XVIII IMEKO WORLD CONGRESS
Metrology for a Sustainable Development
September, 17 – 22, 2006, Rio de Janeiro, Brazil

ACCOUNTING FOR PHYSICAL KNOWLEDGE IN OBTAINING MEASUREMENT RESULTS AND ASSOCIATED UNCERTAINTIES

Maurice Cox[1], *Alistair Forbes*[1], *Peter Harris*[1], *João Sousa*[2]

[1] National Physical Laboratory, Teddington TW11 0LW, UK, maurice.cox@npl.co.uk
[2] Laboratório Regional de Engenharia Civil, 9000-264 Funchal, Madeira, Portugal, jasousa@lrec.pt

Abstract: This paper considers the determination of measurement results and associated uncertainties when prior physical knowledge of the quantities concerned is available. The scientific concepts that provide a basis for determining realistic solutions to such problems are discussed, and implementations of these concepts are considered. To illustrate the concepts, an example concerning the determination of an analyte concentration in chemical metrology is used. This will be the basis for a discussion concerning the various approaches available to deal with such constraints. Results stating a coverage interval containing infeasible values (values the quantity cannot physically take) should be avoided, and this fact will assist in the comparison of the relative merits of each method.

Keywords: modelling, uncertainty, solution feasibility

1. INTRODUCTION

This paper is concerned with the need to ensure that the results of measurement conform with physical knowledge, e.g., that an estimate of a chemical concentration lies in the range zero to 100 %, and that a coverage interval determined for the concentration contains only values in that range.

It might not be possible, when using a conventional method for solving such a problem, to guarantee feasibility of the solution. For instance, an application of the GUM uncertainty framework [1] might provide 99.70 % as an estimate of a concentration and 0.25 % as the associated standard uncertainty. If the quantity concerned were characterized by a Gaussian distribution, the expanded uncertainty corresponding to a 95 % coverage probability would be $2 \times 0.25\ \% = 0.50\ \%$, and hence a 95 % coverage interval for the concentration would be (99.70 ± 0.50) %. Since the part of this interval that exceeds 100 % is infeasible, it is difficult to interpret this result in a meaningful way for an application. A correctly computed coverage interval would have an upper limit no greater than 100 %. Such an interval cannot be based on an underlying assumption of normality in this instance.

This paper considers the use of all available information to assign appropriate prior probability distributions to the input quantities in a model of measurement and possibly a prior distribution to the model output quantity (the measurand). If this information correctly characterizes the physical knowledge of the quantities concerned, the posterior distribution for the measurand will embrace feasible values only, with the consequence that an estimate of the measurand and a coverage interval for the measurand will be meaningful for any required coverage probability.

2. APPROACH

Appropriate functional and probabilistic modelling can be used to encode available physical knowledge within a model of a measurement and within the probability density functions (PDFs) for the model input quantities. Additionally, a Bayesian treatment can be used to account for prior knowledge of the measurand. Solution of the model then provides a posterior PDF for the measurand that relates to measurand values that are feasible in terms of that knowledge.

Several approaches can be used to provide an estimate of the measurand and the associated uncertainty, and a coverage interval for the measurand for a prescribed coverage probability. The approaches include:

1. *Principle of maximum entropy*, in which a unique PDF for the measurand is selected from among all PDFs having specified properties, e.g., specified expectation and standard deviation and specified intervals for which the PDF is non-zero [2].
2. *Bayesian treatment*, in which a *probabilistic model* for the measurement, expressed as a likelihood function, is used to update prior information about the measurand, expressed as a prior PDF, to provide a posterior PDF. In some cases the posterior PDF can be obtained analytically. Otherwise, numerical approaches, such as Monte Carlo Markov chain methods [3], can be used.
3. *Propagation of distributions*, in which a *functional model* is used to relate the measurand to model input quantities about which information is available, and as the basis of obtaining the PDF for the measurand from the PDFs assigned to the input quantities. In some cases the PDF for the measurand can be obtained analytically. Otherwise, approximate and numerical implementations of the propagation of distributions are available, such as the GUM uncertainty framework [1] and a Monte Carlo method [4].

