VALIDATING THE USE OF THE GUM UNCERTAINTY FRAMEWORK

Maurice G Cox and Peter M Harris
National Physical Laboratory
Teddington, Middlesex TW11 0LW, UK
Tel +44 20 8943 6096, Fax +44 20 8977 7091, E-mail maurice.cox@npl.co.uk

Abstract: The uncertainty framework presented in the *Guide to the Expression of Uncertainty in Measurement* (GUM) is widely used. Whether the conditions attached to it hold for a problem of concern is not always verified. The possibility thus exists that some uncertainty evaluations based on the framework are invalid. This paper considers the application of a procedure for validating the framework in any particular instance. The procedure is based on an implementation of the propagation of distributions and constitutes part of Supplement 1 to the GUM, which has been developed by the Joint Committee for Guides in Metrology.

1. INTRODUCTION

The *Guide to the Expression of Uncertainty in Measurement* (GUM) [1] is founded on probability theory. Information concerning the values of quantities of concern is represented in terms of probability density functions (PDFs) for these values. In recognition, at the time of its development, of the difficulties of carrying out uncertainty evaluation for general (linear or non-linear) models using general PDFs, the GUM provided a simplified approach, termed here the *GUM uncertainty framework*.

To overcome working with models that can be arbitrarily complicated, the GUM uncertainty framework primarily operates in terms of a model linearized about the best estimates of the values of the input quantities. Rather than working with the PDFs themselves, the framework operates with *summary parameters* of the PDFs, viz., expectations (means), standard deviations (and covariances in the case of joint PDFs), and degrees of freedom when appropriate. This information is used to propagate uncertainties and to assign a Gaussian PDF or a PDF related to the *t*-distribution to the value of the output quantity in order to obtain a coverage interval for that value.

There are consequently conditions associated with the use of the framework, which ideally should be verified in any particular case. Since doing so is generally far from easy, it is appropriate to make use of a form of independent validation.

This paper outlines guidance prepared in this regard. The probabilistic basis of the GUM and an approach termed the *propagation of distributions* [2] for generating the PDF for the value of the output quantity are outlined (section 2.1). All required uncertainty information concerning the value of the output quantity can be inferred from this PDF. The GUM uncertainty framework and conditions for its application are reviewed (section 2.2). A numerical approach, viz., Monte Carlo simulation (MCS) [2], to implement the propagation of distributions is summarized (section 2.3). A procedure based on MCS for validating the GUM uncertainty framework is presented (section 3) and an illustrative example from mass metrology given (section 4). The attitude taken to uncertainty evaluation and its validation is discussed and concluding remarks made (section 5).

2. THE PROPAGATION OF DISTRIBUTIONS AND ITS IMPLEMENTATION

2.1 The probabilistic basis of the GUM

The GUM focuses on the use of a measurement model $Y = f(\boldsymbol{X})$ to relate *input quantities* $\boldsymbol{X}$, about which information is available, to an *output quantity* Y, about which information is required. Probability density functions (PDFs) are assigned to the values of $\boldsymbol{X}$. These PDFs are obtained from an analysis of series of observations [GUM clauses 2.3.2, 3.3.5][1] or based on scientific judgment using all the relevant information available [GUM 2.3.3, 3.3.5]. There is a unique PDF for the value of Y given the model f and PDFs for the values of $\boldsymbol{X}$.

The determination of the PDF for the value of Y from the model f and the PDFs for the values of $\boldsymbol{X}$ constitutes the propagation of distributions. The PDF for the value of Y is a fundamental entity in that (a) its expectation (mean) is taken as the best estimate y of the value of Y, (b) its standard deviation is taken as the standard uncertainty $u(y)$

[1] Subsequently, citations to clauses of the GUM are given, e.g., as [GUM 2.3.2].

associated with y, and (c) it enables a 95 % (say) coverage interval for the value of Y to be formed.

2.2 The GUM uncertainty framework and conditions for its application

The GUM provides a framework for calculating $u(y)$ using the law of propagation of uncertainty [GUM 5] and a coverage interval for the value of Y by assigning a particular PDF to that value.

