

DE LA RECHERCHE À L'INDUSTRIE

cea



www.cea.fr

www.nucleide.org

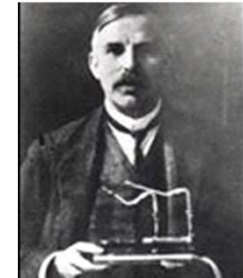
Evaluation of uncertainties, the classical approach (GUM) and their application in radionuclide metrology

Philippe Cassette

Laboratoire national Henri Becquerel, France



“If your experiment needs statistics, you ought to have done a better experiment “ **Ernest Rutherford (1871 – 1937)**



« *Life is only good for two things : developing mathematics and teaching mathematics* »





Siméon Denis Poisson, 1781-1840

Between these two extreme opinions, there is perhaps a way. If it is true that metrology is too serious to be left to statisticians, a reasonable use of statistical tools could help in doing good measurements...

The holy Bible: the GUM



- Creation of a working group on uncertainties (CIPM 1977)
- Questionnaire to National Metrology Institutes (February 1978)
- BIPM 80-3 report (1980)
- Recommendations INC-1 1980 « expression of experimental uncertainties »
- Approbation of INC-1 in 1981 by CIPM
- First publication of GUM in 1993
- Re-edition in 1995 with minor corrections
- Re-edition in 2008 with minor corrections (JCGM 100:2008)

↙	<i>Evaluation of measurement data – An introduction to the "Guide to the expression of uncertainty in measurement" and related documents</i>	JCGM 104:2009	
↙	<i>Evaluation of measurement data – Supplement 1 to the "Guide to the expression of uncertainty in measurement" – Propagation of distributions using a Monte Carlo method</i>	JCGM 101:2008	
↙	<i>Evaluation of measurement data – Supplement 2 to the "Guide to the expression of uncertainty in measurement" – Extension to any number of output quantities</i>	JCGM 102:2011	
↙	<i>Evaluation of measurement data – The role of measurement uncertainty in conformity assessment</i>	JCGM 106:2012	
↙	<i>Evaluation of measurement data – Concepts and basic principles</i>		

The uncertainty must be:

Unique : it quantifies the doubt on the measurement

Universal : similar for any kind of measurement

Internally consistent : deduced from the analysis of the measurement process

Transferable : can be used in the frame of another measurement

Measurand: quantity intended to be measured

Measurement: process of experimentally obtaining one or more quantity values that can be reasonably attributed to a quantity

Measurement results: set of quantity values being attributed to a measurand together with any other available relevant information

Measurement uncertainty: non-negative parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used

*The result of a measurement is a random variable...
... and must be treated accordingly*

Reminder : errors and uncertainties are not equivalent:

Measurement error: measured quantity value minus a reference quantity value

If a systematic error is suspected, it must be corrected... this does not suppress the doubt, thus the uncertainty on the result

UNCERTAINTY = DOUBT

The result of a measurement is a random variable (RV)

The value of the measurement is the expectation of this RV

Uncertainty is the doubt on the result... it is related to the dispersion of the RV, so to its variance

The main goal of uncertainty evaluation is the determination of the optimum estimator of the variance of the measurement (even if the expression of this uncertainty can depend on its practical use)

STEPS OF UNCERTAINTY EVALUATION

- Definition of the measurand (very important!)
- Identification of the parameters of influence (experimental data, correction coefficients...)
- Relation between the input parameters and the result (transfer function)
- Evaluation of the variance of each input parameter, and the covariances between these parameters
- Propagation of variances and covariances to obtain the combined standard uncertainty
- If necessary, expand the uncertainty
- Express the measurement result and its associated uncertainty

“A measurement without uncertainty is not a measurement” (me)

DEFINITION OF THE MEASURAND

- Before measuring, define precisely and exhaustively the measurand (this is not a trivial task!)
- Include in the definition all the parameters which could have a significant influence on the results
- Think about the possible use of the measurement results, to avoid any ambiguity on the definition of the measurand

Experimental data:

- Count rates, masses, volumes, peak area,...