A solution obtained using the GUM uncertainty framework can, as stated, be infeasible, although sound solutions can be anticipated for many problems. Solutions obtained using the other approaches will be feasible. This paper indicates classes of problem for which the GUM uncertainty framework can be unsatisfactory, and therefore for which the other approaches can be considered.

The approaches considered have a wide range of applicability, as the following three problems exemplify:

1. *Expansion coefficient determination in thermal metrology*: the coefficient is estimated as the gradient of a straight line fitted to measured data. For the material concerned it is known *a priori* that the coefficient is positive, and therefore the gradient of the line should likewise be constrained to be positive.
2. *Slit width measurement in dimensional metrology*: a narrow slit width is determined as the difference between two numerically close quantities X_1 and X_2, with $X_2 > X_1$, for which measurements are available.
3. *Analyte concentration determination in chemical metrology*: the concentration of a trace element in a sample is obtained from a series of indications [5][6].

For illustration, problem 3 will be considered in more detail. The real analyte concentration takes values between zero and one. However, when some instruments are used to measure, in an unbiased way, a sample with no analyte present, about half the indicated values can be expected to be negative, and, when the concentration is very small, a proportion of the values can be expected to be negative.

2.1 Principle of maximum entropy

The "entropy" (related to uncertainty in metrology applications) of a continuous distribution with PDF $g(\eta)$ for a quantity Y is defined as

$$S = -\int g(\eta)\, \ln g(\eta)\, \mathrm{d}\eta. \tag{1}$$

The principle of maximum entropy (PME) is to select that function that gives maximum entropy subject to a set of constraints. For example, the PDF that maximizes S from among those PDFs with specified expectation and standard deviation is the PDF for the Gaussian distribution.

Suppose measurement of real analyte concentration delivers an estimate x with associated standard uncertainty $u(x)$. In the application of PME, a PDF $g(\eta)$ for real analyte concentration is sought that maximises S in expression (1) subject to the constraints

$$\begin{aligned} \textstyle\int_0^1 g(\eta)\,\mathrm{d}\eta &= 1, \\ \textstyle\int_0^1 \eta\, g(\eta)\,\mathrm{d}\eta &= x, \\ \textstyle\int_0^1 \eta^2 g(\eta)\,\mathrm{d}\eta &= u^2(x) + x^2. \end{aligned} \tag{2}$$

Lira [7] provides the general form of the maximum entropy PDF in this case, viz.,

$$g(\eta) = A \exp(\lambda_1 \eta + \lambda_2 \eta^2), \tag{3}$$

with values for the parameters A, λ_1 and λ_2 to be determined from the equations (2). Since the equations are non-linear in the parameters, a numerical method is used to solve them. Initial estimates of the parameters are provided by choosing values that define $g(\eta)$ to be the Gaussian PDF with expectation equal to x and standard deviation equal to $u(x)$.

2.2 Bayesian treatment

Bayes' theorem takes the form

$$g(\eta \mid x) = K\, l(x \mid \eta)\, g(\eta), \tag{4}$$

where $g(\eta)$ is a prior PDF for Y, $l(x \mid \eta)$ is the likelihood function for the data x, $g(\eta \mid x)$ is the posterior PDF for Y, and K constitutes a normalization factor. In words, the degree of belief for a given value η of the measurand Y, expressed as the posterior PDF for Y given data x, is proportional to (a) the likelihood that η will produce the observed data x, and (b) the degree of belief attributed to η before the observation, the so-called prior PDF for Y, expressed as $g(\eta)$ [8].