The law of propagation of uncertainty is based on an expansion of f about the best estimates $\boldsymbol{x}$ of $\boldsymbol{X}$ as a first-order Taylor series [GUM 5.1, 5.2]. Although this approximation is usually adopted, when the non-linearity of f is significant, higher-order terms in the Taylor series expansion must be included in the expression for $u(y)$ [GUM 5.1.2]. However, formulae for only the most important terms of next highest order are provided in the GUM. Moreover, although it is not stated there, these terms apply specifically to cases where all the values of $\boldsymbol{X}$ follow mutually independent Gaussian distributions. Thus, in particular, no corresponding formula for correlated input quantities is given, even in the Gaussian case.

In cases where at least one of the standard uncertainties associated with the estimates of the values of the input quantities has a finite degrees of freedom, the GUM establishes an 'effective degrees of freedom' of $u(y)$. The Welch-Satterthwaite formula is used for this purpose. This formula is an approximation. No formula is provided when the values of $\boldsymbol{X}$ are mutually dependent.

To establish a coverage interval for the value of Y, the Gaussian PDF $N(y, u^2(y))$ is assigned to this value, and the symmetric interval about y that embraces 95 % of the distribution taken. (A PDF based on the t-distribution is used when the degrees of freedom of u(y) are finite.) The validity of this assignment generally depends on the applicability of the Central Limit Theorem. This theorem does not hold, for example, for a small number of input quantities whose values are assigned PDFs rather different from Gaussian, or when there is an input quantity that is dominant and the value of which is assigned a PDF that departs appreciably from Gaussian.

In general, the GUM uncertainty framework can be regarded as an approximate implementation of the propagation of distributions, where there is no 'user' control over the quality of the approximation.

2.3 Monte Carlo implementation of the propagation of distributions

An estimate of the value of Y is usually obtained by evaluating the model at the estimates $\boldsymbol{x}$ of the input quantities $\boldsymbol{X}$. However, since the values of $\boldsymbol{X}$ are described by PDFs rather than single numbers, other realizations of the value of Y can be obtained by drawing values at random from these PDFs.

Monte Carlo simulation (MCS) is a *numerical* implementation of the propagation of distributions based on this consideration. A particular value at random from the PDF for the value of each input quantity is generated and the corresponding value of Y formed by evaluating $f(\boldsymbol{X})$ for these particular values. Repeating this process many times, M (= 10^5, say) values of Y are obtained. These values are used to approximate the distribution function (the indefinite integral of the PDF) for the value of Y.

The value chosen for M controls the quality of the approximation obtained. The adaptive determination of M, rather than using an *a priori* value, in principle enables any numerical accuracy in the distribution function for the value of Y to be delivered. This statement applies to the *formulated* problem, viz., the model itself and the PDFs assigned to the values of $\boldsymbol{X}$. It does not relate to the quality of the model as an adequate description of the measurement.

A basic implementation of an adaptive MCS procedure carries out an increasing number of Monte Carlo trials until the quantities of interest have stabilized in a statistical sense. A quantity is deemed to have stabilized if twice the standard deviation associated with the estimate of its value is less than the degree of numerical approximation required in the standard uncertainty $u(y)$.

The process consists of undertaking a *sequence* of Monte Carlo calculations, each containing a small number, say, $M = 10^4$ trials. For each such calculation, form y, $u(y)$ and the endpoints of a 95 % coverage interval from the results obtained. Denote by $y^{(h)}$, $u(y^{(h)})$, $y_{low}^{(h)}$ and $y_{high}^{(h)}$ the values of these quantities for the hth member of the sequence.

After the hth Monte Carlo calculation (apart from the first) in the sequence, form the arithmetic mean of the values $y^{(1)}, \ldots, y^{(h)}$ and the standard deviation $s(y)$ associated with this mean. Determine the counterparts of these statistics for $u(y)$, y_{low} and y_{high}. Regard the overall computation as having stabilized if the largest of $2s(y)$, $2s(u(y))$, $2s(y_{low})$ and $2s(y_{high})$

does not exceed the degree of numerical approximation required in $u(y)$. Use the results from the total number of Monte Carlo trials taken to provide an approximate distribution function for the value of Y, from which an estimate of the value of Y, the associated standard uncertainty and a coverage interval for the value of Y are obtained.

The Joint Committee for Guides in Metrology (JCGM) is the body responsible for maintaining the GUM. It is developing supplements to the GUM, rather than explicitly changing the GUM itself. The first supplement [3] concerns the use of MCS as an implementation of the propagation of distributions, including the use of MCS as a validation facility.

