EDITORIAL

EURACHEM/CITAC workshop on recent developments in measurement uncertainty

Alex Williams

This was the topic of a workshop held in Lisbon in June 2011, which was organised by the Faculty of Sciences of the University of Lisbon, EURACHEM-Portugal and RELACRE, the Portuguese Association of Accredited Laboratories, on behalf of the EURACHEM/CITAC Measurement Uncertainty and Traceability Working Group. This topical issue of ACQUAL contains a selection of the contributed papers at this workshop. These papers and the posters presented at the workshop show how the evaluation of uncertainty is now being applied to a wide range of analyses. Also, it is interesting to see in how many cases the evaluation used method performance data and measurements on reference materials. There were also a number of invited speakers who reported on recent developments and these together with some of the items arising from the discussion are summarised below.

Measurement results near zero

The evaluation of measurement uncertainty for results close to zero is a problem that has been of concern for some time. It was not considered in the ISO Guide to the Expression of Uncertainty in Measurement (GUM) [1]. When the results are close to zero, the uncertainty interval calculated according to the procedures given in GUM could include values below zero, even when the measurand is, for example, a concentration which by definition cannot take a negative value. *Stephen L R Ellison* in his presentation giving an overview of the revised EURACHEM/CITAC guide "Quantifying Uncertainty in Analytical Measurement" [2] pointed out that the guide now has a much larger section on how uncertainty can be evaluated and reported in this case. Two procedures are described one utilising classical statistics [3] and the other utilising Bayes theorem [4].

The procedure for calculating the coverage interval using classical statistics is quite simple. If the expanded uncertainty has been calculated to have for example a 95 % coverage and would have extended below zero, then it is just truncated at zero and this truncated classical confidence interval maintains exact 95 % coverage. This truncated interval becomes progressively more asymmetric as the result approaches zero. However, the observed mean is the best estimate of the value of the measurand until the observed mean falls below zero, when the value of zero should be used. Also, as the observed mean value falls further below zero, the simple truncated interval becomes unreasonably small, but results in this region may indicate that something is wrong with the measurement.

The Bayesian method allows the combination of information from the measurements with the information that the value of the concentration cannot be negative. For measurement results that can be described in the form of a t-distribution, then as shown in [4], the resulting probability distribution of the values attributable to the measurand is approximately a truncated *t*-distribution. The observed mean or zero if the observed mean value is below zero should again be used as the reported value, and the expanded uncertainty interval is calculated as the maximum density interval containing the required fraction of this truncated distribution. How to calculate this interval is given in [2]. This Bayesian interval provides the same minimal bias as the classical approach, with the useful property that as the observed mean value falls further below zero, the reported uncertainty increases. Willink [5]

A. Williams (✉)
Camberley, Surrey, UK
e-mail: aw@camberley.demon.co.uk

pointed out that simulations showed that some Bayesian 95% intervals did not always contain the true value 95 % of the time. The Analytical Method Committee, RSC [6] in responding to this showed that for the maximum density interval, while the coverage is not exactly 95 % for each individual value of the true mean, the deviations from 95 % are unlikely to be serious in practice.

Monte Carlo simulation

The revised EURACHEM/CITAC guide [2] has a new section on The Monte Carlo Simulation (MCS), and this was covered in presentations by *Stephen L R Ellison* and *Matthias Rösslein*. The MCS method described in GUM Supplement 1 [7] is simple in principle and easy to use given appropriate software. MCS requires the probability distributions (called the PDFs or probability density functions) for all the input quantities in the measurement equation used to calculate the result. MCS calculates the result corresponding to one value of each input quantity drawn at random from its PDF and repeats this calculation a large number of times (trials), typically 10^5 to 10^6. This process produces a set of simulated results which, under certain assumptions, forms an approximation to the PDF for the value of the measurand. From this set of simulated results, the mean value and standard deviation are calculated. In GUM Supplement 1, these are used, respectively, as the estimate of the measurand and the standard uncertainty associated with this estimate.