The posterior PDF may be used to provide summary information about the measurand Y, such as its expectation (mean) $\mathrm{E}(Y)$ and variance $\mathrm{V}(Y)$, defined by

$$\mathrm{E}(Y) = \int \eta\, g(\eta \mid x)\, \mathrm{d}\eta, \quad \mathrm{V}(Y) = \mathrm{E}\big(Y - \mathrm{E}(Y)\big)^2, \tag{5}$$

and its most probable value (the mode). Although $\mathrm{E}(Y)$ is the minimum-variance estimator of Y, the mode M, say, may have other desirable properties.

The prior distribution represents the information about values η available before the measurement x was taken, while the posterior represents an aggregation of the prior information and that supplied by the data. In "data-rich" experiments, the information supplied by the data is much more comprehensive that the prior information, so that the posterior is essentially proportional to the likelihood. In other circumstances, the prior distribution can contain information that the data cannot supply.

Consider an observation x of real analyte concentration η, described by the model

$$x = \eta + \varepsilon, \quad \varepsilon \sim \mathrm{N}(0, \sigma^2). \tag{6}$$

In the absence of any other information, the prior information about values η is represented by a rectangular distribution on the interval [0, 1]. In essence, the prior information excludes a value less than zero or greater than one. Given an observation x, the posterior distribution is given by $kp(\eta)$, where $p(\eta)$ is the PDF for the normal distribution $\mathrm{N}(x, \sigma^2)$ restricted to the interval [0, 1], and k is a normalizing constant. The application of Bayes' theorem in this case gives a truncated normal distribution.

It is informative to ask whether this distribution solely represents the result of applying an algebraic identity or whether it indeed provides an adequate representation of the state of knowledge about the measurand. In fact, the

distribution can also be assessed in frequentist terms. Imagine performing a large number of experiments for values η_i of real analyte concentration uniformly distributed between zero and one. For each value, make a single measurement

$$x_i = \eta_i + \varepsilon_i, \quad \varepsilon_i \sim \mathrm{N}(0, \sigma^2), \tag{7}$$

according to the model (6). For a given measurement result x and some small positive value δ, the set

$$\{\eta_i : x_i \in [x - \delta, x + \delta]\} \tag{8}$$

can be used to approximate the frequency with which a value η of the concentration leads to the measurement result x, i.e., the frequentist interpretation of the posterior distribution.

Figures 1 and 2 show (as smooth curves) the posterior distributions for, respectively, measurement results $x = 0.1$ and $x = -0.1$, with associated uncertainty $u(x) = \sigma = 0.2$. The figures also show (as jagged curves) the frequency distributions calculated for each value of x and $\delta = 0.02$ using ten million experiment simulations. In both cases the results of the frequentist experiment agree with the posterior distribution.

2.3 Propagation of distributions

A functional model for the measurement is $Y = \max\{X, 0\}$, where X is measured concentration and Y real analyte concentration [5][6]. The rationale for this model is that for a given value ξ of X the value η of Y is taken to be the closest value to ξ that satisfies the constraint $\eta \geq 0$. Suppose an estimate x of X with associated standard uncertainty $u(x)$ is available. On the basis of this information, a Gaussian distribution with mean x and standard deviation $u(x)$ is assigned to X [4]. (If x and $u(x)$ are obtained from an analysis of a small number of repeated indications, assigning to X a scaled and shifted t-distributions with finite degrees of freedom would be more appropriate [4]). The distribution for Y would resemble that for X for positive values of concentration, but with the fraction of those values that are negative concentrated at 0 %. The coverage interval obtained from the distribution would then contain only feasible values.

