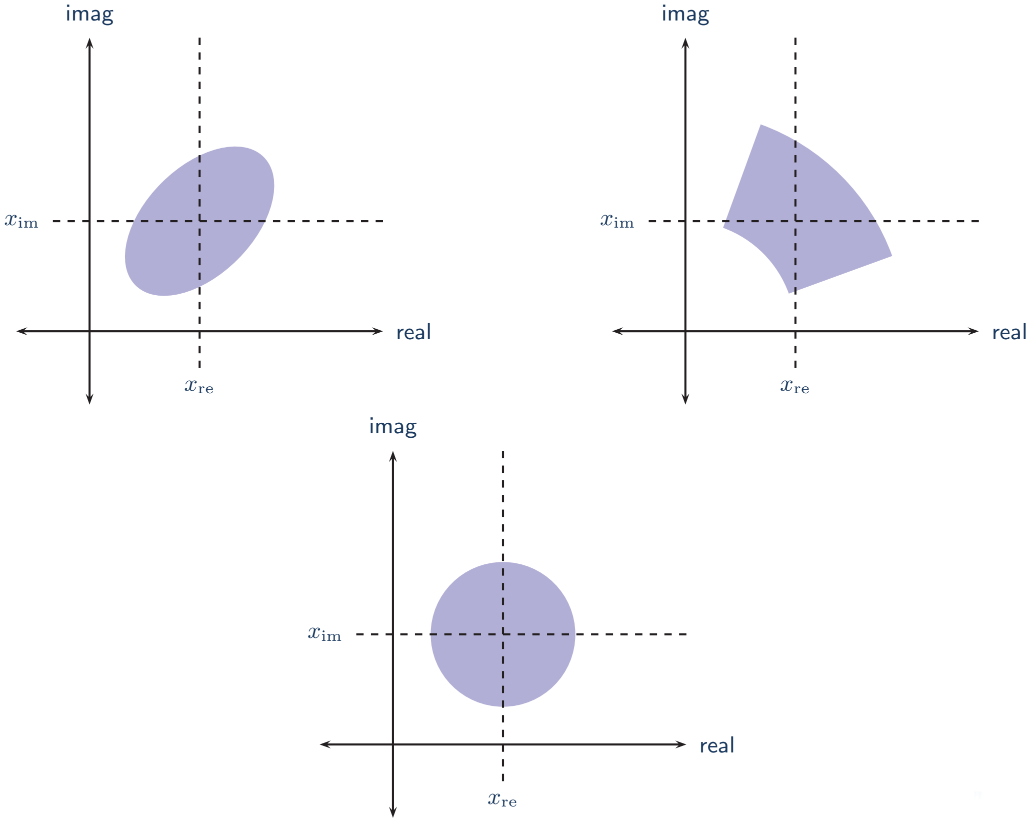
$$\boldsymbol{x} = x_{\mathrm{re}} + \mathrm{i}\, x_{\mathrm{im}}$$

- Univariate coverage factor, k_p, at a specific level of confidence
- Complex uncertainty is a region in the complex plane
- Bivariate coverage factor, $k_{2,p}$, is different from k_p
- $k_{2,p}$ depends on the shape of the uncertainty region too!



When expressed in terms of real and imaginary components, a region of uncertainty can be formed, for example, by the intersection of two bands in the complex plane.

The coverage probability of that region is clearly not the same (less) than the coverage probability of the individual bands, so we see that coverage factors for uncertainty regions must be different from univariate coverage factors.



The circular region has been used since about 1992 by NPL.[1]

There are rather severe conditions for its use to be satisfactory. However, when it is reasonable to make those assumptions there is also a greatly simplified method of propagating measurement uncertainty.[2]

The annular sector is implied by stating uncertainty in terms of an uncertainty interval of the magnitude and a simultaneous interval of the phase.

Note that different coverage factors may apply to different shapes!

[1] Ridler, N. M. and Medley, C. J. *An uncertainty budget for VHF and UHF reflectometers*, National Physical Laboratory, 1992

[2] Yhland, K. and Stenarson, *J. A simplified treatment of uncertainties in complex quantities*, CPEM Conference Digest, 2004, 652-653

CallaghanInnovation

RF measurement is traditionally an engineering discipline.

A variety of "convenient" units can describe essentially the same quantity

Reflection coefficient: $\boldsymbol{\Gamma} = \boldsymbol{b}/\boldsymbol{a} \;=\; \Gamma_{\mathrm{re}} + \mathrm{j}\,\Gamma_{\mathrm{im}}$

Magnitude: $\rho = |\boldsymbol{\Gamma}| \;=\; \sqrt{\Gamma_{\mathrm{re}}^2 + \Gamma_{\mathrm{im}}^2}$

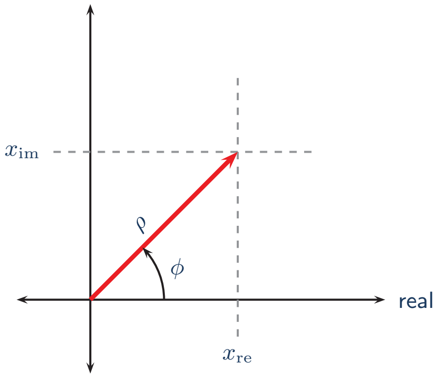
Return loss: $RL = -20\log_{10}(|\boldsymbol{\Gamma}|)$

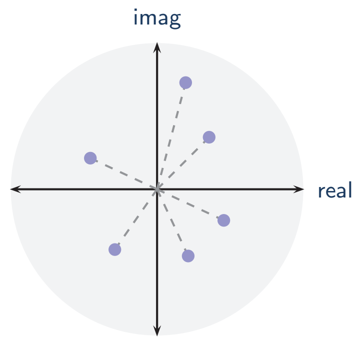
VSWR: $r = (1+\rho)/(1-\rho)$

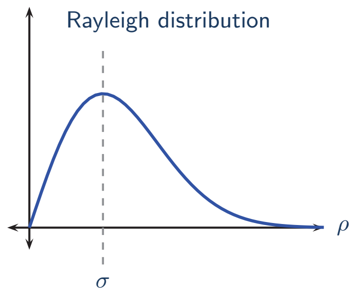
Phase: $\phi = \arg(\boldsymbol{\Gamma}) \;=\; \tan^{-1}(\Gamma_{\mathrm{im}}/\Gamma_{\mathrm{re}})$

The common use of engineering units is a complicating factor in RF metrology. It is not widely appreciated that the non-linear transforms involved are problematic when dealing with measurement uncertainty.

