Good practice in evaluating measurement uncertainty

Compendium of examples

Adriaan M.H. van der Veen and Maurice G. Cox (editors)

31 May 2020
Good practice in evaluating measurement uncertainty

A.M.H. van der Veen\textsuperscript{1}, M.G. Cox\textsuperscript{2}, L.L. Martins\textsuperscript{3}, A.S. Ribeiro\textsuperscript{3}, M.G. Cox\textsuperscript{2}, J.A. Sousa\textsuperscript{4}, D. Loureiro\textsuperscript{3}, M.C. Almeida\textsuperscript{3}, M.A. Silva\textsuperscript{3}, R. Brito\textsuperscript{3}, A.C. Soares\textsuperscript{3}, F. Pennecci\textsuperscript{5}, S. Demeyer\textsuperscript{6}, N. Fischer\textsuperscript{6}, O. Pellegrino\textsuperscript{4}, C. Elster\textsuperscript{7}, A. Bošnjanci\textsuperscript{8}, V. Karahodžić\textsuperscript{8}, P.M. Harris\textsuperscript{2}, and S.L.R. Ellison\textsuperscript{9}

\textsuperscript{1}VSL, Department of Chemistry, Mass, Pressure and Viscosity, Thijsseweg 11, 2629 JA Delft, the Netherlands
\textsuperscript{2}National Physical Laboratory, Data Science Department, Hampton Road, Teddington, Middlesex, United Kingdom, TW11 0LW
\textsuperscript{3}LNEC, National Laboratory for Civil Engineering, Av. do Brasil, 101, 1700-066 Lisbon, Portugal
\textsuperscript{4}IPQ, Portuguese Institute for Quality, Rua António Gião, 2, 2829-513 Caparica, Portugal
\textsuperscript{5}INRIM, Istituto Nazionale di Ricerca Metrologica, Strada delle Cacce 91, 10135 Torino, Italy
\textsuperscript{6}LNE, Laboratoire national de métrologie et d’essais, 29 avenue Roger Hennequin 78197 Trappes Cedex, France
\textsuperscript{7}PTB, Physikalisch-Technische Bundesanstalt, Braunschweig and Berlin, Germany
\textsuperscript{8}IMBIH, Institute of Metrology of Bosnia and Herzegovina, Sarajevo, Bosnia and Herzegovina
\textsuperscript{9}LGC, Laboratory of the Government Chemist, Queens Road, Teddington, TW11 0LY, UK

31 May 2020
Preface

Measurement uncertainty evaluation is at the heart of science and industry as a cross-cutting discipline, impacting on all areas of measurement. Consistent evaluation and use of measurement uncertainty are vital to the implementation of trade agreements, legislation, directives and regulations. The Joint Committee on Guides in Metrology (JCGM) provides authoritative guidance documents to address the needs of the measurement community. The evaluation and expression of measurement uncertainty are essential for the interpretation of measurement data. Even if not explicitly expressed, knowledge about the dispersion of measurement results is important to distinguish between effects from the measurement procedure and effects from other causes.

This suite of examples illustrates the use of the methods described in the Guide to the expression of Uncertainty in Measurement (GUM), and several other methods that have not yet been included in this suite of documents. The examples address issues such as the choice of the mechanism for propagating measurement uncertainty from the input quantities to the output quantities, the evaluation of standard uncertainty, modelling, reporting, and conformity assessment. Where useful, reference is made to software that supports the reproduction and implementation of the examples in practice.

This suite of examples illustrates good practice in evaluating measurement uncertainty in a variety of fields including calibration, testing, comparison and conformity, and relate to sectors that include environment, energy, quality of life, industry and society. Where useful, reference is made to software that supports the reproduction and implementation of the examples in practice.

As many practitioners benefit more quickly from worked examples than from guidance documents, the provided set of carefully selected comprehensive examples facilitates the take up of uncertainty principles as well as improving the state of the art in measurement uncertainty evaluation in the respective disciplines.

The examples are provided “as is”, without any warranty. All examples have been peer-reviewed and assessed for internal consistency and compliance with guidance in the GUM.
Disclaimer

This suite of examples has been developed as a joint effort by experts in the field of measurement. Greatest care has been exercised in the selection and development of the examples. The consortium developing this compendium uses its best efforts to deliver a high quality compendium illustrating best practice in evaluating measurement uncertainty as described in the Guide to the expression of Uncertainty in Measurement. Neither the consortium, its members, or Euramet makes no warranties with regard to the material provided however. The examples are provided “as is”. No liability is assumed for any use that is made of the Compendium.

Software, equipment and other resources identified in the examples are not necessarily the best available for the purpose. The project consortium feels however that these resources are adequate for the context in which they have been used.

Any mention of commercial products is for information only; it does not imply a recommendation or endorsement by the authors, nor by Euramet or its members.

Feedback

The consortium seeks actively feedback on this Compendium from readers. Any feedback can be sent to the editors Adriaan van der Veen (avdveen@vsl.nl) and/or Maurice Cox (maurice.cox@npl.co.uk).

Acknowledgement

The project 17NRM05 “Examples of Measurement Uncertainty Evaluation” leading to this compendium has received funding from the EMPIR programme co-financed by the Participating States and from the European Union’s Horizon 2020 research and innovation programme.
Contents

Preface iii

Disclaimer v

Feedback v

Acknowledgement v

1 Introduction 1

ADRIAAN M.H. VAN DER VEE N, MAURICE G. COX

2 Using the Monte Carlo method 3

ADRIAAN M.H. VAN DER VEE N, MAURICE G. COX

2.1 Preamble . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 3
2.2 Monte Carlo method . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 4
2.3 Software environment . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 5
2.4 Generating random numbers . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 6
2.5 Simple additive model: calculation of the molar mass of phenol . . . . . . . . . . 7
2.6 Mass example from EA 4/02 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 9
2.7 Law of propagation of uncertainty . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 13

3 Bayesian inference in R and RStan 17

ADRIAAN M.H. VAN DER VEE N

3.1 Preamble . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 17
3.2 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 17
3.3 Bayesian evaluation of the mass example of EA 4/02 . . . . . . . . . . . . . . . . . 18

Example E1 Bayesian approach applied to the mass calibration example in JCGM 101:2008 25

S. DEMEYER, N. FISCHER, M.G. COX, A.M.H. VAN DER VEE N, J.A. SOUSA, O. PELLEGRINO, A. BOSJANOVIC, V. KARAHODZIC, C. ELSTER

E1.1 Summary . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 25
E1.2 Introduction of the application . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 25
E1.3 Specification of the measurand . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 25
E1.4 Measurement model . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 26
E1.5 Input quantities of the measurement model . . . . . . . . . . . . . . . . . . . . . . . 26
E1.6 Uncertainty propagation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 26
E1.6.1 Bayesian analysis: generalities . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 26
E1.6.2 Prior distributions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 27
E1.6.3 Posterior distributions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 27
E1.7 Reporting the result . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 28

vii
<table>
<thead>
<tr>
<th>Example</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1.7.1</td>
<td>Bayesian analysis of the mass calibration example in JCGM 101:2008</td>
<td>28</td>
</tr>
<tr>
<td>E1.8</td>
<td>Conclusion</td>
<td>29</td>
</tr>
<tr>
<td>E2</td>
<td>Evaluation of measurement uncertainty in the calibration of a mobile optical measurement system</td>
<td>31</td>
</tr>
<tr>
<td>E3</td>
<td>Conformity assessment of an influenza medication as a multicomponent material</td>
<td>39</td>
</tr>
<tr>
<td>E3.1</td>
<td>Summary</td>
<td>39</td>
</tr>
<tr>
<td>E3.2</td>
<td>Introduction of the application</td>
<td>39</td>
</tr>
<tr>
<td>E3.3</td>
<td>Specification of the measurands</td>
<td>39</td>
</tr>
<tr>
<td>E3.4</td>
<td>Measurement uncertainty and correlations</td>
<td>40</td>
</tr>
<tr>
<td>E3.5</td>
<td>Specification or tolerance limits</td>
<td>40</td>
</tr>
<tr>
<td>E3.6</td>
<td>Decision rule and conformity assessment</td>
<td>40</td>
</tr>
<tr>
<td>E3.6.1</td>
<td>Bayesian framework</td>
<td>41</td>
</tr>
<tr>
<td>E3.6.2</td>
<td>Total specific risk</td>
<td>41</td>
</tr>
<tr>
<td>E3.6.3</td>
<td>Total global risk</td>
<td>42</td>
</tr>
<tr>
<td>E3.7</td>
<td>Interpretation of results</td>
<td>42</td>
</tr>
<tr>
<td>E4</td>
<td>Evaluation of measurement uncertainty in thermal comfort</td>
<td>43</td>
</tr>
<tr>
<td>E5</td>
<td>Evaluation of measurement uncertainty in SBI - Single burning item reaction to fire test</td>
<td>57</td>
</tr>
<tr>
<td>References</td>
<td></td>
<td>71</td>
</tr>
<tr>
<td>Index</td>
<td></td>
<td>75</td>
</tr>
<tr>
<td>Glossaries</td>
<td></td>
<td>75</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

Adriaan M.H. van der Veen, Maurice G. Cox

The evaluation of measurement uncertainty is an essential part of the experimenter’s task to provide a measurement result which comprises, alongside a value for the measurand, a stated uncertainty. The JCGM published a suite of documents covering various aspects of measurement uncertainty evaluation, expression and use. In many areas, measurement results are used to assess compliance with regulatory limits. To understand the risks associated with decision taking, and to apply this knowledge in conformity assessment, it is essential that the stated uncertainty is taken into account.

Many laboratories implement ISO/IEC 17025 to underpin their competence. Producers of (certified) reference materials implement in many cases both ISO/IEC 17025 and ISO 17034 for the same purpose. In proficiency testing, the requirements for demonstrating competence are laid down in ISO/IEC 17043. These standards have in common, among others, that measurement uncertainty shall be evaluated and as appropriate be expressed. Issuing CRMs (certified reference materials) with property values without uncertainty is not permitted according to ISO 17034, as it would for the user be impossible to make a proper assessment of the quality of its result when using the CRM for quality control, nor would it be possible to propagate it when using the CRM in calibration.

In this document, the examples illustrate various aspects of uncertainty evaluation and the use of uncertainty statements in conformity assessment. These aspects include, but are not limited to:

- choice of the mechanism for propagating measurement uncertainty,
- reporting measurement results and measurement uncertainty,
- conformity assessment, and
- evaluating covariances between input quantities.

Most examples cover multiple aspects. The index aids the reader to locate such aspects in the examples.

The first part of this compendium is devoted to generic aspects, which are presented in the form of tutorials that aim at helping the reader to get started with the various methods and examples presented in this compendium. They do not replace the guidance provided in the GUM suite of documents, but rather supplement the general guidance given there. The use of Bayes’ rule is not (yet) contained in the GUM, it is however recognised as one of the ways to evaluate measurement uncertainty, consistent with the spirit of the GUM, and the best mechanism to combine prior knowledge about one or more model parameters with data.
The use of software is essential for anyone performing uncertainty calculations. Most professionals rely on “off the shelf” spreadsheet software or laboratory information management system (LIMS) to perform the bulk of the relevant calculations. Such software systems have largely not been designed for the calculations necessary to evaluate, propagate and express measurement uncertainty. Some examples can nonetheless be implemented readily in this general purpose software, whereas others describe the use of other software. Some of the tutorials describe the use of R \[40\], which is an open source software package for statistical computing and data visualisation. Other examples describe the use of MATLAB or other commercial software. In all cases, these choices have been made for illustration only. If an example describes how to perform the calculation in one software package, it does not imply that it could not have been done in another. The same holds for the selection of libraries and other resources.
Chapter 2

Using the Monte Carlo method

Adriaan M.H. van der Veen, Maurice G. Cox

2.1 Preamble

One of the complicating factors in the evaluation and propagation of measurement uncertainty is the competence in mathematics and statistics required to perform the calculations. Nevertheless, standards such as ISO/IEC 17025 [7], ISO 15189 [3] and ISO 17034 [6] that specify requirements for laboratories to enable them to demonstrate they operate competently, and are able to generate valid results, require that measurement uncertainty is evaluated and reported. The well-known law of propagation of uncertainty (LPU) from the Guide to the expression of uncertainty in measurement (GUM) [9] requires the calculation of the partial derivatives of the measurement model with respect to each of the input variables.

In this tutorial, we (re)introduce the Monte Carlo method of GUM Supplement 1 (GUM-S1) [10], which takes the same measurement model and the probability density functions assigned to the input variables to obtain (an approximation to) the output probability density function. We show, based on some well-known examples illustrating the evaluation of measurement uncertainty, how this method can be implemented for a single measurand and how key summary output, such as the estimate (measured value), the associated standard uncertainty, the expanded uncertainty, and a coverage interval for a specified coverage probability, can be obtained. The Monte Carlo method of GUM-S1 [10] is a versatile method for propagating measurement uncertainty using a measurement model. It performs generally well for any measurement model, as it does not – unlike the law of propagation of uncertainty – depend on a linearisation of the model.

The use of probability density functions is well covered in the GUM [9] and further elaborated in GUM-S1 [10]. In this tutorial, the emphasis is on setting up an uncertainty evaluation using the Monte Carlo method for a measurement model with one output quantity (a “univariate” measurement model). GUM Supplement 2 (GUM-S2) [12] provides an extension of the Monte Carlo method to measurement models with two or more output quantities (“multivariate” measurement models) as well as giving a generalisation of LPU to the multivariate case.

The vast majority of the uncertainty evaluations in calibration and testing laboratories are performed using the LPU [9]. This mechanism takes the estimates (values) and associated standard uncertainties of the input quantities as input to obtain an estimate for the output quantity and the associated standard uncertainty. The measurement model is used to compute (1) the value of the output quantity and (2) the sensitivity coefficients, i.e., the first partial derivatives of the output quantity with respect to each of the input quantities. The second part of the calculation involving the partial derivatives is perceived as being cumbersome and requires skills that are often beyond the capabilities of laboratory staff and researchers. The computation of
the sensitivity coefficients can also be performed numerically \cite{16,37}. One of the advantages of the Monte Carlo method is that no sensitivity coefficients are required. All that is needed is a measurement model, which can be in the form of a computer algorithm, and a specification of the probability distributions for the input quantities. These probability distributions (normal, rectangular, etc.) are typically already specified in uncertainty budgets when the LPU is used.

In this tutorial, we show how the Monte Carlo method of GUM-S1 can be implemented in R \cite{40}. This environment is open source software and specifically developed for statistical and scientific computing. Most of the calculations in laboratories, science and elsewhere are still performed using mainstream spreadsheet software. An example of using the Monte Carlo method of GUM-S1 with MS Excel is given in the Eurachem/CITAC Guide on measurement uncertainty \cite{4}. It is anticipated that this tutorial will also be useful for those readers who would like to get started using other software tools or other languages.

2.2 Monte Carlo method

The heart of the Monte Carlo method of GUM-S1 can be summarised as follows \cite{10}. Given a measurement model of the form

$$Y = f(X_1, \ldots, X_N)$$

and probability density functions assigned to each of the input quantities $X_1, \ldots, X_N$, generate $M$ sets of input quantities $X_{1,r}, \ldots, X_{N,r}$ ($r = 1, \ldots, M$) and use the measurement model to compute the corresponding value for $Y_r$. $M$, the number of sets of input quantities should be chosen to be sufficiently large so that a representative sample of the probability density function of the output quantity $Y$ is obtained. The approach here applies to independent input quantities and a scalar output quantity $Y$. For its extension to dependent input quantities, see GUM-S1 \cite{10}, and a multivariate output quantity, see GUM-S2 \cite{12}.

GUM-S1 \cite{10, clause 6.4} describes the selection of appropriate probability density functions for the input quantities, thereby supplementing the guidance given in the GUM \cite{9, clause 4.3}. GUM-S1 also provides guidance on the generation of pseudo-random numbers. Pseudo-random numbers rather than random numbers are generated by contemporary software since the latter are almost impossible to obtain. However, comprehensive statistical tests indicate that the pseudo-random numbers generated cannot be distinguished in behaviour from truly random numbers.

Considerable confidence has been gained by the authors over many years concerning the performance of the Monte Carlo method of uncertainty evaluation from a practical viewpoint. For measurement models that are linear in the input quantities, for which the law of propagation of uncertainty produces exact results, agreement with results from the Monte Carlo method to the numerical accuracy expected has always been obtained. Thus, weight is added to the above point: there is evidence that the effects of working with pseudo-random numbers and truly random numbers are identical.

If needed, the performance of a random number generator can be verified \cite{30,39}. For the purpose of this tutorial, it is assumed that the built-in random number generator in R is fit for purpose.

A refinement of the Monte Carlo method concerns selecting the number of trials automatically so as to to achieve a degree of assurance in the numerical accuracy of the results obtained. An adaptive Monte Carlo procedure for this purpose involves carrying out an increasing number of Monte Carlo trials until the various results of interest have stabilised in a statistical sense. Details are provided in \cite{10, clause 7.9} and since then an improved method has been developed and published \cite{49}.
In many software environments, random number generators for most common probability density functions are already available; if not, they can be readily developed using random numbers from a rectangular distribution [10] annex C. (The rectangular distribution is also known as the uniform distribution.) Should even a random number generator for the rectangular distribution not be available in the software environment, then the one described in GUM-S1 can be implemented as a basis for generating random numbers. The default random number generator in R is the Mersenne Twister [33], which is also implemented in many other programming environments, including MATLAB and MicroSoft Excel (since version 2010, see [35]). Based on this random number generator, there are generators are available for a number of probability distributions [40].

The output of applying the Monte Carlo method is an array (vector) $Y_1, \ldots, Y_M$ characterising the probability density function of the output quantity. This sample is however not the form in which a measurement result is typically communicated (reported). From the output $Y_1, \ldots, Y_M$, the following can be computed:

- the measured value, usually taken as the arithmetic mean of $Y_1, \ldots, Y_M$
- the standard uncertainty, usually computed as the standard deviation of $Y_1, \ldots, Y_M$
- a coverage interval containing the value of the output quantity with a stated probability, obtained as outlined below
- the expanded uncertainty
- the coverage factor

The last two items apply when the coverage interval can be reasonably approximated by a symmetric probability density function.

The most general way of representing a coverage interval is by specifying its upper and lower limits. This representation is always appropriate whether the output distribution is symmetric or not. In many instances however, the output probability density function is (approximately) symmetric, and then the expanded uncertainty can be computed as the half-width of the coverage interval. The coverage factor can be computed from the expanded uncertainty $U(y)$ and the standard uncertainty $u(y)$, i.e., $k = U(y)/u(y)$. The symmetry of the output probability density function can be verified by examining a histogram of $Y_1, \ldots, Y_M$, or obtaining a kernel density plot, a smooth approximation to the probability density function.

### 2.3 Software environment

R is an open source language and environment for statistical computing and graphics. It is a GNU project, similar to the S language and environment, which was developed at Bell Laboratories (formerly AT&T, now Lucent Technologies) by John Chambers and colleagues. R can be considered as a different implementation of S [40]. It is available for Windows, MacOS and a variety of UNIX platforms (including FreeBSD and Linux) [41].

Users of Windows, MacOS, and a number of Linux distributions may also wish to download and install RStudio [42], which provides an integrated development environment, in which code can be written, the values of variables can be monitored, and separate windows for the console and graphics output are available. The R code provided in this primer has been developed in RStudio (version 1.2.1335, build 1379 (f1ac3452)).
2.4 Generating random numbers

In R, it is straightforward to generate a sample of random numbers from most common probability density functions. For example, the following code generates a sample of a normal distribution with mean \( \mu = 10.0 \) and standard deviation \( \sigma = 0.2 \) and a sample size \( M = 10\,000 \):

```r
M = 10000
mu = 10.0
sigma = 0.2
set.seed(2926)
X1 = rnorm(M, mu, sigma)
```

The function to be called to generate an array (vector) of random numbers with the normal distribution and mean \( \mu \) and standard deviation \( \sigma \) is called `rnorm`. The line `set.seed(2926)` is useful for debugging purposes, as it ensures that the random number generator starts at the same point every time. Any other value for the seed would also ensure the exact reproduction of the series of numbers obtained from the random number generator. If that is not required, the line can be omitted. In this tutorial, the seed is set, so that the reader can exactly reproduce the output. The output is collected in a variable named \( X1 \). It is an array with 10,000 elements.

The following code snippet shows the mean and standard deviation of the 10,000 generated numbers, using R’s built-in functions `mean` and `sd` respectively.

```r
mean(X1)
## [1] 10.00131
sd(X1)
## [1] 0.2006594
```

Using R’s functions `plot` and `density`, the kernel density of variable \( X1 \) can be plotted (see figure 2.1). The code to generate the figure is as follows:

```r
plot(density(X1), xlab = "X1", ylab = "density", main = "")
```

where `density` calculates the kernel density from the array \( X1 \) and `plot` generates the figure. The plotted density resembles that of a normal distribution. The larger the number of samples drawn from the random number generator, the closer the resemblance with the normal distribution will be.