The GUM uncertainty framework provides an approximate implementation of the propagation of distributions. The approach is generally based on a linearization of the model $Y = f(\boldsymbol{X})$ relating input quantities $\boldsymbol{X}$ to the measurand Y through a first-order Taylor series expansion. Cases exist where higher-order terms of the series are taken into account, but it is the simpler formulation that is most often applied. Whatever the functional model (an additive model is often assumed), the first order partial derivatives of f with respect to $\boldsymbol{X}$, evaluated at the best estimates of $\boldsymbol{X}$, are combined with the standard uncertainties associated with these estimates to determine the standard uncertainty $u(y)$ associated with the measurement result y. Finally, the expanded uncertainty U and a coverage interval for Y are computed in terms of an effective degrees of freedom associated with $u(y)$ and characterizing $(Y - y)/u(y)$ by a t-distribution.

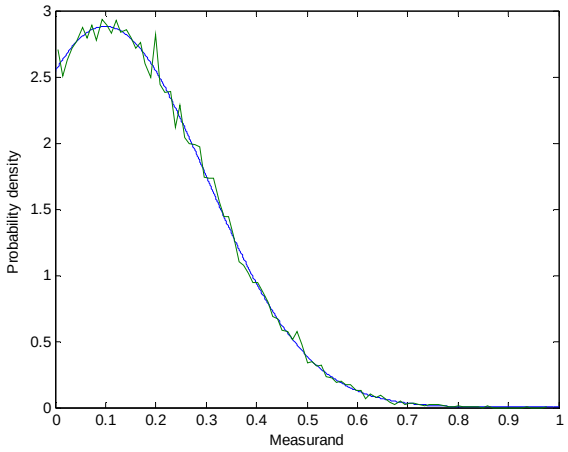

Figure 1 The posterior distribution and its frequentist interpretation for real analyte concentration for the observation $x = 0.1$ with associated uncertainty $u(x) = 0.2$.

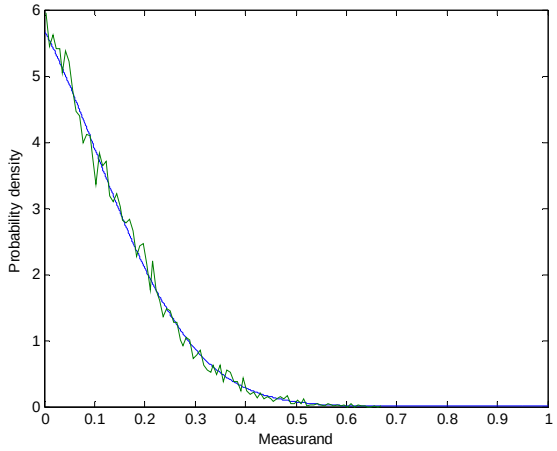

Figure 2 As figure 1, but for the observation $x = -0.1$.

Consider a measurement x of analyte concentration close to zero and having an associated standard uncertainty $u(x)$ of the same order of magnitude as x. The application of the above procedure is likely to yield a coverage interval for real analyte concentration that includes negative values, e.g., (1.0 ± 2.0) ppm or equivalently [–1.0, 3.0] ppm, which is difficult to interpret physically. Consequently, blind application of the GUM uncertainty framework should not be considered for such problems.

In cases where the applicability of the GUM uncertainty framework is questionable, the use of a numerical implementation of the propagation of distributions may be considered [4]. Generally, a Monte Carlo method (MCM) can validly be applied in such situations, e.g., to models having strongly non-linear behaviour or an input PDF based on a small sample size. The approach performs repeated random sampling from the PDFs assigned to the input quantities and evaluation of the measurement model to give a discrete representation of the distribution for the measurand. Summary information about the measurand, including expectation, variance and a coverage interval for

the measurand, are obtained from the discrete representation.

3. RESULTS

Suppose an estimate of measured concentration is $x = 0.1$. Consider three possible values for the associated standard uncertainty $u(x)$, viz., 0.05, 0.1 and 0.2. Figures 3–5 show, for each case, the PDF for the Gaussian distribution with expectation x and standard deviation $u(x)$. This PDF corresponds to the solution provided by an application of the GUM uncertainty framework (GUF). The figures also show the maximum entropy PDF constrained to be zero outside the interval [0, 1] and to have expectation x and standard deviation $u(x)$.