Problems with polar: small magnitudes







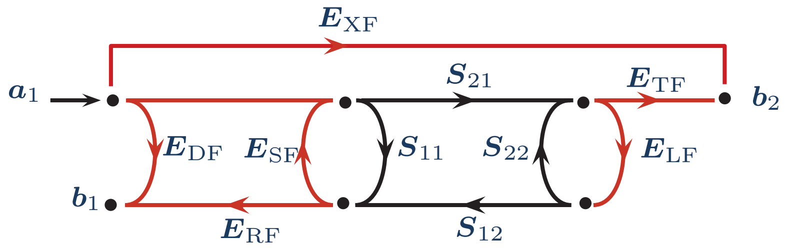
Polar coordinates are widely used in RF and microwave engineering. However, there are problems when dealing with uncertainties.

For example, when observations are distributed about the origin in a Gaussian distribution (so the mean is at the origin), the corresponding distribution of magnitudes is biased and the mode of the distribution occurs at the standard deviation of the Gaussian distribution.

… measurement models are complicated

The ubiquitous measurement instrument for complex quantities in RF metrology is a *Vector Network Analyser* (VNA).

VNA measurements are subject to many systematic errors: the "measurement model" is complicated. For instance,



$$s_{11} = \frac{\left\{\left(\frac{s_{11\mathrm{m}}-e_{\mathrm{df}}}{e_{\mathrm{rf}}}\right)\left[1+\left(\frac{s_{22\mathrm{m}}-e_{\mathrm{dr}}}{e_{\mathrm{rr}}}\right)e_{\mathrm{sr}}\right]\right\} - \left[\left(\frac{s_{21\mathrm{m}}-e_{\mathrm{xf}}}{e_{\mathrm{tf}}}\right)\left(\frac{s_{12\mathrm{m}}-e_{\mathrm{xr}}}{e_{\mathrm{tr}}}\right)e_{\mathrm{lf}}\right]}{\left[1+\left(\frac{s_{11\mathrm{m}}-e_{\mathrm{df}}}{e_{\mathrm{rf}}}\right)e_{\mathrm{sr}}\right]\left[1+\left(\frac{s_{22\mathrm{m}}-e_{\mathrm{dr}}}{e_{\mathrm{rr}}}\right)e_{\mathrm{sr}}\right] - \left[\left(\frac{s_{21\mathrm{m}}-e_{\mathrm{xf}}}{e_{\mathrm{tf}}}\right)\left(\frac{s_{12\mathrm{m}}-e_{\mathrm{xr}}}{e_{\mathrm{tr}}}\right)e_{\mathrm{lf}}e_{\mathrm{lr}}\right]}$$

Systematic errors are estimated by elaborate "calibration" procedures

Estimates are significant in the overall uncertainty budgets and also have important correlations

Many other errors must be considered too: some are "meta-stable" (enduring for shorter intervals)

In practice, it appears "much too hard" to express, and analyse, the complete measurement equation for the purpose of uncertainty propagation following GUM methods

… small samples and ‘type-B’ uncertainties

Using the original GUM as a template, we should also consider

Small samples (type-A)

For repeated measurements, sample estimate of covariance between complex components will usually not be zero.

How can the small sample size (finite degrees of freedom) together with correlation be handled?

Type-B uncertainty

What are suitable models of complex-valued measurement error?

Uncertainty propagation

How does the *Law of Propagation of Uncertainty* apply in two dimensions?

CallaghanInnovation

In the late 1990’s and early 2000’s it was clear that certain notions introduced by the GUM needed to be extended to allow their use with complex (bivariate) quantities.

Comparisons and CMC claims

CallaghanInnovation

It is interesting to look at the structure and reporting of international measurement comparisons. The structure is chosen to reflect best-practice among the best metrologists in the field at the time, so it is a reflection of currently accepted ideas, perhaps also involving a slight stretch for some participants.

CMCs on the other hand are the client facing interface of different NMI capabilities. They are reviewed by the responsible CC, so they also must reflect a certain conformity to current thinking.

GT-RF/83-4 (1999)

Impedance comparison, S-parameters 2-18 GHz; 9 participants

- Magnitudes reported in linear and dB units; phase in degrees
- No uncertainty budgets, no expanded uncertainty ($k = 1$)
- Many labs describe repeating measurements and re-calibrating *each* time.

Final report

- discusses results in magnitude and phase separately
- traceability not discussed
- emphasis on agreement, or otherwise, of results

Full repetition of measurement procedure would capture as many errors as possible statistically.

Systematic errors are not mentioned. Not clear how these may have been treated in uncertainty statements.

Uncertainty used as a measure of possible measurement *error* to explain or highlight agreement or significant differences between results.

Treating magnitude and phase separately suggests a univariate focus, as though they are two independent properties of the artefacts.