From the first code fragment in this section, it is readily seen that R has a function for generating random numbers with a normal distribution. It also has functions for generating random numbers with a rectangular distribution (\`runif\), the \( t \) distribution (\`rt\), exponential distribution (\`rexp\) and gamma distribution (\`rgamma\). There exists a package (extension) called “trapezoid” \[29\] implementing among others the trapezoidal distribution, a package called “mvt-norm” \[28\] implementing the multivariate normal distribution (useful when some of the input quantities are dependent \[10\]), and a package called “triangle” \[19\] implementing the triangular distribution. So, apart from the curvilinear trapezoidal distribution and the arc sine distribution, random numbers for all probability density functions mentioned in GUM-S1 \[10\] table 1 are available in R.

The arc sine distribution can be implemented as follows in R. According to GUM-S1 \[10\] clause 6.4.6.1, a U-shaped random variable \( X \) on the interval \( [a, b] \) can be obtained through

\[
X = \frac{a + b}{2} + \frac{b - a}{2} \sin \Phi
\]
Figure 2.1: Density plot of the random variable X1 having a normal distribution with mean 10.0 and standard deviation 0.2.

where $\Phi$ is a random variable with a rectangular distribution on $[0, 2\pi]$. In R, a function `rarcsin` that provides such a random variable, and a call to that function, can be coded as follows:

```r
rarcsin <- function(n,a,b) {
  X = (a+b)/2 + (b-a)/2 * sin(runif(n,0,2*pi))
  return(X)
}
X2 = rarcsin(M,-1.0,1.0)
```

The argument n determines the number of random numbers returned; a and b denote the lower and upper limits respectively of the interval over which the arcsine distribution has a non-zero density. If $n > 1$, the function returns an array; if $n = 1$ it returns a single number. This behaviour mimics the behaviour of the other functions implemented in R to generate random numbers.

The last line in the code snippet creates an array $X2$ of $M$ elements ($M = 10\,000$ in this instance) of a random variable having an arcsine distribution over the interval $[-1, 1]$. A histogram (obtained through the R function `hist`) is shown in figure 2.2.

2.5 Simple additive model: calculation of the molar mass of phenol

In this example, the molar mass of phenol (molecular formula $C_6H_5OH$) is computed. The example shows how an output quantity with an uncertainty is obtained from input quantities with uncertainty. There is no experiment involved. The example is pivotal for many calculations involving reference data, such as atomic weights, molar masses and enthalpies of formation.

The molar mass is computed from the atomic masses and the coefficients appearing the molecular formula, which for the elements involved are 6 for carbon, 6 $(5+1)$ for hydrogen and 1 for oxygen. The current relative atomic masses are used as published by IUPAC (International Union
Chapter 2. Using the Monte Carlo method

Figure 2.2: Histogram of the random variable X² containing M = 10 000 samples having an arcsine distribution between -1 and 1

of Pure and Applied Chemistry [34]. The relative atomic masses that apply to “normal materials” are called standard atomic weights [22, 34]. Their interpretation is described in an IUPAC technical report [38].

The molar mass of phenol (chemical formula C₆H₅OH) is computed as

\[ M_r(C_6H_5OH) = 6A_r(C) + 6A_r(H) + A_r(O) \]

The Monte Carlo method is implemented in R using M = 100 000 trials. The R code that performs the evaluation reads as

```r
M = 100000
C = runif(M, 12.0096, 12.0116)
H = runif(M, 1.00784, 1.00811)
O = runif(M, 15.99903, 15.99977)
MW = 6*C + 6*H + O
MW.val = mean(MW)
MW.unc = sd(MW)
MW.Unc = (quantile(MW, probs = 0.975) - quantile(MW, probs = 0.025))/2.0
```

The first line declares a variable M that holds the number of trials to be carried out by the Monte Carlo method. Then, for each of the elements, M samples are drawn using the rectangular distribution (using R’s function runif) and the lower and upper limits provided by the standard atomic weights of IUPAC [34]. These arrays have respectively the names C, H and O for the atomic masses of carbon, hydrogen and oxygen. The molar mass is then computed in the line defining MW. R is very efficient with vectors (arrays) and matrices (tables) [15]. The value of the molar mass (MW.val) is computed by taking the average of MW, the standard uncertainty by taking the standard deviation of MW and the expanded uncertainty by taking the half-width of the 95% coverage interval. The latter is obtained by calculating the 0.025 and 0.975 quantiles (which provides a probabilistically-symmetric coverage interval).
Chapter 2. Using the Monte Carlo method

9.100 94.105 94.110 94.115 94.120 94.125
0 20 40 60 80 100
Molar mass (g/mol)
Density (mol/g)

Figure 2.3: Output probability density function of the molar mass of phenol and superimposed a normal distribution with the same mean and standard deviation.

The code to plot the output probability density function of the molar mass (MW) and to superimpose a normal distribution with the same mean and standard deviation is given below:

```r
x = seq(from = MW.val-4*MW.unc, to=MW.val+4*MW.unc, by=8*MW.unc/100)
hx = dnorm(x, MW.val, MW.unc)
{
  plot(density(MW), xlab = "Molar mass (g/mol)",
       ylab = "Density (mol/g)", main="",
       xlim = c(min(x), max(x)), ylim = c(0, max(hx)))
  lines(x, hx, lwd = 2, lty = 2, col = "red")
}
```

The first two lines compute the relevant part of the normal distribution around the mean ± 4 standard deviations. The subsequent lines plot the output probability density function and the normal distribution respectively.

The figure is shown as figure 2.3. It is obvious that the normal distribution is not an appropriate approximation of the probability density function of the output quantity, which is much narrower than the normal distribution. The molar mass is 94.1108 g mol\(^{-1}\) with standard uncertainty 0.0035 g mol\(^{-1}\). The expanded uncertainty is 0.0059 g mol\(^{-1}\). The coverage factor is 1.67.

2.6 Mass example from EA 4/02

In most instances, the Monte Carlo method is implemented using a measurement model (or measurement equation). In this section, the mass calibration example of EA 4/02 [25] is taken and the implementation of the Monte Carlo method is described. The evaluation using the Monte Carlo method rests on the same assumptions for the input quantities as in that example. The example is developed in such a way that for any measurement model having one output quantity the same steps can be followed. The measurement model is coded in the form of a function,
which promotes writing tidy code. It also allows iterative calculations to be readily implemented when the measurement model is defined implicitly [12]. This example describes the calibration of a 10 kg weight by comparison with a standard 10 kg weight. The weighings are performed using the substitution method. This method is implemented in such a way that three mutually independent observations for the mass difference between the two weights are obtained.

The measurement model is given by [25, S2]:

\[
m_X = m_S + \delta m_D + \delta m + \delta m_C + \delta B,
\]

(2.1)

where the symbols have the following meaning

\( m_X \) conventional mass of the weight being calibrated,
\( m_S \) conventional mass of the standard,
\( \delta m_D \) drift of the value of the standard since its last calibration,
\( \delta m \) observed difference in mass between the unknown mass and the standard,
\( \delta m_C \) correction for eccentricity and magnetic effects,
\( \delta B \) correction for air buoyancy.

For using the Monte Carlo method, probability density functions are assigned to each of the five input quantities [10]. These probability density functions are described in the original example [25].

The conventional mass of the standard \( m_S \) is modelled using the normal distribution with mean 10 000.005 g and standard deviation 0.0225 g. The standard deviation (standard uncertainty) is calculated from the expanded uncertainty and the coverage factor provided on the calibration certificate. This interpretation is also described in GUM-S1 [10, 6.4.7]. The drift of the mass of the standard weight \( \delta m_D \) is modelled using a rectangular distribution, centred at 0 g and with a half-width of 0.015 g. The corrections for eccentricity and magnetic effects, and that for air buoyancy are both modelled using a rectangular distribution with midpoint 0.000 g and half-width 0.010 g.

The mass difference \( \delta m \) between the two weights computed from the indications of the balance is calculated as the mean of \( n = 3 \) independent observations. EA 4/02 explains that the associated standard uncertainty is computed from a pooled standard deviation 0.025 g, obtained from a previous mass comparison, divided by \( \sqrt{n} \).

In the implementation of the Monte Carlo method, the three observations are simulated using normal distributions with means of the observed values (i.e., 0.010 g, 0.030 g and 0.020 g respectively) and a standard deviation of 0.025 g for each. The mass difference is formed by calculating the arithmetic average of the three simulated observations.

The measurement model (equation (2.1)) can be coded in R as follows:

```r
# measurement function
mass.x <- function(m.std,dm.d,diff,dm.c,dm.B) {
  m.std + dm.d + diff + dm.c + dm.B
}
```

where \( m \_ \text{std} \) denotes the conventional mass of the standard weight, \( dm \_ d \) the drift correction of the conventional mass of the standard weight, \( diff \) the mass difference obtained from the substitution weighing, \( dm \_ c \) the correction due to eccentricity and magnetic effects, and \( dm \_ B \) the correction due to air buoyancy. The function is called \( mass \_ x \) and returns the value of the output quantity \( m_X \).

Most programming languages implement a “for” loop, which enables executing a block of code a defined number of times. Anyone familiar with this “for” loop in computer programming would now use this kind of loop to code the recipe given in GUM-S1 clause 7.2.2 [10]. An implementation of the Monte Carlo method with a fixed value for the number of samples \( M \) would then read as follows:
Chapter 2. Using the Monte Carlo method

# implementation of the procedure of GUM-S1 with fixed M
prob = 0.95
M = 10000 * ceiling(1.0/(1.0-prob))  # GUM-S1 7.2.2
m.x = numeric(M)
m.data = numeric(3)
for (i in 1:M) {
    m.std = rnorm(1,10000.005,0.0225)
dm.d = runif(1,-0.015,+0.015)
dm.c = runif(1,-0.010,+0.010)
dm.B = runif(1,-0.010,+0.010)
m.data[1] = rnorm(1,0.01,0.025)
m.data[2] = rnorm(1,0.03,0.025)
m.data[3] = rnorm(1,0.02,0.025)
m.diff = mean(m.data)
m.x[i] = mass.x(m.std,dm.d,m.diff,dm.c,dm.B)
}

On the first line, the probability level of the coverage interval \( \text{prob} \) is defined to be 0.95. In accordance with the guidance in clause 7.2.2 of GUM-S1 \([10]\), \( M \) is calculated using the built-in function ceiling which returns the smallest integer not less than its argument. With \( \text{prob} = 0.95 \) the net effect of calling ceiling is that the floating point number is converted to an integer, as the result of \( 1/(1-\text{prob}) \) is 20, hence the minimum number of Monte Carlo trials is \( M = 10000 \cdot 20 = 200000 \). Then an array (vector) \( m.x \) is declared that will hold the values calculated for the mass of the weight being calibrated. The vector \( m.data \) is a temporary storage for simulating the mass differences between the standard weight and the weight being calibrated. In the for loop, at each iteration a sample is drawn of the input quantities \( m_S \) (m.std), \( \delta m_D \) (dm.d), \( \delta m_C \) (dm.c), and \( \delta B \) (dm.B). The mass difference from comparing the two weights (m.diff) is simulated by drawing from a normal distribution with different means, but the same standard deviations, the three readings and taking the average. The measured value of the output quantity \( m_X \) (m.x) is finally obtained by calling the measurement model with as arguments the input quantities.

Running the above code provides the following output for the mean, standard deviation (standard uncertainty) and the coverage interval of \( m_X \):

\[
\text{print(\text{mean(m.x)}, digits = 9)}
\]

\[
\text{## [1] 10000.025}
\]

\[
\text{print(\text{sd(m.x)}, digits = 2)}
\]

\[
\text{## [1] 0.029}
\]

\[
\text{quantile(m.x,probs = c(0.025,0.975))}
\]

\[
\text{## 2.5% 97.5%}
\]

\[
\text{## 9999.968 10000.082}
\]

where the argument \text{probs} holds the probabilities corresponding to the lower and upper ends of the probabilistically symmetric 95% coverage interval.

This way of coding an implementation of the Monte Carlo method would work in a large number of computer languages, including Python, MATLAB, Fortran, C, C++ and Pascal. While the above code in R does what is intended, the same task can be performed with greater effectiveness in R, exploiting the fact that R is very efficient in working with vectors and matrices \([15]\). Computational efficiency is especially important with more complex models and larger numbers.
of Monte Carlo trials, as it can greatly reduce the required computing time. The following code implements the same simulation, using vectors and matrices where possible:

```r
# implementation of the procedure of GUM-S1 with fixed M
prob = 0.95
M = 10000 * ceiling(1.0/(1.0-prob)) # GUM-S1 7.2.2
m.std = rnorm(M,10000.005,0.0225)
dm.d = runif(M,-0.015,+0.015)
dm.c = runif(M,-0.010,+0.010)
dm.B = runif(M,-0.010,+0.010)
m.data = matrix(rep(c(0.01,0.03,0.02),M), nrow = M, byrow = TRUE)
m.data = m.data + matrix(rnorm(3*M,0,0.025),nrow = M,byrow = TRUE)
m.diff = apply(m.data,1,mean)
m.x = mass.x(m.std,dm.d,m.diff,dm.c,dm.B)
```

Now the variables `m.std`, `dm.d`, `dm.c`, and `dm.B` are vectors holding all `M` values for the input quantities. The data from comparing the weights is summarised in a matrix called `m.data` of `M` rows and 3 columns. The matrix is constructed by adding the means (0.01, 0.03, and 0.02) to the simulated data which have been generated using the normal distribution with mean 0 and standard deviation 0.025. The mass differences are computed by calculating the row means and storing these in `m.diff` using the R function `apply`. Note also that the measurement model can be called with vectors rather than scalars as arguments (last line of the code); in this case also `m.x` is a vector of length `M`.

The second code runs in less than half the time of the first implementation. For this simple example, the difference is a matter of a few seconds, but for more complex models the difference in speed will be of more practical significance. Especially the steps that are repeated often should be carefully thought about. Another issue is memory use. The second implementation consumes appreciably more memory (for it holds all generated values for the input quantities) than the first (which only holds the last value for each of the input quantities).

The second code provides the following output for the mean, standard deviation (standard uncertainty) and the coverage interval of `m.x`:

```r
print(mean(m.x),digits = 9)
## [1] 10000.0249
print(sd(m.x),digits = 2)
## [1] 0.029
quantile(m.x,probs = c(0.025,0.975))
## 2.5% 97.5%
## 9999.967 10000.082
```

The output probability density function is shown in figure 2.4. The form of the probability density function resembles that of a normal distribution with mean 10 000.025 g and standard deviation 0.029 g. The following code computes the expanded uncertainty by taking the half-width of the 95% coverage interval and the coverage factor by dividing the expanded uncertainty by the standard uncertainty:

```r
m.x.Unc = (quantile(m.x,probs = 0.975) - quantile(m.x,probs = 0.025))/2.0
m.x.k = m.x.Unc/sd(m.x)
```

Examples of evaluating measurement uncertainty  First edition (M18)
The expanded uncertainty is 0.057 g and the coverage factor is 1.96. This coverage factor is that of a 95% coverage interval of the normal distribution. The coverage factor differs from that used in EA 4/02 which uses $k = 2$ for obtaining (at least) 95% coverage probability. The difference is readily explained, as the dominating uncertainty contributions are modelled using the normal distribution, and the sum of two normal distributions is also normally distributed (see also the measurement model, equation (2.1)). That the output quantity has an (approximately) normal distribution is reflected in the coverage factor obtained from the Monte Carlo method.

Now all results are obtained that commonly appear on a calibration certificate (as well as in many test reports), as described in ISO/IEC 17025[7]:

- the measured value (= value of the output quantity)
- the expanded uncertainty
- the coverage factor

In this case, one might also be willing to state that the output probability density function is a normal distribution. Whereas in this case such a statement can be made, in most cases the output probability density function cannot directly be approximated by a well-known analytic probability density function. Comparison of the three results listed above with those from the LPU would imply that for comparable data LPU would be fit for purpose in a subsequent uncertainty evaluation. In a subsequent uncertainty evaluation, with $m_X$ as one of the input quantities, the above information suffices to apply the law of propagation of uncertainty, say [9].

### 2.7 Law of propagation of uncertainty

The law of propagation of uncertainty (LPU) is the most widely used mechanism for propagating uncertainty. Whereas with the Monte Carlo method the lack of computing and programming skills can form a bottleneck, with the LPU it is often the calculation of the sensitivity coefficients, i.e., the partial derivatives of the output quantity with respect to the input quantities,
that provides a difficulty. Most guidance documents, such as the GUM [9], GUM-S2 [12] and EA 4/02 [25] direct their readers to analytic differentiation of the measurement model to obtain the expressions for calculating the sensitivity coefficients. Whilst this guidance is fully appropriate, it is not always practicable, for many people have lost their skills in differentiation. The fact that there are tables with derivatives of common functions (such as [45,48]) is barely mentioned in such documents. Numerical approximation of the sensitivity coefficients [16,37] is a very good alternative, provided that it is done properly. In this section, we show how to use numerical differentiation and the law of propagation of uncertainty to perform the uncertainty evaluation of the mass example of EA 4/02 [25].

The R package numDeriv provides the function \texttt{grad} (from gradient) that returns from a function a generally good approximation, using Richardson extrapolation [18], of the partial derivatives of the input variables. The function returns a vector holding the values of these partial derivatives. The function passed to \texttt{grad} should have only one argument, namely a vector holding all input variables. Hence, the measurement model needs to be reformulated as follows:

```r
# measurement function
mass2.x <- function(x) {
  m.std = x[1]; dm.d = x[2];
  diff = x[3]; dm.c = x[4]; dm.B = x[5]
  m.std + dm.d + diff + dm.c + dm.B
}
```

where \( x \) denotes the vector with input variables. For clarity and convenience, in the function body of \texttt{mass2.x} the same symbols have been used as in \texttt{mass.x} shown previously. The convenience extends to easier debugging the code as necessary. The penultimate line calculates the result of the function as the sum of the five input variables, just as in the case of the Monte Carlo method.

The uncertainty evaluation itself can be coded as follows:

```r
require(numDeriv)
m.std = 10000.005; dm.d = 0.0; diff = mean(c(0.01,0.03,0.02))
dm.c = 0.0; dm.B = 0.0;
sens = grad(func=mass2.x,x=c(m.std,dm.d,diff,dm.c,dm.B))
m.std.u = 0.0225
dm.d.u = 0.015/sqrt(3); dm.c.u = 0.010/sqrt(3)
diff.u = 0.025/sqrt(3); dm.B.u = 0.010/sqrt(3)
m.x = mass2.x(c(m.std,dm.d,diff,dm.c,dm.B))
m.x.unc = sqrt(sum(sens^2*c(m.std.u,dm.d.u,diff.u,dm.c.u,dm.B.u)^2))
```

The first line loads the package \texttt{numDeriv} (which needs to be installed in RStudio. The next two lines define the values of the input quantities. The vector \texttt{sens} on the fourth line holds the sensitivity coefficients returned by calling \texttt{grad}. The subsequent three lines calculate the standard uncertainties associated with the five input quantities. The penultimate line calculates the estimate of the output quantity \( m.x \) and the last line its associated standard uncertainty \( m.x.unc \). Again, this last line shows the flexibility of R working with vectors.

The mass of the calibrated weight is 10 000.025 g with standard uncertainty 0.029 g. Using a coverage factor \( k = 2 \), the expanded uncertainty becomes 0.059 g. These results reproduce those in example S.2 of EA 4/02 to the number of decimal digits given.

The values of the sensitivity coefficients are

```r
## [1] 1 1 1 1 1
```

and are identical to those given in EA 4/02 [25]. The code is also valid for measurement models with non-trivial sensitivity coefficients [37].
The approach described also works with correlated input variables. In that case, the calculation of the standard uncertainty associated with $m_X$ is performed as follows:

\[
D = \text{diag}(c(\text{m.std.u, dm.d.u, diff.u, dm.c.u, dm.B.u}))
\]
\[
CM = D \times \times D
\]
\[
tmp = t(\text{sens}) \times \times CM \times \times \text{sens}
\]
\[
m.x.unc = \sqrt{\text{tmp}[1,1]}
\]

The first two lines form the covariance matrix, diagonal in this case, associated with the five input quantities. (These are only needed to create the covariance matrix; if there were correlations between the five input variables, the code for creating it would have to be adapted accordingly.) The actual implementation of the LPU for correlated input variables is given in the last two lines of the previous code. By vector/matrix multiplication (see also the law of propagation of uncertainty in GUM-S2 [12]) a covariance matrix of dimension $1 \times 1$ associated with the output quantity is returned ($\text{tmp}$). The last line takes the square root of the only element in this matrix (holding the variance of $m_X$) to obtain the standard uncertainty associated with $m_X$. This standard uncertainty is 0.029 g.
Chapter 3

Bayesian inference in \texttt{R} and \texttt{RStan}

Adriaan M.H. van der Veen

3.1 Preamble

In this tutorial, we revisit the well-known example of an uncertainty evaluation of the calibration of a 10 kg weight, published in the guidance document EA 4/02 from European co-operation for Accreditation (EA) to illustrate how a Bayesian evaluation of measurement uncertainty can be set up using \texttt{R} \cite{R} and \texttt{rstan} \cite{rstan}, including the use of Markov Chain Monte Carlo (MCMC). The example shows how type A and type B methods of evaluating standard uncertainty are coded, how the calculations are performed and how from the posterior of the measurand the value, standard uncertainty, coverage interval and coverage factor can be determined.

