Measurement 43 (2010) 392–399

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

Can coverage factor 2 be interpreted as an equivalent to 95% coverage level in uncertainty estimation? Two case studies

Martin Vilbaste [a], Georgi Slavin [b], Olev Saks [a], Viljar Pihl [a], Ivo Leito [a,*]

[a] Institute of Chemistry, University of Tartu, Jakobi 2, 51014 Tartu, Estonia
[b] Asper Biotech, Oru 3, 51014 Tartu, Estonia

ARTICLE INFO

Article history:
Received 9 April 2009
Received in revised form 4 September 2009
Accepted 7 December 2009
Available online 4 January 2010

Keywords:
Monte Carlo simulation
Coverage interval
Effective degrees of freedom
ISO GUM

ABSTRACT

The GUM modelling, its Bayesian modification and the Monte Carlo method (MCM) to estimate the uncertainty are compared in two practical measurement situations (finding reference value of relative humidity and a generic chemical instrumental analysis procedure). The results of the three approaches agree very well when there are no dominant input quantities with type A evaluated uncertainty estimated from small number of repeated measurements. In the opposite case the GUM gives underestimated expanded uncertainties (by up to 20–25%), compared to both other approaches. Analysis of the practical measurement situations reveals that even in the case of several dominating input quantities of similar uncertainty contributions, if one of them is distributed according to the *t*-distribution and has a low number (3–4) of degrees of freedom, the output quantity cannot be safely assumed Normally distributed and in such a case coverage factor 2 is not an equivalent to 95% coverage level.

1. Introduction

Indirect measurements, where the output quantity value is calculated from one or more input quantity values (parameters measured directly, such as temperatures, liquid volumes, masses and others) [1] are very common in physics and chemistry. Besides calculating the best estimate of the output quantity it is also very important to characterize its uncertainty. The traditional way of finding the uncertainty of an output quantity is via the GUM uncertainty framework [1] whereby the uncertainties of the input quantities are converted to standard uncertainties u and are combined using the measurement model (measurement equation) to give the combined standard uncertainty of the output quantity u_c. Because the coverage level of u_c is too low for many applications, uncertainty is usually reported as expanded uncertainty U, obtained via multiplying u_c by a suitable coverage factor k.

In many cases this approach works well. However, it has certain disadvantages. The most important of these is that, except in the simplest cases, the probability density function (PDF) of the output quantity remains unknown (here and below we assume that the output quantity, as well as all the input quantities have probability density functions as is done in Ref. [2]). Therefore, rigorous calculation of U corresponding to a specified coverage level is not possible and some assumptions or simplifications have to be introduced. Two cases are most widespread:

1. The most common assumption is that the input quantities are independent and that their combination leads to a Normal distribution of the output quantity. This has led to the "*de facto* standard practice" of presenting measurement uncertainty as $k = 2$ expanded uncertainty and implicitly interpreting it (based on the properties of Normal distribution) as corresponding to roughly 95% coverage level. It is true that in many cases this approach is fully justified.

* Corresponding author. Tel.: +372 5 184 176; fax: +372 7 375 264.
E-mail address: ivo.leito@ut.ee (I. Leito).

0263-2241/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.measurement.2009.12.007

2. The most frequent exceptions from the assumptions of the case 1 are believed to be those where there is a non-normally distributed dominating input quantity, in particular, one that is evaluated as the mean of a limited number of individual measurement results. In such a case one of the most frequent approaches is to assume that the PDF of the output quantity can be modelled by a shifted and scaled *t*-distribution. The effective number of degrees of freedom *df*, necessary to characterize this PDF is usually estimated from the Welch–Satterthwaite (WS) approach [1] or some of its modifications [3–5]. Even though the WS approach has serious limitations that have been repeatedly pointed out [6–8] it is certainly useful if the measurement model is linear and output quantity is distributed Student like [9]. In particular, the European Cooperation for Accreditation (EA) have united both these cases into their uncertainty guide [10] and there has been a detailed study arriving at a conclusion that the WS approach performs very well [11].