A CIPM Recommendation in 1986 stipulated that in comparisons the combined standard uncertainty would be reported in terms of "one standard deviation". However, in 1999 the MRA stipulated that CMCs would be stated at a coverage level of 95%. We see here the older recommendation is being followed.

EUROMET.EM.RF-S16 (2005)...

Impedance comparison, S-parameters 50 MHz - 50 GHz; 12 participants

- Magnitudes reported in linear and dB units; phase in degrees
- Uncertainty budgets requested, but not required
- Exemplar budgets provided, based on EA-10/12 Guide

Final report

- data processed in rectangular coordinates; reported in polar
- problems with transformation between these coordinates
- comparison reference values calculated (unweighted)
- CRV uncertainty an elliptical region (finite DoF, based on participants)
- traceability not discussed

CallaghanInnovation

Because of the MRA, we now have a complex-valued CRV and a region of uncertainty.

The decision to process data in rectangular coordinates comes from the 2002 Ridler + Salter paper (see slide 18)

EA Guide is mentioned in a following slide

Comments from the report:

> " The complexity of VNA measurements, the limited S-parameter measurement experience [in the new system] of some participants, the variety of different calibration methods and the ongoing discussion about the relevant contributions make it almost impossible to [compare] different uncertainty budgets in a [meaningful way]. "

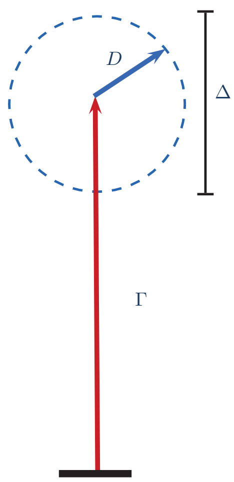
> " There is however potential for improvement in the field of measurement uncertainties. The different laboratories do not employ a [common approach] and methods are in use, which tend to fail under certain conditions. This reflects the fact that no truly established methods exist yet to evaluate measurement uncertainties in the multivariate case. "

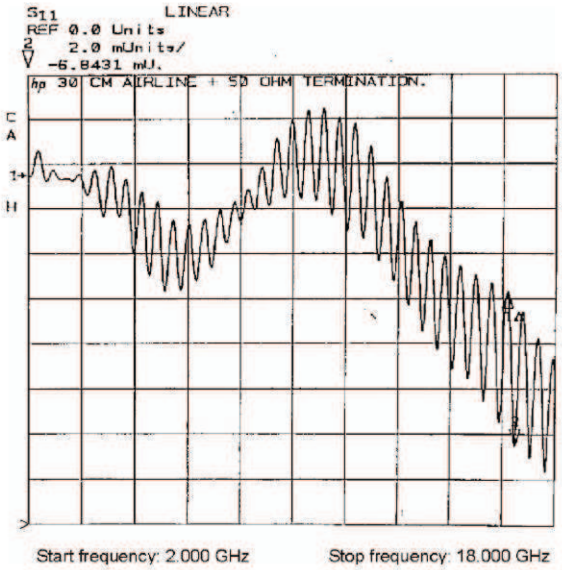
These comments show that current practice lacked uniformity, let alone objective meaning.

This has been a very influential document. First published in 2000, it was revised (and renamed as EURAMET/cg-12/v.01) in 2007.

The approach is 'top-down': a simple error model is assumed.

The magnitude (but, not phase) of residual errors are estimated





This is a consensus document complied by RF metrologists in Europe.

It was intended to be used by advanced commercial calibration laboratories, not just NMIs. So it describes fairly general methods.

Unfortunately, it is now widely recognised that the document contains mistakes. A new guide is being prepared.

Impedance comparison, S-parameters 2-18 GHz; 19 participants

- Report in rectangular coordinates
- Uncertainty budgets required
- Covariance between components (requested)

Final report

- fewer than half of participants provided covariance data
- many assume $u(\mathrm{real}) = u(\mathrm{imag})$
- many associate univariate distributions with influences (c.f., GUM)
- many claim traceability through check-standards
- KCRV uncertainty an elliptical region ($k = 2.5$)

Some labs used *ad hoc* software to process data and did not provide detailed uncertainty budgets. This was probably because such calculations generate rather too much information to communicate concisely in a report.

It is clear that few, if any, labs are adopting a GUM ‘bottom-up’ approach to uncertainty calculation.

When $u(\mathrm{real}) = u(\mathrm{imag})$ and there is no correlation we meet the conditions for circular uncertainty regions. This also suggests that the EA guide was being used.

CCEM.RF.K5c.CL (started 2012, on-going)

Impedance comparison, S-parameters 100 MHz - 33 GHz; 18 participants

- Results to be supplied in rectangular coordinates, in linear units
- Uncertainties required (covariance requested)
- A statement about traceability is required
- Uncertainty budgets requested

> " Ideally the participant would evaluate the uncertainty based on the characterization of [basic influences] that are propagated through the measurement model[3]. The uncertainty budget should list the contributions to the total uncertainty in terms of these basic influences. The residual error method should only [be used] if the participant does not have the capability of proper uncertainty propagation. "

[3]This is the method that is compliant with international guidance (GUM) and there are software tools available that support this method.

So in the space of about 15 years we see remarkable change in the RF community.

The initial comparison looked for consistency in results among different participants. Uncertainty is clearly a measure of the typical error magnitude, so that results can be compared. Measurands appear to be more univariate than bivariate.

The most recent comparison treats measurands as bivariate. Traceability is demanded. It is expected that uncertainty will be propagated, in other words that an unbroken chain of calibrations and measurements will be made.

However, what does the uncertainty mean? Is the intention still to resolve differences in results by accounting for reasonable measurement errors?

