Good practice in evaluating measurement uncertainty

Compendium of examples

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

12 November 2020

EMUE

Examples of Measurement
Uncertainty Evaluation

Good practice in evaluating measurement uncertainty

A.M.H. van der Veen[1], M.G. Cox[2], L.L. Martins[3], A.S. Ribeiro[3], J.A. Sousa[4],
D. Loureiro[3], M.C. Almeida[3], M.A. Silva[3], R. Brito[3], A.C. Soares[3], F. Pennecchi[5],
S. Demeyer[6], N. Fischer[6], O. Pellegrino[4], C. Elster[7], A. Bošnjaković[8], V. Karahodžić[8],
P.M. Harris[2], S.L.R. Ellison[9], S. Martens[7], K. Klauenberg[7], F. Rolle[5], M. Sega[5],
P.G. Spazzini[5], I. de Krom[1], J. Neukammer[7], S. Cowen[9], M. Čaušević[8], H. Meuzelaar[1],
K. Shirono[10], B. Mickan[7], C. Yardin[6], and J. Greenwood[11]

[1]VSL, Department of Chemistry, Mass, Pressure and Viscosity, Thijsseweg 11, 2629 JA Delft, the Netherlands
[2]National Physical Laboratory, Data Science Department, Hampton Road, Teddington, Middlesex, United Kingdom, TW11 0LW
[3]LNEC, National Laboratory for Civil Engineering, Av. do Brasil, 101, 1700-066 Lisbon, Portugal
[4]IPQ, Portuguese Institute for Quality, Rua António Gião, 2, 2829-513 Caparica, Portugal
[5]INRIM, Istituto Nazionale di Ricerca Metrologica, Strada delle Cacce 91, 10135 Torino, Italy
[6]LNE, Laboratoire national de métrologie et d'essais, 29 avenue Roger Hennequin 78197 Trappes Cedex, France
[7]PTB, Physikalisch-Technische Bundesanstalt, Braunschweig and Berlin, Germany
[8]IMBIH, Institute of Metrology of Bosnia and Herzegovina, Sarajevo, Bosnia and Herzegovina
[9]LGC, Laboratory of the Government Chemist, Queens Road, Teddington, TW11 0LY, UK
[10]NMIJ, National Metrology Institute of Japan, Japan
[11]UKAS, United Kingdom Accreditation Service, 2 Pine Trees, Chertsey Lane, Staines-upon-Thames TW18 3HR, UK

12 November 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 is 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.

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, nor Euramet makes any warranty 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

v

Contents

E3   Evaluation of measurement uncertainty in SBI – Single Burning Item reaction to fire test   **33**

L.L. Martins, A.S. Ribeiro, M.G. Cox, J.A. Sousa, D. Loureiro, M.C. Almeida, M.A. Silva, R. Brito, A.C. Soares

E4   Statistical reassessment of calibration and measurement capabilities based on key comparison results   **47**

K. Shirono, M.G. Cox

E5   Measurement uncertainty when using quantities that change at a linear rate — use of quartz He reference leaks to calibrate an unknown leak   **55**

J. Greenwood, M.G. Cox

E6   Factoring effects such as calibration corrections and drift into uncertainty evaluations   **77**

J. Greenwood, M.G. Cox, N. Fischer

E7   Conformity assessment of an influenza medication as a multicomponent material   87

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

E8   Conformity assessment of mass concentration of total suspended particulate matter in air   91

F. Pennecchi, F. Rolle, A. Allard, S.L.R Ellison

E9   Uncertainty evaluation for the quantification of low masses of benzo[a]pyrene   97

F. Pennecchi, F. Rolle, M. Sega, S.L.R. Ellison, A.M.H van der Veen

E10   Calibration of an analyser for NO_x using gas mixtures prepared with mass flow controllers   105

F. Pennecchi, F. Rolle, M. Sega, P.G. Spazzini, I. de Krom, A.M.H. van der Veen

E11   Calibration of a sonic nozzle as an example for quantifying all uncertainties involved in straight-line regression   115

S. Martens, K. Klauenberg, B. Mickan, C. Yardin, N. Fischer, C. Elster

E12 Bayesian evaluation of a between-bottle homogeneity study in the production of reference materials 127

A.M.H. van der Veen, S.L.R. Ellison

E13 Quantifying uncertainty when comparing measurement methods – Haemoglobin concentration as an example of correlation in straight-line regression 135

S. Martens, K. Klauenberg, J. Neukammer, S. Cowen, S.L.R. Ellison, C. Elster

E14 Calibration of a torque measuring system – GUM uncertainty evaluation for least-squares versus Bayesian inference 145

S. Martens, K. Klauenberg, C. Elster

E15 Evaluation of measurement uncertainty in the calibration of a mobile optical measurement system 153

L.L. Martins, A.S. Ribeiro, M.G. Cox, J.A. Sousa, D. Loureiro, M.C. Almeida, M.A. Silva, R. Brito, A.C. Soares

E16 Evaluation of measurement uncertainty in thermal comfort   159

J.A. Sousa, A.S. Ribeiro, M.G. Cox, L.L. Martins

E17 Preparation of calibration gas mixtures of NH$_3$ in nitrogen using permeation   171

M. Čaušević, H. Meuzelaar, A.M.H. van der Veen, M.G. Cox

E18 Pressure drop measurement   185

M. Čaušević, M.G. Cox, J. Greenwood

E19 Flow meter calibration using the master meter method   195

M. Čaušević, M.G. Cox, A.M.H. van der Veen

References   218

Chapter 1

Introduction

A.M.H. van der Veen, M.G. Cox

The evaluation of measurement uncertainty is an essential part of the experimenter's task to obtain for the quantity of interest, the measurand, a value and a stated uncertainty. The JCGM published a suite of documents covering various aspects of measurement uncertainty evaluation, expression and use [51–56]. 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 [55].

Many laboratories implement ISO/IEC 17025 [33] to underpin their competence. Producers of (certified) reference materials implement in many cases both ISO/IEC 17025 and ISO 17034 [27] for the same purpose. In proficiency testing, the requirements for demonstrating competence are laid down in ISO/IEC 17043 [15]. 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 [25].

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, yet it is 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 [149], 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

A.M.H. van der Veen, M.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 [33], ISO 15189 [20] and ISO 17034 [27] 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) [51] 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) [52], 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 [52] 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 [51] and further elaborated in GUM-S1 [52]. 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) [54] 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 [51]. 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 [59, 145]. 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 [149]. 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 [18]. 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 [52, clause 7]. 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 [52], and a multivariate output quantity, see GUM-S2 [54].

GUM-S1 [52, clause 6.4] describes the selection of appropriate probability density functions for the input quantities, thereby supplementing the guidance given in the GUM [51, 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 [103, 148]. 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 [52, clause 7.9] and since then an improved method has been developed and published [177].

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 [52, 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 [123], which is also implemented in many other programming environments, including MATLAB and MicroSoft Excel (since version 2010, see [127]). Based on this random number generator, there are generators available for a number of probability distributions [149].

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 [149]. It is available for Windows, MacOS and a variety of UNIX platforms (including FreeBSD and Linux) [150].

Users of Windows, MacOS, and a number of Linux distributions may also wish to download and install RStudio [155], 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 $M = 10\,000$:

```
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.

```
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:

```
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" [99] implementing among others the trapezoidal distribution, a package called "mvtnorm" [93] implementing the multivariate normal distribution (useful when some of the input quantities are dependent [52]), and a package called "triangle" [63] 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 [52, table 1] are available in R.

The arc sine distribution can be implemented as follows in R. According to GUM-S1 [52, clause 6.4.6.1], a U-shaped random variable X on the interval $[a, b]$ can be obtained through

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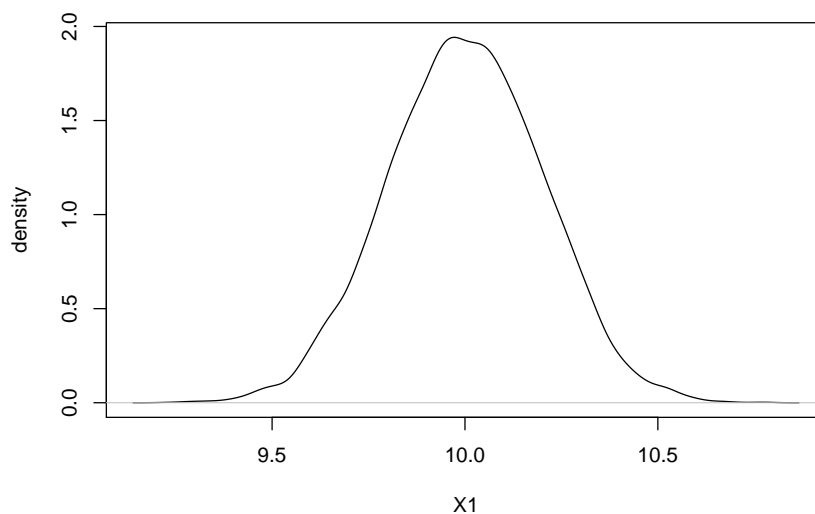
Figure 2.1: Density plot of the random variable X1 having a normal distribution with mean 10.0 and standard deviation 0.2

where Φ 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:

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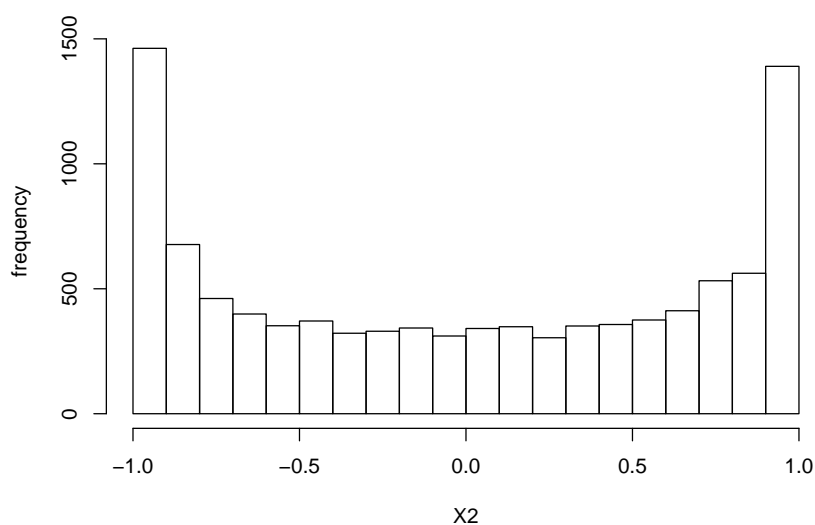
Figure 2.2: Histogram of the random variable X2 containing $M = 10\,000$ samples having an arcsine distribution between -1 and 1

of Pure and Applied Chemistry) [126]. The relative atomic masses that apply to "normal materials" are called standard atomic weights [69, 126]. Their interpretation is described in an IUPAC technical report [147].

The molar mass of phenol (chemical formula C_6H_5OH) 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

```
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 [126]. 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) [57]. 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).
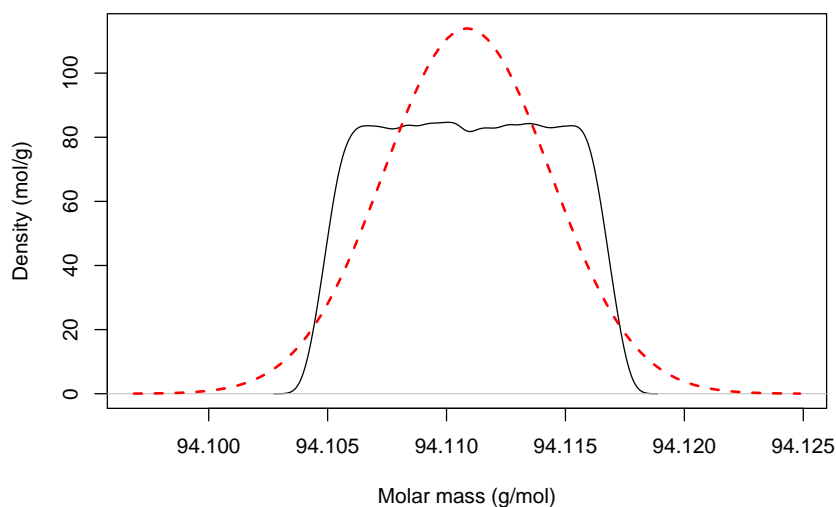
Figure 2.3: Output probability density function of the molar mass of phenol and superimposed a normal distribution with the same mean and standard deviation

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:

```
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 $94.1108\,\mathrm{g\,mol^{-1}}$ with standard uncertainty $0.0035\,\mathrm{g\,mol^{-1}}$. The expanded uncertainty is $0.0059\,\mathrm{g\,mol^{-1}}$. 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 [82] 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 [54]. 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 [82, S2]:

$$m_\mathrm{X} = m_\mathrm{S} + \delta m_\mathrm{D} + \delta m + \delta m_\mathrm{C} + \delta B, \tag{2.1}$$

where the symbols have the following meaning

m_X	conventional mass of the weight being calibrated,
m_S	conventional mass of the standard,
δm_D	drift of the value of the standard since its last calibration,
δm	observed difference in mass between the unknown mass and the standard,
δm_C	correction for eccentricity and magnetic effects,
δB	correction for air buoyancy.

For using the Monte Carlo method, probability density functions are assigned to each of the five input quantities [52]. These probability density functions are described in the original example [82].

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 [52, 6.4.7]. The drift of the mass of the standard weight δm_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 δm 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.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 [52]. An implementation of the Monte Carlo method with a fixed value for the number of samples M would then read as follows:

```
# 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 (`prob`) is defined to be 0.95. In accordance with the guidance in clause 7.2.2 of GUM-S1 [52], M is calculated using the built-in function `ceiling` which returns the smallest integer not less than its argument. With `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-prob)` is 20, hence the minimum number of Monte Carlo trials is $M = 10\,000 \cdot 20 = 200\,000$. 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`), δm_D (`dm.d`), δm_C (`dm.c`), and δB (`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:

```
print(mean(m.x),digits = 9)

## [1] 10000.025

print(sd(m.x),digits = 2)

## [1] 0.029

quantile(m.x,probs = c(0.025,0.975))

##      2.5%     97.5%
## 9999.968 10000.082
```

where the argument `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 [57]. 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:

```
# 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:

```
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:

```
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)
```
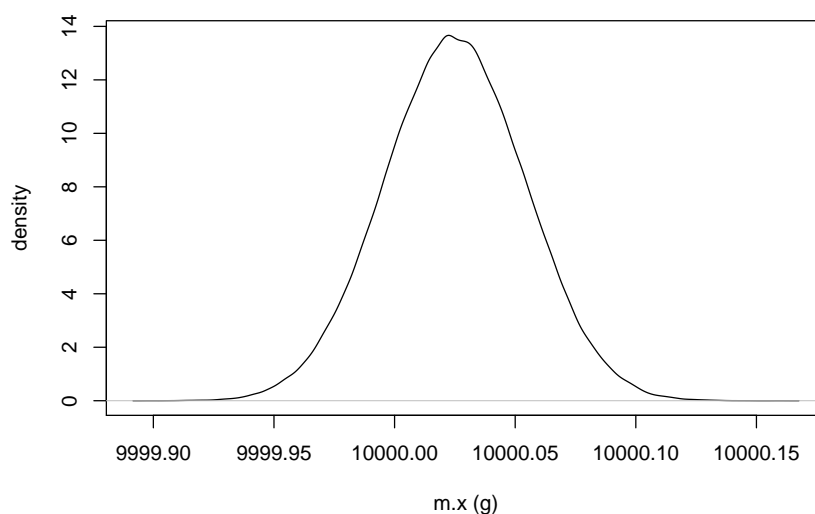
Figure 2.4: Probability density function of the output quantity m.x

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 [33]:

  – 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 [51].

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 [51], GUM-S2 [54] and EA 4/02 [82] 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 [164, 178]) is barely mentioned in such documents. Numerical approximation of the sensitivity coefficients [59, 145] 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 [82].

The R package `numDeriv` provides the function `grad` (from gradient) that returns from a function a generally good approximation, using Richardson extrapolation [61], of the partial derivatives of the input variables. The function returns a vector holding the values of these partial derivatives. The function passed to `grad` should have only one argument, namely a vector holding all input variables. Hence, the measurement model needs to be reformulated as follows:

```
# 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 `mass2.x` the same symbols have been used as in `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:

```
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 `numDeriv` (which needs to be installed in RStudio. The next two lines define the values of the input quantities. The vector `sens` on the fourth line holds the sensitivity coefficients returned by calling `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

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

and are identical to those given in EA 4/02 [82]. The code is also valid for measurement models with non-trivial sensitivity coefficients [145].

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 = diag(c(m.std.u,dm.d.u,diff.u,dm.c.u,dm.B.u))
CM = D %*% D
tmp = t(sens) %*% CM %*% sens
m.x.unc = sqrt(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 [54]) a covariance matrix of dimension 1×1 associated with the output quantity is returned (`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 `R` and `RStan`

A.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 `R` [149] and `rstan` [161], 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 [82] was introduced in chapter 2 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) [52]. 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 [51] or the Monte Carlo method of GUM Supplement 1 (GUM-S1) [52]. 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 [51, 82] 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 `R`, an environment for statistical computation [149], and the package `rstan` [64] that enables writing Bayesian models in a straightforward manner. This environment and the use of `RStan` for Bayesian inference have been introduced previously [170, 174].

    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 [52]. 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 [51, 82] 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 [82] 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 [52] (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 Δm_X denotes the departure from its nominal mass for the weight being calibrated, and Δm_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 [170, 174]. This sensitivity is one of the hurdles to be taken when performing iterative calculations (as MCMC is [92]).

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

```
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 [51] are declared as model parameters in the `parameters` block. By default, Stan assigns these variables a rectangular distribution over their domain of validity [64]. 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 [92]) . 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 Δm_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 [51] deals with type B evaluations of standard uncertainty is 'weakly Bayesian' [50] 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 (±15 mg for drift, and ±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 [64, 161].

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 [82]. 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 [51, 174]).

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 [92]. More details have been given elsewhere [170,174]. Running the model with 21 000 iterations and a warmup of 1000 iterations, using 4 chains yields the fol-

lowing output:

```
## Inference for Stan model: 091dba697d92e3c49746850cfc395085.
## 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. `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. `n_eff` provides a crude estimate of the effective number of samples [92]. The final column, labelled `Rhat`, gives the ratio of the between-chain and within-chain variance. For convergence, it should be close to one [92, 174].

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 (`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 $10\,\mathrm{mg}/\sqrt{3} \approx 5.8\,\mathrm{mg}$, respectively. The same can be said about the correction due to air buoyancy (`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 (`dm_d`) the expected standard deviation is $15\,\mathrm{mg}/\sqrt{3} \approx 8.7\,\mathrm{mg}$, and the mean is zero [82]; the results obtained from the MCMC are 8.7 mg and 0 mg respectively.

The mass difference of the standard (Δm_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 `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 [82]).

The final hurdle in this example is the reproduction of the expanded uncertainty, which is stated to be 59 mg [82]. The MCMC calculation provides for all parameters the 95 % coverage
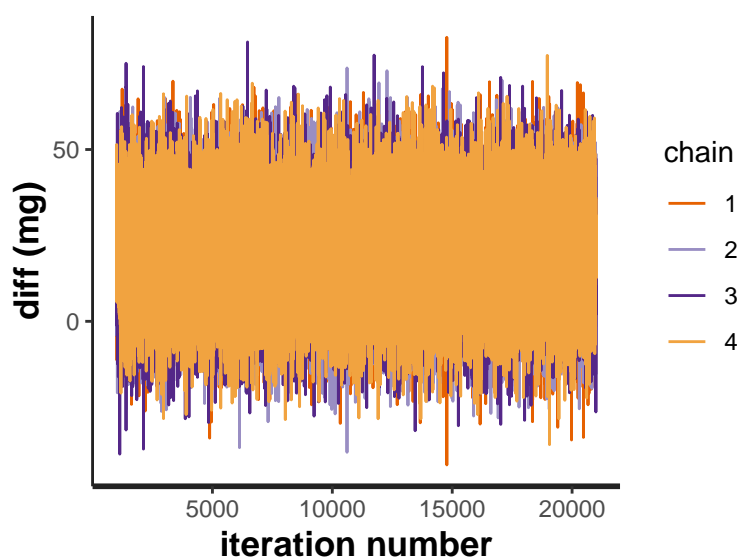
Figure 3.1: Trace plot of the model parameter `diff`

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 Δm_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 Δm_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

```
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) [82]. 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 [71, 92, 174]);
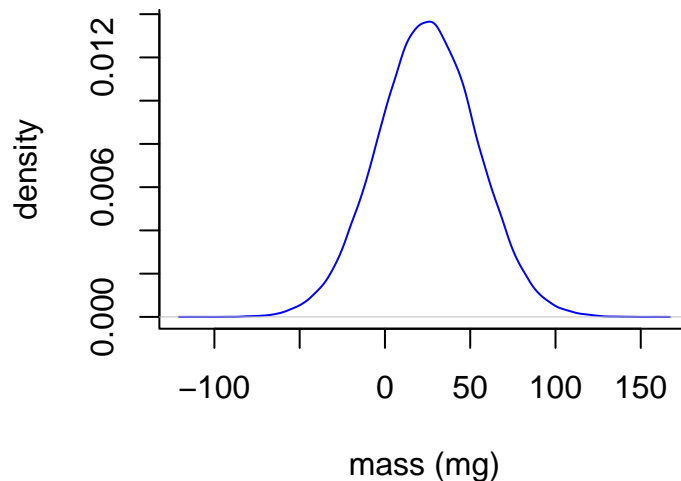
Figure 3.2: Posterior of mass difference of the weight being calibrated from its nominal mass

3. 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! [52, 92]).

A concern for those favouring classical statistical methods could be the weakly informative prior assigned to the variable `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 $(-\infty, +\infty)$ [92]. The corresponding model is obtained by removing the weakly informative prior from the `model` block and takes the form

```
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

```
## Inference for Stan model: d370744d73ed5069a780210ed9d07c6e.
## 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   5.04    0.08 22.54 -39.23 49.06   84178    1
## dm_d -0.01    0.03  8.65 -14.26 14.25   88047    1
## diff 20.09    0.05 14.39  -8.23 48.36   85716    1
## dm_c -0.03    0.02  5.78  -9.51  9.51  102528    1
## dm_B -0.01    0.02  5.75  -9.49  9.49   99109    1
## m_x  25.09    0.10 29.30 -32.43 82.70   86586    1
## lp__  2.24    0.01  1.77  -2.08  4.60   31782    1
##
## Samples were drawn using NUTS(diag_e) at Fri Mar 29 19:17:19 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).
```

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 `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 [170, 174]. 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 Δm_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 [51].

Example E1

Straight-line calibration in errors-in-variables models

S. Martens, K. Klauenberg, C. Elster

E1.1 Summary

In calibration practice, regression problems often include uncertainties in both the dependent and independent variables, which are also called errors-in-variables models. The parameters of such regression models can be estimated with the help of weighted total least squares methods. The uncertainty for these regression parameters can be determined by the GUM approaches of propagating uncertainties [51, 54] or propagating distributions [52, 54]. Alternatively Bayesian inference can be applied.

Comparing these three approaches for straight-line calibration in errors-in-variables models resulted in the examples

– "Calibration of a sonic nozzle as an example for quantifying all uncertainties involved in straight-line regression" (see E11),

– "Quantifying uncertainty when comparing measurement methods – Haemoglobin concentration as an example of correlation in straight-line regression" (see E13), and

– "Calibration of a torque measuring system – GUM uncertainty evaluation for least-squares versus Bayesian inference" (see E14).

Example E2

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. Bošnjaković, V. Karahodžić, C. Elster

E2.1 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.

E2.2 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 [52], the available information is a best estimate and its associated uncertainty. A comparison of results between law of propagation of uncertainty (LPU), Monte Carlo method (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 [77].

E2.3 Specification of the measurand

As described in JCGM 101:2008 [52], the application concerns the calibration of a weight W of mass density ρ_W against a reference weight R of mass density ρ_R having nominally the same mass m_{nom}, using a balance operating in air of mass density ρ_a. Let δm_R be the mass of a small weight of density ρ_R added to R to balance it with W.

27

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 $\rho_0 = 8 \times 10^3 \, \text{kg} \, \text{m}^{-3}$ that balances W in air at density $\rho_{a_0} = 1.2 \, \text{kg} \, \text{m}^{-3}$.

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

E2.4 Measurement model

According to JCGM 101:2008 [52], in terms of conventional masses $m_{R,c}$, $m_{R,c}$ and $\delta m_{R,c}$, an approximation adequate for most purposes is

$$m_{W,c} = (m_{R,c} + \delta m_{R,c})\left[1 + (\rho_a - \rho_{a_0})\left(\frac{1}{\rho_W} - \frac{1}{\rho_R}\right)\right]. \tag{E2.1}$$

The measurement model used in the mass calibration example of [52] is

$$\delta m = (m_{R,c} + \delta m_{R,c})\left[1 + (\rho_a - \rho_{a_0})\left(\frac{1}{\rho_W} - \frac{1}{\rho_R}\right)\right] - m_{nom}. \tag{E2.2}$$

E2.5 Input quantities of the measurement model

Table E2.1 summarizes the input quantities $m_{R,c}$, $\delta m_{R,c}$, ρ_a, ρ_W and ρ_R, and the PDFs assigned from [52]. In the table, a Gaussian distribution $\text{N}(\mu, \sigma^2)$ is described in terms of expectation μ and standard deviation σ, and a rectangular distribution $\text{R}(a, b)$ with endpoints a and b ($a < b$) in terms of expectation $(a + b)/2$ and semi-width $(b - a)/2$.

Table E2.1: The input quantities and PDFs assigned to them for the mass calibration model (E2.2), from JCGM 101:2008 [52].

Quantity	Distribution	Parameters			
		Expectation μ	Standard deviation σ	Expectation $(a+b)/2$	Semi-width $(b-a)/2$
$m_{R,c}$	$\text{N}(\mu, \sigma^2)$	$100\,000.000 \, \text{mg}$	$0.050 \, \text{mg}$		
$\delta m_{R,c}$	$\text{N}(\mu, \sigma^2)$	$1.234 \, \text{mg}$	$0.020 \, \text{mg}$		
ρ_a	$\text{R}(a,b)$			$1.20 \, \text{kg} \, \text{m}^{-3}$	$0.10 \, \text{kg} \, \text{m}^{-3}$
ρ_W	$\text{R}(a,b)$			$8 \times 10^3 \, \text{kg} \, \text{m}^{-3}$	$1 \times 10^3 \, \text{kg} \, \text{m}^{-3}$
ρ_R	$\text{R}(a,b)$			$8.00 \times 10^3 \, \text{kg} \, \text{m}^{-3}$	$0.05 \times 10^3 \, \text{kg} \, \text{m}^{-3}$

Note that the input quantity $\delta m_{R,c}$ is usually associated with fresh calibration results but that in the JCGM 101:2008 [52] treatment of mass calibration, a Type B uncertainty evaluation of $\delta m_{R,c}$ is performed resulting in a Gaussian distribution $\delta m_{R,c} \sim \text{N}(d, u^2(d))$ where d is a best estimate with associated uncertainty $u(d)$.

E2.6 Uncertainty propagation

E2.6.1 Bayesian analysis: generalities

To set up a Bayesian framework [92], [60], a statistical model is needed for which we choose to revise notation, as in [84], so that random variables are now represented by Greek letters. In this document, we consider statistical models of the form

$$D|\eta, \boldsymbol{\theta} \sim \text{N}\left([\eta - G(\boldsymbol{\theta})]/C(\boldsymbol{\theta}), u^2(d)\right) \tag{E2.3}$$

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

The statistical model (E2.3) is equivalent to the measurement model (E2.2)

$$\eta = G(\boldsymbol{\theta}) + C(\boldsymbol{\theta})\zeta \tag{E2.4}$$

with

$$\zeta = \delta m_{\mathrm{R,c}}, \tag{E2.5}$$

$$\boldsymbol{\theta} = \left(\rho_{\mathrm{a}}, \rho_{\mathrm{W}}, \rho_{\mathrm{R}}, m_{\mathrm{R,c}}\right), \tag{E2.6}$$

$$C(\boldsymbol{\theta}) = 1 + (\rho_{\mathrm{a}} - \rho_{\mathrm{a}_0})\left(\frac{1}{\rho_{\mathrm{W}}} - \frac{1}{\rho_{\mathrm{R}}}\right), \tag{E2.7}$$

$$G(\boldsymbol{\theta}) = C(\boldsymbol{\theta}) m_{\mathrm{R,c}} - m_{\mathrm{nom}}. \tag{E2.8}$$

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

$$\pi(\eta|d) = \int_{\boldsymbol{\theta}} \pi(\eta, \boldsymbol{\theta}|d)\mathrm{d}\boldsymbol{\theta}, \tag{E2.9}$$

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

In this document, point estimates are derived from equation (E2.9) for comparison with LPU and MCM. We introduce the following quantities $\widehat{\eta} = E(\eta|d) = \int \eta \pi(\eta|d)\mathrm{d}\eta$ to denote the posterior mean of the measurement result and $u^2(\widehat{\eta}) = V(\eta|d) = \int (\eta - \widehat{\eta})^2 \pi(\eta|d)\mathrm{d}\eta$ to denote the posterior variance of the measurement result. Coverage intervals are computed as shortest intervals as described in [52], similar to highest posterior density (HPD) intervals in Bayesian statistics.

E2.6.2 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 η_0 is to use a prior Gaussian distribution $\pi(\eta) \sim \mathrm{N}(\eta_0, \sigma_0^2)$ where the standard deviation σ_0 controls the degree of belief in η_0. For instance, if $|\eta_0|$ is much larger than σ_0, a small value of the relative uncertainty σ_0/η_0 gives an informative prior distribution whereas a large 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.).

E2.6.3 Posterior distributions

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

$$\pi(\eta, \boldsymbol{\theta}|d) = \frac{l(d|\eta, \boldsymbol{\theta})\pi(\eta, \boldsymbol{\theta})}{m(d)}, \tag{E2.10}$$

where $m(d) = \int l(d|\eta, \boldsymbol{\theta})\pi(\eta, \boldsymbol{\theta})\mathrm{d}\eta$ is the marginal distribution of d, $\pi(\eta, \boldsymbol{\theta}) = \pi(\eta)\pi(\boldsymbol{\theta})$ and $\pi(\boldsymbol{\theta})$ is the probability distribution of the input quantities contained in $\boldsymbol{\theta}$.

Equivalently, (E2.10) can be translated into the proportionality relation as follows

$$\pi(\eta, \boldsymbol{\theta}|d) \propto l(d|\eta, \boldsymbol{\theta})\pi(\eta, \boldsymbol{\theta}). \tag{E2.11}$$

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

$$l(d|\eta, \boldsymbol{\theta}) \propto \frac{1}{([C(\boldsymbol{\theta})]^2 s^2)^{\frac{1}{2}}} \exp\left\{-\frac{(\eta - m(\boldsymbol{\theta}))^2}{2[C(\boldsymbol{\theta})]^2 s^2}\right\}, \tag{E2.12}$$

where $m(\boldsymbol{\theta}) = C(\boldsymbol{\theta})d + G(\boldsymbol{\theta})$.

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

$$\pi(\eta, \boldsymbol{\theta}|d) \sim \mathrm{N}\left(m(\boldsymbol{\theta}), [C(\boldsymbol{\theta})]^2 s^2\right)\pi(\boldsymbol{\theta}). \tag{E2.13}$$

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

$$\pi(\eta, \boldsymbol{\theta}|d) \sim \mathrm{N}\left(m_\mathrm{p}(\boldsymbol{\theta}), \sigma_\mathrm{p}^2(\boldsymbol{\theta})\right)\pi(\boldsymbol{\theta}), \tag{E2.14}$$

where the posterior mean and variance of η are, respectively,

$$m_\mathrm{p}(\boldsymbol{\theta}) = \sigma_\mathrm{p}^2(\boldsymbol{\theta})\left\{\frac{\eta_0}{\sigma_0^2} + \frac{m(\boldsymbol{\theta})}{[C(\boldsymbol{\theta})]^2 s^2}\right\}, \qquad \sigma_\mathrm{p}^2(\boldsymbol{\theta}) = \left\{\frac{1}{\sigma_0^2} + \frac{1}{[C(\boldsymbol{\theta})]^2 s^2}\right\}^{-1}.$$

The posterior mean is a weighted mean between the prior η_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, $1/\{[C(\boldsymbol{\theta})]^2 s^2\}$.

The integration according to (E2.9) is performed with a Monte Carlo method. The total number of Monte Carlo trials is decomposed as follows: n_MC draws according to $\pi(\boldsymbol{\theta})$ and n_post draws from the Gaussian distributions (E2.13) or (E2.14) giving a total of $n_\mathrm{MC} \times n_\mathrm{post}$ simulations.

E2.7 Reporting the result

E2.7.1 Bayesian analysis of the mass calibration example in JCGM 101:2008

Results obtained with LPU, MCM and the Bayesian approach with non-informative prior (Bayes-NI) are displayed in Table E2.2 (LPU$_1$ and LPU$_2$ denote respectively the first and second order Taylor approximations) and plotted in Figure E2.1. The comparison shows a good agreement between methods[1].

Table E2.2: 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 [52].

Method	$\widehat{\delta}m$ /mg	$u(\widehat{\delta}m)$ /mg	Shortest 95 % coverage interval CI/mg
LPU$_1$	1.234 0	0.053 9	$[1.128\,5, 1.339\,5]$
LPU$_2$	1.234 0	0.075 0	$[1.087\,0, 1.381\,0]$
MCM	1.234 0	0.075 4	$[1.083\,4, 1.382\,5]$
Bayes-NI	1.234 0	0.075 5	$[1.084\,5, 1.383\,0]$

[1]For the so-called non informative prior, [84] showed that Bayesian marginal posterior uncertainty coincides with the MCM uncertainty estimate when the model is linear.
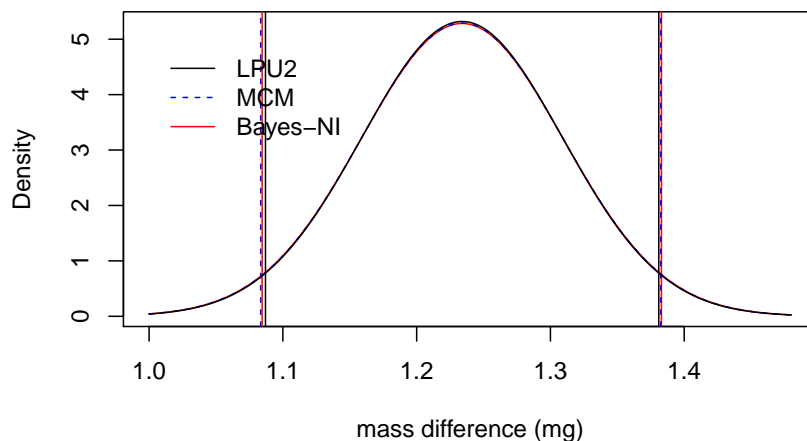
Figure E2.1: Distributions of δm obtained under Gaussian approximation with LPU2, MCM and Bayes-NI from the values in Table E2.2.

Results obtained with a Gaussian prior distribution are displayed in Table E2.3 and plotted in Figure E2.2. It can be observed that, when the prior standard deviation σ_0 increases, the weight of the prior distribution decreases and the resulting posterior distribution tends to the non informative case.

Table E2.3: Comparison of results obtained with the Bayesian analysis under Gaussian prior distributions.

η_0 /mg	σ_0 /mg	$\widehat{\delta}m$ /mg	$u(\widehat{\delta}m)$ /mg	Shortest 95 % coverage interval CI/mg
1.134	0.020	1.184 0	0.039 0	$[1.106\,9, 1.261\,3]$
1.134	0.01	1.153 9	0.017 1	$[1.127\,2, 1.152\,5]$
1.134	0.040	1.214 0	0.061 0	$[1.093\,6, 1.334\,5]$

In this section, all the results obtained with the Bayesian approach involve 2×10^7 Monte Carlo trials ($n_{\mathrm{MC}} = 20\,000$, $n_{\mathrm{post}} = 1000$).

E2.8 Conclusion

This document shows the main features of a Bayesian approach of uncertainty evaluation applied to the mass calibration example in JCGM 101:2008 [52]. 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 [52] and by the Gaussian distribution from LPU [51].

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
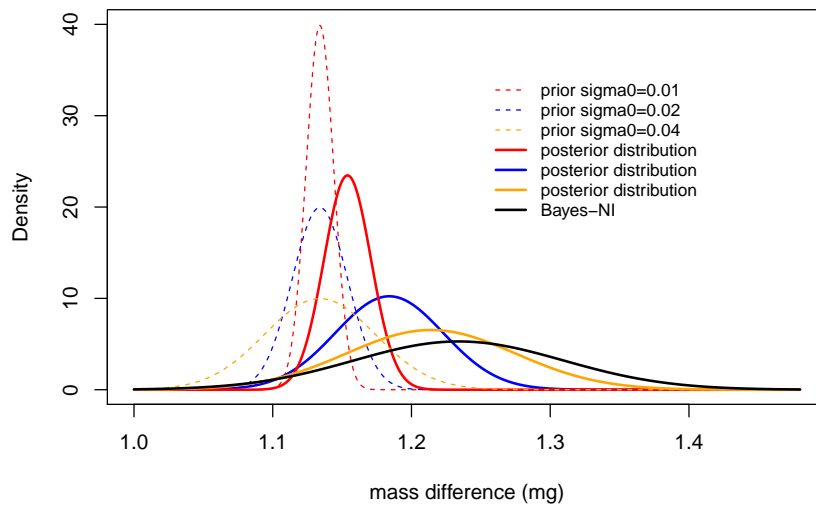
Figure E2.2: Posterior distributions of δm obtained under Gaussian prior distributions from the values in Table E2.3.

MCM determines a sample for linear measurement models, see for instance [84] and [159] 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 [66], usually used to sample from high-dimensional integrals, as in [159] and [146], which can be helpful for any practitioner already familiar with MCM willing to perform a Bayesian uncertainty analysis.

Example E3

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

L.L. Martins, A.S. Ribeiro, M.G. Cox, J.A. Sousa, D. Loureiro, M.C. Almeida, M.A. Silva, R. Brito, A.C. Soares

E3.1   Summary

This example illustrates the application of the Monte Carlo Method (MCM) in measurement uncertainty propagation related to the Single Burning Item (SBI) test, within the European normative framework of reaction to fire tests for building products, namely, the EN 13823:2010+A1 [17]. 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 Uncertainty Framework (GUF) [51] described in [26].

E3.2   Introduction of the application

The objective of the SBI standard test [17] is to measure a set of quantities which determine the evaluation and classification of a construction material (excluding floorings), aiming to characterise 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 deg 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.

E3.3   Specification of the measurand(s)

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 $m^2 s^{-1}$). Due to the applied test method [17], the definitions of these quantities
are related to different time periods of the SBI test, namely:

$$\text{HRR}(t) = \begin{cases} 0, & t \leq 300s \\ \max[0, \text{HRR}_{\text{total}}(t) - \text{HRR}_{\text{burner}}], & 300s < t \leq 312s \\ \text{HRR}_{\text{total}}(t) - \text{HRR}_{\text{burner}}, & 312s < t \end{cases} \quad \text{(E3.1)}$$

where $\text{HRR}_{\text{total}}(t)$ is the total thermal power released by the specimen and the main burner in the
time instant t, while $\text{HRR}_{\text{burner}}$ is the average thermal power released only by the main burner;
and

$$\text{SPR}(t) = \begin{cases} 0, & t \leq 300s \\ \max[0, \text{SPR}_{\text{total}}(t) - \text{SPR}_{\text{burner}}], & 300s < t \leq 312s \\ \text{SPR}_{\text{total}}(t) - \text{SPR}_{\text{burner}}, & 312s < t \end{cases} \quad \text{(E3.2)}$$

where $\text{SPR}_{\text{total}}(t)$ is the total smoke production rate of the specimen and the main burner in the
time instant t, while $\text{SPR}_{\text{burner}}$ 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 $\text{HRR}_{\text{burner}}$ and $\text{SPR}_{\text{burner}}$, 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 speci-
men in a certain time exposure to the main burner flames (namely, in the first 600 s), and FIGRA,
the fire growth rate (expressed in $W s^{-1}$), 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
$m^2 s^{-2}$), 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 characterised 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.

E3.4 Measurement model

The heat release rate measurement model is derived from the studies performed by [137] 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 $MJ m^{-3}$) is considered constant regardless of the combustion material, which can be expressed

by

$$\text{HRR} = E' \cdot x_{O_2}^0 \cdot \varphi \cdot q_{V_{s'}} \tag{E3.3}$$

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

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

$$E' = \rho_{O_2} \cdot E \tag{E3.4}$$

The amount fraction oxygen in the ambient is given by

$$x_{O_2}^0 = x_{O_2}^{\text{initial}} \cdot \left(1 - x_{H_2O}^{\text{o}}\right) \tag{E3.5}$$

$x_{O_2}^{\text{initial}}$ being the amount fraction oxygen measured in the initial stage of the SBI test (in the time period between 30 s and 90 s), with a gas analyser[1], and $x_{H_2O}^0$, the amount fraction water 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 vaporisation of water)

$$x_{H_2O}^0 = \frac{\text{rh}}{100 \cdot p_{\text{atm}}} \exp\left[23.2 - \frac{3816}{T_{\text{initial}} - 46}\right] \tag{E3.6}$$

where rh is the relative humidity in moist air (as a percentage), T_{initial} is the initial air temperature inside the exhaust duct (in K) and p_{atm} is the atmospheric pressure (in Pa). The oxygen depletion factor φ is calculated by

$$\varphi = \frac{x_{O_2}^{\text{initial}}(1 - x_{CO_2}) - x_{O_2}(1 - x_{CO_2}^{\text{initial}})}{x_{O_2}^{\text{initial}}(1 - x_{O_2} - x_{CO_2})} \tag{E3.7}$$

where $x_{O_2}^{\text{initial}}$ and $x_{CO_2}^{\text{initial}}$ are, respectively, the amount fractions of oxygen and carbon dioxide measured in the initial stage of the SBI test with the gas analyser, while x_{02} 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 rate of air in the ambient is indirectly measured based on the expression

$$q_{V_S} = \frac{q_{V_c}}{1 + (\alpha - 1) \cdot \varphi} \tag{E3.8}$$

in which q_{V_c} is the volumetric flow rate of the gases in the exhaust duct (in $\text{m}^3\,\text{s}^{-1}$) and α 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.

$$\alpha = 1 + x_{O_2}^0(\beta - 1) \tag{E3.9}$$

---

[1]This equipment receives a gas sample from a normalised exhaust duct in which all water vapour and water soluble gases are eliminated before measurement.

---

where $x_{O_2}^0$ is obtained from expression (E3.5) and β is the ratio between the amount of substance of combustion products and of consumed oxygen. A combustion reaction involving reactants such as hydrocarbons ($C_a H_b O_c N_d X_e$) 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

$$C_a H_b O_c N_d X_e + \left(a - \frac{g}{2} + \frac{b-e}{4} - \frac{c}{2} \right) O_2 \rightarrow (a-c)CO_2 + \frac{b-e}{2} H_2O + e HX + g CO + \frac{d}{2} N_2 \quad \text{(E3.10)}$$
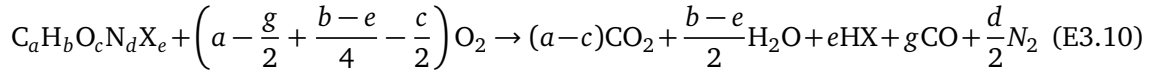
where particular constants a to g apply in any specific instance. Therefore, based on expression (E3.10) and by definition, the β ratio is given by

$$\beta = \frac{4a + 2b + 2e + 2d}{4a + b - e - 2c - 2g} \quad \text{(E3.11)}$$

Depending on the type of hydrocarbon subjected to combustion, several estimates are known for the β ratio usually values between one and two[2]. The volumetric flow rate of the gases in the exhaust duct is obtained by the expression

$$q_{V_c} = \frac{k_t}{k_p} \cdot \sqrt{\frac{2\Delta p}{\rho_{T_0}} \cdot \frac{T_0}{T}} \cdot A \quad \text{(E3.12)}$$

where Δp is the differential pressure measured in a bidirectional pressure sensor located inside the exhaust duct (in Pa); ρ_{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

$$A = \frac{\pi}{4} \cdot d^2 \quad \text{(E3.13)}$$

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 [104] is used:

$$\rho_{T_0} = \frac{0.34848 \cdot p_{Atm} - 0.009024 \text{rh} \exp\left[0.0612 \cdot (T_0 - 273.15) \right]}{T_n} \quad \text{(E3.14)}$$

The differential pressure correction factor is considered in expression (E3.12) due to the use of a bidirectional sensor [124] instead of a conventional Pitot tube (vulnerable to solid particles in the flow). This quantity is defined by

$$k_p = \frac{\sqrt{\frac{\Delta p}{\rho_{T_{amb}}}}}{v_c} \quad \text{(E3.15)}$$

where v_c is the linear flow velocity in the centre of the exhaust duct cross-section (in $m\,s^{-1}$) and $\rho_{T_{amb}}$ is the moist air density [104] for ambient temperature, T_{amb} (in K) given by

$$\rho_{T_{amb}} = \frac{0.34848 p_{atm} - 0.009024 \text{rh} \exp\left[0.0612 (T_{amb} - 273.15) \right]}{T_{amb}} \quad \text{(E3.16)}$$

---

[2]Examples of β ratio estimates for the combustion of: carbon (C, $\beta = 1$); ethylene (C_2H_4, $\beta = 1.3$); propene (C_3H_6, $\beta = 1.3$); butane (C_4H_{10}, $\beta = 1.4$); heptane (C_7H_{16}, $\beta = 1.4$); propane (C_3H_8, $\beta = 1.4$); ethane (C_2H_6, $\beta = 1.4$); methane (CH_4, $\beta = 1.5$); 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 $kg\,m^{-3}$).

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

$$k_{t,v} = \frac{\sum_{i=1}^{5} v_i}{5 \cdot v_c} \tag{E3.17}$$

considering the average[4] 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 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.

$$k_{t,\text{propane}} = k_t' \cdot \frac{\sum_i \text{HRR}_i^{\text{theoretical}}}{\sum_i \text{HRR}_i^{\text{experimental}}} \tag{E3.18}$$

considering the several testing steps indexed by i of this normalised test [17], 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 the step is given by

$$\text{HRR}_i^{\text{theoretical}} = q_{m_i} |\Delta h_c^l|_{\text{propane}'} \tag{E3.19}$$

where q_{m_i} is the propane mass flow in the i th testing step (expressed in kg s^{-1}), and $\left|\Delta h_c^l\right|_{\text{propane}}$ is the low enthalpy of propane combustion per unit of mass (in kJ kg^{-1}). It should be noted that, in the calculation of $\text{HRR}_i^{\text{experimental}}$ by expression (E3.3), 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 [26] and used for construction materials in the SBI test. Regarding the k_{theptane} correction, this quantity is obtained from the expression

$$k_{t\ \text{heptane}} = k_t' \cdot \frac{|\Delta h_c^l|_{\text{heptane}} \cdot m_{\text{heptane}}}{\text{THR}} \tag{E3.20}$$

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, $\left|\Delta h_c^l\right|_{\text{heptane}}$ is the low enthalpy of heptane combustion per unit of mass (in kJ kg^{-1}) and m_{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 (E3.3).

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

$$\text{SPR} = \frac{q_{V_c}}{L} \cdot \frac{T}{T_0} \cdot \ln\left(\frac{1}{\tau}\right) \tag{E3.21}$$

[4]The measurement sample is composed of four velocity measurements in each of five normalised 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 (E3.18).

where q_{V_c} is the volumetric flow rate of gases in the exhaust duct (in $m^3\,s^{-1}$), obtained from expression (E3.12); 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 τ 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 photo detector 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 E3.1 shows a schematic representation of the heat release rate calculation process, while figure E3.2 refers to the smoke production rate. Particular attention is given to the global correction factor and to its calculation process, schematically represented in figure E3.3.

E3.5   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 characterised, 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 [52, 54], 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 [123] 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 E3.1 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.
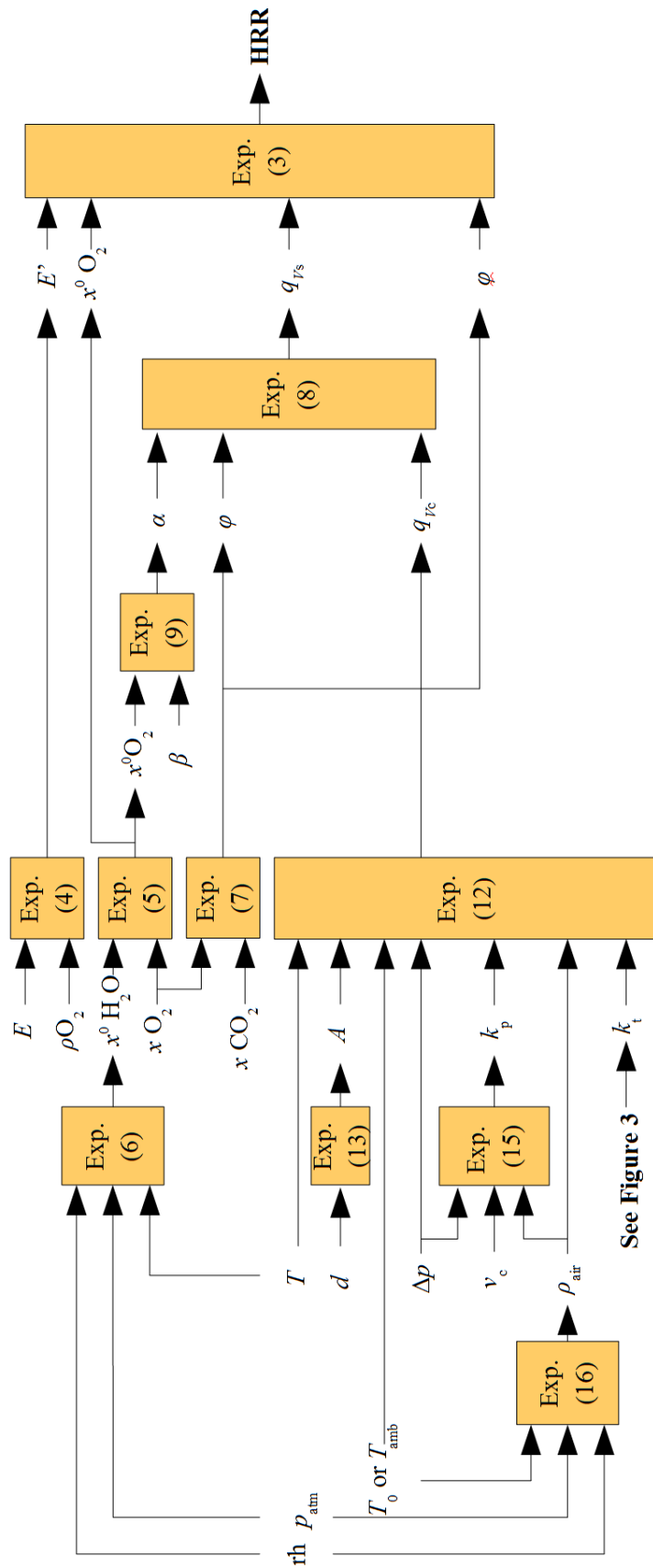
Figure E3.1: Functional diagram of the heat release rate quantity
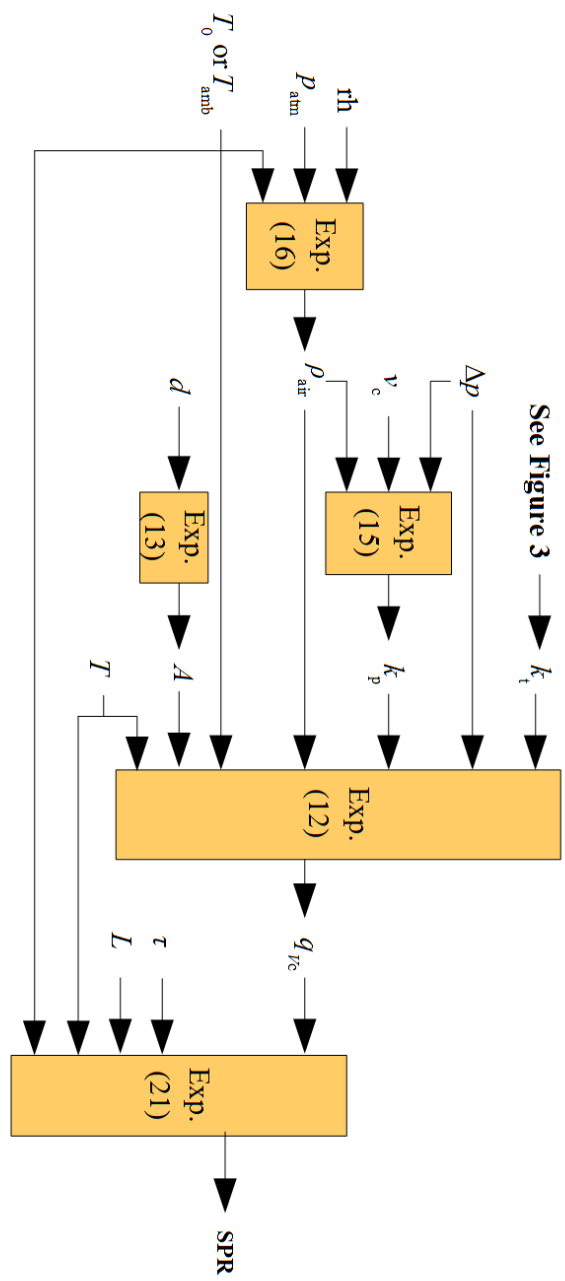
Figure E3.2: Functional diagram of the smoke production rate
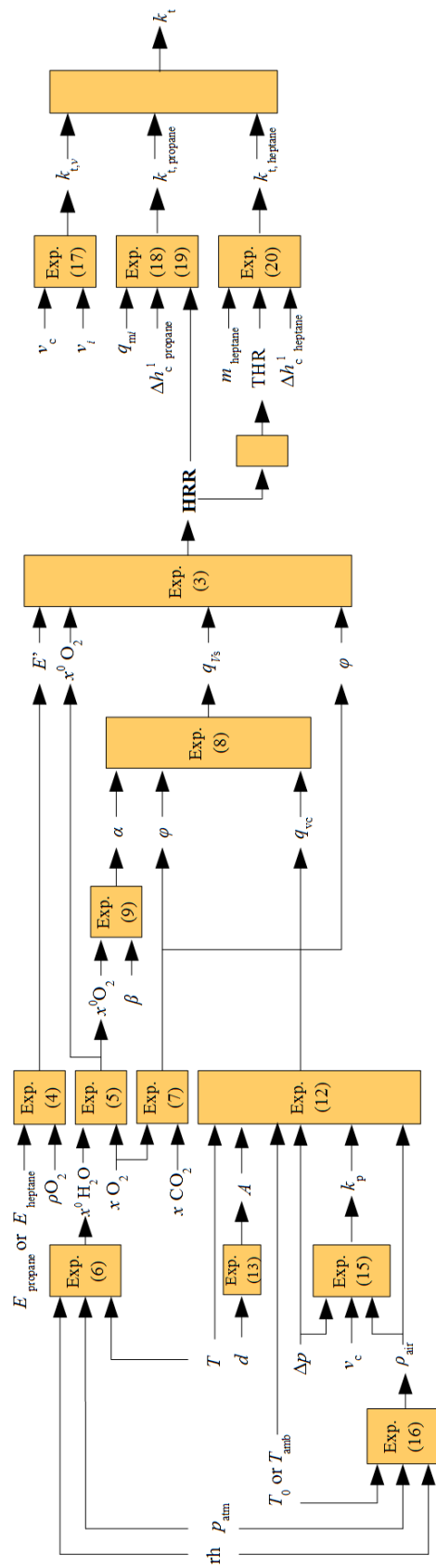
Figure E3.3: Functional diagram of the global correction factor quantity rate quantity

Table E3.1: Probabilistic formulation of the input quantities related to the heat release rate measurement

Quantity	Symbol	PDF	Estimate	Standard uncertainty
Relative humidity	rh	Gaussian	60.1 %	1.1 %
Atmospheric pressure	p_{atm}	Gaussian	101.4 kPa	0.2 kPa
Initial air temperature inside the exhaust duct	$T_{initial}$	Gaussian	288.3 K	0.1 K
Oxygen density	ρ_{O_2}	Gaussian	1.308 kg m^{-3}	0.003 kg m^{-3}
Heat released per unit of consumed oxygen mass	E	Gaussian	13.1 MJ kg^{-1}	0.3 MJ kg^{-1}
Initial amount fraction of oxygen	$x_{O_2}^{initial}$	Gaussian	0.2095	0.000 04
Amount fraction of oxygen	x_{O_2}	Gaussian	0.206 7	0.000 2
Initial amount fraction of carbon dioxide	$x_{CO_2}^{initial}$	Gaussian	0.000 3	0.000 005
Amount fraction of carbon dioxide	x_{CO_2}	Gaussian	0.001 8	0.000 02
Exhaust duct diameter	d	Gaussian	0.315 m	0.001 m
Exhaust gas temperature	T	Gaussian	313.8 K	0.4 K
Ambient temperature	T_{amb}	Gaussian	288.6 K	0.7 K
Differential pressure	Δp	Gaussian	68.6 Pa	2.1 Pa
Linear flow velocity in the centre of the exhaust duct cross-section	v_c	Gaussian	9.6 m s^{-1}	0.7 m s^{-1}
Ratio between the amount of substance of combustion products and of consumed oxygen	β	Uniform	1.5	0.3
Global correction factor	k_t	Gaussian	0.77	0.02

Regarding the smoke production quantity, table E3.2 presents the adopted probabilistic formulation of the input quantities which supported the MCM simulations.

Table E3.2: Probabilistic formulation of the input quantities of the smoke production rate

Quantity	Symbol	PDF	Estimate	Standard uncertainty
Relative humidity	rh	Gaussian	60.1 %	1.1 %
Atmospheric pressure	p_{atm}	Gaussian	101.4 kPa	0.2 kPa
Ambient temperature	T_{amb}	Gaussian	288.6 K	0.9 K
Exhaust duct diameter	d	Gaussian	0.315 m	0.001 m
Optical path length	L	Gaussian	0.315 m	0.001 m
Transmittance	τ	Gaussian	0.974	0.005
Exhaust gas temperature	T	Gaussian	313.8 K	0.4 K
Differential pressure	Δp	Gaussian	68.6 Pa	2.1 Pa
Linear flow velocity in the centre of the exhaust duct cross-section	v_c	Gaussian	9.6 m s^{-1}	0.7 m s^{-1}
Global correction factor	k_t	Gaussian	0.77	0.02

Table E3.3: MCM simulation results for intermediate quantities in the calculation of the heat release rate

Quantity	Symbol	PDF	Estimate	Standard uncertainty
Water vapour amount fraction	$x^0_{H_2O}$	Gaussian	0.0146	0.0004
Moist air density for ambient temperature	$\rho_{T_{amb}}$	Gaussian	$1.180\,\mathrm{kg\,m^{-3}}$	$0.005\,\mathrm{kg\,m^{-3}}$
Differential pressure correction factor	k_p	Gaussian	1.15	0.09
Expansion factor	α	Gaussian	1.1	0.05
Volumetric flow rate of gases in the exhaust duct	q_{V_c}	Gaussian	$0.55\,\mathrm{m^3\,s^{-1}}$	$0.04\,\mathrm{m^3\,s^{-1}}$
Heat released per unit of consumed oxygen volume	E'	Gaussian	$17.1\,\mathrm{MJ\,m^{-3}}$	$0.4\,\mathrm{MJ\,m^{-3}}$
Ambient oxygen molar fraction	$x^0_{O_2}$	Gaussian	0.2074	0.00005
Oxygen depletion factor	φ	Gaussian	0.015	0.001
Ambient volumetric flow rate	q_{V_s}	Gaussian	$0.55\,\mathrm{m^3\,s^{-1}}$	$0.04\,\mathrm{m^3\,s^{-1}}$
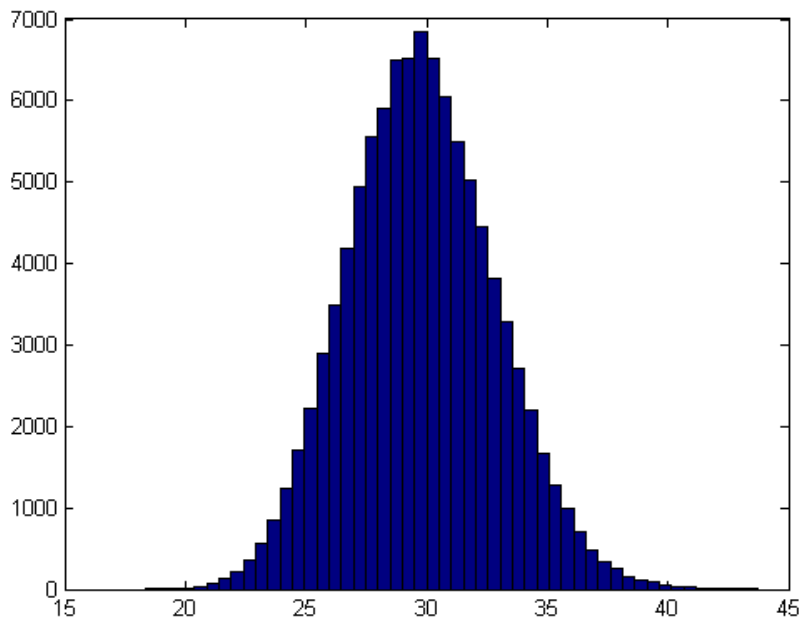
Figure E3.4: Output PDF of the heat release rate quantity
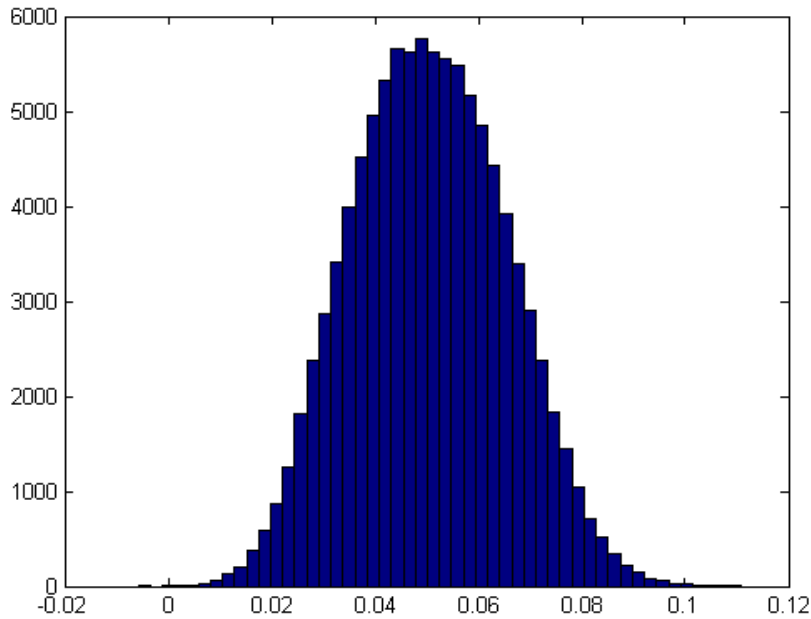
Figure E3.5: Output PDF of the smoke production rate quantity

Table E3.4: MCM simulation results for the heat release rate

Quantity	Symbol	PDF	Estimate	Standard uncertainty	Required accuracy	Simulation accuracy
Heat release rate	HRR	Gaussian	30 kW	3 kW	0.5 kW	0.1 kW

Table E3.5: MCM simulation results for intermediate quantities in the calculation of the smoke release rate

Quantity	Symbol	PDF	Estimate	Standard uncertainty
Moist air density for ambient temperature	$\rho_{T_{\text{amb}}}$	Gaussian	$1.220 \, \text{kg m}^{-3}$	$0.005 \, \text{kg m}^{-3}$
Differential pressure correction factor	k_{p}	Gaussian	1.11	0.09
Volumetric flow rate of gases in the exhaust duct	$q_{V_{\text{c}}}$	Gaussian	$0.55 \, \text{m}^3 \, \text{s}^{-1}$	$0.04 \, \text{m}^3 \, \text{s}^{-1}$

Table E3.6: MCM simulation results for the smoke production rate ($0.05 \, \text{m}^2 \, \text{s}^{-1}$ level)

Quantity	Symbol	PDF	Estimate	Standard uncertainty	Required accuracy	Simulation accuracy
Smoke production rate	SPR	Gaussian	$0.05 \, \text{m}^2 \, \text{s}^{-1}$	$0.02 \, \text{m}^2 \, \text{s}^{-1}$	$0.005 \, \text{m}^2 \, \text{s}^{-1}$	$0.0005 \, \text{m}^2 \, \text{s}^{-1}$

E3.6 Reporting the result

The measurement uncertainties of the input quantities shown in table E3.1 were propagated by the MCM to the intermediate quantities (results shown in table E3.3) and, posteriorly, to the total heat release rate quantity (see table E3.4 for an example of a 30 kW thermal power level). Figure E3.4 shows the output PDF obtained for the heat release rate quantity.