3. VALIDATION PROCEDURE

Although the GUM uncertainty framework can be expected to work well in many circumstances, it is generally difficult to quantify the effects of the approximations involved in (a) model linearization, (b) the Welch-Satterthwaite formula for the effective degrees of freedom, and (c) taking the value of the output quantity as Gaussian. Indeed, the degree of complication of doing so would typically be considerably greater than that required to apply MCS. Therefore, since these circumstances cannot readily be tested, cases of doubt should be validated. To this end, since the propagation of distributions is more general and applies without approximation other than that related to random sampling, it is recommended that both the GUM uncertainty framework and MCS be applied and the results compared. If the comparison is favourable, the GUM uncertainty framework can be used (and also for sufficiently similar problems). Otherwise, consideration can be given to using MCS instead.

Specifically, a comparison procedure is recommended based on determining whether the coverage intervals obtained by the GUM uncertainty framework and MCS agree to a stipulated degree of numerical approximation. This degree of approximation is assessed in terms of the endpoints of the coverage intervals and corresponds to that given by expressing the standard uncertainty $u(y)$ to what is regarded as a meaningful number of *significant decimal digits*.

The concentration is on the coverage interval rather than $u(y)$, since the former is almost invariably a more sensitive entity because of its distributional dependence. The procedure is as follows:

1. Apply the GUM uncertainty framework to yield a 95 % coverage interval $y \pm U(Y)$ for the value of Y.

2. Apply MCS to yield the standard uncertainty $u(y)$ associated with an estimate of the value of Y and the endpoints y_{low} and y_{high} of a 95 % coverage interval for the value of Y.

3. Let n_{dig} denote the number of significant digits regarded as meaningful in the numerical value of $u(y)$. Usually, $n_{\mathrm{dig}} = 1$ or $n_{\mathrm{dig}} = 2$. Express the value of $u(y)$ in the form $a \times 10^{r}$, where a is an n_{dig}–digit integer and r an integer. The comparison accuracy is $\delta = ½ \times 10^{-r}$.

4. Compare the coverage intervals obtained by the GUM uncertainty framework and MCS to determine whether the required numerical accuracy in the coverage interval provided by the former has been obtained. Specifically, determine the quantities $|y - U(Y) - y_{\mathrm{low}}|$ and $|y + U(Y) - y_{\mathrm{high}}|$, viz., the absolute values of the differences of the respective endpoints of the two coverage intervals. Then, if both these quantities are no larger than δ the comparison is successful and the GUM uncertainty framework has been validated in this instance.

As an example [GUM 7.2.2], the estimate of the value of a nominally 100 g standard of mass is y = 100.021 47 g and the associated standard uncertainty $u(y)$ = 0.000 35 g. Thus, $n_{\mathrm{dig}} = 2$ and $u(y)$ is expressed as 35×10^{-5} g, and so $a = 35$ and $r = -5$. Take $\delta = ½ \times 10^{-5}$ g = 0.000 005 g.

4. EXAMPLE: MASS CALIBRATION

A model for the calibration of a weight W of mass density ρ_{W} against a reference weight R of mass density ρ_{R} having the same nominal mass using a balance operating in air of mass density a can be expressed [4] as

$$m_{\mathrm{W,c}} = (m_{\mathrm{R,c}} + \delta m_{\mathrm{R,c}})\{1 + (a - a_0)(1/\rho_{\mathrm{W}} - 1/\rho_{\mathrm{R}})\}. \quad (1)$$

Here, $a_0 = 1.2$ kg/m^3 and $m_{\mathrm{W,c}}$, $m_{\mathrm{R,c}}$ and $\delta m_{\mathrm{R,c}}$ are conventional masses. For instance, the conventional mass $m_{\mathrm{W,c}}$ of W is the mass of a (hypothetical) weight of density ρ_0 = 8 000 kg/m^3 that balances W in air at density a_0. As such, a_0 and ρ_0 are regarded as having no associated uncertainty.

Table 1 lists the input quantities, regarded as mutually independent, and the PDFs assigned to their values for model (1). The semi-widths of the PDFs for the values of a, ρ_W and ρ_R (and hence the corresponding uncertainties) are 'large' because a is monitored rather than measured and of the extent of the knowledge of the values of ρ_W and ρ_R

Table 1. The input quantities and the PDFs assigned to their values for the mass calibration model (1). A Gaussian PDF (G) is described by its expectation and standard deviation, and a rectangular PDF (R) by its expectation and semi-width.