Calibration and measurement capabilities

An example from the CMC database,[4]

> *"Scattering parameters (vectors), Reflection coefficient in coaxial line (real and imaginary)"*

	$k = 2$	$k = 2.45$	rect	polar
NRC	*		*	
NIM	*			*
CMI	*			*
MIKES		*	*	
LNE	*		*	
PTB	*		*	
SCL	*		*	*
INRIM	*		*	
NMIJ	*			*
KRISS	*			*
CENAM	*			*
VSL	*		*	
SNIIM	*		*	
A*STAR	*			*
NMISA	*		*	*
INTA	*			*
SP		*	*	
METAS	*		*	*
UME	*		*	
NPL[5]		*	*	
NIST	*			*

[4] from `http://kcdb.bipm.org/`, May 2015
[5] $k = 2.5$

It is striking that only 3 of 21 NMIs report the expanded uncertainty of a bivariate quantity. By implication, $k = 2$ is a univariate coverage factor for 95% coverage probability.

Reporting in polar coordinates suggests that the magnitude and phase are considered as independent univariate quantities (c.f., GT-RF/83-4). However, to do so does not seem consistent with the CMC statement about the *vector* quantity being measured.

Furthermore, it is not clear what useful purpose would be served by an uncertainty interval with 95% coverage on just one of the rectangular components of a complex quantity. So, one has to wonder what the choice of $k = 2$ and 'rectangular' coordinates is intended to mean.

Also, importantly, we see no mention of the shape of uncertainty regions. This too affects the value of k that will give 95% coverage probability.

What has happened since 1999?

I began to work in this field in 1998-99. At that time, it seemed obvious that extensions to the GUM for complex quantities were needed. There was already talk of a GUM-Supplement to provide these extensions, but that supplement did not appear until 2011.

LPU

An equation involving physical quantities describes a measurement

$$\boldsymbol{Y} = \boldsymbol{f}(\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_N)$$

The uncertainty of $\boldsymbol{y}$ (as an estimate of $\boldsymbol{Y}$) is characterised by a symmetric 2×2 covariance matrix

$$\mathbf{v}(\boldsymbol{y}) = \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix}, \; v_{12} = v_{21}$$

calculated from the $2 \times 2N$ Jacobian matrix $\mathbf{J}(\boldsymbol{y})$ and the $2N \times 2N$ covariance matrix of the inputs $\mathbf{V}(\mathbf{x})$[6]

$$\boxed{\mathbf{v}(\boldsymbol{y}) = \mathbf{J}(\boldsymbol{y})\,\mathbf{V}(\mathbf{x})\,\mathbf{J}(\boldsymbol{y})'}$$

Degrees of freedom

An effective degrees of freedom ν_{eff} can be evaluated when there is correlation between the real and imaginary components of individual complex quantities[7]

[6] N. M. Ridler and M. F. Salter, *An approach to the treatment of uncertainty in complex S-parameter measurements*, Metrologia, 2002, **39**, 295-302

[7] R. Willink and B. D. Hall, *A classical method for uncertainty analysis with multidimensional data*, Metrologia, 2002, **39**, 361-9

CallaghanInnovation

The paper by Ridler + Salter is the most influential publication about measurement uncertainty for RF and microwave metrology. It does not contain new knowledge as such, but clearly presents general statistical methods in a context that can be recognised and used by member of the RF metrology community.

The paper on degrees of freedom complements the generalised LPU, just as the univariate LPU and the Welch-Satterthwaite equation are needed to propagate uncertainty in real-valued problems.

Propagation can be expressed in terms of individual influences[8]

$$\mathbf{v}(\boldsymbol{y}) = \sum_{i=1}^{N} \sum_{j=1}^{N} \mathbf{u}_i(\boldsymbol{y})\, \mathbf{r}_{ij}\, \mathbf{u}_j(\boldsymbol{y})'$$

The component of uncertainty matrices $\mathbf{u}_i(\boldsymbol{y})$ represent the uncertainty contribution from each influence quantity

$$\begin{aligned}\mathbf{u}_i(\boldsymbol{y}) &= \begin{bmatrix} \frac{\partial Y_{\mathrm{re}}}{\partial X_{i \cdot \mathrm{re}}} & \frac{\partial Y_{\mathrm{re}}}{\partial X_{i \cdot \mathrm{im}}} \\ \frac{\partial Y_{\mathrm{im}}}{\partial X_{i \cdot \mathrm{re}}} & \frac{\partial Y_{\mathrm{im}}}{\partial X_{i \cdot \mathrm{im}}} \end{bmatrix} \begin{bmatrix} u(x_{i \cdot \mathrm{re}}) & 0 \\ 0 & u(x_{i \cdot \mathrm{im}}) \end{bmatrix} \\ &= \left[\frac{\partial \boldsymbol{Y}}{\partial \boldsymbol{X}_i} \right] \mathbf{u}(\boldsymbol{x}_i)\end{aligned}$$

The matrix $\mathbf{r}_{ij}$ contains the four correlation coefficients between the real and imaginary components of the i^{th} and j^{th} influences

$$\mathbf{r}_{ij} = \begin{bmatrix} r_{i \cdot \mathrm{re},\, j \cdot \mathrm{re}} & r_{i \cdot \mathrm{re},\, j \cdot \mathrm{im}} \\ r_{i \cdot \mathrm{im},\, j \cdot \mathrm{re}} & r_{i \cdot \mathrm{im},\, j \cdot \mathrm{im}} \end{bmatrix}$$

[8] B. D. Hall, *On the propagation of uncertainty in complex-valued quantities*, Metrologia, 2004, **41**, 173-177

The Ridler-Salter presentation of the LPU was in a very succinct form, that does not identify individual components of uncertainty. An alternative formulation of the LPU explicitly identifies these components (which are 2×2 matrices) and allows them to be evaluated in terms of complex derivatives of the measurement equation.