Physical parameters of influence:

- Temperature, pressure, humidity, density...

Correction factors:

- Calibration coefficients, detection efficiency, coincidence summing...

Standards:

- Of activity, masses, time, frequency...

Theoretical data:

- Emission intensities, equilibrium factors, decay scheme data...

TRANSFER FUNCTION

$$y = f(x_1, x_2, \dots, x_n)$$

y is the measurement result and the x_i are the input quantities (experimental data, coefficients, decay data...)

Each input quantity is considered as a random variable

The standard deviation of each input quantity can be evaluated by 2 methods:

Type A evaluation method: repetition of the measurement and calculation of the experimental standard deviation of the mean of a quantity

Type B evaluation method: evaluation of the standard deviation of the input quantity by scientific judgment based on all the available information on the possible variability of this quantity. This information can include:

- previous measurement data
- experience with or general knowledge of the behavior and properties of relevant materials and instruments
- manufacturer's specifications
- data provided in calibration and other certificates
- uncertainties assigned to reference data taken from handbooks

- Independent repeated measurements
- Calculation of the arithmetic mean (if similar measurements)
- Calculation of the experimental variance and standard deviation
- Calculation of the standard deviation of the mean

If the measurement is repeated under the same conditions (i.e. if the measurement concerns the same measurand)... the standard deviation of the arithmetic mean is a good estimator of the standard uncertainty of the input quantity

From n samples of the random variable $X : x_1, x_2, x_3, x_4, \dots, x_n$

Expectation

$$E(x) = \sum_{i=1}^n x_i P(x = x_i)$$

If all the samples have the same probability (weight):

$$P(x_i) = \frac{1}{n}$$

$$E(x) = \sum_{i=1}^n x_i \frac{1}{n} = \frac{1}{n} \sum_{i=1}^n x_i = \text{arithmetic mean}$$

Is the best estimate of the input quantity

If the samples do not have the same weight, (e.g. different uncertainties) a weighted mean is calculated, the unbiased estimator is a mean, weighted by the inverse of the variances:

$$P(x_i) \propto \frac{1}{u_{x_i}^2}$$

Weighted mean:

$$M = \frac{1}{\sum_{i=1}^n \frac{1}{u_{x_i}^2}} \sum_{i=1}^n \frac{x_i}{u_{x_i}^2}$$

n samples: $x_1, x_2, x_3, x_4, \dots, x_n$

$$u_x^2(\mathbf{x}) = E[(\mathbf{x} - E(\mathbf{x}))^2]$$

$$u_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - M)^2}$$

Describes the variability of the RV

Variance = u^2

Standard deviation = $\sqrt{\text{variance}} = u$

STANDARD DEVIATION OF THE MEAN

$$u_{\bar{x}} = \frac{s}{\sqrt{n}}$$

$$u_{\bar{x}} = \sqrt{\frac{1}{n(n-1)} \sum_{i=1}^n (x_i - M)^2}$$

Describes the variability of the estimated mean

$$u_M = \sqrt{\frac{p(x_i) \cdot (x_i - M)^2}{(n-1) \sum_{i=1}^n P(x_i)}} \quad \text{with} \quad P(x_i) = \frac{1}{u_{x_i}^2}$$

u_M^2 Is also called interclass variance

The mean of the variances (intraclass variance) is:

$$u_M^{2'} = \sqrt{\frac{1}{\sum_{i=1}^n p(x_i)}}$$

The experimental variance is the sum of interclass and intraclass variances

The experimental standard deviation gives the dispersion of the measurements. *When the number of measurements increases, it does not vary significantly, but is more precisely known...*

The standard deviation of the mean gives the dispersion of the estimated mean.