Standard deviations evaluated from a limited number of repeated measurements (measurements carried out under the same conditions) are very common in e.g. chemical analyses/measurements, where repetitions are often costly. The chemical analysis example that we give below is a typical one: the repeated measurements that are made are not just instrumental measurements (measurement with a certain analytical instrument, such as spectrophotometer, and gas chromatograph). Every such repeated measurement needs the whole sample preparation sequence that involves adding reagents, waiting for reaction to complete, etc. It is well possible in a scientific laboratory to make many repeated measurements also in such a case, but this is not possible at a routine analysis laboratory that does analysis work as business. Thus, limited number of repeated measurements is rather common at routine analysis laboratories.

The main problem is that it can be very difficult, especially at routine laboratory level, to recognize the situations when the limitations of the WS approach apply. The second disadvantage of the GUM approach is that it is sometimes (especially in the case of strongly non-linear models) difficult to calculate the sensitivity coefficients of input quantities.

Recognizing these problems of the GUM approach the Joint Committee for Guides in Metrology (JCGM) and its member organizations have prepared Supplement 1 to the "Guide to the expression of uncertainty in measurement" – Propagation of distributions using a Monte Carlo method (MCM) [2]. MCM has several advantages compared to the GUM uncertainty framework [4,5]. It enables to find a numerical approximation of the PDF of the output quantity, thus giving significantly richer information about the output quantity than the GUM approach. With the output quantity PDF available it is easy to calculate the mean, standard deviation, and to find the coverage interval with a preset coverage level (which allows also to find the coverage factor). This method also saves researchers from calculating sensitivity coefficients of input quantities.

Besides the GUM uncertainty framework and MCM other approaches have been proposed for uncertainty evaluation. According to [2] every input quantity has an individual PDF and so does the output quantity. This is not the only possible assumption. According to the Bayesian statistics the measurement results are constants and the value of the measurand is a random variable [9,12]. If the number of degrees of freedom is finite the uncertainty is believed to be uncertain according to GUM uncertainty framework [1,9,12,13]. According to the Bayesian approach the uncertainty does not have statistical uncertainty [9] and it is more appropriate to use in cases if one or more input quantities have low number of degrees of freedom.

Interestingly, the classical uncertainties can be used in the Bayesian framework if they are treated from the Bayesian point of view [12]. The relationship between the classical type A evaluated uncertainty $u_A(x)$ and Bayesian uncertainty $u_{A,Bayes}(x)$ is given by the following formula [12]:

$$u_{A,Bayes}(x) = \sqrt{\frac{(n-1)}{(n-3)}} \cdot u_A(x) \tag{1}$$

where n is the number of repeated measurements. This equation allows to "correct" for a low number of degrees of freedom and to express the standard uncertainty as $u_{A,Bayes}$ corresponding to approximately Normal distribution: the uncertainties $u_{A,Bayes}$ when combined with each other and with B-type uncertainties (the latter are used unchanged), will yield a combined standard uncertainty, which assumes approximate Normal distribution of the output quantity. In this case there is no need to estimate the number of effective degrees of freedom when calculating the expanded uncertainty, because the output quantity can be assumed to be approximately Normally distributed [9]. Below we term this approach as the Bayesian modification of the GUM. If the requirements for using the WS approach are not fulfilled (or it is difficult to find out) this approach is very useful.

In order to be accepted and used confidently, every new approach strongly benefits from application examples. Several application examples on the MCM approach have been included and are analyzed in the Supplement 1 [2]. There have been reports of using MCM for uncertainty calculation [4,5,14–18] many of them emphasizing its advantages over the GUM uncertainty framework. There has, however also been a report on an opposite finding [19]. Examples on the application of the Bayesian approach about different measurement fields have also been published [20–22].

In this paper we present the analysis of two different measurement cases, both from physics and chemistry. These are analyzed in parallel in terms of the GUM approach, GUM modified by the Bayesian approach (using the assumption of approximate Normal PDFs of output quantities) and by using the MCM. The PDFs resulting from different measurement models are computed and uncertainties are evaluated. The expanded uncertainties obtained using the different approaches are compared. These examples reveal that the number and relative weight of influential input quantities are not sufficient criteria to judge whether the output quantity can be assumed to be normally distributed.