This provides for a more intuitive description of the measurement uncertainty, which allows dominant components to be associated with individual influences.

This reference also identified ways to summarise the 2×2 matrices representing uncertainty as single numbers. This helps when interpreting results and removes some of the variability that can arise as uncertainty shifts among different matrix components (e.g., as the frequency is varied).

... uncertainty propagation

Complicated procedures may be represented by a sequence of intermediate results[9]

$$\boldsymbol{Y}_1 = \boldsymbol{f}_1(\Lambda_1)$$

$$\vdots$$

$$\boldsymbol{Y}_p = \boldsymbol{f}_p(\Lambda_p)$$

a component of uncertainty in $\boldsymbol{y}$ can be evaluated recursively; the i^{th} step being[10]

$$\mathbf{u}_j(\boldsymbol{y}_i) = \sum_{\boldsymbol{Y}_k \in \Lambda_i} \left[\frac{\partial \boldsymbol{Y}_i}{\partial \boldsymbol{Y}_k} \right] \mathbf{u}_j(\boldsymbol{y}_k)$$

This can be used to develop computational methods that assume the burden of propagating uncertainty during data processing.

[9] B. D. Hall, *On the propagation of uncertainty in complex-valued quantities*, Metrologia, 2004, **41**, 173-177

[10] Recursion stops when $\boldsymbol{y}_k$ refers to an influence estimate, not an intermediate result.

The propagation of uncertainty in long complicated models can also be handled as a series of intermediate calculations, with uncertainty components propagated at each step. This formulation allows software to automate the propagation of uncertainty.

Development of software for VNA uncertainty calculations has enabled the GUM approach to be applied to VNA measurements. It solved the problem posed by the very complicated measurement equations that arise in this area of metrology, especially the problems related to handling many systematic uncertainty terms.

Complex influences of real quantities

The original GUM may be used to evaluate uncertainty when a real measurand is subject to complex influence quantities. For example, the comparison loss

$$
\begin{aligned}
M = \frac{P_\mathrm{M}}{P_\mathrm{S}} &= \frac{1-|\boldsymbol{\Gamma}_\mathrm{M}|^2}{1-|\boldsymbol{\Gamma}_\mathrm{S}|^2}\,\frac{|1-\boldsymbol{\Gamma}_\mathrm{S}\boldsymbol{\Gamma}_\mathrm{G}|^2}{|1-\boldsymbol{\Gamma}_\mathrm{M}\boldsymbol{\Gamma}_\mathrm{G}|^2} \\
&\approx 1-|\boldsymbol{\Gamma}_\mathrm{M}|^2 \\
&= 1-\Gamma^2_{\mathrm{M\cdot re}}-\Gamma^2_{\mathrm{M\cdot im}}
\end{aligned}
$$

A problem arises when there is a type-A component of uncertainty.

When the estimate of $\boldsymbol{\Gamma}_\mathrm{M}$ is obtained from a sample of size N from bivariate Gaussian distribution, then[11]

$$DoF(|\boldsymbol{\Gamma}_\mathrm{M}|^2) = N-1$$

[11] R. Willink, *A generalzation of the Welch-Satterthwaite formula for use with correlated uncertainty components*, Metrologia, 2007, **44**, 340-349

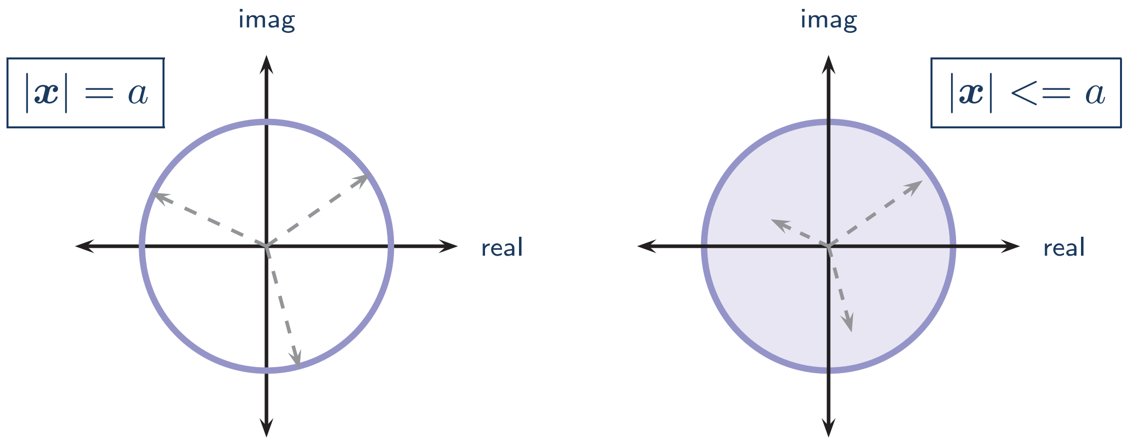
Measurements of reflection coefficients are usually repeated, because there is a possible source of error in the repeatability of connections. However, the number of repetitions is typically small (e.g., 6) so degrees of freedom are important.

When the measurand is real-valued but a reflection coefficient is an influence quantity, then a non-zero sample correlation coefficient between estimates of the real and imaginary components will prevent the application of GUM methods. This is because the Welch-Satterthwaite equation cannot be used when influence quantities are correlated.

The 2007 paper by Willink showed that, under the assumption that measurement errors in the real and imaginary components were Gaussian, and that the real and imaginary components were measured together, a modified version of Welch-Satterthwaite could be applied.