When the number of measurements increases, it decreases (as the inverse of the square root of the number of measurements)

To derive the value of the standard deviation, it is necessary to assume a probability density function (pdf) of the random variable

Examples : if the value is known to be in the interval $[M-a, M+a]$

- Uniform distribution:
$$u = \frac{a}{\sqrt{6}}$$

- Triangular distribution:
$$u = \frac{a}{\sqrt{3}}$$

- Normal distribution $\pm 3 \sigma$:
$$u = \frac{a}{\sqrt{9}}$$

Guesstimation is a type B evaluation method...

- Experimental covariance (type A evaluation method)

$$u_{x,y} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})(y_i - \bar{Y})}$$

- Otherwise, estimated covariance (type B evaluation method)

Examples :

- Correlated parameters: P_K et ω_K
- Coincidences, correlated data
- I_β and IC, emission intensities in γ spectrometry
- Efficiency curve obtained using multi- γ radionuclides

Hypothesis:

- the transfer function is continuous, derivable and can be approximated by a Taylor series
- First order approximation
- Low fluctuations of each input parameter

$$y = f(x_1, x_2, \dots, x_n)$$

$$u_c^2(y) = \sum_{i=1}^n \left[\frac{\partial f}{\partial x_i} \right]^2 \cdot u^2(x_i) + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \frac{\partial f}{\partial x_i} \cdot \frac{\partial f}{\partial x_j} \cdot u(x_i, x_j)$$

← Sensitivity coefficient

If the input parameters are not correlated:

$$u_c(y) = \sqrt{\sum_{i=1}^n \left[\frac{\partial f}{\partial x_i} \right]^2 \cdot u^2(x_i)}$$

The combined uncertainty $u_c(y)$ quantifies the dispersion of the measurand y

EXPANDED UNCERTAINTY

Although the standard uncertainty can be universally used to express the uncertainty of a measurement result, in some commercial, industrial, and regulatory applications, and when health and safety are concerned, it is often necessary to give a measure of uncertainty that defines an interval about the measurement result that may be expected to encompass a large fraction of the distribution of values that could reasonably be attributed to the measurand.

$$U = k u_c(y)$$

The value of k is usually between 2 and 3, depending on the confidence level corresponding to this interval

Expression of the uncertainty

Uncertainty can be expressed as:

- standard uncertainty (square root of the variance)
- or k times the standard uncertainty (k must be stated)
- or by the determination of a confidence interval, with associated probability, after assuming the statistical distribution of the results, and considering the effective degrees of freedom

The GUM is a consensual document, and no strict rule is given on the way to express the uncertainties, even if the Gaussian distribution of the results is sometimes implicit and allows a simple calculation of confidence intervals

From a statistical point of view, the GUM mixes frequential and Bayesian definitions of probabilities. The GUM supplement 1 is based on a Bayesian approach.

Nevertheless, even if “pure statisticians” can be frustrated, the GUM approach works and gives a sound way to express uncertainties

- For international comparisons, the uncertainty is generally reported as a standard uncertainty (CIPM recommendation)
- For calibration certificates, the accreditation bodies generally impose expanded uncertainties ($k=2$)... even if this does not give any supplementary information versus the standard uncertainty*
- Thus, in practice, the determination of the standard uncertainty is the main goal of uncertainty evaluation, and we will see that GUM is perfectly suitable for that

** Supplementary information (if any) is thus implicit, preventing the necessity of justification. Nevertheless, one can consider that this approach is not a sign of intellectual probity...*

If one wants to derive the confidence interval from the standard uncertainty, it is necessary to assume a statistical distribution of the results

Generally, this distribution is unknown... but the central limit theorem can be evoked as an argument for using a Gaussian distribution!