In most cases, using the first-order Taylor series approximation given in GUM and the MCS method will give virtually the same value for the standard uncertainty associated with the estimate of the measurand. Differences become apparent when distributions are far from Normal or where the measurement result depends nonlinearly on one or more input quantities, and the uncertainty on these input quantities is large. When the measurement model is nonlinear and, the relative standard uncertainty is larger than 10% the MCS PDF is likely to be asymmetric. In this case, the mean value computed from the simulated results will be different from the value of the measurand calculated using the estimates of the input quantities (as in GUM). However, in most cases, the difference is likely to be less than the standard uncertainty. For most practical purposes, in chemical measurement, the result calculated from the original input values should be reported; the MCS estimate of standard uncertainty can, however, be used.

MCS gives a good estimate of the standard uncertainty even for simulations with a few hundred trials, and simulations of 500–5000 MCS samples are likely to be adequate for evaluating the standard uncertainty. It is also possible to use the MCS result to determine confidence intervals without the use of effective degrees of freedom. In this case, to obtain sufficient information about the tails of the PDF for the output quantity can require calculating the result for at least 10^6 trials. However, it is important not to be misled by the apparent detail in the PDF obtained for the result. The lack of detailed knowledge about the PDFs for the input quantities, because the information on which these PDFs are based is not always reliable, needs to be borne in mind. The tails of the PDFs are particularly sensitive to such information. Therefore, as is pointed out in GUM, section G 1.2, "it is normally unwise to try to distinguish between closely similar levels of confidence [say a 94 % and a 96 % level of confidence]". In addition, the GUM indicates that obtaining intervals with levels of confidence of 99 % or greater is especially difficult. Further, to obtain sufficient information about the tails of the PDF for the output quantity can require calculating the result for at least 10^6 trials.

Compliance assessment

An important application of measurement uncertainty is its use in the assessment of compliance. When making an assessment of compliance, the presence of unavoidable measurement uncertainty introduces the risk of making incorrect decisions, i.e. of accepting a batch of material that is outside the specification or rejecting one that is within. The probability of making a wrong decision depends both upon the size of the measurement uncertainty and on how the uncertainty is taken into account when assessing compliance.

Because of its importance, a section has been retained in the revised guide [2], although the EURACHEM/CITAC guide [8] on how to take uncertainty into account when assessing compliance was published in 2007. Also, JCGM Working Group 1 recently circulated a final draft of a document for comment that gives similar guidance but with an additional section on the use of prior knowledge of the distribution of the parameter being measured on the product being assessed.

In his presentation, *Bertil Magnusson* pointed out that the key to the assessment of compliance is the concept of decision rules. These rules give a prescription for the acceptance or rejection of a test result for a sample based on the measurement result, its uncertainty and the limits, taking into account the acceptable level of the probability of making a wrong decision. A simple decision rule is that if a result exceeds the upper limit by more than the expanded uncertainty, then this shows noncompliance. This rule gives a low probability of false rejection. However, the larger the uncertainty the larger the value of the result that will be accepted as showing compliance and the possibility of such a large value of the measurand may not be acceptable. It is, therefore, necessary also to set a limit on

the value uncertainty but guidance on how to do this has yet to be developed.

Traceability

Traceability is formally defined in the International Vocabulary of Metrology (VIM) [9] as:

> property of a measurement result whereby the result can be related to a reference through a documented unbroken chain of calibrations, each contributing to the measurement uncertainty.

The reference to uncertainty arises because the agreement between laboratories is limited, in part, by uncertainties incurred in each laboratory's traceability chain. Traceability is accordingly intimately linked to uncertainty. In his presentation, *Ricardo J. N. Bettencourt da Silva* pointed out that traceability provides the means of placing all related measurements on a consistent measurement scale, while uncertainty characterises the 'strength' of the links in the chain and the agreement to be expected between laboratories making similar measurements.