In RF and microwave measurement, a common case is the ignorance of phase [12] [13]

There are several possibilities



[12] B. D. Hall, *Some considerations related to the evaluation of measurement uncertainty for complex-valued quantities in radio frequency measurements*, Metrologia, 2007, **44**, L62-L67

[13] B. D. Hall, *On the expression of measurement uncertainty for complex quantities with unknown phase*, Metrologia 2011, **48**, 324-332

There is a predominant form of type-B uncertainty that occurs in RF and microwave measurements, which is due to a complete lack of information about the phase of a quantity.

The associated uncertainty component matrices have a simple diagonal form.

When the magnitude is known, this is an extension of the 1-D arcsine distribution[14]

- possible values of $\boldsymbol{x}$ form a uniform ring around the origin

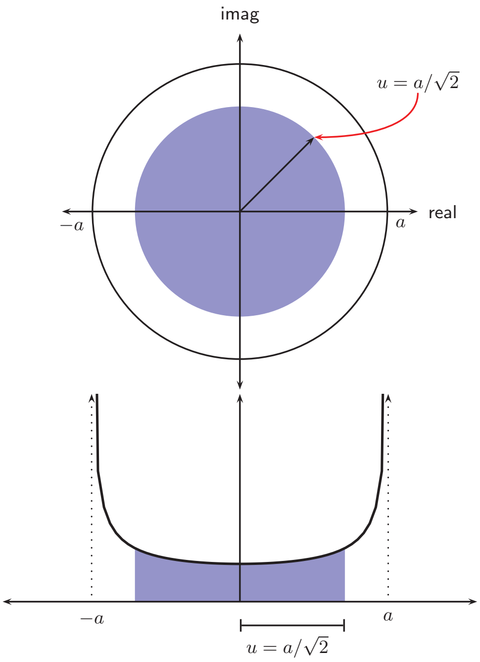
$$|\boldsymbol{x}| = a$$

- covariance matrix

$$\mathbf{v}(\boldsymbol{x}) = \frac{1}{2}\begin{bmatrix} a^2 & 0 \\ 0 & a^2 \end{bmatrix}$$

- uncertainty matrix

$$\mathbf{u}(\boldsymbol{x}) = \frac{1}{\sqrt{2}}\begin{bmatrix} a & 0 \\ 0 & a \end{bmatrix}$$



[14] I. A. Harris and F. Warner, *Re-examination of mismatch uncertainty when measuring microwave power and attenuation*, IEE Proceedings, 1981, **128**, 35-41

The applications considered by Harris and Warner usually involve the combination of more than one 'mismatch' term, which in fact tends to produce a combined effect close to a Gaussian distribution.

Dobbert and Gorin have shown recently that in practical applications, provided a wide range of frequencies are included, the distribution is essentially a bivariate Gaussian at the origin with a diagonal covariance matrix.[15]

A very early paper predicted bivariate normal distributions centred at the origin for mismatch.[16]

[15] M. Dobbert and J. Gorin, *Revisiting mismatch uncertainty with the Rayleigh distribution*, NCSL International Workshop and Symposium, 2011

[16] Mullen, J. A. and Pritchard, W. L. *The statistical prediction of voltage standing-wave ratio*, IRE Transactions on Microwave Theory and Techniques, 1957, 127-130

GUM Supplements

CallaghanInnovation

S1 deals only with univariate problems and introduces the Monte Carlo Method (MCM). It does not address the question of correlation that arises when a complex influence quantity is estimated with finite degrees of freedom during the evaluation of a real-valued measurand.

S2 has two distinct parts. The first deals with multivariate extensions to the analytic GUM methods.

- LPU - in matrix form
- implicit problems
- elliptical uncertainty regions and Bonferroni rectangular regions
- case of finite DoF is given

The notion of uncertainty budgets is not mentioned, nor are component of uncertainty matrices, and their summary magnitudes.

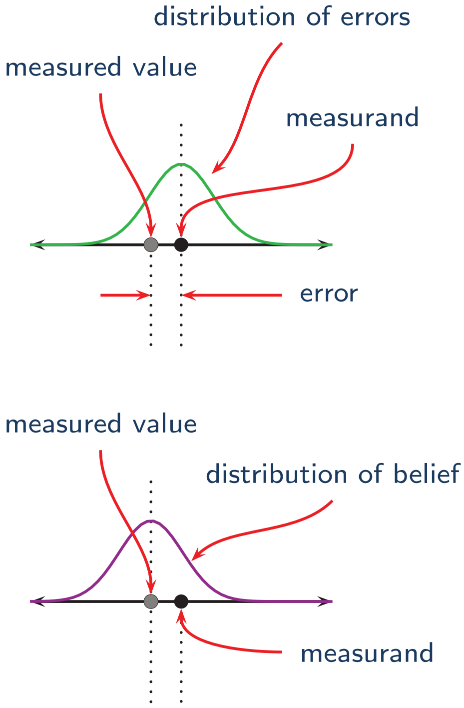
The method of evaluating an effective DoF is not mentioned either.

Such omissions are unfortunate. They do not give an accurate picture of the tools available to evaluate measurement uncertainty in complex quantities.

The second part of S2 introduces the MCM for multivariate problems.

Probability, errors and uncertainties

- The *error* is the difference between an observed and the actual value.
- A statistical distribution can model our limited knowledge about error
- This is a *physical* model, for which we have some understanding and intuition
- Distributions of belief about the uncertainty of influence quantities are not part of a physical model!



"[the Monte Carlo Method] is regarded as a means for providing a numerical representation of the distribution for the output quantity, rather than a simulation per se. In the context of the propagation stage of uncertainty evaluation, the problem to be solved is deterministic, there being no random process to be simulated."
(from the first GUM supplement [Section 5.4.1, Note 2])

S1 and S2 advocate the use of a numerical method (the MCM) for propagating 'distributions of belief' about influence quantities. It is important to understand that a 'distribution of belief' is not the same thing as a model of physical error. Distributions of belief are expressions of Bayesian probability: they are not observable by any hypothetical experiment!