The results indicate that when $u(x)$ is small compared with the magnitude of x, the solution PDFs provided by the two approaches are similar (figure 3). However, when $u(x)$ is comparable to the magnitude of x, the solution PDF provided by PME resembles one-half of a "U-shaped" distribution, and is very different in character from the Gaussian distribution provided by the GUM uncertainty framework (figure 4). For the largest value of $u(x)$, the solution PDF provided by PME is unexpected, showing an increase in probability density in the neighbourhood of unity (figure 5). In all three cases, the solution PDF provided by PME incorporates the physical knowledge about real analyte concentration (that the quantity takes values in the interval [0, 1]), whereas the solution PDF provided by the GUM uncertainty framework does not.

Figures 6 shows, for the case $x = 0.1$ and $u(x) = 0.2$, the solution PDFs provided by a Bayesian treatment and an application of MCM as a numerical implementation of the propagation of distributions. The height of the left-most bin is in fact greater than 20 rather than as shown. For purposes of illustration, the chosen scale was used.

Figure 7 is the counterpart of figure 6 for the case $x = -0.1$ and $u(x) = 0.2$. This case, for which the estimate of measured concentration is negative, provides an example of a problem that cannot properly be treated using the GUM uncertainty framework. In the application of that framework to the model $Y = \max\{X, 0\}$ in this case, the sensitivity coefficient is calculated as zero and, consequently, irrespective of the value of $u(x)$, the measurand is invalidly characterized by a distribution with zero standard uncertainty. The case also provides an example of a problem that cannot be treated by PME, because no PDF exists that is zero outside the interval [0, 1] and has an expectation that is less than zero (or greater than one).

Finally, figure 8 shows, and compares, for the case $x = 0.1$ and $u(x) = 0.1$, the results obtained from three of the approaches considered. The solution PDF provided by the GUM uncertainty framework is shown as the Gaussian distribution (continuous curve), that provided by the Bayesian treatment as the dotted curve, and that provided by the Monte Carlo method as a scaled frequency distribution. For each distribution the endpoints of the (shortest) 95 % coverage interval are shown, as broken, dotted and continuous vertical lines, respectively.

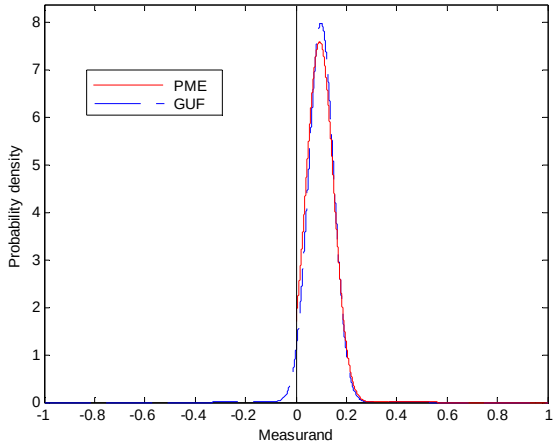

Figure 3 Gaussian PDF from the use of the GUM uncertainty framework (GUF) and maximum entropy PDF (PME) corresponding to an estimate $x = 0.1$ and associated standard uncertainty $u(x) = 0.05$.

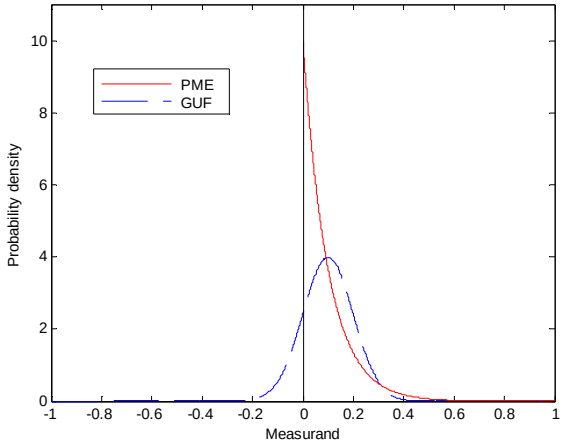

Figure 4 As figure 3, but for $u(x) = 0.1$.