3.2 Introduction

The mass example in EA 4/02 \cite{EA402} was introduced in chapter \cite{Chapter2} and this Bayesian inference builds forth on the example as already described. The Bayesian evaluation using MCMC highlights that the type B evaluation of standard uncertainty in such a Bayesian setting is very similar to the same evaluation using the Monte Carlo method of GUM Supplement 1 (GUM-S1) \cite{GUMS1}. The greatest difference is usually in those uncertainty components that are evaluated using type A methods. There is no technical reason for using MCMC in this instance, for the same result (measured value and expanded uncertainty) can be obtained by much simpler means (i.e., the law of propagation of uncertainty \cite{GUM} or the Monte Carlo method of GUM Supplement 1 (GUM-S1) \cite{GUMS1}). For this reason, it is an excellent case for assessing whether an implementation of the MCMC provides valid results. In this revisit of the mass example, the type A evaluation of standard uncertainty \cite{GUM,Edition25} of the mass differences is fairly straightforward, as the original example assumes a known standard deviation. This known standard deviation can be viewed as a kind of “prior knowledge”, which justifies a Bayesian treatment (the treatment in EA 4/02 is in this respect Bayesian, for it utilises the information about the repeatability standard deviation of the weighings.

The calculations in this tutorial have been performed using \texttt{R}, an environment for statistical computation \cite{R}, and the package \texttt{rstan} \cite{rstan} that enables writing Bayesian models in a straightforward manner. This environment and the use of \texttt{RStan} for Bayesian inference have been introduced previously \cite{RStan1,RStan2}.

From the posterior probability density function obtained through a Bayesian inference, as has been and will be shown, all essential information can be retrieved, including the measured value,
its associated standard uncertainty, and a 95% coverage interval, just as in the case of the Monte Carlo method of GUM-S1 \[10\]. It is worth noting that the posterior is not necessarily symmetric, so that obtaining an expanded uncertainty can turn out to be impossible. The expanded uncertainty is the half-width of a symmetric coverage interval \[9,25\] and obviously only makes sense if that interval is (approximately) symmetric.

### 3.3 Bayesian evaluation of the mass example of EA 4/02

The re-evaluation of the mass example from EA 4/02 is performed by mimicking the assumptions made in EA 4/02 \[25\] as closely as possible. The example describes for all type B evaluations the probability density functions used (rectangular and normal distributions). For the repeated observations of the mass difference, the normal distribution is used with a known standard deviation, which is consistent with the original evaluation as presented in EA 4/02. The measurement model is given in equation (2.1). For Bayesian inference, probability density functions need to be assigned to each of the five input quantities. This aspect of the evaluation is similar to the use of the Monte Carlo method of GUM-S1 \[10\] (see also chapter 2).

The conventional mass of the standard is modelled using a normal distribution with mean 5 mg (the deviation from the nominal value of 10 kg) and standard deviation 22.5 mg. The subtraction of the nominal value is necessary to obtain stable output in the Markov Chain Monte Carlo (MCMC) calculation; it does not in any way change the outcome of the inference, apart from that we have redefined the measurand to be the departure from the nominal mass, rather than the mass of the 10 kg weight itself. The measurement model could be written as

\[
\Delta m_X = \Delta m_S + \delta d_D + \delta m + \delta m_C + \delta B \tag{3.1}
\]

where \(\Delta m_X\) denotes the departure from its nominal mass for the weight being calibrated, and \(\Delta m_S\) the departure from its nominal mass for the standard weight. The fact that the outcome of the MCMC calculation is sensitive to the choice of variables (`parametrisation`) in the model has been discussed previously already \[46,47\]. This sensitivity is one of the hurdles to be taken when performing iterative calculations (as MCMC is \[27\]).

In Stan code, the model of the mass calibration reads as

```stan
data {
  int<lower=1> N;
  vector[N] diffs;
}
parameters{
  real m_s;
  real<lower=-15,upper=15> dm_d;
  real diff;
  real<lower=-10,upper=10> dm_c;
  real<lower=-10,upper=10> dm_B;
}
model {
  m_s ~ normal(5,22.5);
  diff ~ normal(0,500.0); // weak prior
  diffs ~ normal(diff,25.0);
}
generated quantities{
  real m_x;
  m_x = m_s + dm_d + diff + dm_c + dm_B;
}
```
In the data block, a vector of length $N$ is declared called `diffs` which holds the recorded mass differences from comparing the masses of the standard and the weight being calibrated. The input quantities evaluated using type B methods for evaluating standard uncertainty [9] are declared as model parameters in the `parameters` block. By default, Stan assigns these variables a rectangular distribution over their domain of validity [20]. If no constraints on the variable are specified, the domain is $(-\infty, +\infty)$ and thus the assigned prior is improper (i.e. not integrating to one over its domain [27]). This default can be overridden by specifying another prior in the `model` block.

The first variable not having assigned a rectangular distribution, $m_s$, denotes $\Delta m_s$ the departure of its nominal mass of the standard weight (see equation (3.1)). In the `model` block, it is assigned a normal distribution with mean 5 mg and standard deviation 22.5 mg. In Bayesian models, this way of coding a probability distribution would be the same as assigning a prior to the parameter $m_s$. It is not combined with data, so the probability distribution of this parameter does not change as part of the Bayesian inference. Hence, it is sometimes argued that the way in which the GUM [9] deals with type B evaluations of standard uncertainty is 'weakly Bayesian' [8] by nature. The 'weakly' aspect lies in the fact that only an informative prior is assigned, and that it is not combined with (new) measurement data, as no data are generated for this parameter during the measurement. The same applies to the other model parameters in equation (3.1) evaluated using type B methods.

The corrections for drift ($dm_d$), eccentricity and magnetic effects ($dm_c$), and buoyancy ($dm_B$) are all declared with upper and lower limits ($\pm 15$ mg for drift, and $\pm 10$ mg for the other two). As Stan assigns these a rectangular distribution taking into account the limits, there is no need to assign these three variables explicitly a rectangular distribution in the `model` block. Actually, there are computational advantages to write the model as shown; these advantages are well covered in the description of the Stan language [20,44].

The mass difference between the weight being calibrated and the standard weight is called `diff` in the model. It is assigned a weakly informative prior in the form of a normal distribution (that is implied by the example as well) with zero mean and a large standard deviation. This prior does not do more than saying that we expect, before observing the data, that the mass difference between the two weights will be close to zero, given a large standard deviation (500 mg in this case, much larger than any of the uncertainties considered). If the OIML class of a weight is known, the maximum departure from the nominal mass can be presumed to be known, unless the weight is out-of-specification. The specification of the OIML class can be used to elicit a value for the standard deviation of the prior. In the last line of the `model` block, the data (held in `diffs`) is used to update the probability distribution of `diff`, given a fixed standard deviation of 25 mg. The latter is also given in the example in EA 4/02 [25]. This is the only part of the model where Bayes’ rule is applied, and also the only part that differs in nature from the evaluation in the original example, where a frequentist method is used (just as for other type A methods in the GUM [9,47]).

The measurement model finally appears in the `generated quantities` block. The mass (difference from the nominal mass) of the weight being calibrated is declared as $m_x$ and its value is calculated as described in equation (3.1). Note that only $m_x$ needs to be specified using the measurement model. When evaluating the model, Stan will compute a value for $m_x$ during each cycle of the MCMC, thus providing a sample of its posterior.

When running the MCMC, a number of iterations are necessary to enable the sampler to configure itself. This is called the “warmup phase”. Furthermore, several series of samples (“chains”) are generated, as one of the criteria for convergence is that the ratio of the between- and within-chain variances is close to one [27]. More details have been given elsewhere [46,47]. Running the model with 21,000 iterations and a warmup of 1000 iterations, using 4 chains yields the...
following output:

```r
## Inference for Stan model: 091dba697d92e3c497468850cfc395085.
## 4 chains, each with iter=21000; warmup=1000; thin=1;
## post-warmup draws per chain=20000, total post-warmup draws=80000.
##
## mean se_mean sd 2.5% 97.5% n_eff Rhat
## m_s 4.93 0.07 22.43 -39.20 49.10 94501 1
## dm_d 0.03 0.03 8.66 -14.25 14.26 99537 1
## diff 20.01 0.05 14.42 -8.29 48.20 93814 1
## dm_c -0.01 0.02 5.77 -9.49 9.48 104833 1
## dm_B -0.02 0.02 5.77 -9.51 9.49 120707 1
## m_x 24.95 0.09 29.19 -32.23 82.20 97247 1
## lp__ 2.24 0.01 1.77 -2.16 4.59 31166 1
##
## Samples were drawn using NUTS(diag_e) at Fri Mar 29 19:17:03 2019.
## For each parameter, n_eff is a crude measure of effective sample size,
## and Rhat is the potential scale reduction factor on split chains (at
## convergence, Rhat=1).
```

In the output, the first column lists the parameters. \( \text{lp\_} \) denotes the log of the joint posterior. The second column, labelled ‘mean’ provides the estimates of the parameters. The next column gives the standard error of the mean due to the MCMC calculation. The standard error generally decreases as the number of iterations increases. It should be small enough to produce sufficiently accurate results. A simple (yet not always sufficient way) is to repeat the calculation and to see how well the results agree. In the column ‘sd’, the standard deviation (= standard uncertainty) of the parameters is given. The following two columns contain the lower and upper limits of the probabilistically-symmetric 95 % coverage interval. \( \text{n\_eff} \) provides a crude estimate of the effective number of samples \( [27] \). The final column, labelled Rhat, gives the ratio of the between-chain and within-chain variance. For convergence, it should be close to one \( [27,47] \).

A more thorough way of looking at the results of the MCMC calculation is to inspect the traceplots of the parameters. These show the parameter values for each chain and each iteration in the calculation. There is in this example only one variable that warrants looking at its traceplot (diff), which is shown in figure 3.1.

The traceplot shows good convergence: the parameter values fluctuate around a mean value and there are no meaningful differences between the chains.

The value of the correction due to eccentricity and magnetic effects (\( \text{dm\_c} \)) is 0.0 mg with standard uncertainty 5.8 mg. Both values are very close to the values obtained using the rectangular distribution: 0.0 mg and \( 10\text{mg}/\sqrt{3} \approx 5.8\text{mg} \), respectively. The same can be said about the correction due to air buoyancy (\( \text{dm\_B} \)), which has the value 0 mg with standard uncertainty 5.8 mg; the values that are obtained using the rectangular distribution directly are the same as for the correction due to eccentricity and magnetic effects. For the third correction, that due to drift (\( \text{dm\_d} \)), the expected standard deviation is \( 15\text{mg}/\sqrt{3} \approx 8.7\text{mg} \), and the mean is zero \( [25] \); the results obtained from the MCMC are 8.7 mg and 0 mg respectively.

The mass difference of the standard (\( \Delta\text{m}_s \)) is evaluated as 4.9 mg with standard uncertainty 22 mg; the ones given in the original example are 5 mg and 22.5 mg respectively. The calculated mass difference is evaluated as 20 mg with standard uncertainty 14 mg; the ones given in the original example are 20 mg and 14.4 mg respectively. In both cases, the agreement is excellent.

The mass difference between the weights is returned as \( \text{m\_x} \); its value is 24.9 mg and its standard uncertainty is 29 mg. We can see that the value and standard deviation are very close to the ones given in the original example (25 mg and 29.3 mg respectively \( [25] \)).

The final hurdle in this example is the reproduction of the expanded uncertainty, which is stated to be 59 mg \( [25] \). The MCMC calculation provides for all parameters the 95 % coverage.
intervals (see the output discussed previously). Before attempting to compute the expanded uncertainty as the half-width of an approximately symmetric coverage interval, the shape of the posterior of $\Delta m_x$ should be assessed for symmetry. This posterior is shown in figure 3.2.

From figure 3.2, it can be seen that the posterior of $\Delta m_x$ is fairly symmetric. One way to compute the expanded uncertainty would be to compute the difference between the mean (= measured value) and the lower end of the 95% coverage interval and the difference between the upper end of the said interval and the mean, and to use whichever is the greater. The R code to perform the calculation takes the form

```r
Lower = quantile(fitout$m_x, probs = 0.025)
Upper = quantile(fitout$m_x, probs = 0.975)
m_x = mean(fitout$m_x)
U.val = max(Upper-m_x, m_x-Lower)
U.k = U.val/sd(fitout$m_x)
```

where the variable fitout holds the extracted samples of the MCMC calculation. The expanded uncertainty thus obtained is 57 mg and the coverage factor is 1.96. The latter is obtained by dividing the expanded uncertainty by the standard uncertainty. This coverage factor is consistent with that for the normal distribution, which should not come as a surprise, as the two dominating uncertainty contribution have the normal distribution (the mass of the standard and the mass difference between the two weights) \[25\]. Alternatively, the expanded uncertainty could also be computed as the half-width of the 95% coverage interval.

The reprocessing of this example in a computational environment for Bayesian inference highlights that

1. type B evaluations of standard uncertainty can be viewed as assigning only a prior distribution to the parameter concerned; as there is no ‘fresh’ data, the distribution is not updated using Bayes’ rule;

2. the normal distribution naturally arises under the assumption that the standard deviation is known (if the latter were assumed to be completely unknown, the $t$ distribution arises \[23, 27, 47\]).
the propagation of distributions is performed in a similar fashion as in the Monte Carlo method of GUM-S1 (but the Monte Carlo methods are different! [10, 27]).

A concern for those favouring classical statistical methods could be the weakly informative prior assigned to the variable $\text{diff}$. There are different ways to assess the influence of assigning this prior. One of the ways would be to replace it by a reference prior, which in this case would be a rectangular distribution over the interval $(-\infty, +\infty)$ [27]. The corresponding model is obtained by removing the weakly informative prior from the $\text{model}$ block and takes the form

```r
data {
  int<lower=1> N;
  vector[N] diffs;
}
parameters{
  real m_s;
  real<lower=-15,upper=15> dm_d;
  real diff;
  real<lower=-10,upper=10> dm_c;
  real<lower=-10,upper=10> dm_B;
}
model {
  m_s ~ normal(5,22.5);
  diffs ~ normal(diff,25.0);
}
generated quantities{
  real m_x;
  m_x = m_s + dm_d + diff + dm_c + dm_B;
}
```

Fitting the amended model with the same number of chains and chain lengths yields
Comparing the results of the MCMC with those obtained previously shows that they are very close, which underlines the ‘weakly-informative’ behaviour of the assigned prior to \( \text{diff} \) in the original model. Another way to assess the influence of the assigned prior would be to choose other values for the standard deviation (now 500 mg). A larger standard deviation would cause a reduction in the influence of the prior (it becomes less informative); a smaller standard deviation would cause it to become more influential \[46, 47\]. It is left to the reader to confirm that the chosen prior indeed behaves as a weakly-informative prior.

Finally, it is worth noting that for \( \Delta m_x \), the departure of its nominal mass of the weight being calibrated, no prior is assigned. Its probability distribution is obtained in a calculation from the other parameters using the measurement model \((3.1)\). This part of the model behaves in the same way as it would do when using the Monte Carlo method of GUM-S1 \([9]\).
Example E1

Bayesian approach applied to the mass calibration example in JCGM 101:2008

S. Demeyer, N. Fischer, M.G. Cox, A.M.H. van der Veen, J.A. Sousa, O. Pellegrino, A. Bosjanovic, V. Karahodzic, C. Elster

E1.1 Summary

This example describes the calibration of a conventional mass of a weight W against a reference weight R with a nominal mass of 100 g. The example builds on that given in JCGM 101:2008. This time a Bayesian evaluation of the measurement is performed. A Bayesian approach differs from the Monte Carlo method (MCM) of JCGM 101:2008 and the law of propagation of uncertainty (LPU) in JCGM 100:2008 in that it combines prior knowledge about the measurand with the data obtained during calibration. From the joint posterior probability density function which is obtained from this combination, a value and a coverage interval for the measurand are obtained.

E1.2 Introduction of the application

A Bayesian approach to the mass calibration example consists in updating a prior state of knowledge on the measurand by the means of new information obtained during calibration.

In JCGM 101:2008 [10], the available information is a best estimate and its associated uncertainty. A comparison of results between LPU, MCM and the Bayesian approach is given in this document. We show that the three methods give similar results when the Bayesian approach is conducted under a non-informative prior distribution. We also show the effect of various prior parameter values for Gaussian prior distributions.

The data and sources of this example are available electronically [24].

E1.3 Specification of the measurand

As described in JCGM 101:2008 [10], the application concerns the calibration of a weight W of mass density $\rho_W$ against a reference weight R of mass density $\rho_R$ having nominally the same mass $m_{\text{nom}}$, using a balance operating in air of mass density $\rho_a$. Let $\delta m_R$ be the mass of a small weight of density $\rho_R$ added to R to balance it with W.

It is usual to work in terms of conventional masses. The conventional mass $m_{W,c}$ of W is the mass of a (hypothetical) weight of density $\rho_0 = 8 \times 10^3 \text{ kg m}^{-3}$ that balances W in air at density...
\( \rho_{a_0} = 1.2 \text{ kg m}^{-3} \).

The measurand \( \delta m = m_{W,c} - m_{\text{nom}} \) is the deviation of \( m_{W,c} \) from the nominal mass \( m_{\text{nom}} = 100 \text{ g} \).

### E1.4 Measurement model

According to JCGM 101:2008\[10\], in terms of conventional masses \( m_{W,c}, m_{R,c} \) and \( \delta m_{R,c} \), an approximation adequate for most purposes is

\[
m_{W,c} = (m_{R,c} + \delta m_{R,c}) \left[ 1 + (\rho_a - \rho_{a_0}) \left( \frac{1}{\rho_W} - \frac{1}{\rho_R} \right) \right].
\]

(E1.1)

The measurement model used in the mass calibration example of \[10\] is

\[
\delta m = (m_{R,c} + \delta m_{R,c}) \left[ 1 + (\rho_a - \rho_{a_0}) \left( \frac{1}{\rho_W} - \frac{1}{\rho_R} \right) \right] - m_{\text{nom}}.
\]

(E1.2)

### E1.5 Input quantities of the measurement model

Table E1.1 summarizes the input quantities \( m_{R,c}, \delta m_{R,c}, \rho_a, \rho_W \) and \( \rho_R \), and the PDFs assigned from \[10\]. In the table, a Gaussian distribution \( N(\mu, \sigma^2) \) is described in terms of expectation \( \mu \) and standard deviation \( \sigma \), and a rectangular distribution \( R(a, b) \) with endpoints \( a \) and \( b \) \((a < b)\) in terms of expectation \((a + b)/2\) and semi-width \((b - a)/2\).

Table E1.1: The input quantities and PDFs assigned to them for the mass calibration model (E1.2), from JCGM 101:2008\[10\].

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Distribution</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Expectation ( \mu )</td>
</tr>
<tr>
<td>( m_{R,c} )</td>
<td>( N(\mu, \sigma^2) )</td>
<td>( 100 \text{ 000.000 mg} )</td>
</tr>
<tr>
<td>( \delta m_{R,c} )</td>
<td>( N(\mu, \sigma^2) )</td>
<td>( 1.234 \text{ mg} )</td>
</tr>
<tr>
<td>( \rho_a )</td>
<td>( R(a,b) )</td>
<td>( 1.20 \text{ kg m}^{-3} )</td>
</tr>
<tr>
<td>( \rho_W )</td>
<td>( R(a,b) )</td>
<td>( 8 \times 10^3 \text{ kg m}^{-3} )</td>
</tr>
<tr>
<td>( \rho_R )</td>
<td>( R(a,b) )</td>
<td>( 8.00 \times 10^3 \text{ kg m}^{-3} )</td>
</tr>
</tbody>
</table>

Note that the input quantity \( \delta m_{R,c} \) is usually associated with fresh calibration results but that in the JCGM 101:2008\[10\] treatment of mass calibration, a Type B uncertainty evaluation of \( \delta m_{R,c} \) is performed resulting in a Gaussian distribution \( \delta m_{R,c} \sim N(d, u^2(d)) \) where \( d \) is a best estimate with associated uncertainty \( u(d) \).