Input quantity	PDF	Parameter values
$m_{R,c}$	G	100 000.000 mg, 0.050 mg
$\delta m_{R,c}$	G	1.234 mg, 0.020 mg
a	R	1.20 kg/m^3, 0.10 kg/m^3
ρ_W	R	8 000 kg/m^3, 1 000 kg/m^3
ρ_R	R	8 000 kg/m^3, 50 kg/m^3

Let $\delta m = m_{W,c} - m_{nom}$ be the deviation of $m_{W,c}$ from the nominal value $m_{nom} = 100$ g. The GUM uncertainty framework, and MCS with $M = 10^5$ trials, were each used to obtain an estimate δm of the value of the output quantity, the associated standard uncertainty $u(\delta m)$ and a 95 % coverage interval for the value of the output quantity. The results obtained from these approaches are shown in the rows of Table 2 labelled GUF1 and MCS. Figure 1 shows the approximations to the PDF for the value of the output quantity obtained from the two approaches.

The results show that, although the GUM uncertainty framework and MCS give estimates of δm in good agreement, the values for $u(\delta m)$ are noticeably different. The value (0.075 5 mg) of $u(\delta m)$ returned by MCS is 40 % larger than that (0.053 9 mg) returned by the GUM uncertainty framework. The latter is thus optimistic in this respect. The calculations using the GUM uncertainty framework were repeated, but including higher-order terms (using partial derivatives of up to third order) [GUM 5.1.2]: see the row of Table 2 labelled GUF2. The results agree much better with those from MCS.

Table 2 also shows in the right-most two columns the results of applying the comparison procedure of section 3 in the case where one significant digit is regarded as meaningful, i.e., $n_{dig} = 1$ using the terminology of that section. Hence, $u(\delta m) = 0.08 = 8 \times 10^{-2}$, and so $a = 8$ and $r = -2$. Thus, $\delta = ½ \times 10^{-2} = 0.005$. The magnitudes of the endpoint differences are shown, and whether the use of the GUM uncertainty framework has been validated. If only first-order terms are accounted for, the application of the GUM uncertainty framework is not validated. If higher-order terms are included, the GUM uncertainty framework is validated.

Table 2. Results for the mass calibration model from (a) the GUM uncertainty framework (GUF1), (b) MCS with 10^5 trials, and (c) GUF with higher order terms (GUF2). δm denotes the mass deviation, u(δm) the associated standard uncertainty, 95 % CI the endpoints of the 95 % coverage interval for the value of the output quantity value, d_{low} and d_{high} the magnitudes of the endpoint differences and V whether the results were validated.

Method	δm /mg $u(\delta m)$ /mg	95 % CI	d_{low} /mg d_{high} /mg	V
GUF1	1.234 0	1.128 4	0.043 9	No
	0.053 9	1.339 6	0.044 6	
MCS	1.234 3	1.084 5		
	0.075 5	1.384 2		
GUF2	1.234 0	1.087 0	0.002 5	Yes
	0.075 0	1.381 0	0.003 2	

5. DISCUSSION AND CLOSING REMARKS

The GUM uncertainty framework, although seeming to work well in many circumstances, has some limitations. Regarding quality management systems and laboratory accreditation, it is appropriate that the extent to which this framework is fit for purpose for the tasks to which it can validly be applied is established. When the conditions for its use do not hold or, more likely, it is not known whether they hold, it is especially fitting that appropriate validation is carried out. If the validation fails, it is necessary to use instead an approach having the flexibility to provide the required uncertainties to a stipulated degree of numerical approximation.

The GUM uncertainty framework can be regarded as an approximate implementation of the propagation of distributions. If the propagation of distributions could be implemented perfectly, it would provide a 'reference solution', with which results obtained by applying the GUM uncertainty framework, or other approximate approaches, could be compared. Since such an implementation is not generally possible, it is reasonable that results from the GUM uncertainty framework be compared with those from an implementation of the propagation of distributions having a degree of approximation that is under some control. 'Adaptive Monte Carlo' endeavors to provide such control. The convergence of an adaptive approach is stochastic rather than mathematical. Consequently, there is no guarantee that a stipulated degree of numerical accuracy has been achieved in any instance. However, the approach permits a *form* of control, which does not exist in general for the GUM uncertainty framework.