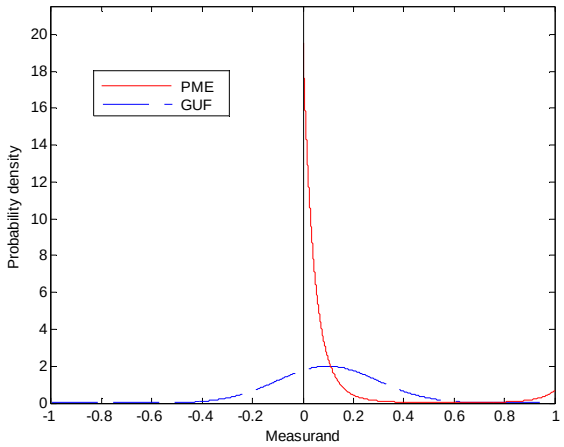

Figure 5 As figure 3, but for $u(x) = 0.2$.

4. DISCUSSION

The results presented are intended to stimulate discussion regarding the suitability of the approaches considered to the problem of measurement near a physical boundary. An apparently simply-stated problem can be far more complicated than anticipated.

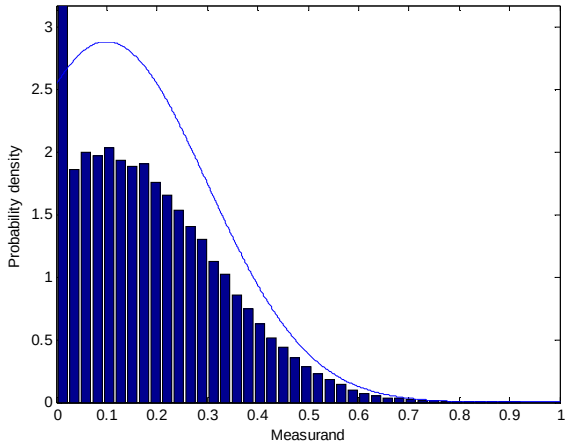

Figure 6 Solution PDFs obtained from a Bayesian treatment (continuous curve) and a Monte Carlo method (scaled frequency distribution) corresponding to an estimate $x = 0.1$ and associated standard uncertainty $u(x) = 0.2$.

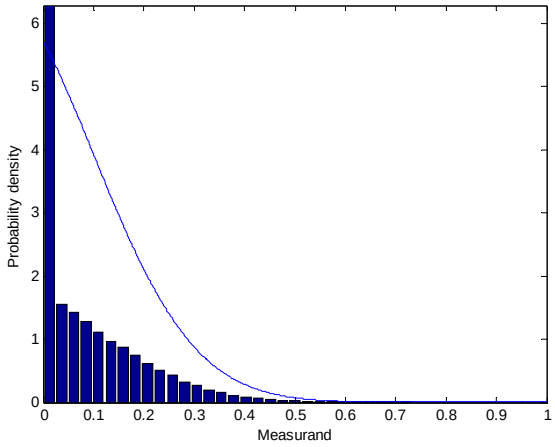

Figure 7 As figure 6, but for $x = -0.1$ and $u(x) = 0.2$.