### E1.6 Uncertainty propagation

#### E1.6.1 Bayesian analysis: generalities

To set up a Bayesian framework \[27, 17\], a statistical model is needed for which we choose to revise notation, as in \[26\], so that random variables are now represented by Greek letters. In this document, we consider statistical models of the form

\[
D|\eta, \theta \sim N\left( [\eta - G(\theta)] / C(\theta), u^2(d) \right)
\]

(E1.3)
in which the observed data \( d \) is modelled as a realization of a random variable \( D \) having a Gaussian distribution with mean \( [\eta - G(\theta)]/C(\theta) \) and variance \( u^2(d) \), \( C(\theta) \neq 0 \) and \( G(\theta) \) are smooth functions. The measurand is denoted by \( \eta \) and \( \theta \) is a vector of further parameters.

The statistical model (E1.3) is equivalent to the measurement model (E1.2)

\[
\eta = G(\theta) + C(\theta) \zeta
\]  

(E1.4)

with

\[
\zeta = \delta m_{R,c},
\]

(E1.5)

\[
\theta = (\rho_a, \rho_W, \rho_R, m_{R,c}),
\]

(E1.6)

\[
C(\theta) = 1 + (\rho_a - \rho_{a_0}) \left( \frac{1}{\rho_W} - \frac{1}{\rho_R} \right),
\]

(E1.7)

\[
G(\theta) = C(\theta)m_{R,c} - m_{\text{nom}}.
\]

(E1.8)

The measurement result (accounting for uncertainty in \( \theta \)) is represented by the marginal posterior probability distribution \( \pi(\eta|d) \), resulting from the (potentially) high-dimensional integration

\[
\pi(\eta|d) = \int \pi(\eta, \theta|d)d\theta,
\]

(E1.9)

where \( \pi(\eta, \theta|d) \) is the joint posterior distribution of \((\eta, \theta)\).

In this document, point estimates are derived from equation (E1.9) for comparison with LPU and MCM. We introduce the following quantities \( \hat{\eta} = E(\eta|d) = \int \eta \pi(\eta|d)d\eta \) to denote the posterior mean of the measurement result and \( u^2(\hat{\eta}) = V(\eta|d) = \int (\eta - \hat{\eta})^2 \pi(\eta|d)d\eta \) to denote the posterior variance of the measurement result. Coverage intervals are computed as shortest intervals as described in [10], similar to highest posterior density (HPD) intervals in Bayesian statistics.

E1.6.2 Prior distributions

In the Bayesian paradigm, a prior state of knowledge is described by a prior distribution \( \pi(\eta) \). For instance, a way to express the prior belief that the measurand is close to a specified value \( \eta_0 \) is to use a prior Gaussian distribution \( \pi(\eta) \sim N(\eta_0, \sigma_0^2) \) where the standard deviation \( \sigma_0 \) controls the degree of belief in \( \eta_0 \). For instance, if \( |\eta_0| \) is much larger than \( \sigma_0 \), a small value of this ratio leads to a poorly informative prior. Another way of modelling poor prior information is to use the so-called non informative prior \( \pi(\eta) \propto 1 \). Alternative prior distributions can be used (uniform, truncated, etc.) to model particular features of the measurand (bounds, non negativity, etc.).

E1.6.3 Posterior distributions

Bayes’ formula gives the expression of the posterior distribution \( \pi(\eta, \theta|d) \) as a function of the likelihood \( l(d|\eta, \theta) \) and the prior distribution \( \pi(\eta, \theta) \):

\[
\pi(\eta, \theta|d) = \frac{l(d|\eta, \theta)\pi(\eta, \theta)}{m(d)},
\]

(E1.10)
where \( m(d) = \int l(d|\eta, \theta) \pi(\eta, \theta) \, d\eta \) is the marginal distribution of \( d \), \( \pi(\eta, \theta) = \pi(\eta) \pi(\theta) \) and \( \pi(\theta) \) is the probability distribution of the input quantities contained in \( \theta \).

Equivalently, (E1.10) can be translated into the proportionality relation as follows

\[
\pi(\eta, \theta|d) \propto l(d|\eta, \theta) \pi(\eta, \theta).
\]

(E1.11)

Letting \( s = u(d) \), the likelihood is

\[
l(d|\eta, \theta) \propto \frac{1}{(\sqrt{C(\theta)} \sigma d^2)^2} \exp \left\{ -\frac{(\eta - m(\theta))^2}{2[C(\theta)]^2 \sigma d^2} \right\},
\]

(E1.12)

where \( m(\theta) = C(\theta) d + G(\theta) \).

Under the non-informative prior distribution \( \pi(\eta) \propto 1 \), Bayes' formula gives

\[
\pi(\eta, \theta|d) \sim N(m(\theta), [C(\theta)]^2 s^2) \pi(\theta).
\]

(E1.13)

Under the Gaussian prior distribution, \( \pi(\eta) \sim N(\eta_0, \sigma_0^2) \), the Bayes's formula gives

\[
\pi(\eta, \theta|d) \sim N(m_p(\theta), \sigma_p^2(\theta)) \pi(\theta),
\]

(E1.14)

where the posterior mean and variance of \( \eta \) are, respectively,

\[
m_p(\theta) = \sigma_p^2(\theta)^{-1} \left\{ \frac{\eta_0}{\sigma_0^2} + \frac{m(\theta)}{[C(\theta)]^2 s^2} \right\}, \quad \sigma_p^2(\theta) = \left\{ \frac{1}{\sigma_0^2} + \frac{1}{[C(\theta)]^2 s^2} \right\}^{-1}.
\]

The posterior mean is a weighted mean between the prior \( \eta_0 \) and the best estimate \( m \) and the inverse posterior variance, also called precision, is the sum of the prior precision, \( 1/\sigma_0^2 \), and the precision from the best estimate, \( 1/\{[C(\theta)]^2 s^2\} \).

The integration according to (E1.9) is performed with a Monte Carlo method. The total number of Monte Carlo trials is decomposed as follows: \( n_{MC} \) draws according to \( \pi(\theta) \) and \( n_{post} \) draws from the Gaussian distributions (E1.13) or (E1.14) giving a total of \( n_{MC} \times n_{post} \) simulations.

### E1.7 Reporting the result

#### E1.7.1 Bayesian analysis of the mass calibration example in JCGM 101:2008 [10]

Results obtained with LPU, MCM and the Bayesian approach with non informative prior (Bayes-NI) are displayed in Table E1.2 (LPU1 and LPU2 denote respectively the first and second order Taylor approximations) and plotted in Figure E1.1. The comparison shows a good agreement between methods.

Table E1.2: Comparison of results obtained with LPU1, LPU2, MCM and Bayes-NI, the Bayesian analysis conducted with non informative prior distribution. Results from LPU1, LPU2, MCM are taken from [10].

<table>
<thead>
<tr>
<th>Method</th>
<th>( \hat{\delta}m ) /mg</th>
<th>( u(\hat{\delta}m) ) /mg</th>
<th>Shortest 95% coverage interval CI/mg</th>
</tr>
</thead>
<tbody>
<tr>
<td>LPU1</td>
<td>1.234 0</td>
<td>0.053 9</td>
<td>[1.128 5, 1.339 5]</td>
</tr>
<tr>
<td>LPU2</td>
<td>1.234 0</td>
<td>0.075 0</td>
<td>[1.087 0, 1.381 0]</td>
</tr>
<tr>
<td>MCM</td>
<td>1.234 0</td>
<td>0.075 4</td>
<td>[1.083 4, 1.382 5]</td>
</tr>
<tr>
<td>Bayes-NI</td>
<td>1.234 0</td>
<td>0.075 5</td>
<td>[1.084 5, 1.383 0]</td>
</tr>
</tbody>
</table>

¹For the so-called non informative prior, [26] showed that Bayesian marginal posterior uncertainty coincides with the MCM uncertainty estimate when the model is linear.

---

Examples of evaluating measurement uncertainty  
First edition (M18)
Figure E1.1: Distributions of $\delta m$ obtained under Gaussian approximation with LPU2, MCM and Bayes-NI from the values in Table E1.2.

Results obtained with a Gaussian prior distribution are displayed in Table E1.3 and plotted in Figure E1.2. It can be observed that, when the prior standard deviation $\sigma_0$ increases, the weight of the prior distribution decreases and the resulting posterior distribution tends to the non informative case.

Table E1.3: Comparison of results obtained with the Bayesian analysis under Gaussian prior distributions.

<table>
<thead>
<tr>
<th>$\eta_0$ (mg)</th>
<th>$\sigma_0$ (mg)</th>
<th>$\delta m$ (mg)</th>
<th>$u(\delta m)$ (mg)</th>
<th>Shortest 95% coverage interval CI (mg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.134</td>
<td>0.020</td>
<td>1.184</td>
<td>0.039</td>
<td>[1.106 9, 1.261 3]</td>
</tr>
<tr>
<td>1.134</td>
<td>0.010</td>
<td>1.153</td>
<td>0.017</td>
<td>[1.127 2, 1.152 5]</td>
</tr>
<tr>
<td>1.134</td>
<td>0.040</td>
<td>1.214</td>
<td>0.061</td>
<td>[1.093 6, 1.334 5]</td>
</tr>
</tbody>
</table>

In this section, all the results obtained with the Bayesian approach involve $2 \times 10^7$ Monte Carlo trials ($n_{\text{MC}} = 20000$, $n_{\text{post}} = 1000$).

**E1.8 Conclusion**

This document shows the main features of a Bayesian approach of uncertainty evaluation applied to the mass calibration example in JCGM 101:2008 [10]. The measurement result is represented by the marginal posterior distribution of the measurand which accounts for both uncertainty sources and prior information on the measurand, and is comparable in nature with the PDFs provided by MCM [10] and by the Gaussian distribution from LPU [9].

In general, the Bayesian approach provides a flexible tool for statistical modelling and achieves added value through prior information, at some computational price. In many circumstances, reduced uncertainties are obtained.

This example illustrates the well known property that, if a non-informative prior distribution is chosen, the Bayesian posterior distribution is essentially the same distribution from which the
Example E1. Bayesian approach applied to the mass calibration example in JCGM 101:2008

Figure E1.2: Posterior distributions of $\delta m$ obtained under Gaussian prior distributions from the values in Table E1.3.

MCM determines a sample for linear measurement models, see for instance [26] and [43] for the mass calibration problem.

This example shows that prior distributions can be chosen to allow a simplified Bayesian uncertainty analysis using a Monte Carlo method instead of a Markov Chain Monte Carlo method [21], usually used to sample from high-dimensional integrals, as in [43] and [36], which can be helpful for any practitioner already familiar with MCM willing to perform a Bayesian uncertainty analysis.
Example E2

Evaluation of measurement uncertainty in the calibration of a mobile optical measurement system

E2 Evaluation of measurement uncertainty in the calibration of a mobile optical measurement system

L. L. Martins\textsuperscript{1}, A. S. Ribeiro\textsuperscript{1}, M. G. Cox\textsuperscript{2}, J. A. Sousa\textsuperscript{3}, D. Loureiro\textsuperscript{1}, M. C. Almeida\textsuperscript{1}, M. A. Silva\textsuperscript{1}, R. Brito\textsuperscript{1}, A. C. Soares\textsuperscript{1}

\textsuperscript{1}LNEC, National Laboratory for Civil Engineering, Av. do Brasil, 101, 1700-066 Lisbon, Portugal
\textsuperscript{2}NPL, National Physical Laboratory, Hampton Road, Teddington, TW11 0LW, UK
\textsuperscript{3}IPQ, Portuguese Institute for Quality, Rua António Gião, 2, 2829-513 Caparica, Portugal

E2.1 Summary
This example illustrates the evaluation of measurement uncertainty related to the calibration of a mobile optical measurement system, based on the use of an SI-traceable reference standard bar measured in specific spatial positions. The measurement system studied (Krypton, model K610) contains three linear CCD cameras, in different spatial positions and orientations, with overlapping fields of view, permitting the simultaneous observation of an infrared LED located in a region of interest. By applying triangulation techniques, the measurement system can determine the static and dynamic spatial position of a set of observed LEDs.

E2.2 Introduction to the application
Mobile optical measurement systems (MOMS) are currently used in different laboratories and industries, namely, in automotive, motorsport, aerospace, and naval and structural engineering. In these contexts, MOMS support the static and dynamical dimensional measurement of objects with complex geometrical shapes, allowing in situ non-contact manual or automatic measurements of their position or motion.

This example is focused in one type of MOMS – the Krypton K610 [1-2] – which comprises a camera system and control unit, acquisition computer, measurement probe, multiplexer boxes and infrared LED. The camera system has three linear CCD cameras, in different spatial positions and orientations with overlapping fields of view, which results in a pyramidal measurement volume characterized by a depth range between 1.5 m and 6.0 m and a corresponding cross-section area ranging between \((0.90 \text{ m} \times 0.55 \text{ m})\) and \((3.6 \text{ m} \times 2.6 \text{ m})\). Using triangulation techniques, the location of an infrared LED can be determined with a measurement accuracy \([2]\) variable between \((60 + 7 \text{ m}^{-1} \cdot L) \mu m\) and \((130 + 17 \text{ m}^{-1} \cdot L) \mu m\), where \(L\) is the distance from the location to the camera, expressed in metres. The acquisition frequency depends on the number of LEDs targeted, varying between 1 kHz for one LED and 232 Hz for 15 LED, for example [1-2].

Regular calibration of this MOMS is advisable, before and after in situ measurements, since it is vulnerable to effects such as transportation, assembly, installation and temperature variation, all of which would introduce uncertainty. This metrological operation is supported by the use of a carbon fibre composite bar, with an SI-traceable reference length (close to 1550 mm), placed in specific spatial positions in front of the camera system, as shown in figure 1, which displays seven spatial distances \(d_1, ..., d_7\) that are measured.
Using a measurement probe composed of a ruby tip and nine LEDs spatially distributed by three sets in the same plane, the position of each end-point in the standard bar can be determined and their relative distance compared with the reference length value. If required for instrumental accuracy improvement, the performed measurements can support the adjustment of the MOMS [2].

E2.3 Specification of the measurand
In this example, the measurand is the length reading, $l$, obtained in the MOMS at a reference temperature of 20 °C. The calibration of the MOMS involves quantifying the difference, $d$, between the measurand and the reference value, $l_s$, related to the measurement standard:

$$d = l - l_s(1 + \alpha_s \cdot \theta_s),$$

where $\alpha_s$ is the coefficient of thermal expansion of the carbon fibre composite bar and $\theta_s$ is its temperature deviation from the 20 °C reference temperature during calibration. Since the MOMS performs non-contact dimensional measurements, the length reading is not directly related to any linear thermal expansion effects.

E2.4 Measurement model
The measurement model can be obtained from expression (1) and is given by

$$l = d + l_s(1 + \alpha_s \cdot \theta_s).$$

E2.5 Uncertainty propagation
The application of the law of propagation of uncertainty [3] to expression (2) yields

$$u^2(l) = c_d^2 \cdot u^2(d) + c_{ls}^2 \cdot u^2(l_s) + c_{\alpha s}^2 \cdot u^2(\alpha_s) + c_{\theta s}^2 \cdot u^2(\theta_s),$$

with

$$c_d = \frac{\partial l}{\partial d} = 1,$$

$$c_{ls} = \frac{\partial l}{\partial l_s} = 1 + \alpha_s \cdot \theta_s,$$

$$c_{\alpha s} = \frac{\partial l}{\partial \alpha_s} = l_s \cdot \theta_s,$$

$$c_{\theta s} = \frac{\partial l}{\partial \theta_s} = l_s \cdot \alpha_s,$$

and thus
Example E2. Evaluation of measurement uncertainty in the calibration of a mobile optical measurement system

\[
u^2(l) = u^2(d) + (1 + \alpha_s \cdot \theta_s)^2 \cdot u^2(l_s) + l_s^2 \cdot \theta_s^2 \cdot u^2(\alpha_s) + l_s^2 \cdot \alpha_s^2 \cdot u^2(\theta_s).
\]

(8)

A reference standard bar such as that used in the calibration of the MOMS is designed to be a rigid body characterized by a null coefficient of thermal expansion at room temperature. It is composed of carbon fibres (related to a reduced negative coefficient of thermal expansion) in a polymer matrix (with a coefficient of thermal expansion of opposite sign). Therefore, if a null coefficient of thermal expansion is considered for the reference standard bar, expression (8) can be simplified:

\[
u^2(l) = u^2(d) + u^2(l_s) + l_s^2 \cdot \theta_s^2 \cdot u^2(\alpha_s).
\]

(9)

Table 1 shows the differences between reading and reference values obtained in one calibration of the MOMS, being composed of four individual tests where the reference standard bar was placed in seven spatial positions (transverse, vertical, longitudinal and four diagonals; see figure 1) in the measurement volume, at a nominal observation distance of 3.5 m. Table 1 also mentions the corresponding average, \(\bar{d}_i\), and experimental standard deviation, \(s(d_i)\), for each of the seven spatial positions \((i = 1, 2, \ldots, 7)\).

<table>
<thead>
<tr>
<th>Test number</th>
<th>Differences between readings and reference values in the calibration of the MOMS, in mm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(d_1)</td>
</tr>
<tr>
<td>1</td>
<td>-0.017</td>
</tr>
<tr>
<td>2</td>
<td>0.020</td>
</tr>
<tr>
<td>3</td>
<td>-0.011</td>
</tr>
<tr>
<td>4</td>
<td>0.001</td>
</tr>
<tr>
<td>(\bar{d}_i)</td>
<td>-0.002</td>
</tr>
<tr>
<td>(s(d_i))</td>
<td>0.016</td>
</tr>
</tbody>
</table>

Based on the results shown in Table 1, correlation coefficients, \(r(d_i, d_j)\), were determined between pairs of spatial positions of the reference standard bar,

\[
r(d_i, d_j) = \frac{u(d_i, d_j)}{s(d_i) \cdot s(d_j)},
\]

(10)

where \(u(d_i, d_j)\) is the covariance, which can be calculated by

\[
u(d_i, d_j) = \frac{1}{n(n-1)} \sum_{k=1}^{n} (d_{i_k} - \bar{d}_i)(d_{j_k} - \bar{d}_j),
\]

(11)

with \(n\) being the number of independent pairs of observations of \(d_i\) and \(d_j\), in this case four. The results obtained are shown in Table 2. Correlation between differences is present since the same physical measurement standard (the reference standard bar) is used in their determination, although in different spatial positions, but having a specific measurement uncertainty related to its reference value.
Example E2. Evaluation of measurement uncertainty in the calibration of a mobile optical measurement system

Table 2 Correlation coefficients between obtained differences.

<table>
<thead>
<tr>
<th>r(d_i, d_j)</th>
<th>d_1</th>
<th>d_2</th>
<th>d_3</th>
<th>d_4</th>
<th>d_5</th>
<th>d_6</th>
<th>d_7</th>
</tr>
</thead>
<tbody>
<tr>
<td>d_1</td>
<td>1</td>
<td>0.20</td>
<td>0.20</td>
<td>-0.20</td>
<td>-0.14</td>
<td>0.01</td>
<td>0.06</td>
</tr>
<tr>
<td>d_2</td>
<td>0.20</td>
<td>1</td>
<td>0.16</td>
<td>-0.14</td>
<td>-0.22</td>
<td>0.06</td>
<td>0.17</td>
</tr>
<tr>
<td>d_3</td>
<td>0.20</td>
<td>0.16</td>
<td>1</td>
<td>-0.25</td>
<td>-0.16</td>
<td>0.10</td>
<td>0.13</td>
</tr>
<tr>
<td>d_4</td>
<td>-0.20</td>
<td>-0.14</td>
<td>-0.25</td>
<td>1</td>
<td>0.12</td>
<td>-0.10</td>
<td>-0.09</td>
</tr>
<tr>
<td>d_5</td>
<td>-0.14</td>
<td>-0.22</td>
<td>-0.16</td>
<td>0.12</td>
<td>1</td>
<td>-0.17</td>
<td>-0.23</td>
</tr>
<tr>
<td>d_6</td>
<td>0.01</td>
<td>0.06</td>
<td>0.10</td>
<td>-0.10</td>
<td>-0.17</td>
<td>1</td>
<td>0.22</td>
</tr>
<tr>
<td>d_7</td>
<td>0.06</td>
<td>0.17</td>
<td>0.13</td>
<td>-0.09</td>
<td>-0.23</td>
<td>0.22</td>
<td>1</td>
</tr>
</tbody>
</table>

The four tests performed in the calibration of the MOMS contributed to the measurement samples of differences between reading and reference values, related to the seven adopted spatial positions of the reference standard bar, for which individual average values and experimental standard deviations were obtained, as shown in table 1. In a global perspective, an estimate of the difference between reading and reference values can be obtained by averaging. The corresponding standard uncertainty [3] is given by

\[
u^2(d) = \sum_{i=1}^{7} c_i^2 \cdot u^2(d_i) + 2 \sum_{i=1}^{6} \sum_{j=i+1}^{7} c_i \cdot c_j \cdot u(d_i) \cdot u(d_j) \cdot r(d_i, d_j), \tag{12}\]

where \(c_i = c_j = \frac{1}{7}\), \(r(d_i, d_j)\) is the correlation coefficient and \(u^2(d_i) = s^2(d_i)\) (see table 2), allowing to simplify expression (12) to

\[
u^2(d) = \frac{1}{49} \left[ \sum_{i=1}^{7} s^2(d_i) + 2 \sum_{i=1}^{6} \sum_{j=i+1}^{7} s(d_i) \cdot s(d_j) \cdot r(d_i, d_j) \right]. \tag{13}\]

The use of the values in tables 1 and 2 in expression (13) results in a standard uncertainty equal to 0.0065 mm.