Warning: the statistical distribution of the mean of a RV could differ from the distribution of the RV:

Examples:

- Poisson distribution: the distribution of the evaluated mean is Gaussian
- Gaussian distribution: the distribution of the evaluated mean is a Student distribution with ν degrees of freedom

Type A evaluation method: if the mean and the standard deviation are evaluated from n measurements:

$$\nu = n - 1$$

Type B evaluation method : no general rule but

If the RV is Gaussian:

$$\nu_i \approx \frac{1}{2} \frac{u^2(x_i)}{\sigma^2(u(x_i))} \approx \frac{1}{2} \left[\frac{\Delta u(x_i)}{u(x_i)} \right]^{-2}$$



Relative uncertainty of the uncertainty (sometimes subjective...)

Welch-Satterthwaite formula:

$$v_{eff} \approx \frac{u_c^4(y)}{\sum_{i=1}^N \frac{u_i^4(y)}{v_i}} \quad \text{With} \quad v_{eff} \leq \sum_{i=1}^N v_i$$

UNCERTAINTY OF THE UNCERTAINTY

*The evaluated standard deviation (i.e. the uncertainty) from repeated measurements is a RV, so has an uncertainty
For a Gaussian distribution, this relative uncertainty can be calculated:*

Number of observations	Relative uncertainty (%)
2	76
3	52
4	42
5	36
10	24
20	16
30	13
50	10

- Uncertainty must be rounded
- Bear in mind that the uncertainty of the uncertainty also concerns all its applications, e.g. :
 - the boundaries of the confidence intervals,
 - the decision thresholds and detection limits...

Exercise 1: what are the boundaries of a confidence interval with a probability of 95,5 % of a Gaussian RV derived from the measurement of 10 samples ?

Comment: what is the point of giving a Student factor with 3 significant digits? (the one who never saw a 1.96 factor in the calculation of detection limits can blame me!)

Correct expressions

$A = 104.7 \text{ Bq/g}$ with a combined standard uncertainty of 1.5 Bq/g

$A = 104.7(15) \text{ Bq/g}$

$A = (104.7 \pm 3.0) \text{ Bq/g} (k=2)$

If an expanded uncertainty is given, it is mandatory to state the value of k

Not so correct expressions

$A = (104.7 \pm 1.5) \text{ Bq/g} (k=1)$

$A = (104.73 \pm 3.02) \text{ Bq/g} (k=2)$

Although the GUM provides a framework for assessing uncertainty, it cannot substitute for critical thinking, intellectual honesty and professional skill.

The evaluation of uncertainty is neither a routine task nor a purely mathematical one; it depends on detailed knowledge of the nature of the measurand and of the measurement.

The quality and utility of the uncertainty quoted for the result of a measurement therefore ultimately depend on the understanding, critical analysis, and integrity of those who contribute to the assignment of its value.

- Gaussian hypothesis is not needed to use type A evaluation methods (the estimators of the mean and of the variance are unbiased and of minimum variance, even if the distribution is unknown)
- Gaussian hypothesis is not imposed for type B evaluation methods
- Gaussian hypothesis is not necessary for the propagation of variances
- Gaussian hypothesis is not required if the uncertainty is expressed as a standard or expanded uncertainty... but in this case the expanded uncertainty does not give more information than the standard uncertainty

Activity determination of the activity of a solution

n sources are prepared

Each source is measured m times

Questions:

Why n sources, how to choose n ?

Why m measurements of each source, how to choose m ?

What are the uncertainties in each case ?

WHY n SOURCES ?

To answer the following questions:

1. When an activity measurement is made from a sampling process, is the sampling representative of the measurand?
2. Is there a hidden variability in the source fabrication?

Several samples giving several sources. It can be expected that any sampling bias will be cancelled out when using the mean, for a large number of samples

Generally this is not obvious, but if there is a variability in the sampling process, there is little chance that the samples are similar. Thus in this case an extra variance should be observed

Practical reasons: solution inhomogeneity or source instability

Example : evaporation during the weighing process, droplet losses during the pycnometer manipulation

This can be checked through the preparation and measurement of several sources, If there is a variability in the source preparation process, an extra variance of the results should be observed

How?