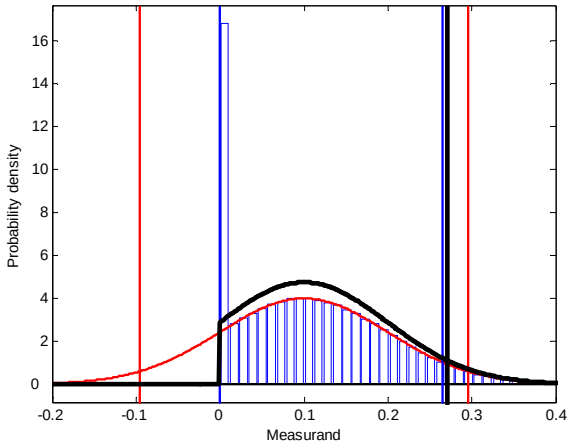

Figure 8 Comparison of solution PDFs provided by the GUM uncertainty framework (continuous curve), Bayesian approach (dotted curve) and a Monte Carlo method (scaled frequency distribution), for the case $x = 0.1$ and $u(x) = 0.1$. The corresponding (shortest) 95 % coverage intervals are also shown as broken, dotted and continuous vertical lines, respectively.

Figures 3–5 show clearly that the GUM uncertainty framework can be inadequate for such problems. A coverage interval provided by this approach can include negative values, which are physically infeasible. If that part of the PDF corresponding to negative values is simply discarded, then the area under the resulting PDF would be less than unity and an invalid solution would emerge. However, recent work [9] has advocated obtaining a feasible coverage interval from an infeasible coverage interval provided by the GUM uncertainty framework by truncating (in this case) the left-hand endpoint at zero. The estimate of the measurand and the associated standard uncertainty are retained. This approach has the advantage of being straightforward to implement, but a detailed comparison with the approaches considered here has yet to be undertaken.

There are some advantages of the approach based on PME compared with the GUM uncertainty framework (figures 3–5). Firstly, the PDF for real analyte concentration provided by the approach is feasible, as will be summary information, such as coverage intervals, derived from the PDF. Secondly, both the estimate of measured analyte concentration and the associated standard uncertainty are preserved as the expectation and standard deviation of the solution PDF.

On the other hand, a problem formulation based on PME, may not be well-posed. As noted before, the equations (2) do not permit a solution of the form (3) when the estimate x lies outside the interval [0, 1]. Lira [7] also notes that, for feasible values of x, there are values of $u(x)$ for which the equations (2) do not permit a solution. It can be expected that for values of x and $u(x)$ near these cases, PME yields problems that are ill-posed and, consequently, having solutions that might be difficult to obtain numerically.

For certain problems, such as where x is close to a physical boundary and $u(x)$ is large, the approach provides results that appear "unrealistic" (see the "U-shaped" distribution in figure 5). However, such problems may themselves be viewed as "unrealistic", in the sense that the problem relates to a "poor" measurement of a quantity whose value is very small (or very large).

Both the Bayesian approach to the problem and the use of a Monte Carlo method deliver solution PDFs that are feasible. Furthermore, unlike the GUM uncertainty framework and PME, neither the Bayesian approach nor the Monte Carlo method make an assumption about x or $u(x)$, e.g., that x is itself feasible.

The Bayesian approach and the Monte Carlo method treat the physical knowledge about the problem differently. In the Bayesian approach the knowledge is treated *probabilistically*. The prior PDF for real analyte concentration Y encapsulates the knowledge about Y independently of any measurement, and through the likelihood function a negative value of measured concentration may arise for a positive value of real analyte concentration with non-zero probability. In the use of a Monte Carlo method (and generally in the application of the propagation of distributions with the proposed functional model) the knowledge is treated *functionally*: real analyte concentration regarded as a quantity can never be negative, even though measured analyte concentration also regarded as a quantity can be positive or negative.

The results obtained from the different approaches considered are compared in figure 4 and figure 8 for the case

$x = 0.1$ and $u(x) = 0.1$. Table 1 shows the corresponding numerical values for the mean, mode and (shortest) 95 % coverage intervals, provided by each method.

Table 1 Results for $x = 0.1$ and $u(x) = 0.1$.