The reference standard bar was calibrated in a controlled laboratory environment, using an SI-traceable coordinate measuring machine, which allowed the determination of the reference length between the two end-points of this measurement standard: \(l_0 = 1550,313\) mm. The calibration certificate issued mentioned an expanded measurement uncertainty equal to 0.016 mm, corresponding to the product of a standard uncertainty of 0.0079 mm and the expansion factor of 2.02, evaluated using a Student t distribution with 102 degrees of freedom, in order to achieve a coverage probability of 95 %.

The estimate of the coefficient of thermal expansion of the reference standard bar is considered, as above, to be equal to zero \((\alpha_s = 0.0 \cdot 10^{-6} \, ^\circ C^{-1})\). However, it has an associated standard uncertainty represented by a uniform distribution with a semi-amplitude of \(2 \cdot 10^{-6} \, ^\circ C^{-1}\), taking into consideration the dispersion of known values for the two main components (carbon fibre and polymer matrix) of the composite bar. Therefore, the standard uncertainty [3] corresponds to

\[
u(\alpha_s) = \frac{1}{\sqrt{3}} 2 \cdot 10^{-6} \, ^\circ C^{-1} = 1.2 \cdot 10^{-6} \, ^\circ C^{-1}. \tag{14}\]

The calibration of the MOMS was performed in a controlled laboratory environment, with a nominal temperature of 20 °C. Room temperature time records show an average temperature deviation of \(\theta_s = 0.1 \, ^\circ C\), and a cyclic variation following an arcsine distribution of temperature with a semi-amplitude of 0.5 °C. In addition, these temperature measurements in time were
performed by a digital thermo-hygrometer with an instrumental standard uncertainty of 0,2 °C, related to a normal distribution. The combination of these two temperature measurement uncertainties is given by

$$u(\theta_s) = \sqrt{\left(0.5 \, ^\circ C / \sqrt{2}\right)^2 + (0.2 \, ^\circ C)^2} = 0.41 \, ^\circ C.$$  (15)

Table 3 shows a summary of the abovementioned standard uncertainty components of the length reading performed by the MOMS during calibration.

| Standard uncertainty component $u(x_i)$ | Source of uncertainty | Standard uncertainty $u(x_i)$ | $c_i \equiv \frac{\partial l}{\partial x_i}$ | $u_i(l) \equiv |c_i| \cdot u(x_i)$ | Degrees of freedom |
|----------------------------------------|-----------------------|------------------------------|---------------------------------|---------------------------------|-------------------|
| $u(d)$ | Difference between reading and reference values | 0,0065 mm | 1 | 0,0065 mm | 6 |
| $u(l_s)$ | Calibration of the standard bar | 0,0079 mm | 1 | 0,0079 mm | 102 |
| $u(\alpha_s)$ | Thermal expansion coefficient of the standard bar | $1,2 \cdot 10^{-6} \, ^\circ C^{-1}$ | 155 mm°C | 0,0002 mm | 50 |
| $u(\theta_s)$ | Temperature deviation from reference value | 0,41 °C | 0 | 0 | $\infty$ |

**E2.6 Reporting the result**

Based on the results shown in table 3, the combined standard uncertainty, $u_c(l)$, of the length reading is determined from expression (9), corresponding to 0,010 mm, with 32 effective degrees of freedom.

Considering an interval having a level of confidence of approximately 95% in a Student t distribution, the expansion factor is 2,04, which results in an expanded measurement uncertainty of

$$U_{95\%}(l) = k \cdot u_c(l) = 2,04 \cdot 0,010 \, mm = 0,021 \, mm.$$  (16)

**E2.7 Interpretation of results**

Table 3 shows that the calibration of the reference standard bar is the major contribution to the output measurement uncertainty, followed closely by the measured difference between reading and reference values. The remaining uncertainty components have a negligible contribution to the combined measurement uncertainty.

If no correlation effect was considered in the measured difference between reading and reference values, $u(d)$ would increase to 0,007 mm and $U_{95\%}(l)$ would be slightly higher (0,022 mm). Therefore, if the correlation between measurements performed in different positions of the reference standard bar is not considered, the expanded measurement uncertainty of the calibration is only overestimated by approximately 5%.
Although the uncertainty components related to the thermal influence on the performed measurements were considered negligible, some significant considerations can be made based on the established probabilistic formulation and calculation method.

For instance, suppose a steel bar (characterized by a thermal expansion coefficient estimate of \(11.5 \cdot 10^{-6} \, ^\circ\text{C}^{-1}\), with the same standard uncertainty as mentioned before) were used as the measurement standard instead of the carbon fibre composite bar (with a null estimate). The expanded measurement uncertainty would then increase by 19%.

In a similar way, if the estimate of the temperature deviation from the reference temperature would increase to 2 °C, keeping the same measurement uncertainty as before, this would be reflected in a 4.5% increase of the calibration expanded measurement uncertainty.

References


Example E3

Conformity assessment of an influenza medication as a multicomponent material

F. Pennecchi, M.G. Cox, P.M. Harris, A.M.H. van der Veen and S.L.R. Ellison

E3.1 Summary

The main goal of the present study is to show how to calculate risks of false decisions in the conformity assessment of a multicomponent material, taking into account both the measurement uncertainties and the covariances for the measured content values of the components. As a case study, a particular influenza medication (NyQuil tablets) is here considered.

E3.2 Introduction of the application

Medicinal products are typical examples of multicomponent materials, since they are made of several active compounds and excipients. Conformity assessment has to be performed in the content of each of its components. However, even when conformity assessment is successful for all the components individually and relevant consumer's and producer's risks are acceptable, the total probability of a false decision (total risk) on the conformity of the material as a whole might still be significant.

A IUPAC Project [31] was dedicated to the modelling of total risks of false decisions due to measurement uncertainty for multicomponent materials or objects. The mathematical framework was developed as a generalization of that suggested in [13] for conformity assessment of a single item. For this reason, the notation used here is consistent with the notation in [13] and [32], the latter being a relevant paper in which the reader can find more details on this case study.

E3.3 Specification of the measurands

This case study concerns test results for NyQuil tablets [1], a cold medication containing four active components:

- acetaminophen (APAP), as a pain reliever and fever reducer;
- dextromethorphan hydrobromide (DEX), as a cough suppressant;
Example E3. Conformity assessment of an influenza medication as a multicomponent material

- doxylamine succinate (DOX), as an antihistamine and hypnotic;
- phenylephrine hydrochloride (PE), as a nasal decongestant.

The measurands are the content values $c_i, i = 1, \ldots, 4,$ of the components of the tested medication tablets. Corresponding measured values (test results) $c_{im}$, obtained according to the test method described in [32], undergo conformity assessment. Quantities are masses of the components in a tablet expressed as a fraction (%) of the corresponding labelled amount $l_i$. Labelled amounts are $l_1 = 325$ mg for APAP, $l_2 = 10$ mg for DEX, $l_3 = 6.25$ mg for DOX, and $l_4 = 5$ mg for PE, respectively, per tablet (a tablet mass is 775 mg on average).

### E3.4 Measurement uncertainty and correlations

A full uncertainty budget for the test results of the components’ content is available in [32]. Relative measurement uncertainty is evaluated as 2.8% of $c_{im}$.

A total of 105 lots of the medication produced and released at the same factory during a year are tested in the same laboratory belonging to the factory. Linear correlation among the test results for different components is estimated by the Pearson’s correlation coefficients $r_{ij}$ [9, sec. C.3.6], $i < j$. Only APAP test results are not significantly correlated with the other component contents’, whereas test results for the low-dose active components – DEX, DOX and PE – show to be significantly correlated (at a 99% level of confidence) [32]. Correlation coefficients are reported in table E3.1.

<table>
<thead>
<tr>
<th>Component</th>
<th>Index</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>APAP</td>
</tr>
<tr>
<td>i/j</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>DEX</td>
</tr>
<tr>
<td>i</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>0.107</td>
</tr>
<tr>
<td>i</td>
<td>DOX</td>
</tr>
<tr>
<td>2</td>
<td>0.125</td>
</tr>
<tr>
<td>i</td>
<td>PE</td>
</tr>
<tr>
<td>3</td>
<td>0.177</td>
</tr>
<tr>
<td></td>
<td>PE</td>
</tr>
<tr>
<td>4</td>
<td>0.311</td>
</tr>
<tr>
<td></td>
<td>0.404</td>
</tr>
<tr>
<td></td>
<td>0.539</td>
</tr>
<tr>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

### E3.5 Specification or tolerance limits

The lower and upper tolerance limits, $T_{Li}$ and $T_{Ui}$, for the product release are 95.0% and 105.0% of the labelled amount $l_i$ for each active component, $i = 1, \ldots, 4.$ The tolerance limits derive from regulatory authorities controlling the quality of marketed medicinal products.

### E3.6 Decision rule and conformity assessment

In the present case study, the “simple acceptance”, or “shared risk”, rule is considered as the decision rule for conformity assessment [13, sec. 8.2.1], i.e., acceptance limits of test results coincide with tolerance limits ($A_{Li} = T_{Li}$ and $A_{Ui} = T_{Ui}$).

The producer of the medication is the pharmaceutical company producing the drug, whereas the consumer is any individual who may take that medication. In the present example, only the calculation of consumer’s risks is shown, but the counterpart models for the producer’s risks are easily obtainable as well.
Example E3. Conformity assessment of an influenza medication as a multicomponent material

E3.6.1 Bayesian framework

In the framework of the IUPAC project [31], evaluation of total risks of false decision for multicomponent materials is based on a multivariate version of the evaluation of specific and global risks for a single characteristic of an item, as defined in [13, sec. 9.3.2 and 9.5.2]. The underlying Bayesian approach requires defining a multivariate prior probability density function (PDF) \( g_0(c) \) for “true” values of the components’ content, where \( c = [c_1, \ldots, c_4] \), and a multivariate likelihood function \( h(c_m | c) \) for the corresponding test results, where \( c_m = [c_{1m}, \ldots, c_{4m}] \).

As discussed in [32], a multivariate normal distribution is used for modelling both the prior knowledge and the likelihood function. The former multivariate normal PDF, \( g_0(c) \), has vector mean \( m = [m_1, \ldots, m_4] \), where \( m_i \) is the \( i \)th experimental sample mean, calculated from the available production data (see table E3.2), and covariance matrix \( S \), made by terms \( S_{c_{ij}} = r_{ij} s_i s_j \), where \( r_{ij} \) are the correlation coefficients in table E3.1 and \( s_i \) is the \( i \)th experimental standard deviation (see table E3.2). For each fixed vector value \( c \), the multivariate normal PDF modelling the likelihood function \( h(c_m | c) \) has vector mean \( c \) and covariance matrix \( S_{cm} \), made by terms \( S_{cm_{ij}} = r_{ij} u_i u_j \), where \( u_i = 0.028 c_{im_i} \), % of labelled amount, is the \( i \)th associated standard uncertainty. The same correlation coefficients are used for both the prior PDF and the likelihood function since it is supposed that no further correlation effect is attributable to the analytical measurement process: just the correlation between “true” values, maybe due to technological conditions in the production of the medication, is effective and induces, consequently, a correlation between the corresponding test results.

<table>
<thead>
<tr>
<th>Component</th>
<th>Index</th>
<th>Mean % of labelled amount</th>
<th>Standard deviation % of labelled amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>APAP</td>
<td>1</td>
<td>99.18</td>
<td>1.37</td>
</tr>
<tr>
<td>DEX</td>
<td>2</td>
<td>97.70</td>
<td>1.02</td>
</tr>
<tr>
<td>DOX</td>
<td>3</td>
<td>99.33</td>
<td>1.05</td>
</tr>
<tr>
<td>PE</td>
<td>4</td>
<td>98.94</td>
<td>1.22</td>
</tr>
</tbody>
</table>

Table E3.2: Experimental mean and standard deviation of the components’ content values of 105 lots of the medication

E3.6.2 Total specific risk

For a vector of test results \( c_m \) of a specific multicomponent item, when all the \( c_{im_i} \) are measured within their own acceptance interval and hence the material is accepted as conforming, the total specific consumer’s risk \( R_{i}^{tot} \) is defined as the probability of at least one of the “true” \( c_i \) values of the components’ content being outside its tolerance interval. Therefore, it is calculated as one minus the probability that all the “true” values \( c_i \) are inside their tolerance interval. Such a probability is provided by the posterior probability density function integrated over the multivariate tolerance domain \([T_{L1}, T_{U1}]×[T_{L2}, T_{U2}]×[T_{L3}, T_{U3}]×[T_{L4}, T_{U4}]\). The integral can be obtained by calculation of the corresponding cumulative distribution function at the desired limits.

In the current study, since both prior \( g_0(c) \) and likelihood function \( h(c_m | c) \) are modelled by multivariate normal PDFs, also the joint posterior function for the “true” components’ content values results in a multivariate normal PDF with covariance matrix \( S_{post} \) and vector of posterior

1Standard deviations \( s_i \) are smaller than measurement uncertainties \( u_i \), since each released lot has passed several quality tests (any out-of-specification test result preventing the lot release), whereas 2.8 % is a target relative standard uncertainty (hence, the actual measurement standard uncertainty may be smaller).
Example E3. Conformity assessment of an influenza medication as a multicomponent material

means \( c_{\text{post}} \) respectively equal to [27] eq. 3.13:

\[
\begin{align*}
S_{\text{post}} & = \left( S_c^{-1} + n_{\text{rep}} S_{cm}^{-1} \right)^{-1} \\
c_{\text{post}} & = S_{\text{post}} \left( S_c^{-1} \bar{m} + n_{\text{rep}} S_{cm}^{-1} \bar{c}_m \right),
\end{align*}
\]

(E3.1)

(E3.2)

where \( \bar{m} \) is the vector of the prior mean values, \( \bar{c}_m \) is the vector of the arithmetic means of replicated measurement/test results and \( n_{\text{rep}} \) is the number of such replicates (in this study, since each component is measured once, \( n_{\text{rep}} = 1 \) and \( \bar{c}_m = [c_{1m}, \ldots, c_{4m}] \)).

Considering, for example, the special case in which all the test results \( c_{im} \) are exactly equal to the corresponding prior mean values \( m_i \), the total specific consumer’s risk is \( R_{\text{tot}}^* = 0.0029 \). When \( c_{im} = T_{Li} \) for each \( i \), hence \( R_{\text{tot}}^* = 0.0117 \); when \( c_{im} = T_{Ui} \), \( R_{\text{tot}}^* = 0.0002 \). Details of the calculation can be found in the code file A121_Medicine_total_specific_risk.r, where the “pmvnorm”\(^2\) function from the R package “mvtnorm” is used for the calculation of the posterior cumulative distribution.

E3.6.3 Total global risk

The total global consumer’s risk \( R_{\text{tot}}^* \) is defined as the probability that test results \( c_{im} \) of all the components’ contents of an item, drawn at random from the item population, are in their respective acceptance intervals and at least one of the corresponding “true” values \( c_i \) is outside its tolerance interval. Such probability is the integral of the joint multivariate PDF of “true” and test results, which is given by the product \( g_0(c) h(c_m|c) \). It can be calculated by a Monte Carlo (MC) simulation in which, for each vector \( c \) randomly drawn from \( g_0(c) \), a corresponding vector \( c_m \) is drawn from \( h(c_m|c) \). Hence, the total risk is approximated by the frequency of cases in which, within randomly generated vectors \( [c_m, c] \), all the \( c_{im} \) values are within their respective acceptance intervals but at least one \( c_i \) value is outside its tolerance interval.

In the present study, for a number \( N = 10^7 \) of MC simulations, such risk value is equal to \( R_{\text{tot}}^* = 0.0018 \), being numerically stable up to the fourth decimal digit. Details of the calculation are in the code file A121_Medicine_total_global_risk.r. The obtained result is slightly different from that reported in [32] \( (R_{\text{tot}}^* = 0.0019) \), which was obtained by a composition of several probability terms, arranged according to the law of total probability, each calculated by the “adaptIntegrate” function of the R package “cubature”.

E3.7 Interpretation of results

The above-reported values of total specific risk are for illustrative purposes. Value \( R_{\text{tot}}^* = 0.0029 \) means that, whenever test results coincided with prior mean values, for instance, there would be a probability of 0.29% of selling a nonconforming product, in the sense that at least one of the “true” values of the components’ content would be actually out-of-specification. The dependence of total specific risk on the test result of a particular component at a time (while the other \( c_{im} \) values are fixed and equal to the prior mean values) is depicted in [32], showing that the risk behaviour is not easily predictable.

The obtained total global risk \( R_{\text{tot}}^* = 0.0018 \) indicates that, out of 10 000 tablets chosen at random from the whole medication production, 18 of them might be assessed as conforming without actually being (i.e., presenting conforming test results for all the four component contents, while actually having at least an out-of-specification “true” value).

---

\(^2\)The absolute error of the reported values, provided as an output of the function, is about \( 10^{-5} \) for \( R_{\text{tot}}^* = 0.0029 \) and for \( R_{\text{tot}}^* = 0.0117 \), and \( 10^{-6} \) for \( R_{\text{tot}}^* = 0.0002 \).
Example E4

Evaluation of measurement uncertainty in thermal comfort

J.A. Sousa, A.S. Ribeiro, M.G. Cox, L.L. Martins
E4 Evaluation of measurement uncertainty in thermal comfort

J. A. Sousa¹, A. S. Ribeiro², M. G. Cox³, L. L. Martins²

¹ IPQ, Portuguese Institute for Quality, Rua António Gião, 2, 2829-513 Caparica, Portugal
² LNEC, National Civil Engineering Laboratory, Av. do Brasil, 101, 1700-066 Lisboa, Portugal
³ NPL, National Physical Laboratory, Hampton Road, Teddington, TW11 0LW, UK

E4.1 Summary
The Monte Carlo method for uncertainty evaluation is particularly suitable to handle the complexity of the mathematical model that specifies the relation between the quantities involved in the evaluation of thermal comfort. The standard ISO 7730:2005 is the main document in this field and, besides the application and uncertainty evaluation, the limitations of this standard will also be discussed.

E4.2 Introduction to the application
The example, which is based in part on the paper [10], is concerned with the evaluation of thermal comfort as defined in the international standard ISO 7730:2005 – the condition of mind that expresses the degree of satisfaction with the thermal environment [8], which inevitably differs from person to person and thus entails a probabilistic approach. The main parameter to be evaluated is a thermal comfort index named predicted mean vote (PMV), which predicts the average thermal sensation of a large group of persons exposed to the same environment, based on principles of heat balance and experimental data collected in a controlled climate chamber under steady-state conditions.

Although the PMV formula is widely recognized and adopted, little has been done to establish measurement uncertainties associated with its use, bearing in mind that the formula depends on measured values and tabulated values given to limited numerical accuracy. Knowledge of these uncertainties is invaluable when values provided by the formula are used in making decisions in various health and civil engineering situations. Energy efficiency is an example where thermal perception plays an important role in influencing the thermal performance of buildings, which in turn has enormous impact on energy consumption worldwide.

This example examines this formula, providing a general mechanism for evaluating the uncertainties associated with values of the quantities on which the formula depends. Further, consideration is given to the propagation of these uncertainties through the formula to provide the uncertainty associated with the value obtained for the index. Current international guidance on uncertainty evaluation is utilized and discussed.