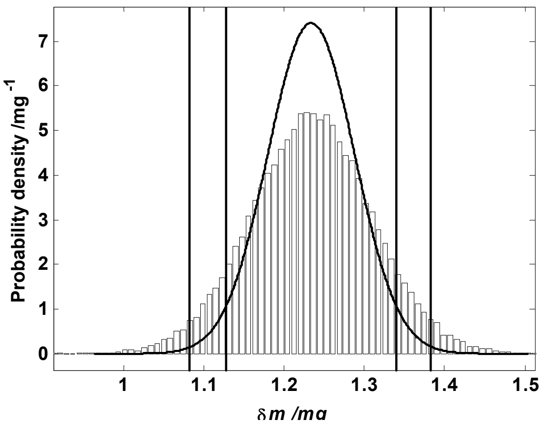

Figure 1. Approximations to the PDF for the mass deviation δm obtained using the GUM uncertainty framework and MCS. The solid curve represents a Gaussian PDF with parameters given by this framework. The histogram is derived from the MCS values as an approximation to the PDF. The solid vertical lines are the endpoints of a 95 % coverage interval for δm returned by the GUM uncertainty framework and the broken lines those from MCS.

It is recommended that if a comparison of the results from MCS and from the GUM uncertainty framework is favourable, the GUM uncertainty framework can validly be applied to the current problem and to problems sufficiently close to it. If the converse is true, another approach should be considered, MCS itself constituting a natural candidate.

Even if the use of the GUM uncertainty framework with higher-order terms is validated using MCS, it might be preferred to use MCS. A reason is that the partial derivatives required can be algebraically complex, even for the relatively simple mass calibration model of section 4. The possible difficulty associated with the maintenance of the resulting software is a consideration. Conversely, MCS is no more complicated, because only a means for forming model *values* need be provided.

A mass example illustrated some of the issues. It has a feature in common with other models involving relative corrections that the standard uncertainty $u(y)$ is underestimated by the GUM uncertainty framework based on the use of first order terms. Letting $\boldsymbol{X} = (\boldsymbol{X}_1^{\mathrm{T}}, \boldsymbol{X}_2^{\mathrm{T}}, \boldsymbol{X}_3^{\mathrm{T}})^{\mathrm{T}}$, such models can be expressed as $Y = f_1(\boldsymbol{X}_1)\{1 + f_2(\boldsymbol{X}_2)f_3(\boldsymbol{X}_3)\}$, where $f_2(\boldsymbol{X}_2)$ or $f_3(\boldsymbol{X}_3)$ or both take zero values at the estimates $\boldsymbol{x}$ of the values of $\boldsymbol{X}$. The resulting sensitivity coefficients associated with some of the uncertainties $u(x_i)$ are zero, as are the consequent contributions to $u(y)$.

This work constitutes part of the Software Support for Metrology (SS*f*M) programme within the UK's National Measurement System. This paper is based largely on current and previous editions of an SS*f*M best-practice guide [4] and software specifications [5] for uncertainty evaluation. The latter document aims to contain sufficient explicit information to permit implementation of MCS as outlined in this paper. The current paper also reflects material embodied within Supplement 1 to the GUM [3].

REFERENCES

[1] BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, and OIML. Guide to the Expression of Uncertainty in Measurement, International Standards Organization, 1995.

[2] M. G. Cox, M. P. Dainton, A. B. Forbes, P. M. Harris, P. M. Schwenke, B. R. L. Siebert, and W. Wöger. Use of Monte Carlo simulation for uncertainty evaluation in metrology. In P. Ciarlini, M. G. Cox, E. Filipe, F. Pavese, and D. Richter, editors, Advanced Mathematical Tools in Metrology V. Series on Advances in Mathematics for Applied Sciences, 57, 93–104, Singapore, 2001. World Scientific.

[3] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. Guide to the Expression of Uncertainty in Measurement. Supplement 1: Numerical Methods for the Propagation of Distributions. Technical report, Joint Committee for Guides in Metrology, 2004. Draft.

[4] M. G. Cox and P. M. Harris. SS*f*M Best Practice Guide No. 6. Uncertainty evaluation. Technical report, National Physical Laboratory, Teddington, UK, 2004. www.npl.co.uk/ssfm/download/documents/ssfmbpg6.pdf.

[5] M. G. Cox and P. M. Harris. Software specifications for uncertainty evaluation. Technical Report CMSC 40/04, National Physical Laboratory, Teddington, UK, 2004. www.npl.co.uk/ssfm/download/documents/cmsc40_04.pdf.