Approach	Mean	Mode	Coverage interval
PME	0.100	0	[0, 0.301]
Bayesian	0.129	0.100	[0, 0.274]
MCM	0.109	0.004	[0, 0.264]
GUM	0.100	0.100	[–0.096, 0.296]

The GUM uncertainty framework and PME are illustrated in figure 4, and compared in table 1: a coverage interval with negative values for GUM and a longer tail on the right-hand side for PME. Moreover, the latter preserves the mean, as expected, and has a mode equal to zero. Similarly, the comparison in figure 8 shows the coverage interval of the Bayesian approach to be slightly longer than that provided by MCM, and both shorter than the GUM coverage interval.

The mode is preserved in the case of the Bayesian approach. Indeed, there is a case when applying the Bayesian treatment for using the mode of a quantity characterized by a PDF rather than the expectation as the estimate of the measurand. The reason is that the mode, the most probable value, is uninfluenced, when $x \geq 0$, by the truncation of the PDF resulting from the prior used, which excludes negative values. When $x < 0$, the mode is at zero, the closest feasible value to x.

4. CONCLUSIONS

The solution approaches considered are capable of treating functional or probabilistic models to the degree of approximation typically required in practice. However, the modelling itself constitutes a critical stage. The choice of model (our representation of reality) dictates the solution. For the analyte concentration example, the differences in the results obtained by the various methods might be appreciable. Similar remarks might apply in other circumstances.

Metrology is not about using recipes, but rather [1] about understanding and critical analysis, and depends on detailed knowledge of the nature of the measurand and the measurement. Therefore, the purpose of this study is not to advocate the blind use of any particular approach, but to draw attention to a class of problems where one particular method can prove better suited in providing an adequate solution.

This paper has raised issues relating to the feasibility of measurement results and the associated uncertainties, and coverage intervals for the measurands. Both functional and probabilistic modelling have roles to play in the determination of feasible solutions. Basic concepts were indicated, the propagation of distributions and Bayesian treatments being central to consideration. Implementations of the concepts were outlined and their application to chemical metrology considered.

Many problems in metrology can benefit from modelling that accounts for prior physical knowledge, and therefore future work will include (a) the treatment of small samples, using *t*-distributions accordingly, in the presence of constraints, (b) further comparisons of approaches such as PME, Bayesian and MCM, to help determine their applicability to various classes of problem and (c) further work on reliable numerical methods for the different approaches.

The National Measurement System Directorate of the UK's Department of Trade and Industry supported the work of the National Physical Laboratory as part of its Software Support for Metrology programme. It supported the work of João Sousa as part of its National Measurement Secondment Scheme.

REFERENCES

[1] BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, and OIML, Guide to the Expression of Uncertainty in Measurement, 1995. ISBN 92-67-10188-9, Corrected and reprinted.

[2] W. Wöger, "Probability assignment to systematic deviations by the principle of maximum entropy", *IEEE Trans. Intrum. Meas.*, **36**, 665-8, 1987.

[3] H. S. Migon and D. Gamerman, Statistical Inference: an Integrated Approach, Arnold, London, UK, 1999.

[4] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML, Guide to the Expression of Uncertainty in Measurement, Supplement 1, Propagation of distributions using a Monte Carlo method, Draft, 2006

[5] A. M. H. van der Veen, Annarita Baldan and M. G. Cox, "Application of uncertainty evaluation principles to purity verification", submitted to Analytica Chimica Acta.

[6] LGC Ltd., "Uncertainty treatment in statements of purity for reference materials", Report LGC/VAM/ 2004/010, LGC Ltd., 2004.

[7] I. Lira, Evaluating the Measurement Uncertainty, IoP Publishing Ltd., Bristol, UK, 2002.

[8] G. D'Agostini, "Bayesian reasoning in high energy physics – Principles and applications", CERN Yellow Report 99-03, July 1999.

[9] S. Cowen and S. L. R. Ellison, "Reporting measured uncertainty and confidence and confidence intervals near natural limits", to appear in *The Analyst*.