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 E3.5 (intermediate quantities) and table E3.6 (output quantity). Figure E3.5 shows the PDF obtained by the MCM for the smoke production rate.

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

E3.7 Interpretation of results

As seen in figures E3.4 and E3.5, 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 E3.1) were taken as Gaussian. In terms of validation of results, tables E3.4 and E3.6 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 E4

Statistical reassessment of calibration and measurement capabilities based on key comparison results

K. Shirono, M.G. Cox

E4.1   Summary

This example illustrates the minimal adjustment of CMC (calibration and measurement capability) uncertainty claims so they are supported by the results of a key comparison (KC). According to the CIPM Mutual Recognition Arrangement (MRA) [68, clause T.7], CMC uncertainties are normally expressed at a 95 % level of confidence. CMC uncertainties are the expanded measurement uncertainties available to customers under normal conditions of measurement. When laboratories' CMC claims are unsupported by the relevant KC, modified values must be assigned to their declared CMC uncertainties.

In the vast majority of cases when CMCs apply to a continuous interval of values such as mass fraction or wavelength, KCs are carried out for selected discrete values of the quantity concerned. Since the comparison at each discrete value strictly only supports the CMC uncertainty at that value, it is not immediately apparent how to modify the CMC uncertainties. Under realistic assumptions, we apply a method that is applicable in such an instance and for which the reported CMC uncertainties are amplified so that they are underpinned by the results of the KC. The amplification factors depend on the laboratories' degrees of equivalence (DoEs) for these discrete values, adjusted to achieve consistency with the key comparison reference values (KCRVs).

The method is based on the patterns in the individual behaviour of the DoEs of the participating laboratories for the discrete values, implying the presence of correlation associated with the DoE values. It applies when the weighted mean of some or all of the measured values reported by the participating laboratories in the KC is used to obtain the KCRV.

Full details of the example are provided in [158].

E4.2   Introduction of the application

CMCs must be consistent with results derived from KCs [68, clause T.7], a requirement interpreted in the sense that a CMC uncertainty claimed by a participating laboratory must be no smaller than the expanded uncertainty associated with the corresponding laboratory's reported value in the KC. The extent of agreement of that reported value to the reference value in the KC is assessed

by a DoE calculated in accordance with the MRA [68, clause T.2]. Such an interpretation is straightforward when there exists a 'one-to-one' relationship between the KC and the CMC claim, that is, when the KC and the CMC relate to the same measurand [73]. In such a case, it is straightforward to obtain an appropriate uncertainty that should be reported in the KC for the performance evaluated by the DoE to be satisfactory [73].

This example relates to the commonest class of CMC claims in which laboratories provide uncertainty for a measurand that depends continuously on a quantity (parameter) having an interval of values, termed here the 'measurement interval'. This parameter could, for example, be frequency, wavelength or mass concentration.

The corresponding KC provides DoEs for each participating laboratory for each of a discrete set of values of the parameter within the measurement interval. An analysis based on the one-to-one relationship could be applied separately for each of these discrete parameter values. A consequence of doing so is that any structure present in the data across these parameter values is not taken into consideration: the analysis of these discrete cases are independent exercises. Generally there would be a different expansion factor for the CMC uncertainties corresponding to each discrete value, particularly in cases when the same measuring system is used for each such case, perhaps due only to random variation. The provision of a single expansion factor for each NMI based on the completed and published KC results would be helpful for the reassessment of the CMC uncertainties.

Importantly, KC results for these discrete values almost invariably display some degree of correlation that cannot be taken into consideration by an analysis for the one-to-one relationship. The existence of correlation is often evidenced by patterns in the individual behaviour of the DoEs for each participating laboratory across the discrete values of the parameter within the measurement interval. An instance is given in [158] relating to a KC of free-field hydrophone calibrations in the frequency interval 1 kHz to 500 kHz [153]. Such correlations relate to the biases often associated with individual participating laboratories' measured values.

An approach for CMC uncertainty reassessment involving the estimation of correlations is exemplified by providing a single multiplicative expansion factor for the CMC uncertainties for each laboratory. The method described applies Bayesian principles under the assumption that the observed pattern in each laboratory's DoE value components can largely be explained by a single correlation coefficient, specific to that laboratory. Since no specific physical adjustments are assumed, the approach is generally applicable to a wide range of practices in metrology. As part of the approach, for each laboratory a single common expansion factor for CMC uncertainties is estimated that applies across the measurement interval. Since some estimated expansion factors may prove to be unity, the corresponding laboratories can be regarded as already having CMC uncertainties that are consistent with the relevant KC. Thus, the approach is discriminatory: only some laboratories are required to adjust their CMC uncertainties depending on their DoEs.

E4.3   Specification of the measurand(s)

Suppose there are N laboratories participating in the KC, each providing a measured value at p stipulated values of a parameter in the measurement interval. The measurand is a vector measurand consisting of adjusted CMC uncertainties $U_i^{\mathrm{CMC,adj}}(j)$, $i = 1, \ldots, N$, $j = 1, \ldots, p$.

There are intermediate measurands, especially the CMC uncertainty expansion factors, in an according multi-stage measurement model as described in section E4.4.

E4.4   Measurement model

The measurement model is multi-staged[1] [56] comprising various steps in the analysis of existing KC data and corresponding CMC data. The measurement model uses the following data.

Each laboratory participating in the KC reports a measured value and an associated standard uncertainty for each prescribed parameter value within the measurement interval. Specifically, for each laboratory i, $i = 1, \ldots, N$, the value $x_i(j)$, $j = 1, \ldots, p$, and the associated standard uncertainty $u(x_i(j))$ are provided. The corresponding (unadjusted) CMC uncertainties are also provided. It is assumed that for each j the KCRV $x_{\mathrm{ref}}(j)$ relating to the jth measurand is given as the weighted mean[2] (WM) of all or some of the reported values $x_i(j)$, $j = 1, \ldots, p$ [72]. Thus, the according DoEs $(d_i(j), U(d_i(j)))$, $i = 1, \ldots, N$, $j = 1, \ldots, p$, defined as follows are available.

E4.4.1   Degrees of equivalence

The DoE value component for laboratory i and parameter j is

$$d_i(j) = x_i(j) - x_{\mathrm{ref}}(j), \qquad i = 1, \ldots, N, \tag{E4.1}$$

and the corresponding uncertainty component is

$$U(d_i(j)) = k_i u(d_i(j)) = \begin{cases} k_i[u^2(x_i(j)) - u^2(x_{\mathrm{ref}}(j))]^{1/2}, & i \in I_{\mathrm{ref}}, \\ k_i[u^2(x_i(j)) + u^2(x_{\mathrm{ref}}(j))]^{1/2}, & \text{otherwise}, \end{cases} \tag{E4.2}$$

where I_{ref} denotes the set of values of i for which $x_i(j)$ and $u(x_i(j))$ are used in the computation of $x_{\mathrm{ref}}(j)$ and $u^2(x_{\mathrm{ref}}(j))$ is the variance (squared standard uncertainty) associated with $x_{\mathrm{ref}}(j)$ [72]. Under the assumption of normality, the coverage factors for the DoE uncertainties are taken as $k_i = 1.96$.

If the DoE for any participating laboratory has an E_n score[3] that is in magnitude greater than unity, that laboratory's performance is unsatisfactory and the according CMC uncertainty may have to be re-assessed.

Exclusive statistics $d_{i,\mathrm{ex}}(j)$ rather than $d_i(j)$ [130, 163], where $d_{i,\mathrm{ex}}(j) = x_i - x_{\mathrm{ex}}(j)$ and $x_{\mathrm{ex}}(j)$ is the exclusive weighted mean, as given in [158], are used to describe the DoEs because of algebraic advantages over conventional statistics. The associated standard uncertainty $u(x_{\mathrm{ex}}(j))$ is provided.

E4.4.2   Assumptions

The following assumptions are made:

1. An individual common expansion factor specific to each laboratory applies for the uncertainty over its measurement interval. The expansion factors for the uncertainties reported in a KC and the CMC uncertainties are considered identical.

---

[1]In many stepwise processes in metrology, quantities from intermediate measurements are naturally used in a subsequent measurement. Each stage in the process can be described by a measurement model with input quantities and output quantities. This set of measurement models constitutes a multi-stage measurement model and can be used as such.

[2]It is assumed that no information on correlation was employed in the computation of the WMs. Otherwise, the treatment here would require modification.

[3]As a measure of the performance of laboratory i, a normalized error ratio or 'E_n score'

$$E_n^{(i)} = \frac{d_i}{U(d_i)} = \frac{d_i}{k_i u(d_i)} \tag{E4.3}$$

is used: If $|E_n^{(i)}| \leq 1$, laboratory i's performance is regarded as 'satisfactory'; otherwise it is 'unsatisfactory' [15].

2. The measurement errors in the reported values $x_i(1), \ldots, x_i(p)$ from laboratory i can be regarded as being drawn from a multivariate normal distribution whose covariance matrix depends on a single correlation coefficient associated with those values.

E4.4.3   Steps in the multi-stage model

1. Establish a statistical model for the DoEs for the participating laboratories. The statistical model contains the following parameters to be estimated from the KC data for $i = 1, \ldots, N$:

   - Expansion factor L_i for laboratory i.
   - Correlation coefficient ρ_i for laboratory i: see section E4.4.4.
   - Technical parameter λ_i related to the standard deviations S_i in section E4.4.4.

   Although the ρ_i are not primary measurands, they are of interest in understanding the extent of the correlations involved for the individual laboratories.

2. Solve the statistical model for expansion factors L_i for the participating laboratories.

3. Apply the expansion factors to the existing CMC uncertainties to provide adjusted CMC uncertainties that are supported by the KC.

E4.4.4   Statistical model

Let the vector $\boldsymbol{d}_{i,\mathrm{ex}} = [d_{i,\mathrm{ex}}(1), \ldots, d_{i,\mathrm{ex}}(p)]^{\top}$ denote the value components of the exclusive DoEs for laboratory i for the p discrete values of the parameter. The probability distribution used to describe the vector quantity for which $\boldsymbol{d}_{i,\mathrm{ex}}$ is a realization is assumed to be multivariate normal:

$$\boldsymbol{d}_{i,\mathrm{ex}} \sim \mathrm{N}(\boldsymbol{0}, \mathring{}_i), \tag{E4.4}$$

where $\boldsymbol{0}$ is the column vector having p zero elements and $\mathring{}_i$ is a covariance matrix of dimension $p \times p$. Consider the decomposition

$$\mathring{}_i = \boldsymbol{S}_i \boldsymbol{P}_i \boldsymbol{S}_i,$$

where $\boldsymbol{S}_i$ is the diagonal matrix whose jth diagonal element is a standard deviation $\sigma_i(j)$ equal to the square root of the jth diagonal element of $\mathring{}_i$ and $\boldsymbol{P}_i$ is a correlation matrix [54]. Neither $\boldsymbol{S}_i$ nor $\boldsymbol{P}_i$ is typically available from the KC and must be estimated from reported results.
   Define

$$r_i(j) = u^2(x_{\mathrm{ex}}(j))/u^2(x_i(j)), \qquad r_{i,\mathrm{min}} = \min_j r_i(j).$$

Then, for the matrix $\boldsymbol{S}_i$, $\sigma_i^2(j)$ is approximately given by

$$\sigma_i^2(j) = \lambda_i[u^2(x_i(j)) + u^2(x_{\mathrm{ex}}(j))] \approx (1 + r_{i,\mathrm{min}})\lambda_i u^2(x_i(j)). \tag{E4.5}$$

The expression in the right side of (E4.5) is obtained through applying the relationship

$$\tau_i + [r_i(j) - r_{i,\mathrm{min}}] \approx \tau_i, \qquad \tau_i = (1 + r_{i,\mathrm{min}})\lambda_i. \tag{E4.6}$$

Although we cannot say that expression (E4.6) is always a reasonable approximation, we can confirm the extent of its validity after estimating the parameters specified in section E4.4.3. More

details are given in [158], where τ_i is employed as a parameter to be estimated rather than λ_i. No essential change happens because of the transformation from τ_i to λ_i.

The correlation matrix $\boldsymbol{P}_i$ used in [158] has the form

$$\boldsymbol{P}_i = \boldsymbol{P}_i(\rho_i) = (1 - \rho_i)\boldsymbol{I} + \rho_i \boldsymbol{1}\boldsymbol{1}^\top,$$

where $\boldsymbol{I}$ is the identity matrix of dimension $p \times p$ and $\boldsymbol{1}$ is the column vector containing p ones. $\boldsymbol{P}_i$ is thus a matrix with ones on the main diagonal and ρ_i elsewhere. The parameters ρ_i and λ_i are obtained using Bayesian estimation. A uniform distribution over $[0,1]$ is used as the prior for ρ_i since the correlation between DoEs is expected to be non-negative, and a Jeffreys' prior is used for λ_i:

$$p(\rho_i) \propto \begin{cases} 1, & 0 \le \rho_i \le 1, \\ 0, & \text{otherwise} \end{cases} , \qquad p(\lambda_i) \propto \begin{cases} \lambda_i^{-1}, & \lambda_i \ge 1, \\ 0, & \text{otherwise}. \end{cases} \tag{E4.7}$$

E4.4.5 Data

The data used in this example is for KC CCL.K-2 [114] relating to gauge block measurements. Four gauge blocks with nominal lengths 175 mm, 500 mm, 500 mm and 900 mm were circulated to 12 participating laboratories. Because the data from a particular laboratory were "…known to contain errors and is not representative of their standard measurement technique, its data was withdrawn from the comparison" [114]. The reported deviations of the remaining 11 laboratory values from the nominal lengths of the gauge blocks and their associated standard uncertainties are summarized in table E4.1 and figure E4.1. The reference values and their associated standard uncertainties are also given in table E4.1.

Table E4.1: Reported values $x_i(j)$ and associated standard uncertainties $u_i(j)$ with reference values $x_{\text{ref}}(j)$ and associated standard uncertainties $u(x_{\text{ref}}(j))$. Numbers in square brackets were not used in the determination of reference values in accordance with reference [114]

| | Nominal length | | | | | | | |
| | 175 mm | | 500 mm | | 500 mm | | 900 mm | |
i	$x_i(1)$ /nm	$u_i(1)$ /nm	$x_i(2)$ /nm	$u_i(2)$ /nm	$x_i(3)$ /nm	$u_i(3)$ /nm	$x_i(4)$ /nm	$u_i(4)$ /nm
1	140	28	916	33	814	33	2033	42
2	122	13	915	16	807	15	1983	21
3	161	30	962	38	861	38	2057	52
4	142	16	908	23	781	23	2075	60
5	150	20	930	20	830	20	2020	35
6	125	27	881	67	786	66	2004	118
7	148	19	938	39	858	39	2070	68
8	194	19	1007	60	912	60	2160	136
9	154	23	885	50	818	50	1982	87
10	180	110	980	150	870	150	2010	250
11	[312]	[21]	952	56	868	56	2165	100

$x_{\text{ref}}(1)$ /nm	$u(x_{\text{ref}}(1))$ /nm	$x_{\text{ref}}(2)$ /nm	$u(x_{\text{ref}}(2))$ /nm	$x_{\text{ref}}(3)$ /nm	$u(x_{\text{ref}}(3))$ /nm	$x_{\text{ref}}(4)$ /nm	$u(x_{\text{ref}}(4))$ /nm
145	7	923	9	818	9	2016	14
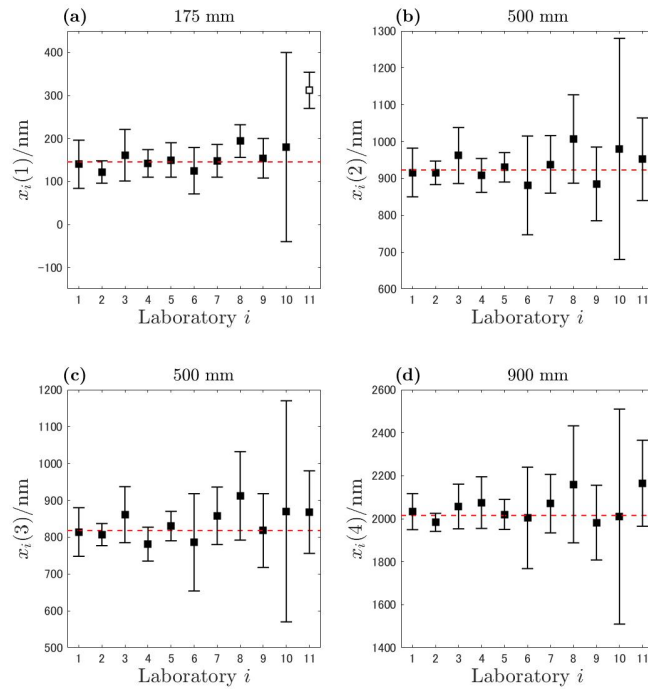
Figure E4.1: CCL.K-2 gauge block data for four nominal lengths and weighted means as KCRVs (broken horizontal lines). Vertical bars depict coverage intervals with coverage factor $k_i = 1.96$

E4.5   Uncertainty analysis

Bayesian estimation with modestly informative priors for the quantities to be estimated was used to obtain a single factor for each laboratory to expand (only when necessary) its CMC uncertainties. Bayesian modelling allows unknown correlations between reported values to be taken into consideration by estimating them and to include constraints by using priors. The maximum a posteriori (MAP) estimator was used because of several advantages [158]:

(a) An expansion factor given by MAP estimation is close to that obtained by the conventional method when $p = 1$ (only one stipulated value in the measurement interval),

(b) MAP estimation suggests that no expansion of the CMC uncertainty is required for a laboratory whose performance is satisfactory in the KC, and

(c) The MAP estimator has an analytic solution (given in [158]).

E4.6   Reporting the result

Defining L_i as the MAP estimator of λ_i, the variance of $d_{\mathrm{ex}}(j)$ is estimated as $[1 + r_i(j)]L_i u^2(x_i)$. Since the variance is supposed to be $[K_i + r_i(j)]u^2(x_i)$ using the expansion factor $K_i^{1/2}$ for the standard uncertainty $u(x_i(j))$, the following relation holds between L_i and K_i:

$$(K_i + r_{i,\min})u^2(x_i) = (1 + r_{i,\min})L_i u^2(x_i)$$

when relation (E4.6) holds. Thus,

$$K_i^{1/2} = [L_i(1 + r_{i,\min}) - r_{i,\min}]^{1/2}$$

is the expansion factor for the CMC standard uncertainty in this study.

Table E4.2 shows the values of the expansion factors $K_i^{1/2}$ and the MAP estimates ρ_i^{MAP} of ρ_i. The symbol "–" indicates that the computed values are not recommended to be used in the reassessment because relationship (E4.6) does not hold in these results. In [158], we gave the criterion $K_i/r_{i,\max} > 4$ to check the appropriateness of relationship (E4.6).

Table E4.2: Estimated expansion factors $K_i^{1/2}$ and MAP estimates ρ_i^{MAP} of ρ_i

Laboratory i	$K_i^{1/2}$	ρ_i^{MAP}
1	1.0	0.90
2	–	–
3	1.0	0.92
4	1.0	0.00
5	–	–
6	1.0	0.91
7	1.0	0.83
8	1.3	0.67
9	1.0	0.72
10	1.0	0.96
11	3.3	0.19

E4.7 Interpretation of results

For seven of the 11 laboratories (1, 3, 6, 7, 8, 9, 10), the estimated correlation coefficient was appreciable (between 0.67 and 0.96), implying the presence of systematic effects or biases in the measured values provided by those laboratories.[4]

Laboratory 4. The deviations from the reference values are nearly zero for two cases [$d_4(1)$ and $d_4(2)$], and considerably negative and positive respectively for the other two cases [$d_4(3)$ and $d_4(4)$]. The fact that no systematic effect can be seen in these deviations implies that the correlations are small, and ρ_4^{MAP} is actually zero to two decimal places.

Laboratory 8. For the conventional method, the minimum permissible expansion factors are 1.4, 1.0, 1.0 and 1.0 for the four measurands, suggesting that for that laboratory only the CMC uncertainty for the shortest gauge block requires expansion. However, if a common expansion factor for the CMC uncertainty throughout the measurement interval is required, the conventional method cannot suggest an appropriate value.

Laboratory 10. The correlation for laboratory 10, whose deviations are nearly zero for all four cases, is estimated to be very large ($\rho_{10}^{\mathrm{MAP}} = 0.96$). Because the small deviations suggest small random effects, the large uncertainties must depend on their systematic effects. Thus, the large correlation is theoretically reasonable.

Laboratory 11. Laboratory 11 reported a value for the shortest gauge block that was far from the KCRV, which is a likely cause for the resulting expansion factor of 3.3 for that laboratory. According to the final report on this KC [114], the laboratory found a problematic issue with its measuring system (see section E4.4.5). For that laboratory, whilst all the deviations are

---

[4]Since the model (E4.4) expresses the variation between the reported values, no bias is shown directly.

positive, the magnitudes are largely different. Consequently, the correlation is estimated to be not so large ($\rho_{11}^{\mathrm{MAP}} = 0.19$).

We conclude from these results and those for other KCs and CMC claims that the method exemplified here might also be useful in checking the validity of the measuring system over an interval of the parameter. We also believe that these results show the rationale of the MAP estimator to provide values for the expansion factors, and support the validity of the proposed method. Further, the results indicate that the correlation existing among each laboratory's measured values can be estimated.

Example E5

Measurement uncertainty when using quantities that change at a linear rate — use of quartz He reference leaks to calibrate an unknown leak

J. Greenwood, M.G. Cox

E5.1 Summary

There are numerous practical situations in which, a quantity of interest changes linearly with respect to another quantity. The mass flow rate from a reference leak as a function of time is an example of such a quantity. It is described here in terms of the depletion of helium from quartz membrane reference leaks.

However, the main purpose of the work is to demonstrate what is a generally applicable process for modelling the quantity and establishing the uncertainty associated with measured values of the quantity, including those situations where there is covariance within the data.

The intention when presenting this example is to include many of the intervening steps that, in published examples, might normally be omitted in providing the final result. Although this may make the treatment rather protracted for those who already have sufficient understanding of the subject, it is hoped that this approach will be useful to those readers wishing to gain understanding by following the evaluation in smaller steps. In addition, the cases are presented in terms of matrices and vectors (as in GUM Supplement 2 [54]), and in the perhaps more familiar notation of subscripted summations (as in the GUM [51]). The matrix notation can be ignored with no loss of completeness in the examples.

E5.2 Introduction of the application

E5.2.1 General

Leak detectors are commonly used instruments for identifying and quantifying the rate of gaseous material leaving (or entering) an otherwise sealed system. They are routinely used in non-destructive testing and as analytical tools in the vacuum industry.

At the heart of many such instruments is a detector that is selectively sensitive to a gas of interest. These detectors can be based upon a variety of principles ranging from solid state

chemical sensors to particle counters. One of the most commonly found types of leak detector is the helium mass spectrometer leak detector (MSLD).

Gas reference leaks, such as the quartz membrane He reference leak, are often found within mass spectrometer leak detectors where they are used to perform an 'internal' calibration of the gain of the system.

Quartz membrane leaks usually consist of a sealed reservoir containing the gas; the reservoir has an outlet connection that incorporates the membrane through which helium is able to permeate at a rate that depends on temperature. To ensure a steady depletion rate the leak is stored under stable conditions whilst not in use and is left 'open' (that is, not sealed) to maintain a stable gradient of He across the membrane. An example of such reference leaks is depicted in figure E5.1.

The leaks can be calibrated using gas flow meters of the type usually found in national metrology institutes. This would be normal practice for calibration of the 'master' reference leaks belonging to a calibration laboratory.

Alternatively, an 'unknown' leak can be calibrated by using two such reference leaks, which are chosen to 'bracket' the unknown leak (see Case 3 — section E5.8). Typically, the two reference leaks would be used to establish a linear calibration function for a measuring instrument over the intervening range. This would be normal practice for a calibration laboratory measuring 'unknown' leaks on behalf of its customers and is the subject of the scenarios presented here.

Both types of calibration are described in ISO 20486 [30], which in addition recommends that uncertainty in calibrated leak rate should be evaluated according to GUM [51] principles, but does not provide details of the evaluation process. A more general description of leaks and leak detectors can be found in [70] and the references therein.

Figure E5.1: Reference leaks externally mounted on a leak detector (photograph courtesy of Vaseco Ltd.)

E5.2.2 Scenarios

This example provides several scenarios that demonstrate the evaluation and use of values of a quantity that change at a linear rate. The scenarios are presented in terms of the depletion of a reference quantity over time, specifically, the depletion of He for a quartz membrane reference leak.

In all cases leak rate is the dependent variable. The independent variable is time or instrument response and measured values of both (time and response) have negligible uncertainty. Values of leak rate do have associated uncertainty and in some cases are correlated.

This example supports the related example EMUE A1.1.1; however the scenarios are intended to have general applicability for analogous measurements.

Case 1 sets out a basic situation in which there is no correlation. It is treated using the method

described in the ISO Technical Specification (ISO/TS 28037) concerned with the determination and use of straight-line calibration functions [16, clause 6].

Case 2 has correlations present. It is addressed by following the process described in the GUM [51, annex F.1.2] and in GUM Supplement 2 [54] to calculate covariance. The covariance is then taken into account in the evaluation of the fitting parameters using the method described in ISO/TS 28037 [16, clause 9].

Case 3 considers a situation where there are *two* independent reference leaks (each individually corresponding to leaks described in Case 1). These are used together to calibrate a leak measuring instrument, a mass spectrometer leak detector that is subsequently used to calibrate an 'unknown' leak. This scenario demonstrates how correlation arises between the values assigned to each leak when both leaks are in use together.

Case 3 goes on to provide a demonstration of the treatment of correlation in the *use* of these leaks. In practice this correlation is usually neglected. This example will demonstrate how it can be appropriately incorporated in a GUM law of propagation of uncertainty (LPU) type evaluation.

E5.3   Specification of the measurands

In all three cases, the measurand of primary interest is the leak (flow) rate Q of helium when the leak is operating at reference temperature T_0. There are also other measurands of interest at intermediate stages within each scenario — these are the coefficients a and b of various straight-line calibration functions for two reference leaks L_1 and L_2 and for the MSLD.

E5.4   Measurement model

The measurement model embodied in the following scenarios consists (in the first part) of steps to establish estimates for the measurands a and b, the coefficients of a straight-line fit through the given calibration data, and subsequently (in the second part) use of these coefficients and other data to calculate a value for the measurand Q corresponding to leak rate at a defined reference temperature. An underpinning concept, employed throughout, is that of a straight-line calibration function as defined and elaborated in ISO/TS 28037 [16]. Case 1 uses clause 6, and Case 2 uses clause 9 of that Technical Specification. Case 3 makes use of both clauses.

E5.5   Uncertainty propagation

There is uncertainty associated with each leak rate value and, in cases 2 and 3, there is correlation between these quantities. The independent quantities are either time or detector response; it is assumed that there is negligible uncertainty in their associated values during fitting. (If this is not the case then the treatment of clause 7 in ISO/TS 28037 [16] applies in the absence of correlations; otherwise the more general treatment of clause 10 becomes necessary.)

Measurement uncertainty evaluation follows the standard LPU approach outlined in ISO/TS 28037 [16], the GUM [51] and GUM Supplement 2 [54]. In particular, it follows the guidance on treatment of correlations elaborated in GUM Annex F.1.2 and in GUM Supplement 2 clause 6.2.

E5.6   Case 1: No correlation within the data

Consider a reference leak L_1. When not in use the leak is stored under fixed and stable conditions, which are sufficient to maintain a linear depletion rate over the course of time. It is periodically calibrated and it is assumed that there is no correlation within the calibration data. The calibrated results for reference leak L_1 are given in figure E5.2 and table E5.1.
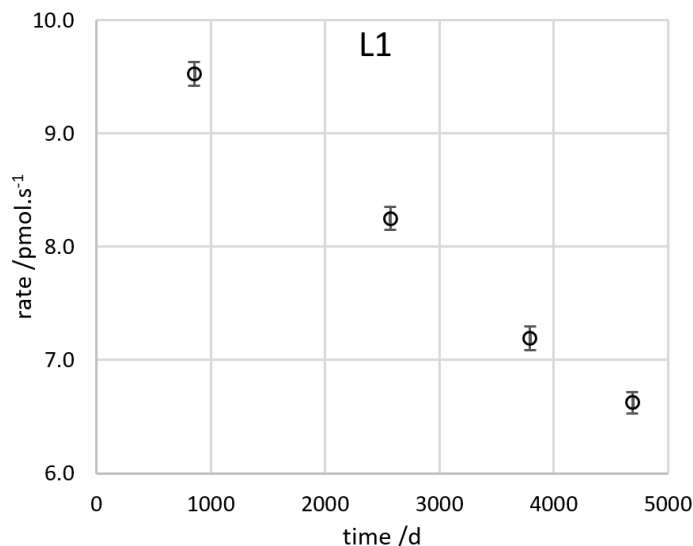
Figure E5.2: Calibration results for a reference leak L_1. Data points represent the reference value with error bars corresponding to ± 1 standard uncertainty

Table E5.1: Calibrated results for reference leak L_1

t/d	Q/pmol s^{-1}	$u(Q)$/pmol s^{-1}
857	9.525	0.105
2571	8.250	0.103
3792	7.192	0.103
4689	6.623	0.094

The reference value Q corresponding to time t and temperature T_0 is to be established by forward evaluation using a straight-line calibration function for the reference leak:

$$Q = a_1 + b_1 t, \tag{E5.1}$$

where (a_1, b_1) are the coefficients of the function.

Since there is no covariance in the data, a model corresponding to clause 6 of ISO/TS 28037 [16] is assumed to apply. The associated evaluation can be readily implemented in a spreadsheet. It should be noted that, if in addition there had been uncertainty in the time (independent variable) data but still no correlations, the approach in clause 7 of [16] could instead be followed. This approach is also relatively straightforward to implement in a spreadsheet calculation.

The results are found to be:

$$a_1 = 10.185 \, \text{pmol s}^{-1},$$
$$b_1 = -7.678 \times 10^{-4} \, \text{pmol s}^{-1},$$
$$u(a_1) = 0.119 \, \text{pmol s}^{-1},$$
$$u(b_1) = 3.506 \times 10^{-5} \, \text{pmol s}^{-1} \text{d}^{-1},$$
$$u(a_1, b_1) = -3.785 \times 10^{-6} \, \text{pmol}^2 \text{s}^{-2} \text{d}^{-1}.$$

The computed value Q has associated uncertainty $u(Q)$ given by clause 11 of ISO/TS 28037 which is concerned with the use of the calibration function:

$$u^2(Q) = \boldsymbol{C}_Q^\top \boldsymbol{V}_Q \boldsymbol{C}_Q,$$

where $\boldsymbol{C}_Q$ is an array containing the sensitivity coefficients, and $\boldsymbol{V}_Q$ is the corresponding covariance matrix:

$$
\boldsymbol{C}_Q = \begin{bmatrix} \frac{\partial Q}{\partial a_1} \\ \frac{\partial Q}{\partial b_1} \\ \frac{\partial Q}{\partial t} \end{bmatrix} = \begin{bmatrix} 1 \\ t \\ b_1 \end{bmatrix}, \qquad
\boldsymbol{V}_Q = \begin{bmatrix} u^2(a_1) & u(a_1, b_1) & 0 \\ u(a_1, b_1) & u^2(b_1) & 0 \\ 0 & 0 & u^2(t) \end{bmatrix},
$$

which is equivalent to

$$u^2(Q) = u^2(a_1) + 2tu(a_1, b_1) + t^2 u^2(b_1) + b_1^2 u^2(t),$$

where $u^2(t)$ is the variance associated with the time of use t.

The expression for $u^2(Q)$ is the same as that found by applying GUM equation (13) to equation (E5.1).

Suppose that the leak is to be used at $t = 5000 \, \text{d}$ and $u(t) = 1 \, \text{d}$. Applying forward evaluation using the above parameter values, the result for the computed value of the reference leak is

$$Q = 6.346 \, \text{pmol s}^{-1}, \qquad u(Q) = 0.084 \, \text{pmol s}^{-1}.$$

E5.7   Case 2: Correlation between leak rate data

Suppose there is a degree of correlation between each of the calibration results for reference leak L_1.

E5.7.1   Measurement model

In this situation (following example in Annex D of [16]) the leak rate data Q_j can be modelled in terms of the observed rate Q_{oj} and a common systematic effect, represented by e_s:

$$Q_j = Q_{oj} + e_s, \tag{E5.2}$$

where $j = 1, \ldots, m$ for m measurement data points.

All known corrections are assumed to have been made; therefore, the best estimate of e_s (its expectation) is zero, with a standard uncertainty $u(e_s)$. In this scenario, as will shortly be seen, a value of $u(e_s)$ is determined from available knowledge of the systematic effects contributing to a calibration correction.

The observed values Q_{oj} have uncertainties $u(Q_{oj})$ that are established in the normal GUM-LPU manner for all effects other than e_s.

Suppose that the estimate Q_{oj} is based upon an observed value to which a calibration correction c_j has been applied, where c_j has standard uncertainty $u(c_j)$. Suppose also that the calibration process is itself subject to certain effects that are essentially random in nature contributing a standard uncertainty $u(r_j)$ to the overall standard uncertainty; and to ~~other~~ poorly understood systematic effects that will be *the same each time* a calibration is performed, contributing a standard uncertainty $u(s)$ to the overall uncertainty. The calibration standard uncertainty $u(c_j)$ is therefore given by

$$u^2(c_j) = u^2(r_j) + u^2(s).$$

Suppose further (for sake of realistic demonstration) that besides calibration effects there are two other, independent effects influencing the measurement of Q_{oj} with corresponding uncertainties, $u(e_{1j})$ and $u(e_{2j})$. These might for example be the uncertainty associated with correction of a known bias and the uncertainty associated with finite resolution of observed indications.

In this scenario we therefore have

$$u^2(Q_{oj}) = u^2(e_{1j}) + u^2(e_{2j}) + u^2(r_j),$$
$$u^2(e_s) = u^2(s),$$

which when combined give the result

$$\begin{aligned}
u^2(Q_j) &= u^2(Q_{oj}) + u^2(e_s), \\
&= u^2(e_{1j}) + u^2(e_{2j}) + u^2(r_j) + u^2(s).
\end{aligned} \tag{E5.3}$$

To illustrate this scenario, consider the calibration results in table E5.2.

Table E5.2: Uncertainty contributions for reference leak L_1

t/d	$Q_j/\mathrm{pmol\,s^{-1}}$	$u(Q_{oj})/\mathrm{pmol\,s^{-1}}$	$u(e_s)/\mathrm{pmol\,s^{-1}}$
857	9.525	0.090	0.055
2571	8.250	0.087	0.055
3792	7.192	0.087	0.055
4689	6.623	0.076	0.055

As in Case 1, the reference value Q corresponding to a time t and temperature T_0 is to be established by forward evaluation using a straight-line calibration function for the reference leak:

$$Q = a + bt, \tag{E5.4}$$

where (a, b) are the coefficients of the function.

Since the quantity e_s is common to all leak rate measurements, there will be correlation between the quantities Q_j; so a measurement model corresponding to clause 9 of ISO/TS 28037:2010 [16] is adopted to establish values for the coefficients.

Firstly though, a covariance matrix V_Q is needed that describes the correlations within the Q data. This is established by following the process described, for example, in annex F.1.2.3 of the GUM and in GUM Supplement 2, clause 6.2. This process involves defining functions f_j of quantities x_i such that

$$Q_j = f_j(x_i),$$

with $i = 1, \ldots, N$ and $j = 1, \ldots, m$; thus

$$
\begin{aligned}
Q_1 &= f_1(x_i) = f_1(Q_{o1}, Q_{o2}, Q_{o3}, Q_{o4}, e_s) = Q_{o1} + e_s, \\
Q_2 &= f_2(x_i) = f_2(Q_{o1}, Q_{o2}, Q_{o3}, Q_{o4}, e_s) = Q_{o2} + e_s, \\
Q_3 &= f_3(x_i) = f_3(Q_{o1}, Q_{o2}, Q_{o3}, Q_{o4}, e_s) = Q_{o3} + e_s, \\
Q_4 &= f_4(x_i) = f_4(Q_{o1}, Q_{o2}, Q_{o3}, Q_{o4}, e_s) = Q_{o4} + e_s,
\end{aligned}
$$

that is, the functions f_j are defined in terms of *all* quantities x_i that influence *all* Q_j, even though some of the quantities only have an effect in one or other functions.

In terms of matrices (as used in GUM Supplement 2, clause 6.2)

$$ Y = Q = f(X), $$

where

$$
Y = Q = \begin{bmatrix} Q_1 \\ Q_2 \\ Q_3 \\ Q_4 \end{bmatrix}; \qquad
X = \begin{bmatrix} Q_{o1} \\ Q_{o2} \\ Q_{o3} \\ Q_{o4} \\ e_s \end{bmatrix}.
$$

The covariance matrix V_Q is given by

$$
V_Q = C_x V_x C_x^{\mathsf{T}} =
\begin{bmatrix}
u^2(Q_1) & u(Q_1, Q_2) & u(Q_1, Q_3) & u(Q_1, Q_4) \\
u(Q_2, Q_1) & u^2(Q_2) & u(Q_2, Q_3) & u(Q_2, Q_4) \\
u(Q_3, Q_1) & u(Q_3, Q_2) & u^2(Q_3) & u(Q_3, Q_4) \\
u(Q_4, Q_1) & u(Q_4, Q_2) & u(Q_4, Q_3) & u^2(Q_4)
\end{bmatrix},
$$

where

$$
C_x =
\begin{bmatrix}
\frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_N} \\
\vdots & \ddots & \vdots \\
\frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_N}
\end{bmatrix}
=
\begin{bmatrix}
1 & 0 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 & 1 \\
0 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 1 & 1
\end{bmatrix}
$$

and

$$
V_x =
\begin{bmatrix}
u^2(Q_{o1}) & 0 & 0 & 0 & 0 \\
0 & u^2(Q_{o2}) & 0 & 0 & 0 \\
0 & 0 & u^2(Q_{o3}) & 0 & 0 \\
0 & 0 & 0 & u^2(Q_{o4}) & 0 \\
0 & 0 & 0 & 0 & u^2(e_s)
\end{bmatrix}.
$$

Alternatively, the components of the covariance matrix can be evaluated in terms of subscripted summations. Thus, the variance, $u^2(Q_j)$ for Q_j can be calculated using GUM equation (F.1) [that is, GUM equation (10)]:

$$u^2(Q_j) = \sum_{i=1}^{N} \left(\frac{\partial f_j}{\partial x_i} \right)^2 u^2(x_i),$$

and the covariance terms $u(Q_j, Q_{k \neq j})$ can be calculated using GUM (F.2):

$$u(Q_j, Q_{k \neq j}) = \sum_{i=1}^{N} \frac{\partial f_j}{\partial x_i} \frac{\partial f_k}{\partial x_i} u^2(x_i).$$

Note that in cases where any of the terms $u(x_i, x_{k \neq i}) \neq 0$, that is, off-diagonal terms are not equal to zero, then GUM formulæ (F.1) and (F.2) can no longer be used and, noting that $u(Q_j, Q_j) = u^2(Q_j)$, all terms of the covariance matrix V_Q are instead given by

$$u(Q_j, Q_k) = \sum_{i=1}^{N} \sum_{\ell=1}^{N} \frac{\partial f_j}{\partial x_i} \frac{\partial f_k}{\partial x_\ell} u(x_i, x_\ell). \tag{E5.5}$$

Whichever approach is used, matrix or subscripted summations, the result is that

$$u^2(Q_1) = u^2(Q_{o1}) + u^2(e_s),$$
$$u^2(Q_2) = u^2(Q_{o2}) + u^2(e_s),$$
$$u^2(Q_3) = u^2(Q_{o3}) + u^2(e_s),$$
$$u^2(Q_4) = u^2(Q_{o4}) + u^2(e_s).$$

and

$$u(Q_j, Q_{k \neq j}) = u^2(e_s).$$

E5.7.2   Model fitting

In matrix form, the data for fitting by method ISO/TS 28037 clause 9, expressed in terms of the quantities used therein, correspond to

$$x = t = \begin{bmatrix} t_1 \\ t_2 \\ t_3 \\ t_4 \end{bmatrix}, \quad y = Q = \begin{bmatrix} Q_1 \\ Q_2 \\ Q_3 \\ Q_4 \end{bmatrix}, \quad V_y = \begin{bmatrix} u^2(Q_1) & u(Q_1,Q_2) & u(Q_1,Q_3) & u(Q_1,Q_4) \\ u(Q_2,Q_1) & u^2(Q_2) & u(Q_2,Q_3) & u(Q_2,Q_4) \\ u(Q_3,Q_1) & u(Q_3,Q_2) & u^2(Q_3) & u(Q_3,Q_4) \\ u(Q_4,Q_1) & u(Q_4,Q_2) & u(Q_4,Q_3) & u^2(Q_4) \end{bmatrix}.$$

In this example we have

$$x = \begin{bmatrix} 857 \\ 2571 \\ 3792 \\ 4689 \end{bmatrix} \text{d}, \quad y = \begin{bmatrix} 9.525 \\ 8.250 \\ 7.192 \\ 6.623 \end{bmatrix} \text{pmol s}^{-1},$$

$$V_y = \begin{bmatrix} 0.0111 & 0.0030 & 0.0030 & 0.0030 \\ & 0.0106 & 0.0030 & 0.0030 \\ & & 0.0106 & 0.0030 \\ \text{sym.} & & & 0.0088 \end{bmatrix} (\text{pmol/s})^2.$$

In practice, the *correlation matrix* may be of more intuitive interest than the covariance matrix (and has the advantage of being dimensionless). This is defined in terms of the covariance matrix and component uncertainties by

$$R(y_j, y_k) = \frac{u(y_j, y_k)}{u(y_j)\, u(y_k)};$$

hence

$$\boldsymbol{R_y} = \begin{bmatrix} 1 & 0.279 & 0.279 & 0.306 \\ & 1 & 0.286 & 0.313 \\ & & 1 & 0.313 \\ \text{sym.} & & & 1 \end{bmatrix}. \tag{E5.6}$$

The results of the fitting are

$$a = 10.184\,\mathrm{pmol\,s^{-1}},$$
$$b = -7.671 \times 10^{-4}\,\mathrm{pmol\,s^{-1}\,d^{-1}},$$
$$u(a) = 0.115\,\mathrm{pmol\,s^{-1}},$$
$$u(b) = 2.939 \times 10^{-5}\,\mathrm{pmol\,s^{-1}\,d^{-1}},$$
$$u(a, b) = -2.708 \times 10^{-6}\,\mathrm{pmol^2\,s^{-2}\,d^{-1}}.$$

ISO/TS 28037 [16] provides algorithms to perform the necessary calculations to evaluate a, b, $u(a)$, $u(b)$ and $u(a, b)$. Unfortunately, they are not generally amenable to implementation using spreadsheet cell formulae and some other means of solving, such as a mathematical software package or user-written code, is required. For example, ISO/TS 20837 Annex F describes software and source code that is provided free by NPL.

The standard uncertainty $u(Q)$ for a forward evaluation using equation (E5.4) is evaluated by a standard GUM-LPU approach. This can be expressed in matrix format as in ISO/TS 28037, clause 11.2:

$$u^2(Q) = C_Q^\top V_Q C_Q,$$

where C_Q is an array containing the sensitivity coefficients, and V_Q is the corresponding covariance matrix:

$$C_Q = \begin{bmatrix} \frac{\partial Q}{\partial a_1} \\ \frac{\partial Q}{\partial b_1} \\ \frac{\partial Q}{\partial t} \end{bmatrix} = \begin{bmatrix} 1 \\ t \\ b \end{bmatrix}, \qquad V_Q = \begin{bmatrix} u^2(a) & u(a, b) & 0 \\ u(a, b) & u^2(b) & 0 \\ 0 & 0 & u^2(t) \end{bmatrix},$$

which equates to

$$u^2(Q) = u^2(a) + t^2 u^2(b) + b^2 u^2(t) + 2tu(a, b).$$

This is the same expression that is found by applying GUM equation (13).

For example, a forward evaluation using equation (E5.4) at say $t = 5000\,\mathrm{d}$ with $u(t) = 1\,\mathrm{d}$ gives $Q = 6.349\,\mathrm{pmol\,s^{-1}}$ and standard uncertainty $u(Q) = 0.088\,\mathrm{pmol\,s^{-1}}$.

If there was no correlation in the data ...

The corresponding results of fitting can be evaluated for the situation where there is *no correlation in the data*, that is, the effect characterised by $u(s)$ is in this case *not* common to each flow calibration measurement. The data model is now described by:

$$Q_j = Q_{oj}, \tag{E5.7}$$
$$u^2(Q_j) = u^2(Q_{oj}) = u^2(e_{1j}) + u^2(e_{2j}) + u^2(c_j), \tag{E5.8}$$

which corresponds to the data model in Case 1. For the data in table E5.3, the process described in clause 6 of [16] can be used to calculate the fitting results in this case, giving the following:

$$
\begin{aligned}
a &= 10.185 \, \text{pmol s}^{-1} & &= a_1, \\
b &= -7.678 \times 10^{-4} \, \text{pmol s}^{-1} \, \text{d}^{-1} & &= b_1, \\
u(a) &= 0.119 \, \text{pmol s}^{-1} & &= u(a_1), \\
u(b) &= 3.506 \times 10^{-5} \, \text{pmol s}^{-1} \, \text{d}^{-1} & &= u(b_1), \\
u(a,b) &= -3.785 \times 10^{-6} \, \text{pmol}^2 \text{s}^{-2} \text{d}^{-1} & &= u(a_1, b_1), \tag{E5.9}
\end{aligned}
$$

for which a forward evaluation at $t = 5000 \, \text{d}$ with $u(t) = 1 \, \text{d}$ gives the estimate $Q = 6.346 \, \text{pmol s}^{-1}$ and standard uncertainty $u(Q) = 0.084 \, \text{pmol s}^{-1}$.

Table E5.3: Calibration results for reference leak L_1

t/d	$Q_j/\text{pmol s}^{-1}$	$u(Q_j)/\text{pmol s}^{-1}$
857	9.525	0.105
2571	8.250	0.103
3792	7.192	0.103
4689	6.623	0.094

The difference between the results in the two different scenarios (correlation and no correlation) is not large *in this particular example*; however, the extent of the difference is entirely dependent upon the data.

Further data for a second leak

For later reference (in Case 3 — section E5.8), consider a second leak L_2 for which the data in table E5.4 is available, where the data model for L_2 is as described above for L_1 in equations (E5.2) and (E5.3).

Table E5.4: Data for reference leak L_2

t/d	$Q_j/\text{pmol s}^{-1}$	$u(Q_{oj})/\text{pmol s}^{-1}$	$u(e_s)/\text{pmol s}^{-1}$
100	4.391	0.046	0.055
474	4.293	0.045	0.055
856	4.190	0.044	0.055
2568	3.724	0.041	0.055
3791	3.531	0.040	0.055
4692	3.402	0.037	0.055

The results of the fitting for leak L_2 are

$$a = 4.376 \,\text{pmol s}^{-1},$$
$$b = -2.183 \times 10^{-4} \,\text{pmol s}^{-1}\,\text{d}^{-1},$$
$$u(a) = 0.062 \,\text{pmol s}^{-1},$$
$$u(b) = 9.706 \times 10^{-6} \,\text{pmol s}^{-1}\,\text{d}^{-1},$$
$$u(a, b) = -2.208 \times 10^{-7} \,\text{pmol}^2\text{s}^{-2}\text{d}^{-1},$$

with

$$
\boldsymbol{R_y} =
\begin{bmatrix}
1 & 0.594 & 0.599 & 0.615 & 0.620 & 0.636 \\
 & 1 & 0.604 & 0.621 & 0.626 & 0.642 \\
 & & 1 & 0.626 & 0.632 & 0.648 \\
 & & & 1 & 0.648 & 0.665 \\
 & & & & 1 & 0.671 \\
\text{sym.} & & & & & 1
\end{bmatrix}.
\tag{E5.10}
$$

Correlation in this case is considerably higher than previously as seen by comparing the off-diagonal terms in the correlation matrix with those in the matrix (E5.6). Such a statement could not easily be made by examining covariance matrices.

Forward evaluation using equation (E5.4), again at $t = 5000 \,\text{d}$ with $u(t) = 1 \,\text{d}$ gives the estimate $Q = 3.284 \,\text{pmol s}^{-1}$ and standard uncertainty $u(Q) = 0.063 \,\text{pmol s}^{-1}$.

The corresponding results of fitting a straight line can again be evaluated for the situation where there is no correlation, as detailed in equations (E5.7) and (E5.8). In this case the data is given in table E5.5.

Table E5.5: Calibration results for reference leak L_2

t/d	$Q_j/\text{pmol s}^{-1}$	$u(Q_j)/\text{pmol s}^{-1}$
100	4.391	0.072
474	4.293	0.071
856	4.190	0.070
2568	3.724	0.069
3791	3.531	0.068
4692	3.402	0.066

The process described in clause 6 of [16] can again be used to calculate the fitting results in this case, giving

$$a_2 = 4.380 \,\text{pmol s}^{-1},$$
$$b_2 = -2.200 \times 10^{-4} \,\text{pmol s}^{-1}\,\text{d}^{-1},$$
$$u(a_2) = 0.045 \,\text{pmol s}^{-1},$$
$$u(b_2) = 1.620 \times 10^{-5} \,\text{pmol s}^{-1}\,\text{d}^{-1},$$
$$u(a_2, b_2) = -5.713 \times 10^{-7} \,\text{pmol}^2\text{s}^{-2}\text{d}^{-1}.
\tag{E5.11}$$

A forward evaluation at $t = 5000 \,\text{d}$ with $u(t) = 1 \,\text{d}$ gives the estimate $Q = 3.280 \,\text{pmol s}^{-1}$ and standard uncertainty $u(Q) = 0.054 \,\text{pmol s}^{-1}$.

E5.8   Case 3: Use of two reference leaks to calibrate a third unknown leak

In this scenario, the two reference leaks L_1 and L_2 are used to calibrate a leak detector at points bracketing the value of an uncalibrated leak L_x. The previously determined calibration functions for the reference leaks are used to establish reference values at the time of use. Each leak rate is then calculated for its *prevailing* temperature and the corresponding MSLD response is observed. A linear fit is then performed to this stimulus-response data to calibrate the MSLD. Finally, taking the MSLD response from the 'unknown' leak, the corresponding leakage rate is evaluated and expressed in terms of a defined reference temperature.

E5.8.1   Specification of the measurands

In this scenario the principal measurand is the reference value Q_x for the 'unknown' leak L_x. During the evaluation process it is necessary to evaluate intervening measurands a_M and b_M, the coefficients of the MSLD calibration function.

E5.8.2   Measurement model

Reference values

The reference values Q_1 and Q_2 corresponding to a time t and temperature T_0 are established by forward evaluation using the straight-line calibration functions for each reference leak:

$$Q_1 = a_1 + b_1 t, \qquad Q_2 = a_2 + b_2 t. \tag{E5.12}$$

Measured values

In use at temperatures T_1 and T_2 respectively, the two reference leaks L_1 and L_2 produce helium at rates q_1 and q_2 given by:

$$q_1 = Q_1[1 + \alpha(\Delta T_1 + \delta T)] = (a_1 + b_1 t)[1 + \alpha(\Delta T_1 + \delta T)],$$
$$q_2 = Q_2[1 + \alpha(\Delta T_2 + \delta T)] = (a_2 + b_2 t)[1 + \alpha(\Delta T_2 + \delta T)], \tag{E5.13}$$

where

  α is the temperature coefficient for the depletion rate in the region of reference temperature T_0, assumed to be the same for both reference leaks,

$\Delta T_1 = T_1 - T_0$,

$\Delta T_2 = T_2 - T_0$,

  T_1 is the temperature assigned to leak L_1, for example, measured temperature of its leak housing or coupling,

  T_2 is the temperature assigned to leak L_2, for example, measured temperature of its leak housing or coupling,

  T_0 is the reference temperature for the leaks,

δT is the (unknown) temperature measurement error, corresponding to the difference between the assigned temperature and the actual temperature of the quartz membrane (which controls the rate of helium permeation). The best estimate (expectation) of δT is zero but the uncertainty is finite. This is an example of a poorly known systematic effect, as described in GUM Supplement 3 [56, clause 10.4].

MSLD Calibration

The MSLD responses corresponding to q_1 and q_2 are p_1 and p_2, respectively, and it is assumed here that the associated standard uncertainties $u(p_1)$, $u(p_2)$ are negligible. The calibration function for the MSLD established from the data (p_1, q_1) and (p_2, q_2) and the associated covariance is

$$q = a_M + b_M p, \tag{E5.14}$$

where p is the MSLD response and q is the corresponding leak rate.

Calibration of unknown leak

The leak rate q_x corresponding to MSLD response p_x for a leak L_x operating at temperature T_x can now be evaluated and the value Q_x can be established that is referenced to a temperature T_0; thus

$$q_x = a_M + b_M p_x \tag{E5.15}$$

and

$$Q_x = \frac{q_x}{[1 + \alpha_x(\Delta T_x + \delta T)]}, \tag{E5.16}$$

where

α_x is the temperature coefficient for the depletion rate in the region of the reference temperature,

$\Delta T_x = T_x - T_0$

T_x is the temperature assigned to leak L_x, for example, measured temperature of its leak housing or coupling,

T_0 is the reference temperature for the leaks, assumed to be the same for all three leaks,

δT is the (unknown) temperature error, also assumed to be the same for all three leaks.

E5.8.3 Uncertainty Propagation

Reference values

Correlation within the data for the reference leaks could appear in various forms that are, for reasons of space, not considered here but are nevertheless treatable using the methods in ISO/TS 28037 [16]. The most likely two such scenarios are, firstly, the common effect described in Case 2 is present for *both* leaks and for *all* values (as might arise when the leaks are calibrated using the same method with the equipment having the same traceability for all reported results); or, secondly, there is a 'pair-wise' common effect between corresponding values for the two leaks, but little or no correlation *within* the data for each leak (as might arise if the leaks are both calibrated at the same time but the method, equipment and traceability are not fixed as in the first case).

In our example scenario we shall assume that the correlations within and between the data for each reference leak are not significant and the results evaluated in Case 2, equations (E5.9) and (E5.11), will be used.

Measured values

Since the quantities t, α and δT are common to both leak rate expressions [equations (E5.13)], there will be correlation between the estimates q_1 and q_2 of those quantities.

A covariance evaluation is needed that represents the correlations in the data. This can be established by following the process described in matrix form in clause 6.2 of GUM Supplement 2 [54] and in subscripted summation form in GUM Annex F.1.2.3.

The process begins by defining two functions f_1 and f_2 from equations (E5.13) such that

$$q_1 = f_1(x_i) = f_1(a_1, b_1, a_2, b_2, t, \alpha, \Delta T_1, \Delta T_2, \delta T) = (a_1 + b_1 t)[1 + \alpha(\Delta T_1 + \delta T)],$$
$$q_2 = f_2(x_i) = f_2(a_1, b_1, a_2, b_2, t, \alpha, \Delta T_1, \Delta T_2, \delta T) = (a_2 + b_2 t)[1 + \alpha(\Delta T_2 + \delta T)],$$

that is, f_1 and f_2 are defined in terms of *all* quantities x_i that influence *both* q_1 and q_2, even though some of the quantities only have an effect in one or other function.

In the terminology of GUM Supplement 2 [54] clause 6.2, the quantities are

$$X = (a_1 \ b_1 \ a_2 \ b_2 \ t \ \alpha \ \Delta T_1 \ \Delta T_2 \ \delta T)^{\mathsf{T}},$$
$$Y = (q_1 \ q_2)^{\mathsf{T}}.$$

The covariance matrix V_y is given by

$$V_y = C_x V_x C_x^{\mathsf{T}} = \begin{bmatrix} u^2(q_1) & u(q_1, q_2) \\ u(q_2, q_2) & u^2(q_2) \end{bmatrix},$$

where, in this example, $N = 9$, $m = 2$ and we have

$$C_x^{\mathsf{T}} = \begin{bmatrix} \frac{\partial f_1}{\partial a_1} & \frac{\partial f_2}{\partial a_1} \\ \frac{\partial f_1}{\partial b_1} & \frac{\partial f_2}{\partial b_1} \\ \frac{\partial f_1}{\partial a_2} & \frac{\partial f_2}{\partial a_2} \\ \frac{\partial f_1}{\partial b_2} & \frac{\partial f_2}{\partial b_2} \\ \frac{\partial f_1}{\partial t} & \frac{\partial f_2}{\partial t} \\ \frac{\partial f_1}{\partial \alpha} & \frac{\partial f_2}{\partial \alpha} \\ \frac{\partial f_1}{\partial \Delta T_1} & \frac{\partial f_2}{\partial \Delta T_1} \\ \frac{\partial f_1}{\partial \Delta T_2} & \frac{\partial f_2}{\partial \Delta T_2} \\ \frac{\partial f_1}{\partial \delta T} & \frac{\partial f_2}{\partial \delta T} \end{bmatrix} = \begin{bmatrix} 1 + \alpha(\Delta T_1 + \delta T) & 0 \\ t\,[1 + \alpha(\Delta T_1 + \delta T)] & 0 \\ 0 & 1 + \alpha(\Delta T_2 + \delta T) \\ 0 & t\,[1 + \alpha(\Delta T_2 + \delta T)] \\ b_1\,[1 + \alpha(\Delta T_1 + \delta T)] & b_2\,[1 + \alpha(\Delta T_2 + \delta T)] \\ (a_1 + b_1 t)(\Delta T_1 + \delta T) & (a_2 + b_2 t)(\Delta T_2 + \delta T) \\ \alpha\,(a_1 + b_1 t) & 0 \\ 0 & \alpha\,(a_2 + b_2 t) \\ \alpha\,(a_1 + b_1 t) & \alpha\,(a_2 + b_2 t) \end{bmatrix}$$

and

$$
V_x = \begin{bmatrix}
u^2(a_1) & u(a_1, b_1) & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
u(a_1, b_1) & u^2(b_1) & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & u^2(a_2) & u(a_2, b_2) & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & u(a_2, b_2) & u^2(b_2) & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & u^2(t) & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & u^2(\alpha) & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & u^2(\Delta T_1) & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & u^2(\Delta T_2) & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & u^2(\delta T)
\end{bmatrix}.
$$

Alternatively, the covariance matrix can be calculated in terms of subscripted summations in line with annex F.1.2 of the GUM [51], albeit using equation (E5.5) rather than GUM equations (F.1) and (F.2) as several of the quantities are correlated (namely a_1 with b_1 and a_2 with b_2.

This gives (remembering that the expectation of δT is zero),

$$
\begin{aligned}
u^2(q_1) = {} & u(q_1, q_1) \\
= {} & (1 + \alpha\Delta T_1)^2 u^2(a_1) + t^2(1 + \alpha\Delta T_1)^2 u^2(b_1) + b_1^2(1 + \alpha\Delta T_1)^2 u^2(t) \\
& + \Delta T_1{}^2(a_1 + b_1 t)^2 u^2(\alpha) + \alpha^2(a_1 + b_1 t)^2 u^2(\Delta T_1) + \alpha^2(a_1 + b_1 t)^2 u^2(\delta T) \\
& + 2t(1 + \alpha\Delta T_1)^2 u(a_1, b_1),
\end{aligned}
$$

$$
\begin{aligned}
u^2(q_2) = {} & u(q_2, q_2) \\
= {} & (1 + \alpha\Delta T_2)^2 u^2(a_2) + t^2(1 + \alpha\Delta T_2)^2 u^2(b_2) + b_2^2(1 + \alpha\Delta T_2)^2 u^2(t) \\
& + \Delta T_2{}^2(a_2 + b_2 t)^2 u^2(\alpha) + \alpha^2(a_2 + b_2 t)^2 u^2(\Delta T_2) + \alpha^2(a_2 + b_2 t)^2 u^2(\delta T) \\
& + 2t(1 + \alpha\Delta T_2)^2 u(a_2, b_2)
\end{aligned}
$$

and

$$
\begin{aligned}
u(q_1, q_2) = {} & u(q_2, q_1) \\
= {} & b_1 b_2(1 + \alpha\Delta T_1)(1 + \alpha\Delta T_2) u^2(t) \\
& + (a_1 + b_1 t)(a_2 + b_2 t)\Delta T_1 \Delta T_2 u^2(\alpha) \\
& + \alpha^2(a_1 + b_1 t)(a_2 + b_2 t) u^2(\delta T).
\end{aligned}
$$

MSLD Calibration

In matrix form, the data for fitting by method ISO/TS 28037, clause 9 (expressed in terms of the variables used in [16]) correspond to

$$
x = p = \begin{bmatrix} p_1 \\ p_2 \end{bmatrix}, \qquad y = q = \begin{bmatrix} q_1 \\ q_2 \end{bmatrix}, \qquad V_y = V_{q_1, q_2} = \begin{bmatrix} u^2(q_1) & u(q_1, q_2) \\ u(q_1, q_2) & u^2(q_2) \end{bmatrix}.
$$

Solving the model establishes estimates for the coefficients a_M and b_M [for equation (E5.14)] and the elements of the covariance matrix

$$V_{a,b} = \begin{bmatrix} u^2(a_M) & u(a_M, b_M) \\ u(a_M, b_M) & u^2(b_M) \end{bmatrix}.$$

Note on calculations

Note that in general the calculations in ISO/TS 28037 clause 9, Steps 1 and 2 cannot easily be implemented within a spreadsheet and some means of performing matrix algebra is required; however, the solution when fitting to just *two* data points can be written out in a relatively short form that *is* amenable to spreadsheet evaluation.