In general, the uncertainty on a result, which is traceable to a particular reference, will be the uncertainty on that reference together with the uncertainty on making the measurement relative to that reference.

How to establish traceability is given in the EURACHEM/CITAC guide Traceability in Chemical Measurement [10], and this guidance has been widely accepted. However, the latest version of VIM [9] includes the possibility that the reference can be a measurement procedure, which could be taken to mean that an "empirical method" could be such measurement procedure, i.e. the result obtained using a measurement procedure is traceable to that procedure itself. In [10] the empirical method was seen as the definition of the measurand and the traceability was established by establishing the traceability for all the measurements set out in the measurement procedure. This dichotomy led to a heated discussion at the workshop, and the issue has yet to be resolved.

Revision of GUM

Work has started on the revision of GUM, and although at the time of the workshop, there had been no publications from the JCGM Working Group 1, since then the Working Group has recently produced a document "Motivation and Scope for the Revision of the GUM" and survey requesting views on issues to be considered during the revision, http://www.metrologyinfo.org As was pointed by *Adriaan M.H. van der Veen* in his presentation, at the moment, no fundamental change to the basic principles of the current edition is envisaged. One item which is being considered is

improving the consistency of the approaches used in Type A and Type B evaluation, where currently Type A evaluation is based on classical frequentist statistics and Type B on Bayesian degree of belief. The possibility of writing one version using classical statistics and one using Bayesian is also being discussed. However, the main concern of the analytical community is that more emphasis is given to the use of method performance data in evaluating uncertainty and that preferably it should be treated in a separate section. This was emphasised by many speakers during the discussion.

The survey by the WG seeking views gives the analytical community the opportunity to emphasise the importance of giving more emphasis to the use of method performance data and I strongly urge you to partake in this survey.

This was a very successful workshop with very good discussions being held around the posters and following both the contributed and invited lectures. It also achieved its other aim of obtaining comments on the latest draft of the third edition EURACHEM/CITAC Guide on measurement uncertainty [2].

Alex Williams
Chairman, EURACHEM/CITAC WG on Measurement Uncertainty and Traceability

References

1. Guide to the Expression of Uncertainty in Measurement. ISO, Geneva (1993) (ISBN 92-67-10188-9) (Reprinted 1995: Reissued as ISO Guide 98-3 (2008), also available from http://www.bipm.org/en/publications/guides/gum.html as JCGM 100:2008)
2. EURACHEM/CITAC guide quantifying uncertainty in analytical measurement, Third Edition. Available shortly from http://www.eurachem.org
3. Cowen S, Ellison SLR (2006) Analyst 131:710–717
4. Analytical Methods Committee, Royal Society of Chemistry (2008) Measurement uncertainty evaluation for a non-negative measurand. An alternative to limit of detection. Accred Qual Assur 13:29–32
5. Willink R (2010) Uncertainty in repeated measurement of a small non-negative quantity: explanation and discussion of Bayesian methodology. Accred Qual Assur 15:181–188
6. Analytical Method Committee, The Royal Society of Chemistry (2010) Response from the Analytical Methods Committee (AMC) to the paper "Uncertainty in repeated measurement of a small non-negative quantity: explanation and discussion of Bayesian methodology". Accred Qual Assur 15:189–191
7. Evaluation of measurement data—Supplement 1 to the "Guide to the expression of uncertainty in measurement"—Propagation of distributions using a Monte Carlo method, JCGM 101:2008, http://www.bipm.org/en/publications/guides/gum.html
8. EURACHEM/CITAC Guide Use of uncertainty information in compliance assessment (2007). Available from http://www.eurachem.org
9. ISO Guide 99:2007, International vocabulary of metrology-Basic and general concepts and associated terms (VIM) (2007) ISO, Geneva. Also available from http://www.bipm.org/vim as JGCM 200:2012
10. EURACHEM/CITAC Guide (2003) Traceability in chemical measurement. Available from http://www.eurachem.org