Compare the experimental variance of the counting rate of different sources with the counting rate of one source for the same counting time and taking into account the weighing uncertainty

If the results are consistent (within say 30%), no extra variability

If the results are not consistent, look for possible variability

$n=1$: for lazy (and happy) people 😊

$n=2$: in case of significant differences, how to conclude?

$n=3$: one can vote, 2 against 1, but difficult to justify...

$n=5$: minimum sample for sound comparison

$n=10$ or more: statistical tools can be used...

But of course, there are other constraints (available time, manpower, cost, etc.)

WHY m repetitions?

It is better to repeat m times a measurement of duration t , instead of one measurement of duration mt

The statistical uncertainty is the same but this gives the possibility to calculate the experimental standard deviation (and the experimental covariances)

This provides a good way to check that there is no extra variability in the measurement process and allow the detection of:

- Non-steady noise (e.g. parasitic pulses)
- Instrument shift (e.g. from variation of ambient conditions)
- Allow the use of type A evaluation method of uncertainties

Comment: if x pulses are recorded:

- considering x as the variance supposes a hypothesis on the *pdf* (Poisson) and is a type B evaluation method
- Repeating the measurement and using the experimental variance is a type A evaluation method

EXPERIMENTAL STANDARD DEVIATION OR STANDARD DEVIATION OF THE MEAN?

If the measurements are made under strictly similar conditions (i.e. if the measurand is the same and observed fluctuations are under statistical control), the standard deviation of the mean must be used

If there is a hidden variability (from source preparation or from instrumental instability), and if this variability cannot be fixed, it is safer to consider the experimental standard deviation

In each case it is pertinent to compare the observed standard deviation with the expected uncertainty (i.e. considering all the known sources of uncertainty: weighing, counting statistics...)

WHAT IS MORE PRECISE, TYPE A OR TYPE B EVALUATION METHODS?

This question is only pertinent if the choice is possible (e.g. for a counting process)

Type A evaluation method:

- No hypothesis needed, but not very accurate if the number of repetitions is small (remember about 25% uncertainty of the uncertainty for 10 repetitions in case of Gaussian distribution)
- Gives useful experimental information on the variability of the measurement process

Type B evaluation method:

- Could be accurate if there is a good justification on the pdf to use.
- For counting processes, Poisson distribution is generally acceptable. In the presence of dead-time, the variance is lower than the Poisson variance

If you are Bayesian, type B is the only way!

PRACTICAL EXAMPLE: EFFICIENCY CURVE OF A GAMMA SPECTROMETER

N°	Energie [keV]	Efficacité [%] (mesuré)	Efficacité [%] (Courbe)	Delta [%]	Sigma [%] (mesuré)	Sigma [%] (Courbe)
4	59.54	0.6125	0.5923	-3.3	2.02	0.35
5	59.54	0.5977	0.5923	-0.9	2.27	0.35
6	59.54	0.5909	0.5923	0.2	1.85	0.35
7	88.03	1.9924	1.9197	-3.6	2.40	0.61
8	88.03	1.9618	1.9197	-2.1	2.29	0.61
9	88.03	1.9535	1.9197	-1.7	2.46	0.61
10	122.06	2.6336	2.6372	0.1	1.22	0.76
11	122.06	2.6159	2.6372	0.8	0.55	0.76
12	122.06	2.6085	2.6372	1.1	1.28	0.76
13	122.06	2.5840	2.6372	2.1	1.15	0.76
14	136.47	2.7143	2.7117	-0.1	1.40	0.79
15	136.47	2.6445	2.7117	2.5	4.48	0.79
16	136.47	2.6193	2.7117	3.5	4.87	0.79
17	136.47	2.6155	2.7117	3.7	2.87	0.79
18	165.86	2.7529	2.6754	-2.8	1.28	0.80
19	165.86	2.7268	2.6754	-1.9	1.26	0.80
20	165.86	2.7145	2.6754	-1.4	1.19	0.80
21	165.86	2.6940	2.6754	-0.7	0.91	0.80
22	320.08	1.9195	1.9222	0.1	1.42	0.65
23	320.08	1.9174	1.9222	0.2	1.47	0.65
24	320.08	1.9098	1.9222	0.6	1.58	0.65
25	391.70	1.6819	1.6809	-0.1	1.02	0.59
26	391.70	1.6750	1.6809	0.3	1.02	0.59
27	391.70	1.6672	1.6809	0.8	1.01	0.59