Alternative approaches are discussed, e.g., using raw data from enquiries on thermal comfort, to overcoming the limitation of the coarse resolution imposed by the standard ISO 7730 on the thermal sensation felt by a specific individual. Consideration is given to the possibility of using a continuous scale, thus introducing comparability of the scale used in a possibly modified ISO 7730 and an enquiries-based scale.
E4.3 Specification of the measurand

The PMV index is given in ISO 7730 by the following mathematical function of eight quantities:

\[
\text{PMV} = [0.303 \exp(-0.036M) + 0.028] \\
\times [M - W - 3.05 \times 10^{-3}(5733 - 6.99(M - W) - \rho_a)] \\
-0.42(M - W - 58.15) - 1.7 \times 10^{-5}(5867 - \rho_a)M - 0.0014(34 - t_a)M \\
-3.96 \times 10^{-8}[(t_{cl} + 273)^4 - (t_r + 273)^4]f_{cl} - (t_{cl} - t_a)f_{cl}h_c(t_{cl})],
\]

where

\begin{align*}
M & \text{ – metabolic rate in W·m}^{-2}, \\
W & \text{ – effective mechanical power in W·m}^{-2}, \\
\rho_a & \text{ – water-vapour partial pressure in Pa}, \\
t_a & \text{ – air temperature in °C}, \\
f_{cl} & \text{ – clothing surface area factor}, \\
t_{cl} & \text{ – clothing surface temperature in °C}, \\
t_r & \text{ – mean radiant temperature in °C}, \\
h_c & \text{ – convective heat transfer coefficient in W·m}^{-2} K^{-1},
\end{align*}

with the main complication arising from the fact that the quantity \(t_{cl}\) is defined implicitly (see expression (2) below).

The model is clearly non-linear, and depends on (a) fundamental quantities \(M, W\) and \(I_{cl}\) obtained from tables \((t_{cl}\) and \(f_{cl}\) each depend on \(I_{cl}\), the clothing insulation in m²·K·W⁻¹), and (b) quantities \(t_a, t_r, v_{ar}\) and RH obtained by measurement \((v_{ar}\) is the relative air velocity in m·s⁻¹ that influences \(h_c\), and RH is the relative humidity in % that influences \(\rho_a\)).

The expressions involved are quite complicated as described in the next section.

Measurement model

To be able to calculate the PMV index we need to specify the input quantities. The clothing surface temperature is defined implicitly and in terms of other input quantities:

\[
t_{cl} = 35.7 - 0.028(M - W) \\
- I_{cl} \{3.96 \times 10^{-8}[(t_{cl} + 273)^4 - (t_r + 273)^4]f_{cl} + (t_{cl} - t_a)f_{cl}h_c(t_{cl})\},
\]

where the convective heat transfer coefficient is

\[
h_c(t_{cl}) = \max(2.38|t_{cl} - t_{cl}|^{1/4}, 12.1\sqrt{v_{ar}}),
\]

which is mathematically identical to the expression given in the Standard [8], but simpler. We write \(h_c(t_{cl})\) in this equation rather than simply \(h_c\) to emphasize that it depends on \(t_{cl}\), a parameter already existing explicitly in the model.

The clothing surface area factor depends on the clothing insulation:

\[
f_{cl} = \begin{cases} 
1.00 + 1.290I_{cl}, & I_{cl} \leq 0.0775 \text{ m}^2 \cdot \text{K} \cdot \text{W}^{-1}, \\
1.05 + 0.645I_{cl}, & I_{cl} > 0.0775 \text{ m}^2 \cdot \text{K} \cdot \text{W}^{-1}.
\end{cases}
\]

In the cases considered in this study \(\rho_a\) or, more precisely, \(\rho_w(t_a)\), the water-vapour partial pressure, may be obtained from measurements of the relative humidity RH = 100\(\rho_a/\rho_s(t_a)\), using
Example E4. Evaluation of measurement uncertainty in thermal comfort

\[ \rho_a = RH \times \rho_s(t_a) \text{10 Pa}, \]  

where \( \rho_a(t_a) \) is the water-vapour saturation pressure given by the function

\[ \rho_s(t_a) = \exp \left( 16.6536 - \frac{403.0183}{t_a + 235} \right) \text{K \cdot Pa}. \]  

The degree of complexity can be slightly reduced if advantage is taken of the fact that the term

\[ g(t_{cl}) = 3.96 \cdot 10^{-8}[(t_{cl} + 273)^4 - (t_r + 273)^4]f_{cl} + (t_{cl} - t_a)f_{cl}h_c(t_{cl}) \]  

is common to Equations (1) and (2). Again, we indicate \( g \) as a function of \( t_{cl} \) to emphasize that, given values for the other quantities involved, the value of \( g(t_{cl}) \) can readily be obtained knowing \( t_{cl} \). Therefore, Equation (2) can be expressed as

\[ t_{cl} = 35.7 - 0.028(M - W) - l_dg(t_{cl}), \]  

and Equation (1) can be simplified accordingly to

\[ \text{PMV} = [0.303 \exp(-0.036M) + 0.028] \times \{M - W - 3.05 \times 10^{-5}[5733 - 6.99(M - W) - \rho_a] \]  

\[ -0.42(M - W - 58.15) - 1.7 \cdot 10^{-5}(5867 - \rho_a)M - 0.0014(34 - t_a)M \]  

\[ -[35.7 - 0.028(M - W) - t_{cl}]/l_{cl}. \]

Nevertheless, the above expressions remain complicated and working with them using the GUM approach (LPU - Law of Propagation of Uncertainty) [2] is not only questionable from the point of view of the assumptions entailed by that approach, but difficult to implement since it requires the calculation of partial derivatives within an implicit non-trivial formulation.

Concerning the associated uncertainties, the metabolic rate \( M \), the effective power \( W \) and the clothing insulation \( l_d \) were defined according to the conditions and tables given in ISO 7730 and therefore their values have no associated uncertainty. The experimental data for the measured quantities were obtained from calibrated instruments (traceable to national standards with reported measurement uncertainties). The data were used in a thermal comfort study developed for a health institution and obtained from three locations: two offices with different indoor environmental conditions and a customer service room.

For each location the testing procedure included the measurement of the following quantities: air temperature \( t_a \), globe temperature\(^1 \) \( t_g \), relative humidity \( RH \), relative air velocity \( v_{ar} \), and mean radiant temperature \( t_r \). Table 1 gives estimates of the measured quantities as average values of 40 observations (obtained every 2 min).\(^2 \) Their associated uncertainties are shown in Table 2. These standard uncertainties comprise contributions from the averaging process (Type A evaluation of uncertainty) and instrument calibration (Type B evaluation).

\[ ^1 \text{Measured with a globe thermometer as a means of assessing the combined effects of radiation, air temperature and air velocity on human comfort.} \]

\[ ^2 \text{It is assumed that the measuring conditions are stable during the period of measurement and the observations can be regarded as independently obtained values of the quantities.} \]
Example E4. Evaluation of measurement uncertainty in thermal comfort

Table 1 Estimates of the input quantities

<table>
<thead>
<tr>
<th>Location</th>
<th>$M\times58.2/\text{Wm}^{-2}$</th>
<th>$W/\text{Wm}^{-2}$</th>
<th>$I_{cl}\times0.155/\text{m}^{-2}\text{K}^{-1}$</th>
<th>$t_a/\degree\text{C}$</th>
<th>$t_g/\degree\text{C}$</th>
<th>RH/%</th>
<th>$v_{ar}/\text{m}\cdot\text{s}^{-1}$</th>
<th>$t_r/\degree\text{C}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Office 1</td>
<td>1.2</td>
<td>0</td>
<td>0.7</td>
<td>25.8</td>
<td>26.3</td>
<td>47.4</td>
<td>0.01</td>
<td>26.4</td>
</tr>
<tr>
<td>Office 2</td>
<td>1.2</td>
<td>0</td>
<td>1.0</td>
<td>20.9</td>
<td>21.2</td>
<td>68.1</td>
<td>0.02</td>
<td>21.3</td>
</tr>
<tr>
<td>Customer service</td>
<td>1.2</td>
<td>0</td>
<td>0.7</td>
<td>24.0</td>
<td>24.3</td>
<td>46.4</td>
<td>0.07</td>
<td>24.6</td>
</tr>
</tbody>
</table>

Table 2 Standard uncertainties associated with the estimates of the measured quantities in Table 1

<table>
<thead>
<tr>
<th>Location</th>
<th>Type</th>
<th>$t_a/\degree\text{C}$</th>
<th>$t_g/\degree\text{C}$</th>
<th>RH/%</th>
<th>$v_{ar}/\text{m}\cdot\text{s}^{-1}$</th>
<th>$t_r/\degree\text{C}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Office 1</td>
<td>A</td>
<td>0.0</td>
<td>0.0</td>
<td>0.1</td>
<td>0.00</td>
<td>0.0</td>
</tr>
<tr>
<td>Office 2</td>
<td>A</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.00</td>
<td>0.0</td>
</tr>
<tr>
<td>Customer service</td>
<td>A</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.01</td>
<td>0.0</td>
</tr>
<tr>
<td>All</td>
<td>B</td>
<td>0.1</td>
<td>0.1</td>
<td>0.5</td>
<td>0.05</td>
<td>0.2</td>
</tr>
</tbody>
</table>

The type A entries in the table are given to the same number of decimal places as the Type B entries. As a result, many of the non-zero entries are reported as zero.

Another important comment refers to the difference between the definitions and use of standard uncertainty in the GUM and GUM Supplement 1 (GUM-S1) [3]. In the former the sample standard deviation is used as the standard uncertainty whereas in the latter a factor of \([n-1]/(n-3)]^{1/2}\) is included to obtain the standard deviation of the state-of-knowledge distribution assigned to the corresponding input quantity. In this case, for a sample size of \(n = 40\) the effect is small (less than 3%), not affecting the standard uncertainties in Table 2 to the number of decimal digits reported.

**Uncertainty propagation**

Two approaches for the evaluation of uncertainty are considered: the GUM uncertainty framework, based on LPU [2], and the propagation of distributions of GUM-S1, based on the Monte Carlo method [3].

Both approaches depend on knowledge of the probability density functions (PDFs) for the input quantities, but whereas the GUM uncertainty framework uses summary information – estimates and associated standard uncertainties – obtained from the PDFs, the propagation of distributions uses the PDFs themselves. The simplification inherent in the GUM approach, however, imposes limitations on its applicability, which are irrelevant to the GUM-S1 approach, making the latter more reliable and which should be used for validation, when the conditions for the use of the GUM approach are not fully met.

**GUM Uncertainty Framework**

The GUM uncertainty framework requires the calculation of sensitivity coefficients $c_i$, the first partial derivatives of the PMV index measurement function with respect to the quantities on which the function depends, evaluated at the estimates of those quantities. These derivatives are determined from expression (9). As for many complicated models, determining the required partial derivatives algebraically is not always practical and a numerical approach is recommended [9]. It is a burden not shared by the GUM-S1 approach.

In this model, special attention must be paid to the derivative $\partial\text{PMV}/\partial t_{cl}$ given values for all other input quantities. Equation (8) [with Equation (7)] is solved for $t_{cl}$, and then PMV is evaluated.
using Equation (9). The partial derivatives of PMV with respect to the input quantities, required by LPU, are formed in the usual manner apart from the partial derivative of PMV with respect to $t_{cl}$. The fact that PMV defined by Equation (1) involves $g(t_{cl})$, defined by Equation (7), and hence the derivative $g'(t_{cl})$ is required, which is not the case when using Equation (9), is a simplification that poses no problem since $g(t_{cl})$ is already used when solving Equation (8) numerically for $t_{cl}$. No further numerical operations are necessary in evaluating the partial derivative.

Another aspect of this non-trivial model is that there are two instances where estimates of the input quantities are close to the breakpoints (derivative discontinuities) of the respective model functions. They relate to the clothing surface area factor $f_{cl}$ [expression (4)] and to the convective heat transfer coefficient $h_c$ [expression (3)].

The first of these model functions (the second is similar) is illustrated in Figure 1, showing $f_{cl}$ as a function of clothing insulation $I_{cl}$. Equation (4) can be also be expressed as

$$f_{cl} = 1.1 + \min\{0.645(I_{cl} - 0.0775), 1.290(I_{cl} - 0.0775)\},$$

which displays explicitly the fact that the function is continuous at $I_{cl} = 0.0775$ m$^2$·K·W$^{-1}$ with $f_{cl}$ taking the value 1.1. The sensitivity coefficient $\partial f_{cl}/\partial I_{cl}$ changes from 1.290 to 0.645 at $I_{cl} = 0.0775$ m$^2$·K·W$^{-1}$, halving its value.

Figure 1: Clothing surface area factor $f_{cl}$ as a function of clothing insulation $I_{cl}$

Figure 2 shows the PDF for $f_{cl}$ when $I_{cl}$ is assigned a Gaussian PDF $N(0.0775, (0.01)^2)$, as produced by a Monte Carlo calculation. It is generally far easier to use such a calculation to provide (or at least approximately) a PDF, even though an analytical solution could also be obtained using the "change of variables" formula [11], applying the formula separately in both branches. The PDF obtained agrees to graphical accuracy with that provided by the Monte Carlo calculation. The PDF in this case is discontinuous, in fact a mixture of two "half-Gaussians", with standard deviations in the ratio 2:1.
Even when the measurement model involves derivative discontinuities, as here, the standard deviation of the measurand ($f_{cl}$ or subsequently PMV index) is a continuous function of the input quantities in the model. Figure 3 shows the standard deviation of $f_{cl}$ (equal to the standard uncertainty associated with an estimate of $f_{cl}$) for $I_{cl}$ ranging from 0.001 m$^2$·W·K$^{-1}$ to 0.200 m$^2$·W·K$^{-1}$. The smooth but rapid change of the standard deviation from 0.0129 to 0.0065 over this interval is apparent. Convection plays a predominant role in the thermal comfort perception, as expected.

**Reporting the result**

Table 3 shows for the PMV model the input quantities, the PDFs that characterizes them, and their estimates and associated standard uncertainties. Some of the input quantities are experimental, while other quantities have tabulated values [8] for which values are regarded as fixed and exact. Together with the above partial derivatives evaluated at the estimates of the input quantities, LPU is applied to produce results given in the table. In the case of $t_a$ the PDF results from the combination of data from measurement and instrument calibration.

A relevant conclusion that can be drawn from Table 3 is that relative air velocity is a dominant factor in the perception of thermal comfort (expressed as PMV) in this particular case, which is a well-known phenomenon. In this model its influence is through quantity $h_c$. The PDF for $h_c$ is very sensitive to relative air velocity and its shape is like that of the output quantity PMV, which corroborates the finding on the predominant role of relative air velocity in the perception of
thermal comfort. The influence of $h_c$ can also be concluded from a sensitivity analysis of the quantities present in Equation (1).

The Monte Carlo calculation was based on the experimental input indicated in Tables 1 and 2. The number of Monte Carlo trials was taken as $10^5$. Thus, samples of $10^5$ drawn from the PDFs for the input quantities were used to obtain a PDF for the output quantity as described in GUM-S1 [3].

<table>
<thead>
<tr>
<th>Quantity</th>
<th>PDF</th>
<th>Estimate</th>
<th>Standard uncertainty</th>
<th>$c_t$</th>
<th>$c_t$(PMV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M/W \cdot m^{-2}$</td>
<td>Ref. value</td>
<td>70</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
</tr>
<tr>
<td>$W/W \cdot m^{-2}$</td>
<td>Ref. value</td>
<td>0</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
</tr>
<tr>
<td>$t_a/{}^\circ C$</td>
<td>Combined</td>
<td>22.0</td>
<td>0.1</td>
<td>0.228</td>
<td>0.013</td>
</tr>
<tr>
<td>$I_{cl}/m^2 \cdot K \cdot W^{-1}$</td>
<td>Ref. value</td>
<td>0.078</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
</tr>
<tr>
<td>$t_r/{}^\circ C$</td>
<td>Gaussian</td>
<td>22.0</td>
<td>negligible</td>
<td>$-$</td>
<td>$-$</td>
</tr>
<tr>
<td>$RH/%$</td>
<td>Rectangular</td>
<td>60.0</td>
<td>0.3</td>
<td>0.0059</td>
<td>0.0017</td>
</tr>
<tr>
<td>$v_{ar}/m \cdot s^{-1}$</td>
<td>Rectangular</td>
<td>0.10</td>
<td>0.03</td>
<td>3.27</td>
<td>0.094</td>
</tr>
<tr>
<td>PMV</td>
<td></td>
<td>$-0.75$</td>
<td>$\mu$(PMV) = 0.094</td>
<td>$U_{0.95}(PMV) = 0.18$</td>
<td></td>
</tr>
</tbody>
</table>

Table 3 GUM uncertainty budget for the PMV model

Figure 4 shows the PDF for the PMV index provided by the Monte Carlo (MC) method for the same location as Table 3. The striking asymmetry in the PDF is evident, with a very long right-hand tail. This long tail implies there is non-negligible probability of having a different (higher) value for the PMV index than would have been obtained by applying the GUM with its assumption that the measurand is Gaussian.

The advantages of MC over GUM are apparent in this application. The latter only delivers an estimate (expectation), an associated standard uncertainty (standard deviation) and a coverage interval based on the assumption of normality. MC gives considerable insight, providing much richer information, through the display of any given shape for the PDF for the measurand, allowing characteristics such as the tails to be considered.

Results related to the figure are presented in Table 4 in which "Estimate" is taken as the expectation of the PDF for the corresponding measurand. We note that this parameter can be
somewhat misleading in the case of an extremely asymmetric PDF although it does indeed formally give the expectation (mean) of the distribution. The mode (point at which the probability density is greatest) might be more meaningful, but it is recommended that only the expectation is used for purposes of uncertainty propagation [2].

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Estimate</th>
<th>Standard uncertainty</th>
<th>95 % coverage interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>PMV</td>
<td>0.23</td>
<td>0.04</td>
<td>0.18 0.34</td>
</tr>
</tbody>
</table>

A complementary study was carried out to provide a sensitivity analysis for the input parameters used to obtain the PMV index. For this purpose, small variations of the input quantities were introduced successively, keeping all other input quantities fixed at their estimates, to approximate the partial derivatives near the measurement point. The analysis showed that relative air velocity has the largest impact. As pointed out before, the convective heat transfer coefficient also has a significant impact on the perception of thermal comfort, and thus any changes related to air temperature and air velocity impact appreciably on the perception of thermal comfort.

The same analysis was applied to those quantities whose values were taken from tables [8]. The partial derivatives and the effect of uncertainty related to these quantities (based on assuming an error magnitude of at most one half in the last stated decimal place) on PMV index uncertainty showed non-linear behaviour in the neighbourhood of the estimates of the input quantities, which is another reason in favour of the application of Monte Carlo. In relation to the relative influence of uncertainty contributions, the sensitivity analysis, assuming the Gaussian PDF $\mathcal{N}(1.2, (0.05)^2)$ for $M$ and the Gaussian PDF $\mathcal{N}(0.7, (0.05)^2)$ for $I_{cl}$, enabled it to be concluded that the variation of $M$ has little influence on PMV, whereas $I_{cl}$ greatly influences it. Care should thus be taken when selecting these tabular values from Standard ISO 7730 [8], especially those related to clothing insulation.

An important point to be made relates to the validation of the Monte Carlo implementation, which was made using five reference test sets taken from Table D1 in Annex D of ISO 7730 [8], to cover various testing conditions. The comparison showed strong agreement between the reference values and the values provided by GUM-S1.

Comparing the results for the estimate of PMV and its associated standard uncertainty as provided by the GUM uncertainty framework and Monte Carlo, using two of those reference test sets, showed a surprising close agreement between both approaches, which is not always the case. Bearing in mind that the opposite can also occur, it emphasizes the need for validation whenever possible: any difference may affect decision making in conformity assessment in thermal comfort.

This decision making has an immediate application in this standard if thermal environments are to be classified in various categories, as in Annex A of [8]. A detailed evaluation of measurement uncertainty applied to PMV index is required, since this parameter will have a direct impact on the classification to be attributed to a specific thermal environment, by affecting the possible values of PMV, with the corresponding consequences on value and suitability of different building spaces. Detailed information on conformity assessment can be found in JCGM 106 [5].
E4.4 Interpretation of results
Since this project EMUE is being developed within the framework of pre-normative projects, it is important to evaluate aspects not only of the standard ISO 7730 [8], but also of standard ISO 7726 [7], which relates closely to the former standard by specifying the requirements for the accuracy of the measuring instruments used in ISO 7730.

In terms of ISO 7730, the PMV index is defined on a continuous scale, whereas the interpretation is often translated into a discrete scale, on a 7-point thermal sensation scale (−3/cold, −2/cool, −1/slightly cool, 0/neutral, +1/slightly warm, +2/warm, +3/hot), values relating to the subjective thermal perception of a large sample of individuals exposed to the same thermal conditions.

On one hand, the resolution of this scale is too coarse, leading to an increase in uncertainty that would be totally artificial. For example, if a value of PMV is halfway between two points on its seven-point scale (≈ 0.5, say), a substantial increase in the uncertainty of a rounded PMV value over and above the PMV uncertainty based on a continuous scale would occur.