Step 1: described in clause 9.2.2 of ISO/TS 28037 requires factorisation of the covariance matrix. For a matrix V such as V_{q_1, q_2} established in Case 3 above, this involves calculating the components of a lower left matrix L such that

$$V = \begin{pmatrix} v_1 & v_2 \\ v_2 & v_3 \end{pmatrix} = LL^\top$$

and

$$L = \begin{pmatrix} l_1 & 0 \\ l_2 & l_3 \end{pmatrix},$$

which is satisfied when

$$l_1 = \sqrt{v_1},$$
$$l_2 = \frac{v_2}{\sqrt{v_1}},$$
$$l_3 = \sqrt{v_3 - \frac{v_2^2}{v_1}}.$$

Step 2: described in clause 9.2.2 of ISO/TS 28037 requires solving several systems of equations to establish values for variables identified as f, g and h. For the 2-point systems described above the values are found to be:

$$f_1 = \frac{1}{l_1},$$
$$f_2 = \frac{1 - (l_2/l_1)}{l_3},$$
$$g_1 = \frac{p_1}{l_1},$$
$$g_2 = \frac{p_2 - p_1(l_2/l_1)}{l_3},$$
$$h_1 = \frac{q_1}{l_1},$$
$$h_2 = \frac{q_2 - q_1(l_2/l_1)}{l_3}.$$

Calibration of unknown leak

Forward evaluation, to establish a value for an unknown leak rate q_x and its associated standard uncertainty $u(q_x)$ from an observed MSLD response p_x and associated standard uncertainty $u(p_x)$, uses equation (E5.15):

$$q_x = a_M + b_M p_x,$$

and again follows the process described in clause 11 of ISO/TS 28037 [16], giving in matrix form

$$u^2(q_x) = C_x^\top V_x C_x,$$

where C_x is an array containing the sensitivity coefficients and V_x is the corresponding covariance matrix

$$C_x = \begin{bmatrix} \frac{\partial q_x}{\partial a_M} \\ \frac{\partial q_x}{\partial b_M} \\ \frac{\partial q_x}{\partial p_x} \end{bmatrix} = \begin{bmatrix} 1 \\ p_x \\ b_M \end{bmatrix}, \qquad V_x = \begin{bmatrix} u^2(a_M) & u(a_M, b_M) & 0 \\ u(a_M, b_M) & u^2(b_M) & 0 \\ 0 & 0 & u^2(p_x) \end{bmatrix},$$

which is equivalent to

$$u^2(q_x) = u^2(a_M) + 2p_x u(a_M, b_M) + p_x^2 u^2(b_M) + b_M^2 u^2(p_x),$$

as is found by applying GUM equation (13) to equation (E5.15)

Calculation of reference value

The reference value Q_x calculated using (E5.16), namely,

$$Q_x = \frac{q_x}{[1 + \alpha_x(\Delta T_x + \delta T)]},$$

has an associated uncertainty $u(Q_x)$ given in matrix form by

$$u^2(Q_x) = C_{Q_x}^\top V_{Q_x} C_{Q_x},$$

where

$$C_{Q_x} = \begin{bmatrix} C_{q_x} \\ C_{\alpha_x} \\ C_{\Delta T_x} \\ C_{\delta T} \end{bmatrix} = \begin{bmatrix} \frac{Q_x}{q_x}, \\ -\frac{Q_x^2}{q_x}(\Delta T_x + \delta T), \\ -\frac{Q_x^2}{q_x}\alpha, \\ -\frac{Q_x^2}{q_x}\alpha, \end{bmatrix}; \qquad V_{Q_x} = \begin{bmatrix} u^2(q_x) & 0 & 0 & 0 \\ 0 & u^2(\alpha_x) & 0 & 0 \\ 0 & 0 & u^2(\Delta T_x) & 0 \\ 0 & 0 & 0 & u^2(\delta T) \end{bmatrix},$$

which is equivalent to

$$u^2(Q_x) = C_{q_x}^2 u^2(q_x) + C_{\alpha_x}^2 u^2(\alpha) + C_{\Delta T_x}^2 u^2(\Delta T_x) + C_{\delta T}^2 u^2(\delta T),$$

as is found by applying GUM equation (13) to equation (E5.16)

E5.8.4   Numerical illustration

To illustrate, consider a calibration that is performed using the two reference leaks for which
calibration data is available as depicted in figure E5.3 and tables E5.6 and E5.7.
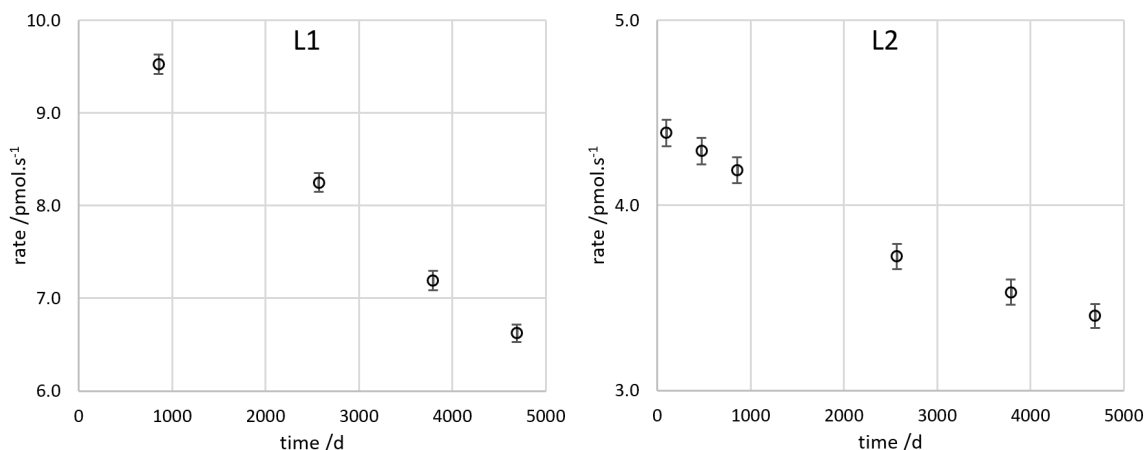
Figure E5.3: Calibration data for the two reference leaks. Data points represent the reference
value with error bars corresponding to ±1 standard uncertainty

Table E5.6: Calibrated reference values for reference leak L_1

t/d	$Q_j/\mathrm{pmol\,s^{-1}}$	$u(Q_j)/\mathrm{pmol\,s^{-1}}$
857	9.525	0.105
2571	8.250	0.103
3792	7.192	0.103
4689	6.623	0.094

Table E5.7: Calibrated reference values for reference leak L_2

t/d	$Q_j/\mathrm{pmol\,s^{-1}}$	$u(Q_j)/\mathrm{pmol\,s^{-1}}$
100	4.391	0.072
474	4.293	0.071
856	4.190	0.070
2568	3.724	0.069
3791	3.531	0.068
4692	3.402	0.066

Reference values

Since there is no covariance in the data, a straight line can be fitted for each set of data following the approach detailed in clause 6 of ISO/TS 28037 [16]. The results of these operations are

$$a_1 = 10.185 \, \mathrm{pmol\,s^{-1}},$$
$$b_1 = -7.678 \times 10^{-4} \, \mathrm{pmol\,s^{-1}},$$
$$u(a_1) = 0.119 \, \mathrm{pmol\,s^{-1}},$$
$$u(b_1) = 3.506 \times 10^{-5} \, \mathrm{pmol\,s^{-1}\,d^{-1}},$$
$$u(a_1, b_1) = -3.785 \times 10^{-6} \, \mathrm{pmol^2 s^{-2} d^{-1}}$$

and

$$a_2 = 4.380 \, \mathrm{pmol\,s^{-1}},$$
$$b_2 = -2.200 \times 10^{-4} \, \mathrm{pmol\,s^{-1}},$$
$$u(a_2) = 0.045 \, \mathrm{pmol\,s^{-1}},$$
$$u(b_2) = 1.620 \times 10^{-5} \, \mathrm{pmol\,s^{-1}\,d^{-1}},$$
$$u(a_2, b_2) = -5.713 \times 10^{-7} \, \mathrm{pmol^2 s^{-2} d^{-1}}.$$

Note that values for the reference leaks do not need to be enumerated in this example; however, for completeness the values found by applying forward evaluation using the above parameters and the process described in clause 11 of ISO/TS 28037 for equations (E5.12) above are found to be:

$$Q_1 = 6.346 \, \mathrm{pmol\,s^{-1}}, \qquad u(Q_1) = 0.084 \, \mathrm{pmol\,s^{-1}},$$
$$Q_2 = 3.280 \, \mathrm{pmol\,s^{-1}}, \qquad u(Q_2) = 0.054 \, \mathrm{pmol\,s^{-1}}$$

when $t = 5000 \, \mathrm{d}$ and $u(t) = 1 \, \mathrm{d}$.

Measured values

Suppose that the leak is to be used at $t = 5000 \, \mathrm{d}$, $u(t) = 1 \, \mathrm{d}$ and the calibration conditions are as in table E5.8 for which it is calculated [equation (E5.13)] that

$$q_1 = 6.735 \, \mathrm{pmol\,s^{-1}}, \qquad q_2 = 3.473 \, \mathrm{pmol\,s^{-1}}.$$

Table E5.8: Conditions during use of reference leaks L_1 and L_2

Quantity	Value	Standard uncertainty
ΔT_1	2.11 K	0.52 K
ΔT_2	2.03 K	0.53 K
α	0.029 K^{-1}	0.005 K^{-1}

The covariance matrix for the data is then

$$\boldsymbol{V_y} = \boldsymbol{V}_{q_1, q_2} = \begin{bmatrix} u^2(q_1) & u(q_1, q_2) \\ u(q_1, q_2) & u^2(q_2) \end{bmatrix} = \begin{bmatrix} 0.030 & 0.0066 \\ 0.0066 & 0.009 \end{bmatrix} (\mathrm{pmol/s})^2.$$

Hence, the correlation matrix is

$$\boldsymbol{R}_{q_1, q_2} = \begin{bmatrix} 1 & 0.398 \\ 0.398 & 1 \end{bmatrix}.$$

MSLD Calibration

Suppose that the MSLD indications (in display units, du) corresponding to q_1 and q_2 are observed:

$$p_1 = 149.2\,\text{du}, \qquad p_2 = 52.1\,\text{du}.$$

The parameter values in the measurement equation (E5.14) are then found to be

$$a_M = 1.722\,\text{pmol s}^{-1},$$
$$b_M = 0.034\,\text{pmol/s/du},$$
$$u(a_M) = 0.139\,\text{pmol s}^{-1},$$
$$u(b_M) = 0.0017\,\text{pmol/s/du},$$
$$u(a_M, b_M) = -1.701 \times 10^{-4}\,\text{pmol}^2/\text{s}^2/\text{du}.$$

Calibration of unknown leak

Suppose now that when an unknown reference leak L_x is connected to the MSLD the response
is 120 du. Forward evaluation using the MSLD calibration function then estimates a leak rate of
$q_x = 5.754\,\text{pmol s}^{-1}$ and an associated standard uncertainty of $u(q_x) = 0.135\,\text{pmol s}^{-1}$.

Calculation of reference value

Finally, a value for the unknown leak rate can be established that is referenced to a temperature
T_0. Suppose that the calibration conditions are those in table E5.9.

Table E5.9: Conditions during measurement of Q_x

Quantity	Value	Standard uncertainty
ΔT_x	2.01 K	0.55 K
α_x	0.030 K^{-1}	0.005 K^{-1}

The reference value Q_x at time t is therefore calculated to be

$$Q_x = 5.427\,\text{pmol s}^{-1}; \qquad u(Q_x) = 0.158\,\text{pmol s}^{-1}.$$

E5.9   Reporting the result

The estimate of the measurand Q and the associated standard uncertainty are directly reported
in the conventional manner according to the GUM [51] including the less common additional
reporting of covariance where required.

In practice two situations might arise. In the one case the evaluation of a leak rate may be
a multi-step process, in which case the intermediate measurands a and b will be reported and
taken as explicit inputs to the next stage of the evaluation process (perhaps by a different party).
In the other case a and b may not be explicitly evaluated at all; instead they may be directly
incorporated into the evaluation process which reports a result for the measurand Q.

In the scenarios presented here, for the sake of completeness, the first case is taken to apply.

E5.10   Interpretation of results

Case 2 demonstrates that when correlation is present the correlation matrix gives greater insight as an indicator than the covariance matrix. Correlation in this case is considerably greater for L_2 than for L_1 as seen by comparing the off-diagonal terms in the correlation matrix (E5.10) with those in the matrix (E5.6). Such a statement could not easily be made by examining covariance matrices.

The overall significance of correlation is dependent on the specific data and it cannot easily be evaluated without a measurement model and a proper analysis.

The data for L_2 show signs of curvature, visually evident in figure E5.3, even though a chi-squared test for linearity is passed (a straight line just about passes through all error bars). A higher order function such as a quadratic [39] would likely result in lower and more random residuals. A similar approach to that described here could be applied, but this is beyond the scope of the present work.

Example E6

Factoring effects such as calibration corrections and drift into uncertainty evaluations

J. Greenwood, M.G. Cox, N. Fischer

E6.1 Summary

This activity comprises two examples that demonstrate potential danger in the practice of factoring effects such as calibration corrections and drift into uncertainty evaluations as rectangular distributions, and presents ways of handling these effects that is consistent with the GUM suite of documents. These examples illustrate that, in spite of the availability of appropriate guidance, significant known bias as a result of effects such as calibration corrections, drift or consumption, hysteresis and non-linearity is often not properly handled. This abuse could bias conformity decisions and thereby place either the consumer or supplier at an unfair disadvantage.

E6.2 Introduction of the application

Poor practice in the evaluation of measurement uncertainty can influence decisions on which it depends. Still, known corrections are often not applied to observed values when computing a measurement result and instead the uncertainty is enlarged in an attempt to compensate. This poor practice inflates coverage intervals and could bias conformity decisions and therefore place either the consumer or supplier at a disadvantage. The consequences of such poor practice are demonstrated in several examples.

E6.3 Specification of the measurand(s)

Denote an output quantity by Z and an input quantity by Y representing an indicated value. The measurand generically is

$$Z_{\text{uncor}} = Y \tag{E6.1}$$

for an uncorrected model as in section E6.4. For a corrected model it is

$$Z_{\text{cor}} = Y + X, \tag{E6.2}$$

where X is the quantity regarded as a correction.

77

E6.4   Measurement model

A 'best estimate' of Y is the arithmetic mean

$$y = \frac{1}{n} \sum_{i=1}^{n} y_i, \qquad i = 1, \ldots, n, \tag{E6.3}$$

where the y_i are unbiased observations made independently under repeatability conditions of measurement.

A 'best estimate' of Z_{uncor} is then

$$z_{\text{uncor}} = y. \tag{E6.4}$$

A value x of a correction quantity X is often incorporated to account for a systematic effect:

$$z_{\text{cor}} = y + x. \tag{E6.5}$$

x is the correction for a known bias or systematic error in the measuring system. A further term can be included in model (E6.4) or (E6.5) relating to the resolution of the measuring instrument. Its inclusion is straightforward and is treated by Lira and Wöger [116]. We do not consider that term here.

The 'known systematic error' can arise from a variety of sources including calibration, effects due to temperature deviation, drift, hysteresis, consumption of material and 'wear', and effects due to method or operator bias.

E6.5   Uncertainty propagation

Knowledge concerning Y is in terms of a set of repeated observations made under repeatability conditions (section E6.4). The arithmetic mean y of the observations is taken in (E6.3) as an estimate of Y. The associated standard uncertainty $u(y)$ is given by [51, clause 4.2]

$$u^2(y) = \frac{1}{n(n-1)} \sum_{i=1}^{n} (y_i - y)^2.$$

Knowledge concerning the systematic error is that a value x and an associated standard uncertainty $u(x)$ are available. The use of this knowledge in practice depends on whether a correction is or is not to be made to y.

In expressions (E6.4) and (E6.5), y can be considered as a realized value of a random variable with that variable typically modelled by the normal distribution $N(y, u^2(y))$, which is strictly valid only for large n. In cases where n is small, the t-distribution should be used [51, annex G.3].

E6.5.1   Case 1: Good practice

Good practice, as assumed by JCGM guidance, dictates that a correction is made for a known systematic error. This is the situation represented in expression (E6.5), in which x is a realized value of a random variable. That variable is often modelled by a rectangular distribution with mean x and standard deviation $u(x)$.

The standard uncertainty $u(z_{\text{cor}})$ associated with the corrected value z_{cor} in expression (E6.5) is given by

$$u^2(z_{\text{cor}}) = u^2(y) + u^2(x). \tag{E6.6}$$

There is a correspondence between the right-hand sides of expressions (E6.5) and (E6.6) in that
the terms in the former are the means of the corresponding random variables and the terms in
the latter are the according variances of these random variables. That is, expression (E6.6) can
be regarded as giving the squared standard uncertainty associated with the corrected measured
value z_{cor} in (E6.5). Said another way, expression (E6.5) is a realisation of measurement model
(E6.2) and expression (E6.6) gives the squared standard uncertainties corresponding to the terms
in the model.

For purposes of conformance assessment, we can regard the corrected value z_{cor} as modelled
by a probability distribution with mean z_{cor} and standard deviation $u(z_{cor})$. When n is large and x
has a normal distribution, the distribution relating to z_{cor} can be taken as normal: $N(z_{cor}, u^2(z_{cor}))$.
If that is not the case, a Monte Carlo method [52] can be used to obtain its probability distribution
given probability distributions for Y (for example, Student's t) and X (for example, rectangular).

E6.5.2   Case 2: Poor practice

Common practice [97,116,117,142,165] often involves making no correction and increasing the
uncertainty associated with the value of the measurand. On this basis the reported uncorrected
value would be

$$z_{uncor} = y \tag{E6.7}$$

A variety of approaches have been adopted for evaluating the associated uncertainty [117]. For
the purposes of demonstration let us suppose that the associated 'standard uncertainty' $u(z_{uncor})$
is given by one of the commonly used approaches, for which it is assumed that

$$u^2(z_{uncor}) = u^2(y) + u^2(x) + x^2/3. \tag{E6.8}$$

The term $x^2/3$ is included in the uncertainty evaluation as a consequence of modelling the sys-
tematic effect as a rectangular distribution with mean of zero and half-width equal to the mag-
nitude of x. Other assumptions would give rise to generally somewhat different contributions.

'Standard uncertainty' is given in quotation marks since, as stated in [102], it is *not* a standard
uncertainty as defined in the GUM [51]. There is no one-to-one correspondence between the
terms in expressions (E6.7) and (E6.8) as in the expressions for the corrected value (E6.5) and
the associated variance (squared standard uncertainty) (E6.6). It does not possess the properties
of internal consistency, transferability and universality: see the GUM [51, clause 0.4] and the
strong comments in [102].

The use of expression (E6.8) to obtain a 'standard uncertainty' for an uncorrected value con-
stitutes *poor practice*.

It is not proper to attach a probability distribution to z_{uncor}. However, poor practice might
assume the normal distribution $N(z_{uncor}, u^2(z_{uncor}))$.

Not correcting for a systematic effect can be serious. Molinar et al. [132] report in the context
of methods for evaluation of uncertainty increase due to chemical impurities:

> 'If no correction is applied, an additional type B uncertainty component $u \approx 0.2\,mK$
> if a rectangular probability distribution is assumed. The additional component is
> nearly one order of magnitude larger than the other uncertainty components of the
> fixed-point realization.'

Further, Westgard et al. [179] state in the context of laboratory medicine:

> 'To characterize analytical quality of a laboratory test, common practice is to estimate
> Total Analytical Error (TAE) which includes both imprecision and trueness (bias).

The metrologic approach is to determine measurement uncertainty, which assumes
bias can be eliminated, corrected, or ignored. Resolving the differences in these
concepts and approaches is currently a global issue […]

Elimination or correction of […] biases is not always possible, even with calibration
based on comparative patient results; therefore, bias must still be measured and
monitored and should not be ignored or assumed to be accommodated by long-term
estimates of measurement uncertainty.'

E6.6   Reporting the result

E6.6.1   Impact on tests against specification or tolerance limits

General

Conformance probability [55, definition 3.3.7] is the probability p that the measurand Z lies in
a tolerance interval $[a, b]$, with $a < b$, that is

$$p = \Pr(a \leq Z \leq b) = \int_a^b g(\eta)\mathrm{d}\eta,$$

where g is the probability density function for Z [55, clause 7.4]. An interval such as $[a, b]$ is
called a coverage interval for Z and p is the associated coverage probability. Guidance on con-
structing a coverage interval with a desired coverage probability given the probability distribution
for Z is contained in JCGM 101:2008 [52].

Using the recommended practice in section E6.5.1, when g is normal, resulting from Y and
X being normal, the integral can straightforwardly be computed. Otherwise, a Monte Carlo
calculation [52] can be used to establish an approximation to g since a false assumption of
normality might lead to an invalid indication of conformance probability.

Conformity decisions based on poor and good practice

The consequences of the poor practice regarding corrections in subsection E6.5.2 can be demon-
strated by example. Suppose the primary length of a product is tested using an appropriate
measuring instrument and there is an upper tolerance limit $T_\mathrm{U} = 100\,\mathrm{cm}$ on the length. Suppose
the known correction and its associated standard uncertainty are

$$x = 0.20\,\mathrm{cm}, \qquad u(x) = 0.05\,\mathrm{cm}.$$

A measured length value and its associated standard uncertainty are

$$y = 99.75\,\mathrm{cm}, \qquad u(y) = 0.15\,\mathrm{cm}.$$

Both x and y have Gaussian distributions.

Using expressions (E6.5) and (E6.6), the corrected value and associated standard uncertainty
are

$$z_\mathrm{cor} = y + x = 99.95\,\mathrm{cm}, \qquad u(z_\mathrm{cor}) = [u^2(y) + u^2(x)]^{1/2} = 0.16\,\mathrm{cm}.$$

The conformance probability is the area to the left of $T_\mathrm{U} = 100\,\mathrm{cm}$ under the normal curve having
mean z_cor and standard deviation $u(z_\mathrm{cor})$. That probability is 0.62.

On the other hand, working with uncorrected values, expressions (E6.7) and (E6.8) give

$$z_{\text{uncor}} = y = 99.75\,\text{cm}, \qquad u(z_{\text{uncor}}) = [u^2(y) + u^2(x) + x^2/3]^{1/2} = 0.20\,\text{cm}.$$

The conformance probability is now the area to the left of $T_U = 100\,\text{cm}$ under the normal curve
having these values as mean and standard deviation. That probability is 0.90, implying that a
significantly greater proportion of non-conforming items might be accepted.

Figure E6.1 depicts these two situations.
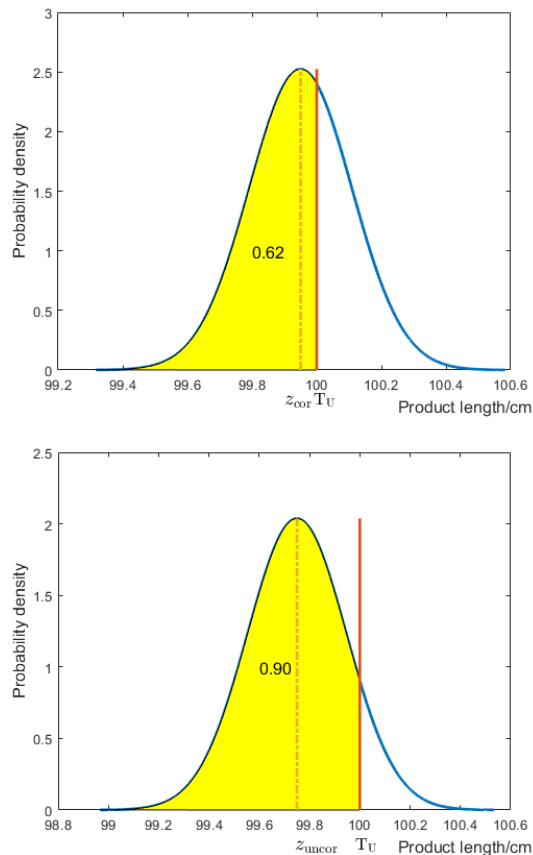
Figure E6.1: Using a corrected value, the conformance probability (0.62) is the shaded area in
the left figure, whereas using the uncorrected value, the conformance probability (0.90) is the
shaded area in the right figure; the latter (poor-practice) approach allows a greater proportion
of non-conforming items to be accepted

Figure E6.2 shows the conformance probability for a range of values y.
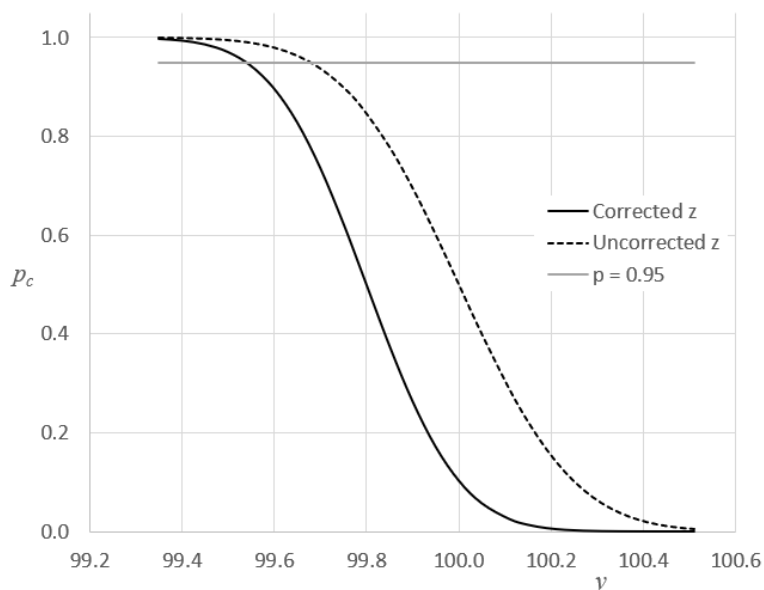
Figure E6.2: Conformance probability for a range of measured values y with corrected and uncorrected bias in the example described in section E6.6.1

E6.7 Treatment of drift

In many practical situations it is difficult to model drift with any degree of confidence – in most cases it is probably one of the least reliable influence quantities.

In general terms calibration drift usually corresponds to a change in a calibration value over the course of time. This variation might occur in a predictable or not so predictable fashion depending upon the underlying source of the variation.

In situations where a calibration function is established from data consisting of reference values and corresponding observed values (as elaborated in [16] and [38]). The drift could be modelled in terms of a time dependent relationship for the fitted coefficients, perhaps resulting in a linear function

$$Z(t) = a(t) + b(t)Y, \tag{E6.9}$$

in which a and b are time-dependent parameters whose value is influenced by historical calibration data as well as the most recent data.

This procedure is generally not straightforward and is unlikely to be widely adopted. In these circumstances some other approach is usually adopted, such as analysis of any trend in estimated values of the parameters.

However, in many situations such a calibration function that directly relates an observation to an estimate of the measurand is not established. Instead the available calibration data is used to estimate an additive (or multiplicative) correction, as for example

$$Z = Y + C. \tag{E6.10}$$

In use, the measurand is estimated by

$$z(t) = y + c(t),$$

where the estimate for the calibration correction $c(t)$ at time of use t is based upon calibration data

$$c_i = c(t_i) = z_i - y_i,$$

and where $z_i = z(t_i)$ and $y_i = y(t_i)$ are calibration values corresponding to a reference value and observation obtained at time t_i

These approaches usually seem intuitively more reasonable to laboratory practitioners, but in practice problems remain due to the generally small amount of information available (say 3 or 4 successive annual observations) and the question of how much weight to give to historical data.

Where there is enough data to perform a fit to a trend in the c_i, such a fit and use of an extrapolated calibration value $c(t)$ based upon a functional fit to historical data seems appropriate. This approach is demonstrated in Models 1 and 4 below.

In other situations there is insufficient data to draw any strong conclusions about how c_i varies over time. Consequently, the most recent calibration correction c_n is often taken to be the best estimate of C. In other words, an estimate

$$c(t) = c_n + d(t), \tag{E6.11}$$

is made, in which $d(t)$ is a poorly understood bias effect with an assumed mean value of zero. This is the approach demonstrated in Models 2, 3, and 5 below. (Arguably the mean calibration value could be chosen rather than c_n; however, there is usually a preference to give more weight to the most recent value.)

The question that remains is how to evaluate the uncertainty associated with equation (E6.11). If the expectation of the drift is genuinely zero (rather than this simply being a convenient assumption) then assigning a distribution centred on zero is quite reasonable. Some guidance [96] suggests that in this case the data should be considered as a Type A contribution (see Model 5 below). In practice, however, the available data is often used to identify limit values for a rectangular distribution with expectation value zero and limits $\pm a_d$.

Common estimates for a_d are

$$a_d = \max(|d_i|),$$

where d_i is the difference between successive values c_i and c_{i-1} (see Model 2 below). Occasionally the estimate is more sophisticated, for example (Model 3 below), the larger value when comparing the average $|\bar{d}|$ of the absolute differences $|d_i|$, and the most recent value d_n, that is

$$a_d = \max(|\bar{d}|, \ |d_n|).$$

E6.7.1 Example of treatment of drift

The issue is demonstrated in the following example in which the conformance probability p_c is calculated using GUM-consistent measurement models (E6.10) with various approaches taken for evaluating drift.

In this example suppose that a test is defined with an upper tolerance $T_U = 10$ for the measurand Z.

Suppose that the available calibration data consists of four equally spaced results c_i corresponding to times t_i for $i = 1, \ldots, n = 4$ as shown in Table E6.1. Let $u(c_i) = 0.15$ and the evaluation be conducted for some later time $t > t_n$, say $t = 42$.

Table E6.1: Annual calibration correction data c_i and difference $d_i = c_i - c_{i-1}$

i	t_i	c_i	d_i
1	0	0.3	
2	12	0.3	0.00
3	24	−0.15	−0.45
4	36	0.1	0.25

The conformance probability p_c is calculated for a range of measured values y using different drift models as described below. Figure E6.3 summarises the results. These drift models are some that are used in practice. Note that many other variations on these models are possible and are also encountered.

Model M1:
A straight-line fit to the data is performed to establish $c(t)$ at the time of use; hence $c(t) = a_0 + a_1 t$, where a_0 and a_1 are the coefficients of the fitted function and $u(c_t)$ is the uncertainty associated with $c(t)$, all of which can be established using ISO/TS 28037 [16]; hence

$$z(t) = y + c(t),$$
$$u^2(z) = u^2(y) + u^2(c_t).$$

Model M2:
The most recent calibration result c_n is used. There is a genuine belief (perhaps due to some metrological knowledge or experience) that the expectation of drift is zero despite the recent albeit sparse evidence to the contrary. The drift is therefore assumed to have mean value of zero. Its associated standard uncertainty is evaluated as the standard deviation of a rectangular distribution with semi-width corresponding to the maximum absolute difference between successive calibration results; hence

$$z(t) = y + c_n + 0,$$
$$u^2(z) = u^2(y) + u^2(c_n) + a_d{}^2/3,$$
$$a_d = \max(|d_i|).$$

Model M3:
The most recent calibration result c_n is used. Again, there is a genuine belief that the expectation of drift is zero despite the recent evidence to the contrary. The drift is therefore assumed to have mean value of zero. In order to give more weight to the most recent data, its associated standard uncertainty is evaluated as the standard deviation of a rectangular distribution with semi-width corresponding to the larger of a) the most recent absolute difference, and b) the mean of all absolute differences between successive calibration results.

$$z(t) = y + c_n + 0,$$
$$u^2(z) = u^2 + u^2(c_n) + a_d{}^2/3,$$
$$a_d = \max(|d_n|, \bar{d}),$$
$$\bar{d} = \frac{1}{n} \sum_{i=2}^{n} |d_i|.$$

Model M4:
A straight-line fit $c(t) = a_0 + a_1 t$ to the data is performed to establish $c(t)$ at the time of use.

The coefficients a_0 and a_1 of the fitted function are established using ordinary least squares (for example, using the Excel SLOPE and INTERCEPT functions). The standard uncertainty associated with $c(t)$ is evaluated as the standard deviation of a rectangular distribution with semi-width corresponding to the maximum fitting residual; hence

$$z(t) = y + c(t),$$
$$u^2(z) = u^2(y) + u^2(c_t),$$
$$u^2(c_t) = u^2(c_n) + r_{\max}^2/3,$$
$$r_{\max} = \max(|c_i - (a_0 + a_1 t)|).$$

Model M5:

There is a genuine belief that the expectation of drift is zero despite the recent evidence to the contrary. The drift is therefore assumed to have mean value of zero. Its standard uncertainty $u(d)$ is taken as the standard deviation of the set of available data for d_i. It is assumed to be characteristic of quantity with a t-distribution centred at zero having standard deviation equal to $u(d)$. The best estimate of the appropriate correction at the time of use is assumed to be the most recent value, c_n. (Arguably the mean value should be chosen; however, there is usually a preference to weight towards the most recent value.)

$$z(t) = y + c_n + 0,$$
$$u^2(z) = u^2(y) + u^2(c_n) + u^2(d),$$
$$\nu_{\text{eff}} = (n-1)\frac{u^4(z)}{u^4(d)}.$$

The conformance probability is evaluated on the basis that z has a t-distribution with ν_{eff} effective degrees of freedom (evaluated using the Welch-Satterthwaite formula [51]).
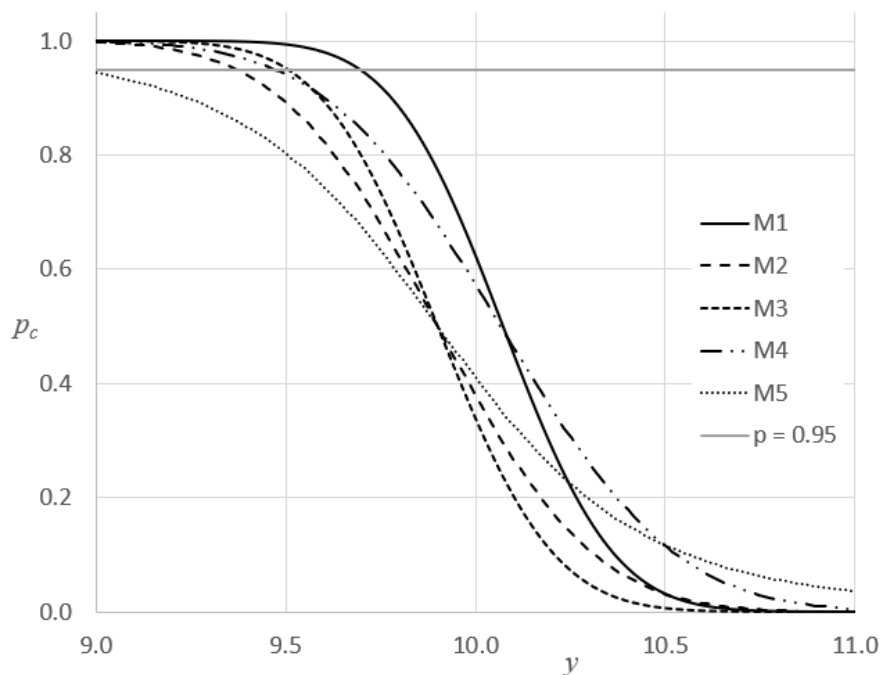
Figure E6.3: conformance probability for a range of observed values y, evaluated at $t = 42$ using the various models, M1 to M5 and the data in Table E6.1

The conformance probability p_c has been calculated for a range of measured values y with the different drift models described above. The conformance probabilities are depicted in Figure E6.3 in which it is evident that the choice of drift model can have a significant impact upon the measurement uncertainty and upon any subsequent decisions.

All of these models, when based upon little data, make more or less arbitrary choices for $u(c)$ and $u(d)$. For this reason any significant assumptions should be clearly stated with the results and the choice of model should be justified, either by additional measurements, or with supporting information based upon metrological experience and expertise. For example, where a linear model seems appropriate M1 might seem justified, whereas M2 is more conservative and may be preferable if the risk of false acceptance is a key concern and no additional information about the drift is available.

E6.8   Interpretation of results

As an alternative to correction, a number of methods have been proposed or adopted that increase the expanded uncertainty to take account of bias. In [133], 'all sensible combinations' of correcting or enlarging uncertainty for bias, whether considered significant or not, were modelled by a Latin hypercube simulation of 1.25×10^5 'iterations' for a range of bias values. The fraction of results for which the value and the associated expanded uncertainty contained the true value of a simulated test measurand was used to assess the various methods. The strategy of estimating the bias and always correcting is consistently the best throughout the range of biases.

Laboratories are routinely faced with the question of whether they need to correct for biases (such as temperature effects, calibration corrections or drift) with the associated investment of time and effort in maintaining such a process. The attraction of a simple approach whereby such biases are factored into an uncertainty budget makes this a commonly adopted approach in which there is usually no appreciation of the potential problems that are created for others further along the measurement chain, as demonstrated in section E6.6.1 and in section E6.7.1.

Unfortunately, however (as explained in section E6.5.2), there is no way to state the uncertainty associated with an uncorrected value that is consistent with the GUM [51].

On the basis of the explanation in section E6.5.2 and supported by these simulations it is strongly recommended that whenever possible a corrected value and the associated uncertainty is reported as in section E6.5.1.

Otherwise, when an uncorrected value and an uncertainty are reported, it should be stated that the result is inconsistent with the GUM in that a measurement model has not been used but the result follows the advice of a publication that is cited. Reference [117] usefully categorises several approaches.

The consequent impact on conformity decisions must also be considered. In some cases the effect on conformance probabilities would be considerable, as seen in section E6.6.1 and in section E6.7.1.

Even when a model *is* consistent with the GUM, the choice of how to treat drift can have significant impact on conformance intervals and upon decisions, as discussed in the previous section. It is strongly recommended that the practitioner gives sufficient details in the report on how drift has been handled.

Example E7

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

E7.1   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.

E7.2   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 [110] 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 [55] for conformity assessment of a single item. For this reason, the notation used here is consistent with the notation in [55] and [112], the latter being a relevant paper in which the reader can find more details on this case study.

E7.3   Specification of the measurands

This case study concerns test results for NyQuil tablets [2], a cold medication containing four active components:

- acetaminophen (APAP), as a pain reliever and fever reducer;

- dextromethorphan hydrobromide (DEX), as a cough suppressant;

87

- 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 [112], 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).

E7.4   Measurement uncertainty and correlations

A full uncertainty budget for the test results of the components' content is available in [112]. 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} [51, 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) [112]. Correlation coefficients are reported in table E7.1.

Table E7.1: Correlation coefficients between components' content values

Component	Index	APAP	DEX	DOX	PE
	i/j	1	2	3	4
APAP	1	1	0.107	0.125	0.177
DEX	2		1	0.311	0.404
DOX	3			1	0.539
PE	4				1

E7.5   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.

E7.6   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 [55, 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.

E7.6.1   Bayesian framework

In the framework of the IUPAC project [110], 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 [55, sec. 9.3.2 and 9.5.2]. The underlying Bayesian approach requires defining a multivariate prior probability density function (PDF) $g_0(\boldsymbol{c})$ for "true" values of the components' content, where $\boldsymbol{c} = [c_1, \ldots, c_4]$, and a multivariate likelihood function $h(\boldsymbol{c}_m | \boldsymbol{c})$ for the corresponding test results, where $\boldsymbol{c}_m = [c_{1m}, \ldots, c_{4m}]$.

As discussed in [112], a multivariate normal distribution is used for modelling both the prior knowledge and the likelihood function. The former multivariate normal PDF, $g_0(\boldsymbol{c})$, has vector mean $\boldsymbol{m} = [m_1, \ldots, m_4]$, where m_i is the ith experimental sample mean, calculated from the available production data (see table E7.2), and covariance matrix $\boldsymbol{S}_c$ made by terms $S_{c_{ij}} = r_{ij} s_i s_j$, where r_{ij} are the correlation coefficients in table E7.1 and s_i is the ith experimental standard deviation (see table E7.2). For each fixed vector value $\boldsymbol{c}$, the multivariate normal PDF modelling the likelihood function $h(\boldsymbol{c}_m | \boldsymbol{c})$ has vector mean $\boldsymbol{c}$ and covariance matrix $\boldsymbol{S}_{cm}$ made by terms $S_{cm_{ij}} = r_{ij} u_i u_j$, where $u_i = 0.028\, c_{im}$, % of labelled amount, is the ith associated standard uncertainty[1]. 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 E7.2: Experimental mean and standard deviation of the components' content values of 105 lots of the medication

Component	Index i	Mean m_i, % of labelled amount	Standard deviation s_i, % of labelled amount
APAP	1	99.18	1.37
DEX	2	97.70	1.02
DOX	3	99.33	1.05
PE	4	98.94	1.22

E7.6.2   Total specific risk

For a vector of test results $\boldsymbol{c}_m$ of a specific multicomponent item, when all the c_{im} are measured within their own acceptance interval and hence the material is accepted as conforming, the total specific consumer's risk R_{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 $[T_{L1}, T_{U1}] \times [T_{L2}, T_{U2}] \times [T_{L3}, T_{U3}] \times [T_{L4}, T_{U4}]$. The integral can be obtained by calculation of the corresponding cumulative distribution function at the desired limits.

In the current study, since both prior $g_0(\boldsymbol{c})$ and likelihood function $h(\boldsymbol{c}_m | \boldsymbol{c})$ 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 $\boldsymbol{S}_{post}$ and vector of posterior

---

[1]Standard 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).

means $\boldsymbol{c}_{\text{post}}$ respectively equal to [92, eq. 3.13]:

$$S_{\text{post}} = \left(S_c^{-1} + n_{\text{rep}}S_{cm}^{-1}\right)^{-1} \tag{E7.1}$$

$$\boldsymbol{c}_{\text{post}} = S_{\text{post}}\left(S_c^{-1}\boldsymbol{m} + n_{\text{rep}}S_{cm}^{-1}\bar{\boldsymbol{c}}_{\text{m}}\right), \tag{E7.2}$$

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

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.

E7.6.3 Total global risk

The total global consumer's risk R_{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 $g_0(\boldsymbol{c})\,h(\boldsymbol{c}_{\text{m}}|\boldsymbol{c})$. It can be calculated by a Monte Carlo (MC) simulation in which, for each vector $\boldsymbol{c}$ randomly drawn from $g_0(\boldsymbol{c})$, a correspoding vector $\boldsymbol{c}_{\text{m}}$ is drawn from $h(\boldsymbol{c}_{\text{m}}|\boldsymbol{c})$. Hence, the total risk is approximated by the frequency of cases in which, within randomly generated vectors $[\boldsymbol{c}_{\text{m}}, \boldsymbol{c}]$, 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 [112] ($R_{\text{tot}} = 0.0019$), 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".

E7.7 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 [112], 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 E8

Conformity assessment of mass concentration of total suspended particulate matter in air

F. Pennecchi, F. Rolle, A. Allard, S.L.R Ellison

E8.1 Summary

The main goal of the present study is to show how to calculate risks of false decisions in the conformity assessment of test results, according to the framework of [55], in the case in which a normal distribution is not a valid assumption for modelling prior information on the measurand. As a case study, test results of mass concentration of Total Suspended Particulate Matter (TSPM) in ambient air are considered.

E8.2 Introduction of the application

A total of 496 test results of mass concentration of TSPM in ambient air, collected in 2009 in the proximity of three stone quarries located in Israel, were obtained according to the Environmental Protection Agency (EPA) method IO-2.1 [5]. Such results were compared with the national (Israeli) regulation limit for air quality to study the occurrence of Out-Of-Specification (OOS) test results, as detailed in [109] and in [111].

In the present example, the focus is on the calculation of global and specific risks of false decision in the conformity assessment of such kind of test results. The risk of underestimating the pollutant concentration is the consumer's/inhabitants' risk and that of overestimating is the producer's risk. Calculation of such risks is as important for the Regulator (the Ministry of Environmental Protection) protecting the inhabitants' quality of life in the area surrounding the quarries, as for the Manufacturers' Association acting in the interests of the stone producers in the country.

Risk values of false decisions on conformity of the TSPM concentration are here calculated for each quarry separately. Nonetheless, total risks of false decisions concerning the environmental compartment as a whole can also be calculated, hence characterizing the conformity of the TSPM concentration in the overall region encompassing the three quarries. Such total risks were modelled on the basis of the law of total probability in [141], but are out of the scope of the present example.

E8.3 Specification of the measurand

For characterization of TSPM, the EPA method IO-2.1 [5] indicates the use of a high-volume
sampler for collection of particles with aerodynamic diameters of 100 μm or less. A large volume
V of air, in the range 1600 m³ to 2400 m³, was typically sampled at an average rate and the mass
m of the matter in the sampled air volume, collected on the sampler filter, was measured as the
difference between the results of weighing the filter before and after sampling. The measurand is
the average value of the TSPM mass concentration over the sampling period: $c = m/V$ (mg m^{-3}).
In this study, TSPM from the i-th quarry, $i = 1, 2, 3$, is considered as the i-th pollutant.

E8.4 Test results and associated measurement uncertainty

Three quarries were monitored by the Israeli National Physics Laboratory (INPL) at four points in
the compass approximately 1 km to 3 km from each quarry, four to five times per month. A total
of 496 test results were collected (220 relevant to quarry 1, 176 to quarry 2 and 100 to quarry
3), each test lasting 24 h. In [109] it was demonstrated, by means of the analysis of variance
(ANOVA), that the monthly variation was not a significant factor in the data variability, whereas
TSPM mass concentration seemed significantly influenced by the factor 'quarry'. Thus, it was
concluded that the anthropogenic contributions to TSPM mass concentration due to the activity
of the quarries were dominant and the test results for each quarry had to be studied separately.

Measured TSPM concentration values c_m are reported (in mg m^{-3}) within Q1data.txt, Q2data
and Q3data.txt files for quarry 1, 2 and 3, respectively (available in the repository [88]), and
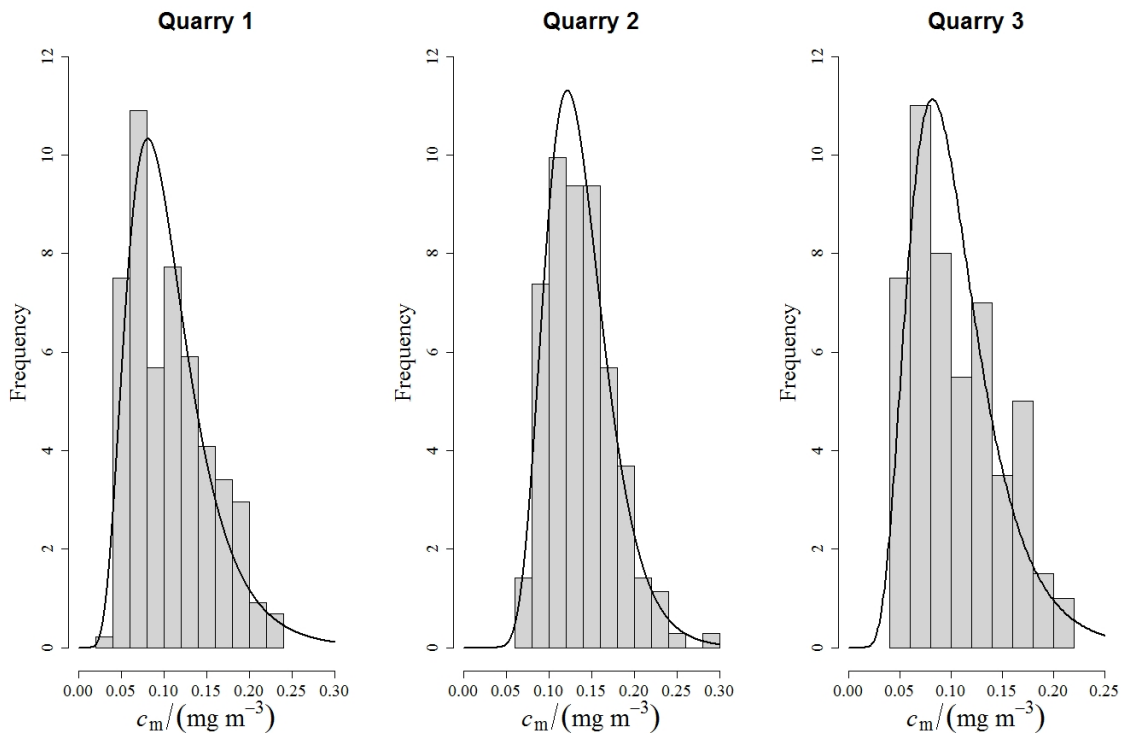depicted in figure E8.1.

Figure E8.1: Histograms of the measured TSPM mass concentration values for each quarry and
corresponding lognormal probability density functions smoothing the data.

A full uncertainty budget for the considered test results is available in [109], where it was

shown that the major contribution to the combined measurement uncertainty associated with the results is that coming from the measurement of the sampled air volume. The combined relative standard uncertainty associated with a typical test result was evaluated as 7.0 %. No correlation among test results from different quarries was observed.

E8.5   Tolerance limits

The Israeli national regulations of ambient air quality prescribe an upper tolerance (regulation) limit $T_U = 0.2\,\mathrm{mg\,m^{-3}}$ for TSPM mass concentration for 24 h sampling. This limit holds for any location, also close to the quarry. Hence, for each quarry and at any sampling point, $T_{Ui} = 0.2\,\mathrm{mg\,m^{-3}}$, for $i = 1, 2, 3$.

E8.6   Decision rule and conformity assessment

Regulations require direct comparison of measured values c_{im} with T_{Ui}. In the present example, acceptance limits A_{Ui} will be made varying in order to show their impact on the risk values of false decisions. When acceptance limits are taken to coincide with the tolerance limits (that is, $A_{Ui} = T_{Ui}$), a "shared risk" rule is considered as the decision rule for conformity assessment [55, sec. 8.2.1].

In the present example, the consumers are the inhabitants living in the area surrounding the quarries, whereas the producers are the owners of the stone quarries.

The global and specific risks of false decisions in conformity assessment are defined in [55, sec. 3.3] for both the consumer and the producer, and have different interpretations. While a specific risk is the risk of an incorrect decision made for a particular measurement result, global risks refer to the probability of an incorrect decision based on a future measurement. Both kinds of risks rely on a Bayesian framework but require the calculation of different probability objects. Indeed, the posterior distribution (obtained through Bayes' theorem) is used for specific risks while the joint distribution is used for global risks.

E8.6.1   Bayesian framework

In the framework of the JCGM document on the role of measurement uncertainty in conformity assessment, the evaluation of risks of false decisions on a characteristic of an item is described in [55, clause 9.3.2 and 9.5.2] for specific and global risks, respectively.

The underlying Bayesian approach requires defining a prior probability density function (PDF) $g_0(c_i)$ for the "true" values of TSPM mass concentration. Based on the Kolmogorov–Smirnov criterion of goodness-of-fit, the widely-used null hypothesis of a normal PDF was tested on the data available for each quarry and had to be rejected [109]. The normal distribution was found instead to be the best-fitting distribution for the experimental results after their logarithmic transformation. Therefore, for each quarry i, a lognormal distribution was chosen for modelling the actual values of TSPM mass concentration c_i:

$$g_0(c_i) = \frac{1}{c_i \sigma_i \sqrt{2\pi}} \exp\left[-\frac{(\ln(c_i) - \mu_i)^2}{2\sigma_i^2} \right], \tag{E8.1}$$

whose distributional parameters are reported (on the logarithmic scale) in table E8.1. They were taken respectively as the mean and the standard deviation of the log-transformed data. The corresponding lognormal prior PDFs are the curves approximating the histograms in figure E8.1.

Table E8.1: Location and scale parameters of the prior PDF for each quarry.

Quarry i	Location parameter μ_i (adimensional)	Scale parameter σ_i (adimensional)
1	−2.325	0.434
2	−2.031	0.279
3	−2.337	0.402

The distribution of the measurement results c_{im} at an actual concentration c_i was modelled by a normal distribution with expectation equal to c_i and standard deviation equal to the standard measurement uncertainty $u_i = 0.07c_{im}$ [109]. The corresponding likelihood for each quarry is hence a normal PDF:

$$h(c_{im}|c_i) = \frac{1}{u_i\sqrt{2\pi}} \exp\left[-\frac{(c_{im}-c_i)^2}{2u_i^2}\right]. \tag{E8.2}$$

When both the prior PDF and the likelihood are normal distributions, the posterior PDF [55, Eq. (1)] is also normal [55, Sec. 7.2.1][1]. In such a case, the evaluation of specific and global risks is straightforward, as detailed in [55]. In the present example, instead, the prior PDF is lognormal, for each quarry, hence requiring some numerical approximation of the consumer's and producer's risks.

E8.6.2 Global risks

For each quarry, and for any considered (upper) acceptance limit A_U, global risks for the consumer and the producer were calculated as a numerical approximation of the (double) integral of the product of the prior PDF (E8.1) and the likelihood (E8.2), according to [55, equations (19) and (20)]. In the considered case, since all the involved PDFs were defined on the positive axis only, the lower integration limits (both T_L and U_L) were taken as zero. Details of the calculation are in the code file A123_Global_risk_TSPM.r (available in the repository [88]), where the R-function `dlnorm` was used for evaluating the density of the considered lognormal distributions, whose logarithms have the mean and the standard deviation, reported in table E8.1 for each quarry, of the data distributions on the log scale (note that the log-transformed data have a normal distribution by the definition of the lognormal distribution). The integration of the joint PDF was performed by means of the R function `integrate`.

The obtained consumer's (red line) and producer's (blue line) global risks are displayed in figure E8.2, for A_U values varying in the interval $[T_U - 0.05, T_U + 0.05]$ mg m^{-3}. Considering, for example, the special case in which $A_U = T_U$, consumer's and producer's global risks were respectively 0.58 % and 0.74 % for quarry 1, 1.04 % and 1.52 % for quarry 2, and 0.46 % and 0.62 % for quarry 3. Focusing on quarry 1, for example, one could be interested in finding the maximum acceptable A_U in order to have a desired small consumer's risk, let us say 0.01 %: it turns out that such an acceptance limit should not exceed 0.17 mg m^{-3}. However, in this case, the global producer's risk would increase from 0.74 % to about 5 %. The other way round, A_U should be at least equal to 0.23 mg m^{-3} in order to assure a producer's risk smaller than 0.01 %, again. In this case, the global consumer's risk would increase from 0.58 % to about 2 %.

---

[1]If the prior information is meagre and the likelihood function is characterised by a normal distribution, then the posterior PDF is approximately normal [55, Sec. 7.2.2].
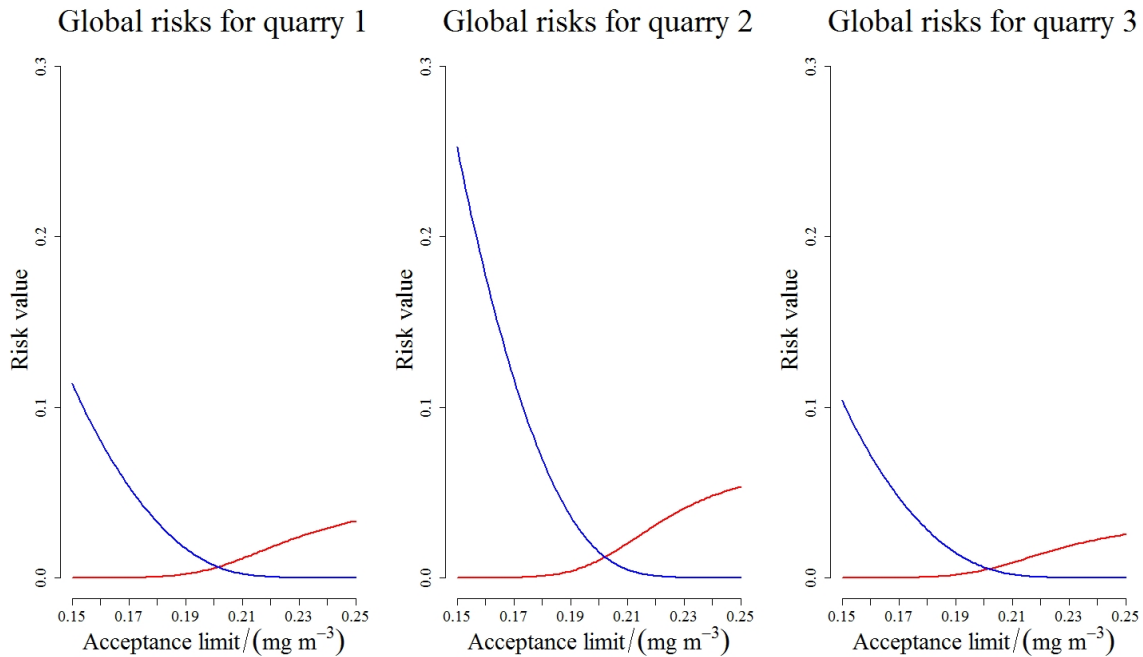
Figure E8.2: Consumer's (red line) and producer's (blue line) global risks versus acceptance limit values.

E8.6.3   Specific risks

For each quarry i, and just for the special case $A_U = T_U$, specific risks for the consumer and the producer were calculated according to the framework of [55, Sec. 9.3.2]. For a specific value $c_{im} < A_U$ (that is, the measured TSPM mass concentration is assessed as conforming to the regulation limit), the consumer's specific risk is the integral of the posterior PDF $h(c_i|c_{im})$ on the region $[T_U, \infty]$, that is on the region of true values which would not be actually conforming. For a specific value $c_{im} > A_U$ (that is, the test result is not conforming to the regulation limit), the producer's specific risk is the integral of the posterior PDF on the region $[0, T_U]$, the region of actually conforming true values. In both cases, the posterior PDF $h(c_i|c_{im})$ [55, equation A.11] was needed, but in the considered case it does not have a closed form because the prior PDF is lognormal.

Details of the calculation are in the code file A123_Specific_risk_TSPM.r (available in the repository [88]), where, for each c_{im} value, the posterior PDF was evaluated as the exponential of the log-posterior PDF, the latter being implemented as the sum of the log-prior PDF, evaluated in c_i, and the corresponding log-likelihood function at c_{im} (i.e., the logarithm of a normal PDF, with mean c_i and standard deviation equal to $0.07 c_{im}$, evaluated at c_{im}). The integral of the posterior PDF was calculated by means of the R function `integrate`.

The obtained consumer's and producer's specific risks are displayed in Figure E8.3 (when $A_U = T_U$) for quarry 1 – blue line, quarry 2 – green line and quarry 3 – red line. They are plotted versus values c_{im} varying in the interval $[0.15, T_U]$ mg m^{-3} and $[T_U, 0.25]$ mg m^{-3} for the consumer and the producer, respectively.

These results have been validated against those from CASoft [44], which relies on simulation using a Metropolis-Hastings algorithm to estimate the posterior distribution used to calculate the specific risks. The results agreed within the small random variation expected for Monte Carlo estimates of small probabilities.
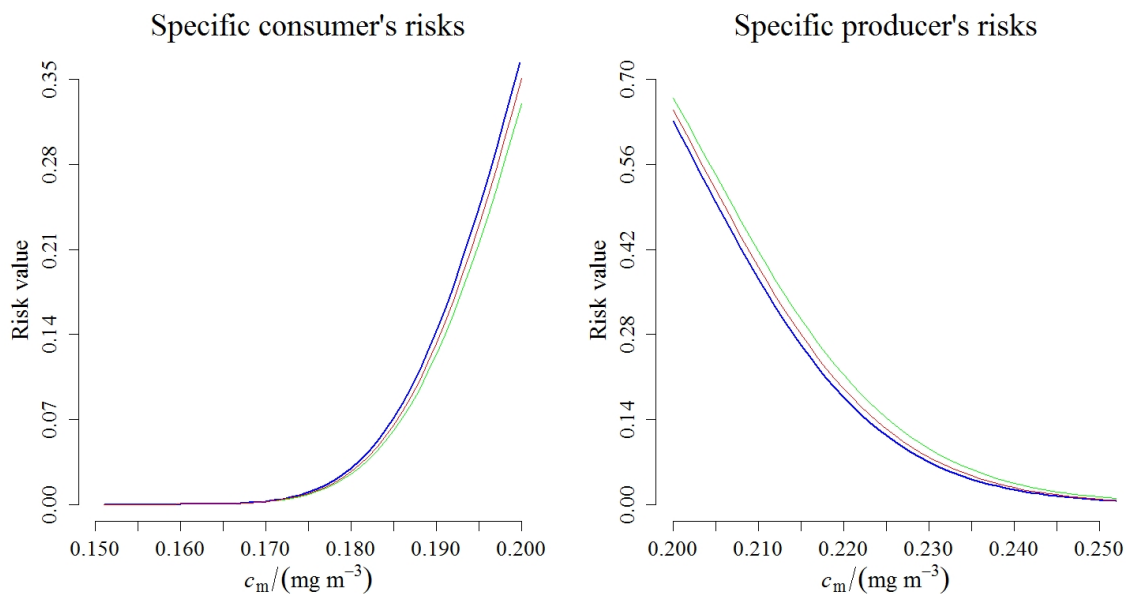
Figure E8.3: Consumer's and producer's specific risks versus test results, for quarry 1 (blue line),
2 (green line) and 3 (red line).

E8.7   Interpretation of results

Studies on global risks, such as that conducted in section E8.6.2, can allow the involved parties (consumers and producers) to agree on an acceptance limit (balancing the safeguarding of the inhabitants' health and the economical interests of the quarries' owners, in the considered example).

The approach in section E8.6.3 provides risks of false decision for a specific test result and for a particular acceptance limit ($A_U = T_U$, in the considered case). From a practical point of view, no action will be undertaken when a measurement result is under the acceptance limit, that is when it is conforming with the requirements. However, when a test result exceeds the limit, it will be declared as non conforming and some corrective action will be required. In this case, the producer has at hand a tool for assessing the extent of his/her responsibility for such failure and possibly elaborate an appropriate reaction. As an example, for a non-conforming test result c_{1m} = 0.225 mg m^{-3}, the specific producer's risk for quarry 1 is about 12 %, meaning that there is a non-negligible 12 % probability of such a test result to correspond to an actually conforming true value c_1.

Example E9

Uncertainty evaluation for the quantification of low masses of benzo[a]pyrene

F. Pennecchi, F. Rolle, M. Sega, S.L.R. Ellison, A.M.H van der Veen

E9.1 Summary

The aim of the present example is to show the uncertainty evaluation for the quantification of low masses of benzo[a]pyrene (BaP), which is an important Polycyclic Aromatic Hydrocarbon (PAH) for ambient air monitoring. Comparison between the results obtained according to the GUM uncertainty framework [51] and the Monte Carlo method for the propagation of distributions [52, 54], applied to both real and simulated data sets, are shown and discussed.

E9.2 Introduction of the application

The quantification of low masses of PAHs is an important issue as they are ubiquitous toxic contaminants which can be present in all the environmental compartments even at trace levels. The evaluation of the uncertainty associated with the quantification of such micro pollutants plays an important role for the reliability of their measurements. Among PAHs, BaP is classified as carcinogenic agent and is listed in the current European legislation [86] as marker of the carcinogenic risk for the whole class of PAHs in ambient air.

The present example aims at comparing the results obtained by application of the Law of Propagation of Uncertainty (LPU) [51] and the Monte Carlo method (MCM) for propagation of distributions [52] to real data sets derived from the quantification of a low mass of BaP spiked on filters commonly used for airborne particulate matter sampling. The comparison is performed also on simulated data corresponding to a BaP mass at trace level.

The description of the analytical method to quantify BaP in ambient air can be found in [154], whereas details on the uncertainty evaluation process, not explicitly reported in the present example, can be found in [157].

E9.3 Specification of the measurand

A glass fiber filter (Pall & Whatman) having diameter of 47 mm, a type of filter commonly used for the sampling of airborne particulate matter, was spiked with the Certified Reference Material

(CRM) NIST SRM 2260a, containing 36 PAHs in an organic solution. The spiked filter was extracted by Soxhlet, following the extraction procedure described in [154]. The same filter was subsequently extracted a second time thus obtaining a diluted sample. BaP masses in the two extracts were quantified by means of a gas chromatograph coupled with a mass spectrometer (GC-MS) Focus DSQ II (Thermo Fisher Scientific).

The measurand of interest in the present example is the mass of BaP contained in a nominal volume of $1\,\mu L$ of the second extract. Moreover, in order to consider very low mass values of BaP, a numerical simulation was carried out by decreasing the chromatographic areas corresponding to the BaP in the sample of the second extraction by a common constant term, hence reaching a (simulated) measurand value close to the minimum mass of BaP detectable with the method described in [154], i.e., $2.5 \cdot 10^{-3}$ ng.