The supplements assert that the propagation of distributions approach is very accurate and can be used to validate the approximate GUM methods.

However, simulation studies do not support this assertion.

A comparison comparison

(GUM S1 §8.1.1, S2 §8.1)

> *" Since the domain of validity for MCM is broader than that for the GUM uncertainty framework, it is recommended that both the GUM uncertainty framework and MCM be applied and the results compared. Should the comparison be favourable, the GUM uncertainty framework could be used on this occasion and for sufficiently similar problems in the future. Otherwise, consideration should be given to using MCM or another appropriate approach instead. "*

We chose a very simple comparison and compared the nominal coverage probability with the observed coverage frequency using many simulated data sets.

Measurands: $|\Gamma|$ and $|\Gamma|^2$

Required: expanded uncertainties at a 95% level of confidence

To test metrological software, one can generate data representing problems that have known answers. Running software algorithms on such data sets then tests the algorithm performance. The data can be generated in ways that control how 'difficult' it is numerically to obtain an acceptable result.

Simulated observations

- Fixed 'target' value

$$\mathbf{\Gamma} = \Gamma_{\mathrm{re}} + \mathrm{j}0$$

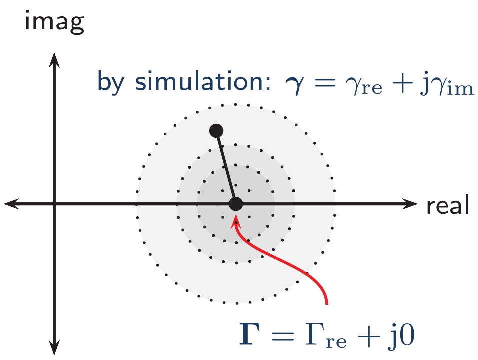
- Simulated observations

$$\gamma_{\mathrm{re}\cdot i} = \Gamma_{\mathrm{re}} + \varepsilon_{\mathrm{re}\cdot i}\,,$$
$$\gamma_{\mathrm{im}\cdot i} = \Gamma_{\mathrm{im}} + \varepsilon_{\mathrm{im}\cdot i}$$

- Random errors

$$\varepsilon_{\mathrm{re}\cdot i} \sim N(0, u^2)\,,$$
$$\varepsilon_{\mathrm{im}\cdot i} \sim N(0, u^2)$$

- observe coverage with fixed $\mathbf{\Gamma}$
- change $\mathbf{\Gamma}$ to observe different physical scenarios

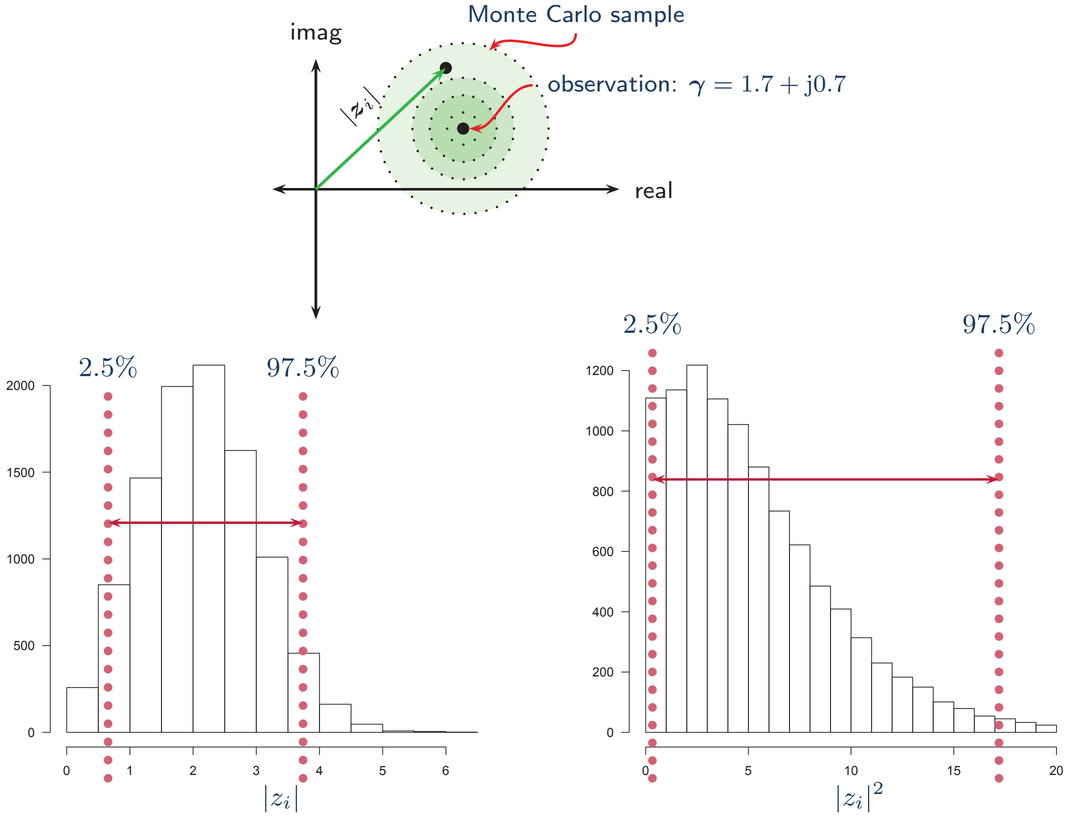


In this case, we simulate experimental data in which the measurand is fixed and noise is added to generate independent observations

We will apply both the usual GUM method and the S1 MCM method to obtain uncertainty statements for estimates of the quantities $|\mathbf{\Gamma}|$ and $|\mathbf{\Gamma}|^2$. Since we know the value of $\mathbf{\Gamma}$ used to simulate the observational data, we are able to classify each uncertainty interval as either containing the measurand or not (a success or failure).