On the other hand, based on experience, few individuals vote for the extreme values of the discrete scale, the majority concentrating their votes on central values. Thus, the scale should probably be changed to increase the resolution of the scale on the central part of acceptable thermal condition. However, the treatment of ordinal scales is a branch of science on its own, using e.g., Rasch models [1], and close collaboration should therefore be pursued with researchers from the social sciences.

With respect to ISO 7726, one study has shown [12] that of the two possible requirements for the measurement instruments, namely "required" and "desirable", only the latter permits to obtain the accuracy assumed for the values used in ISO 7730, and therefore these points should be conveyed to the standardization committee responsible for these documents.
Example E4. Evaluation of measurement uncertainty in thermal comfort

E4.5 Annex: GUM uncertainty framework and GUM-S1 propagation of distributions

Explicit Model

In the more common explicit univariate measurement model, a single real output quantity \( Y \) is related to a number of input quantities \( X = (X_1, \ldots, X_N)^T \) by a functional relationship \( f \) in the form \( Y = f(X) \), as stated in the GUM [2]. The estimate of the output quantity is taken as \( y = f(x) \).

The standard uncertainty \( u(y) \) associated with \( y \) is evaluated from

\[
\begin{align*}
\sum_{i=1}^{N} \sum_{j=1}^{N} c_i u(x_i, x_j) c_j,
\end{align*}
\]

where \( c_i \) is the partial derivative \( \partial f / \partial X_i \) evaluated at \( X = x \) and is known as the \( i \)th sensitivity coefficient, \( u(x_i) \) is the standard uncertainty associated with \( x_i \), and \( u(x_i, x_j) \) the covariance associated with \( x_i \) and \( x_j \).

A compact way of writing the sum in expression (10), better suited for scientific software based on matrix formulation, e.g., MATLAB, is

\[
\begin{align*}
u^2(y) = c^T V_x c,
\end{align*}
\]

where \( V_x \) is the covariance matrix of dimension \( N \times N \) containing the covariances \( u(x_i, x_j) \),

\[
\begin{align*}V_x = \begin{bmatrix} u(x_1, x_1) & \cdots & u(x_1, x_N) \\ \vdots & \ddots & \vdots \\ u(x_N, x_1) & \cdots & u(x_N, x_N) \end{bmatrix},\end{align*}
\]

and the (row) vector \( c^T = [c_1, \ldots, c_N] \) of dimension \( 1 \times N \) contains the sensitivity coefficients. Both expressions (10) and (11) are equivalent representations of LPU of the GUM [2].

For independent input quantities, we would obtain the better-known simplified expression (equivalent to using \( V_x \) with its off-diagonal elements replaced by zeros)

\[
\begin{align*}
u^2(y) = \sum_{i=1}^{N} [c_i u(x_i)]^2 = \sum_{i=1}^{N} u_i^2(y),
\end{align*}
\]

The \( u_i(y) \) are often used in uncertainty budgets to identify which input quantities, with respect to their corresponding standard uncertainties, have significant influence on the standard uncertainty \( u(y) \) associated with the estimate \( y \) of the output quantity.

Implicit Model

For an implicit univariate measurement model, however, a single output quantity \( Y \) is related to real input quantities \( X \) in a way that cannot readily or stably be represented by a direct functional relationship. Instead, a model for the measurement takes the form \( h(Y, X) = 0 \), in which \( Y \) is not expressed directly as a function of \( X \), often requiring a numerical implementation to obtain a solution [4].

The estimate \( y \) of \( Y \) is the value of \( \eta \) that solves the equation \( h(\eta, x) = 0 \). This equation is to be solved numerically with a suitable zero-finding algorithm [9]. The standard uncertainty \( u(y) \) associated with \( y \) is evaluated from

\[
\begin{align*}u^2(y)c_y^2 = c_x^T U_x c_x,\end{align*}
\]
where $c^T x$ is the (row) vector of dimension $1 \times N$ of partial derivatives $\partial h / \partial X_i$, and $c_y$ is the partial derivative $\partial h / \partial Y$, with all derivatives evaluated at $X = x$ and $Y = y$ [2].

Conditions for valid application
There are a number of conditions for valid application of the GUM uncertainty framework for non-linear models. They include [3] that $f$ must be continuously differentiable with respect to the elements $X_i$ of $X$ in the neighbourhood of the estimates $x_i$ of the $X_i$, for all derivatives up to the appropriate order, and that higher-order terms not included in the Taylor series approximation to $f(X)$ are negligible. The differentiation issue was treated above.

Propagation of distributions
The most general and reliable approach for uncertainty propagation is the propagation of distributions, where the PDFs for the input quantities are propagated through the measurement model to provide the PDF for the output quantity. The expectation of this PDF is then used as the estimate of the measurand and the standard deviation of the PDF is used as the standard uncertainty associated with that estimate.

A Monte Carlo method is an implementation of the propagation of distributions. It still requires a functional relationship, but it does not suffer from the limitations imposed by the GUM uncertainty framework, namely the differentiable issues, the compliance with the central limit theorem, the requirement of symmetrical input PDFs, Gaussian output PDF or the non-existence of a non-Gaussian dominant source of uncertainty. It should provide valid results, provided an adequate number of samples is drawn, whenever the applicability of the GUM uncertainty framework is questionable. It should always be checked that any given target uncertainty has been attained [6], which is a further feature the GUM uncertainty framework cannot provide.

Once the PDF for the output quantity $Y$ is available, a coverage interval for $Y$ corresponding to any stipulated coverage probability $p$ can be obtained. Commonly, $p$ is taken as 0.95. Such a coverage interval contains the value of $Y$ with probability $p$. A straightforward method for obtaining a coverage interval from the results of applying a Monte Carlo method is to sort the values of $Y$ in non-decreasing order and use the percentiles to obtain the required interval. The shortest 95% coverage interval includes values with the highest density and can be obtained by the procedure given in [3].

The GUM uncertainty framework does not provide the PDF for $Y$, but instead assumes that $Y$ can be described by a Gaussian PDF $N(y, u^2(y))$, namely, with expectation $y$ and standard deviation $u(y)$ (or a scaled and shifted t-distribution). Specifically, the GUM defines a coverage interval for $Y$ as $y \pm U_p$, where $U_p$ is an expanded uncertainty corresponding to coverage probability $p$ given by $U_p = k_p u(y)$. The factor $k_p$ is known as a coverage factor, which is obtained from the standard Gaussian PDF or the t-distribution [1].
E4.6 References


Example E4. Evaluation of measurement uncertainty in thermal comfort

Examples of evaluating measurement uncertainty

First edition (M18)
Example E5

Evaluation of measurement uncertainty in SBI- Single burning item reaction to fire test

E5 Evaluation of measurement uncertainty in SBI – Single Burning Item reaction to fire test

L. L. Martins¹, A. S. Ribeiro¹, M. G. Cox², J. A. Sousa³, D. Loureiro¹, M. C. Almeida¹, M. A. Silva¹, R. Brito¹, A. C. Soares¹

¹ LNEC, National Laboratory for Civil Engineering, Av. do Brasil, 101, 1700-066 Lisbon, Portugal
² NPL, National Physical Laboratory, Hampton Road, Teddington, TW11 0LW, UK
³ IPQ, Portuguese Institute for Quality, Rua António Gião, 2, 2829-513 Caparica, Portugal

E5.1 Summary
This example illustrates the application of the Monte Carlo Method (MCM) in measurement uncertainty propagation related to the SBI – Single Burning Item test, within the European normative framework of reaction to fire tests for building products, namely, the EN standard 13823:2010+A1 [1]. The use of the MCM is justified by the multivariate, non-linear and complex nature of the functional relations between a large number of input, intermediate and output quantities, thus providing a numerical approach to the validation of the GUM method [2] described in [3].

E5.2 Introduction to the application
The objective of the SBI standard test [1] is to measure a set of quantities which determine the evaluation and classification of a construction material (excluding floorings), aiming to characterize its contribution to the deflagration and propagation of fires in buildings, when exposed to adverse thermal conditions by means of a combustion item.

In this test, the specimen retrieved from the tested material is composed of two plates vertically positioned with a 90° angle between both plates, being exposed to a main burner located in the lower region of the plate’s junction. The specimen’s performance is evaluated for a period of 20 minutes, based on the indirect measurement of quantities related to heat release and smoke production. Complementary observations are also performed regarding lateral flame propagation and the production of drops or particles from the combustion process.

E5.3 Specification of the measurands
Two main measurands are defined in the SBI test: the heat release rate, HRR, which corresponds to the thermal power released in a given time instant (expressed in kW) and, in a similar way, the smoke production rate, SPR, both being related to the combustion of the specimen (expressed in m²·s⁻¹).

Due to the applied test method [1], the definitions of these quantities are related to different time periods of the SBI test, namely:

\[ HRR(t) = \begin{cases} 
0, & t \leq 300 \text{ s}, \\
\max\{0, HRR_{\text{total}}(t) - HRR_{\text{burner}}\}, & 300 \text{ s} < t \leq 312 \text{ s}, \\
HRR_{\text{total}}(t) - HRR_{\text{burner}}, & 312 \text{ s} < t \end{cases} \]  

(1)

where HRR_{total}(t) is the total thermal power released by the specimen and the main burner in the time instant t, while HRR_{burner} is the average thermal power released only by the main burner; and
Example E5. Evaluation of measurement uncertainty in SBI- Single burning item reaction to fire test

SPR(t) = \begin{cases} 
0, & t \leq 300 \text{ s}, \\
\max[0, SPR_{total}(t) - SPR_{burner}], & 300 \text{ s} < t \leq 312 \text{ s}, \\
SPR_{total}(t) - SPR_{burner}, & 312 \text{ s} < t,
\end{cases} \tag{2}

where SPR_{total}(t) is the total smoke production rate of the specimen and the main burner in the time instant t, while SPR_{burner} is the average smoke production rate related only to the main burner.

In both cases, the initial stage of the SBI test time period (between 210 s and 270 s) is used to determine the quantities HRR_{burner} and SPR_{burner}, based on average values obtained when combustion occurs only in an auxiliary burner (identical to the main burner) installed in the experimental apparatus.

The heat release rate is a key intermediate quantity in the determination of two main output quantities of the SBI test – THR, the total heat release (usually expressed in MJ) from the specimen in a certain time exposure to the main burner flames (namely, in the first 600 s), and FIGRA, the fire growth rate (expressed in W\cdot s^{-1}), and defined as the maximum value of the quotient of heat release rate from the specimen and the time of its occurrence using a THR threshold (such as 0.2 MJ or 0.4 MJ).

In a similar way, the smoke production rate is also a significant intermediate quantity in the SBI test since it contributes for the determination of two other main output quantities – TSP, the total smoke production (in m^2) from the specimen in a certain time exposure to the main burner flames (namely, in the first 600 s), and SMOGRA, the smoke growth rate (expressed in m^2\cdot s^{-2}), which is defined as the maximum value of the quotient of smoke production rate from the specimen and the time of its occurrence.

This example only addresses the measurement uncertainty evaluation of the quantities heat release rate and smoke production rate, since the posterior uncertainty propagation from these intermediate key quantities to the output quantities of the SBI test (total heat release, fire growth rate, total smoke production and smoke growth rate) is straightforward and characterized by simple linear mathematical models. Both the heat release rate and the smoke production rate quantities are indirectly measured, in a given time instant, based on a large number of input quantities and mathematical models, as described in the following sections.

**Measurement model**

The heat release rate measurement model is derived from the studies performed by [4] in the oxygen consumption calorimetry research field. The measurement principle states that the amount of heat released per unit of consumed oxygen volume, E', during a combustion process (in MJ\cdot m^{-3}) is considered constant regardless of the combustion material, which can be expressed by

\[ HRR = E' \cdot x_{O_2}^0 \cdot \varphi \cdot q_{VS}, \tag{3} \]

where x_{O_2}^0 is the oxygen molar fraction in the ambient, \varphi is the oxygen depletion factor and q_{VS} is the volumetric flow of air in the ambient (expressed in m^3\cdot s^{-1}).

The amount of heat released per unit of consumed oxygen volume quantity can be determined by the product between the oxygen density, \rho_{O_2} (in kg\cdot m^{-3}) and the heat release per unit of consumed oxygen mass, E (in MJ\cdot kg^{-1}), i.e.

\[ E' = \rho_{O_2} \cdot E. \tag{4} \]
The oxygen molar fraction in the ambient is given by

\[ x_{O_2}^0 = x_{O_2}^{\text{initial}} \cdot (1 - x_{H_2O}^0), \]  

\[ x_{O_2}^{\text{initial}} \] being the oxygen molar fraction measured in the initial stage of the SBI test (in the time period between 30 s and 90 s), with a gas analyser, and \( x_{H_2O}^0 \) the water molar fraction in the ambient, which can be determined by the following model (derived from the Clausius-Clapeyron equation for water vapour saturation pressure and based on conventional values for the gas constant and the heat vaporization of water)

\[ x_{H_2O}^0 = \frac{r_h}{100} \frac{p_{\text{atm}}}{T_{\text{initial}}} \exp \left[ 23.2 - \frac{3816}{T_{\text{initial}}-46} \right], \]

where \( r_h \) is the relative humidity in moist air (as a percentage), \( T_{\text{initial}} \) is the initial air temperature inside the exhaust duct (in K) and \( p_{\text{atm}} \) is the atmospheric pressure (in Pa).

The oxygen depletion factor is calculated by

\[ \varphi = \frac{x_{O_2}^{\text{initial}} (1-x_{CO_2}) - x_{O_2}^{\text{final}} (1-x_{CO_2})}{x_{O_2}^{\text{final}} (1-x_{O_2} - x_{CO_2})}, \]

where \( x_{O_2}^{\text{initial}} \) and \( x_{CO_2}^{\text{initial}} \) are, respectively, the molar fractions of oxygen and carbon dioxide measured in the initial stage of the SBI test with the gas analyzer, while \( x_{O_2} \) and \( x_{CO_2} \) are, respectively, the molar fractions of oxygen and carbon dioxide measured with the same equipment in a given time instant after the initial stage.

The volumetric flow of air in the ambient is indirectly measured based on the expression

\[ q_{V_c} = \frac{q_{V_c}}{1 + (\alpha - 1) \varphi} \]

in which \( q_{V_c} \) is the volumetric flow of the gases in the exhaust duct (in m³·s⁻¹) and \( \alpha \) represents the expansion factor. This last quantity reflects the fact that, in a combustion chemical reaction, the amount of substance related to combustion products is not identical to the amount of substance related to the oxygen consumed in the reaction process, i.e.

\[ \alpha = 1 + x_{O_2}^0 \cdot (\beta - 1), \]

where \( x_{O_2}^0 \) is obtained from expression (5) and \( \beta \) is the ratio between the amount of substance of combustion products and of consumed oxygen. A combustion reaction involving reactants such as hydrocarbons \( (C_nH_{2n+2}O_{n}X_{e}) \) and oxygen \( (O_2) \) originates products such as carbon dioxide \( (CO_2) \), water \( (H_2O) \), hydrates \( (HX) \), carbon monoxide \( (CO) \) and nitrogen \( (N_2) \), the overall chemical reaction formula being given by

\[ C_nH_{2n+2}O_{n}X_{e} + \left( \alpha - \frac{a}{2} + \frac{b-e}{4} - \frac{c}{2} \right)O_2 \rightarrow \left( \alpha - c \right)CO_2 + \frac{b-e}{2}H_2O + eHX + gCO + \frac{d}{2}N_2, \]  

1 This equipment receives a gas sample from a normalized exhaust duct in which all water vapour and water-soluble gases are eliminated before measurement.
where particular constants \( a \) to \( g \) apply in any specific instance. Therefore, based on expression \( (10) \) and by definition, the \( \beta \) ratio is given by

\[
\beta = \frac{4a+2b+2e+2d}{4a+b-e-2c-2g} \tag{11}
\]

Depending on the type of hydrocarbon subjected to combustion, several estimates are known for the \( \beta \) ratio, usually values between one and two.\(^2\)

The volumetric flow of the gases in the exhaust duct is obtained by the expression

\[
q_v = \frac{k_t}{k_p} \cdot \sqrt{\frac{2\Delta p}{\rho_{T_0}}} \cdot T_0 \cdot A, \tag{12}
\]

where \( \Delta p \) is the differential pressure measured in a bidirectional pressure sensor located inside the exhaust duct (in Pa); \( \rho_{T_0} \) is the moist air density\(^3\) for a reference temperature, \( T_0 \), equal to 298.15 K; \( T \) is the gas temperature in the exhaust duct (in K); \( A \) is the area (in m\(^2\)) of the exhaust duct circular cross-section; \( k_p \) is the differential pressure correction factor; and \( k_t \) is the global correction factor.

Since the exhaust duct as a circular cross-section, its area corresponds to

\[
A = \frac{\pi}{4} \cdot d^2, \tag{13}
\]

where \( d \) is the exhaust duct diameter (in m).

For the quantification of the moist air density (considering the reference temperature \( T_0 \) in K), the following expression [5] is used:

\[
\rho_{T_0} = \frac{0.34848 \cdot p_{atm} - 0.009024 \cdot r_h \cdot \exp[0.0612 \cdot (T_0 - 273.15)]]}{T_0}. \tag{14}
\]

The differential pressure correction factor is considered in expression (12) due to the use of a bidirectional sensor [6] instead of a conventional Pitot tube (vulnerable to solid particles in the flow). This quantity is defined by

\[
k_p = \frac{\sqrt{\frac{\Delta p}{\rho_{T_amb}}}}{v_c}, \tag{15}
\]

where \( v_c \) is the flow velocity in the centre of the exhaust duct cross-section (in m\( \cdot \)s\(^{-1}\)) and \( \rho_{T_amb} \) is the moist air density [5] for ambient temperature, \( T_{amb} \) (in K) given by

\[
\rho_{T_{amb}} = \frac{0.34848 \cdot p_{atm} - 0.009024 \cdot r_h \cdot \exp[0.0612 \cdot (T_{amb} - 273.15)]]}{T_{amb}}. \tag{16}
\]

\(^2\)Examples of \( \beta \) ratio estimates for the combustion of: carbon (C, \( \beta = 1 \)); ethylene (C\(_2\)H\(_4\), \( \beta = 1.3 \)); propene (C\(_3\)H\(_6\), \( \beta = 1.3 \)); butane (C\(_4\)H\(_10\), \( \beta = 1.4 \)); heptane (C\(_7\)H\(_{16}\), \( \beta = 1.4 \)); propane (C\(_3\)H\(_8\), \( \beta = 1.4 \)); ethane (C\(_2\)H\(_6\), \( \beta = 1.5 \)); hydrogen (H\(_2\), \( \beta = 2 \)).

\(^3\)Since the density of the gas mixture inside the exhaust duct is unknown, this quantity is assumed to be close to the moist air density (expressed in kg\( \cdot \)m\(^{-3}\)).
The global correction factor, $k_t$, corresponds to the average of three individual corrections, $k_{t,v}$, $k_{t,\text{propane}}$, $k_{t,\text{heptane}}$, related to the periodic testing of the SBI experimental apparatus aiming, respectively, at the determination of the non-uniformity of the flow velocity in the exhaust duct and the comparison between experimental and theoretical heat release rate values, concerning the combustion of known pure substances such as propane and heptane.

In the case of the $k_{t,v}$ correction, its quantification is supported by

$$k_{t,v} = \frac{\sum_{i=1}^{5} v_i}{5 \cdot v_c}, \quad (17)$$

considering the average flow velocities measured in the $i$ radius of the exhaust duct, $v_i$, and in its centre, $v_c$, all these quantities being expressed in m·s$^{-1}$.

The $k_{t,\text{propane}}$ correction is expressed by the ratio between the theoretical and the experimental heat release rate values of the propane combustion (in kW) respectively, $\text{HRR}_{i}^{\text{theoretical}}$ and $\text{HRR}_{i}^{\text{experimental}}$, i.e.

$$k_{t,\text{propane}} = k'_t \cdot \frac{\sum_{i=1}^{5} \text{HRR}_{i}^{\text{theoretical}}}{\sum_{i=1}^{5} \text{HRR}_{i}^{\text{experimental}}}, \quad (18)$$

considering the several testing steps indexed by $i$ of this normalized test [1], where $k'_t$ is the global correction used in the experimental determination of the heat release rate$^5$. The theoretical heat release rate at the $i$th step is given by

$$\text{HRR}_{i}^{\text{theoretical}} = q_{m_i} \cdot |\Delta h_{c}^{l,\text{propane}}|,$$

where $q_{m_i}$ is the propane mass flow in the $i$th testing step (expressed in kg·s$^{-1}$), and $|\Delta h_{c}^{l,\text{propane}}|$ is the low enthalpy of propane combustion per unit of mass (in kJ·kg$^{-1}$). It should be noticed that, in the calculation of $\text{HRR}_{i}^{\text{experimental}}$ by expression (3), the heat released per unit of consumed oxygen volume adopts a specific estimate and measurement uncertainty known for propane, instead of the value mentioned in [3] and used for construction materials in the SBI test.