E9.4   Measurement model

Quantification of the mass of BaP contained in $1\,\mu L$ of the second extract was performed according to the Internal Standard method prescribed by [13]. An aliquot of the NIST SRM 2270, containing perdeuterated benzo[a]pyrene (BaP-d_{12}), was added to the solution in order to obtain a concentration of BaP-d_{12} equal to $0.2455\,\mu g\,ml^{-1}$, to be used as the internal standard. Then, an aliquot of $1\,\mu L$ of the solution was injected three times in the GC-MS. The ratio of peak areas corresponding to the internal standard and those corresponding to the analyte was used to determine the mass m_E of BaP present in the injected volume of the extracted sample, according to the following model:

$$m_E = \left(f\, \overline{A_E}\, m_{ISE} \right) / \left(\overline{A_{ISE}} \right), \tag{E9.1}$$

where f is the GC-MS calibration factor, $\overline{A_E}$ is the mean area (a.u.) of the chromatographic peak corresponding to BaP in the extract, and m_{ISE} and $\overline{A_{ISE}}$ are the mass (ng) and the mean chromatographic area (a.u.) for the internal standard in the extract (ISE).

Calibration factor f was obtained as the arithmetic mean of three calibration factors corresponding to three reference solutions at different BaP concentrations. Details on the calibration procedure are reported in [157]. In the evaluation of the uncertainty associated with f (characterized by 9 degrees of freedom), covariance terms between the three factors were taken into account: they were due to the same mass of the internal standard used in the calibration model for each of the factors and to the same CRM used for preparing the three necessary reference solutions. For the same reason, f and m_{ISE}, as input quantities of measurement model (E9.1), were correlated because of the use of the same internal standard both in the calibration and in the analysis process.

The value and uncertainty associated with the mass of the internal standard m_{ISE} were derived from its calibration certificate (the uncertainty was considered as having a very high number of degrees of freedom, so that it did not give contribution to the effective degrees of freedom for the uncertainty of the measurand estimate).

$\overline{A_E}$ and $\overline{A_{ISE}}$ were evaluated as the arithmetic means of three repetitions of the area measurement of the relevant chromatographic peaks. Their uncertainty was calculated as the standard deviation of such mean [51, Sec. 4.2.3] (hence, having two degrees of freedom). A strong linear relationship was observed between the areas of the BaP and those of the ISE in the same run, hence a corresponding covariance term for the two mean areas was calculated according to [51, Sec. 5.2.3].

Estimates, uncertainties $u(x)$ and covariances $u(x, y)$ of the input quantities in model (E9.1) are reported in Table (E9.1), together with other parameters relevant to the uncertainty eval-

uation performed by MCM. The corresponding estimate for the measurand is $m_E = 0.014\,$ng.

Table E9.1: Estimate, uncertainty, covariance and distributional parameters of the input quantities in model (E9.1).

Parameter	Value
f	0.616
$u(f)$	0.017
m_{ISE}	0.2455 ng
$u(m_{ISE})$	0.0036 ng
$u(f, m_{ISE})$	$-3.3 \cdot 10^{-5}$ ng
$\overline{A_E}$	85 114 a.u.
$u(\overline{A_E})$	9564.35 a.u.
$\overline{A_{ISE}}$	917 545.67 a.u.
$u(\overline{A_{ISE}})$	44 492.21 a.u.
$u(\overline{A_E}, \overline{A_{ISE}})$	$-203\,436\,959.5$ (a.u.)^2
S_{11}	548 861 202 (a.u.)^2
$S_{12} = S_{21}$	$-1\,220\,621\,757$ (a.u.)^2
S_{22}	11 877 338 582 (a.u.)^2

Note that, in order to simulate a smaller value of BaP mass, the experimental results obtained for the sample of the second extraction were re-used as they were, but the areas corresponding to the BaP were all decreased by a common constant term equal to 67 000 a.u.. Therefore, all the values in Table (E9.1) are still valid for the simulated case[1], except for the $\overline{A_E}$ value which becomes equal to 18 114 a u.. The corresponding estimate for the (simulated) measurand is $m_E = 0.003$ ng.

E9.5   Uncertainty propagation

For calculating the uncertainty associated with the estimates of the measurands (i.e. $m_E = 0.014$ ng and $m_E = 0.003$ ng, respectively), both the LPU [51] and the MCM for the propagation of probability distributions [52, 54] were applied and compared. Details of the calculation (expressed to at least six significant figures to avoid rounding errors) are available in the data elaboration file "A212_BaP_example.r".

E9.5.1   GUM uncertainty framework

Applying the LPU to model (E9.1), taking into account the uncertainty and covariance contributions of the input quantities reported in Table (E9.1), the resulting uncertainty $u(m_E)$ was 0.0020 ng and 0.0017 ng for the experimental and the simulated case, respectively. Table (E9.2) reports the uncertainty budget for the mass $m_E = 0.014$ ng of BaP of the sample obtained with the second extraction (the first two columns repeat part of the information already available in Table (E9.1)). Effective degrees of freedom ν_{eff} were calculated, according to the Welch-Satterthwaite formula [51, eqn. G2.b] applied to the input uncertainties and the corresponding degrees of freedom discussed in Sec. E9.4. They were equal to 3.07 and 2.05, giving coverage factors k

---

[1]The uncertainty assumed for the simulated arithmetic mean of the peak areas is probably larger than that expected for a material actually close to the detection limit.

Table E9.2: Uncertainty budget for the mass of BaP (in 1 μL) of the sample obtained with the second extraction: associated combined uncertainty was $u(m_E) = 0.002$ ng.

Component	$u(x_i)$	$\frac{\partial m_E}{\partial x_i}$	$\left[\frac{\partial m_E}{\partial x_i} u(x_i) \right]^2$
$u(f)$	$1.7 \cdot 10^{-2}$	$2.3 \cdot 10^{-2}$ ng	$1.4 \cdot 10^{-7}$ ng^2
$u(m_{ISE})$	$3.6 \cdot 10^{-3}$ ng	$5.7 \cdot 10^{-1}$	$4.3 \cdot 10^{-8}$ ng^2
$u(\overline{A_E})$	$9.6 \cdot 10^3$ a.u.	$1.6 \cdot 10^{-7}$ ng (a.u.)$^{-1}$	$2.5 \cdot 10^{-6}$ ng^2
$u(\overline{A_{ISE}})$	$4.4 \cdot 10^4$ a.u.	$-1.5 \cdot 10^{-8}$ ng (a.u.)$^{-1}$	$4.6 \cdot 10^{-7}$ ng^2
	$u(x_i, x_j)$	$\frac{\partial m_E}{\partial x_i} \frac{\partial m_E}{\partial x_j}$	$2 \frac{\partial m_E}{\partial x_i} \frac{\partial m_E}{\partial x_j} u(x_i, x_j)$
$u(f, m_{ISE})$	$-3.3 \cdot 10^{-5}$ ng	$1.3 \cdot 10^{-3}$ ng	$-8.5 \cdot 10^{-8}$ ng^2
$u(\overline{A_E}, \overline{A_{ISE}})$	$-2.0 \cdot 10^8$ (a.u.)^2	$-2.5 \cdot 10^{-15}$ ng^2 (a.u.)$^{-2}$	$1.0 \cdot 10^{-6}$ ng^2
$u^2(m_E)$			$4.1 \cdot 10^{-6}$ ng^2

of 3.1 and 4.2 for the real and the simulated case, respectively. Coverage factors of a Student t-distribution with an integer number ν of degrees of freedom are given in [51, Table G.2], otherwise, i.e. for a non-integer ν, they can be recovered by means of common statistical software. For obtaining a 95 % coverage interval for the distribution, the 97.5[th] percentile of the distribution is used for calculating the expanded uncertainty according to [51, eqn. G.1d]. In the present case, the expanded uncertainties $U = k\,u(m_E)$ at a 95 % coverage probability were equal to 0.006 ng and 0.007 ng for the real and the simulated case, respectively. Results of application of the LPU to both the experimental and the simulated case are summarized in Table (E9.3). Note that the expanded uncertainty at the lower level is larger than that at the higher level because of the larger coverage factor multiplying the corresponding standard uncertainty. Indeed, even if neither the number of measurement repetitions nor the uncertainties involved in the application of the Welch-Satterthwaite formula change from one model to the other, the different values of the sensitivity coefficients lead to different effective degrees of freedom in the two cases.

Table E9.3: Estimate of the measurand, with associated standard and expanded uncertainty, for the experimental and the simulated case.

	m_E (ng)	$u(m_E)$ (ng)	$U(m_E)$ (ng)
Experimental case	0.014	0.0020	0.006
Simulated case	0.003	0.0017	0.007

E9.5.2 Monte Carlo method

The MCM for propagation of probability distributions of the input quantities was applied in order to obtain an approximated distribution for the measurand, i.e. the mass of BaP in the extract and in the simulated case. For this purpose, suitable probability distributions were assigned to the input quantities of model (E9.1), according to the criteria prescribed in [52, 54].

Since the available information on f and m_{ISE} were their best estimates and their associated covariance matrix, a bivariate Gaussian distribution was assigned to these quantities [52, Sec. 6.4.8]. Hence, the bivariate normal distribution had a (vector) expectation equal to $[f, m_{ISE}]$ and

a covariance matrix Σ equal to

$$\Sigma = \begin{bmatrix} u^2(f) & u(f, m_{ISE}) \\ u(f, m_{ISE}) & u^2(m_{ISE}) \end{bmatrix},$$

whose components are available in Table (E9.1).

Since the two ($N = 2$) quantities A_E and A_{ISE} were considered as following a bivariate normal distribution and, for each quantity, ($n = 3$) repeated measurements were available, a scaled and shifted bivariate t-distribution with one degree of freedom ($\nu = n - N$) was assigned to them, according to [54, Sec. 5.3.2]. Hence, the bivariate t-distribution had a (vector) expectation equal to $[\overline{A_E}, \overline{A_{ISE}}]$ and the scale matrix $\mathbf{S}/n$ with $\mathbf{S}$ defined by:

$$\mathbf{S} = \frac{1}{\nu} \begin{bmatrix} \sum_{i=1}^{3}(A_{E_i} - \overline{A_E})^2 & \sum_{i=1}^{3}(A_{E_i} - \overline{A_E})(A_{ISE_i} - \overline{A_{ISE}}) \\ \sum_{i=1}^{3}(A_{E_i} - \overline{A_E})(A_{ISE_i} - \overline{A_{ISE}}) & \sum_{i=1}^{3}(A_{ISE_i} - \overline{A_{ISE}})^2 \end{bmatrix},$$

whose components S_{ij} are available in Table (E9.1), for $i, j = 1, 2$. Note that for $\nu = 1$, the mean value and the covariance matrix of the t-distribution are not defined, anyway a coverage (hyper) interval for the distribution can always be determined [54, Section 5.5.2, Note 1].

The numerical simulation of the input probability distributions and their propagation through measurement model (E9.1) were implemented in R environment [149] by applying R functions "rmvnorm" and "rmvt" available in the "mvtnorm" package [93]. For each input quantity, $M = 10^7$ values were drawn. Since only positive values of measurand are feasible, the joint input probability density functions were numerically truncated at zero by disregarding negative values drawn during the MCM simulation [52, Sec. 9.4.2.1.1, Note], thus obtaining a number of corresponding simulated BaP mass values smaller than M. The number of MCM replicates retained, however, was about $9 \cdot 10^6$ and $7 \cdot 10^6$ for the real and the simulated case, respectively, hence still providing a reliable numerical approximation for the measurand distribution.

From the MCM distribution distribution, the shortest 95 % coverage interval was obtained, for both real and simulated data, and reported in Table (E9.4).

E9.6 Reporting the result

Figs. (E9.1) and (E9.2) show the approximate numerical representation of the pdf for the BaP mass corresponding to the second extraction and to the simulated case, respectively, indicating the 95 % coverage intervals for the measurand m_E produced according to the MCM and to the GUM uncertainty framework. The relevant interval limits are also reported in Table (E9.4).

Table E9.4: Measurand estimate, 95 % coverage interval according to the GUM uncertainty framework and the MCM for propagation of distributions. (ng)

	m_E	GUM 95 % C.I.	MCM 95 % C.I.
Second extraction	0.014	[0.008, 0.020]	[$8.3 \cdot 10^{-7}$, 0.032]
Simulated very low extraction	0.003	[-0.004, 0.010]	[$5.9 \cdot 10^{-10}$, 0.020]

In the present example, the MCM simulations involved an input bivariate distribution with 1 degree of freedom, leading to an output pdf with an extreme right tail. The standard deviations of the MC output distribution were unreliable (6.8 ng and 5.8 ng for the experimental and the simulated case, respectively) and, because of the truncation effect, the corresponding sample means were heavily biased (0.028 ng and 0.017 ng, respectively). Hence, this is a case in which

neither the MCM mean nor the standard deviation are reliable, but just the coverage interval at a desired coverage level should be reported. Incidentally, both the MC medians (0.014 ng and 0.004 ng, respectively) resulted very close to the measurand estimates in Table (E9.4), proving themselves as robust and sensible estimates for the measurand.
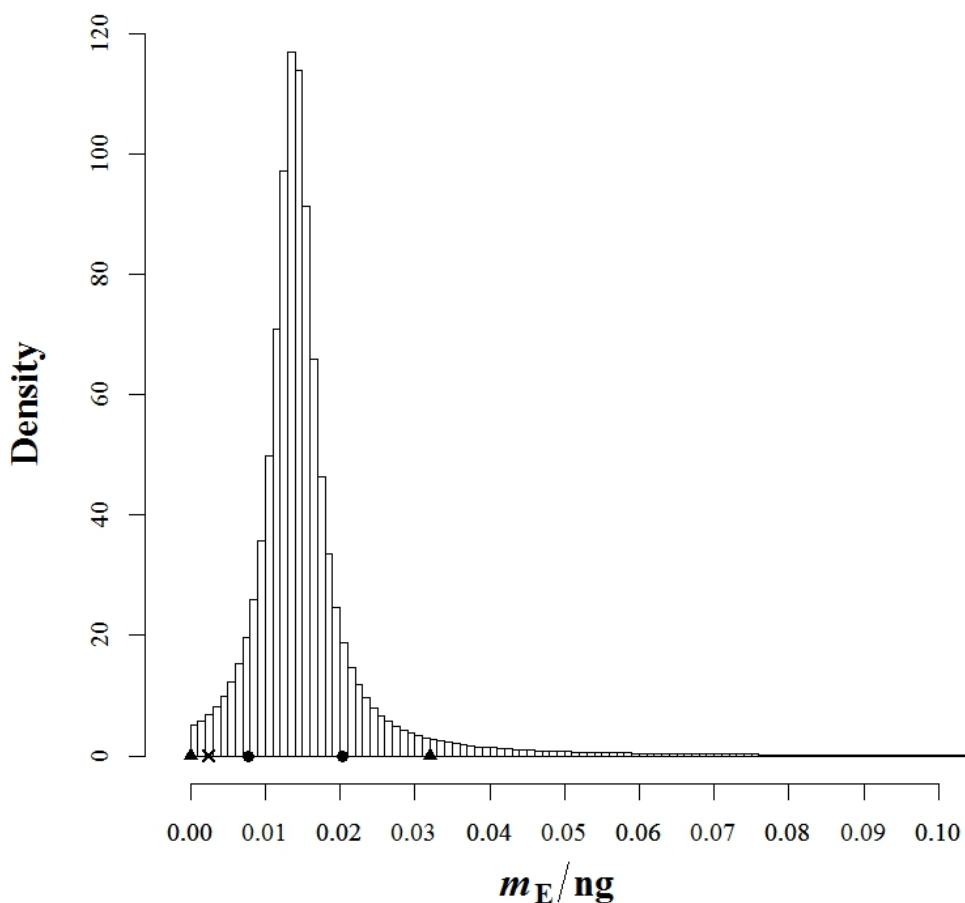
Figure E9.1: Numerical representation of the pdf associated with the mass of BaP in the nominal volume of 1 μL of the sample obtained with the second extraction ($m_E = 0.014$ ng). Circle and triangle symbols indicate the limits of the 95 % coverage interval obtained according to the GUM uncertainty framework and by MCM, respectively. Symbol x indicates the minimum detectable mass of the analytical method.
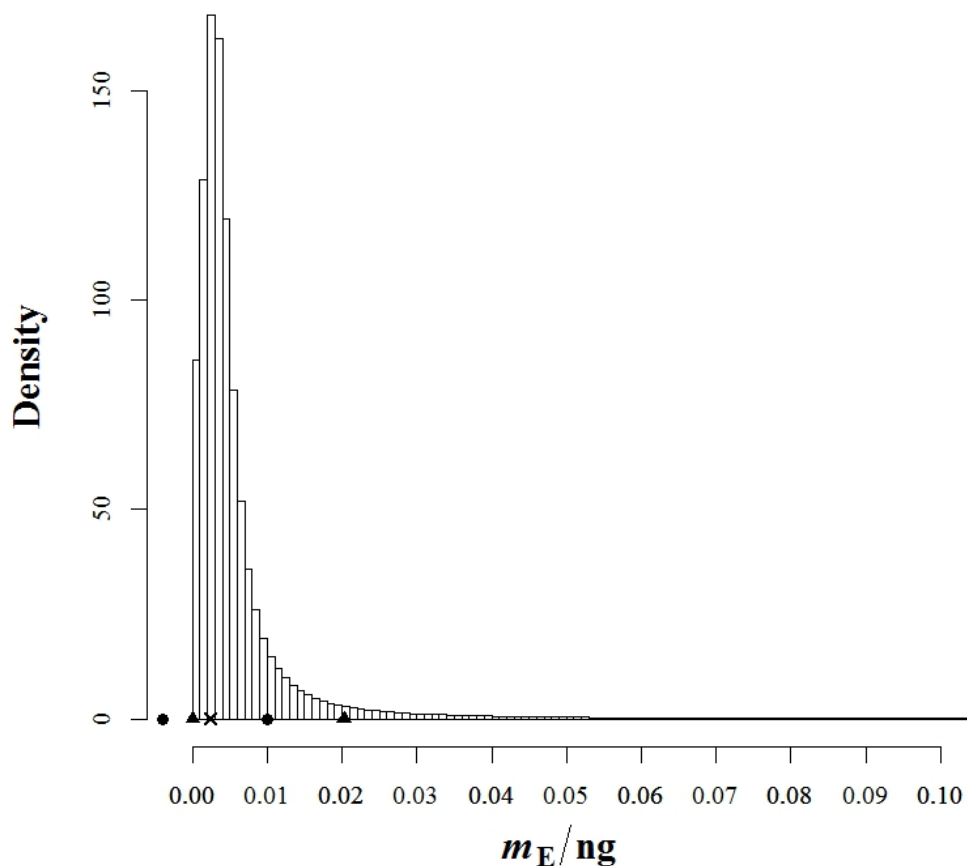
Figure E9.2: Numerical representation of the pdf associated with a simulated very small mass value of BaP ($m_E = 0.003$ ng). Circle and triangle symbols indicate the limits of the 95 % coverage interval obtained according to the GUM uncertainty framework and by MCM, respectively. Symbol x indicates the minimum detectable mass of the analytical method.

E9.7   Interpretation of results

When applying MCM, the measurand estimate and the associated uncertainty are usually taken as the mean and the standard deviation of the simulated output results, according to [52, eqs. 16 and 17]. Nonetheless, NOTE 2 in [52, Sec. 6] states that in some special circumstances, such as when one of the input quantities has been assigned a PDF based on the t-distribution with fewer than three degrees of freedom, the expectation and standard deviation of the output quantity might not exist and the above-cited equations (16) and (17) might not then provide meaningful results. A coverage interval for the measurand can, however, be formed, since the simulated output distribution is meaningful. This is exactly the situation of the present example, for which, in fact, plausible estimates and corresponding standard uncertainties are those obtained within the GUM uncertainty framework, as reported in Table E9.3, whereas feasible coverage intervals are those provided by MCM, as discussed in the following.

From both Figs. (E9.1) and (E9.2), it is evident that the two approaches give quite different

results in terms of coverage intervals. Although the assumed output distribution in the GUM uncertainty framework is a Student t-distribution with few degrees of freedom, hence leading to a large coverage factor for the calculation of the corresponding expanded uncertainty, the MCM coverage interval is about 2.5 and 1.5 times larger than that obtained in the GUM uncertainty framework, respectively. Moreover, it is asymmetric with respect to the measurand estimate because of the left censoring of simulated results. Due to the very few degrees of freedom of the input bivariate Student t-distribution of the mean areas and due to the fact that both the bivariate Student t and the Gaussian input distributions were feasibly truncated at zero, the MCM output distribution has in fact a very long right tail, resulting in a net positive bias of the mean value and a considerable inflation of the standard deviation. This is a clear example of those situations in which the conditions required by the Central Limit Theorem are not met, since the pdf for the output quantity is not a Gaussian distribution nor a scaled and shifted t-distribution.

Moreover, at the lower simulated mass value (Fig. 2), the GUM uncertainty framework would lead to a coverage interval stretching into a region of negative unfeasible values. MCM simulation, instead, can provide a realistic asymmetric interval. It is worth mentioning that the EURACHEM/CITAC guidelines on uncertainty evaluation for analytical measurements [18] recommend truncating the expanded uncertainty interval at zero whenever a negative lower limit is found for a positive-defined quantity.

Summarizing, this example is a clear case in which blind adherence either to the approach in the GUM [51] or to the MCM in [52, 54] would be dangerous. Careful considerations on estimates, standard uncertainties and coverage intervals are always needed, according to the specific problem under study.

Example E10

Calibration of an analyser for NO$_\text{X}$ using gas mixtures prepared with mass flow controllers

F. Pennecchi, F. Rolle, M. Sega, P.G. Spazzini, I. de Krom, A.M.H. van der Veen

E10.1   Summary

The present example shows the uncertainty evaluation of the calibration of a chemiluminescence analyser for nitrogen oxides (NO$_\text{x}$) using a multi-point calibration as described in ISO 6143 [6] with dynamically prepared calibration gas mixtures obtained by dynamic dilution of standard gas mixtures performed by means of calibrated mass flow controllers (MFCs) [37].

This example addresses the need for a more advanced treatment of correlations arising in such measurements, especially those caused by the use of the same equipment for calibration gas mixture preparation and the use of one calibration gas mixture from which the dilutions are made.

E10.2   Introduction of the application

The European Directive on ambient air quality [87] prescribes the monitoring of NO$_\text{x}$ by means of chemiluminescence as the reference method [19], which requires the use of proper calibration gas mixtures for instrument calibration. To prepare such mixtures, dynamic dilution is a primary method considered as a valid alternative to the static gravimetric method: it allows preparing ready-to-use gas mixtures at low amount fractions by diluting a standard mixture (parent mixture) with a proper diluent gas, thus avoiding stability problems related to diluted mixtures of reactive gases in high-pressure cylinders.

The work consists in the following steps:

1. **Use of two calibrated MFCs** to dilute a static calibration gas mixture with a diluent gas to obtain reference gas mixtures having the analyte amount fraction in the range of interest (for environmental monitoring applications). The uncertainty associated with and the covariance between the flow values generated by the MFCs are evaluated by taking into account the calibration and the repeatability contributions.

2. Employing the classic model equation of the **dynamic dilution**, the uncertainty associated with and the covariance between the amount fractions of the analyte in the different mix-

tures are evaluated by taking into account contributions arising from i) the flow of the parent mixture and that of the dilution gas, ii) the amount fraction of the analyte in the parent mixture and iii) the (possible) impurities of the analyte in the diluent gas. Detailed calculation of the relevant results are shown in the Excel spreadsheet A213_data_elaboration.xls.

3. Use of the obtained NO_x reference mixtures for **calibration of a chemiluminescence analyser** in the desired range of amount fractions. Weighted Total Least-Squares (WTLS) regression is applied, taking into account uncertainties associated with and covariances among the values of both the dependent and independent variables.

In the following sections, each step will be addressed in detail.

E10.3   Specification of the measurand

E10.3.1   Use of two calibrated MFCs

In the present example, two MFCs from MKS with full scale range (FSR) of $500\,\mathrm{cm}^3\,\mathrm{min}^{-1}$ and $2000\,\mathrm{cm}^3\,\mathrm{min}^{-1}$ were employed, after calibration, for the parent mixture and the diluent gas, respectively. The calibration of the MFCs was performed against the INRIM Microgas station, the Italian primary flow standard for low flow rates. The MFCs were characterised in terms of their calibration coefficient $C = q_{VR}/q_{VN}$, where q_{VR} is the (reference) volume flow rate at standard conditions (often expressed in "standard cubic centimetres per minute" (SCCM)) supplied by the MFC under calibration as it is read by the Microgas, whereas q_{VN} is the set (nominal) volume flow rate of the MFC. For the MFC with FSR $500\,\mathrm{cm}^3\,\mathrm{min}^{-1}$ the following model was found appropriate

$$C_1 = \alpha_1/q_{VN} + \beta_1 + \gamma_1 q_{VN} + \delta_1 q_{VN}^2 \,, \tag{E10.1}$$

whereas for the MFC with FSR $2000\,\mathrm{cm}^3\,\mathrm{min}^{-1}$, the appropriate model was

$$C_2 = \alpha_2/q_{VN} + \beta_2/\sqrt{q_{VN}} + \gamma_2 + \delta_2\sqrt{q_{VN}} + \epsilon_2 q_{VN} \,, \tag{E10.2}$$

where C_i, for $i = 1, 2$, indicates the calibration coefficient of MFC1 and MFC2, respectively. Guidance on this kind of model selection is given in ISO/TS 28038 [38].

Weighted Least-Squares regression was employed for fitting eqs. (E10.1) and (E10.2) to experimental data, considering q_{VN} as not uncertain (being the flow rate set at the MFC), whereas C was affected by several uncertainty contributions: that due to the measurement repeatability (evaluated by the standard deviation of the repeated measurements of C) and that associated with the reference Microgas (accounting for the uncertainty in the involved measurements of temperature, pressure and volume). The curve parameter estimates are reported in table E10.1, whereas associated (squared) uncertainties and covariances are shown in covariance matrices (E10.3) and (E10.4), respectively.

$$V_{C_1} = \begin{pmatrix} 9.2 \times 10^{-4} & -1.84 \times 10^{-5} & 9.5 \times 10^{-8} & -1.3 \times 10^{-10} \\ -1.84 \times 10^{-5} & 4.4 \times 10^{-7} & -2.4 \times 10^{-9} & 3.4 \times 10^{-12} \\ 9.5 \times 10^{-8} & -2.4 \times 10^{-9} & 1.4 \times 10^{-10} & -2.1 \times 10^{-14} \\ -1.3 \times 10^{-10} & 3.4 \times 10^{-12} & -2.1 \times 10^{-14} & 3.3 \times 10^{-17} \end{pmatrix} \tag{E10.3}$$

$$V_{C_2} = \begin{pmatrix} 18.60 & -4.0 & 3.0 \times 10^{-1} & -9.3 \times 10^{-3} & 9.9 \times 10^{-5} \\ -4.0 & 8.9 \times 10^{-1} & -6.7 \times 10^{-2} & 2.1 \times 10^{-3} & -2.2 \times 10^{-5} \\ 3.0 \times 10^{-1} & -6.7 \times 10^{-2} & 5.1 \times 10^{-3} & -1.6 \times 10^{-4} & 1.7 \times 10^{-6} \\ -9.3 \times 10^{-3} & 2.1 \times 10^{-3} & -1.6 \times 10^{-4} & 4.9 \times 10^{-6} & -5.3 \times 10^{-8} \\ 9.9 \times 10^{-5} & -2.2 \times 10^{-5} & 1.7 \times 10^{-6} & -5.3 \times 10^{-8} & 5.8 \times 10^{-10} \end{pmatrix} \tag{E10.4}$$

Table E10.1: Calibration curve parameter estimates for the two MFCs according to eqs. (E10.1) and (E10.2) (measurement units are such that each parameter times its unit is adimensional).

Parameter	MFC1	MFC2
α	-2.493	-39.967
β	1.021	5.306
γ	-1.590×10^{-4}	6.104×10^{-1}
δ	1.975×10^{-7}	1.195×10^{-2}
ϵ		-1.249×10^{-4}

As an example, the calibration curve for the MFC1 with FSR of $500 \, cm^3 \, min^{-1}$ is shown in figure E10.1 (all volumes here and in the following are referred to temperature and pressure standard conditions, i.e. $0 \, °C$ and $1013.25 \, mbar$).
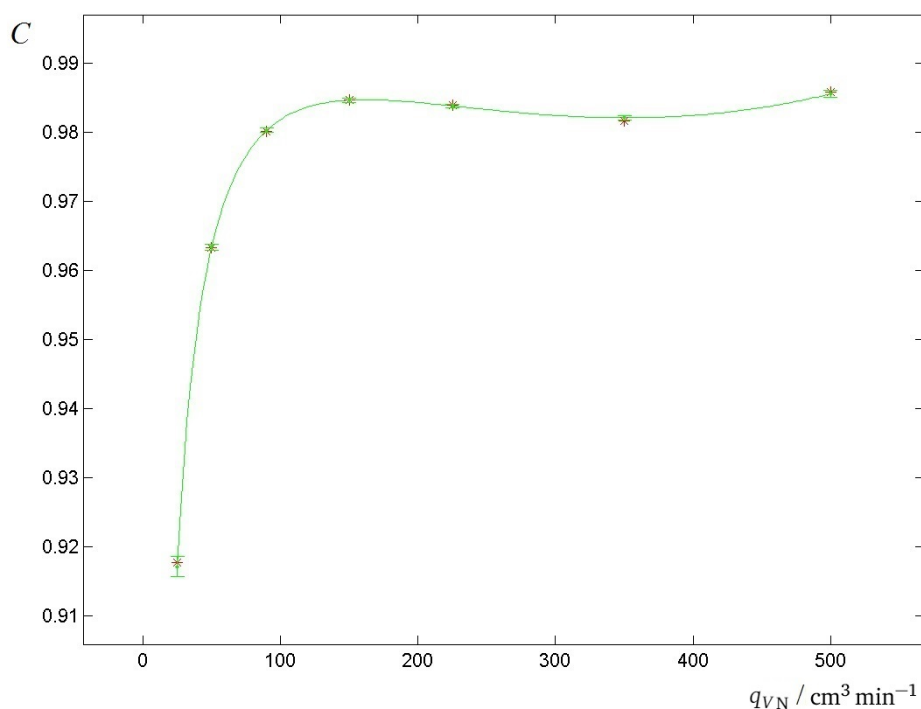
Figure E10.1: Calibration curve of the MFC1 with FSR of $500 \, cm^3 \, min^{-1}$

The calibration coefficient, obtained after calibration of the MFCs, is used as correction factor for the volume flow rate set at the MFCs to obtain the volume flow rate actually provided.

E10.3.2 Dynamic dilution

In the present example, the preparation of reference gas mixtures of nitrogen dioxide (NO_2) in synthetic air (NO_2/SA) by dynamic dilution is addressed. The measurand is the amount fraction of NO_2 in the prepared calibration gas mixture. This mixture is obtained by mixing the flow of a parent mixture, containing a known amount fraction of NO_2 supplied by MFC1, with a flow of

the dilution gas (SA), a high pure gas containing just impurities (possibly also NO_2), supplied by MFC2. The method is described in ISO 6145-7 [37].

E10.3.3 Calibration of a chemiluminescence analyser

A Thermo Fisher Scientific 42i chemiluminescence analyser is calibrated for NO_2/SA, in the range 700 nmol mol^{-1}-1300 nmol mol^{-1}. To this aim, a parent mixture with an amount fraction of NO_2 of 10.252 µmol mol^{-1} in SA is diluted with SA (grade 4.7) in order to dynamically prepare three calibration gas mixtures with different amount fractions of NO_2. At each amount fraction level, three repeated readings from the analyser are collected: their mean and the corresponding standard deviation are taken as the estimate provided by the analyzer and the associated uncertainty, respectively. WTLS regression is performed [6] to fit a straight line to the calibration data points. The measurand is the set of the parameters of the instrument analysis function. The analysis function relates the amount fraction to the instrument response and can be used to calculate the amount fraction, given an instrument response.

E10.4 Measurement model

E10.4.1 Use of two calibrated MFCs

After the MFCs are calibrated, the flow q_{Vi} they provide at a nominal flow q_{ViN} is given by the following model:

$$q_{Vi} = q_{ViN} C_i, \tag{E10.5}$$

where C_i is the calibration coefficient calculated according to eq. (E10.1) or (E10.2).

E10.4.2 Dynamic dilution

In the presented case of a binary mixture, the dynamic dilution involves two MFCs (MFC1 and MFC2) which regulate two different gas flows, i.e. MFC1 is used for the parent mixture and MFC2 for the diluent gas. The amount fraction x_a of the analyte gas in the mixture prepared by dynamic dilution is calculated according to the following model equation:

$$x_a = \frac{x_1 q_{V1a} + x_2 q_{V2a}}{q_{V1a} + q_{V2a}}, \tag{E10.6}$$

where x_1 is the amount fraction (nmol mol^{-1}) of the analyte in the parent mixture, x_2 is the amount fraction (nmol mol^{-1}) of the analyte potentially present in the diluent gas (impurity), q_{V1a} is the flow (cm^3 min^{-1}) of the parent mixture supplied by MFC1, q_{V2a} is the flow (cm^3 min^{-1}) of the diluent gas supplied by MFC2[1]. Assuming that the analyte is not present in the diluent gas and that this is certified with zero uncertainty (as in the present case study), equation (E10.6) can be simplified into the following:

$$x_a = \frac{x_1 q_{V1a}}{q_{V1a} + q_{V2a}}. \tag{E10.7}$$

Three different mixtures having amount fractions x_a, x_b and x_c, respectively, are generated.

---

[1]Equation (E10.6) applies if the compressibility factors of the parent gases are equal. This condition is usually met if (1) the matrix of the parent gases is the same and (2) the amount fraction of the other components is low, say below 10 µmol mol^{-1}. The latter limit depends, among other, on the target measurement uncertainty.

E10.4.3   Calibration of a chemiluminescence analyser

As the analysis curve of the instrument, a straight line

$$y = A + Bw \tag{E10.8}$$

is fitted to the data, which are the three different amount fractions x_a, x_b, x_c (the y values in equation (E10.8)), and the sample mean of three repeated measurement at each amount fraction level (w values). An analysis function (reference amount fractions on the ordinate axis and means of the repeated readings on the abscissa axis), rather than a calibration curve, is determined, since it allows to easily employ the calibration output when the analyser is subsequently used in field: for each new reading, the instrument analysis curve provides a straightforward estimate of the amount fraction of an unknown sample under analysis, with an associated uncertainty. In order to fit the analysis curve, a WTLS regression was performed by means of the CCC software [140], taking into account the covariance matrices associated with both w and y values. The main advantage of the WTLS algorithm is indeed the possibility to deal with regression problems involving uncertain and correlated variables.

In the present case, the y values are characterised by a covariance matrix whose terms are later defined by eqs. (E10.14) and (E10.16), whereas the covariance matrix associated with the w values is diagonal (instrumental readings at different NO_2 amount fractions are not correlated) with elements equal to the (square of the) standard deviation of the means of the three repeated readings obtained at each amount fraction.

E10.5   Uncertainty propagation

E10.5.1   Use of two calibrated MFCs

Using to the law of propagation of uncertainty (LPU) from [51], the uncertainty associated with a flow q_{Vi} (E10.5) produced by the i-th calibrated MFC is given by

$$u(q_{Vi}) = q_{ViN}\, u(C_i). \tag{E10.9}$$

Uncertainty $u(C_i)$ can be expressed as the sum in quadrature of a systematic contribution, $u(C_i)_{cal}$, due to the MCF calibration (and calculated by applying the LPU to eq. (E10.1) or (E10.2), respectively, taking into account uncertainties of and covariances between the curve parameters), and a repeatability contribution, $u(C_i)_{rep}$, of the MFC when it is used (in the specific case, the repeatability experienced in the dynamic dilution was similar to that typically encountered within the calibration process at approximately the same flow values). Therefore,

$$u(C_i) = \sqrt{u^2(C_i)_{cal} + u^2(C_i)_{rep}}\,. \tag{E10.10}$$

Concerning covariances, covariance term $u(q_{V1a}, q_{V2a})$ between flow rates provided by two MFCs at the same nominal value q_{VNa} is considered negligible, in this context, since the MFCs, even if calibrated against the same reference standard, are different instruments, calibrated in different moments, by means of different calibration functions. On the other hand, covariance terms $u(q_{V1a}, q_{V1b})$ and $u(q_{V2a}, q_{V2b})$ between flows generated by the same MFC at two different nominal values q_{VNa} and q_{VNb} are not negligible since flow estimates by the same MFC are recovered by the application of the very same calibration curve. By considering expression (E10.5), one has

$$u(q_{Via}, q_{Vib}) = u(q_{ViNa}C_{ia}, q_{ViNb}C_{ib}) = q_{ViNa}\, q_{ViNb}\, u(C_{ia}, C_{ib})\,, \tag{E10.11}$$

where, when $i = 1$, for example, and hence applying eq. (E10.1),

$$u(C_{1a}, C_{1b}) = u(\alpha_1/q_{V1Na} + \beta_1 + \gamma_1 q_{V1Na} + \delta_1 q_{V1Na}^2, \ \alpha_1/q_{V1Nb} + \beta_1 + \gamma_1 q_{V1Nb} + \delta_1 q_{V1Nb}^2).$$
(E10.12)

Employing the covariance property for linear combinations of variables, eq. (E10.12) becomes

$$u(C_{1a}, C_{1b}) = 1/(q_{V1Na}q_{V1Nb})u^2(\alpha_1) + 1/q_{V1Na} u(\alpha_1, \beta_1) + \ldots + (q_{V1Na}q_{V1Nb})^2 u^2(\delta_1).$$ (E10.13)

Therefore, uncertainties associated with and covariances between parameters of calibration curve (E10.1) influence the covariance between two different flows produced by the same MFC1. Analogous expressions are derived for MFC2 as well.

E10.5.2 Dynamic dilution

Using the LPU, the (squared) uncertainty associated with the amount fraction x_a of the analyte (E10.7) is given by

$$u^2(x_a) = \left(\frac{\partial x_a}{\partial x_1}\right)^2 u^2(x_1) + \left(\frac{\partial x_a}{\partial q_{V1a}}\right)^2 u^2(q_{V1a}) + \left(\frac{\partial x_a}{\partial q_{V2a}}\right)^2 u^2(q_{V2a}),$$
(E10.14)

where $u(x_1)$ is provided by the certificate of the reference parent mixture, while $u(q_{V1a})$ and $u(q_{V2a})$ are calculated according to expression (E10.9). Note that $u(q_{V1a}, q_{V2a}) = 0$. Analogous expressions hold for x_b and x_c as well.

Covariances between two different amount fractions x_a and x_b are calculated as:

$$\begin{aligned} u(x_a, x_b) \ &= \ u\left(\frac{x_1 q_{V1a}}{q_{V1a} + q_{V2a}}, \frac{x_1 q_{V1b}}{q_{V1b} + q_{V2b}}\right) \approx \\ &\approx \ \frac{\partial x_a}{\partial x_1}\frac{\partial x_b}{\partial x_1}u(x_1, x_1) + \frac{\partial x_a}{\partial x_1}\frac{\partial x_b}{\partial q_{V1b}}u(x_1, q_{V1b}) + \\ &\quad + \frac{\partial x_a}{\partial x_1}\frac{\partial x_b}{\partial q_{V2b}}u(x_1, q_{V2b}) + \ldots + \frac{\partial x_a}{\partial q_{V2a}}\frac{\partial x_b}{\partial q_{V2b}}u(q_{V2a}, q_{V2b}). \end{aligned}$$
(E10.15)

Recalling that $u(q_{V1a}, q_{V2b}) = u(q_{V2a}, q_{V1b}) = 0$, and considering that there is no covariance between x_1 and any of MFC flow values, eq. (E10.15) reduces to:

$$\begin{aligned} u(x_a, x_b) \ &\approx \ \frac{q_{V1a}q_{V1b}}{(q_{V1a} + q_{V2a})(q_{V1b} + q_{V2b})}u^2(x_1) + \\ &\quad + \frac{x_1^2 q_{V2a}q_{V2b}}{(q_{V1a} + q_{V2a})^2(q_{V1b} + q_{V2b})^2}u(q_{V1a}, q_{V1b}) + \\ &\quad + \frac{x_1^2 q_{V1a}q_{V1b}}{(q_{V1a} + q_{V2a})^2(q_{V1b} + q_{V2b})^2}u(q_{V2a}, q_{V2b}), \end{aligned}$$
(E10.16)

where $u(q_{Via}, q_{Vib})$ are calculated according to eq. (E10.11). Analogous expressions hold for $u(x_a, x_c)$ and $u(x_b, x_c)$, as well.

E10.5.3 Calibration of a chemiluminescence analyser

Estimate of coefficients A and B of analysis curve (E10.8) and the associated covariance matrix are the main output of the applied WTLS software. Details on such estimates and covariance matrix are available in the User Manual of the CCC software [140].

E10.6   Reporting the result

Nominal and measured flow values of MFC1 and MFC2 are reported in table (E10.2) together with associated uncertainties. These are calculated by eqs. (E10.9) and (E10.10), where contribution $u(C_i)_{rep}$ is equal to 0.013 % and 0.037 % of the measured flow q_{Vi} value for MFC1 and MFC2, respectively.

Table E10.2: Nominal and measured flow values of MFC1 and MFC2 with associated uncertainties ($cm^3\,min^{-1}$).

	q_{V1N}	q_{V2N}	q_{V1}	$u(q_{V1})$	q_{V2}	$u(q_{V2})$
a	84	1116	82.243	0.017	1108.49	0.54
b	115	1085	113.088	0.023	1077.17	0.53
c	152	1048	149.677	0.028	1039.78	0.57

Covariance terms between measured flow values of the same MFC are calculated according to eqs. (E10.11) and (E10.12) (and corresponding ones for amount fraction values b and c, and for MFC2), and reported in table (E10.3).

Table E10.3: Covariance terms between measured flow values of the same MFC (MFC1 and MFC2) at different fraction levels a, b and c (($cm^3\,min^{-1})^2$).

MFC	$u(q_{Va}, q_{Vb})$	$u(q_{Va}, q_{Vc})$	$u(q_{Vb}, q_{Vc})$
1	$2.31 \cdot 10^{-4}$	$2.54 \cdot 10^{-4}$	$3.58 \cdot 10^{-4}$
2	$1.26 \cdot 10^{-1}$	$1.22 \cdot 10^{-1}$	$1.20 \cdot 10^{-1}$

The parent mixture of NO_2 has an amount fraction $x_1 = 10.252\,\mu mol\,mol^{-1}$, with associated uncertainty $u(x_1) = 0.016\,\mu mol\,mol^{-1}$.

Applying eq. (E10.7), and corresponding ones for amount fractions b and c, the three reference amount fractions are obtained as reported in table (E10.4).

Table E10.4: amount fractions of the three calibration mixtures obtained by dynamic dilution and associated uncertainties.

amount fraction	x ($nmol\,mol^{-1}$)	$u(x)$ ($nmol\,mol^{-1}$)
a	708.1	6.3
b	974.1	6.1
c	1290.1	6.4

Relevant squared uncertainties and covariances (in $nmol^2\,mol^{-1}$), reported in the covariance matrix (E10.17), are calculated according to eqs. (E10.14) and (E10.16), exploiting uncertainty and covariance values associated with q_{Vi} reported in tables (E10.2) and (E10.3).

$$V_{x_{a,b,c}} = \begin{pmatrix} 39.5 & 1.7 & 2.2 \\ 1.7 & 37.0 & 3.0 \\ 2.2 & 3.0 & 41.2 \end{pmatrix}. \tag{E10.17}$$

Each calibration mixture is analyzed three times: the sample mean and the associated uncertainty values are reported in table (E10.5). A straight line is fitted to the data by means of WTLS

Table E10.5: Sample mean of repeated readings and associated uncertainties.

Mean Reading (a.u.)	u(Mean Reading) (a.u.)
146.6	0.75
213.7	0.67
285.0	1.53

regression. Parameter estimates of model (E10.8) are $A = 88.94$ nmol mol^{-1} and $B = 4.19$ nmol mol^{-1} a.u.$^{-1}$, respectively, and the associated covariance matrix $V_{A,B}$ is (E10.18).

$$V_{A,B} = \begin{pmatrix} 2.85 \cdot 10^2 & -1.30 \\ -1.30 & 6.36 \cdot 10^{-3} \end{pmatrix} . \tag{E10.18}$$

Validation of the obtained analysis curve is then performed by analysing a known gas mixture with the calibrated instrument and comparing its output with this value: the validation is passed if the two values are consistent within their expanded uncertainties. In the present case, an independent gas mixture of NO_2 at the amount fraction of 975.5 nmol mol^{-1} (with uncertainty equal to 1.5 nmol mol^{-1}) is used, obtained by dynamic dilution starting from a gas mixture of NO_2 at amount fraction of 5.113 µmol mol^{-1} in SA diluted with SA 4.7. Applying model (E10.8) to the mean of three repeated readings of the instrument ($w = 213.33$ a.u. and $u(w) = 0.11$ a.u.) corresponding to the independent gas mixture, the estimate $y = 982.0$ nmol mol^{-1} is obtained with associated uncertainty $u(y) = 4.7$ nmol mol^{-1}. Such uncertainty is obtained by application of the LPU to model (E10.8) propagating uncertainty $u(w)$ and terms of the covariance matrix (E10.18) through the model, that is applying the following equation:

$$u^2(y) = u^2(A) + u^2(B) w^2 + u^2(w) B^2 + 2 u(A, B) w . \tag{E10.19}$$

The validation result is reported in fig. (E10.2).

E10.7   Interpretation of results

Covariances in matrix (E10.17) between amount fractions of the three calibration mixtures are mainly due to the term proportional to $u(x_1)$ (first term of eq. (E10.16)), the uncertainty of the amount fraction of the common parent mixture used for obtaining all the calibration mixtures. Such contribution is of practically the same order of magnitude of the resulting covariance term, whereas contributions relevant to $u(q_{V\,a}, q_{V\,b})$, $u(q_{V\,a}, q_{V\,c})$ and $u(q_{V\,b}, q_{V\,c})$ in table (E10.3) (second and third terms of eq. (E10.16)) are smaller by one or two orders of magnitude.

If covariances between the three amount fractions were ignored in the calibration of the chemiluminescence analyser, i.e., if a diagonal version of matrix (E10.17) were used in the WTLS regression, slightly different estimates of A and B for the analysis curve (E10.8) would be obtained, with a different associated covariance matrix. The corresponding results of the validation process would be $y = 982.0$ nmol mol^{-1} and $u(y) = 4.5$ nmol mol^{-1}, showing an undervaluation of the uncertainty up to 3.4 %.
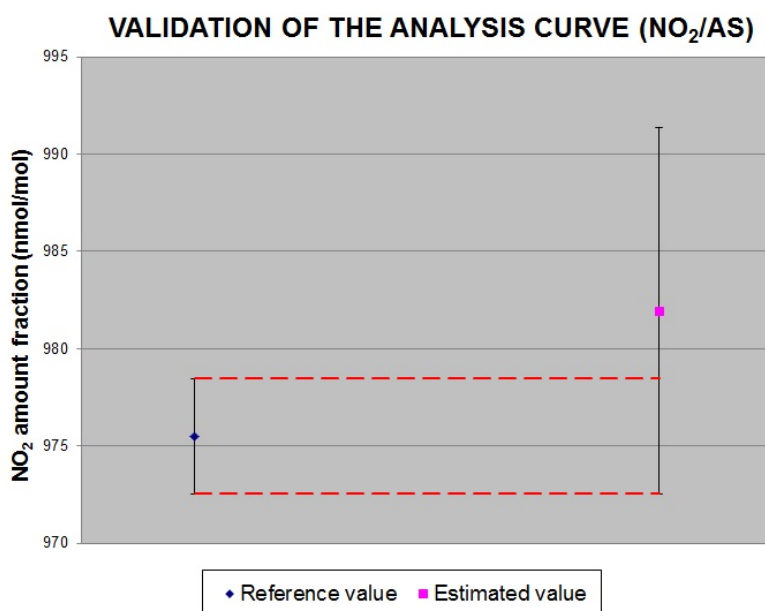
Figure E10.2: Validation of the analysis curve for the quantification of NO_2 in SA. Estimate of a gas mixture of 975.5 nmol mol^{-1} of NO_2/SA as provided by the calibrated instrument (Estimated value) in comparison with the actual value of the mixture as reported in its calibration certificate (Reference value).

Example E11

Calibration of a sonic nozzle as an example for quantifying all uncertainties involved in straight-line regression

S. Martens, K. Klauenberg, B. Mickan, C. Yardin, N. Fischer, C. Elster

E11.1   Summary

When calibrating a sonic nozzle, it is recommended to estimate the straight-line relationship between the discharge coefficient of the nozzle and the square root of the inverse Reynolds number for a gas. The slope and intercept of this relation characterise the nozzle, and reliable estimates and uncertainties for this multivariate measurand are mandatory for its use as transfer or working standard.

This example emphasises the importance of accounting for correlation for a reliable uncertainty evaluation. The use of common reference standards and instruments causes correlation among and between the discharge coefficient and the Reynolds number, and impacts significantly on the uncertainty of the characteristic parameters of the nozzle. To show this, a measurement model based on the weighted total least-squares (WTLS) method is applied and its input quantities are fully characterised. In particular, we demonstrate in detail how to jointly evaluate the correlation, uncertainties and estimates for the input quantities of least-squares methods applying the Monte Carlo method. As a result, sonic nozzles can be characterised in line with the *Guide to the expression of uncertainty in measurement* (GUM).

E11.2   Introduction of the application

Sonic nozzles are widely used to determine gas flow rates with high precision and excellent reproducibility. Technically called Critical Flow Venturi Nozzles, they are internationally recognised as a calibration standard for gas flow meters and other flow measurement devices, which in turn facilitate traceable measurements e.g. in gas and oil pipelines, in chemical, pharmaceutical and food industries as well as for fuel dispensers, water, heat and gas meters at home. Sonic nozzles are also employed in dilution systems for the preparation of calibration gas mixtures [31], for flow limiting and overspeed protection of gas flow meters, to name a few.

115
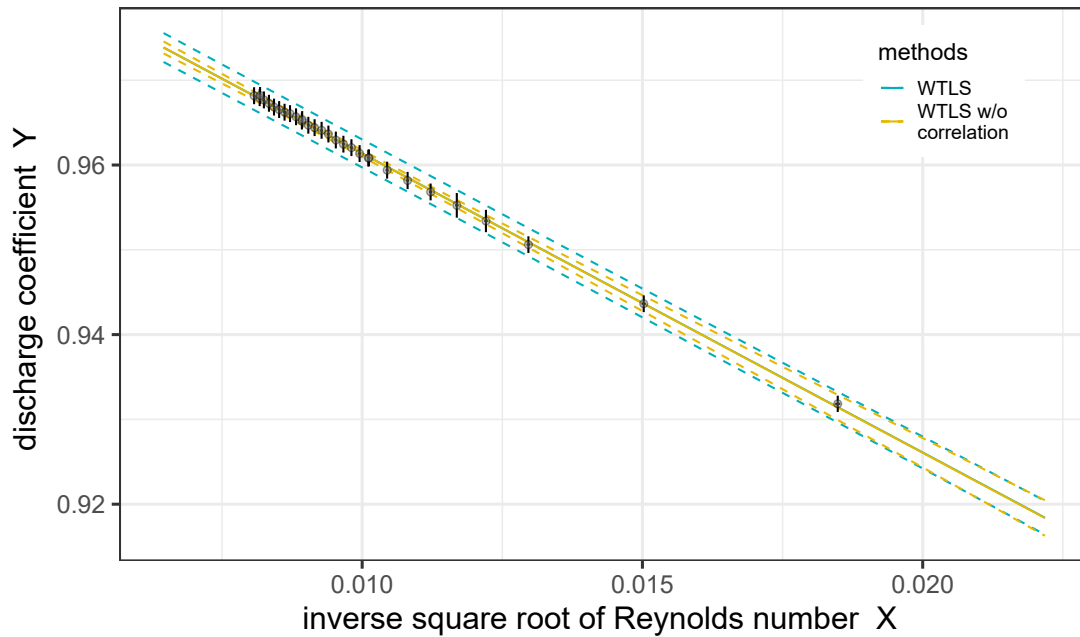
Figure E11.1: Estimates and uncertainties for the discharge coefficient Y_n as a function of the
inverse square root of Reynolds number X_n (both dimensionless) for measurements $n = 1, \ldots, 30$.
The data are available online in repository [120]. Superimposed are the estimated straight-line
relationship (solid lines) and the associated 95 % coverage bands (dashed lines) obtained by
WTLS (blue) and by WTLS without correlation (yellow).

For toroidal sonic nozzles, the standard ISO 9300 [12] specifies the following relation be-
tween the discharge coefficient Y and the Reynolds number Re

$$Y = \beta_0 + \frac{\beta_1}{\sqrt{Re}}, \tag{E11.1}$$

see Appendix E11.A for some background information. The two parameters β_0 and β_1 charac-
terise a specific nozzle (depending on its inner contour curvature and surface structure [94,95]).
The use of the nozzle as transfer standard depends on the reliability of the estimates and uncer-
tainties of these parameters.

In this example, a toroidal sonic nozzle shall be characterized based on $N = 30$ pairs of dis-
charge coefficients Y_n and inverse square root of Reynolds numbers $X_n = Re_n^{-0.5}$, $n = 1, \ldots, N$. Es-
timates and uncertainties for these quantities, as measured at PTB and displayed in figure E11.1,
rely on common input quantities which cause correlation (cf. clause 5.2.4 in [51]). Such corre-
lations among and between the quantities X_n and Y_n along with their uncertainties need to be
evaluated and accounted for, to reliably quantify the characteristic slope β_1 and intercept β_0 of
the straight-line relation (E11.1).

The aim of this example is to derive reliable estimates and uncertainties for the characteristic
values of a sonic nozzle following the GUM. As illustrated in figure E11.2, we will proceed in two
stages:

1. Estimates, uncertainties, and correlations for all quantities X_n, Y_n are evaluated. For this
   purpose, probability distributions for all input quantities are propagated through a first,
   joint measurement model with the help of the Monte Carlo method.
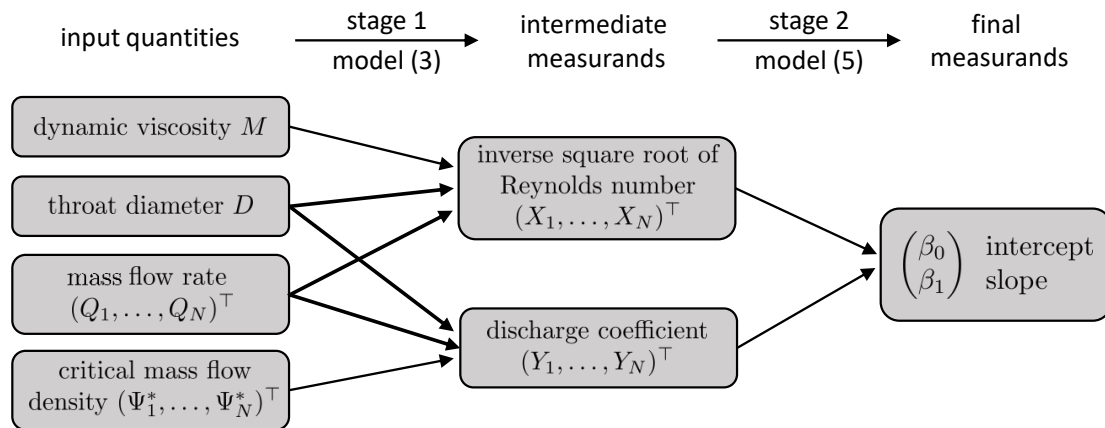
Figure E11.2: Illustration of the two consecutive measurement models and their input quantities.

2. Estimates and uncertainties for the measurands β_0 and β_1 are evaluated from a second measurement model. The latter is based on the weighted total least-squares (WTLS) method which accounts for all uncertainties and correlations derived in stage 1.

E11.3   Specification of the measurands

Let X denote the inverse square root of the Reynolds number, and Y the discharge coefficient. According to equation (E11.1) and equivalently standard [12], Y depends linearly on X,

$$Y = \beta_0 + \beta_1 X. \tag{E11.2}$$

The measurands are the intercept parameter β_0 and slope parameter β_1 of this straight-line model.

The quantities $\boldsymbol{X} = (X_1,\ldots,X_N)^\top$ and $\boldsymbol{Y} = (Y_1,\ldots,Y_N)^\top$ influence the measurands and in turn are influenced by further input quantities. For stage 1 of this example, $(\boldsymbol{X},\boldsymbol{Y})$ will itself be a measurand – a $2N$-dimensional, intermediate one. The full covariance matrix $\boldsymbol{U}$ as well as the estimates $(\boldsymbol{x},\boldsymbol{y})$ for this intermediate measurand shall be evaluated and in turn used to evaluate estimates and uncertainties for the final measurands β_0 and β_1. The covariance matrix $\boldsymbol{U}$ contains on its diagonal the squared standard uncertainties associated with the estimates, and on its off-diagonal positions, the covariances associated with pairs of these estimates, see [54, clause 3.20]. Thus, uncertainties and correlations can be derived directly from $\boldsymbol{U}$.

E11.4   Measurement models

The uncertainty evaluation in this example consists of two consecutive measurement models, see figure E11.2. Section E11.4.1 first describes the measurement model for the intermediate measurand $(\boldsymbol{X},\boldsymbol{Y})$ and characterises its input quantities. Section E11.4.2 describes the subsequent measurement model for the final measurands β_0 and β_1. Equivalently, both models could be viewed as a single, multi-stage model.

Table E11.1: Characterization of the input quantities in measurement model (E11.3). The estimate and Type B evaluated, relative standard uncertainty as well as the assigned distribution is given for each input quantity.

Input Quantity	Distribution	Estimate	Unit	Type B relative stand. uncertainty $\times 10^{-3}$	References
D throat diameter	Gaussian	1.2067×10^{-3}	m	0.4165	Cal. Cert.[1]
M dynamic viscosity	Gaussian	$1.82 \quad \times 10^{-5}$	$\mathrm{kg\,m^{-1}\,s^{-2}}$	5	[113]
Q_n mass flow rate	Gaussian	–	$\mathrm{kg\,s^{-1}}$	0.55	[1]
Ψ_n^* critical mass flow	Gaussian	–	$\mathrm{kg\,m^{-2}\,s^{-1}}$	0.25	[12, clause 8.3] [108, 168]

E11.4.1 Intermediate measurands X and Y (stage 1)

Referring to appendix E11.A, the inverse square root of the Reynolds number $\boldsymbol{X} = (X_1, \ldots, X_N)^\top$ and the discharge coefficient $\boldsymbol{Y} = (Y_1, \ldots, Y_N)^\top$ are modelled as follows

$$X_n = \sqrt{\frac{\pi M D}{4 Q_n}}, \quad Y_n = \frac{4 Q_n}{\pi D^2 \Psi_n^*} \quad \text{with } n = 1, \ldots, N, \tag{E11.3}$$

and denoting $\boldsymbol{Q} = (Q_1, \ldots, Q_N)^\top$ and $\boldsymbol{\Psi}^* = (\Psi_1^*, \ldots, \Psi_N^*)^\top$. Equations (E11.3) define the joint, $2N$-variate measurement model for the intermediate measurand $(\boldsymbol{X}, \boldsymbol{Y})$. The throat diameter of the nozzle D, the dynamic viscosity of the gas M, and the n-th real mass flow rate Q_n influence the quantity X_n. The same diameter D, the same flow rate Q_n, and the n-th critical mass flow density Ψ_n^* impact on the quantity Y_n. (See also figure E11.2.) The intermediate measurands $\boldsymbol{X}$ and $\boldsymbol{Y}$ are correlated due to the common quantities $\boldsymbol{Q}$ and D. The common quantities M and D cause additional correlation among the $\boldsymbol{X}$ and among the $\boldsymbol{Y}$, respectively.

Let us characterise these input quantities for the considered example. The estimates and uncertainties for the throat diameter D and the dynamic viscosity of the gas M are displayed in table E11.1. A Gaussian distribution is assigned to each of these two input quantities. The quantities Q_n and $\Psi_n^*, n = 1, \ldots, N$, are derived from measurements which were calibrated against references. The estimates and Type A evaluated uncertainty contributions $u_A(.)$ are obtained from repeated measurements and are provided online in the repository [120]. The Type B uncertainty contributions, say $u_B(.)$, are given in table E11.1. The Type A and Type B evaluated uncertainties are added quadratically $\sqrt{u_A^2(.) + u_B^2(.)}$ for each quantity Q_n and Ψ_n^*, and a Gaussian distribution is assigned to each. In addition, we assume independence between the Type B contributions to the components of the quantities $\boldsymbol{Q}$ and $\boldsymbol{\Psi}$ (which could be revised using [12, clause 9] and [108] if necessary). For general guidance, how to assign distributions to input quantities (also multivariate ones), see Supplements 1 and 2 to the GUM [52, 54].

Having specified all input quantities, the assigned distributions can be propagated through model (E11.3) with the help of the Monte Carlo method to arrive at a joint $2N$-dimensional distribution of the intermediate measurand $(\boldsymbol{X}, \boldsymbol{Y})$, see section E11.5.1 and Supplement 2 to the GUM [54]. Estimates $(\boldsymbol{x}, \boldsymbol{y})$ and the full covariance matrix $\boldsymbol{U}$ can then be derived from the Monte Carlo samples of this distribution.

---

[1]The dimensions of the inner geometry are measured by an accredited laboratory using a coordinate measuring machine (CMM). The uncertainty is calculated based on the method of "virtual CMM" [166].

E11.4.2   Weighted total least-squares method (stage 2)

The measurement model for the straight-line regression (E11.2) can be constructed from the appropriate least-squares method. The frequently applied ordinary and the weighted least-squares method may be inappropriate here, because they assume the measured values x to be exact. Weighted Total Least-Squares (WTLS) takes into account uncertainties associated with the estimates x and y for the intermediate measurands X and Y as well as their associated covariances. WTLS is recommended by multiple standards [16, 38] and applied here.

The WTLS method is based on minimizing the generalized sum of squares

$$S = \begin{pmatrix} x - \widetilde{\xi} \\ y - (\widetilde{\beta}_0 + \widetilde{\beta}_1 \widetilde{\xi}) \end{pmatrix}^\top U^{-1} \begin{pmatrix} x - \widetilde{\xi} \\ y - (\widetilde{\beta}_0 + \widetilde{\beta}_1 \widetilde{\xi}) \end{pmatrix}, \tag{E11.4}$$

with respect to $\widetilde{\beta}_0$, $\widetilde{\beta}_1$ and the unknown, "true" values of x called $\widetilde{\xi}$. Here, the vector x contains the elements $x = (x_1, \ldots, x_N)^\top$ and the vectors y and $\widetilde{\xi}$ are likewise defined. The minimizer of (E11.4) defines the solution $(\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\xi}^\top)$ of the WTLS method.

The measurement model is then defined by replacing the estimates x and y in the minimization of S by the underlying quantities $X = (X_1, \ldots, X_N)^\top$ and $Y = (Y_1, \ldots, Y_N)^\top$, respectively. That is,

$$(\beta_0, \beta_1, \xi^\top)^\top = \arg \min_{\widetilde{\beta}_0, \widetilde{\beta}_1, \widetilde{\xi}} \begin{pmatrix} X - \widetilde{\xi} \\ Y - (\widetilde{\beta}_0 + \widetilde{\beta}_1 \widetilde{\xi}) \end{pmatrix}^\top U^{-1} \begin{pmatrix} X - \widetilde{\xi} \\ Y - (\widetilde{\beta}_0 + \widetilde{\beta}_1 \widetilde{\xi}) \end{pmatrix}, \tag{E11.5}$$

where only (β_0, β_1) define the final measurand in this example.