For a nominal coverage probability of 95%, we expect to observe approximately 95 out of 100 successes as a long-term average as we generate many independent sets of observations.

S1 method for $|\Gamma|$ and $|\Gamma|^2$



It is important to note (and usually confusing at first impression) that the MCM also generates a data set by drawing from random number generators. This however, is NOT a physical simulation (see footnote to slide 25), it is merely a numerical process used to perform the desired calculation.

Performance results

The performance of both methods gave us some surprises[17]

Successes from 1000 simulated experiments:

$\mathbf{\Gamma}/u$	$\lvert\mathbf{\Gamma}\rvert$		$\lvert\mathbf{\Gamma}\rvert^2$	
	GUM	SUP	GUM	SUP
0.1	884	0	1000	0
0.2	880	0	999	0
0.5	920	735	999	765
1.0	955	908	984	913
2.0	956	934	943	959
5.0	948	945	942	950
10.0	943	943	948	950

- Monte Carlo (MC) sample size of $L = 10^4$
- Standard uncertainty in these coverages ≈ 7

[17] B. D. Hall, *Assessing the performance of uncertainty calculations by simulation*, Microwave Measurement Symposium Digest, 74th ARFTG, 2009

This is a problematic measurement.[18] [19]

The point here is not to suggest that GUM is superior to S1; it is to show that S1 fails, rather badly, when coverage is used as a performance criterion. That fact casts doubt on the recommendation in S1 (and in S2) that MCM should be used to validate uncertainty calculations.

This simple counter example shows that the accuracy claimed is not guaranteed by the MCM. It follows that validation by MCM (as described in S1 and S2) is no guarantee of satisfactory performance in any real sense (related to outcomes of real measurements).

A satisfactory validation scheme could be based on the simulation method used here to compare the two methods of uncertainty calculation.

[18] Oberto, L. and Pennecchi, F. *Estimation of the modulus of a complex-valued quantity*, Metrologia, 2006, 43, 531-538

[19] Pennecchi, F. and Oberto, L. *Uncertainty evaluation for the estimate of a complex-valued quantity modulus*, Metrologia, 2010, 47, 157-166

S2 small sample estimates

The second GUM supplement uses a particular multivariate t-distribution to describe the uncertainty of an estimate based on a small sample of observations.

How does that compare with an extension to the classical Welch-Satterthwaite calculation? [20]

$$\boldsymbol{Y} = \boldsymbol{X}_1 + \boldsymbol{X}_2$$

n_1	n_2	classical	H distribution	n_1	n_2	classical	H distribution
3	3	984	999	8	8	961	972
3	4	974	995	8	10	958	966
3	8	953	982	8	15	956	966
3	10	951	981	10	10	959	964
3	15	948	978	10	15	952	959
4	4	973	992	15	15	955	958
4	8	958	979				
4	10	953	976				
4	15	952	973				

[20] B. D. Hall, *Monte Carlo uncertainty calculations with small-sample estimates of complex quantities*, Metrologia, 2006, 43, 220-226

CallaghanInnovation

The standard uncertainty due to simulation size is again about 7.

Note that the coverage probability for H is consistently high, which also implies that the uncertainty regions are much larger than they need to be.

Also the choice of the distribution H is somewhat arbitrary. This distribution gives 'exact' coverage when the measurement equation is Y=X, but becomes conservative here. The marginals of H will also be conservative in this case (the real or imaginary components).

Klauneberg and Elster[21] associate the H distribution with the 'independence Jeffreys prior' (independent under re-parametrisation)

According to K and E, we would need about 24 observations before the (marginal) uncertainty estimates from other types of multivariate t distribution would be less than 5%. This is alarming. It means that there is a sensitivity in the results to the type of t-distribution chosen (in one dimension there is only one type of t distribution but there is a large number in the multidimensional case).

Again, we see that the coverage of the method proposed in S2 is inaccurate when compared with an objective measure of coverage probability. This is unacceptable.

[21] Klauenberg, K. and Elster, C. *The multivariate normal mean - sensitivity of its objective Bayesian estimates*, Metrologia, 2012, 49, 395-400

Final comments

- The RF metrology community is working hard to adopt notions of uncertainty expressed in the GUM.
- Basic conceptual issues must still be clarified.
  - When reporting reported measurement results, or CMCs, what information needs to be conveyed?
  - How should the associated uncertainty statement be formulated?
  - What is a suitable objective interpretation for uncertainty?
- A different approach, adopted in GUM S1 & S2, is difficult to relate to physical measurement errors. Claims about the accuracy of this approach seem overstated.

CallaghanInnovation

There are now adequate tools available for uncertainty specification and propagation for complex quantities (DoF, LPU, correlation and type-B uncertainty). These are extensions of the familiar GUM methods, and retain the so-called ‘inconsistent’ features of the GUM (linear approximation, finite degrees of freedom, blurring of frequentist and Bayesian interpretations).

Nevertheless, they offer a familiar framework for metrologists to work with. They also perform better than the MCM methods proposed in S1 and S2 in all cases that we have explored, which are related to RF and microwave metrology. Hence the so-called ‘inconsistencies’ of the extended GUM methods do not appear to be problematic for RF and MW metrology.

Using simple simulations of physical measurements, where the ideal result (measurand) is known, we find that S1 and S2 (Bayesian) methods do not perform well in terms of coverage in situations of interest to RF and MW metrology. That is, they are not good at accounting for the effects of physical measurement error on the measurement result.

This is a serious shortcoming, given the claims made about their high accuracy. We think that those claims should be revised.