Regarding the $k_{t,\text{heptane}}$ correction, this quantity is obtained from the expression

$$k_{t,\text{heptane}} = k'_t \cdot \frac{|\Delta h_{c}^{l,\text{heptane}}| m_{\text{heptane}}}{\text{THR}}, \quad (20)$$

where $k'_t$ is the global correction used in the experimental determination of the total heat release$^6$, THR, (in MJ) during the heptane combustion test, $|\Delta h_{c}^{l,\text{heptane}}|$ is the low enthalpy of heptane combustion per unit of mass (in kJ·kg$^{-1}$) and $m_{\text{heptane}}$ is heptane mass used as burning combustible (in kg). As in the case of propane combustion, the estimate and measurement uncertainty of the heat released per unit of consumed oxygen volume adopts known values for heptane, when using expression (3).

---

4 The measurement sample is composed of four velocity measurements in each of five normalized distances from the exhaust duct centre, in addition to four velocity measurements at the centre.
5 This quantity is also included in $\text{HRR}_{i}^{\text{experimental}}$; therefore, it can be removed from expression (18).
6 This quantity is also included in the quantification of THR; therefore, it can be removed from expression (20).
The smoke production rate, SPR, is defined in a similar way to the heat release rate quantity. However, its measurement is based on the light attenuation phenomenon resulting from the presence of smoke in an optical path. In this case, the measurement model corresponds to

\[
SPR = \frac{q_{Ve}}{L} \cdot \frac{T}{T_0} \cdot \ln \left( \frac{1}{\tau} \right),
\]

(21)

where \(q_{Ve}\) is the volumetric flow of gases in the exhaust duct (in m\(^3\)s\(^{-1}\)), obtained from expression (12); \(L\) is the optical path length (in m); the factor \(T/T_0\) is a correction for the temperature difference between the gases in the exhaust duct, \(T\) (in K) and the reference temperature, \(T_0\), equal to 298.15 K; and \(\tau\) is transmittance, defined as the ratio between the luminous intensity measured in a given time instant and in the initial testing stage, \(I\) and \(I_0\), respectively. In the SBI test, the luminous intensity that reaches the photodetector installed in the exhaust duct, is considered proportional to the electrical tension between its terminals so that the transmittance quantity is determined by electrical tension measurements.

In order to improve the comprehension of the functional relations related to the presented measurement models, Figure 1 shows a schematic representation of the heat release rate calculation process, while Figure 2 refers to the smoke production rate. Particular attention is given to the global correction factor and to its calculation process, schematically represented in Figure 3.

**Uncertainty propagation**

The measurement uncertainty evaluation shown in this example is composed of two main stages: (i) the formulation stage, in which all the input quantities of the mathematical models involved in the measurements are identified and characterized, through the assignment of a probability density function (PDF) which better represents the dispersion of values related to its measurement; (ii) the calculation stage, from which the measurement uncertainty of the quantities of interest (heat release rate and smoke production rate) is obtained, based on the propagation of the measurement uncertainties of the input quantities through the above described mathematical models.

In the presented case, the MCM was used in the calculation stage [7, 8], justified by the multivariate, non-linear and complex nature of the functional relations between a large number of input, intermediate and output quantities. For this purpose, the Mersenne Twister pseudo-random number generator [9] was used to obtain numerical sequences with a typical dimension (number of trials) of \(10^6\), in order to give a good assurance in obtaining convergent solutions. In addition, validated computational tools for converting and sorting the generated numerical sequences were also used.

In the SBI test, the heat release rate quantity is measured in different test stages, firstly in the preliminary periodic combustion of propane and heptane and, afterwards, during the combustion of the tested specimen. The only significant difference is related to the heat release per unit of consumed oxygen mass quantity, which assumes different estimates and measurement uncertainties in each test case (propane, heptane or specimen combustion).

Table 1 shows the adopted probabilistic formulation of the input quantities required for the determination of the total heat release rate related to the combustion of a certain specimen, which already includes (in the global correction factor) the measurement uncertainty of the heat release rate measured in the propane and heptane combustions.
Example E5. Evaluation of measurement uncertainty in SBI - Single burning item reaction to fire test
Figure 2 Functional diagram of the smoke production rate quantity

Let $\Delta p$, $v_c$, $P_{ak}$, $\rho_{sm}$, $T_0$ or $T_{amb}$, $d$, $A$, $\tau$, $L$ and $T$ be the variables. The output variable is $SPR$.

See Figure 3
Example E5. Evaluation of measurement uncertainty in SBI - Single burning item reaction to fire test
### Table 1 Probabilistic formulation of the input quantities related to the heat release rate measurement

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Symbol</th>
<th>PDF</th>
<th>Estimate</th>
<th>Standard uncertainty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative humidity</td>
<td>rh</td>
<td>Gaussian</td>
<td>60.1 %</td>
<td>1.1 %</td>
</tr>
<tr>
<td>Atmospheric pressure</td>
<td>( p_{\text{atm}} )</td>
<td>Gaussian</td>
<td>101.4 kPa</td>
<td>0.2 kPa</td>
</tr>
<tr>
<td>Initial air temperature inside the exhaust duct</td>
<td>( T_{\text{initial}} )</td>
<td>Gaussian</td>
<td>288.3 K</td>
<td>0.1 K</td>
</tr>
<tr>
<td>Oxygen density</td>
<td>( \rho_{\text{O}_2} )</td>
<td>Gaussian</td>
<td>1.308 kg( \cdot )m(^{-3} )</td>
<td>0.003 kg( \cdot )m(^{-3} )</td>
</tr>
<tr>
<td>Heat released per unit of consumed oxygen mass</td>
<td>( E )</td>
<td>Gaussian</td>
<td>13.1 MJ( \cdot )kg(^{-1} )</td>
<td>0.3 MJ( \cdot )kg(^{-1} )</td>
</tr>
<tr>
<td>Initial molar fraction of oxygen</td>
<td>( x_{\text{O}_2}^{\text{initial}} )</td>
<td>Gaussian</td>
<td>0.2095</td>
<td>0.00004</td>
</tr>
<tr>
<td>Molar fraction of oxygen</td>
<td>( x_{\text{O}_2} )</td>
<td>Gaussian</td>
<td>0.2067</td>
<td>0.0002</td>
</tr>
<tr>
<td>Initial molar fraction of carbon dioxide</td>
<td>( x_{\text{CO}_2}^{\text{initial}} )</td>
<td>Gaussian</td>
<td>0.0003</td>
<td>0.000005</td>
</tr>
<tr>
<td>Molar fraction of carbon dioxide</td>
<td>( x_{\text{CO}_2} )</td>
<td>Gaussian</td>
<td>0.0018</td>
<td>0.0002</td>
</tr>
<tr>
<td>Exhaust duct diameter</td>
<td>( d )</td>
<td>Gaussian</td>
<td>0.315 m</td>
<td>0.001 m</td>
</tr>
<tr>
<td>Exhaust gas temperature</td>
<td>( T )</td>
<td>Gaussian</td>
<td>313.8 K</td>
<td>0.4 K</td>
</tr>
<tr>
<td>Ambient temperature</td>
<td>( T_{\text{amb}} )</td>
<td>Gaussian</td>
<td>288.6 K</td>
<td>0.7 K</td>
</tr>
<tr>
<td>Differential pressure</td>
<td>( \Delta p )</td>
<td>Gaussian</td>
<td>68.6 Pa</td>
<td>2.1 Pa</td>
</tr>
<tr>
<td>Flow velocity in the centre of the exhaust duct cross-section</td>
<td>( v_c )</td>
<td>Gaussian</td>
<td>9.6 m( \cdot )s(^{-1} )</td>
<td>0.7 m( \cdot )s(^{-1} )</td>
</tr>
<tr>
<td>Ratio between the amount of substance of combustion products and of consumed oxygen</td>
<td>( \beta )</td>
<td>Uniform</td>
<td>1.5</td>
<td>0.3</td>
</tr>
<tr>
<td>Global correction factor</td>
<td>( k_t )</td>
<td>Gaussian</td>
<td>0.77</td>
<td>0.02</td>
</tr>
</tbody>
</table>

Regarding the smoke production quantity, Table 2 presents the adopted probabilistic formulation of the input quantities which supported the MCM simulations.

### Table 2 Probabilistic formulation of the input quantities of the smoke production rate

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Symbol</th>
<th>PDF</th>
<th>Estimate</th>
<th>Standard uncertainty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative humidity</td>
<td>rh</td>
<td>Gaussian</td>
<td>60.1 %</td>
<td>1.1 %</td>
</tr>
<tr>
<td>Atmospheric pressure</td>
<td>( p_{\text{atm}} )</td>
<td>Gaussian</td>
<td>101.4 kPa</td>
<td>0.2 kPa</td>
</tr>
<tr>
<td>Ambient temperature</td>
<td>( T_{\text{amb}} )</td>
<td>Gaussian</td>
<td>288.6 K</td>
<td>0.9 K</td>
</tr>
<tr>
<td>Exhaust duct diameter</td>
<td>( d )</td>
<td>Gaussian</td>
<td>0.315 m</td>
<td>0.001 m</td>
</tr>
<tr>
<td>Optical path length</td>
<td>( L )</td>
<td>Gaussian</td>
<td>0.315 m</td>
<td>0.001 m</td>
</tr>
<tr>
<td>Transmittance</td>
<td>( \tau )</td>
<td>Gaussian</td>
<td>0.974</td>
<td>0.005</td>
</tr>
<tr>
<td>Exhaust gas temperature</td>
<td>( T )</td>
<td>Gaussian</td>
<td>313.8 K</td>
<td>0.4 K</td>
</tr>
<tr>
<td>Differential pressure</td>
<td>( \Delta p )</td>
<td>Gaussian</td>
<td>68.6 Pa</td>
<td>2.1 Pa</td>
</tr>
<tr>
<td>Flow velocity in the centre of the exhaust duct cross-section</td>
<td>( v_c )</td>
<td>Gaussian</td>
<td>9.6 m( \cdot )s(^{-1} )</td>
<td>0.7 m( \cdot )s(^{-1} )</td>
</tr>
<tr>
<td>Global correction factor</td>
<td>( k_t )</td>
<td>Gaussian</td>
<td>0.77</td>
<td>0.02</td>
</tr>
</tbody>
</table>
**Reporting the result**

The measurement uncertainties of the input quantities shown in Table 1 were propagated by the MCM to the intermediate quantities (results shown in Table 3) and, posteriorly, to the total heat release rate quantity (see Table 4 for an example of a 30 kW thermal power level). Figure 4 shows the output PDF obtained for the heat release rate quantity.

### Table 3 MCM simulation results for intermediate quantities in the calculation of the heat release rate

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Symbol</th>
<th>PDF</th>
<th>Estimate</th>
<th>Standard uncertainty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water vapour molar fraction</td>
<td>$x_{H_2O}^0$</td>
<td>Gaussian</td>
<td>0.014 6</td>
<td>0.000 4</td>
</tr>
<tr>
<td>Moist air density for ambient temperature</td>
<td>$\rho_{T_\text{amb}}$</td>
<td>Gaussian</td>
<td>1.180 kg\cdot m^{-3}</td>
<td>0.005 kg\cdot m^{-3}</td>
</tr>
<tr>
<td>Differential pressure correction factor</td>
<td>$k_p$</td>
<td>Gaussian</td>
<td>1.15</td>
<td>0.09</td>
</tr>
<tr>
<td>Expansion factor</td>
<td>$\alpha$</td>
<td>Gaussian</td>
<td>1.10</td>
<td>0.05</td>
</tr>
<tr>
<td>Volumetric flow of gases in the exhaust duct</td>
<td>$q_{V_c}$</td>
<td>Gaussian</td>
<td>0.55 m$^3$\cdot s^{-1}</td>
<td>0.04 m$^3$\cdot s^{-1}</td>
</tr>
<tr>
<td>Heat released per unit of consumed oxygen volume</td>
<td>$E'$</td>
<td>Gaussian</td>
<td>17.1 MJ\cdot m^{-3}</td>
<td>0.4 MJ\cdot m^{-3}</td>
</tr>
<tr>
<td>Ambient oxygen molar fraction</td>
<td>$x_{O_2}^0$</td>
<td>Gaussian</td>
<td>0.207 4</td>
<td>0.000 05</td>
</tr>
<tr>
<td>Oxygen depletion factor</td>
<td>$\varphi$</td>
<td>Gaussian</td>
<td>0.015</td>
<td>0.001</td>
</tr>
<tr>
<td>Ambient volumetric flow</td>
<td>$q_{V_s}$</td>
<td>Gaussian</td>
<td>0.55 m$^3$\cdot s^{-1}</td>
<td>0.04 m$^3$\cdot s^{-1}</td>
</tr>
</tbody>
</table>

### Table 4 MCM simulation results for the heat release rate

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Symbol</th>
<th>PDF</th>
<th>Estimate</th>
<th>Standard uncertainty</th>
<th>Required accuracy</th>
<th>Simulation accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heat release rate</td>
<td>HRR</td>
<td>Gaussian</td>
<td>30 kW</td>
<td>3 kW</td>
<td>0.5 kW</td>
<td>0.1 kW</td>
</tr>
</tbody>
</table>

![Figure 4 Output PDF of the heat release rate quantity](image)

Additional simulations were performed for higher thermal power levels (up to 250 kW), showing similar results. The obtained relative standard uncertainty varies between 8 % and 9 %.
The obtained results for the smoke production quantity are shown in Table 5 (intermediate quantities) and Table 6 (output quantity). Figure 5 shows the PDF obtained by the MCM for the smoke production rate.

**Table 5** MCM simulation results for intermediate quantities in the calculation of the smoke release rate

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Symbol</th>
<th>PDF</th>
<th>Estimate</th>
<th>Standard uncertainty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moist air density for ambient temperature</td>
<td>$\rho_{T_{amb}}$</td>
<td>Gaussian</td>
<td>1.220 kg m$^{-3}$</td>
<td>0.005 kg m$^{-3}$</td>
</tr>
<tr>
<td>Differential pressure correction factor</td>
<td>$k_p$</td>
<td>Gaussian</td>
<td>1.11</td>
<td>0.09</td>
</tr>
<tr>
<td>Volumetric flow of gases in the exhaust duct</td>
<td>$q_{V_{c}}$</td>
<td>Gaussian</td>
<td>0.55 m$^3$s$^{-1}$</td>
<td>0.04 m$^3$s$^{-1}$</td>
</tr>
</tbody>
</table>

**Table 6** MCM simulation results for the smoke production rate (0.05 m$^2$s$^{-1}$ level)

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Symbol</th>
<th>PDF</th>
<th>Estimate</th>
<th>Standard uncertainty</th>
<th>Required accuracy</th>
<th>Simulation accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smoke production rate</td>
<td>SPR</td>
<td>Gaussian</td>
<td>0.05 m$^2$s$^{-1}$</td>
<td>0.02 m$^2$s$^{-1}$</td>
<td>0.005 m$^2$s$^{-1}$</td>
<td>0.0005 m$^2$s$^{-1}$</td>
</tr>
</tbody>
</table>

Additional simulations were performed for higher smoke levels (up to 6.8 m$^2$s$^{-1}$), showing similar results. The obtained relative standard uncertainty varies between 9% and 12%.

**E5.4 Interpretation of results**

As seen in Figures 4 and 5, the output PDF of both the heat release and smoke production quantities have a geometrical shape close to a Gaussian PDF, which was expected since all the input quantities (with the exception of the ratio between the amount of substance of combustion products and of consumed oxygen, see Table 1) were taken as Gaussian. In terms of validation of results, Tables 4 and 6 show that the number of performed simulations allowed achieving a computational accuracy quite lower than the required accuracy needed to perform the SBI test. In this particular example, the major advantage of using the MCM, when compared with the GUM approach, relies on its greater simplicity and accuracy when dealing with a large number of input quantities.
Example E5. Evaluation of measurement uncertainty in SBI- Single burning item reaction to fire test

References


Bibliography


Examples of evaluating measurement uncertainty First edition (M18)


Examples of evaluating measurement uncertainty First edition (M18)


# Index

<table>
<thead>
<tr>
<th>C</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>11</td>
</tr>
<tr>
<td>C++</td>
<td>11</td>
</tr>
<tr>
<td>correlated input variables</td>
<td>15</td>
</tr>
<tr>
<td>covariance matrix</td>
<td>15</td>
</tr>
<tr>
<td>I</td>
<td>3</td>
</tr>
<tr>
<td>ISO 15189</td>
<td>3</td>
</tr>
<tr>
<td>ISO 17034</td>
<td>3</td>
</tr>
<tr>
<td>ISO/IEC 17025</td>
<td>3</td>
</tr>
<tr>
<td>L</td>
<td>3</td>
</tr>
<tr>
<td>law of propagation of uncertainty</td>
<td>13</td>
</tr>
<tr>
<td>multivariate</td>
<td>3</td>
</tr>
<tr>
<td>M</td>
<td>3</td>
</tr>
<tr>
<td>MATLAB</td>
<td>11</td>
</tr>
<tr>
<td>MCMC</td>
<td>17</td>
</tr>
<tr>
<td>Convergence</td>
<td>20</td>
</tr>
<tr>
<td>Validation</td>
<td>17</td>
</tr>
<tr>
<td>measurement model</td>
<td>3</td>
</tr>
<tr>
<td>multivariate</td>
<td>3</td>
</tr>
<tr>
<td>measurement uncertainty</td>
<td>11</td>
</tr>
<tr>
<td>molar mass</td>
<td>7</td>
</tr>
<tr>
<td>Monte Carlo method</td>
<td>3</td>
</tr>
<tr>
<td>adaptive procedure</td>
<td>4</td>
</tr>
<tr>
<td>computational efficiency</td>
<td>11</td>
</tr>
<tr>
<td>coverage factor</td>
<td>12</td>
</tr>
<tr>
<td>expanded uncertainty</td>
<td>12</td>
</tr>
<tr>
<td>multivariate</td>
<td>3</td>
</tr>
<tr>
<td>number of trials</td>
<td>11</td>
</tr>
<tr>
<td>reporting</td>
<td>5</td>
</tr>
<tr>
<td>P</td>
<td>3</td>
</tr>
<tr>
<td>Parameterisation</td>
<td>18</td>
</tr>
<tr>
<td>Pascal</td>
<td>11</td>
</tr>
<tr>
<td>Posterior</td>
<td>21</td>
</tr>
<tr>
<td>Prior</td>
<td>19</td>
</tr>
<tr>
<td>Improper</td>
<td>19</td>
</tr>
<tr>
<td>Informative</td>
<td>19</td>
</tr>
<tr>
<td>Proper</td>
<td>19</td>
</tr>
<tr>
<td>type B evaluation</td>
<td>19</td>
</tr>
<tr>
<td>Prior information</td>
<td>17</td>
</tr>
<tr>
<td>measurand</td>
<td>3</td>
</tr>
<tr>
<td>probability density function</td>
<td>3</td>
</tr>
<tr>
<td>input variable</td>
<td>3</td>
</tr>
<tr>
<td>output variable</td>
<td>3</td>
</tr>
<tr>
<td>Python</td>
<td>11</td>
</tr>
<tr>
<td>R</td>
<td>5</td>
</tr>
<tr>
<td>R (software environment)</td>
<td>5</td>
</tr>
<tr>
<td>random number generator</td>
<td>5</td>
</tr>
<tr>
<td>arcsine distribution</td>
<td>6</td>
</tr>
<tr>
<td>exponential distribution</td>
<td>6</td>
</tr>
<tr>
<td>gamma distribution</td>
<td>6</td>
</tr>
<tr>
<td>multivariate normal distribution</td>
<td>6</td>
</tr>
<tr>
<td>normal distribution</td>
<td>6</td>
</tr>
<tr>
<td>rectangular distribution</td>
<td>5</td>
</tr>
<tr>
<td>t distribution</td>
<td>5</td>
</tr>
<tr>
<td>trapezoidal</td>
<td>5</td>
</tr>
<tr>
<td>triangular distribution</td>
<td>5</td>
</tr>
<tr>
<td>RStan</td>
<td>17</td>
</tr>
<tr>
<td>S</td>
<td>13</td>
</tr>
<tr>
<td>sensitivity coefficient</td>
<td>13</td>
</tr>
<tr>
<td>analytic differentiation</td>
<td>14</td>
</tr>
<tr>
<td>numerical differentiation</td>
<td>13</td>
</tr>
</tbody>
</table>
Acronyms

**EA** European co-operation for Accreditation. 17

**GUM** Guide to the expression of Uncertainty in Measurement. iii, 1

**GUM-S1** GUM Supplement 1. 17

**JCGM** Joint Committee on Guides in Metrology. iii, 1

**LIMS** laboratory information management system. 2

**MCMC** Markov Chain Monte Carlo. 17, 18