E11.5   Estimation and uncertainty evaluation

Following the GUM [51, 54], estimates $\widehat{\beta}_0$ and $\widehat{\beta}_1$ of the final measurands are obtained by evaluating measurement model (E11.5) at the estimates x and y of the intermediate measurand (X, Y). The uncertainties associated with $(\widehat{\beta}_0, \widehat{\beta}_1)$ result from propagating the covariance matrix U associated with these intermediate estimates through the same model, which will be detailed in section E11.5.2. Before, section E11.5.1 describes how to arrive at the intermediate estimates x and y, and the full covariance matrix U following the GUM [54].

E11.5.1   Intermediate measurands X and Y (stage 1)

The probability distributions of the input quantities $M, D, \boldsymbol{Q}$ and $\boldsymbol{\Psi}^*$, as assigned in section E11.4.1, shall be propagated through model (E11.3) to evaluate estimates, uncertainties and correlations for the intermediate measurand (X, Y). Application of the multivariate law of propagation of uncertainty [54, clause 6] is lengthy due to the full correlation structure resulting from model (E11.3), and it may not be adequate due to non-linearities in model (E11.3). Instead, the Monte Carlo method [54, clause 7] is applied to approximate the $2N$-dimensional distribution for (X, Y) numerically and summary information is obtained subsequently.

In particular, R repeated samples are drawn from the $2N + 2$ independent Gaussian distributions of the input quantities $M, D, \boldsymbol{Q}$ and $\boldsymbol{\Psi}^*$. Model (E11.3) is applied to each of these samples, generating R samples of the intermediate measurand (X, Y) – each sample representing a random realisation from the $2N$-dimensional distribution. If these output samples form a matrix of dimension $2N \times R$, an estimate (x, y) is obtained by averaging over its columns, and the covariance matrix U is obtained by calculating the covariance between all rows (cf. [54, clause 7.6]).
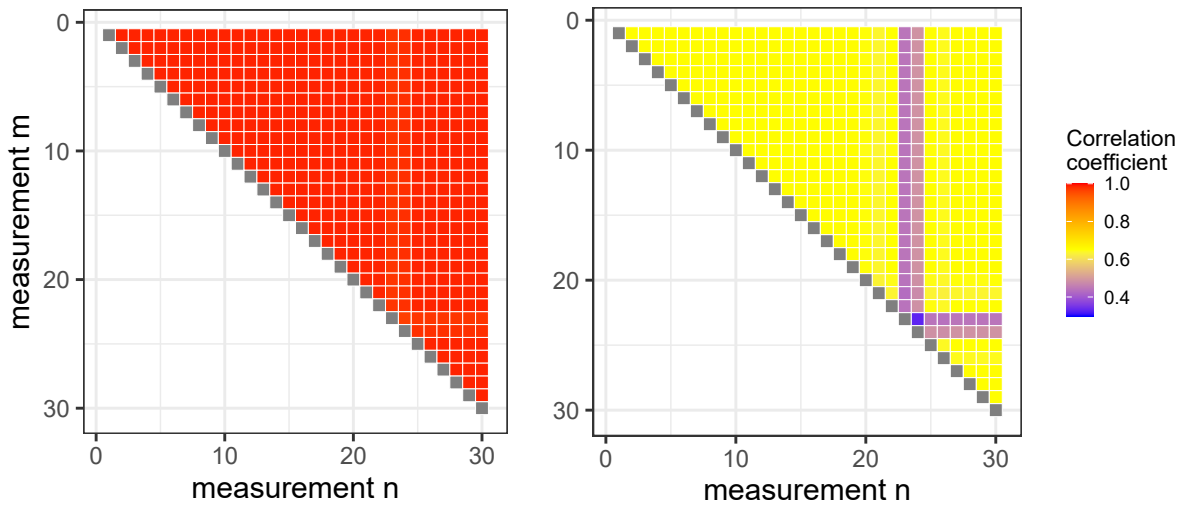
Figure E11.3: Upper triangle of the correlation matrices associated with x (left panel) and with y (right panel), obtained from model (E11.3) and by the Monte Carlo method, cf. sections E11.4.1 and E11.5.1. To increase the visibility of strongly correlated values, values with a perfect correlation of 1 are marked in grey.

This summary information for (X, Y) is available online in repository [120]. The estimates and uncertainties are displayed in figure E11.1 and the correlation matrices for X as well as for Y are displayed in figure E11.3. One observes that the correlation coefficients associated with pairs of estimates x_n and x_m, $n \neq m$, are close to one. This is due to the fact that x_n and x_m are affected by a common, dominating source of uncertainty; in fact, the covariance element $u(x_n, x_m)$ and the uncertainty $u(x_n)$ are governed by the squared relative uncertainties associated with the estimate of M (cf. table E11.1). The correlation coefficients associated with pairs of estimates y_n and y_m are mostly about 2/3 (see right panel in figure E11.3). Only the correlation associated with the components $n = 23$ and $n = 24$ of Y is smaller, because the contributing Type A uncertainties associated with Q_n and Ψ_n^* are larger. Consequently, the combined uncertainties $u(y_{23})$ and $u(y_{24})$ are larger than for the other components of Y, while the covariance element $u(y_n, y_m)$ remains unaffected. The correlation between X and Y is much smaller (in the range between -0.067 and -0.045) and not displayed here.

Let us note, that model (E11.3) theoretically involves a division by 0 and a square root of negative numbers for normally distributed input quantities M, D, Q_n and Ψ_n^*. Practically, this does not cause problems here because the uncertainty of the input quantities is much smaller than the estimate (less than 0.5 %). Formally, truncated Normal distributions could be assigned to each input quantity instead.

The above Monte Carlo procedure was implemented in R Markdown [43] code, which is also available online in repository [120]. We chose to implement $R = 10^8$ Monte Carlo trials providing a relative numerical accuracy of smaller than 10^{-7} for the estimates (x, y), but as high as $5 \cdot 10^{-3}$ for the covariances U (especially for the covariance between X and Y). Nevertheless, the results of the final measurand (β_0, β_1) vary little when repeating the Monte Carlo procedure (no more than 10^{-3} times the associated standard uncertainty).

E11.5.2   Weighted total least-squares method (stage 2)

The measurement model (E11.5) based on the WTLS method is implicit, multivariate, non-linear and usually no closed form is available for its solution. An iterative scheme for deriving estimates

Table E11.2: Results obtained by weighted total least-squares (WTLS), by weighted least-squares (WLS), and by WTLS regression without correlation (WTLS w/o correlation). Listed are the estimates for slope and intercept, their associated uncertainties and covariance.

Method	$\widehat{\beta}_0$	$u(\widehat{\beta}_0)$	$\widehat{\beta}_1$	$u(\widehat{\beta}_1)$	$\text{cov}(\widehat{\beta}_0, \widehat{\beta}_1)$
WTLS	0.996 63	0.000 961	−3.5267	0.048 27	-2.53×10^{-5}
WLS	0.996 64	0.000 966	−3.5275	0.048 55	-2.58×10^{-5}
WTLS w/o correlation	0.996 66	0.000 842	−3.5295	0.082 23	-6.75×10^{-5}

$\widehat{\beta}_i$ with $i = 0, 1$ and their associated uncertainties $u(\widehat{\beta}_i)$ is described in clause 10 of the standard [16] and implemented in the R Markdown code in [120]. This simple scheme also provides correlations between β_0 and β_1. The iterative algorithm requires the covariance matrix U to be symmetric and positive definite, which is fulfilled by construction. For numerically semidefinite matrices the Cholesky decomposition of U could be replaced by a modified version [156], or other decomposition methods.

Assuming a Gaussian distribution[2] for the measurand, a 95 % coverage interval for each measurand β_i is given by

$$\left[\widehat{\beta}_i - 1.96 \, u(\widehat{\beta}_i), \widehat{\beta}_i + 1.96 \, u(\widehat{\beta}_i) \right]. \tag{E11.6}$$

A two-dimensional, joint 95 % coverage region can be calculated following [54, clause 6.5.2].

E11.6 Reporting the result

Table E11.2 contains the estimates $\widehat{\beta}_0$ and $\widehat{\beta}_1$ for the final measurands and their associated standard uncertainties. These results are obtained following the GUM by applying measurement model (E11.5) for the weighted total least-squares method to the estimates x, y and their associated covariance matrix U; which in turn are obtained by applying measurement model (E11.3) using the Monte Carlo method.

For comparison, measurement model (E11.5) is applied ignoring any correlation (WTLS w/o correlation) by assuming a diagonal covariance matrix U (see also [16, clause 8]). In addition, measurement model (E11.5) is applied omitting the uncertainty and correlation in the quantity X, as well as the correlation between X and Y. That is, a simple weighted least-squares (WLS) fit is applied accounting only for the uncertainty and correlation in quantity Y (see also [16, clause 9]). Figure E11.4 shows the estimates $\widehat{\beta}_0$ and $\widehat{\beta}_1$ (markers), their joint 95 % coverage region (ellipses), and the associated 95 % coverage interval for these three least-squares methods.

Nearly identical results have been obtained by applying the Monte Carlo method [54] to measurement model (E11.5) and the algorithm in Ref. [16, clause 10]. The non-linearity of (E11.5) or a non-Normality behind (E11.6) could cause differences, however, this was not observed.

We stress that ordinary and weighted least-squares (OLS and WLS) regression using standard software often involves the estimation of a multiple for the variance of the input quantity Y. Such a procedure is rarely adequate in metrology and estimating variance components of input quantities cannot be formulated as a measurement model. Thus it is not covered by the current GUM. It would usually cause significantly different uncertainties for the regression parameters, which would be only a fifth in this example. A measurement model for OLS regression, which

---

[2]Cf. clause 6.3.3 in [51] and section 10.2.3 in [16] for the approximate validity of this Normality assumption.

enables the user to propagate the uncertainties of the input quantities in the sense of the GUM, is presented and discussed in a related example [119].
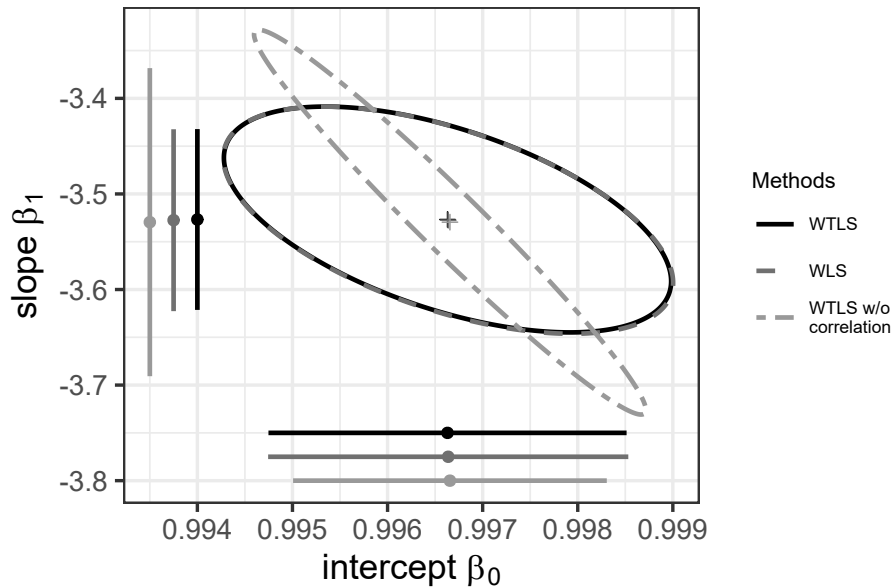
Figure E11.4: Displayed are the estimates $\widehat{\beta}_0$ and $\widehat{\beta}_1$ (markers), their joint 95 % coverage region (ellipses), and the associated 95 % coverage intervals (horizontal and vertical lines) for the least-squares methods listed in table E11.2.

Before interpreting the results of a regression, the data as well as the assumptions contributing to the analysis should be assessed critically. For instance, the graphical analysis of the weighted residuals did not indicate a violation of the straight-line assumption (E11.2), see figure E11.5. Also a more flexible, polynomial model $Y = \beta_0 + \beta_1 X + \beta_2 X^2$ does not improve the fit significantly.

The sonic nozzle used in this example was repeatedly calibrated for many years. For five measurement series the estimates $\widehat{\beta}_0$ and $\widehat{\beta}_1$ and their associated 95 % coverage intervals are displayed in figure E11.6. All five, temporally separate, measurement series display consistent straight-line relationships. Similar equivalence statements are relevant when approving, or disapproving, calibration and measurement capabilities in interlaboratory comparisons and require reliable uncertainties complying with the GUM.

E11.7   Discussion and conclusion

Ignoring correlation in the calibration of the sonic nozzle under consideration causes only minor differences in the estimates of its characteristic parameters β_0 and β_1 (compare WTLS and WTLS w/o correlation in table E11.2 and figure E11.4). However, ignoring any correlation impacts markedly on the uncertainty of these parameters. The standard uncertainty for the intercept $u(\widehat{\beta}_0)$ is understated by about 12 %, while the standard uncertainty for the slope $u(\widehat{\beta}_1)$ is overstated by almost 70 %. In addition, the covariance between β_0 and β_1 increases by a factor of roughly 2.5 if correlations are omitted. As a consequence, the 95 % coverage band around the

---

[3]In contrast to the other measurements, this series has been measured with pressurized air. The associated Type B relative standard uncertainty for Q_n is 0.75×10^{-3}.
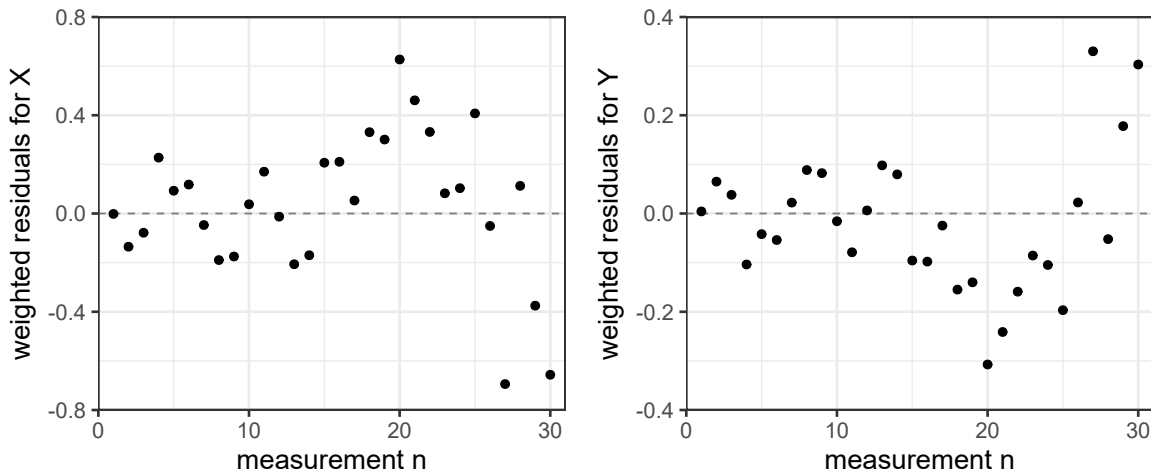
Figure E11.5: Displayed are the residuals $\left(\left(\boldsymbol{x}-\widehat{\boldsymbol{\xi}}\right)^{\top},\left(\boldsymbol{y}-\widehat{\beta}_0-\widehat{\beta}_1\widehat{\boldsymbol{\xi}}\right)^{\top}\right)$ weighted by the Choleski decomposed covariance matrix U. No systematic behaviour indicates a violation of the straight-line relation (E11.2).

estimated straight line has a different shape and is much broader over the range of the data when correlation is accounted for; see figure E11.1.

Ignoring the uncertainty and correlation in the quantity $\boldsymbol{X}$ yields almost identical results in this example (compare WTLS and WLS in table E11.2 and figure E11.4). The reason is twofold: First, the correlation between $\boldsymbol{X}$ and $\boldsymbol{Y}$ is small, and implicitly ignoring it in the simpler weighted least-squares method is inconsequential. Second, also the squared standard uncertainties and co-variances in $\boldsymbol{X}$ relative to the variance of all measurements $x_1, \ldots, x_n$ are two magnitudes smaller than in $\boldsymbol{Y}$ (relative to the variance of $y_1, \ldots, y_n$). For this example, it may thus be sufficient to consider only the uncertainty and correlation in $\boldsymbol{Y}$. While this is not known in advance here, it might be for future calibrations of this nozzle.

Reliable calibration results thus require accounting for correlation. For calibrating a sonic nozzle, we demonstrated how these correlations can be evaluated and accounted for. We provide reliable estimates and uncertainties for the characteristic parameters of the nozzle and contribute to its use as a transfer standard.

Acknowledgment

E11.A Background information on sonic nozzles

Sonic nozzles (also called Critical Flow Venturi Nozzles) consist of a smooth rounded inlet section converging to the throat area $A_{\text{throat}} = 0.25\pi D^2$ (the area of minimum size with throat diameter D) and then diverging along a pressure recovery section. Construction details are described e.g. in the standard ISO 9300 [12]. Here, we follow the notation of the GUM (cf. Note 1–3 in [51, clause 4.1]) and denote all quantities which are used in the main part of this example by capital letter symbols.
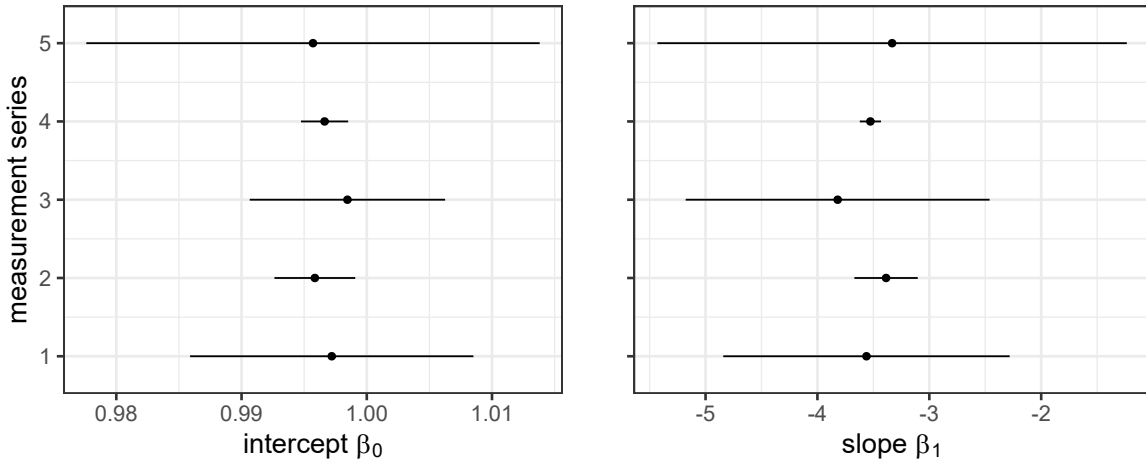
Figure E11.6: Displayed are the estimates $\widehat{\beta}_0$ and $\widehat{\beta}_1$ (dots) and their 95 % coverage intervals obtained by the WTLS method (E11.5) for each of five, temporally separate, measurement series. These were measured (1) at 06/21/2010 with $N = 3$, (2) at 06/28/2012 with $N = 6$, (3) at 10/08/2012[3] with $N = 9$, (4) at 07/15/2018 with $N = 30$, and (5) at 02/21/2019 with $N = 3$.

The gas flow through the nozzle is driven by the pressure difference between the upstream stagnation pressure p_0 and the downstream pressure p_d. If the ratio p_d/p_0 is below a certain critical limit, the maximum flow velocity, which is achieved at the throat, is identical to the local speed of sound of the gas. This velocity cannot be exceeded, even if the pressure ratio p_d/p_0 is further decreased.

The theoretical, ideal mass flow rate $Q^{(\text{ideal})}$ through the nozzle equals the product of the throat area A_{throat} and the so-called critical mass flow density Ψ^*, i.e., $Q^{(\text{ideal})} = A_{\text{throat}}\Psi^*$. Thereby, $\Psi^*(T_0, p_0, \text{gas})$ is purely a function of the stagnation pressure p_0, the stagnation temperature T_0 upstream of the nozzle, and the gas composition. For many gases, this functionality is well known [108, 168].

Because the gas velocity is zero at the wall of the nozzle due to non-slipping conditions, there exists a boundary layer of thickness δ_1 between the core flow and the wall; the gas velocity (and with this the mass flow density) is decreasing from sonic speed down to zero within this layer. Hence, the overall average of the mass flow density Q in the throat is in reality smaller than the theoretical one. The correction is expressed by the discharge coefficient $Y = Q/Q^{(\text{ideal})}$ (in the flow community, the discharge coefficient is usually denoted by c_D). The discharge coefficient is connected with the displacement thickness δ_1 of the boundary layer at the throat via [134, p. 250] [129]

$$Y = \beta_0 \left(1 - 2\frac{\delta_1}{D}\right)^2.$$ (E11.7)

From the Navier-Stokes equation it can be derived that δ_1 is inversely proportional to the square root of Reynolds number Re (i.e. $\delta_1 \propto Re^{-0.5}$) as long as the flow in the boundary layer is laminar. The Reynolds number is a dimensionless flow number, which is given by $Re = Q D/(A_{\text{throat}} M)$ with M being the dynamic viscosity of the gas. Ignoring the term proportional to $(\delta_1/D)^2$ in (E11.7), one obtains the functional relationship between the discharge coefficient Y and the Reynolds number Re provided by the standard ISO 9300 [12], namely,

$$Y = \beta_0 + \beta_1 Re^{-0.5}.$$ (E11.8)

In [12], two nozzle types, the cylindrical and the toroidal one, are defined with tight requirements to keep to specifications. Thereby, equation (E11.8) is valid for both types as long as the flow in the boundary layer is laminar.

Sonic nozzles have a great advantage over sub-sonic flowmeters such as Venturi tubes or orifice plates. In a sonic nozzle, any downstream pressure disturbances cannot move upstream past the throat of the nozzle because the throat velocity is higher (speed of sound of the gas) and in the opposite direction. Therefore, they cannot affect the speed or density of the flow through the nozzle. This is in contrast to Venturi's or orifice plates, where any change in downstream pressure will affect the differential pressure across the flowmeter, which in turn, affects the flow.

Example E12

Bayesian evaluation of a between-bottle homogeneity study in the production of reference materials

A.M.H. van der Veen, S.L.R. Ellison

E12.1 Summary

This example shows how a Bayesian hierarchical model can be used to determine the between-bottle standard deviation of the amount fraction of a component in a set of synthetic natural gas mixtures. The gas mixture are used in a proficiency test. The model takes as input an $n \times k$ table with amount fractions of a selected component, where n denotes the number of gas mixtures and k the number of replicates per mixture. It computes the mean μ, between-group standard deviation τ and within-group standard deviation σ. The model assumes that the amount fractions are, conditionally on the model parameters, normally distributed and uses weakly informative prior probability density functions for the three parameters. The elicitation of the parameters is based on experience in previous proficiency tests. The model is demonstrated for two datasets, one for ethane and a second for nitrogen; in the former case, classical one-way analysis of variance can be used well, but in the latter the classical analysis does not provide a solution. The example also illustrates how such a model can be set up using `R` and `Stan`.

E12.2 Introduction of the application

An essential element in the production of certified reference materials and proficiency test materials in batch form is the evaluation of the between-bottle homogeneity [169]. This form of (in)homogeneity accounts for the (small) differences in the property of interest between the bottles (or more generally, items [32]) and including it in the uncertainty budget of the property value ensures that the value and associated uncertainty are valid for each bottle in the batch, rather than for the batch as a whole [169,173]. The evaluation of between-bottle homogeneity is both a requirement in reference material production [27] as well as in proficiency testing [15,23].

Traditionally, classical analysis of variance (ANOVA) is used [23,32] for this purpose, which is more fully described elsewhere [169]. The parameter of prime interest is the between-group standard deviation, which in this specific case is called the between-bottle standard deviation [169, 170]. Whereas classical ANOVA works well if the between-bottle homogeneity effect is of similar magnitude as the measurement repeatability or greater, difficulties arise when the

between-bottle homogeneity effect is (substantially) smaller than the repeatability effect [170].
Such situations should be avoided [32, 169] but that is not always possible [170].

In this example, a dataset is used that suffers from poor repeatability which justifies the use
of a Bayesian hierarchical model. This model is, apart from the use of prior probability density
functions for the parameters, very similar to the traditional one-way ANOVA model widely used in
the evaluation of homogeneity studies [169]. The background of the model is briefly summarised
here; a more elaborate treatise is available elsewhere [170,174]. A similar model is also available
in the NIST Consensus Builder [105, 106].

E12.3   Specification of the measurand(s)

The measurand in this example is the between-bottle standard deviation (τ) of the amount frac-
tion of a component in a batch of gas mixtures.

E12.4   Measurement model

The statistical model relating the observed amount fractions y_{ij} for mixture i and replicate j to
the mean amount fraction μ, the error in the amount fraction in mixture i, B_i and the random
measurement error ε_{ij} takes the form [169]

$$y_{ij} = \mu + B_i + \varepsilon_{ij} \tag{E12.1}$$

The objective of the evaluation is to determine $\tau^2 = \text{var}(B_i)$ and $\sigma^2 = \text{var}(\varepsilon_{ij})$. If no pooling is
used, then $\sigma_i^2 = \text{var}(\varepsilon_{ij})$, i.e., a standard deviation is computed for each mixture. In a between-
bottle homogeneity study, it would usually make sense to make the assumption that all standard
deviations σ are equal in principle, so to use pooling [170]. The Bayesian treatise presented here
is using pooling of the within-group standard deviations.

In this example, a Bayesian model is used, which implies that a joint prior probability density
function should be chosen for the model parameters. In the case that the parameters are assumed
to be mutually independent, then this joint prior probability density function can be replaced
by the product of three probability density functions, one for each of the parameters. These
probability density functions are specified as follows

$$\mu \sim \text{N}(\mu_0, \sigma_{\text{target}}^2) \tag{E12.2}$$
$$\tau \sim \text{Cauchy}(0, \tau_0) \tag{E12.3}$$
$$\sigma \sim \text{Cauchy}(0, \sigma_0) \tag{E12.4}$$

The prior probability density function (hereafter *prior*) for μ is a normal distribution with mean μ_0
(elicited from the specification of the composition of the gas mixtures) and a standard deviation
that reflects how close the amount fraction for the component of interest of the batch is expected
to be to the specified value. The manufacturer specifies that the actual amount fraction will not
differ more than 5 % from the specified amount fraction. This specification is interpreted as a
95 % coverage interval, and hence a relative standard deviation of 2.5 % is used. This standard
deviation is sufficiently large to ensure that the posterior probability density function (hereafter
posterior) will be dominated by the data [170, 174].

The prior for the between-bottle standard deviation τ is chosen to be the Cauchy distribution
with location parameter 0 and scale parameter τ_0. The latter is obtained from the specification
for the production of the batch gas mixtures, which is usually larger than the value expected for
τ. By using this approach, it is ensured that the scale parameter not smaller than the anticipated

standard deviation [91]. The Cauchy distribution concentrates most of the density between 0 and the scale parameter. Due to the restriction put on τ (and σ), the lower end of the prior is 0, as a standard deviation cannot be negative. A similar approach is used for the prior of σ. The scale parameter σ_0 is set to be equal to the repeatability standard deviations of the amount fractions in this type of mixtures, as observed in previous measurements.

The likelihood is, as stated previously, conditionally on the parameters, a normal distribution [174]

$$y_{ij}|\mu, \tau, \sigma \sim \mathrm{N}(\mu, \tau^2 + \sigma^2) \tag{E12.5}$$

The likelihood of $\bar{y}_i|\theta_i, \sigma_i$ can be described as [92]

$$\bar{y}_i|\theta_i \sim \mathrm{N}(\theta_i, \sigma^2/k)$$

where θ_i denotes the group mean and $\bar{y}_i$ the mean of the y_{ij}, averaged over the replicates. The marginal distributions of the group means $\bar{y}_i$, averaged over the θ_i are independent normal

$$\bar{y}_i|\mu, \tau \sim \mathrm{N}(\mu, \tau^2 + \sigma^2/k)$$

E12.5 Data evaluation

The model as described in the previous sections is used with Bayes' rule. From the weakly informative priors for μ, τ, and σ (equations (E12.2)-(E12.4)), using the data and the likelihood, a joint posterior for the model parameters is obtained. From this posterior, the value for τ, the between-bottle standard deviation, is calculated.

E12.6 Implementation

In Stan code, the model of the between-bottle homogeneity study with pooling of the within-group standard deviations reads as

```
data {
  int<lower=1> N;
  int<lower=1> K;
  matrix[N,K] y;
  int<lower=1> n[N];
  real mu0;
  real tau0;
  real sig0;
}
transformed data {
  matrix[N,K] y_;
  real scale;
  scale = mean(y);
  y_ = y / scale;
}
parameters {
  real mu_;
  real<lower=0> tau_;
  real<lower=0> sig_;
  vector[N] eta;
}
transformed parameters {
  vector[N] theta_;
  theta_ = mu_ + tau_*eta;
}
model {
```

```
  tau_ ~ cauchy(0,tau0/scale);
  sig_ ~ cauchy(0,sig0/scale);
  mu_ ~ normal(mu0/scale,0.025*mu0/scale);
  eta ~ normal(0, 1);
  for (i in 1:N) {
    y_[i,] ~ normal(theta_[i], sig_);
  }
}
generated quantities {        // computation of unscaled parameters
  real mu;
  real<lower=0> tau;
  real<lower=0> sig;
  vector[N] theta;
  mu = mu_ * scale;
  tau = tau_ * scale;
  sig = sig_ * scale;
  theta = theta_ * scale;
}
```

The model consists of the following blocks

1. `data`, declaring the data used by the model

2. `transformed data`, used here to rescale the data by dividing the observed amount fractions by their mean

3. `parameters`, declaring the model parameters and any auxiliary parameters needed for running the calculations

4. `transformed parameters`, declaring the scaled group means

5. `model`, specifying the Bayesian model in terms of the priors and the likelihood

6. `generated quantities`, declaring and computing the unscaled model parameters

In the `data` block, the variables are declared needed for transferring the data. In this block, the number of gas mixtures ("bottles") `N` (n) and the (maximum) number of replicates `K` (k) are declared, following by the table with amount fractions y. `mu0`, `tau0` and `sigma0` are the (hyper)parameters of the priors assigned to μ, τ, and σ respectively.

The next block, `transformed data`, performs a rescaling on the data in y. The transformation consists of calculating the mean of all observed amount fractions and to use this to rescale the data (variable `scale`). The rescaled variables is `y_`. This transformation could also have been performed in R before transferring the data to the Bayesian model. Including it in the model enables for the user of the model to transfer the original data, and as we will see the model also returns the unscaled model parameters (see the `generated quantities block`. The last line in the `transformed data` performs the rescaling of the data and is written in vectorised form, as this is the fastest way to perform the rescaling [64, 161].

In the `parameters` block, the (rescaled) parameters `mu_`, `tau_` and `sigma_` are declared, as well as an auxiliary variable called `eta`, which is used for an efficient implementation of the hierarchical model. Sampling `eta` and then using it is more efficient than directly trying to obtain the group means [92]. These group means are declared in the block `transformed parameters`.

The data for ethane are shown in table E12.1. Using traditional analysis of variance, for the dataset of ethane, the between-bottle homogeneity standard deviation is 3.19 μmol mol^{-1} and the (pooled) repeatability standard deviation is 10.88 μmol mol^{-1} [170].

Experience from previous between-bottle homogeneity studies for ethane in natural gas has indicated that the repeatability standard deviation for the amount fraction ethane is 0.10 % and

Table E12.1: Amount-of-substance fraction of ethane (%) of the 10 mixtures

	Replicate 1	Replicate 2	Replicate 3	Replicate 4	Replicate 5
D520472	3.496632	3.498528	3.495981	3.497959	3.495900
H95396	3.499540	3.496899	3.498857	3.497719	3.498985
VSL190663	3.499200	3.497277	3.496565	3.496385	3.499073
D520467	3.498073	3.496472	3.496522	3.496959	3.497474
D520834	3.499093	3.496120	3.496482	3.498150	3.497656
D520361	3.497502	3.498803	3.498913	3.499660	3.499122
D520270	3.497358	3.498859	3.497683	3.497349	3.498252
D520446	3.498206	3.497208	3.499195	3.498182	3.497208
VSL190485	3.498724	3.497021	3.496768	3.498014	3.495973
VSL190977	3.499762	3.498026	3.498264	3.495620	3.499327

the specification for the between-bottle homogeneity is 0.4 %. The latter is interpreted as an expanded uncertainty with coverage factor $k = 2$, thus $\tau_0 = \mu_0 \cdot 0.2$ %. Running the model with 25 000 iterations and a warmup of 5000 iterations, using 4 chains [170, 174] on the dataset of ethane yields the following output:

```
## Inference for Stan model: 3c7d78ea6604265a562e4008c783360f.
## 4 chains, each with iter=25000; warmup=5000; thin=1;
## post-warmup draws per chain=20000, total post-warmup draws=80000.
##
##              mean se_mean      sd    2.5%   97.5% n_eff    Rhat
## mu        3.49779       0 0.00022 3.49736 3.49823 60910 0.99997
## tau       0.00039       0 0.00026 0.00002 0.00099 25435 1.00010
## sig       0.00112       0 0.00012 0.00090 0.00139 52679 1.00002
## theta[1]  3.49751       0 0.00037 3.49668 3.49813 51074 1.00005
## theta[2]  3.49801       0 0.00035 3.49740 3.49880 58512 1.00002
## theta[3]  3.49776       0 0.00032 3.49710 3.49840 89856 1.00000
## theta[4]  3.49755       0 0.00036 3.49674 3.49816 54818 1.00004
## theta[5]  3.49769       0 0.00033 3.49699 3.49830 81013 1.00003
## theta[6]  3.49815       0 0.00040 3.49749 3.49905 42563 1.00003
## theta[7]  3.49783       0 0.00032 3.49719 3.49850 93936 0.99996
## theta[8]  3.49786       0 0.00032 3.49725 3.49855 85007 0.99999
## theta[9]  3.49762       0 0.00034 3.49686 3.49822 67517 0.99998
## theta[10] 3.49793       0 0.00033 3.49733 3.49867 74363 1.00001
##
## Samples were drawn using NUTS(diag_e) at Thu Aug 13 09:31:33 2020.
## 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).
```

The posterior probability density functions for μ, τ, and σ computed from the dataset of ethane are shown in figure E12.1. From the output of the MCMC, also the shortest coverage intervals can be computed. The following code, requiring the packages `coda` and `HDIntervals` performs the calculation:

```
fit.mcmc = As.mcmc.list(fit)
ethane.hpdi = hdi(fit.mcmc, credMass = 0.95)
```

The first line converts the output from `Stan` into the appropriate form [143]. Then the function `hdi` is used to compute the highest posterior density intervals [128]. The lower (L) and upper (H) limits of the 95 % highest posterior density intervals are shown in table E12.2. The lower bound on the coverage interval for τ is 1.13×10^{-10} cmol mol^{-1}, which is very close to zero.
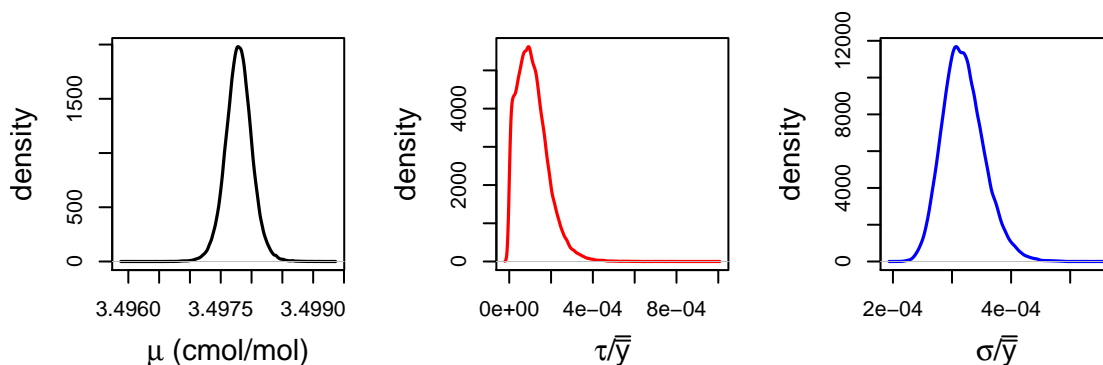
Figure E12.1: Posterior densities for the mean, between-group and within-group standard deviations for ethane

As the uncertainty evaluation in accordance with the GUM [51] and ISO Guide 35 [32] requires only a point estimate, there are several ways to obtain such an estimate from the posterior probability density function. Candidates include the mean, the mode, and the median. If the probability density function is symmetric and unimodal, these candidates will all have the same value.

Given the skewness of the posterior of τ (see figure E12.1), these three options are not equivalent. The mode is $3.22\,\mu\text{mol}\,\text{mol}^{-1}$, the median is $3.54\,\mu\text{mol}\,\text{mol}^{-1}$, and the mean is $3.89\,\mu\text{mol}\,\text{mol}^{-1}$.

Table E12.2: 95 % highest posterior density intervals for μ, τ and σ for ethane(expressed as amount fractions, %)

Parameter	L	H
μ	3.49737	3.49823
τ	0.00000	0.00086
σ	0.00089	0.00137

Table E12.3: Amount-of-substance fraction of nitrogen (%) in the 10 mixtures

	Replicate 1	Replicate 2	Replicate 3	Replicate 4	Replicate 5
D520472	0.424577	0.425167	0.425379	0.424522	0.424805
H95396	0.425572	0.425411	0.423638	0.425301	0.424527
VSL190663	0.424152	0.425517	0.425638	0.424207	0.425135
D520467	0.426320	0.424672	0.425211	0.425533	0.425864
D520834	0.424855	0.425079	0.425413	0.424729	0.424725
D520361	0.425104	0.424773	0.426424	0.424266	0.424632
D520270	0.425750	0.424917	0.424779	0.425086	0.425318
D520446	0.425547	0.426483	0.424631	0.425968	0.424620
VSL190485	0.426326	0.424646	0.425205	0.426302	0.425020
VSL190977	0.425968	0.424069	0.425988	0.425489	0.423936

The data for nitrogen are shown in table E12.3. Using traditional analysis of variance for the dataset of nitrogen, the between-bottle standard deviation is $0.00\,\mu\text{mol}\,\text{mol}^{-1}$ and the (pooled)
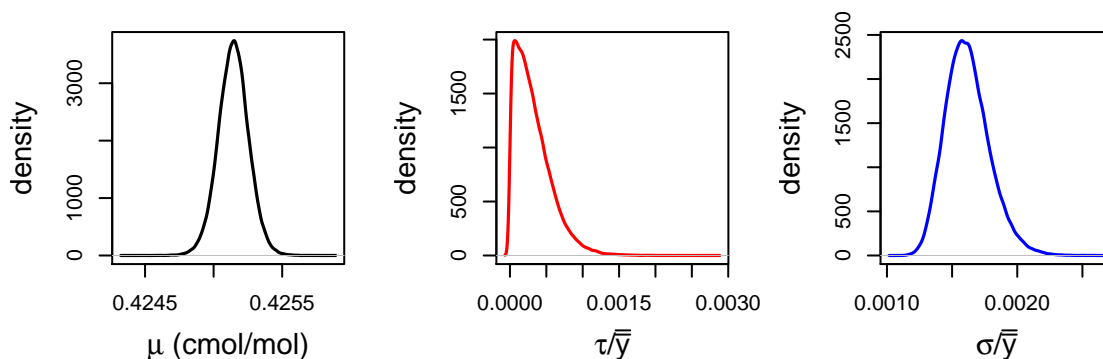
Figure E12.2: Posterior densities for the mean, between-group and within-group standard deviations for nitrogen

repeatability standard deviation is $7.00\,\mu\text{mol}\,\text{mol}^{-1}$ [170]. The zero value for the between-bottle standard deviation is readily explained by considering that $MS_{\text{between}} = 3.249 \times 10^{-7}$ is smaller than $MS_{\text{within}} = 4.903 \times 10^{-7}$.

Experience from previous between-bottle homogeneity studies for nitrogen in natural gas has the repeatability standard deviation for the amount fraction nitrogen is 0.20 % and the specification for the between-bottle homogeneity is 0.3 %. The latter is interpreted as an expanded uncertainty with coverage factor $k = 2$. Running the model with 25 000 iterations and a warmup of 5000 iterations, using 4 chains [170, 174] on the dataset of nitrogen yields the following output:

```
## Inference for Stan model: 3c7d78ea6604265a562e4008c783360f.
## 4 chains, each with iter=25000; warmup=5000; thin=1;
## post-warmup draws per chain=20000, total post-warmup draws=80000.
##
##              mean se_mean     sd    2.5%   97.5% n_eff    Rhat
## mu        0.42514       0 0.00011 0.42492 0.42537 77137 1.00002
## tau       0.00014       0 0.00011 0.00001 0.00040 49866 1.00004
## sig       0.00069       0 0.00007 0.00056 0.00085 85493 0.99999
## theta[1]  0.42510       0 0.00017 0.42473 0.42540 81567 1.00000
## theta[2]  0.42510       0 0.00016 0.42474 0.42540 80768 1.00002
## theta[3]  0.42511       0 0.00016 0.42475 0.42541 85237 1.00003
## theta[4]  0.42521       0 0.00017 0.42492 0.42561 78431 0.99997
## theta[5]  0.42511       0 0.00016 0.42476 0.42541 87385 1.00004
## theta[6]  0.42512       0 0.00016 0.42479 0.42543 89116 1.00002
## theta[7]  0.42515       0 0.00016 0.42484 0.42547 90335 1.00002
## theta[8]  0.42520       0 0.00017 0.42490 0.42558 79099 0.99997
## theta[9]  0.42521       0 0.00017 0.42491 0.42560 78602 0.99999
## theta[10] 0.42513       0 0.00016 0.42480 0.42544 90610 0.99999
##
## Samples were drawn using NUTS(diag_e) at Thu Aug 13 09:33:22 2020.
## 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).
```

The posterior probability density functions for μ, τ, and σ computed from the dataset of nitrogen are shown in figure E12.2. Where traditional analysis of variance fails at quantifying the between-bottle homogeneity effect, the Bayesian counterpart provides a probability density function for τ, from which the between-bottle standard deviation can be derived. The 95 % highest posterior density intervals are given in table E12.4. The lower (L) and upper (H) limits of the 95 % highest posterior density intervals are shown in table E12.4. The lower bound on the coverage interval for τ is 2.18×10^{-9} $\text{cmol}\,\text{mol}^{-1}$.

Table E12.4: 95 % highest posterior density intervals for μ, τ and σ for nitrogen(expressed as amount fractions, %)

Parameter	L	H
μ	0.42493	0.42537
τ	0.00000	0.00034
σ	0.00056	0.00083

Just as in the case of ethane, also for the amount fraction nitrogen there are different options for the between-bottle standard deviation τ. The mode of the posterior of τ is $0.27\,\mu\text{mol}\,\text{mol}^{-1}$, the median is $1.12\,\mu\text{mol}\,\text{mol}^{-1}$, and the mean is $1.36\,\mu\text{mol}\,\text{mol}^{-1}$.

E12.7   Reporting the result

The prime result is the value for the between-bottle standard deviation τ. In a previous paper [170], the mean was chosen as estimate for τ, which is the most cautious option (it leads to the largest value for this uncertainty contribution). For datasets where the between-bottle variability is larger, the differences between the three options become smaller. As also discussed in the cited paper, the width of the posterior makes that several alternatives [83, 169] also fall in the 95 % coverage interval.

Alternatives to using the mean include the use of the median or mode of the posterior probability density function for τ, and even $\tau = 0$ could be justified, as the lower ends of the 95 % highest posterior density intervals are for practical purposes indistinguishable from zero. The between-bottle homogeneity effect in both datasets is small (that is why they were selected for this example in the fist place).

Example E13

Quantifying uncertainty when comparing measurement methods – Haemoglobin concentration as an example of correlation in straight-line regression

S. Martens, K. Klauenberg, J. Neukammer, S. Cowen, S.L.R. Ellison, C. Elster

E13.1 Summary

In metrology, often two methods measuring the same quantity are to be judged whether or not they are in agreement. For measurements across a whole range of values, this can be done by comparing their straight-line fit to the identity line. Such a comparison is only meaningful, when uncertainties are available. Furthermore, the estimates of the straight-line fit and their uncertainties are only reliable when all sources of uncertainty have been accounted for. In particular, the measurements of both methods in a comparison are usually uncertain, and common instruments or standards cause correlation among or between them.

When fitting a straight-line relation, the weighted total least-squares (WTLS) method accounts for correlation and uncertainties in both variables. This example focuses on WTLS and defines a measurement model from it to propagate all uncertainties and correlations through to the estimate of the slope and intercept, and associate uncertainties with them according to the GUM. Using the example of two high accuracy methods measuring the total haemoglobin concentration in blood, i.e. the cyanmethaemoglobin and alkaline haematin method, we indicate how correlations can be inferred, demonstrate how they can be accounted for and show their impact on the regression. The results are discussed and recommendations are given.

E13.2 Introduction of the application

The total haemoglobin (Hb) concentration in blood is one of the most frequently measured analytes in clinical medicine because of its significance for evaluating the state of health of a human. The medical need for this analyte and the different spectrophotometric methods applied are summarized in Appendix E13.A. For external quality assurance of routine laboratories, interlaboratory comparisons are performed in which the deviation from the reference value may not

135

exceed 6 % [41]. To evaluate such round robin tests, ideally reference or "higher order" measurement procedures allowing for standard uncertainties smaller than 0.6 % (an order of magnitude below the allowable deviations) are required. The cyanmethaemoglobin (HiCN) method is the internationally accepted, spectrophotometric reference method [8, 14, 115] to determine the total Hb concentration. Critical issues of the HiCN method are the toxicity of the potassium cyanide involved and that it is not traceable to the International System of Units. An alternative spectrophotometric procedure for the determination of reference values for this quantity is the non-cyanide, alkaline haematin (AHD) method. Among other advantages, the AHD method has the potential as a primary method [9, 90] since a primary calibrator exists.

Previous comparisons of the HiCN and the AHD method with high-accuracy procedures [90, 181] demonstrate a good agreement, but are limited to only one blood sample with a Hb concentration in the normal range, i.e. a healthy person. Studies based on protocols for routine diagnostics[1] also show a good agreement between both methods (see [45, 131, 136] and references therein) and rely, among others, on the regression of a straight-line relationship. However, these comparisons do not consider the uncertainty of measurements. Estimates of regression parameters will usually differ when all uncertainties are accounted for. In addition, these comparison studies do not provide an uncertainty for the regression estimates. It is thus difficult to compare the results of these studies and to quantitatively judge the agreement between the reference and the alternative AHD method.

This example demonstrates how the uncertainties of HiCN and AHD measurements, including correlation, can be propagated to give the uncertainty of their straight-line relation. The total Hb concentrations are used, which PTB measured with both the HiCN and the AHD method for $P = 104$ blood samples over the past 10 years. The data cover the whole range from 60 g L^{-1} to 190 g L^{-1} relevant in clinical diagnosis and include pathologically low as well as pathologically high Hb concentrations. These measurements and their associated uncertainties, say x_p, $u(x_p)$ and y_p, $u(y_p)$, are displayed in figure E13.1 and can be found online in repository [121]. Derivation of the total Hb concentration involves quantities common to both methods and all samples (cf. Appendix E13.B for background information). Some of these common quantities contribute significantly to the uncertainty of the Hb concentration [90, 181]. Therefore, it is reasonable to suspect significant correlation among the HiCN as well as among the AHD method (cf. clause 5.2.4 in [51]).

Also beyond method comparison, uncertainty in all variables of a regression and correlation among or between them is prevalent in metrology. For example in calibrations, the reference and the device under test usually both display uncertainty. Additionally, measurements over the range of use are often performed with the same measuring instrument or physical standard which often contribute a considerable amount of uncertainty.

This example focusses on a measurement model that is based on the weighted total least-squares (WTLS) method. The measurement model allows for uncertainty evaluation following the GUM. The WTLS method accounts for uncertainties in both variables of a regression, as well as, for correlation among and between them. WTLS is recommended by multiple standards [16, 38] and applied in metrology (e.g. Refs. [89, 107, 118]).

---

[1]In routine applications only one value for the absorbance is measured, while reference procedures include dilution series, repeat measurements and centrifugation to reduce uncertainties.
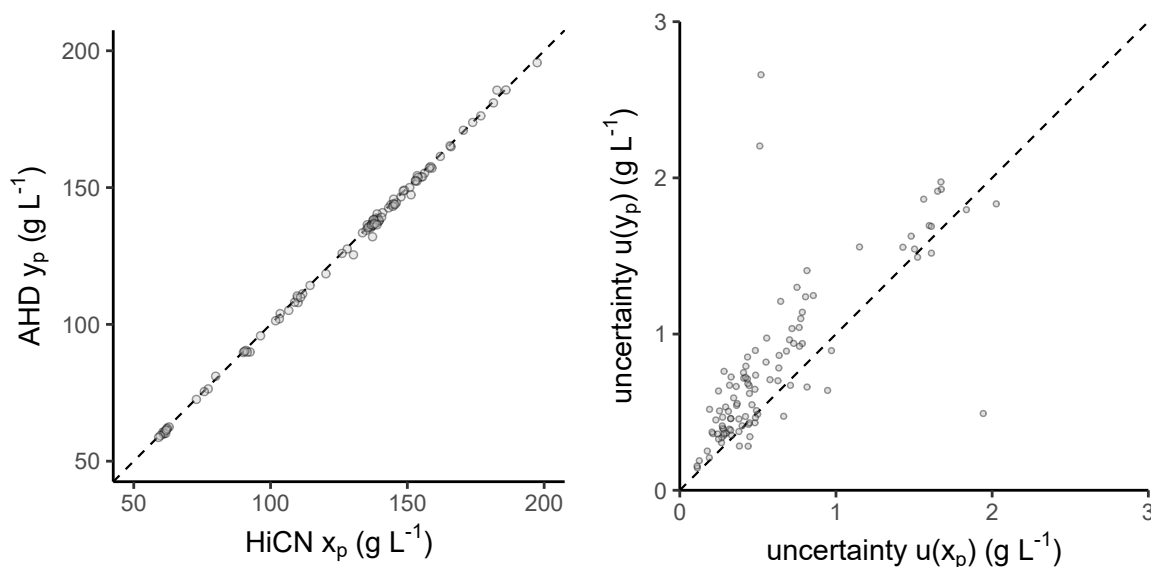
Figure E13.1: Left: Visualization of haemoglobin concentration measurements x_p, y_p performed
at PTB on $P = 104$ blood samples by the two methods HiCN and AHD.
Right: Standard uncertainties $u(x_p)$ and $u(y_p)$ for both methods and all samples. These measurement results are available online in repository [121]. In both panels, the dashed line represents
the identity $y = x$ and the markers are drawn as transparent; thus, overlayed markers appear
darker.

E13.3   Specification of the measurand

Let X denote the total Hb concentrations obtained by HiCN and Y the corresponding quantity
measured by the AHD method. The straight-line relation

$$Y = \beta_0 + \beta_1 X \tag{E13.1}$$

is assumed to model the relationship between the measured values of both methods, and is
supported by previous studies comparing the HiCN and the AHD method (see [45, 131, 136] and
references therein). The measurands are the intercept parameter β_0 and slope parameter β_1 of
the straight-line model (E13.1). If both methods measure the same, uniquely defined quantity,
one usually obtains estimates close to $\widehat{\beta}_0 = 0$ and $\widehat{\beta}_1 = 1$.

The input quantities influencing the measurands are the P pairs (X_p, Y_p). Estimates of these
inputs are the Hb concentration measurements of each method, x_p and y_p. Standard uncertainties $u(x_p)$ and $u(y_p)$ of these inputs are of the same magnitude (cf. figure E13.1). In addition,
any two inputs X_p, X_q are correlated due to the use of common standards in their measurement,
especially of the same molar extinction coefficient ϵ and corrections C_0, C_1 (as detailed in Appendix E13.B). The covariance matrix $\boldsymbol{U}_x$ shall contain these correlations as well as the standard
uncertainties $u(x_p)$. Likewise, the covariance matrix $\boldsymbol{U}_y$ contains the correlations and standard
uncertainties $u(y_p)$ for the inputs Y_p. (For the definition of a covariance matrix, we refer to clause
3.11 in the supplement 1 to the GUM [52].)

E13.4   Measurement model

The measurement model for straight-line regression can be constructed from the appropriate
least-squares method. The frequently applied ordinary and weighted least-squares method are

inappropriate here because they assume that the measured values of one method are exact. Notably, regressing one method over the other will generally result in different estimates than the other way around; especially, when the uncertainties of both methods are similar and non-negligible – as for HiCN and AHD. The measurand would thus be ambiguous. Also Deming regression [78] and Passing-Bablok regression [138], two common methods for method comparison, are not appropriate for this data set. First, the uncertainties $u(y_p)$ cannot be expressed as a common multiple of $u(x_p)$ as Deming regression requires (see right panel in figure E13.1); second, it is important to take account of applicable uncertainty and covariance information where possible and Passing-Bablok regression does not use information on uncertainties.

Weighted total least-squares (WTLS) is the method recommended by multiple standards [16, 38] when the uncertainty associated with the measured values x_p and y_p are both non-negligible. It also addresses correlation. The WTLS method is based on minimizing the generalized sum of squares

$$Q = \left(\boldsymbol{x} - \widetilde{\boldsymbol{\xi}}\right)^\top \boldsymbol{U}_x^{-1}\left(\boldsymbol{x} - \widetilde{\boldsymbol{\xi}}\right) + \left(\boldsymbol{y} - (\widetilde{\beta}_0 + \widetilde{\beta}_1\widetilde{\boldsymbol{\xi}})\right)^\top \boldsymbol{U}_y^{-1}\left(\boldsymbol{y} - (\widetilde{\beta}_0 + \widetilde{\beta}_1\widetilde{\boldsymbol{\xi}})\right) \tag{E13.2}$$

with respect to $\widetilde{\beta}_0$, $\widetilde{\beta}_1$ and the unknown, "true" values of $\boldsymbol{x}$ called $\widetilde{\boldsymbol{\xi}}$. Here, the vector $\boldsymbol{x}$ contains the elements $\boldsymbol{x} = (x_1, \ldots, x_P)^\top$ and the vectors $\boldsymbol{y}$ and $\widetilde{\boldsymbol{\xi}}$ are likewise defined. The minimizer of (E13.2) defines the solution $(\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\boldsymbol{\xi}}^\top)$ of the WTLS method.

The measurement model is then defined by replacing the estimates $\boldsymbol{x}$ and $\boldsymbol{y}$ in the minimization of Q by the underlying quantities $\boldsymbol{X} = (X_1, \ldots, X_P)^\top$ and $\boldsymbol{Y} = (Y_1, \ldots, Y_P)^\top$, respectively. That is,

$$(\beta_0, \beta_1, \boldsymbol{\xi}^\top)^\top = \arg\min_{\widetilde{\beta}_0, \widetilde{\beta}_1, \widetilde{\boldsymbol{\xi}}} \left\{ \left(\boldsymbol{X} - \widetilde{\boldsymbol{\xi}}\right)^\top \boldsymbol{U}_x^{-1}\left(\boldsymbol{X} - \widetilde{\boldsymbol{\xi}}\right) + \left(\boldsymbol{Y} - (\widetilde{\beta}_0 + \widetilde{\beta}_1\widetilde{\boldsymbol{\xi}})\right)^\top \boldsymbol{U}_y^{-1}\left(\boldsymbol{Y} - (\widetilde{\beta}_0 + \widetilde{\beta}_1\widetilde{\boldsymbol{\xi}})\right) \right\},$$
$$\tag{E13.3}$$

where only (β_0, β_1) define the measurand.

E13.5   Estimation and uncertainty evaluation

Following the GUM [51,54], estimates $\widehat{\beta}_0$ and $\widehat{\beta}_1$ of the measurands are obtained by evaluating measurement model (E13.3) at the estimates $\boldsymbol{x}$ and $\boldsymbol{y}$ of the input quantities $\boldsymbol{X}$ and $\boldsymbol{Y}$. The uncertainties associated with $(\widehat{\beta}_0, \widehat{\beta}_1)$ result from propagating the uncertainties in $\boldsymbol{U}_x$ and $\boldsymbol{U}_y$ associated with the estimates of the input quantities through this measurement model.

Measurement model (E13.3) is implicit, multivariate, non-linear and usually no closed form is available for its solution. An iterative scheme for deriving estimates and their associated uncertainties is described in clause 10 of the standard [16]. This simple scheme also provides correlations between β_0 and β_1, and is valid for any covariance matrices $\boldsymbol{U}_x$ and $\boldsymbol{U}_y$ whose eigenvalues are all positive.

Assuming a Gaussian distribution[2], a 95 % coverage interval for each measurand β_i with $i = 0, 1$ is given by

$$\left[\widehat{\beta}_i - 1.96\, u\left(\widehat{\beta}_i\right), \widehat{\beta}_i + 1.96\, u\left(\widehat{\beta}_i\right)\right].$$

A two-dimensional, joint 95 % coverage region can be calculated following clause 6.5.2 in [54].

In order to estimate the slope and intercept of a straight-line relation as well as valid uncertainties and/or coverage intervals, the full covariance matrices $\boldsymbol{U}_x, \boldsymbol{U}_y$ and possible cross-correlation between $\boldsymbol{X}$ and $\boldsymbol{Y}$ need to be known. Annex D in [16] describes how these covariances

---

[2]Cf. section 10.2.3 in [16] for the approximate validity of this Normality assumption.

can be calculated for common, simple measurement models. For more involved measurement
models, like for HiCN and AHD measurements, we recommend the Monte Carlo method [54],
where distributions for all input quantities are propagated through a joint measurement model
to arrive at the $2P$-dimensional, joint distribution for the outputs X and Y.

The uncertainty in HiCN and in AHD measurements is dominated by a common quantity,
namely the molar extinction coefficient ϵ (see [90, 181]). We thus suspect that the covariance
matrices U_x and U_y are governed by a common correlation coefficient ρ. That is, we set their
elements $U_{x,pq} = \rho u(x_p)u(x_q)$ and $U_{y,pq} = \rho u(y_p)u(y_q)$ for all $p \neq q$. The diagonal elements
contain the variances, i.e. $U_{x,pp} = u^2(x_p)$ and $U_{y,pp} = u^2(y_p)$. Further details are given in Ap-
pendix E13.C. First Monte Carlo evaluations of the joint uncertainty budget showed that correla-
tion coefficients up to $\rho = 0.8$ may be realistic. Details on how to jointly evaluate the correlation,
uncertainties and estimates for the input quantities of least-squares methods applying the Monte
Carlo method are illustrated in [120]. The correlation between HiCN and AHD is dominated
by two common quantities, viz., the cuvettes' absorption length d and the mean molar mass
M(Hb). According to [90], the amount of cross-correlation is much smaller compared to corre-
lation between the estimates x_p and x_q as well as between the estimates y_p and y_q. We assume
zero cross-correlation throughout this example. Note that the results reported below are condi-
tional on the plausibility of this correlation structure. The real correlation structure and amount
could be different and is to be inferred from the quite complex measurement model described in
Appendix E13.B.

E13.6   Reporting the result

Let us now apply the measurement model (E13.3) to the estimates and uncertainties presented
in figure E13.1 and to the above covariance structures U_x and U_y. For selected correlation coeffi-
cients ρ, the results are listed in table E13.1. The estimate, associated standard uncertainty and
the covariance for the measurands β_0 and β_1 are obtained by the algorithm in clause 10 of [16]
and application of the law of propagation of uncertainty [51]. R Markdown [43] code for this
algorithm is available online in repository [121]. Figure E13.2 depicts the estimates $\widehat{\beta}_0$ and $\widehat{\beta}_1$
and the corresponding 95 % coverage interval.

Table E13.1: Results obtained by weighted total least-squares with uncertainty evaluation accord-
ing to the GUM for varying correlation coefficients ρ. Listed are the estimates and uncertainties
for slope and intercept.

Correlation a.u.	$\widehat{\beta}_0$ $\mathrm{g\,L^{-1}}$	$u(\widehat{\beta}_0)$ $\mathrm{g\,L^{-1}}$	$\widehat{\beta}_1$ a.u.	$u(\widehat{\beta}_1)$ a.u.	$\mathrm{cov}(\widehat{\beta}_0,\widehat{\beta}_1)$ 10^{-3} a.u.
$\rho = 0.0$	-0.488 6	0.166 7	0.998 4	0.001 6	-0.24
$\rho = 0.6$	-0.489 4	0.105 5	0.998 6	0.001 2	-0.10
$\rho = 0.8$	-0.489 4	0.074 6	0.998 6	0.000 9	-0.05

Nearly identical results have been obtained by applying the Monte Carlo method [54] to the
measurement model (E13.3) and the algorithm in Ref. [16]. The non-linearity of (E13.3) could
cause differences; however, this was not observed. Software is available that implements WTLS
and propagates uncertainties. For example, the CALIBRATION CURVE COMPUTING Software
provided by INRIM [118] also produces the results in table E13.1, although a slightly different
algorithm is implemented (which relies on an implicit set of normal equations).
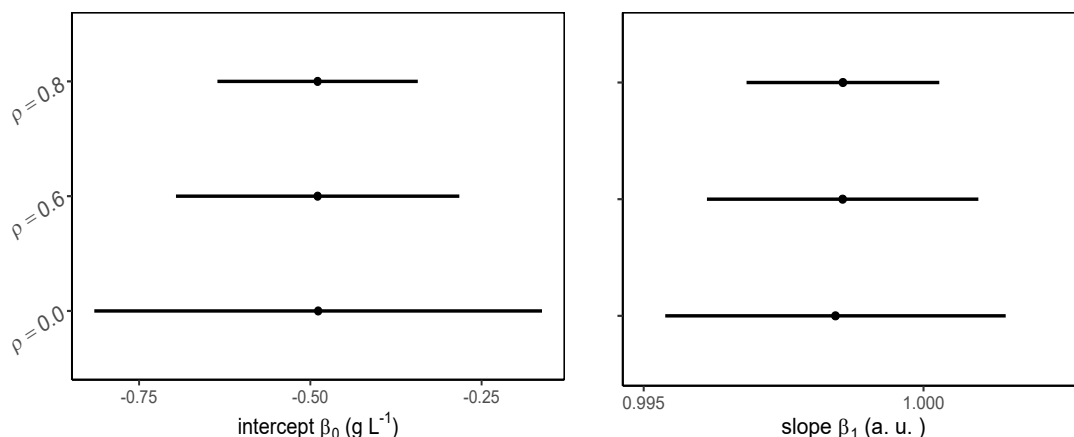
Figure E13.2: Displayed are the estimates $\widehat{\beta}_0$ and $\widehat{\beta}_1$ (dots) and their 95 % coverage intervals for the weighted total least-squares regression results listed in table E13.1.

Before interpreting the results of a regression, the data as well as the assumptions contributing to the analysis should be assessed critically. For instance, graphically analysing the (weighted) residuals did not indicate a violation of the straight-line assumption (E13.1), since no systematic behaviour of these residuals were observed. A significant outcome of the χ^2 test, whose application is recommended in standard [16], does not necessarily indicate departures from the linearity assumption. The χ^2 test assesses, whether the (weighted) residuals are independently normally distributed – an assumption which is not required for WTLS estimation and measurement model (E13.3). Any observed test statistics which exceed the 95 % quantile of the χ^2-distribution, are suspected to be due to non-normally distributed residuals rather than a violation of the straight-line assumption (E13.1). The former does not contradict the assumptions of our analysis.