In this calibration, 4 measurements are considered at 136.47 keV. Is it useful? Is it better to consider only a mean value with the associated standard deviation?

The uncertainty given by InterWinner is lower than the residuals of the fit. This gives the impression that this uncertainty is underestimated


This will be discussed later (fitting session)

Various documents: see <https://www.bipm.org/en/publications/guides/gum.html>





GUM: Guide to the Expression of Uncertainty in Measurement

 In order to benefit fully from the hyperlinking between the documents, the reader is advised to download all JCGM documents presently available in one **ZIP file**.

→ The fundamental reference document is the *Guide to the Expression of Uncertainty in Measurement* (GUM):

<p>↙ Evaluation of measurement data – Guide to the expression of uncertainty in measurement</p> <p style="text-align: right;">JCGM 100:2008 (GUM 1995 with minor corrections)</p> <p>Note: JCGM 100:2008 is also available in HTML form from the JCGM portal on ISO's website.</p>	
---	---

→ The JCGM Working Group 1 (**JCGM-WG1**) is producing a series of documents to accompany the GUM. The first four of these documents have been approved and are available for download as PDF files. Printed copies are available for purchase from ISO.

<p>↙ Evaluation of measurement data – An introduction to the "Guide to the expression of uncertainty in measurement" and related documents</p> <p style="text-align: right;">JCGM 104:2009</p>	
<p>↙ Evaluation of measurement data – Supplement 1 to the "Guide to the expression of uncertainty in measurement" – Propagation of distributions using a Monte Carlo method</p> <p style="text-align: right;">JCGM 101:2008</p>	
<p>↙ Evaluation of measurement data – Supplement 2 to the "Guide to the expression of uncertainty in measurement" – Extension to any number of output quantities</p> <p style="text-align: right;">JCGM 102:2011</p>	
<p>↙ Evaluation of measurement data – The role of measurement uncertainty in conformity assessment</p> <p style="text-align: right;">JCGM 106:2012</p>	
<p>↙ Evaluation of measurement data – Concepts and basic principles</p>	

The following documents are at an early stage of preparation:

<p>↙ Evaluation of measurement data – Supplement 3 to the "Guide to the expression of uncertainty in measurement" – Modelling</p>	
<p>↙ Evaluation of measurement data – Applications of the least-squares method</p>	

Related articles

- GUM:**
- **BIPM Workshop on Measurement Uncertainty**
 - Software related to the GUM and the GUM supplements 1 and 2
 - Tutorial for metrologists on the probabilistic and statistical apparatus underlying the GUM and related documents
 - Bibliography on Uncertainty
 - News from JCGM-WG1
 - JCGM Working Group 1
- VIM:**
- "Annotated VIM3"
 - The rationale for VIM3
 - FAQs on the VIM3
 - News from JCGM-WG2
 - JCGM Working Group 2

In 2015-2016, the JCGM attempted to revise the GUM:

Goal: make it more internally consistent, on a Bayesian frame

Consultation of the members: (BIPM, CEI, IFCC, ILAC, ISO, UICPA, UIPPA et OIML)

Consultation of the users and the NMIs



Major rejection of the new text

Actual situation: status quo... but a new revision is under way

We hope that the JCGM learned the lessons of the first revision process

Thank you for your attention