E13.7   Discussion and conclusion

This example demonstrates how two measurement methods can be compared to judge whether both measure the same quantity over a defined measurement range. If the uncertainties of both methods are non-negligible, ordinary and weighted least-squares methods are inappropriate. Instead, weighted total least-squares is a suitable method which allows for an uncertainty propagation when embedded in a measurement model in line with the GUM.

Using the example of measuring the total haemoglobin concentration in blood, it is reasoned that correlation among and possibly between two measurement methods is not unusual, and likely to be rather frequent in metrology in general. We indicate how these correlations can be inferred and select a common correlation structure for this example.

The reader observes a small but significant offset between the HiCN and AHD method for measuring haemoglobin – irrespective of the amount of correlation. The slope of the linear relation between both methods is compatible with unity for all reasonable values of correlation, but would be significantly smaller than one for higher correlations $\rho \geq 0.9$. For the assumed correlation structure, the estimates of the linear relation vary little with the amount of correlation. However, their uncertainty changes by the factor $\sqrt{1-\rho}$, i.e. it reduces to two thirds for a correlation coefficient of $\rho = 0.6$ and to a half for $\rho = 0.8$, compared to WTLS estimation without correlation. Also the covariance between $\widehat{\beta}_0$ and $\widehat{\beta}_1$ scales with $1-\rho$. In addition, the estimates change with varying correlation coefficient when fewer observations are available. These rela-

tionships are detailed in Appendix E13.C. Other correlation structures, for instance when the correlation within one method is much larger than within the other method, will also change the estimates.

Our analyses show that the HiCN method leads to slightly higher Hb concentrations than the AHD method, if the correlation structure and the amount of correlation are realistic. This has been observed before ([182] and references therein) and may be caused by a background due to bilirubin. However, the differences between the HiCN and the AHD method are sufficiently small. If the correlation assumptions can be confirmed in future, both methods could be applied to determine higher order measurement values to evaluate round robin tests for external quality assurance in laboratory medicine.

We conclude that only stating the uncertainty of a fitted (linear) relation allows for a quantitative comparison of two methods over their measurement range. To derive these uncertainties reliably and to give valid estimates, it is important to account for correlation among and between the measurement methods. Otherwise, the conclusions drawn from such a comparison study could differ and become unreliable.

E13.A   Haemoglobin concentration: Importance and determination

The total haemoglobin (Hb) concentration in blood is part of the complete blood count, which is one of the most frequently measured analytes in clinical medicine. For example, Hb concentrations are needed for screening blood donors to protect their health and to guarantee the quality of the blood product [65]. Deviations of the Hb concentration from the normal range ($137 \, g \, L^{-1}$ – $162 \, g \, L^{-1}$ for men and $123 \, g \, L^{-1}$ – $145 \, g \, L^{-1}$ for women; c.f. [67, table 4, p. 190]) are observed for various diseases. Further diagnostics are initiated to identify the origin of such an anomaly. Iron deficiency could be caused by bleeding in the gastrointestinal tract [85], malaria [180] or thalassemia, the most common genetic disorder worldwide [167]. In addition, haemoglobin concentration is relevant to manage iron deficiency in pregnant women [98].

Total haemoglobin concentration is determined by a variety of methods [160], depending on the specific medical application. In countries where anaemia is widespread, portable instruments are used to estimate haemoglobin concentration using capillary blood for analysis [75]. Measurements with higher precision and accuracy compared to such point-of-care instruments are routinely performed in laboratory medicine and require venous blood and chemical conversion of the different haemoglobin variants to a stable end product, which is subsequently spectrophotometrically analysed. Conversion to cyanmethaemoglobin (HiCN), first applied by Drabkin and Austin [80], has been considered as a gold standard for routine applications [160] and is also internationally accepted as higher-order method [8, 14, 115] to determine reference measurement values in external quality assurance of medical laboratories [41]. However, because of the toxicity of the potassium cyanide involved, the HiCN method is not allowed in most countries and has been replaced by the sodium lauryl sulfate (SLS) procedure [101, 135].

Typically, in laboratory medicine accuracies below 6 % shall be reached for Hb concentration measurements. This value is stated in the guideline of the German Medical Association for Quality Assurance in Medical Laboratory Examinations [41] and indicates the maximum allowable deviation to pass the ring trials mandatory in Germany. To evaluate such external quality assurance schemes, so-called "higher-order measurement methods" or reference procedures are required providing results with expanded uncertainties (95 % confidence level) possibly smaller than 1.5 %. This requirement is specified in DIN 58931 [14, p. 18] and was met in comparison experiments [90, 181]. For such higher-order procedures the same reagents may be used to convert the different Hb variants to a stable end product. Lower uncertainties are achieved by gravimetrical preparation of dilution series and centrifugation to suppress the scattering of residual white

blood cells or agglomerates of membranes of erythrocytes. In addition, high-accuracy absorbance measurements are required, traceable to a national standard [14]. Although the HiCN method is frequently used as a reference method for comparison when evaluating new procedures for the determination of the total Hb concentration, it is presently not traceable to the International System of Units. In particular, material suited as primary calibrator is not available, it is known that verdoglobin is not converted to HiCN and that background due to bilirubin can cause systematic deviations towards higher concentrations. It follows that according to the ISO standard 17511 on metrological traceability [9] the HiCN method can be characterised as an international conventional reference measurement procedure. An alternative spectrophotometric procedure for the determination of total Hb concentration is the non-cyanide, alkaline haematin (AHD) method. In contrast to the HiCN procedure, when applying the AHD method verdoglobin is converted to the end product chlorohaemin and the sensitivity against bilirubin perturbations is much smaller. In addition, the globin protein is destructed and solutions of the end product, the well-defined molecule chlorohaemin, might serve as primary calibrator. Hence, the AHD method may have the potential as a primary method [9, 90].

E13.B   Details of the measurement methods for haemoglobin concentration

The HiCN and the AHD method both rely on the measurement of the spectral absorbance. The photometrical traceability is established by correcting the measured absorbance values[3] $a_{i,p}^k$ using the linear relationship $C_{0,k} + C_{1,k} a_{i,p}^k$ for $k \in \{\text{HiCN}, \text{AHD}\}$, blood sample $p = 1, \dots, P$ and dilution i (cf. [181]). As recommended in DIN 58931:2010 [14], at least four dilutions ϕ_i of each blood sample are prepared and the associated Hb mass fractions $w_{i,p}^k$ are calculated according to

$$w_{i,p}^k = \frac{\left(C_0^k + C_1^k a_{i,p}^k \right) M(\text{Hb})}{d \epsilon^k \phi_i}. \tag{E13.4}$$

Here, d represents the absorption length of the rectangular spectrophotometric cuvette, ϵ^k is the molar decadic absorption coefficient of the reaction product and $M(\text{Hb})$ is the mean molar mass of one Hb subunit. The estimates and associated uncertainties for the input quantities in (E13.4) can be found in [181]. The final reported total Hb concentration for each sample and method, x_p and y_p, are determined by a weighted average of the Hb mass fractions $w_{i,p}^k$ over the dilutions i. The associated uncertainties are discussed in detail in Ref. [181].

E13.C   Influence of correlation for a common structure

If the covariance matrix for the HiCN method is given by

$$\boldsymbol{U}_x = (1 - \rho)\text{diag}(\boldsymbol{u}_x^2) + \rho \boldsymbol{u}_x \boldsymbol{u}_x^\top$$

with $\boldsymbol{u}_x = (u(x_1), \dots, u(x_P))^\top$, as described in section E13.5, the inverse of $\boldsymbol{U}_x$ is determined by

$$\boldsymbol{U}_x^{-1} = \frac{1}{1 - \rho} \left[\text{diag}\left(\frac{1}{u^2(x_1)}, \dots, \frac{1}{u^2(x_P)} \right) - \frac{1}{P - 1 + 1/\rho} \left(\frac{1}{\boldsymbol{u}_x} \right) \left(\frac{1}{\boldsymbol{u}_x} \right)^T \right].$$

---

[3]Each absorbance value $a_{i,p}^k$ in turn is based on a series of repeated measurements and its uncertainty is evaluated following the GUM.

The inverse U_y^{-1} can be determined by analogy. Then, the generalized sum of squares (E13.2)
simplifies to

$$
Q = \frac{1}{1-\rho} \left(\sum_{p=1}^{P} \frac{(x_p - \xi_p)^2}{u^2(x_p)} + \frac{(y_p - \beta_0 - \beta_1 \xi_p)^2}{u^2(y_p)} \right.
$$

$$
\left. -\frac{1}{P-1+1/\rho} \left(\sum_{p,q=1}^{P} \frac{(x_p - \xi_p)(x_q - \xi_q)}{u(x_p)u(x_q)} + \sum_{p,q=1}^{P} \frac{(y_p - \beta_0 - \beta_1 \xi_p)(y_q - \beta_0 - \beta_1 \xi_q)}{u(y_p)u(y_q)} \right) \right).
$$

The factor $1/(1-\rho)$ is irrelevant for the optimization of Q and thus does not influence the
estimates $\widehat{\beta}_0$ and $\widehat{\beta}_1$. At the same time it influences the uncertainties $u(\widehat{\beta}_i)$ and the covariance
$\mathrm{cov}(\widehat{\beta}_0,\widehat{\beta}_1)$ when the number of observations P is large, which change approximately by the
factor $\sqrt{1-\rho}$ and $1-\rho$, respectively, compared to a correlation coefficient of $\rho = 0$.

For a small number of observations, table E13.2 shows the influence of the correlation coeffi-
cient on the estimates (assuming the same, above correlation structure). In particular, the table
lists for a subset of size $P = 20$ of the data in figure E13.1 the estimates and uncertainties for $\widehat{\beta}_0$,
$\widehat{\beta}_1$ and for $\rho = 0$, $\rho = 0.8$. The reader observes, that compared to no correlation, the estimate
for the slope changes by almost half of the uncertainty (i.e. $\widehat{\beta}_1^{\mathrm{corr}} - \widehat{\beta}_1 \approx u(\widehat{\beta}_1)/2$) and at the same
time the uncertainty reduces considerably.

Table E13.2: Results obtained by weighted total least-squares with uncertainty evaluation accord-
ing to the GUM for a subset of size $P = 20$ of the data in figure E13.1. Listed are the estimates
and uncertainties for slope and intercept.

Correlation a.u.	$\widehat{\beta}_0$ $\mathrm{g\,L^{-1}}$	$u(\widehat{\beta}_0)$ $\mathrm{g\,L^{-1}}$	$\widehat{\beta}_1$ a.u.	$u(\widehat{\beta}_1)$ a.u.
$\rho = 0.0$	0.153 3	0.458 8	0.994 5	0.004 2
$\rho = 0.8$	0.176 1	0.206 1	0.996 5	0.002 4

Example E14

Calibration of a torque measuring system – GUM uncertainty evaluation for least-squares versus Bayesian inference

S. Martens, K. Klauenberg, C. Elster

E14.1 Summary

This example addresses the straight-line calibration of a torque measuring sensor against a reference system using measurements taken at different torque values. For each torque value, a single measurement result of the reference system is available, together with results of repeated measurements of sensor that shall be calibrated. The goal is to determine a linear relationship that relates results of the torque measuring sensor with those of the reference system. The data are analysed by applying (i) ordinary and weighted least-squares estimation in combination with an uncertainty evaluation following the GUM and (ii) Bayesian inference. Analytic expressions are given for the Bayesian uncertainty analysis which simplifies its application. The results obtained by the different approaches are discussed and recommendations given.

E14.2 Introduction of the application

Straight-line calibration of a torque measuring sensor, which are made of a strain gauges, against a reference system is addressed. The data are partly taken from example B2 of the guideline VDI/VDE 2600 part 2 [40]. The goal of the calibration is to determine a functional relationship between results obtained by the sensor and those of a reference system. Measurements have been carried out at different values of torque by the considered sensor and the reference system. For each torque value, a single measurement has been conducted by the reference system and a number of repeated measurements by the considered sensor. Table E14.1 shows the data, where repeated measurement results of the sensor have been summarized through their means and standard deviations, respectively. For further details about the measurements the reader is referred to [40].

This example provides guidance for the evaluation of uncertainty in the estimation of a calibration curve from data like those in table E14.1. Two approaches are provided: (i) ordinary and weighted least-squares estimation (see, e.g., ISO 28037 [16]) accompanied with an uncertainty

i	x_i N m	Mean y_i N m	SD S_i N m	n_i
1	0.101	0.095 0	0.005 5	6
2	0.201	0.196 6	0.005 2	6
3	0.305	0.301 6	0.004 1	6
4	0.501	0.498 3	0.004 1	6
5	1.001	1.008 3	0.009 8	6
6	3.000	3.026 6	0.008 2	6
7	4.001	4.046 6	0.012 1	6
8	5.007	5.066 6	0.037 9	3

Table E14.1: Summary statistics for part[1] of the measurement data given in the guideline [40][table B6, p. 43]. The summary statistics include mean $y_i = n_i^{-1} \sum_{j=1}^{n_i} y_{ij}$ and standard deviations (SD) $S_i = \left((n_i - 1)^{-1} \sum_{j=1}^{n_i} (y_{ij} - y_i)^2 \right)^{1/2}$ of the n_i measurement results y_{ij} of the considered sensor at the i-th torque level; x_i denotes the corresponding measurement result of the reference system. These summary statistics are available online in repository [121].
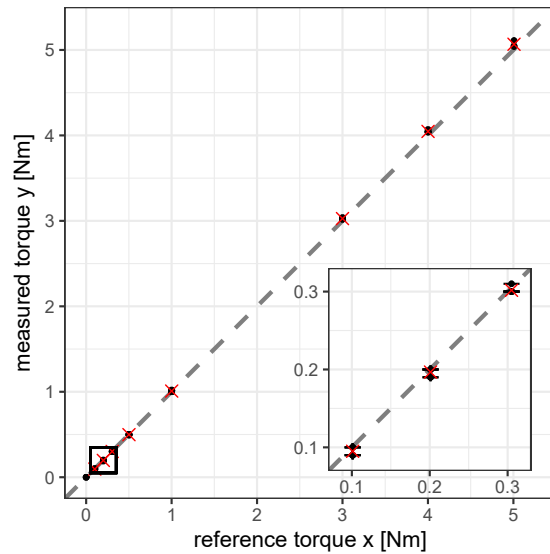
Figure E14.1: Visualization of the data. Repeated measurements (dots) are indistinguishable from the mean values (crosses) in this presentation. The dashed line represents the identity $y = x$.

evaluation based on the GUM [51], and (ii) a statistical approach applying Bayesian inference (cf., for example, [92]). Explicit expressions are given for the Bayesian uncertainty analysis which simplifies its application.

E14.3 Specification of the measurand

Let X denote the applied torque, in what follows called stimulus, and Y the corresponding quantity measured by the considered sensor, below denoted as response. The linear relation

$$Y = \beta X \tag{E14.1}$$

is assumed to model the relationship between the measured responses of the considered sensor and the applied stimulus. Model (E14.1) represents a straight line with zero offset. The latter has been chosen for physical reasons which is supported by the observed data (cf. figure E14.1). The measurand is the slope parameter β of the particular straight line model (E14.1). The input quantities are $Y_1, \ldots, Y_p$ (with $p = 8$ in our example) which correspond to the measurement results at the considered torques $X_1, \ldots, X_p$. The variability associated with the measurement results $x_1, \ldots, x_p$ of the reference device are considered as small enough so that they can be neglected.

---

[1]The original analysis includes data points $(0,0)$ which support the assumed relationship (E14.1). To prevent double counting information, we omit the data point $(0,0)$ in the our consideration.

E14.4 Measurement model

The uncertainty evaluations presented in this example are based on two different models. The uncertainty evaluation following the GUM in connection with ordinary and weighted least-squares estimation is based on a measurement model in which the measurand is represented as a function of the input quantities. An estimate of the measurand is then obtained by evaluating this measurement model using the estimates of the input quantities. The uncertainty associated with the resulting estimate for the measurand results from a propagation of the uncertainties associated with the estimates of the input quantities through this measurement model.

The Bayesian inference is based on a statistical model for the observed data, and the measurand enters as one of the parameters of the statistical model. The Bayesian inference can account for prior knowledge about the measurand. It results in a probability distribution for the measurand which can be viewed as the final complete result. Mean and standard deviation of that distribution can be taken as an estimate and standard uncertainty for the measurand. Bayesian uncertainty analysis can be viewed as being reached through the Bayesian inference, rather than by a process of propagating input uncertainties through a measurement model in the sense of the GUM.

E14.4.1 Ordinary and weighted least-squares

Application of weighted least-squares estimation determines an estimate $\widehat{\beta}$ for the measurand by minimizing

$$Q = \sum_{i=1}^{p} \sum_{j=1}^{n_i} W_i \left(y_{ij} - \beta x_i \right)^2 = \sum_{i=1}^{p} W_i \{ n_i \left(y_i - \beta x_i \right)^2 + (n_i - 1) S_i^2 \} \tag{E14.2}$$

with respect to β. In (E14.2), y_i and S_i represent mean and standard deviation of the repeated measurement results $y_{ij}, j = 1, \ldots, n_i$, of the considered sensor at the i-th stimulus x_i, and W_i denote some weights, $i = 1, \ldots, p$. The solution $\widehat{\beta}$ to this minimization problem is given by

$$\widehat{\beta} = \left(\sum_{i=1}^{p} n_i W_i x_i^2 \right)^{-1} \sum_{i=1}^{p} n_i W_i x_i y_i. \tag{E14.3}$$

The measurement model will now be *defined* as

$$\beta = \left(\sum_{i=1}^{p} n_i W_i x_i^2 \right)^{-1} \sum_{i=1}^{p} n_i W_i x_i Y_i, \tag{E14.4}$$

i.e., by replacing the estimates y_i in (E14.3) with corresponding quantities Y_i. Note that a measurement model in the sense of the GUM is always a model between quantities. Since the estimates x_i of the stimulus are treated as being exact, the actual quantity X_i in (E14.4) has already been replaced with the known values x_i in this example. Ordinary least-squares estimation is obtained by choosing weights $W_i = 1, i = 1, \ldots, p$.

E14.4.2 Statistical model

A statistical model specifies the distribution from which the observed data is taken as a realization. The subsequent statistical model assumes that all single measurements $y_{ij}, j = 1, \ldots, n_i, i = 1, \ldots, p$, are realizations of independently and normally distributed random variables Y_{ij} with means equal βx_i and variances σ_i^2, i.e.

$$Y_{ij} | \beta, \sigma_i^2 \sim N \left(\beta x_i, \sigma_i^2 \right). \tag{E14.5}$$

In (E14.5), the x_i denote the known stimuli and the σ_i^2 the unknown variances to be inferred.

The likelihood function is the probability for the observed data viewed as a function of the unknown parameters. For the statistical model (E14.5) the likelihood function is given by

$$l(\beta, \boldsymbol{\sigma}^2; \text{data}) \propto \prod_{i=1}^{p} (\sigma_i^2)^{-n_i/2} \exp\left(-\frac{1}{2\sigma_i^2}\left((n_i-1)S_i^2 + n_i(y_i - \beta x_i)^2\right)\right), \tag{E14.6}$$

where $\boldsymbol{\sigma}^2 = (\sigma_1^2, \ldots, \sigma_p^2)^\top$, y_i and S_i denote mean and standard deviation of $y_{ij}, j = 1, \ldots, n_i$, and "data" summarizes the sufficient statistics $y_1, \ldots, y_p, S_1, \ldots, S_p$ of the data, see table E14.1.

E14.5 Estimation and uncertainty evaluation

E14.5.1 GUM uncertainty propagation

The measurement model in (E14.4) contains input quantities $Y_i, i = 1, \ldots, p$. For each of these input quantities a series of repeated measurement results $y_{ij}, j = 1, \ldots, n_i$ is available. In following the GUM, mean and scaled standard deviation $S_i/\sqrt{n_i}$ are taken as estimate y_i and associated standard uncertainty $u(y_i)$ for Y_i. The estimates and standard uncertainties for the input quantities are listed in table E14.2.

According to the GUM [51], the estimate $\widehat{\beta}$ for β is taken as value of the measurement model (E14.4) when inserting the estimates from table E14.2 for the input quantities, i.e.

$$\widehat{\beta} = \left(\sum_{i=1}^{p} n_i W_i x_i^2\right)^{-1} \sum_{i=1}^{p} n_i W_i x_i y_i. \tag{E14.7}$$

The (squared) standard uncertainty is obtained by

$$u^2\left(\widehat{\beta}\right) = \sum_{i=1}^{p}\left(\left.\frac{\partial \beta}{\partial Y_i}\right|_{Y_i = y_i}\right)^2 u^2(y_i) = \frac{\sum_{i=1}^{p}(n_i W_i x_i)^2 u^2(y_i)}{\left(\sum_{i=1}^{p} n_i W_i x_i^2\right)^2}. $$

If the weights are chosen according to $n_i W_i = 1/u^2(y_i)$, one obtains

$$u^2\left(\widehat{\beta}\right) = \left(\sum_{i=1}^{p}\frac{x_i^2}{u(y_i)^2}\right)^{-1}, \tag{E14.8}$$

and for ordinary least-squares estimation with $W_i = 1$

$$u^2(\widehat{\beta}) = \frac{\sum_{i=1}^{p}(n_i x_i)^2 u^2(y_i)}{\left(\sum_{i=1}^{p} n_i x_i^2\right)^2}. \tag{E14.9}$$

Assuming a Gaussian distribution for β, a 95% coverage interval is given by

$$\left[\widehat{\beta} - 1.96u\left(\widehat{\beta}\right), \widehat{\beta} + 1.96u\left(\widehat{\beta}\right)\right]. \tag{E14.10}$$

Table E14.2: For the data in table E14.1 and measurement model (E14.4), the estimate and standard uncertainty is listed for each input quantity (N m).

Input quantity	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8
Estimate y_i	0.095 0	0.196 6	0.301 6	0.498 3	1.008 3	3.026 6	4.046 6	5.066 6
Standard uncertainty $u(y_i)$	0.002 2	0.002 1	0.001 7	0.001 7	0.004 0	0.003 3	0.004 9	0.021 9

E14.5.2 Bayesian uncertainty analysis

In a Bayesian inference one combines the prior knowledge about the measurand (and other unknowns) with the information contained in the data through application of Bayes' theorem. The result is the posterior distribution which summarizes the knowledge about the measurand (and other unknowns in the statistical model (E14.5)) conditional on the observed data. In our case, the posterior is given through the following probability density function (PDF)

$$\pi(\beta, \sigma^2|\text{data}) \propto \pi(\beta, \sigma^2)l(\beta, \sigma^2; \text{data}), \tag{E14.11}$$

where $l(\beta, \sigma^2; \text{data})$ denotes the likelihood function (E14.6) for the assumed statistical model (E14.5), and $\pi(\beta, \sigma^2)$ the employed prior for β and $\sigma^2 = (\sigma_1^2, \ldots, \sigma_p^2)^\top$. From the joint posterior (E14.11), the marginal posterior $\pi(\beta|\text{data})$ for the measurand is obtained through marginalization according to

$$\pi(\beta|\text{data}) = \int_0^\infty \ldots \int_0^\infty \pi(\beta, \sigma^2|\text{data})\mathrm{d}\sigma_1^2 \ldots \mathrm{d}\sigma_p^2. \tag{E14.12}$$

The marginal posterior (E14.12) is a PDF that can be seen as the complete Bayesian uncertainty analysis for the measurand. Summary statistics of this PDF may be sufficient in many cases, and one can consider in line with the GUM the posterior mean,

$$\widehat{\beta} = \int_{-\infty}^\infty \pi(\beta|\text{data})\beta\mathrm{d}\beta \tag{E14.13}$$

as the Bayesian estimate, and the posterior standard deviation as the associated standard uncertainty $u(\beta)$, where

$$u^2(\beta) = \int_{-\infty}^\infty \pi(\beta|\text{data})(\beta - \widehat{\beta})^2\mathrm{d}\beta. \tag{E14.14}$$

Note that from a Bayesian point of view the standard uncertainty characterizes the uncertainty about the quantity β, rather than the uncertainty about its estimate $\widehat{\beta}$ (which is known exactly). For this reason, the notation $u(\beta)$ is used in (E14.14) rather than $u(\widehat{\beta})$. Finally, a 95% credible interval $[\underline{\beta}, \overline{\beta}]$ can be calculated from the posterior (E14.12) which satisfies

$$\int_{\underline{\beta}}^{\overline{\beta}} \pi(\beta|\text{data})\mathrm{d}\beta = 0.95. \tag{E14.15}$$

Equation (E14.15) does not uniquely determine a credible interval and further conditions need to be posed, for example that the credible interval is symmetric around the Bayes estimate, probabilistically symmetric, or of shortest length, cf. also [52].

Informative prior

Below, $\pi(\beta)$ denotes a PDF that models the prior knowledge about the measurand β. $\widehat{\sigma}_i^2$ are the prior guesses of the variances σ_i^2, $i = 1, \ldots, p$, which have to be inferred. Assume that the reliability of these variance guesses can be expressed in terms of chosen coefficients of variations c_i. In using inverse Gamma distributions to model such prior knowledge, the parameters a_i and b_i of the inverse Gamma distributions are then determined through

$$a_i = 2 + \frac{1}{c_i^2}, \quad b_i = (a_i - 1)\widehat{\sigma}_i^2, \tag{E14.16}$$

i.e., the prior knowledge about each σ_i^2 is modeled by a distribution with mean $\widehat{\sigma}_i^2$ and variance $c_i^2(\widehat{\sigma}_i^2)^2$. The resulting marginal posterior for the measurand is then obtained as

$$\pi(\beta|\text{data}) \propto \pi(\beta) \prod_{i=1}^{p} t_{n_i-1+2a_i}\left(\beta; y_i/x_i, ((n_i-1)S_i^2 + 2b_i)/(n_i x_i^2[n_i - 1 + 2a_i])\right), \tag{E14.17}$$

where $\pi(\beta)$ denotes the prior PDF for β, and $t_\nu(x; m, s^2)$ stands for the PDF of a scaled and shifted t-distribution with ν degrees of freedom, i.e.

$$t_\nu(x; m, s^2) \propto \left(1 + \frac{1}{\nu}\frac{(x-m)^2}{s^2}\right)^{-\frac{\nu+1}{2}}, \tag{E14.18}$$

cf. also [52]. The univariate PDF (E14.17) is easily evaluated, and the summary statistics (E14.13)–(E14.15) can immediately be obtained through standard procedures of numerical quadrature. Note that for evaluating (E14.17) it is advantageous to calculate the logarithm of $\pi(\beta|\text{data})$ first, and applying the exponential function afterwards.

In this example no true prior knowledge has been available. For the purpose of illustration, hypothetical prior knowledge in form of a normal distribution for β with mean 1 and standard deviation 0.1 has been used, accompanied with guesses $\widehat{\sigma}_i^2$ for the variances that have been taken as the observed variances S_i^2. The reliability of the variance estimates was modelled by a coefficient of variation equal to unity, $c_i = 1, i = 1, \ldots, p$. Note that prior knowledge is information available before measurements are preformed. True prior knowledge will and shall not be deductions of observed data, as in this illustrative case.

Noninformative prior

The case of vague prior knowledge can be modelled by choosing a vague prior for the measurand $\pi(\beta)$ which has large variance, together with parameters a_i and b_i of the inverse Gamma distributions which approach zero. In this case the inverse Gamma distributions (taken to model prior knowledge about the variances) are distributions with huge tails and they do not even possess a finite expectation. The posterior (E14.17) then approaches

$$\pi(\beta|\text{data}) \propto \prod_{i=1}^{p} t_{n_i-1}(\beta; y_i/x_i, S_i^2/(n_i x_i^2)), \tag{E14.19}$$

which is also formally obtained when using the following non-informative prior [92]

$$\pi(\beta, \sigma^2) \propto \prod_{i=1}^{p} \frac{1}{\sigma_i^2} \tag{E14.20}$$

from the start.

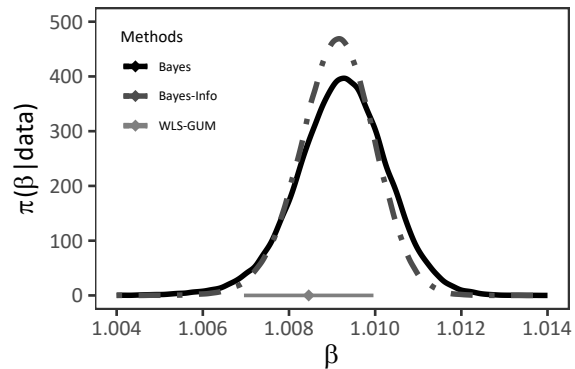| Method	$\widehat{\beta}$	$u(\widehat{\beta})$	95% coverage / credible interval
	a.u.	a.u.	a.u.
OLS-GUM	1.010 7	0.001 5	[1.007 7, 1.013 6]
WLS-GUM	1.008 5	0.000 8	[1.007 0, 1.010 0]
Bayes	1.009 2	0.001 1	[1.007 0, 1.011 2]
Bayes-Info	1.009 1	0.000 9	[1.007 3, 1.010 8]

Figure E14.2: Marginal posterior distribution $\pi(\beta|\text{data})$ for the measurand according to (E14.19) (solid line) and (E14.17) (dashed line) using a non-informative or a vague informative prior, respectively. The symbols in light gray display the estimate $\widehat{\beta}$ and 95% coverage interval of the weighted least-squares approach for comparison.

Table E14.3: Results obtained by ordinary least-squares (OLS-GUM) and weighted least-squares (WLS-GUM) with uncertainty evaluation according to the GUM, as well as results from a Bayesian uncertainty analysis with (Bayes-Info) and without (Bayes) accounting for vague prior knowledge.

E14.6   Reporting the result

Table E14.3 contains the estimate, its associated standard uncertainty, and the 95% coverage interval obtained by application of the GUM to ordinary and weighted least-squares estimation, together with corresponding results for the Bayesian uncertainty analysis. The credible intervals determined by the Bayesian uncertainty analysis were taken as probabilistically symmetric intervals. Figure E14.2 shows the PDFs for the measurand obtained by Bayesian uncertainty analyses in comparison with the results achieved by weighted least-squares estimation with uncertainty evaluated according to the GUM.

E14.7   Discussion and recommendation

The results obtained by application of the GUM to ordinary and weighted least-squares estimation are different. This difference is due to the difference of the corresponding measurement models (E14.4) used. Specifically, weighted least-squares estimation with weights $n_i W_i \propto 1/u^2(y_i)$ leads to a different estimate for the slope and a smaller uncertainty $u(\widehat{\beta})$ than ordinary least-squares estimation. In fact, these weights are "optimal" in the sense that they lead to a minimum uncertainty under all measurement models (E14.4).

On the other hand, ordinary least-squares does not apply "optimal" weights and results in a larger uncertainty associated with its different estimate for the measurand. Since the corresponding measurement model is linear in the data, the squared standard uncertainty provides an unbiased estimate of the variance of the ordinary least-squares estimator under hypothetical repeated sampling from the statistical model (E14.5). From the perspective of the type A evaluation of the GUM, the corresponding uncertainty evaluation can thus be recommended. However, ordinary least-squares estimation utilizes a measurement model that does not account for the fact that different observations have different variability. That is, measurements are assigned the same weight although their variability differs by orders of magnitude.

The weighted least-squares estimate can be also be viewed as a solution to the statistical model (E14.5) if the variances σ_i^2 were known. For unknown variances, however, the optimal

weights
$n_i W_i \propto 1/(\sigma_i^2/n_i)$ are also unknown. The uncertainties $u(y_i)$ approximate the optimal weights $\sigma_i/\sqrt{n_i}$, but they will generally be different – especially for a small number of repeated measurements. Some of the observations are weighted too high and, more importantly, are also treated by the GUM uncertainty evaluation as being more accurate than they actually are. Consequently, the resulting uncertainty associated with the weighted least squares estimate might be too small. This judgment is to be seen from a frequentist point of view, which corresponds to the view of the GUM with respect to type A evaluation.

The Bayesian uncertainty analysis is based on the statistical model (E14.5) and does account for the different variability in the observations. At the same time, it does not use a single estimate of that variability to be used in a subsequent estimation of the measurand, but rather estimates the measurand and the variability in the observations simultaneously. Due to the straight-line model, all observations influence the estimation of all different variabilities in the observations, and observations with large variability will have less influence in the final result for the measurand. Furthermore, Bayesian inference allows prior knowledge about the measurand to be taken into account. For these reasons, we recommend the Bayesian uncertainty analysis for this example. It should be noted that also methods from classical statistics can be used to analyze the data on the basis of the statistical model (E14.5) which has not been considered in this example.

Bayesian inference using our hypothetical informative prior yields very similar results to those using the non-informative prior. The reason is that the data overrule the prior information taken for the measurand, and that the (hypothetical) prior knowledge about the variances has been taken only vaguely and in accordance with the observed variances. If either of these two latter conditions for the prior of the variances is removed, the results of an informative Bayesian inference might look significantly different because each variance is modelled individually for each stimulus value and only a small number of repeated measurement results are available. In this case, the prior about for the variance will be more informative. In other applications it can be reasonable to assume a common variance, which would reduce the sensitivity with respect to the prior for the variance significantly. Furthermore, the proceeding provided for the Bayesian inference would then result in a single t-distribution for the measurand in the non-informative case, or the product of a single t-distribution and an informative prior for the measurand otherwise.

We emphasise that the statistical model (E14.5) does not directly account for possible errors in the measurement results of the reference system. In fact, the example B2 of the guideline VDI/VDE 2600 part 2 [40] reports non-vanishing uncertainties for them. The statistical model (E14.5) account for such an additional variability to a certain extent, as it includes unknown, individual variances for the dependent variable, that are simultaneously inferred together with the parameters of the straight line.

Acknowledgment

Example E15

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

L.L. Martins, A.S. Ribeiro, M.G. Cox, J.A. Sousa, D. Loureiro, M.C. Almeida, M.A. Silva, R. Brito, A.C. Soares

E15.1   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) [10, 139] contains three linear CCD (charge-coupled device) cameras, in different spatial positions and orientations, with overlapping fields of view, permitting the simultaneous observation of an infrared LED (light emitting diode) 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.

E15.2   Introduction of 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 [10,139] – 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 characterised by a depth range between 1.5 m and 6.0 m and a corresponding cross-section area ranging between (0.90 m × 0.55 m) and (3.6 m × 2.6 m). Using triangulation techniques, the location of an infrared LED can be determined with a measurement accuracy [10] variable between $(60 + 7\,\mathrm{m}^{-1} \cdot L)$ μm and $(130 + 17\,\mathrm{m}^{-1} \cdot L)$ μm, 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 [10,139].

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 E15.1, which displays seven spatial distances $d_1 \dots d_7$ that are measured.
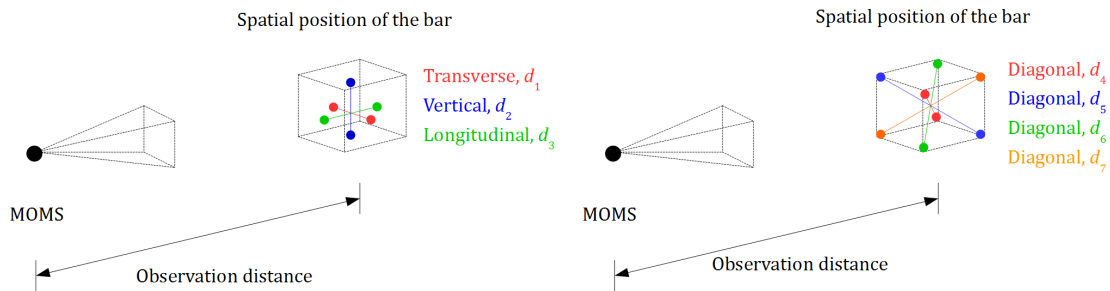
Figure E15.1: Schematic representation of the MOMS calibration

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 [10].

E15.3   Specification of the measurand(s)

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:

$$d = l - l_s (1 + \alpha_s \cdot \theta_s) \tag{E15.1}$$

where α_s is the coefficient of thermal expansion of the carbon fibre composite bar and θ_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.

E15.4   Measurement model

The measurement model can be obtained from expression (E15.1) and is given by

$$l = d + l_s (1 + \alpha_s \cdot \theta_s) \tag{E15.2}$$

E15.5   Uncertainty propagation

The application of the law of propagation of uncertainty [3] to expression (E15.2) yields

$$u^2(l) = c_d^2 \cdot u^2(d) + c_{l_s}^2 \cdot u^2(l_s) + c_{\alpha_s}^2 \cdot u^2(\alpha_s) + c_{\theta_s}^2 \cdot u^2(\theta_s) \tag{E15.3}$$

with

$$c_d = \frac{\partial l}{\partial d} = 1 \tag{E15.4}$$

$$c_{l_s} = \frac{\partial l}{\partial l_s} = 1 + \alpha_s \cdot \theta_s \tag{E15.5}$$

$$c_{\alpha_s} = \frac{\partial l}{\partial \alpha_s} = l_s \cdot \theta_s \tag{E15.6}$$

$$c_{\theta_s} = \frac{\partial l}{\partial \theta_s} = l_s \cdot \alpha_s \tag{E15.7}$$

and thus

$$u^2(l) = u^2(d) + (1 + \alpha_s \cdot \theta_s)^2 \cdot u^2(l_s) + l_s^2 \cdot \theta_s^2 \cdot u^2(\alpha_s) + l_s^2 \cdot \alpha_s^2 \cdot u^2(\theta_s) \tag{E15.8}$$

A reference standard bar such as that used in the calibration of the MOMS is designed to be a rigid body characterised 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 (E15.8) can be simplified:

$$u^2(l) = u^2(d) + u^2(l_s) + l_s^2 \cdot \theta_s^2 \cdot u^2(\alpha_s) \tag{E15.9}$$

Table E15.1 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 E15.1) in the measurement volume, at a nominal observation distance of 3.5 m. Table E15.1 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 E15.1: MOMS calibration results

Test number	Differences between readings and reference values (mm)						
	d_1	d_2	d_3	d_4	d_5	d_6	d_7
1	-0.017	-0.007	-0.021	-0.008	0.015	0.028	0.020
2	0.020	0.011	-0.002	-0.044	-0.007	0.031	0.030
3	-0.011	-0.013	-0.008	-0.034	0.020	0.042	0.021
4	0.001	0.011	-0.003	-0.036	-0.032	0.058	0.084
$\bar{d}_i$	-0.002	0.001	-0.009	-0.031	-0.001	0.040	0.039
$s(d_i)$	0.016	0.012	0.009	0.016	0.024	0.014	0.031

Based on the results shown in table E15.1, correlation coefficients, $r(d_i, d_j)$, were determined between pairs of spatial positions of the reference standard bar,

$$r(d_i, d_j) = \frac{u(d_i, d_j)}{s(d_i) \cdot s(d_j)}, \tag{E15.10}$$

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

$$u(d_i, d_j) = \frac{1}{n(n-1)} \sum_{k=1}^{n} (d_{i_k} - \bar{d}_i)(d_{j_k} - \bar{d}_j) \tag{E15.11}$$

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 E15.2. 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.

Table E15.2: Correlation coefficients between the obtained differences

$r(d_i, d_j)$	d_1	d_2	d_3	d_4	d_5	d_6	d_7
d_1	1	0.20	0.20	-0.20	-0.14	0.01	0.06
d_2	0.20	1	0.16	-0.14	-0.22	0.06	0.17
d_3	0.20	0.16	1	-0.25	-0.16	0.10	0.13
d_4	-0.20	-0.14	-0.25	1	0.12	-0.10	-0.09
d_5	-0.14	-0.22	-0.16	0.12	1	-0.17	-0.23
d_6	0.01	0.06	0.10	-0.10	-0.17	1	0.22
d_7	0.06	0.17	0.13	-0.09	-0.23	0.22	1

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 E15.1. In a global perspective, an estimate of the difference between reading and reference values can be obtained by averaging. The corresponding standard uncertainty [51] is given by

$$u^2(d) = \sum_{i=1}^{7} c_i^2 \cdot u^2(d_i) + 2\sum_{i=1}^{6}\sum_{j=i+1}^{7} c_i \cdot c_j \cdot u(d_i) \cdot u(d_j) \cdot r(d_i, d_j) \tag{E15.12}$$

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 E15.2), allowing to simplify expression (12) to

$$u^2(d) = \frac{1}{49}\left[\sum_{i=1}^{7} s^2(d_i) + 2\sum_{i=1}^{6}\sum_{j=i+1}^{7} s(d_i) \cdot s(d_j) \cdot r(d_i, d_j)\right] \tag{E15.13}$$

The use of the values in tables E15.1 and E15.2 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: $l_s = 1550.313$ 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 coverage 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 $\left(\alpha_s = 0.0 \times 10^{-6}\,°\mathrm{C}^{-1}\right)$. However, it has an associated standard uncertainty represented by a uniform distribution with a semi-width of $2 \times 10^{-6}\,°\mathrm{C}^{-1}$, 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 [51] corresponds to

$$u(\alpha_s) = \frac{1}{\sqrt{3}} 2 \times 10^{-6}\,°\mathrm{C}^{-1} = 1.2 \times 10^{-6}\,°\mathrm{C}^{-1} \tag{E15.14}$$

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 $\theta_s = 0.1$ °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 [3] is given by

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

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

Table E15.3: Summary of the standard uncertainty components

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	$1.2 \times 10^{-6}\,°\text{C}^{-1}$	155 mm °C	0.0002 mm	50
$u(\theta_s)$	Temperature deviation from reference value	0.41 °C	0	0	∞

E15.6 Reporting the result

Based on the results shown in table E15.3, the combined standard uncertainty, $u_c(l)$, of the length reading is determined from expression (E15.9), 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

$$U_{95\%}(l) = k \cdot u_c(l) = 2.04 \cdot 0.010\,\text{mm} = 0.021\,\text{mm}$$

E15.7 Interpretation of results

Table E15.3 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 (characterised by a thermal expansion coefficient estimate of $11.5 \times 10^{-6}\,°C^{-1}$, 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.

Example E16

Evaluation of measurement uncertainty in thermal comfort

J.A. Sousa, A.S. Ribeiro, M.G. Cox, L.L. Martins

E16.1 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.

E16.2 Introduction of the application

The example, which is based in part on the paper [151], 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 [11], 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 recognised 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 utilised 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.

E16.3   Specification of the measurand(s)

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

$$
\begin{aligned}
\text{PMV} =& [0.303\exp(-0.036M) + 0.028] \\
& \times \Big\{ M - W - 3.05 \times 10^{-3}[5733 - 6.99(M - W) - \rho_a] \\
& - 0.42(M - W - 58.15) - 1.7\cdot 10^{-5}(5867 - \rho_a)M - 0.0014(34 - t_a)M \\
& - 3.96\cdot 10^{-8}\big[(t_{cl} + 273)^4 - (t_r + 273)^4\big]f_{cl} - (t_{cl} - t_a)f_{cl}h_c(t_{cl})\Big\}
\end{aligned}
\tag{E16.1}
$$

where
M metabolic rate in $\text{W}\,\text{m}^{-2}$
W effective mechanical power in $\text{W}\,\text{m}^{-2}$
ρ_a water-vapour partial pressure in Pa
t_a air temperature in °C
f_{cl} clothing surface area factor
t_{cl} clothing surface temperature in °C
t_r mean radiant temperature in °C and
h_s convective heat transfer coefficient in $\text{W}\,\text{m}^{-2}\,\text{K}^{-1}$
with the main complication arising from the fact that the quantity t_{cl} is defined implicitly (see expression (E16.2) 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 $\text{m}^2\,\text{K}\,\text{W}^{-1}$), and (b) quantities t_a, t_r, v_{ar} and RH obtained by measurement (v_{ar} is the relative air velocity in $\text{m}\,\text{s}^{-1}$ that influences h_c, and RH is the relative humidity in % that influences ρ_a.

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

E16.4   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:

$$
\begin{aligned}
t_{cl} =& 35.7 - 0.028(M - W) \\
& - I_{cl}\Big\{3.96\cdot 10^{-8}\big[(t_{cl} + 273)^4 - (t_r + 273)^4\big]f_{cl} + (t_{cl} - t_a)f_{cl}h_c(t_{cl})\Big\}
\end{aligned}
\tag{E16.2}
$$

where the convective heat transfer coefficient is

$$
h_c(t_{cl}) = \max\left(2.38\,|t_{cl} - t_a|^{1/4}, 12.1\sqrt{v_{ar}}\right)
\tag{E16.3}
$$

which is mathematically identical to the expression given in the standard [11], but simpler. We write $h_c(t_{cl})$ in this equation rather than simply h_c to emphasise that it depends on t_{cl}, a parameter already existing explicitly in the model. The clothing surface area factor depends on the clothing insulation:

$$
f_{cl} = \begin{cases} 1.00 + 1.290I_{cl}, & I_{cl} \le 0.0775\,\text{m}^2\,\text{K}\,\text{W}^{-1} \\ 1.05 + 0.645I_{cl}, & I_{cl} > 0.0775\,\text{m}^2\,\text{K}\,\text{W}^{-1} \end{cases}
\tag{E16.4}
$$

In the cases considered in this study ρ_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

$$\rho_a = RH \times \rho_s(t_a) 10 \, Pa \tag{E16.5}$$

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

$$\rho_s(t_a) = \exp\left(16.6536 - \frac{4030.183}{t_a + 235}\right) K \, Pa \tag{E16.6}$$

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

$$g(t_{cl}) = 3.96 \cdot 10^{-8}\left[(t_{cl} + 273)^4 - (t_r + 273)^4\right]f_{cl} + (t_{cl} - t_a)f_{cl}h_c(t_{cl}) \tag{E16.7}$$

is common to equations (E16.1) and (E16.2). Again, we indicate g as a function of t_{cl} to emphasise that, given values for the other quantities involved, the value of $g(t_{cl})$ can readily be obtained knowing t_{cl}. Therefore, equation (E16.2) can be expressed as

$$t_{cl} = 35.7 - 0.028(M - W) - I_{cl}g(t_{cl}) \tag{E16.8}$$

and equation (E16.1) can be simplified accordingly to

$$\begin{aligned}
PMV =&[0.303 \exp(-0.036M) + 0.028] \\
&\times \{M - W - 3.05 \times 10^{-3}[5733 - 6.99(M - W) - \rho_a] \\
&- 0.42(M - W - 58.15) - 1.7 \cdot 10^{-5}(5867 - \rho_a)M - 0.0014(34 - t_a)M \\
&- [35.7 - 0.028(M - W) - t_{cl}]/I_{cl}\}
\end{aligned} \tag{E16.9}$$

Nevertheless, the above expressions remain complicated and working with them using the GUM approach (LPU - law of propagation of uncertainty) [51] 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 I_{cl} 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.

In table E16.1, best estimates of the quantities concerned are taken as the average values of 40 observations (obtained every 2 min) for those quantities, and standard uncertainties associated with those estimates are evaluated. It is assumed that the measuring conditions are stable during the period of measurement and thus the observations can be regarded as repeated indication values of the quantities. These standard uncertainties comprise contributions from the averaging process (Type A evaluation of uncertainty) and from instrument calibration (Type B evaluation), as discussed above.

---

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

Table E16.1: Estimates of the input quantities

Location	$M \times 58.2$ $\mathrm{W\,m^{-2}}$	W $\mathrm{W\,m^{-2}}$	$I_{\mathrm{cl}} \times 0.155$ $\mathrm{m^2\,K\,W^{-1}}$	t_{a} °C	t_{g} °C	RH %	v_{ar} $\mathrm{m\,s^{-1}}$	t_{r} °C
Office 1	1.2	0	0.7	25.8	26.3	47.4	0.01	26.4
Office 2	1.2	0	1.0	20.9	21.2	68.1	0.02	21.3
Customer service	1.2	0	0.7	24.0	24.3	46.4	0.07	24.6

In table E16.2, sample standard deviations of the experimental data, taken as the standard uncertainties, are given, which express the repeatability (Type A) contributions obtained from the experimental data used for the studies. The table also gives the Type B contributions, which for each quantity is a constant value because the same measuring instrument was used for all three locations.

Table E16.2: Standard uncertainties associated with the estimates of the measured quantities in table E16.1

Location	t_{a} °C	t_{g} °C	RH %	v_{ar} $\mathrm{m\,s^{-1}}$	t_{r} °C
Office 1	0.04	0.05	0.56	0.03	0.15
Office 2	0.05	0.00	0.28	0.02	0.08
Customer service	0.04	0.03	0.13	0.05	0.05
All	0.1	0.1	0.5	0.05	0.2

All entries in table E16.2 are given to the same number of decimal places. As a result, some of these entries are reported as zero. It would be necessary to use one or two further decimal digits to demonstrate that these values are non-zero albeit negligible.

Another important comment refers to the difference between the definitions and use of standard uncertainty in the GUM and GUM Supplement 1 (GUM S1) [52]. 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 E16.2 to the number of decimal digits reported.

E16.5   Uncertainty propagation

E16.5.1   Preamble

Two approaches for the evaluation of uncertainty are considered: the GUM uncertainty framework, based on LPU [51], and the propagation of distributions of GUM-S1, based on the Monte Carlo method [52].

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.

E16.5.2   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 (E16.8). As for many complicated models, determining the required partial derivatives algebraically is not always practical and a numerical approach is recommended [76]. It is a burden not shared by the GUM-S1 approach.

In this model, special attention must be paid to the derivative $\partial(\text{PMV})/\partial t_{\text{cl}}$ given values for all other input quantities. Equation (E16.8) (with equation (E16.7)) is solved for t_{cl}, and then PMV is evaluated using equation (E16.9). 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 (E16.1) involves $g(t_{\text{cl}})$, defined by equation (E16.7), and hence the derivative $g'(t_{\text{cl}})$ is required. This derivative is not necessary when using equation (E16.8), is a simplification that poses no problem since $g(t_{\text{cl}})$ is already used when solving equation (E16.8) 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} (equation (E16.4)) and to the convective heat transfer coefficient h_{c} (equation (E16.3)).

The first of these model functions (the second is similar) is illustrated in Figure E16.1 , showing f_{cl} as a function of clothing insulation I_{cl}. Equation (E16.4) can be also be expressed as

$$f_{\text{cl}} = 1.1 + \min\left[0.645\,(I_{\text{cl}} - 0.0775),\quad 1.290\,(I_{\text{cl}} - 0.0775)\right]$$

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

Figure E16.2 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 [152], 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 $0.001\,\text{m}^2\,\text{W}\,\text{K}^{-1}$ to $0.200\,\text{m}^2\,\text{W}\,\text{K}^{-1}$. 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.

E16.6   Reporting the result

Table E16.3 shows for the PMV model the input quantities, the PDFs that characterise them, and their estimates and associated standard uncertainties. Some of the input quantities are experi-
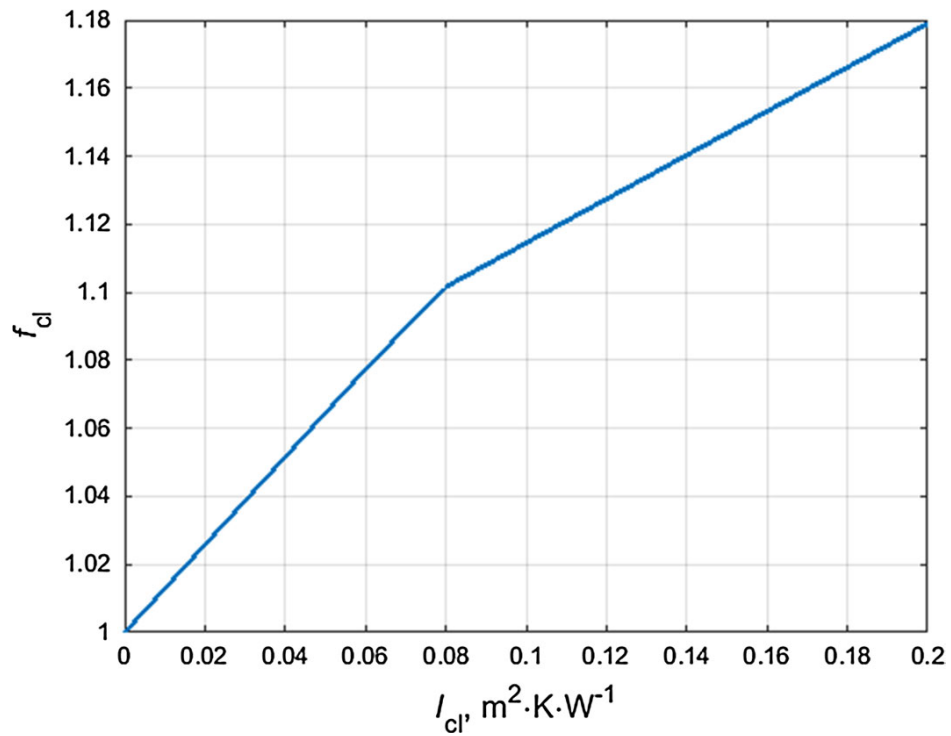
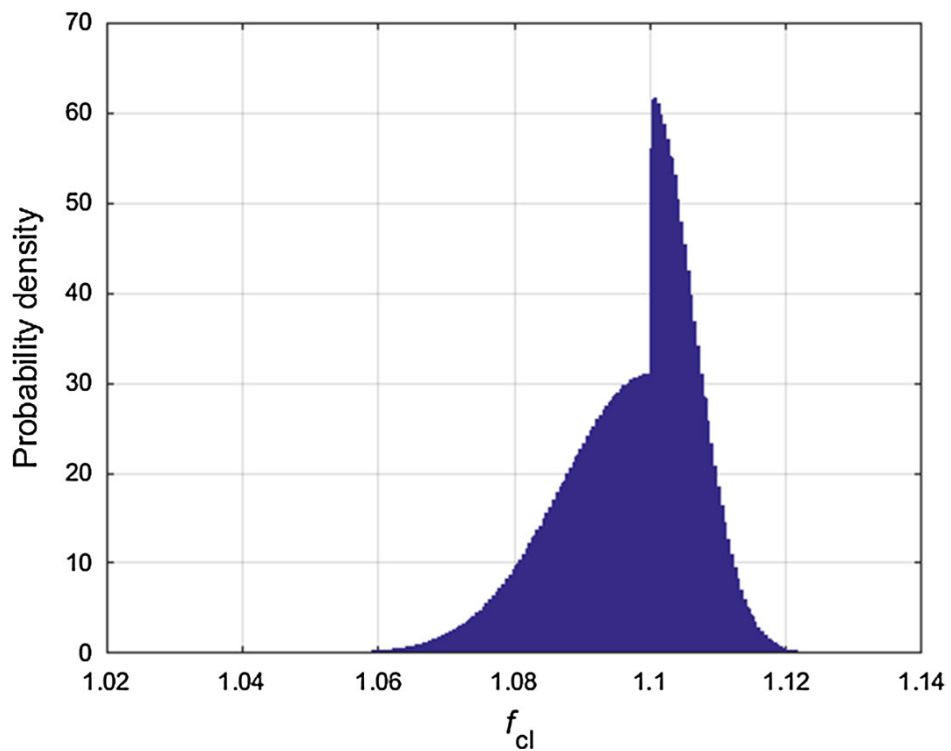Figure E16.1: Clothing surface area factor f_{cl} as a function of clothing insulation I_{cl}

Figure E16.2: PDF for f_{cl} as a function of I_{cl}, as produced by a Monte Carlo calculation
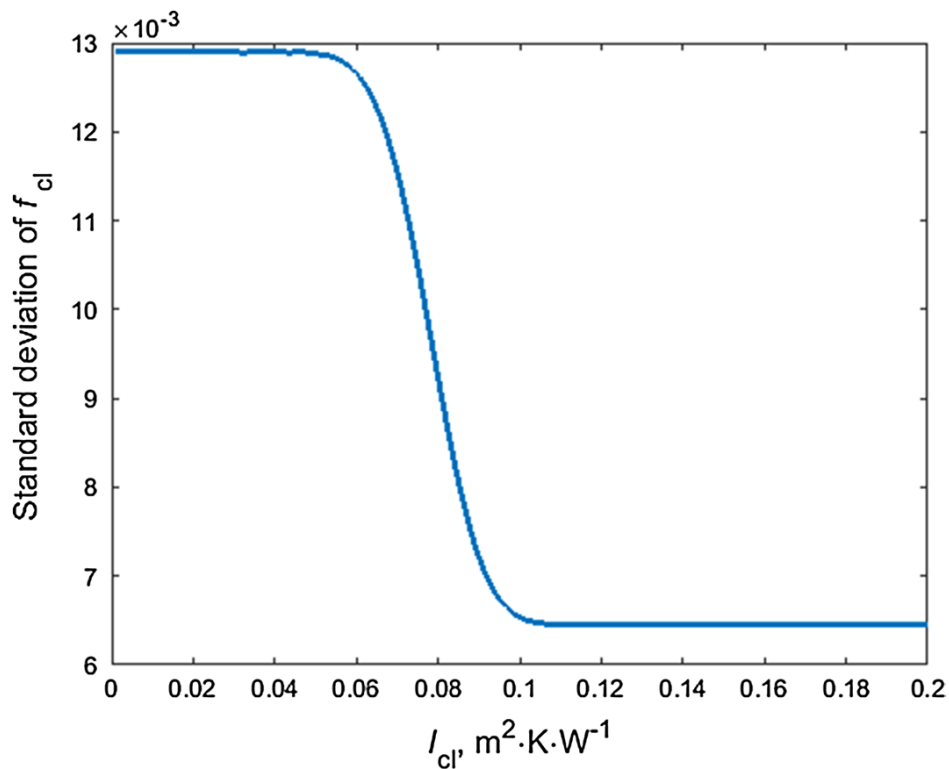
Figure E16.3: Standard deviation of f_{cl} as a function of I_{cl}

mental, while other quantities have tabulated values [11] 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.

Table E16.3: GUM uncertainty budget for the PMV model

Quantity	PDF	Estimate	u_i	c_i	$c_i u_i$
$M/\mathrm{W\,m^{-2}}$	Ref. value	70			
$W/\mathrm{W\,m^{-2}}$	Ref. value	0			
$t_a/°\mathrm{C}$	Combined	22.0	0.1	0.228	0.013
$I_{cl}/\mathrm{m^2\,K\,W^{-1}}$	Ref. value	0.078			
$t_r/°\mathrm{C}$	Gaussian	22.0	negligible		
RH/%	Rectangular	60.0	0.3	0.0059	0.0017
$v_{ar}/\mathrm{m\,s^{-1}}$	Rectangular	0.10	0.03	3.27	0.094
PMV		-0.75	$u(\mathrm{PMV}) = 0.094$		$U_{0.95}(\mathrm{PMV}) = 0.18$

A relevant conclusion that can be drawn from table E16.3 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 (E16.1).

The Monte Carlo calculation was based on the experimental input indicated in tables E16.1
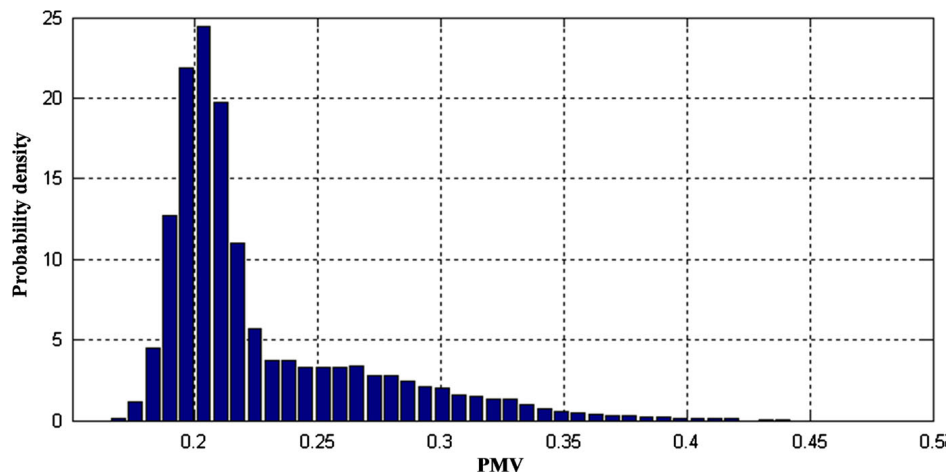
Figure E16.4: PDF for the PMV index from Monte Carlo calculation

and E16.2. 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 [52].

Figure E16.4 shows the PDF for the PMV index provided by the Monte Carlo (MC) method for the same location as table E16.3. 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 E16.4 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 [51].

Table E16.4: Results from the Monte Carlo calculation

Quantity	Estimate	Standard uncertainty	95 % coverage interval		
			Lower limit	Upper limit	Width
PMV	0.23	0.04	0.18	0.34	0.16

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 [11]. 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 $N(1.2, (0.05)^2)$ for M and the Gaussian PDF $N(0.7, (0.05)^2)$ 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 [11], 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 [11], 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 emphasises 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 ISO 7730 [11]. 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 [55].

E16.7   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 [11], but also of ISO 7726 [4], 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 [46], and close collaboration should therefore be pursued with researchers from the social sciences.

With respect to ISO 7726, one study has shown [74] 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.

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 $X = (X_1, \ldots, X_N)^\top$ by a functional relationship f in the form $Y = f(X)$ as stated in the GUM [11]. The estimate of the output quantity is taken as $y = f(x)$ The standard uncertainty $u(y)$ associated with y is evaluated from

$$u^2(y) = \sum_{i=1}^{N} \sum_{j=1}^{N} c_i u(x_i, x_j) c_j \tag{E16.10}$$

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

$$u^2(y) = c^\mathrm{T} V_x c \tag{E16.11}$$

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

$$V_x = \begin{bmatrix} u(x_1, x_1) & \cdots & u(x_1, x_N) \\ \vdots & \ddots & \vdots \\ u(x_N, x_1) & \cdots & u(x_N, x_N) \end{bmatrix} \tag{E16.12}$$

and the (row) vector $c^\top = \lceil c_1, \ldots, c_N \rceil$ of dimension $1 \times N$ contains the sensitivity coefficients. Both expressions (E16.10) and (E16.11) are equivalent representations of LPU of the GUM [11].

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)

$$u^2(y) = \sum_{i=1}^{N} [c_i u(x_i)]^2 = \sum_{i=1}^{N} u_i^2(y), \quad u_i(y) \equiv |c_i| u(x_i) \tag{E16.13}$$

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 [52].

The estimate y of Y is the value of η that solves the equation $h(\eta, x) = 0$. This equation is to be solved numerically with a suitable zero-finding algorithm [46]. The standard uncertainty $u(y)$ associated with y is evaluated from

$$u^2(y)c_y^2 = \boldsymbol{c}_x^\top \boldsymbol{U}_x \boldsymbol{c}_x \tag{E16.14}$$

where $\boldsymbol{c}_x^\top$ 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 $\boldsymbol{X} = \boldsymbol{x}$ and $Y = y$ [11].

Conditions for valid application

There are a number of conditions for valid application of the GUM uncertainty framework for non-linear models. They include [51] that f must be continuously differentiable with respect to the elements X_i of $\boldsymbol{X}$ 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 $f(\boldsymbol{X})$ 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 [152], 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 [51].

The GUM uncertainty framework does not provide the PDF for Y, but instead assumes that Y can be described by a Gaussian PDF $\mathrm{N}\left(y, u^2(y)\right)$, 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 [151].

Example E17

Preparation of calibration gas mixtures of NH_3 in nitrogen using permeation

M. Čaušević, H. Meuzelaar, A.M.H. van der Veen, M.G. Cox

E17.1 Summary

This example describes the uncertainty evaluation of the preparation of a calibration gas mixtures of ammonia in nitrogen by using permeation. The measurand is the amount fraction ammonia.

E17.2 Introduction of the application

Permeation [7] is one of several techniques to prepare dynamically calibration gas mixtures [3]. The method is a dynamic-gravimetric method, which implies that mass flow rates are used, together with information concerning the purity of the materials used [35] and the molar masses of the components to calculate the composition. One of the mass flow rates originates from the permeation tube ("permeation rate"), the other from a thermal mass flow controller. In this example, we describe the calculation of the composition of the calibration gas mixture expressed in amount fractions, as used in many high-end applications [172, 176].

E17.3 Specification of the measurand(s)

The measurand is the amount fraction ammonia, the most abundant component in the permeation tube. A calibration gas mixture is prepared of this component in high-purity nitrogen.

E17.4 Measurement model

The measurement model consists of the following parts:
1. expression to calculate the amount fraction ammonia,
2. expression to calculate the permeation rate (using regression of the recorded mass loss of the permation tube as a function of time), including effects of temperature and pressure,
3. expression of the molar masses of the parent gases,
4. expressions for calculating the composition of the parent gases.

E17.4.1   Principle

The component of interest, which in this case is ammonia (NH_3), is permeated from a permeation tube through a membrane into a flow of the carrier gas [7]. In this instance, it is high-purity nitrogen (N_2; grade 6.0). A permeation tube, containing NH_3 of known purity was hooked inside the temperature and pressure controlled permeation chamber to a magnetic suspension balance, which continuously performs accurate mass measurements. The permeation chamber is purged at a known and controlled flow rate by the carrier gas. The flow rate of the carrier gas is controlled by using thermal mass flow controller [36]. The permeation rate is determined by continuously measuring the mass of the permeation tube. The permeation rate of the component of interest through the membrane depends on the properties of the component and the permeability of the membrane:

  – chemical nature and structure of the membrane,
  – area and thickness of the membrane,
  – temperature and pressure gradient of the calibration component across the membrane.

Generally, different amount fractions are realised by either changing the gas flow of the carrier gas or the permeation flow. The latter can be changed by changing the temperature in the permeation chamber.

The measurement equation to calculate the amount fraction y_k of the component of interest k reads as [7]

$$y_k = \frac{\frac{\dot{m}_1}{\bar{M}_1} x_{1k}}{\frac{\dot{m}_1}{\bar{M}_1} + \frac{\dot{m}_2}{\bar{M}_2}}, \tag{E17.1}$$

where $\dot{m}_1$ denotes the mass flow rate from the permeation tube and $\dot{m}_2$ the mass flow rate measured by the thermal mass flow controller (MFC). $\bar{M}_1$ and $\bar{M}_2$ denote molar masses of the parent gases respectively, and x_{1k} the amount fraction of the component of interest in parent gas 1 (from the permeation tube). As nominally pure substances are used, $x_{1k} \approx 1$, but a purity analysis according to ISO 19229 [35] is necessary to determine the value for x_{1k} and the associated standard uncertainty. Furthermore, equation (E17.1) presumes that the component of interest is not present in the carrier gas, which is for ammonia in nitrogen a reasonable assumption. The most general form of model is that of ISO 6142-1 [24,34], which can be used if the assumptions underlying equation (E17.1) are not met.

E17.4.2   Permeation rate

The permeation rate (mass flow rate from the permeation tube) $\dot{m}_1$ is defined as [7]

$$\dot{m}_1 = q_{m1} = \frac{dm_1}{dt},$$

where m_1 denotes the mass of the permeation tube and t time. In this instance, a calibrated magnetic suspension balance is used to monitor the mass loss of the permeation tube.

Temperature fluctuations in the permeation chamber affect the permeation rate of the tube and the performance of the weighing. Also, pressure fluctuations in the permeation chamber affect the weighing. These effects are all incorporated in the measurement model.

E17.4.3   Material purity

Material purity affects the molar masses $\bar{M}_1$ and $\bar{M}_2$ of the parent gases in equation (E17.1). The only restriction in the modelling is that the component of interest does not occur in the carrier gas. As NH_3 is not a common impurity in high-purity N_2, this assumption is reasonable.

In this example, nominally pure ammonia and nitrogen have been used. So, the composition of the parent gases can be described as follows. The amount fraction of the most abundant component is calculated as [35]

$$x_{sj} = 1 - \sum_{\substack{i=1 \\ i \neq s}}^{q} x_{ij}, \tag{E17.2}$$

where s denotes index of the most abundant component.

E17.4.4   Molar masses

The molar mass of any component can be expressed as [24]

$$M_i = \sum_{z=1}^{Z} \nu_{zi} A_z, \tag{E17.3}$$

where
   – A_z denotes the standard atomic weight of the element z,
   – ν_{zi} coefficient of element z in the molecular formula of component i.
So, for ammonia (NH_3), $\nu_{1i} = 1$ for nitrogen and $\nu_{2i} = 3$ for hydrogen. Covariances between the molar masses of the components arise because of the use of the standard atomic weights of the elements for all molecules. (So, the molar masses of N_2 and NH_3 are correlated through the standard atomic weight of nitrogen.)

The molar mass of parent gas j is calculated as

$$\bar{M}_j = \sum_{i=1}^{q} x_{ij} M_i, \tag{E17.4}$$

where x_{ij} denotes the amount fraction of component i in parent gas j.

E17.5   Uncertainty propagation

E17.5.1   General

Throughout this example, the law of propagation of uncertainty in the GUM [51] is mostly used. For the Bayesian model that takes account of the finite resolution of the balance, Markov Chain Monte Carlo is used to obtain a sample of the posterior probability density function, from which an estimate and standard uncertainty are computed.

E17.5.2   Amount fraction of the component of interest

The evaluation of the standard uncertainty associated with y_k is performed using the law of propagation of uncertainty. A convenient way to obtain the expressions for the sensitivity coefficient (first partial derivatives of y_k with respect to the input quantities) is to use differentials.

Equation (E17.1) can be written as

$$y_k = \frac{\dot{n}_1 x_{1k}}{\dot{n}_1 + \dot{n}_2},$$

where $\dot{n}_j = \dot{m}_j / \bar{M}_j$. The expressions for the partial derivatives with respect to the input variables read as

$$\mathrm{d}y_k = \frac{\dot{n}_1}{\dot{n}_1 + \dot{n}_2}\mathrm{d}x_{1k} + \frac{\dot{n}_2 x_{1k}}{(\dot{n}_1 + \dot{n}_2)^2}\mathrm{d}\dot{n}_1 - \frac{\dot{n}_1 x_{1k}}{(\dot{n}_1 + \dot{n}_2)^2}\mathrm{d}\dot{n}_2,$$

where

$$\mathrm{d}\dot{n}_j = \frac{1}{\bar{M}_j}\mathrm{d}\dot{m}_j - \frac{\dot{m}_j}{\bar{M}_j^2}\mathrm{d}\bar{M}_j.$$

The expression for calculating the squared standard uncertainty (variance) associated with y_k, taking into consideration the covariances between $\bar{M}_1$ and $\bar{M}_2$ as well as between $\bar{M}_1$ and x_{1k} takes the form

$$u^2(y_k) = \left(\frac{\partial y_k}{\partial x_{1k}}\right)^2 u^2(x_{1k}) + \left(\frac{\partial y_k}{\partial \dot{m}_1}\right)^2 u^2(\dot{m}_1) + \left(\frac{\partial y_k}{\partial \bar{M}_1}\right)^2 u^2(\bar{M}_1) + \left(\frac{\partial y_k}{\partial \dot{m}_2}\right)^2 u^2(\dot{m}_2)$$

$$+ \left(\frac{\partial y_k}{\partial \bar{M}_2}\right)^2 u^2(\bar{M}_2) + 2\frac{\partial y_k}{\partial \bar{M}_1}\frac{\partial y_k}{\partial \bar{M}_2}u(\bar{M}_1, \bar{M}_2) + 2\frac{\partial y_k}{\partial x_{1k}}\frac{\partial y_k}{\partial \bar{M}_1}u(x_{1k}, \bar{M}_1),$$

where

$$\frac{\partial y_k}{\partial x_{1k}} = \frac{\dot{n}_1}{\dot{n}_1 + \dot{n}_2}$$

$$\frac{\partial y_k}{\partial \dot{m}_1} = \frac{\dot{n}_2 x_{1k}}{(\dot{n}_1 + \dot{n}_2)^2}\frac{1}{\bar{M}_1}$$

$$\frac{\partial y_k}{\partial \bar{M}_1} = -\frac{\dot{n}_2 x_{1k}}{(\dot{n}_1 + \dot{n}_2)^2}\frac{\dot{m}_1}{\bar{M}_1^2}$$

$$\frac{\partial y_k}{\partial \dot{m}_2} = -\frac{\dot{n}_1 x_{1k}}{(\dot{n}_1 + \dot{n}_2)^2}\frac{1}{\bar{M}_2}$$

$$\frac{\partial y_k}{\partial \bar{M}_2} = \frac{\dot{n}_1 x_{1k}}{(\dot{n}_1 + \dot{n}_2)^2}\frac{\dot{m}_2}{\bar{M}_2^2}$$

and, assuming that the amount fraction x_{1k} is calculated by difference (see equation (E17.2)),

$$u(x_{1k}, \bar{M}_1) = -\sum_{i \neq k} M_i u^2(x_{1i})$$

Uncertainty due to temperature variations in the permeation rate

The temperature dependence of the permeation rate follows Arrhenius' law [48, 49, 125]; thus the temperature dependence takes the form

$$\dot{m} = k_0 \exp\left(-\frac{E_A}{RT}\right), \tag{E17.5}$$

where E_A denotes the activation energy and k_0 is a constant, dependent on, e.g., the component and the polymer used for the permeation tube. From this expression, it follows that

$$\frac{\partial \dot{m}}{\partial T} = \frac{k_0 E_A}{RT^2}\exp\left(-\frac{E_A}{RT}\right). \tag{E17.6}$$

Values for k_0 and E_A can be obtained using the measured permeation rate determined at two different temperatures. From equation (E17.5), for permeation rates $\dot{m}_1$ and $\dot{m}_2$ at temperatures T_1 and T_2, it follows that

$$E_A = \frac{R \ln \dfrac{\dot{m}_1}{\dot{m}_2}}{\dfrac{1}{T_2} - \dfrac{1}{T_1}}$$

and

$$k_0 = \dot{m}_1 \exp\left(\frac{E_A}{RT_1}\right).$$

The uncertainty due to temperature variations in the permeation rate according to the standard [3] requires one more term, which is the standard uncertainty $u(T)$ associated with the measured temperature stability of the temperature in the tube enclosure. It can be determined directly from the measured data by the following equation for the experimental standard deviation of the observations [51].

The uncertainty due to temperature variations in the permeation rate is then calculated as

$$\frac{\partial \dot{m}}{\partial T} = \frac{k_0 E_A}{RT^2} \exp\left(-\frac{E_A}{RT}\right) u(T). \tag{E17.7}$$

At $T = 52.45\,°C$, the permeation rate is $-7.4921\ \mu g\,min^{-1}$ with an associated standard uncertainty of $0.0795\ \mu g\,min^{-1}$, determined using ordinary least squares (OLS). The activation energy is $59.164\ kJ\,mol^{-1}$ and the value of $k_0 = 43.069\ \mu g\,min^{-1}$. The value of the sensitivity coefficient at $T = 62.00\,°C$ is $4.81 \times 10^{-7}\ \mu g\,min^{-1}\,K^{-1}$. The standard uncertainty contribution due to temperature fluctuations on the permeation rate is $0.0022\ \mu g\,min^{-1}$.

Uncertainty of weighing with the balance

Factors that affect the uncertainty of weighing include resolution of the balance, repeatability, and the linearity of the balance. Eccentric loading of the balance does not occur, since the weighted tube is fixed in the same position during the whole measurement. The repeatability of the weighing affects the readings that are used to calculate the permeation rate, so this effect is duly taken into account when calculating the uncertainty of the permeation rate. The resolution of the balance amounts to $1\ \mu g$ and affects the assessment of the linearity of the balance.

Fluctuations of temperature, pressure and convection are believed to be duly reflected in the dispersion of the weighing data and not evaluated separately. Such an evaluation could, if necessary, be performed by determining the effect of temperature and pressure on buoyancy. Temperature and pressure fluctuations give rise to fluctuations in the density of the carrier gas (N_2 in this example). The density can be calculated from a suitable equation of state [144] or a simplified formula such as given in ISO 6976 [28].

Linearity testing verifies the accuracy of the instrument at intermediate values of weight. The balance linearity test assesses the ability of the balance to accurately measure the mass of an added weight. The uncertainty of the balance due to linearity can then be described as deviation of the straight line between two measured values of the same load. The linearity assessment is conducted by verifying whether the balance records a mass loss (mimicked by a weight of $100\ mg$) is, within the uncertainty correctly recorded by the MSB. The permeation tube is substituted by a $20\ g$ weight in this approach.

The balance used for weighing the permeation tube is assessed for linearity as follows. A 20 g weight is suspended in the permeation chamber for a prolonged period of time. The reading of the balance is recorded in the same fashion as for a permeation tube. After allowing for a stabilisation period and taking a sufficient number of readings, a weight of 100 mg is added. Then, again after a stabilisation period, a further ~~sufficient~~ number of readings is taken. The results of the weighings are shown in figure E17.1.

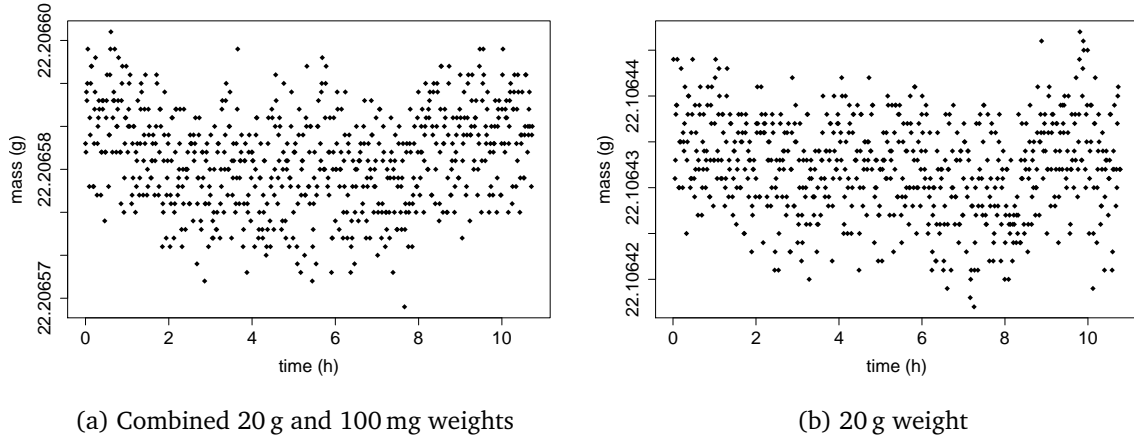(a) Combined 20 g and 100 mg weights

(b) 20 g weight

Figure E17.1: Results of the weighings performed for a linearity assessment of the balance over a range of 100 mg

From the data in figure E17.1, it can be seen that there are at least two effects at play, (1) the repeatability of weighing and (2) the resolution of the balance. To evaluate the linearity of the balance, the recorded mean difference of the two weighings is compared with the conventional mass of the 100 mg weight. The mass difference as recorded by the balance is defined as

$$d = \mu_2 - \mu_1, \tag{E17.8}$$

where μ_i denotes the mean reading of the balance. From both weighings, 700 observations are available, which have been used. A naive evaluation would use the type A evaluation method of either the GUM [51], based on the normal distribution, or that in Supplement 1 of the GUM (GUM-S1) [52], based on the t distribution. Neither of these two methods is capable of addressing duly the two effects (finite resolution and repeatability), as in the recorded readings these effects are inseparable. Hence, the use of a Bayesian that addresses both effects and is fitted to the data is to be preferred.

The Bayesian evaluation used is based on the manual of the R [149] package `RStan` [162]. In this manual, a Bayesian model is given for evaluating the standard deviation of the mean of a series of observations. The likelihood is formulated as follows

$$z_i \sim N(\mu, \sigma^2)$$

with unknown mean μ and unknown variance σ^2. The z_i denote raw (not rounded) observations. The joint prior on the parameters is non-informative and takes the form of a Jeffreys' prior [92]

$$p(\mu, \sigma^2) \propto \sigma^{-2}.$$

The raw observations are distributed as

$$z_i \sim R(y_i, r),$$

```
1  data {
2      int<lower=1> N1;
3      vector[N1] y1;
4      int<lower=1> N2;
5      vector[N2] y2;
6  }
7  parameters {
8      real mu1;
9      real mu2;
10     real<lower=0> sigma_sq1;
11     real<lower=0> sigma_sq2;
12     vector<lower=−0.5, upper=0.5>[N1] y_err1;
13     vector<lower=−0.5, upper=0.5>[N2] y_err2;
14 }
15 transformed parameters {
16     real<lower=0> sigma1;
17     real<lower=0> sigma2;
18     vector[N1] z1;
19     vector[N2] z2;
20     sigma1 = sqrt(sigma_sq1);
21     sigma2 = sqrt(sigma_sq2);
22     z1 = y1 + y_err1;
23     z2 = y2 + y_err2;
24 }
25 model
26 {
27     target += −2 * log(sigma1);
28     z1 ~ normal(mu1, sigma1);
29     target += −2 * log(sigma2);
30     z2 ~ normal(mu2, sigma2);
31 }
32 generated quantities {
33     real diff;
34     diff = mu1−mu2;
35 }
```

Listing E17.1: Model for computing the difference between two series of observations, with flat priors on μ and $\ln \sigma^2$ and modelling the effect of finite resolution

where R denotes the rectangular distribution with mean y_i and semi-width $r = 0.5\,\mu g$ (the resolution of the balance). Save for using z_i instead of y_i, this model is well covered in the literature [92, 100] and also revisited in the framework of evaluating measurement uncertainty [174]. The complete Bayesian model computes the difference d using equation (E17.8). The model as outlined for μ and σ^2 is used for both weighing sequences parameters respectively (μ_1, σ_1^2) and (μ_2, σ_2^2). The full model has been coded as shown in listing E17.1.

Fitting the data shown in figure E17.1 yields the following results (all values are in μg):

```
Inference for Stan model: Model3.
4 chains, each with iter=10000; warmup=5000; thin=1;
post-warmup draws per chain=5000, total post-warmup draws=20000.
```

	mean	se_mean	sd	2.5%	97.5%	n_eff	Rhat
mu1	100151.305	0.001	0.217	100150.880	100151.737	41351	1
sigma1	5.674	0.001	0.153	5.385	5.981	38617	1
mu2	2.268	0.001	0.201	1.871	2.662	39273	1
sigma2	5.242	0.001	0.140	4.978	5.522	39697	1
diff	100149.036	0.001	0.298	100148.454	100149.628	40528	1

```
Samples were drawn using NUTS(diag_e) at Wed Jan 29 12:51:54 2020.
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).
```
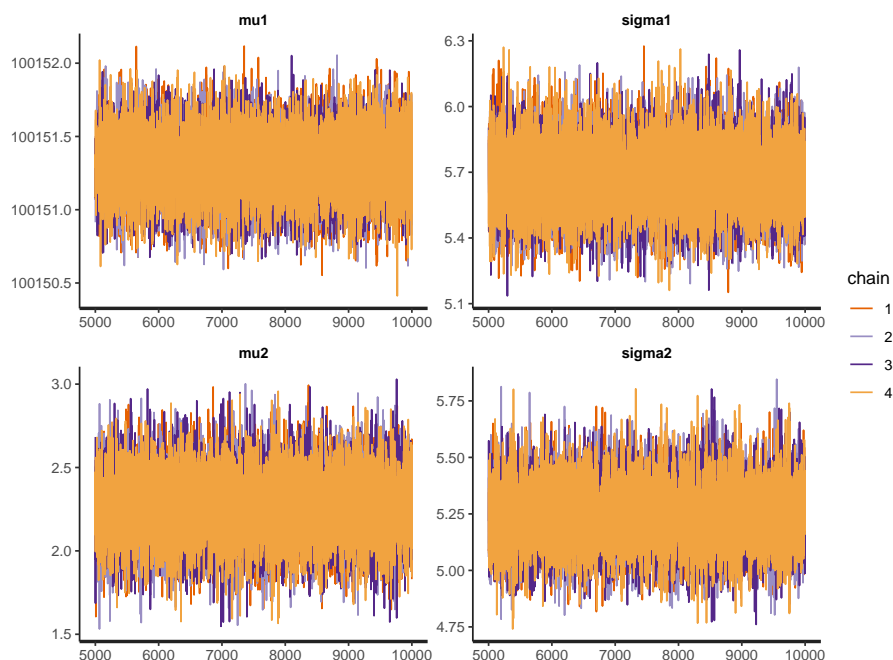

Figure E17.2: Trace plot of the four model parameters in the Bayesian model used to fit the weighing data

Four chains have been used with 10 000 iterations, of which the first 5000 have been used as warm-up of the sampler. The definitions and interpretation of the parameters $\hat{R}$ and the effective chain length n_{eff} in the Markov Chain Monte Carlo method have been discussed elsewhere [174] and are not repeated here. From the values of these parameters, it can be concluded that the fit of the model is satisfactory. This is confirmed by the trace plot of the four parameters (figure E17.2). The calculated mass difference is 100 149.036 µg with an associated standard uncertainty 0.298 µg. The mass of the 100 mg weight is 100.181 mg with associated standard uncertainty 0.005 mg (true mass). The relative deviation is 0.03 %, which is negligible in view of the uncertainty on the permeation rate. Hence, neither a correction is made, nor the uncertainty is incorporated in the calculations.

Uncertainty of the mass flow controller used for measuring flow rate of the dilution gas

The mass flow of the carrier gas is measured with a thermal mass flow controller. The mass flow of the dilution gas is constant and amounts to 250 mL min^{-1}, which corresponds to a mass flow rate of 0.3126 g min^{-1}. The required density of the gas can be computed using the method described in ISO 6976 [28,29]. The standard uncertainty of the volume flow rate is 0.2 %. For the mass flow rate, the same relative standard uncertainty applies, which is as absolute standard uncertainty 0.0006 g min^{-1}. The uncertainty of the gas density at reference conditions is ignored, as it is in the order of 0.02 % [29], which is negligible in view of the standard uncertainty associated with the mass flow rate.

Uncertainty in permeation rate

The permeation data are given at two temperatures, in tables E17.1 and E17.2. According to ISO 6145-10 [7], a number of observations shall be taken so that $u(\dot{m})/|\dot{m}| \approx 1\,\%$. In this example, this requirement corresponds to measurement data taken over an interval of approximately 20 min, which is a reasonable time span considering effects such as the stabilisation time of the permeation system. The calculated permeation rate is an average taken over such a time span.

Table E17.1: Measured values of time t, temperature T, pressure p and mass m of the permeation tube containing ammonia in the permeation chamber at nominally 52 °C

t/min	T/°C	p/mbar	m/g	t/min	T/°C	p/mbar	m/g
49.39	52.45	1018.1	16.467 067	58.44	52.45	1018.7	16.466 999
49.72	52.45	1018.1	16.467 063	58.77	52.45	1018.7	16.466 994
50.06	52.45	1018.2	16.467 059	59.11	52.45	1018.7	16.466 987
50.39	52.45	1018.2	16.467 055	59.44	52.45	1018.7	16.466 986
50.72	52.43	1018.1	16.467 056	59.77	52.45	1018.7	16.466 990
51.06	52.45	1018.1	16.467 044	60.11	52.45	1018.7	16.466 987
51.39	52.45	1018.2	16.467 046	61.63	52.45	1018.7	16.466 976
51.72	52.45	1018.1	16.467 047	61.97	52.45	1018.7	16.466 967
53.25	52.45	1018.7	16.467 035	62.30	52.45	1018.7	16.466 962
53.58	52.45	1018.7	16.467 032	62.63	52.45	1018.7	16.466 965
53.92	52.45	1018.7	16.467 029	62.97	52.45	1018.7	16.466 963
54.25	52.45	1018.7	16.467 026	63.30	52.45	1018.7	16.466 960
54.58	52.45	1018.7	16.467 021	63.63	52.45	1018.7	16.466 956
54.92	52.45	1018.7	16.467 022	63.97	52.45	1018.7	16.466 957
55.25	52.45	1018.7	16.467 021	64.30	52.45	1018.7	16.466 955
55.58	52.45	1018.7	16.467 013	65.83	52.45	1018.7	16.466 939
55.92	52.45	1018.7	16.467 016	66.16	52.45	1018.7	16.466 938
57.44	52.45	1018.7	16.467 003	66.49	52.45	1018.7	16.466 936
57.77	52.45	1018.7	16.467 002	66.83	52.43	1018.7	16.466 931
58.11	52.45	1018.7	16.467 001	67.16	52.45	1018.7	16.466 926

The resolution of the time measurement is 1 s. Hence, modelling this effect with the rectangular distribution leads to standard uncertainty of $1\,\text{s}/\sqrt{12} = 0.29\,\text{s}$. The resolution of the balance is 1 µg. The standard deviation of weighing a 10 g weight for approximately 40 hours is 2.9 µg. These standard uncertainties have been used in errors-in-variables (EIV) regression [6, 62]. This regression method is appropriate here, for it takes into consideration the uncertainties associated with the independent and dependent variables.

The permeation rate is determined using regression. The results using errors-in-variables regression, which takes into account the standard uncertainties associated with the time and mass measurements, are shown in figure E17.3. The top-left panel shows the permeation data and the bottom-left panel the same data with the fitted straight line. The residuals are shown in the right panels, the top figure displaying the residuals in the x-direction (time) and the bottom figure those in the y-direction (mass). Practically all residuals meet the consistency criteria of Deming regression [6, 78], namely that their absolute value does not exceed the expanded uncertainty. From the EIV regression, $\dot{m} = -13.599\,\mu\text{g min}^{-1}$ with standard uncertainty $u(\dot{m}) = 0.088\,\mu\text{g min}^{-1}$. Using OLS [81], $\dot{m} = -13.599\,\mu\text{g min}^{-1}$ with standard uncertainty $u(\dot{m}) = 0.108\,\mu\text{g min}^{-1}$. To the last stated digit, the calculated permeation rates concur, which

Table E17.2: Measured values of time t, temperature T, pressure p and mass m of the permeation tube containing ammonia in the permeation chamber at nominally 62 °C

t/min	T/°C	p/mbar	m/g	t/min	T/°C	p/mbar	m/g
1382.38	62.01	1014.3	16.449 565	1392.65	62.00	1014.2	16.449 430
1383.91	62.00	1014.3	16.449 544	1392.98	62.00	1014.2	16.449 424
1384.25	62.00	1014.3	16.449 540	1393.32	62.00	1014.2	16.449 418
1384.58	62.00	1014.4	16.449 526	1393.65	61.99	1014.2	16.449 418
1384.91	62.00	1014.3	16.449 526	1393.98	62.00	1014.2	16.449 413
1385.25	62.00	1014.3	16.449 528	1394.32	62.01	1014.2	16.449 407
1385.58	62.00	1014.3	16.449 529	1394.65	62.00	1014.2	16.449 404
1385.91	62.00	1014.3	16.449 522	1394.98	61.98	1014.1	16.449 395
1386.25	62.00	1014.3	16.449 518	1396.52	62.00	1014.1	16.449 374
1386.58	62.00	1014.3	16.449 513	1396.85	62.00	1014.1	16.449 369
1388.11	61.99	1014.2	16.449 489	1397.18	62.00	1014.1	16.449 365
1388.45	61.99	1014.2	16.449 491	1397.52	61.99	1014.1	16.449 359
1388.78	62.00	1014.2	16.449 482	1397.85	62.00	1014.1	16.449 356
1389.11	61.99	1014.2	16.449 477	1398.18	62.00	1014.1	16.449 349
1389.45	61.99	1014.2	16.449 473	1398.52	62.00	1014.1	16.449 345
1389.78	62.00	1014.2	16.449 464	1398.85	62.00	1014.2	16.449 339
1390.11	62.00	1014.2	16.449 468	1399.18	62.00	1014.1	16.449 339
1390.45	62.00	1014.2	16.449 465	1400.72	62.00	1014.2	16.449 318
1390.78	61.99	1014.2	16.449 457	1401.05	62.00	1014.2	16.449 317
1392.32	62.00	1014.2	16.449 434	1401.39	62.00	1014.2	16.449 308

implies that it is justified to use OLS.

For OLS, the built-in function LINEST in MS Excel provides identical values to the OLS results above for the slope and associated standard uncertainty. To obtain both the estimate and the standard uncertainty, LINEST should be called with the last two arguments equal to TRUE; the first one to allow for a non-zero intercept, the second for calculating additional statistics.

E17.5.3 Uncertainty due to materials purity

The squared standard uncertainty of x_{sj} is computed as

$$u^2(x_{sj}) = \sum_{\substack{i=1 \\ i \neq s}}^{q} u^2(x_{ij}). \tag{E17.9}$$

The expression for the covariance term of the amount fraction of the most abundant component and all other components in the parent gas j ($i \neq s$) is given by [171]

$$u(x_{sj}, x_{ij}) = -u^2(x_{ij}).$$

Both expressions follow from the applying the law of propagation of uncertainty to equation (E17.2). The composition of the materials used is summarised in table E17.3. The amount fractions ammonia (parent 1) and nitrogen have been computed using equation (E17.2) and the standard uncertainty has been obtained from equation (E17.9). The amount fractions methane, carbon monoxide, carbon dioxide and oxygen in nitrogen stem from the limit of quantification of the
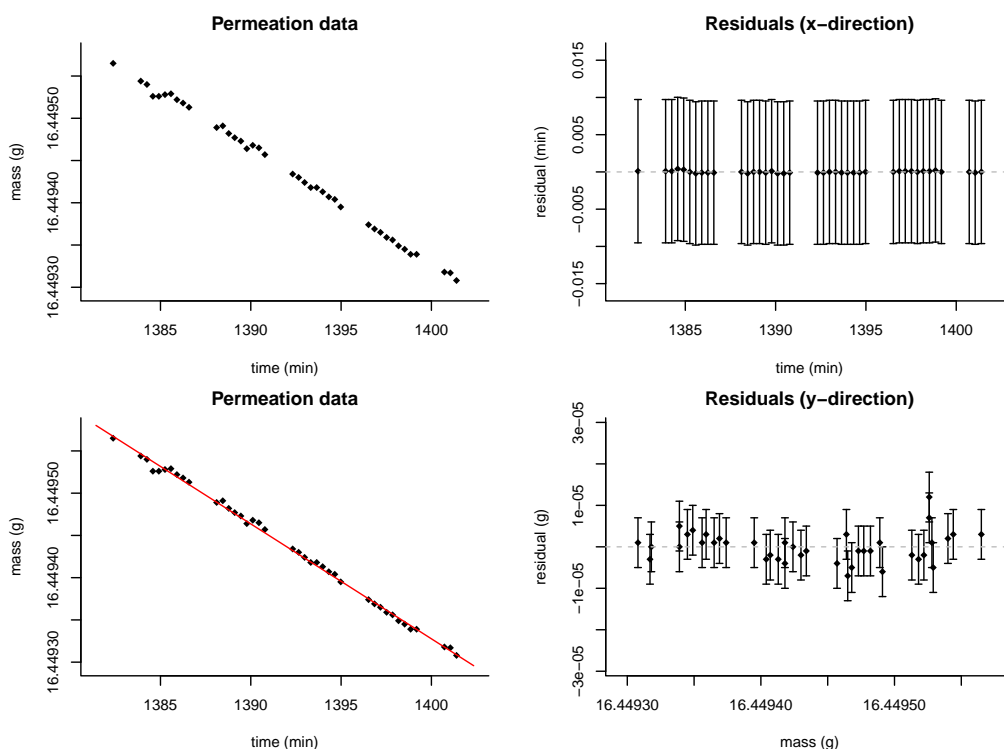
Figure E17.3: Results of the errors-in-variables regression of the permeation data at 62 °C from table E17.2

respective methods. The derivation of the values and standard uncertainties has been performed using the rectangular distribution in accordance with ISO 19229 [35].

Table E17.3: Purity information of ammonia (Parent 1) and nitrogen (Parent 2), expressed in amount fractions (mol mol^{-1})

Component	Parent 1		Parent 2	
	x	$u(x)$	x	$u(x)$
Argon			0.000 005	0.000 003
Methane			8.00×10^{-9}	5.00×10^{-9}
Carbon monoxide			1.50×10^{-8}	9.00×10^{-9}
Carbon dioxide			1.00×10^{-8}	6.00×10^{-9}
Hydrogen			2.50×10^{-8}	1.50×10^{-8}
Water	0.010	0.002	1.00×10^{-8}	6.00×10^{-9}
Nitrogen			0.999995	0.000 003
Oxygen			5.00×10^{-9}	3.00×10^{-9}
Ammonia	0.990	0.002		

E17.5.4 Uncertainty due to the molar masses

The squared standard uncertainty of the molar mass is computed by applying the law of propagation of uncertainty of the GUM [51] to equation (E17.3):

$$u^2(M_i) = \sum_{z=1}^{Z} v_{zi}^2 u^2(A_z), \tag{E17.10}$$

where $u(A_z)$ is obtained by using the rectangular distribution to model the uncertainty of the standard atomic weights [24].

ISO 6142-1 [24] does not take into account covariance between the molar masses of two components, yet according to [171] it can be calculated as:

$$u(M_i, M_j) = \sum_{z=1}^{L} v_{zi} v_{zj} u^2(A_z), \tag{E17.11}$$

where L denotes the number of atoms that components i and j have in common.

Molar masses of each component of ammonia (Parent 1) and nitrogen (Parent 2) have been calculated by using the standard atomic weights of elements (E17.3). As an illustration of equation (E17.11) we can consider the covariance between the molar masses of water and ammonia. The common element for these components is hydrogen (H), where its stochiometric number in molecular formula of water is $v_{zi} = 2$ and in ammonia $v_{zj} = 3$. The standard uncertainty of the standard atomic weight of hydrogen with the assumed rectangular probability distribution amounts to $u(A_z) = 0.00007803$, which gives $u(M_{H_2O}, M_{NH_3}) = 3.64 \times 10^{-8}$. The covariance matrix of the molar masses of each component has been computed as

$$
V_M \times 10^{-8} =
\begin{bmatrix}
33.3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 43.1 & 33.3 & 33.3 & 4.86 & 4.86 & 0 & 0 & 7.29 \\
0 & 33.3 & 37.9 & 42.5 & 0 & 4.56 & 0 & 9.13 & 0 \\
0 & 33.3 & 42.5 & 51.6 & 0 & 9.13 & 0 & 18.3 & 0 \\
0 & 4.86 & 0 & 0 & 2.43 & 2.43 & 0 & 0 & 3.64 \\
0 & 4.86 & 4.56 & 9.13 & 2.43 & 6.99 & 0 & 9.13 & 3.64 \\
0 & 0 & 0 & 0 & 0 & 0 & 2.41 & 0 & 12.0 \\
0 & 0 & 9.13 & 18.3 & 0 & 9.13 & 0 & 18.3 & 0 \\
0 & 7.29 & 0 & 0 & 3.64 & 3.64 & 12.0 & 0 & 11.5
\end{bmatrix},
\tag{E17.12}
$$

where each diagonal element represents the squared standard uncertainty of molar masses and otherwise the covariance between the molar masses of two components.

The components themselves are ordered into rows and columns according to their distribution shown in table E17.3. The matrix multiplication can be carried out in MS Excel by using the in-built function MMULT.

The molar masses of components of parent gases, accompanied by associated standard uncertainties, are given in table E17.4.

The squared standard uncertainty associated with the molar mass of parent gas j is calculated using the law of propagation of uncertainty for correlated input quantities of the GUM [51, eqn (13)]:

$$u^2(\bar{M}_j) = \sum_{i=1}^{q} x_{ij}^2 u^2(M_i) + \sum_{i=1}^{q} M_i^2 u^2(x_{ij}) + 2\sum_{i=1}^{q-1}\sum_{k=i+1}^{q} x_{ij} x_{kj} u(M_i, M_k) + 2\sum_{i=1}^{q-1}\sum_{k=i+1}^{q} M_i M_k u(x_{ij}, x_{kj}),$$

where x_{ij} denotes the amount fraction of component i in parent gas j.

Table E17.4: Molar masses of components of parent gases ammonia and nitrogen and their standard uncertainties

Component	M_i	$u(M_i)$
Argon	39.948	5.77×10^{-4}
Methane	16.0425	6.56×10^{-4}
Carbon monoxide	28.01000	6.16×10^{-4}
Carbon dioxide	44.0094	7.18×10^{-4}
Hydrogen	2.01595	1.56×10^{-4}
Water	18.01535	2.64×10^{-4}
Nitrogen	28.01371	4.91×10^{-4}
Oxygen	31.9988	4.27×10^{-4}
Ammonia	17.03078	3.39×10^{-4}

The expression for the covariance between $\bar{M}_1$ and $\bar{M}_2$ can be derived as follows. Let

$$\bar{M}_1 = \sum_{i=1}^{q} x_{i1} M_i,$$

$$\bar{M}_2 = \sum_{i=1}^{q} x_{i2} M_i.$$

Given that the M_i are correlated, we need to generalise equation (F.1.2) of the GUM [51] to enable working with correlated variables. Suppose we have two functions $Y = f(X_1, \ldots, X_N)$ and $Z = g(X_1, \ldots, X_N)$; then [47]

$$u(Y, Z) = \sum_i \frac{\partial f}{\partial X_i} \frac{\partial g}{\partial X_i} u^2(X_i) + \sum_i \sum_{j \neq i} \frac{\partial f}{\partial X_i} \frac{\partial g}{\partial X_j} u(X_i, X_j). \tag{E17.13}$$

Noting that the molar masses are the shared variables in the expressions for $\bar{M}_1$ and $\bar{M}_2$,

$$u(\bar{M}_1, \bar{M}_2) = \sum_i x_{i1} x_{i2} u^2(M_i) + \sum_i \sum_k x_{i1} x_{k2} u(M_i, M_k).$$

This expression could also have been obtained using the law of propagation of uncertainty for the explicit, multivariate measurement model of GUM-S2 [54], where the above expressions for $\bar{M}_1$ and $\bar{M}_2$ form the multivariate measurement model. The covariance matrix contains $u^2(\bar{M}_1)$, $u^2(\bar{M}_2)$ and $u(\bar{M}_1, \bar{M}_2)$. The molar masses of parent gases have been calculated and amount to $\bar{M}_1 = 17.04063 \, \text{g mol}^{-1}$ for ammonia and $\bar{M}_1 = 28.01377 \, \text{g mol}^{-1}$ for nitrogen. The associated squared standard uncertainties and covariance have been computed as $u^2(\bar{M}_1) = 3.99 \times 10^{-6} \, \text{g}^2 \, \text{mol}^{-2}$, $u^2(\bar{M}_2) = 2.42 \times 10^{-7} \, \text{g}^2 \, \text{mol}^{-2}$ and $u(\bar{M}_1, \bar{M}_2) = 1.19 \times 10^{-7} \, \text{g}^2 \, \text{mol}^{-2}$. Hence, the correlation coefficient is $r(\bar{M}_1, \bar{M}_2) = 0.121$, which implies a weak correlation between the molar masses of the two parent gases.

E17.6 Reporting the result

The amount fraction NH$_3$ is $70.8 \, \mu\text{mol mol}^{-1}$ with standard uncertainty $0.6 \, \mu\text{mol mol}^{-1}$. For the components that have been added intentionally (in this example NH$_3$ and N$_2$), it is appropriate

to assume the normal distribution or t distribution to obtain a coverage factor to calculate the expanded uncertainty. Hence, the result can be stated as $(70.8 \pm 1.2)\,\mu\text{mol mol}^{-1}$ ($k = 2$) with 95 % probability.

The full composition of the calibration gas mixture is given in table E17.5, computed with the measurement model from ISO 6142-1 [24]. The measurement model in this example only provides the amount fraction, which is also given in table E17.5.

Table E17.5: Composition of a calibration gas mixture of NH$_3$ in N$_2$, expressed in amount fractions (mol mol^{-1})

Component	y	$u(y)$	$u(y)$
Argon	5×10^{-6}	3×10^{-6}	60.0 %
Methane	8×10^{-9}	5×10^{-9}	62.5 %
Carbon monoxide	1.5×10^{-8}	9×10^{-9}	60.0 %
Carbon dioxide	1×10^{-8}	6×10^{-9}	60.0 %
Hydrogen	2.5×10^{-8}	1.5×10^{-8}	60.0 %
Water	7.25×10^{-7}	1.43×10^{-7}	19.7 %
Nitrogen	0.999923	0.000003	0.0003%
Oxygen	5×10^{-9}	3×10^{-9}	60.0 %
Ammonia	7.08×10^{-5}	5.99×10^{-7}	0.8%

E17.7   Interpretation of results

Experience has shown that as far as the amount fractions of abundant components are concerned, the use of the law of propagation of uncertainty suffices to obtain an estimate and standard uncertainty of the amount fraction of the component of interest. Also the establishment of coverage intervals for these amount fractions using the normal or t distribution is appropriate. For trace components, it can be necessary to use another probability density function for establishing coverage intervals, such as the beta distribution [35, 175].

In this example, we have shown how the use of OLS can be justified. The principal question was whether the uncertainty associated with the recorded time measurements could be ignored in the calculation of the permeation rate. The resolution of the time measurement is 1 s (see tables E17.1 and E17.2). The comparison between the permeation rates obtained by EIV and OLS shows that the simpler method can be used.

The method of total differentials as employed in section E17.5.2 is a convenient way to obtain the expressions for the partial derivatives. It is particularly useful in multistage measurement models, as it eases the use of the chain rule of differentiation. It can be shown that the chain rule of differentiation can be readily implemented by substituting differentials [58]. The resulting expressions for the sensitivity coefficients are compact. When applying this method to a multivariate measurement model (such as the model in ISO 6142-1 [24] to compute the amount fractions of all components in the calibration gas mixture), the chain rule of differentiation can be implemented in the form of a matrix multiplication [171].

Example E18

Pressure drop measurement

M. Čaušević, M.G. Cox, J. Greenwood

E18.1   Summary

This example demonstrates how correlations can sometimes be removed from an uncertainty evaluation by modifying the measurement equation so that strongly correlated quantity estimates do not appear together. This is demonstrated by an example considering the pressure drop in a pressurised vessel, in which the effect of correlation between temperature measurements is removed. In this example the uncertainty associated with an estimate of pressure drop is appreciably smaller when correlation is taken into consideration.

E18.2   Introduction of the application

In every kind of measurements involving gases, it is necessary to have a controlled situation in terms of gas temperature, pressure and volume. This example is derived from a real-world test that involves the gas at a pressure up to 50 MPa trapped in a vessel having small volume at standard temperature. Probes to measure the pressure and temperature were fixed inside the sealed vessel. The aim of the test was to show that the uncertainty due to gas leakage, i.e. the pressure drop in the vessel, can vary for two cases that involved the same measurement results but different approaches of evaluation. All correlations within this example arise from the application of the gas equation, which is a good approximation of the behaviour of many gases under many conditions, although it has some limitations [42].

E18.3   Specification of the measurand(s)

The measurand is the pressure drop in a pressurised vessel and depends on several physical quantities: pressure, temperature and time at the beginning and end of the test. Measured values of these quantities and their associated standard uncertainties are based upon actual measurements. The period over which the test was performed was 1800 s. Measured values are corrected to a standard temperature, that is, the reference temperature of 20 °C. The mathematical model for pressure drop, which gives the relation between the measurand and all influence quantities is explained in the following section.

   The aim of the example is to show how the use of a temperature difference as opposed to the temperature at the end of the test leads to a much smaller uncertainty.

E18.4   Measurement model

E18.4.1   Measurement principle

The measurement model in this example originates from the gas equation, which is the equation of state of the 'ideal' gas [122]:

$$pV = nRT, \tag{E18.1}$$

where

- p is the pressure in the vessel,
- V is the volume of the vessel,
- n is the number of moles of gas,
- R is the gas constant,
- T is the absolute temperature of the gas.

To obtain the pressure drop in the vessel during the period of the test, we observe quantities in equation (E18.1) with respect to times t_1 and t_2 at the beginning and end of the test. Number of moles, i.e amount of the gas at the start time (t_1), can be expressed as

$$n_1 = \frac{V p_1}{R T_1} \tag{E18.2}$$

and accordingly the amount of gas at the end time (t_2) is

$$n_2 = \frac{V p_2}{R T_2}. \tag{E18.3}$$

Loss of material in the form of gas escaping from the system is therefore provided by the difference

$$\Delta n = n_2 - n_1 = \frac{V}{R}\left(\frac{p_2}{T_2} - \frac{p_1}{T_1}\right). \tag{E18.4}$$

Since pressure change is one of the best indicators of gas leakage in the sealed system it would be useful to show the relation between the number of moles and the pressure change, i.e pressure drop at some reference temperature T_s:

$$\Delta p_{\mathrm{s}} = \Delta n \frac{R}{V} T_{\mathrm{s}}. \tag{E18.5}$$

By substituting equation (E18.4) into (E18.5) we obtain

$$\Delta p_{\mathrm{s}} = \left(\frac{p_2}{T_2} - \frac{p_1}{T_1}\right) T_{\mathrm{s}}. \tag{E18.6}$$

The resulting change of pressure occurring over the period $\Delta t = t_2 - t_1$ can be scaled to some reference interval Δt_s; thus the equation (E18.6) becomes

$$\Delta p_{\mathrm{s}} = \left(\frac{p_2}{T_2} - \frac{p_1}{T_1}\right) T_{\mathrm{s}} \frac{\Delta t_{\mathrm{s}}}{\Delta t}. \tag{E18.7}$$

For test items of a predefined volume, pressure change is often a sufficient measurand. Therefore, expression (E18.7) is the measurement model we will consider.

In general, for arbitrary volumes, pressure change does not fully specify the size of a leak in the way that loss of material Δn does.

More commonly a leak-rate is evaluated in terms of rate of mass flow

$$Q_m = \frac{V}{MR}\left(\frac{p_2}{T_2} - \frac{p_1}{T_1}\right)\frac{1}{\Delta t}, \tag{E18.8}$$

where M represents the molar mass of the gas. Alternatively, a leak rate might be expressed in terms of rate of change of volume for gas at a reference temperature T_s and pressure P_s:

$$Q_v = V\left(\frac{p_2}{T_2} - \frac{p_1}{T_1}\right)\frac{T_s}{p_s}. \tag{E18.9}$$

The arguments presented shortly apply equally well to flow rate measurement equations (E18.8) and (E18.9).

E18.4.2   Correlations in the measurement model

Whenever estimates of a quantity are measured using a common process or common equipment (such as in a comparison of values 'before' and 'after' some change) there is a possibility of correlation between the estimates. In this example correlations could exist between the pressure measurements, between the temperature measurements and between the time measurements. These correlations could arise from a variety of sources such as common errors in traceability, or from metrological effects such as the positioning of gauges, or from gradients in temperature between the location of the measurement and the location of interest.

Where there are significant correlations present, the normal process is to follow the procedure described in GUM Annex F.1.2. However, if the equation can be modified and expressed in such a way that the correlation can be removed, then the more straightforward approach offered by GUM equation (10) can be followed.

For example, if the same measuring instrument were used for obtaining two measured values,

$$y_1 = x_1 + s,$$
$$y_2 = x_2 + s,$$

where the y_i denotes corrected values, the x_i and s represent measured values and systematic error, respectively. By differencing y_1 and y_2 the common systematic error s cancels.

In this example we concentrate on temperature correlation, since this is found to have the most significant effect. Therefore it is worthwhile re-expressing (E18.7) in terms of T_1 and the temperature *difference*

$$\Delta T = T_2 - T_1, \tag{E18.10}$$

in other words, to eliminate T_2, which is strongly correlated with T_1, from explicit consideration. Such an approach prevents any common systematic error appearing multiple times in the measurement equation. The standard uncertainty associated with the temperature difference is much smaller than that associated with a single temperature, that is,

$$u(\Delta T) \ll u(T_1). \tag{E18.11}$$

Making use of (E18.10) and (E18.11), expression (E18.7) becomes

$$\Delta p_s = \left(\frac{T_1 p_2 - T_2 p_1}{T_2 T_1}\right)\frac{T_s}{\Delta t}\Delta t_s \tag{E18.12}$$

$$\approx \left(\frac{p_2 - p_1}{T_1} - \frac{\Delta T p_1}{T_1^2}\right)\frac{T_s}{\Delta t}\Delta t_s. \tag{E18.13}$$

E18.5   Uncertainty propagation

E18.5.1   Law of propagation of uncertainty

For the measurement uncertainty evaluation, the GUM [51] law of propagation of uncertainty (LPU) is applied for the two measurement equations (E18.7) and (E18.13). As described above, the first equation takes no account of correlation, and the second accounts for correlation by expressing the second temperature term as the sum of the first temperature term and the difference between the two terms. Use of equation (E18.13), knowing the standard uncertainty associated with the difference, is shown to decrease significantly the measurement uncertainty associated with the estimated pressure drop.

The mathematical models presented in equations (E18.7) and (E18.13) will be used as a starting point for evaluation of uncertainties.

Both approaches distinguish between Type A (u_A) and Type B (u_B) uncertainty evaluations, where Type A uncertainty is the standard deviation associated with repeatability of measurements. Three consecutive measurements were performed with the same results used for both approaches.

The probability distributions in this example are assigned based on knowledge concerning the measurement.

E18.5.2   Ignoring temperature correlation

In this subsection correlation between temperature values is ignored, that is, it is taken as zero. For model (E18.7) the sensitivity coefficients are

$$\frac{\partial \Delta p_s}{\partial p_1} = -\frac{1}{T_1}\frac{T_s}{\Delta t}\Delta t_s = -\frac{T_s \Delta t_s}{T_1 \Delta t}, \tag{E18.14}$$

$$\frac{\partial \Delta p_s}{\partial p_2} = \frac{1}{T_2}\frac{T_s}{\Delta t}\Delta t_s = \frac{T_s \Delta t_s}{T_2 \Delta t}, \tag{E18.15}$$

$$\frac{\partial \Delta p_s}{\partial T_2} = -\frac{p_2}{T_2{}^2}\frac{T_s}{\Delta t}\Delta t_s = -\frac{p_2 T_s \Delta t_s}{T_2{}^2 \Delta t}, \tag{E18.16}$$

$$\frac{\partial \Delta p_s}{\partial T_1} = \frac{p_1}{T_1{}^2}\frac{T_s}{\Delta t}\Delta t_s = \frac{p_1 T_s \Delta t_s}{T_1{}^2 \Delta t}, \tag{E18.17}$$

$$\frac{\partial \Delta p_s}{\partial t_1} = \left(\frac{p_2}{T_2} - \frac{p_1}{T_1}\right)\frac{T_s}{(\Delta t)^2}\Delta t_s = \frac{\Delta p_s}{\Delta t}, \tag{E18.18}$$

$$\frac{\partial \Delta p_s}{\partial t_1} = \left(\frac{p_2}{T_2} - \frac{p_1}{T_1}\right)\frac{T_s}{(\Delta t)^2}\Delta t_s = -\frac{\Delta p_s}{\Delta t}. \tag{E18.19}$$

The standard measurement uncertainty of individual measurement quantities can be determined as the standard deviations of the according rectangular or normal probability distributions. It is assumed that the expanded uncertainties, denoted by U, for pressure measurements are based on normal probability distributions. Expected values of temperature and time lie within intervals whose lengths are denoted by $2a_T$ and $2a_t$, respectively, and rectangular distributions are

assumed for both quantities. The corresponding standard uncertainties are therefore

$$u(p_1) = \frac{U(p_1)}{2}, \tag{E18.20}$$

$$u(p_2) = \frac{U(p_2)}{2}, \tag{E18.21}$$

$$u(T_2) = \frac{a_T}{\sqrt{3}}, \tag{E18.22}$$

$$u(T_1) = \frac{a_T}{\sqrt{3}}, \tag{E18.23}$$

$$u(t_1) = \frac{a_t}{\sqrt{3}}, \tag{E18.24}$$

$$u(t_2) = \frac{a_t}{\sqrt{3}}. \tag{E18.25}$$

All the results obtained from equations (E18.14)–(E18.25) can be combined into a Type B measurement uncertainty evaluation for the pressure drop in the vessel:

$$u_B^2 = \left[\frac{\partial \Delta p_s}{\partial p_1} u(p_1)\right]^2 + \left[\frac{\partial \Delta p_s}{\partial P_2} u(p_2)\right]^2 + \left[\frac{\partial \Delta p_s}{\partial T_2} u(T_2)\right]^2$$
$$+ \left[\frac{\partial \Delta p_s}{\partial T_1} u(T_1)\right]^2 + \left[\frac{\partial \Delta p_s}{\partial t_1} u(t_1)\right]^2 + \left[\frac{\partial \Delta p_s}{\partial t_2} u(t_2)\right]^2. \tag{E18.26}$$

The combined measurement uncertainty u_c for the pressure drop in the vessel is obtained from

$$u_c^2 = u_A^2 + u_B^2. \tag{E18.27}$$

E18.5.3 Accounting for temperature correlation

In this subsection correlation between temperature values is taken into consideration through use of the model presented in equation (E18.13). The sensitivity coefficients are as follows:

$$\frac{\partial \Delta p_s}{\partial p_1} = -\left(\frac{1}{T_1} - \frac{\Delta T}{T_1^2}\right)\frac{T_s}{\Delta t}\Delta t_s = -\frac{(T_1 + \Delta T)T_s \Delta t_s}{T_1^2 \Delta t} \approx -\frac{T_s \Delta t_s}{T_1 \Delta t}, \tag{E18.28}$$

$$\frac{\partial \Delta p_s}{\partial p_2} = \frac{1}{T_1}\frac{T_s}{\Delta t}\Delta t_s = \frac{T_s \Delta t_s}{T_1 \Delta t}, \tag{E18.29}$$

$$\frac{\partial \Delta p_s}{\partial \Delta T} = -\frac{p_1}{T_1^2}\frac{T_s}{\Delta t}\Delta t_s = -\frac{p_1 T_s \Delta t_s}{T_1^2 \Delta t}, \tag{E18.30}$$

$$\frac{\partial \Delta p_s}{\partial T_1} = \left(\frac{p_1 - p_2}{T_1^2} + \frac{2T_1 \Delta T p_1}{T_1^4}\right)\frac{T_s}{\Delta t}\Delta t_s = \frac{T_s \Delta t_s[T_1(p_1 - p_2) + 2\Delta T p_1]}{T_1^3 \Delta t}, \tag{E18.31}$$

$$\frac{\partial \Delta p_s}{\partial t_1} = -\left(\frac{p_2 - p_1}{T_1} - \frac{\Delta T p_1}{T_1^2}\right)\frac{T_s}{(\Delta t)^2}\Delta t_s = -\frac{\Delta p_s}{\Delta t}, \tag{E18.32}$$

$$\frac{\partial \Delta p_s}{\partial t_2} = \left(\frac{p_2 - p_1}{T_1} - \frac{\Delta T p_1}{T_1^2}\right)\frac{T_s}{(\Delta t)^2}\Delta t_s = \frac{\Delta p_s}{\Delta t}. \tag{E18.33}$$

The standard measurement uncertainty of individual measurement quantities can be determined as the standard deviations of the according rectangular, triangular or normal probability distributions. As in the previous case, when the correlation between temperatures are not taken into account, it is assumed that pressure measurements have normal probability distributions and temperature and time measurements have rectangular. In case of temperature difference,

the length of the interval in which ΔT is expected to lie is denoted by $2a_{\Delta T}$. It is assumed that the temperature difference has a triangular distribution based on the earlier assumption of rectangular distributions for T_1 and T_2. Hence,

$$u(p_1) = \frac{U(p_1)}{2}, \tag{E18.34}$$

$$u(p_2) = \frac{U(p_2)}{2}, \tag{E18.35}$$

$$u(\Delta T) = \frac{a_{\Delta T}}{\sqrt{6}}, \tag{E18.36}$$

$$u(T_1) = \frac{a_T}{\sqrt{3}}, \tag{E18.37}$$

$$u(t_1) = \frac{a_t}{\sqrt{3}}, \tag{E18.38}$$

$$u(t_2) = \frac{a_t}{\sqrt{3}}. \tag{E18.39}$$

All the values obtained from equations (E18.28)–(E18.39) can be combined into a Type B measurement uncertainty evaluation for the pressure drop in the vessel:

$$
\begin{aligned}
u_{\mathrm{B}}^2 = &\left[\frac{\partial \Delta p_s}{\partial p_1} u(p_1) \right]^2 + \left[\frac{\partial \Delta p_s}{\partial p_2} u(p_2) \right]^2 + \left[\frac{\partial \Delta p_s}{\partial \Delta T} u(\Delta T) \right]^2 \\
&+ \left[\frac{\partial \Delta p_s}{\partial T_1} u(T_1) \right]^2 + \left[\frac{\partial \Delta p_s}{\partial t_1} u(t_1) \right]^2 + \left[\frac{\partial \Delta p_s}{\partial t_2} u(t_2) \right]^2.
\end{aligned}
\tag{E18.40}
$$

The combined standard uncertainty for the pressure drop in the vessel is obtained from equation (E18.27).

E18.6   Reporting the result

During the measurement the data in table E18.1 were obtained (values are given as provided). Uncertainties of individual quantities are taken from uncertainty budgets for the use of the pressure, temperature and time measuring equipment. The temperature uncertainty includes a dominant contribution with limits of about 0.15 K for a common (for all such measurements) but poorly understood systematic effect due to temperature gradients between the thermometer and the gas in the vessel.

The results of the second approach are presented in table E18.2.

Table E18.1: Measured values including values of the pressure drop, assigned probability density functions and combined standard uncertainty based upon application of LPU to equation (E18.7)

Quantity	Value		Std. uncertainty		PDF	Sensitivity coefficient		Variance	
p_1	50.001	MPa	0.0056	MPa	norm.	-1.0010		7.856×10^{-6}	MPa^2
p_2	49.951	MPa	0.0056	MPa	norm.	1.0007		7.851×10^{-6}	MPa^2
T_1	292.850	K	0.15	K	rect.	0.1709	MPa K^{-1}	2.191×10^{-4}	MPa^2
T_2	292.950	K	0.15	K	rect.	-0.1706	MPa K^{-1}	2.184×10^{-4}	MPa^2
t_1	1000	s	5	s	rect.	-3.73×10^{-5}	MPa s^{-1}	1.159×10^{-8}	MPa^2
t_2	2800	s	5	s	rect.	3.73×10^{-5}	MPa s^{-1}	1.159×10^{-8}	MPa^2
T_s	293.150	K							
t_s	1800	s							
Δp_s	$-0.067\,12$	MPa							
u_A			0.0025	MPa	norm.	1		0.0025	MPa
u_c								0.0214	MPa
$U(k=2)$								0.0429	MPa

Table E18.2: Measured values including values of the pressure drop, assigned probability density functions and combined standard uncertainty based upon application of LPU to equation (E18.13)

Quantity	Value		Std. uncertainty		PDF	Sensitivity coefficient		Variance	
p_1	50.001	MPa	0.0056	MPa	norm.	-1.0010		7.856×10^{-6}	MPa^2
p_2	49.951	MPa	0.0056	MPa	norm.	1.0007		7.851×10^{-6}	MPa^2
T_1	292.850	K	0.15	K	rect.	0.0030	MPa K^{-1}	6.205×10^{-10}	MPa^2
ΔT	0.100	K	0.02	K	trian.	-0.1709	MPa K^{-1}	1.947×10^{-6}	MPa^2
t_1	1000	s	5	s	rect.	-3.73×10^{-5}	MPa s^{-1}	1.159×10^{-8}	MPa^2
t_2	2800	s	5	s	rect.	3.73×10^{-5}	MPa s^{-1}	1.159×10^{-8}	MPa^2
T_s	293.150	K							
t_s	1800	s							
Δp_s	$-0.067\,14$	MPa							
u_A			0.0025	MPa	norm.	1		0.0025	MPa
u_c								0.0049	MPa
$U(k=2)$								0.0098	MPa

E18.7   Interpretation of results

It can be seen from the results in section E18.6 that both equations (E18.7) and (E18.13) yield essentially the same normalized pressure drop ($-0.067\,12$ MPa and $-0.067\,14$ MPa). An extra decimal digit is quoted beyond that for the associated standard uncertainties (0.0429 MPa and 0.0049 MPa, respectively) to show the difference is two units in the fifth digit, some two orders of magnitude smaller than these uncertainties. However, these standard uncertainties are significantly different from each other, being a factor four smaller when correlation is taken into account. Thus, on the basis of the data used, some means of treating the correlation is necessary to establish a reliable measurement uncertainty.

The question may arise whether it is also necessary to take into account correlation between pressures at the beginning and end of the test. In this example we decided to consider only the correlation between temperature terms, since the uncertainty due to pressure measurements is appreciably lower than that due to temperature measurements.

Besides the GUM approach [51], the Monte Carlo method described within GUM Supple-

ment 1 (GUM-S1) [52] was also used for the evaluation of the measurement uncertainty for the measurement models described by equations (E18.7) and (E18.13). GUM-S1 recommends application of the Monte Carlo method approach to validate the use of the GUM uncertainty framework. In this example, in order to obtain an estimate of the output quantity, a value for the associated standard uncertainty and endpoints of the shortest coverage interval, the number of Monte Carlo trials used was 10^7. The MATLAB programming language was used to perform the calculations for the Monte Carlo method.

Input quantities, together with associated standard uncertainties and assumptions on the probability density function for each input quantity remained the same as listed in table E18.1 and E18.2 in section E18.6. The results of the GUM uncertainty framework and the Monte Carlo method are presented in tables E18.3 and E18.4. Table E18.3 shows results for the first approach in which the temperature correlation was ignored and table E18.4 shows results for the second approach.

Table E18.3: Results obtained by the GUM uncertainty framework (GUF) and the Monte Carlo method (MCM) not accounting for temperature correlation

Approach	Estimate/bar	Std. unc./bar	CI (95%)/bar
GUF	−0.067120	0.02143	(−0.10999, 0.02425)
MCM	−0.067118	0.02143	(−0.10789, 0.02634)

Table E18.4: Results obtained by the GUM uncertainty framework (GUF) and the Monte Carlo method (MCM) accounting for temperature correlation

Approach	Estimate/bar	Std. unc./bar	CI (95%)/bar
GUF	−0.067143	0.00489	(−0.07693, 0.05735)
MCM	−0.067485	0.00471	(−0.07671, −0.00582)

Summary statistics obtained with GUM uncertainty framework and Monte Carlo method do not differ greatly.

Further information can be obtained by considering the full PDF as shown in figures E18.1 and E18.2 (note the different scales).
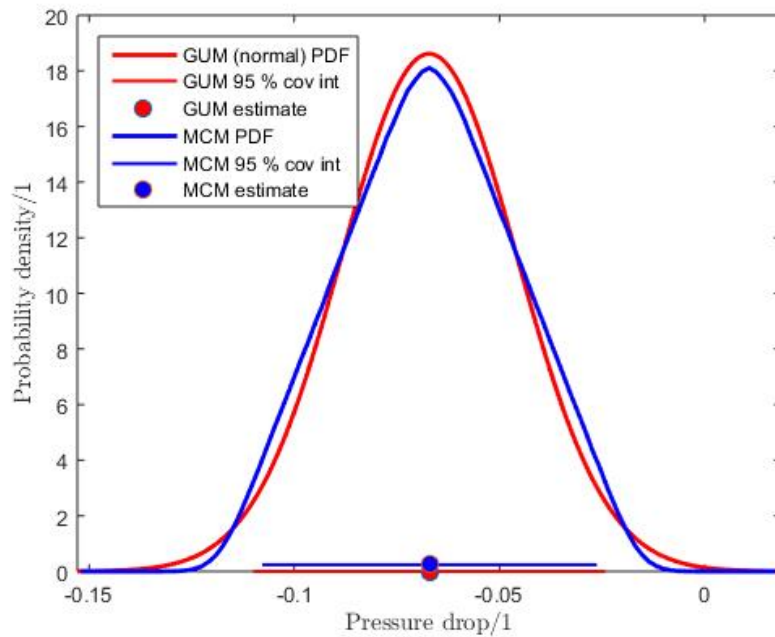
Figure E18.1: Probability density functions for pressure drop obtained by the GUM uncertainty framework and the Monte Carlo method not accounting for temperature correlation
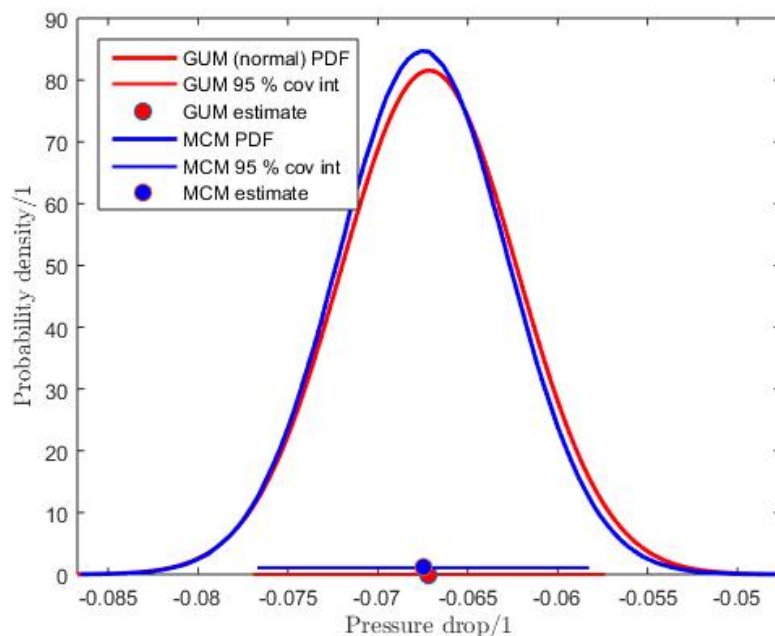

Figure E18.2: Probability density functions for pressure drop obtained by the GUM uncertainty framework and the Monte Carlo method accounting for temperature correlation (note the different scale from figure E18.1)

Figures E18.1 and E18.2 show the Gaussian PDF for pressure drop (red line) resulting from the GUM uncertainty framework. They also show the PDF for the pressure drop obtained as a

result of scaled frequency distribution of $M = 1 \times 10^7$ Monte Carlo trials (blue line). The endpoints of the probabilistically symmetric 95 % coverage interval provided by both methods and both approaches are shown as horizontal lines. Probability density functions, coverage intervals and pressure drop estimates are visually very similar. It can be noticed that the slope of the PDF on figure E18.1 as a result of the Monte Carlo method is steeper than that of the GUM method. The MCM curve agrees well with a triangular distribution. On the other hand, the shape of the PDFs as result of both methods (GUM and MCM) in figure E18.2 show very good agreement. The reason for this could be the correlation between the measured temperature at the beginning and end of the test.

According to [51], if the measurement model is linear in the input quantities and the dominant contributions have normal probability distributions, the GUM uncertainty framework (GUF) will provide reliable results. Even though the measurement models for both approaches (ignoring and accounting for the temperature correlation) were non-linear, both the MCM and GUM provided similar results (see table E18.3 and table E18.4). The Monte Carlo method in this example was used for the purpose of analysis of results and validation of the GUM uncertainty framework, which was successfully achieved.

Example E19

Flow meter calibration using the master meter method

M. Čaušević, M.G. Cox, A.M.H. van der Veen

E19.1 Summary

This example demonstrates the calibration of a gas flow measuring instrument by the so-called "master meter" method, i.e. by comparing the measured flow on a master meter (reference standard) and the measured flow on the device under test. The measurements in this example were performed by using three measurement standards with different measuring ranges and one device under test in the "SARAJEVOGAS" Laboratory. The measurements were performed at 10 different flow rates, where each flow rate was measured three times, which gives in total 30 measurements of flow rate. As a result, this example gives the uncertainty of measurement of the meter under test at each of ten flow rates within this set-up.

E19.2 Introduction of the application

The test facility operates on the so-called "master meter" principle where the meter under test (MUT) is located downstream from the standard meter (figure E19.1). Ambient air is sucked by a fan and the flow rate is adjusted by regulation of the fan and electromotive valve. The testing procedure is controlled by software. The measurement of flow rate for this kind of set-up first
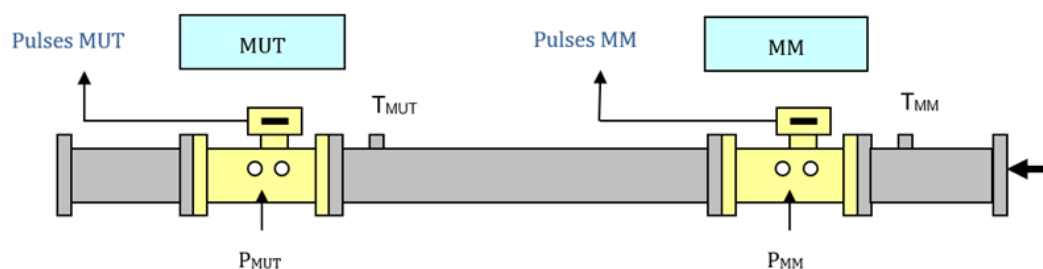

Figure E19.1: Set-up in the SARAJEVOGAS laboratory

starts with entering the desired flow rate into the flow computer. After the first recorded pulse from the MUT, the volume flow rate from the MUT and the reference measurement standard (master meter or MM) are measured and recorded separately on the indicating devices of these

measuring instruments. After two or more MUT pulses (depending on the selected volume) the measurement stops automatically. The volume flow rates from the MUT and the MM are calculated by dividing the number of pulses by the pulse value for each measuring instrument. Figure E19.1 shows the location of the master meter, the meter under test and the measuring instruments for temperature and pressure measurements in the laboratory set-up.

E19.3 Specification of the measurand(s)

The measurement, which in this case was for calibration purposes, was performed at atmospheric conditions with air temperature around 22 °C. The absolute pressure was measured directly with the standard and the meter under test, while the temperatures were measured downstream. Single tests lasted a minimum of 200 s to reach a stable flow rate. The calibration was performed with three standard/master meters with the following measuring ranges given in table E19.1.

Table E19.1: Volume flow rate ranges of the meters involved

G40 Rotary gas meter	G250 Turbine gas meter	G1000 Turbine gas meter
$20\,\mathrm{m^3\,h^{-1}}$–$50\,\mathrm{m^3\,h^{-1}}$	$100\,\mathrm{m^3\,h^{-1}}$–$350\,\mathrm{m^3\,h^{-1}}$	$450\,\mathrm{m^3\,h^{-1}}$–$1000\,\mathrm{m^3\,h^{-1}}$

E19.4 Measurement model

E19.4.1 Main effects

The basic procedure within this example differentiates between two types of quantities that influence the measurement uncertainty. The first type refers to the measurement error of the meter under test and the second type to the measurement standard, repeatability of measurement as well as any other additional influence quantity. The measurement error of the device under test is considered to be the main effect since it includes measurement effects of pressure, temperature and impulses.

The mathematical model for the measurement error of the MUT can be expressed as follows [79]:

$$e = \frac{V_{\mathrm{MUT}} - V_{\mathrm{REF}}}{V_{\mathrm{REF}}} \tag{E19.1}$$

$$= \frac{V_{\mathrm{MUT}}}{V_{\mathrm{REF}}} - 1, \tag{E19.2}$$

where

e is the measurement error of the MUT,

V_{MUT} is the volume of the gas flow that is measured with the MUT,

V_{REF} is the reference volume, i.e. the volume of the gas flow measured with the MM.

The reference volume V_{REF} is not the same as V_{MM}, which is the volume of the gas flow measured with the MM, because it is corrected for the reference conditions (temperature and pressure) at the measurement point where the MUT is placed. V_{REF} is calculated by using the the gas equation [122]

$$pV = ZRT, \tag{E19.3}$$

where

p is the pressure of the gas,
V is the volume of the gas,
Z is the compressibility factor of the gas,
T is the absolute temperature of the gas,
R the ideal gas constant.

Z and R are considered to remain constant for both the measurement point of the MM and the MUT [79]:

$$\frac{p_{\text{MUT}} V_{\text{REF}}}{T_{\text{MUT}}} = \frac{p_{\text{MM}} V_{\text{MM}}}{T_{\text{MM}}} = ZR, \tag{E19.4}$$

$$V_{\text{REF}} = V_{\text{MM}} \frac{p_{\text{MM}}}{p_{\text{MUT}}} \frac{T_{\text{MUT}}}{T_{\text{MM}}}, \tag{E19.5}$$

where

V_{MM} is the volume of gas measured with the MM,
p_{MM} is the gas pressure measured with the MM,
p_{MUT} is the gas pressure measured with the MUT,
T_{MUT} is the gas temperature measured with the MUT,
T_{MM} is the gas temperature measured with the MM.
By substituting equation (E19.5) into equation (E19.2),

$$e = \frac{V_{\text{MUT}}}{V_{\text{MM}}} \frac{p_{\text{MUT}}}{p_{\text{MM}}} \frac{T_{\text{MM}}}{T_{\text{MUT}}} - 1. \tag{E19.6}$$

By using a slightly different notation from that in [79] the volume of the measured gas can be expressed in terms of the number of pulses and the K-factor (pulse value) of the measuring instrument:

$$V_{\text{MUT}} = \frac{I_{\text{MUT}}}{K_{\text{MUT}}}, \tag{E19.7}$$

$$V_{\text{MM}} = \frac{I_{\text{MM}}}{K_{\text{MM}}(1 + f_{\text{MM}})}, \tag{E19.8}$$

where

I_{MUT} is the number of pulses recorded on the MUT,
K_{MUT} is the pulse value directly given on the label of the MUT (a constant value for the individual measuring instrument),
I_{MM} is the number of pulses recorded on the MM,
K_{MM} is the pulse value directly given on the label of the MM (a constant value for the individual measurement standard),
f_{MM} is the MM error according to the calibration certificate.
After substitution, equation (E19.6) becomes

$$e = \frac{I_{\text{MUT}}}{I_{\text{MM}}} \frac{K_{\text{MM}}(1 + f_{\text{MM}})}{K_{\text{MUT}}} \frac{p_{\text{MUT}}}{p_{\text{MM}}} \frac{T_{\text{MM}}}{T_{\text{MUT}}} - 1. \tag{E19.9}$$

E19.4.2   Other influencing factors

Other factors that influence the measurement results are considered to be related to the calibration of the MM, i.e. the measurement standard (Q_{MM}), repeatability of measurement (Q_{REP}) and additional influencing quantities (Q_{AUX}). The additional influencing factors on the measuring results are as follows:
    MASTER METER

- Location of the MM (some MMs are located directly under the ceiling),
- Drift of the MM.

<u>METER UNDER TEST</u>

- Unknown characteristics.

<u>LABORATORY</u>

- Inadequate thermal insulation,
- For large flows, air is drawn from adjacent rooms whose temperature is different from that in the laboratory,
- Low interconnecting room,
- Flow computer,
- Separated pressure and temperature probes from the related transmitters/converters (the probes are on the test bench and the converters are remote and located in the control cabinet).

The measurement model used for the evaluation of measurement uncertainty is obtained by summing all influencing quantities on the measurement result as follows:

$$e_{\text{flow}} = e + Q, \tag{E19.10}$$

where Q denotes other influencing quantities and

$$Q = Q_{\text{MM}} + Q_{\text{REP}} + Q_{\text{AUX}}. \tag{E19.11}$$

E19.5   Uncertainty propagation

Uncertainty propagation follows the procedure described within GUM [51] (although we validate the results obtained using the propagation of distributions in section E19.7). The measurement model used for the uncertainty propagation is described by equation (E19.10), where it is assumed that the quantities Q_{MM}, Q_{REP} and Q_{AUX} have zero mean values and standard deviations equal to the standard uncertainties that will be explained in the following subsections. The standard uncertainty of e_{flow} given in equation (E19.10) can be expressed using the law of propagation of uncertainty [51] as follows:

$$u^2(e_{\text{flow}}) = \left[\left(\frac{\partial e_{\text{flow}}}{\partial e} u(e) \right)^2 + \left(\frac{\partial e_{\text{flow}}}{\partial Q} u(Q) \right)^2 \right], \tag{E19.12}$$

where

partial derivatives denote sensitivity coefficients of the measurement error (e) and other influencing quantities Q,
$u(e)$ is the standard uncertainty of the measurement error,
$u(Q)$ is the standard uncertainty of other influencing quantities.

E19.5.1   Standard measurement uncertainty $u(e)$ of the measurement error of the meter under test

The standard measurement uncertainty of the measurement error can be obtained from the mathematical model (E19.9), and can be expressed as follows:

$$
\begin{aligned}
u^2(e) = & \left[\frac{\partial e}{\partial I_{\text{MUT}}} u(I_{\text{MUT}}) \right]^2 + \left[\frac{\partial e}{\partial I_{\text{MM}}} u(I_{\text{MM}}) \right]^2 + \left[\frac{\partial e}{\partial p_{\text{MUT}}} u(p_{\text{MUT}}) \right]^2 \\
& + \left[\frac{\partial e}{\partial p_{\text{MM}}} u(p_{\text{MM}}) \right]^2 + \left[\frac{\partial e}{\partial T_{\text{MUT}}} u(T_{\text{MUT}}) \right]^2 + \left[\frac{\partial e}{\partial T_{\text{MM}}} u(T_{\text{MM}}) \right]^2.
\end{aligned} \tag{E19.13}
$$

The uncertainties of the pulse values (K) for the MM and the MUT, as well as the uncertainties of the error of the MM in this example were considered negligible.

Using (E19.9), let

$$S = e + 1 = \frac{I_{\text{MUT}}}{I_{\text{MM}}} \frac{K_{\text{MM}}(1 + f_{\text{MM}})}{K_{\text{MUT}}} \frac{p_{\text{MUT}}}{p_{\text{MM}}} \frac{T_{\text{MM}}}{T_{\text{MUT}}}.$$

Then, sensitivity coefficients are calculated from equation (E19.9) as follows:

$$\frac{\partial e}{\partial I_{\text{MUT}}} = \frac{S}{I_{\text{MUT}}}, \tag{E19.14}$$

$$\frac{\partial e}{\partial I_{\text{MM}}} = -\frac{S}{I_{\text{MM}}}, \tag{E19.15}$$

$$\frac{\partial e}{\partial p_{\text{MUT}}} = \frac{S}{p_{\text{MUT}}}, \tag{E19.16}$$

$$\frac{\partial e}{\partial p_{\text{MM}}} = -\frac{S}{p_{\text{MM}}}, \tag{E19.17}$$

$$\frac{\partial e}{\partial T_{\text{MM}}} = \frac{S}{T_{\text{MM}}}, \tag{E19.18}$$

$$\frac{\partial e}{\partial T_{\text{MUT}}} = -\frac{S}{T_{\text{MUT}}}. \tag{E19.19}$$

The standard measurement uncertainty of the following individual measurement quantities can be determined as the standard deviations of the according rectangular probability distributions:

$$u(e_{n_i}) = \frac{L_{n_i}}{\sqrt{3}}, \tag{E19.20}$$

$$u(e_{p_i}) = \frac{L_{p_i}}{\sqrt{3}}, \tag{E19.21}$$

$$u(e_{T_i}) = \frac{L_{T_i}}{\sqrt{3}}, \tag{E19.22}$$

where
  – e_{n_i} is the presumed error due to reading the number of pulses on the measuring instrument. According to [51] it is expected that the value of this error to lie within the interval $[n_{i-}, n_{i+}] = [-0.5, 0.5]$ pulse with length $L_{n_i} = n_{i+} - n_{i-} = 1$ pulse,
  – e_{p_i} is the presumed error due to measurement with pressure tubes, where, as in the previous case for e_{n_i}, the length of the interval is $L_{p_i} = 0.2$ mbar,
  – e_{T_i} is the presumed error due to measurement with temperature tubes, where again as for e_{n_i} the length of the interval is for the MM $L_{T_{\text{MM}}} = 0.132$ K and for the MUT $L_{T_{\text{MUT}}} = 0.163$ K.

Since the location of the pressure and temperature probes on the test bench can be changed, in the measurement uncertainty budget two probes for each quantity and their combination on the test bench are considered. In this way measurement uncertainty is slightly increased, but the measurement uncertainty calculation is simplified and kept on the "safe side" (despite its being not in keeping with the GUM, which recommends the use of realistic values).

The calculated values of the standard measurement uncertainty of individual quantities, accompanied by sensitivity coefficients, are used in equations (E19.14)–(E19.19).

E19.5.2 Standard measurement uncertainties of other influencing quantities Q_{REP}, Q_{MM}, Q_{AUX}

Standard measurement uncertainty u_{REP} of the mean value, obtained by a series of consecutive measurements — repeatability of the measurement

The method used for obtaining the standard deviation follows the principle described within the GUM [51]. The repeatability of measurement is calculated using

$$u_{\text{REP}} = \frac{s}{\sqrt{n}}, \tag{E19.23}$$

where s denotes the standard deviation of the series of n consecutive measurements.

Standard measurement uncertainty u_{MM} of the standard used -- master meter

The standard measurement uncertainty of the MM is calculated by using the expanded measurement uncertainty and a coverage factor, both obtained from the calibration certificate, i.e.

$$u_{\text{MM}} = \frac{U_{\text{MM}}}{k}, \tag{E19.24}$$

where

> U_{MM} is the expanded measurement uncertainty of the MM during the calibration procedure,
> k is the coverage factor ($k = 2$).

Standard measurement uncertainty u_{AUX} of additional influence factors

When evaluating measurement uncertainty it is necessary to include additional factors, which have influence on the measurement results and which influence is hard to quantify. The combined standard measurement uncertainty of other influence factors can be calculated from the estimated measurement error contribution, as well as from the assumption of rectangular distribution for influence factors, i.e.

$$u_{\text{AUX}} = \frac{e_{\text{AUX}}}{\sqrt{3}}, \tag{E19.25}$$

where e_{AUX} is the estimated error, which is usually bounded by $|e_{\text{AUX}}| \leq 0.1\,\%$.

Since the sensitivity coefficients in equation (E19.12) have the value 1,

$$u(e_{\text{flow}}) = \sqrt{u^2(e) + u_{\text{REP}}^2 + u_{\text{MM}}^2 + u_{\text{AUX}}^2}. \tag{E19.26}$$

E19.6   Reporting the result

The measurements were carried out according to the requirements set out in standards [22] and [21] for turbine and rotary gas meters. During the calibration process the measurement data presented in tables E19.2 and E19.3 were obtained.

Table E19.2: Data on flow, pressure and temperature parameters obtained by involving measurement standards G1000, G250 in the calibration procedure

Standard	G1000	G1000	G1000	G250	G250
Flow/(m^3/h)	995.832	800.404	650.997	452.394	349.768
I_{MM}	91 330.33	73 804	59 990	206 280.67	159 517.67
I_{MUT}	25 379.67	20 477.33	16 625	11 539.67	8907.6
K_{MM}	1630.75	1630.75	1630.75	8100	8100
K_{MUT}	450.238	450.238	450.238	450.238	450.238
f_{MM}/%	−0.0212	0.004 962	0.0189	0.1300	0.1200
p_{MM}/mbar	957.5	958.83	959.74	947.51	952.75
p_{MUT}/mbar	954.6	956.93	958.48	942.43	949.59
T_{MM}/K	295.010	294.920	294.980	294.970	295.000
T_{MUT}/K	295.080	294.990	294.980	295.030	295.050

Table E19.3: Data on flow, pressure and temperature obtained by involving measurement standards G250, G40 in the calibration procedure

Standard	G250	G250	G250	G40	G40
Flow/(m^3 h^{-1})	251.649	159.545	100.092	50.232	20.015
I_{MM}	115 029.33	73 059.67	45 954	9452.33	7564
I_{MUT}	6413.67	4066.67	2553.33	1277.67	1022
K_{MM}	8100	8100	8100	3338.82	3338.82
K_{MUT}	450.238	450.238	450.238	450.238	450.238
f_{MM}/%	0.1171	0.1604	0.2796	0.1000	−0.0276
p_{MM}/mbar	956.67	959.35	960.47	960.41	961.03
p_{MUT}/mbar	954.96	958.67	960.23	958.31	960.79
T_{MM}/K	295.060	295.12	295.20	295.370	295.470
T_{MUT}/K	295.070	295.100	295.140	295.170	295.200

In tables E19.2 and E19.3 'Flow' represents the mean value of three observations of flow rate. One flow rate value was selected among the results in these tables in order to present step-by-step calculation of measurement uncertainty in this example. The selected flow rate is 251.649 m^3/h and it was measured by turbine gas meter G250. The data used for calculation appear in Tables E19.2 and E19.3. The error of the measuring instrument for this measuring point can be calculated according to equation (E19.9) as follows:

$$e = \left(\frac{6413.67}{115\,029.33} \frac{8100(1 + 0.1171/100)}{450.238} \frac{954.96}{956.67} \frac{295.06}{295.07} - 1 \right) \times 100\,\% \qquad \text{(E19.27)}$$

$$= 0.2441\,\%. \qquad \text{(E19.28)}$$

The standard measurement uncertainty of the measurement error of the MUT can be calculated from equation (E19.13). However, it is necessary first to calculate the sensitivity coefficients of each influencing parameter, which can be carried out according to formulæ (E19.14) and (E19.15). The value of S was calculated as 100.243.

Sensitivity coefficients are calculated as follows:

$$\frac{\partial e}{\partial I_{\mathrm{MUT}}} = \frac{100.243}{6413.67} = 0.0156, \tag{E19.29}$$

$$\frac{\partial e}{\partial I_{\mathrm{MM}}} = -\frac{100.243}{115\,029.33} = -0.000\,871, \tag{E19.30}$$

$$\frac{\partial e}{\partial p_{\mathrm{MUT}}} = \frac{100.243}{954.96} = 0.149\,70, \tag{E19.31}$$

$$\frac{\partial e}{\partial p_{\mathrm{MM}}} = -\frac{100.243}{956.67} = -0.147\,80, \tag{E19.32}$$

$$\frac{\partial e}{\partial T_{\mathrm{MM}}} = \frac{100.243}{295.06} = 0.339\,74, \tag{E19.33}$$

$$\frac{\partial e}{\partial T_{\mathrm{MUT}}} = -\frac{100.243}{295.07} = -0.339\,729. \tag{E19.34}$$

From equations (E19.20)–(E19.22) and information on errors given in subsection E19.5.1, we obtain

$$u(I_i) = \frac{1}{\sqrt{3}} = 0.5780, \tag{E19.35}$$

$$u(p_i) = \frac{0.2}{\sqrt{3}} = 0.1156, \tag{E19.36}$$

$$u(T_{MUT}) = \frac{0.163}{\sqrt{3}} = 0.0942 \tag{E19.37}$$

$$u(T_{MM}) = \frac{0.132}{\sqrt{3}} = 0.0763. \tag{E19.38}$$

Results (E19.35)–(E19.37) combined with equation (E19.13) give the standard uncertainty of the measuring system measurement error:

$$u(e) = 0.045\,\%. \tag{E19.39}$$

The relative standard deviation (s) of the measured values at this point is $s = 0.0031\,\%$, which gives the relative standard uncertainty due to repeatability of measurements:

$$u_{\mathrm{REP}} = 0.0018\,\%. \tag{E19.40}$$

The standard measurement uncertainty u_{MM} of the standard used –- master meter – is calculated according to equation (E19.24), where the the expanded measurement uncertainty of the MM (obtained from the calibration certificate) is $U_{\mathrm{MM}} = 0.25\,\%$, giving

$$u_{\mathrm{MM}} = \frac{0.25}{2} \tag{E19.41}$$

$$= 0.125\,\%. \tag{E19.42}$$

The relative standard uncertainty u_{AUX} of other influence factors can be calculated according to equation (E19.25), where it is assumed the estimated error is $e_{\mathrm{AUX}} = 0.1\,\%$ and to have a rectangular probability distribution:

$$u_{\mathrm{AUX}} = \frac{0.1}{\sqrt{3}}\,\% \tag{E19.43}$$

$$= 0.0578\,\% \tag{E19.44}$$

By substituting the uncertainty contributions obtained in (E19.40), (E19.42), (E19.39) and (E19.44) into equation (E19.26), the combined relative standard uncertainty becomes

$$u(e_{\text{flow}}) = (0.045^2 + 0.0018^2 + 0.125^2 + 0.0578^2)^{1/2}\,\%, \qquad (\text{E19.45})$$
$$u(e_{\text{flow}}) = 0.1449\,\%. \qquad (\text{E19.46})$$

The expanded measurement uncertainty, with coverage factor $k = 2$ is

$$U = 2 \times u(e_{\text{flow}}) \qquad (\text{E19.47})$$
$$= 0.29\,\%. \qquad (\text{E19.48})$$

The results for the uncertainty contributions at every flow rate, as well as the combined and expanded measurement uncertainty ($k = 2$) are presented in table E19.4.

Table E19.4: Uncertainty contributions for individual flow rates

Standard	$u_{\text{REP}}/\%$	$u_{\text{MM}}/\%$	$u(e)/\%$	$u_{\text{AUX}}/\%$	$u(e_{\text{flow}})/\%$	$U/\%$
G1000	0.044 65	0.125	0.000 25	0.057 73	0.144 74	0.29
G1000	0.044 67	0.125	0.000 87	0.057 73	0.144 76	0.29
G1000	0.044 69	0.125	0.001 50	0.057 73	0.144 77	0.29
G250	0.044 85	0.125	0.002 36	0.057 73	0.144 83	0.29
G250	0.045 01	0.125	0.000 39	0.057 73	0.144 86	0.29
G250	0.045 41	0.125	0.001 78	0.057 73	0.144 99	0.29
G250	0.046 68	0.125	0.005 42	0.057 73	0.145 48	0.29
G250	0.049 82	0.125	0.001 27	0.057 73	0.146 43	0.29
G40	0.063 70	0.125	0.001 82	0.057 73	0.151 72	0.30
G40	0.072 42	0.125	0.001 32	0.057 73	0.155 57	0.31

E19.7   Interpretation of results

The approach for uncertainty evaluation described within this example can be generally used for calibration of gas flow meters by the 'master meter' method, when the meter under test is located downstream of the master meter. Depending on the costumer needs, it can be decided earlier how many flow rate points are necessary for the calculation of the measurement uncertainties.

In this example the uncertainties due to additional influence factors (u_{AUX}), which were not exactly known, were quantified. One way of improving this example would be to quantify other uncertainty sources and include them to the overall uncertainty budget.

According to [51], if the measurement model is linear in the input quantities and the dominant contributions have normal probability distributions, the GUF will provide reliable results. In order to validate the GUF, the MCM, described in GUM Supplement 1 (GUM-S1) [52] was used for the evaluation of measurement uncertainty for the measurement model described by equation (E19.10). For this method the number of Monte Carlo trials was prescribed to be 10^8. Values for the input quantities, their associated standard uncertainties and their probability distributions remained the same as described in section E19.6.

The results of the applied method are shown in table E19.5.

Table E19.5: Results obtained by two approaches for measurement uncertainty evaluation: GUF–GUM uncertainty framework, MCM–Monte Carlo method, CI–Coverage interval (lower limit, higher limit)

Approach	Estimate/%	Std. unc./%	CI (95%)/%
GUF	0.02441	0.1449	(−0.0459, 0.5341)
MCM	0.02441	0.1402	(−0.0309, 0.5185)

Figure E19.2 shows the probability density functions as a result of the evaluation of uncertainty by following the principles described within GUM and GUM Supplement 1 (Monte Carlo method) [52]. It can be noticed, both in table E19.5 and from figure E19.2 that the obtained results for the methods differ insignificantly. The 95 % coverage interval for the MCM method is slightly shorter than that obtained by the GUM method, while the estimates of the flow measurement error are the same.
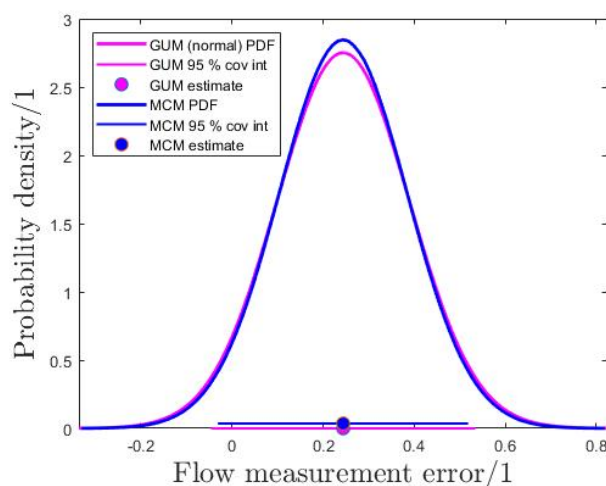
Figure E19.2: Probability density functions for Monte Carlo and GUM approach

Due to the sufficiently large number of Monte Carlo trials and even though the measurement model was not linear in this example, the GUM method provided accurate results.

The GUM approach for evaluation of measurement uncertainty in terms of the measurement error described in previous sections is followed by several national metrology institutes in Europe.

Acknowledgements

We offer our sincerest gratitude to the 'SARAJEVOGAS' Laboratory, which contributed greatly to the project by providing all necessary information for development of this example.

Bibliography

[1] Calibration and Measurement Capabilities of PTB for gas flow rate are documented at www.bipm.org/kcdb/. Last visit on 06/08/2020.

[2] Drugs.com Overview: What is NyQuil Severe Cold & Flu? `https://www.drugs.com/mtm/nyquil-severe-cold-flu.html`. Accessed: 2019-06-26.

[3] ISO 6145 Gas analysis – Preparation of calibration gas mixtures using dynamic methods. ISO 6145, International Organization for Standardization, Geneva, Switzerland. 11 parts.

[4] ISO 7726 Ergonomics of the thermal environment – Istruments for measuring physical quantities. ISO, International Organization for Standardization, Geneva, Switzerland, 1998. Second edition.

[5] EPA method IO-2.1 (1999) Sampling of ambient air for total suspended particulate matter (SPM) and PM10 using high volume (HV) sampler. `http://www.epa.gov/ttnamti1/inorg.html`, 1999.

[6] ISO 6143 Gas analysis – Comparison methods for determining and checking the composition of calibration gas mixtures. ISO, International Organization for Standardization, Geneva, Switzerland, 2001. Second edition.

[7] ISO 6145–10 Gas analysis – Preparation of calibration gas mixtures using dynamic volumetric methods – Part 10: Permeation methods. ISO 6145–10, International Organization for Standardization, Geneva, Switzerland, First edition 2002.

[8] BS 3985:2003 Haemoglobincyanide (cyanmethhaemoglobin) preparation as a standard for spectrometric haemoglobin. BS, British Standards Institution, BSI Group, UK, 2003.

[9] ISO 17511 In vitro diagnostic medical devices - Measurement of quantities in biological samples - Metrological traceability of values assigned to calibrators and control materials. ISO, International Organization for Standardization, Geneva, Switzerland, 2003.

[10] Krypton Help Pages on K400/K600 Hardware & Software Guide, April 2003.

[11] ISO 7730 Ergonomics of the thermal environment – Analytical determination and interpretation of the thermal comfort using calculation of the PMV and PPD indices and local thermal comfort criteria. ISO, International Organization for Standardization, Geneva, Switzerland, 2005. Third edition.

[12] ISO 9300 Measurement of gas flow by means of critical flow Venturi nozzles. ISO, International Organization for Standardization, Geneva, Switzerland, 2005. Second edition.

[13] EN 15549:2008, Air Quality - Standard Method for the Measurement of the Concentration of Benzo[a]pyrene in Ambient Air. EN, European Organization for Standardization, 2008.

[14] DIN 58931:2010 Haematology – determination of haemoglobin concentration in blood – reference method. DIN, German Standards Institution, Beuth-Verlag, Berlin, Germany, 2010.

[15] ISO/IEC 17043 Conformity assessment – General requirements for proficiency testing. ISO, International Organization for Standardization, Geneva, Switzerland, 2010. First edition.

[16] ISO/TS 28037 Determination and use of straight-line calibration functions. ISO, International Organization for Standardization, Geneva, Switzerland, 2010.

[17] EN 13823:2010+A1:2014 Reaction to fire tests for building products. building products excluding floorings exposed to the thermal attack by a single burning item. EN, European Organization for Standardization, Brussels, Belgium, 2010, 2014. First edition.

[18] EURACHEM/CITAC Guide QUAM:2012.P1, 2012. third edition.

[19] EN 14211:2012, Ambient air - Standard method for the measurement of the concentration of nitrogen dioxide and nitrogen monoxide by chemiluminescence. Technical report, 2012.

[20] ISO 15189 Medical laboratories – Requirements for quality and competence. ISO, International Organization for Standardization, Geneva, Switzerland, 2012. Third edition.

[21] EN 12261 Gas meters. Turbine gas meters. CEN, European Committee for Standardization, 2015.

[22] EN 12480 Gas meters. Rotary displacement gas meters. CEN, European Committee for Standardization, 2015.

[23] ISO 13528 Statistical methods for use in proficiency testing by interlaboratory comparison. ISO, International Organization for Standardization, Geneva, Switzerland, 2015. Second edition.

[24] ISO 6142–1 Gas analysis – Preparation of calibration gas mixtures – Gravimetric method for Class I mixtures. ISO, International Organization for Standardization, Geneva, Switzerland, 2015. First edition.

[25] ISO Guide 33 Reference materials – Good practice in using reference materials. ISO, International Organization for Standardization, Geneva, Switzerland, 2015. Third edition.

[26] EN/TR 16988:2016 Estimation of uncertainty in the single burning item test. EN, European Organization for Standardization, Brussels, Belgium, 2016. First edition.

[27] ISO 17034 General requirements for the competence of reference material producers. ISO, International Organization for Standardization, Geneva, Switzerland, First edition 2016.

[28] ISO 6976 Natural gas – Calculation of calorific values, density, relative density and Wobbe indices from composition. ISO, International Organization for Standardization, Geneva, Switzerland, 2016. Third edition,.

[29] ISO/DTR 29922 Natural gas – Supporting information on the calculation of physical properties according to ISO 6976. ISO 29922, International Organization for Standardization, Geneva, Switzerland, 2016. First edition.

[30] ISO 20486:2017 Non-destructive testing — Leak testing — Calibration of reference leaks for gases. ISO, International Organization for Standardization, Geneva, Switzerland, 2017. First edition.

[31] ISO 6145–6 Gas analysis – Preparation of calibration gas mixtures using dynamic volumetric methods – Part 6: Critical flow orifices. ISO 6145–6, International Organization for Standardization, Geneva, Switzerland, Third edition 2017.

[32] ISO Guide 35 Reference materials – Guidance for characterization and assessment of homogeneity and stability. ISO, International Organization for Standardization, Geneva, Switzerland, 2017. Fourth edition.

[33] ISO/IEC 17025 General requirements for the competence of testing and calibration laboratories. ISO, International Organization for Standardization, Geneva, Switzerland, 2017. Third edition.

[34] ISO 14167 Gas analysis – General quality aspects and metrological traceability of calibration gas mixtures. ISO, International Organization for Standardization, Geneva, Switzerland, 2018. First edition.

[35] ISO 19229 Gas analysis – Purity analysis and the treatment of purity data. ISO, International Organization for Standardization, Geneva, Switzerland, 2018. Second edition.

[36] ISO 6145–7 Gas analysis – Preparation of calibration gas mixtures using dynamic volumetric methods – Part 7: Thermal mass-flow controllers. ISO 6145–7, International Organization for Standardization, Geneva, Switzerland, Third edition 2018.

[37] ISO 6145-7:2018 Gas analysis – Preparation of calibration gas mixtures using dynamic methods – Part 7: Thermal mass-flow controllers. ISO, International Organization for Standardization, Geneva, Switzerland, 2018. Third edition.

[38] ISO/TS 28038 Determination and use of polynomial calibration functions. ISO, International Organization for Standardization, Geneva, Switzerland, 2018. First edition.

[39] ISO/TS 28038:2018 Determination and use of polynomial calibration functions. ISO, International Organization for Standardization, Geneva, Switzerland, 2018. First edition.

[40] VDI/VDE 2600 Part 2 Inspection process management – Determination of the measurement uncertainty of complex inspection processes. VDI-Richtlinie, Verband Deutscher Ingeneure, Verband der Elektrotechnik, Elektronik und Informationstechnik, Beuth-Verlag, Berlin, Germany, 2019.

[41] *Richtlinie der Bundesärztekammer zur Qualitätssicherung laboratoriumsmedizinischer Untersuchungen*. Deutsches Ärzteblatt, 2019. See also the unauthorised translation of the previous version of the Guideline of the German Medical Association on Quality Assurance in Medical Laboratory Examinations – Rili-BAEK. J. Lab. Med., 39(1): 26–69, 2019. doi:10.3238/arztebl.2019.rili_baek_QS_Labor20192312.

[42] A.A.Kornhauser. Demonstrating the limitations of the ideal gas law. *International Journal of Applied Engineering Research –IJAER*, 7(5):398–400, 1991. URL: https://www.ijee.ie/articles/Vol07-5/070512.PDF.

[43] J. J Allaire, Y. Xie, J. McPherson, J. Luraschi, K. Ushey, A. Atkins, H. Wickham, J. Cheng, W. Chang, and R. Iannone. *rmarkdown: Dynamic Documents for R*, 2020. R package version 2.1. URL: https://github.com/rstudio/rmarkdown.

[44] A. Allard, N. Fischer, I. Smith, P. Harris, and L. Pendrill. CASoft software. `https://www.lne.fr/en/software/CASoft`, 2020. version 1.1.

[45] V. T. Anchinmane and S. V. Sankhe. Evaluation of hemoglobin estimation with non-cyanide alkaline haematin D-575 method. *Int. J. Res. Med. Sci.*, 4(10):4297–4299, 2016.

[46] David Andrich and Pender Pedler. On a law of ordinal error. *Journal of Physics: Conference Series*, 1044:012055, Jun 2018. `doi:10.1088/1742-6596/1044/1/012055`.

[47] Kai Oliver Arras. An introduction to error propagation: Derivation, meaning and examples of equation $C_Y = F_X C_X F_X^\mathsf{T}$. Technical Report EPFL-ASL-TR-98-01 R3, Swiss Federal Institute of Technology, Lausanne, Switzerland, Sept 1998.

[48] Svante Arrhenius. Über die Dissociationswärme und den Einfluss der Temperatur auf den Dissociationsgrad der Elektrolyte. *Zeitschrift für Physikalische Chemie*, 4U(1), jan 1889. `doi:10.1515/zpch-1889-0408`.

[49] Svante Arrhenius. Über die Reaktionsgeschwindigkeit bei der Inversion von Rohrzucker durch Säuren. *Zeitschrift für Physikalische Chemie*, 4(1), jan 1889. `doi:10.1515/zpch-1889-0116`.

[50] Walter Bich, Maurice G Cox, René Dybkaer, Clemens Elster, W Tyler Estler, Brynn Hibbert, Hidetaka Imai, Willem Kool, Carine Michotte, Lars Nielsen, Leslie Pendrill, Steve Sidney, Adriaan M H van der Veen, and Wolfgang Wöger. Revision of the 'Guide to the expression of Uncertainty in Measurement'. *Metrologia*, 49(6):702–705, 2012. URL: `http://stacks.iop.org/0026-1394/49/i=6/a=702`.

[51] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. *Guide to the Expression of Uncertainty in Measurement, JCGM 100:2008, GUM 1995 with minor corrections*. BIPM, 2008.

[52] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. *Supplement 1 to the 'Guide to the Expression of Uncertainty in Measurement' – Propagation of distributions using a Monte Carlo method, JCGM 101:2008*. BIPM, 2008.

[53] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. *Evaluation of measurement data – Introduction to the "Guide to the expression of Uncertainty in Measurement" and related documents, JCGM 104:2009*. BIPM, 2009.

[54] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. *Supplement 2 to the 'Guide to the Expression of Uncertainty in Measurement' – Extension to any number of output quantities, JCGM 102:2011*. BIPM, 2011.

[55] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. *Evaluation of measurement data – The role of measurement uncertainty in conformity assessment, JCGM 106:2012*. BIPM, 2012.

[56] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. *Guide to the Expression of Uncertainty in Measurement – Developing and using measurement models, CD JCGM 103:2018*. Joint Committee for Guides in Metrology (JCGM), 2018.

[57] Victor Bloomfield. *Using R for numerical analysis in science and engineering*. CRC Press, Taylor & Francis Group, Boca Raton, 2014.

[58] M. L. Boas. *Mathematical Methods in the Physical Sciences*. John Wiley & Sons Inc, 2005.

[59] R. Boudjemaa, M. G. Cox, A. B. Forbes, and P. M. Harris. Automatic differentiation and its applications to metrology. In Patrizia Ciarlini, M. G. Cox, F. Pavese, and G. B. Rossi, editors, *Advanced Mathematical and Computational Tools in Metrology VI*, pages 170–179, Singapore, 2004. World Scientific.

[60] G.E.P. Box and G.C. Tiao. *Bayesian Inference in Statistical Analysis*. Wiley, 1992.

[61] C. Brezinski and M. Redivo-Zaglia. Extrapolation methods. *Applied Numerical Mathematics*, 15(2):123–131, sep 1994. `doi:10.1016/0168-9274(94)00015-8`.

[62] C. A. Cantrell. Technical note: Review of methods for linear least-squares fitting of data and application to atmospheric chemistry problems. *Atmospheric Chemistry and Physics*, 8(17):5477–5487, sep 2008. URL: `http://dx.doi.org/10.5194/acp-8-5477-2008`, `doi:10.5194/acp-8-5477-2008`.

[63] Rob Carnell. *triangle: Provides the Standard Distribution Functions for the Triangle Distribution*, 2017. R package version 0.11. URL: `https://CRAN.R-project.org/package=triangle`.

[64] Bob Carpenter, Andrew Gelman, Matthew D. Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1), 2017. `doi:10.18637/jss.v076.i01`.

[65] R. Chaudhary, A. Dubey, and A. Sonker. Techniques used for the screening of haemoglobin levels in blood donors: current insights and future directions. *J. Blood. Med.*, 8:75–88, 2017. `doi:10.2147/JBM.S103788`.

[66] S. Chib and E. Greenberg. Understanding the metropolis-hastings algorithm. *The American Statistician*, 49(4):327–335, 1995. URL: `http://amstat.tandfonline.com/doi/abs/10.1080/00031305.1995.10476177`, `doi:10.1080/00031305.1995.10476177`.

[67] CIBA-Geigy AG. *Wissenschaftliche Tabellen Geigy: Teilband Hämatologie und Humangenetik*. Basel, 8th edition, 1979. 4. Nachdruck 1985. URL: `https://books.google.de/books?id=ovhlPgAACAAJ`.

[68] CIPM. Mutual recognition of national measurement standards and of calibration and measurement certificates issued by national metrology institutes, October 1999.

[69] E.R. Cohen, T. Cvitas, J.G.Frey, B. Holmström, K. Kuchitsu, R. Marquardt, I. Mills, F. Pavese, M. Quack, J. Stohner, H.L. Strauss, M. Takami, and A.J. Thor. *Quantities, units and symbols in physical chemistry, IUPAC Green Book*. IUPAC & RSC Publishing, Cambridge, 3rd edition, 2nd print edition, 2008.

[70] EMRP IND 12 consortium. Metrology of the leak detection - Practical guide. Technical report, 2015.

[71] M Cox and K Shirono. Informative Bayesian Type A uncertainty evaluation, especially applicable to a small number of observations. *Metrologia*, 54(5):642, 2017. URL: `http://stacks.iop.org/0026-1394/54/i=5/a=642`.

[72] M G Cox. The evaluation of key comparison data. *Metrologia*, 39(6):589–595, 2002. `doi:10.1088/0026-1394/39/6/10`.

[73] Maurice Cox, Peter Harris, and Martin Milton. Method for determining acceptable CMCs to ensure consistency with KC results. CCQM Report 09-15, 2009. URL: `https://www.bipm.org/cc/CCQM/Restricted/15/CCQM-09-15.pdf`.

[74] Manuel Gameiro da Silva, Maria Marrero Santana, and João Alves e Sousa. Uncertainty analysis of the mean radiant temperature measurement based on globe temperature probes. *Journal of Physics: Conference Series*, 1065:072036, Aug 2018. `doi:10.1088/1742-6596/1065/7/072036`.

[75] A. da Silva Pereira, I. R. Ribeiro da Castro, F. F. Bezerra, J. F. N. Neto, and A. C. Feldenheimer da Silva. Reproducibility and validity of portable haemoglobinometer for the diagnosis of anaemia in children under the age of 5 years. *J. Nutr. Sci.*, 9:e3, 2020. `doi:10.1017/jns.2019.43`.

[76] T. J. Dekker. Finding a zero by means of successive linear interpolation. In B. Dejon and P. Henrici, editors, *Constructive aspects of the fundamental theorem of algebra*, pages 37–48. Interscience, New York, 1969.

[77] S. Demeyer, N. Fischer, M.G. Cox, A.M.H. van der Veen, O. Pellegrino, J. Sousa, A. Bošnjaković, V. Karahodžić, and C. Elster. EMUE-D1-2-BayesianMassCalibration, March 2020. `doi:10.5281/zenodo.3726908`.

[78] W.E. Deming. *Statistical Adjustment of Data*. Dover books on elementary and intermediate mathematics. J. Wiley & Sons, Incorporated, 1943. URL: `https://books.google.nl/books?id=QGdAAAAAIAAJ`.

[79] Dietrich H., Hotze H.J., Jarosch B., Juenger F.J., Kaempf M., Kramer R., Mickan B., Nath B., Polzin H., and Wendt G. PTB testing instructions: Measuring instruments for gas:gas meters–testing of gas volume meters with air at atmospheric pressure. Technical report, Braunschweig and Berlin, Germany, 2003.

[80] D. Drabkin and J.H. Austin. Spectrophotometric studies: I. Spectrophotometric constants for common haemoglobin derivatives in human, dog, and rabbit blood. *J. Biol. Chem.*, 98:719–733, 1932.

[81] Norman R. Draper and Harry Smith. *Applied Regression Analysis*. Wiley-Blackwell, apr 1998. URL: `http://dx.doi.org/10.1002/9781118625590`, `doi:10.1002/9781118625590`.

[82] EA Laboratory Committee. EA 4/02 Evaluation of the uncertainty of measurement in calibration. European Cooperation for Accreditation, September 2013.

[83] Stephen L. R. Ellison. Homogeneity studies and ISO Guide 35:2006. *Accreditation and Quality Assurance*, 20(6):519–528, 2015. URL: `http://dx.doi.org/10.1007/s00769-015-1162-z`, `doi:10.1007/s00769-015-1162-z`.

[84] C. Elster. Bayesian uncertainty analysis compared with the application of the GUM and its supplements. *Metrologia*, 51:S159–S166, 2014. URL: `http://stacks.iop.org/0026-1394/54/i=5/a=642`.

[85] R. Enns. Capsule endoscopy in patients with iron deficiency. *J. Gastroenterol. Hepatol.*, 8:847–849, 2012.

[86] European Parliament and Council of the European Union. Directive 2004/107/EC, Directive of the European parliament and of the Council of 15 December 2004 relating to arsenic, cadmium, mercury, nickel and polycyclic aromatic hydrocarbons in ambient air. *Off. J. Eur. Union*, L23:3–16, 2005. URL: `http://data.europa.eu/eli/dir/2004/107/oj`.

[87] European Parliament and Council of the European Union. Directive 2008/50/EC, Directive of the European parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe. *Off. J. Eur. Union*, L152:1–44, 2008. URL: `https://eur-lex.europa.eu/eli/dir/2008/50/oj`.

[88] F. Pennecchi, F. Rolle, A. Allard, and S. L. R. Ellison. EMUE-D2-3-TSPConcentration, 2020. URL: `https://zenodo.org/record/4242988`, `doi:10.5281/ZENODO.4242988`.

[89] R Feistel, J. W Lovell-Smith, P Saunders, and S Seitz. Uncertainty of empirical correlation equations. *Metrologia*, 53(4):1079, 2016.

[90] C. Frank, C. Brauckmann, M. Palos, C. G Arsene, J. Neukammer, M. E. del Castillo Busto, S. Zakel, C. Swart, B. Güttler, and R. Stosch. Comparison of potential higher order reference methods for total haemoglobin quantification – an interlaboratory study. *Anal. Bioanal. Chem.*, 409(9):2341–2351, 2017.

[91] Andrew Gelman. Prior distributions for variance parameters in hierarchical models (Comment on article by Browne and Draper). *Bayesian Analysis*, 1:515–534, 2006. URL: `http://dx.doi.org/10.1214/06-BA117A`, `doi:10.1214/06-ba117a`.

[92] Andrew Gelman, John Carlin, Hal Stern, David Dunson, Aki Vehtari, and Donald Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, Boca Raton, Florida, USA, 3rd edition, 2013.

[93] Alan Genz and Frank Bretz. *Computation of Multivariate Normal and t Probabilities*. Lecture Notes in Statistics. Springer-Verlag, Heidelberg, 2009.

[94] D. Geropp. Laminare Grenzschichten in Ebenen und Rotationssymmetrischen Lavaldüsen. *Deutsche Luft- und Raumfahrt, Forschungsbericht*, pages 71–90, 1971.

[95] D. Geropp. Laminare Grenzschichten in Überschalldüsen. *Deutsche Luft- und Raumfahrt, Forschungsbericht*, 01 TM 8603-AK/PA 1, 1987.

[96] William F. Guthrie. NIST/SEMATECH e-Handbook of Statistical Methods (NIST Handbook 151), 2020. `doi:10.18434/M32189`.

[97] Werner Haesselbarth. Accounting for bias in measurement uncertainty estimation. *Accreditation and Quality Assurance*, 9(8), 2004. `doi:10.1007/s00769-004-0782-5`.

[98] T. Harvey, A. Zkik, M. Auges, and T. Clavel. Assessment of iron deficiency and anemia in pregnant women: an observational french study. *Women's Health*, 12:95–102, 2016. `doi:10.2217/whe.15.91`.

[99] Jeremy Thoms Hetzel. *trapezoid: The Trapezoidal Distribution*, 2012. R package version 2.0-0. URL: `https://CRAN.R-project.org/package=trapezoid`.

[100] Peter D. Hoff. *A First Course in Bayesian Statistical Methods*. Springer New York, 2009. URL: `http://dx.doi.org/10.1007/978-0-387-92407-6`, `doi:10.1007/978-0-387-92407-6`.

[101] A. Karsan, I. Maclaren, D. Conn, and L. Wadsworth. An evaluation of haemoglobin determination using sodium lauryl sulfate. *Am. J. Clin. Pathol.*, 100:123–126, 1993. `doi:10.1093/ajcp/100.2.123`.

[102] Katy Klauenberg, Gerd Wübbeler, and Clemens Elster. About not Correcting for Systematic Effects. *Measurement Science Review*, 19(5):204–208, 2019. `doi:10.2478/msr-2019-0026`.

[103] D.E. Knuth. *The art of computer programming*. The art of computer programming: Seminumerical algorithms. Addison-Wesley, 2001. URL: `https://books.google.nl/books?id=um0eAQAAIAAJ`.

[104] Manfred Kochsiek and Gläser. *Comprehensive mass metrology*. Wiley-VCH, Weinheim New York, 2000.

[105] A. Koepke, T. Lafarge, A. Possolo, and B. Toman. NIST Consensus Builder user's manual. Technical report, National Institute of Standards and Technology, Gaithersburg, MD, 2016. URL: `http://consensus.nist.gov`.

[106] Amanda Koepke, Thomas Lafarge, Antonio Possolo, and Blaza Toman. Consensus building for interlaboratory studies, key comparisons, and meta-analysis. *Metrologia*, 54(3):S34, 2017. URL: `http://stacks.iop.org/0026-1394/54/i=3/a=S34`.

[107] M. Krystek and M. Anton. A least-squares algorithm for fitting data points with mutually correlated coordinates to a straight line. *Meas. Sci. Technol.*, 22(3):035101, 2011.

[108] O. Kunz, R. Klimeck, W. Wagner, and M. Jaeschke. The GERG-2004 Wide-Range Reference Equation of State for Natural Gases and Other Mixtures. GERG Technical Monograph 15, 2007.

[109] I. Kuselman, S. Shpitzer, F. Pennecchi, and C. Burns. Investigating out-of-specification test results of mass concentration of total suspended particulates in air based on metrological concepts - a case study. *Air Qual Atmos Health*, 5:269–276, 2012. URL: `http://dx.doi.org/10.1007/s11869-010-0103-6`, `doi:10.1007/s11869-010-0103-6`.

[110] Ilya Kuselman, D. Brynn Hibbert, Francesca R. Pennecchi, and Ricardo J.N.B. da Silva. IUPAC Project no.: 2016-007-1-500, Risks of conformity assessment of a multicomponent material or object in relation to measurement uncertainty of its test results, 1 June 2016 – 30 June 2018. `https://iupac.org/projects/project-details/?project_nr=2016-007-1-500`. Accessed: 2019-06-26.

[111] Ilya Kuselman, Francesca R. Pennecchi, C. Burns, A. Fajgelj, and P. de Zorzi. IUPAC/CITAC Guide: Investigating out-of-specification test results of chemical composition based on metrological concepts (IUPAC Technical Report). *Pure Appl. Chem*, 84(9):1939–1971, 2012. `doi:10.1351/PAC-REP-11-10-04`.

[112] Ilya Kuselman, Francesca R. Pennecchi, Ricardo J.N.B. da Silva, and D. Brynn Hibbert. Risk of false decision on conformity of a multicomponent material when test results of the components' content are correlated. *Talanta*, 174:789–796, nov 2017. `doi:10.1016/j.talanta.2017.06.073`.

[113] E. W. Lemmon and R. T. Jacobsen. Viscosity and thermal conductivity equations for nitrogen, oxygen, argon,and air. *Int. J. Thermophys.*, 25:21–69, 2004. `doi:10.1023/B:IJOT.0000022327.04529.f3`.

[114] Andrew Lewis. CCL-K2: Long gauge block measurement by interferometry: Final report. *Metrologia*, 40(1A):04004, 2003. URL: `http://stacks.iop.org/0026-1394/40/i=1A/a=04004`.

[115] S. M. Lewis and S. Kumari. Chapter 7: Haemoglobinometry. In *Guidelines on standard operating procedures for haematology*. WHO, 1999.

[116] I. H. Lira and W. Wöger. The evaluation of the uncertainty in knowing a directly measured quantity. *Meas. Sci. Technol.*, 9:1167–1173, 1998.

[117] Bertil Magnusson and Stephen L. R. Ellison. Treatment of uncorrected measurement bias in uncertainty estimation for chemical measurements. *Analytical and Bioanalytical Chemistry*, 390(1):201–213, 2007. `doi:10.1007/s00216-007-1693-1`.

[118] A. Malengo and F. Pennecchi. A weighted total least-squares algorithm for any fitting model with correlated variables. *Metrologia*, 50(6):654, 2013. URL: `http://stacks.iop.org/0026-1394/50/i=6/a=654`, `doi:10.1088/0026-1394/50/6/654`.

[119] Steffen Martens, Katy Klauenberg, and Clemens Elster. EMUE-D6-2-CalibrationUncertaintyGUMvsBayesian, 2020. URL: `https://zenodo.org/record/3858120`, `doi:10.5281/ZENODO.3858120`.

[120] Steffen Martens, Katy Klauenberg, Bodo Mickan, Catherine Yardin, Nicolas Fischer, and Clemens Elster. EMUE-D4-3-QuantifyUncertaintiesInCalibration, 2020. URL: `https://zenodo.org/record/4016915`, `doi:10.5281/ZENODO.4016915`.

[121] Steffen Martens, Katy Klauenberg, Jörg Neukammer, Simon Cowen, Stephen L. R. Ellison, and Clemens Elster. EMUE-D5-4-MethodComparisonWithCorrelation, 2020. URL: `https://zenodo.org/record/3911583`, `doi:10.5281/ZENODO.3911583`.

[122] K.A. Masavetas. The mere concept of an ideal gas. *Mathematical and Computer Modelling*, 12(6):651–657, 1989.

[123] Makoto Matsumoto and Takuji Nishimura. Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation*, 8(1):3–30, jan 1998. `doi:10.1145/272991.272995`.

[124] B.J. McCaffrey and G. Heskestad. A robust bidirectional low-velocity probe for flame and fire application. *Combustion and Flame*, 26:125–127, feb 1976. `doi:10.1016/0010-2180(76)90062-6`.

[125] A. D. McNaught and A. Wilkinson. *Compendium of Chemical Terminology – IUPAC Recommendations (IUPAC Chemical Data) – Gold Book*. Wiley, 2014. URL: `https://www.amazon.com/Compendium-Chemical-Terminology-IUPAC-Recommendations/dp/0865426848?SubscriptionId=0JYN1NVW651KCA56C102&tag=techkie-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=0865426848`.

[126] Juris Meija, Tyler B. Coplen, Michael Berglund, Willi A. Brand, Paul De Bièvre, Manfred Gröning, Norman E. Holden, Johanna Irrgeher, Robert D. Loss, Thomas Walczyk, and Thomas Prohaska. Atomic weights of the elements 2013 (IUPAC technical report). *Pure and Applied Chemistry*, 88(3), jan 2016. `doi:10.1515/pac-2015-0305`.

[127] Guy Mélard. On the accuracy of statistical procedures in Microsoft Excel 2010. *Computational Statistics*, 29(5):1095–1128, apr 2014. `doi:10.1007/s00180-014-0482-5`.

[128] Mike Meredith and John Kruschke. *HDInterval: Highest (Posterior) Density Intervals*, 2020. R package version 0.2.2. URL: `https://CRAN.R-project.org/package=HDInterval`.

[129] B. Mickan, C.-Y. Kuo, and M. Xu. Systematic investigations of cylindrical nozzles acc. ISO 9300 down to throat diameters of 125 μm. 10th International Symposium on Fluid Flow Measurement ISFFM, 2018.

[130] M J T Milton and M G Cox. Evaluating degrees of equivalence using 'exclusive' statistics. *Metrologia*, 40(2):L1, 2003. URL: `http://stacks.iop.org/0026-1394/40/i=2/a=101`.

[131] N. M. M. Moharram, R. El Aouad, S. Al Busaidy, A. Fabricius, S. Heller, W. G. Wood, H. U. Wolf, and C. C. Heuck. International collaborative assessment study of the AHD[575] method for the measurement of blood haemoglobin. *E. Mediterr. Health J.*, 12(5):522–534, 2006.

[132] Gianfranco Molinar, Min Beciet, and Franco Pavese. *Modern Gas-Based Temperature and Pressure Measurements*. Springer US, 2014. URL: `https://www.ebook.de/de/product/23216541/gianfranco_molinar_min_beciet_franco_pavese_modern_gas_based_temperature_and_pressure_measurements.html`.

[133] Gregory E. O'Donnell and D. Brynn Hibbert. Treatment of bias in estimating measurement uncertainty. *The Analyst*, 130(5):721, 2005. `doi:10.1039/b414843f`.

[134] United States. National Bureau of Standards, G. Kulin, and P. H. Gurewitz. *Hydraulic Research in the United States 1970: Including Contributions from Canadian Laboratories*. NBS special publication. U.S. Department of Commerce, National Bureau of Standards, 1971. Available online at http://nvlpubs.nist.gov/nistpubs. Last visit on 06/08/2020. URL: `https://books.google.de/books?id=ruByhuhMjEMC`.

[135] I. Oshiro, T. Takenaka, and J. Maeda. New method for haemoglobin determination by using sodium lauryl sulfate (SLS). *Clin. Biochem.*, 15:83–88, 1982. `doi:10.1016/S0009-9120(82)91069-4`.

[136] S. Parikh, B. Parikh, C. Shah, P. Bhansali, J. Patel, and D. Joshi. Haemoglobinometry by a novel alkaline haematin detergent-575 method. *Gujarat Medical Journal*, 65(1):14–19, 2010.

[137] W. Parker. Calculations of the heat release rate by oxygen consumption for various applications. Technical Report 81-2427-1, National Bureau of Standards, Washington DC, USA, 1982.

[138] H. Passing and W. Bablok. A New Biometrical Procedure for Testing the Equality of Measurements from Two Different Analytical Methods. Application of linear regression procedures for method comparison studies in Clinical Chemistry, Part I. *J. Clin. Chem. Clin. Biochem.*, 21:709–720, 1983. `doi:10.1515/cclm.1983.21.11.709`.

[139] Bart Peeters, Kristof Peeters, Herman Van der Auweraer, Thierry Olbrechts, Francky Demeester, and Luc Wens. Experimental modal analysis using camera displacement measurements: a feasibility study. In Enrico Primo Tomasini, editor, *Sixth International Conference on Vibration Measurements by Laser Techniques: Advances and Applications*, volume 5503, pages 298 – 309. International Society for Optics and Photonics, SPIE, 2004. `doi:10.1117/12.579867`.

[140] F. Pennecchi, P. G. Spazzini, and A. Malengo. Calibration Curves Computing Software (CCC). `https://www.inrim.eu/research-development/quality-life/ccc-software`, 2015. Accessed: 2019-11-13.

[141] Francesca R. Pennecchi, Ilya Kuselman, Ricardo J.N.B. da Silva, and D. Brynn Hibbert. Risk of a false decision on conformity of an environmental compartment due to measurement uncertainty of concentrations of two or more pollutants. *Chemosphere*, 202:165–176, 2018. `doi:10.1016/j.chemosphere.2018.03.054`.

[142] S. Phillips, K. Eberhardt, and B. Parry. Guidelines for expressing the uncertainty of measurement results containing uncorrected bias. *Journal of Research of the National Institute of Standards and Technology*, 102:577–585, 1997.

[143] Martyn Plummer, Nicky Best, Kate Cowles, and Karen Vines. Coda: Convergence diagnosis and output analysis for mcmc. *R News*, 6(1):7–11, 2006. URL: `https://journal.r-project.org/archive/`.

[144] Bruce E. Poling and John M. Prausnitz. *The Properties of Gases and Liquids*. McGraw-Hill Education, fifth edition edition, 2000. URL: `https://www.amazon.com/Properties-Gases-Liquids-Bruce-Poling-ebook/dp/B000T81T94%3FSubscriptionId%3D0JYN1NVW651KCA56C102%26tag%3Dtechkie-20%26linkCode%3Dxm2%26camp%3D2025%26creative%3D165953%26creativeASIN%3DB000T81T94`.

[145] Antonio Possolo. Five examples of assessment and expression of measurement uncertainty. *Applied Stochastic Models in Business and Industry*, 29(1):1–18, oct 2012. `doi:10.1002/asmb.1947`.

[146] Antonio Possolo and Blaza Toman. Assessment of measurement uncertainty via observation equations. *Metrologia*, 44(6):464–475, 2007. `doi:10.1088/0026-1394/44/6/005`.

[147] Antonio Possolo, Adriaan M. H. van der Veen, Juris Meija, and D. Brynn Hibbert. Interpreting and propagating the uncertainty of the standard atomic weights (IUPAC technical report). *Pure and Applied Chemistry*, 90(2):395–424, feb 2018. `doi:10.1515/pac-2016-0402`.

[148] William H. Press, Brian P. Flannery, Saul A. Teukolsky, and William T. Vetterling. *Numerical Recipes in C: The Art of Scientific Computing, Second Edition*. Cambridge University Press, 1992. URL: `https://www.amazon.com/Numerical-Recipes-Scientific-Computing-Second/dp/0521431085%3FSubscriptionId%3D0JYN1NVW651KCA56C102%26tag%3Dtechkie-20%26linkCode%3Dxm2%26camp%3D2025%26creative%3D165953%26creativeASIN%3D0521431085`.

[149] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2019. URL: `https://www.R-project.org/`.

[150] R Core Team. The R Project for Statistical Computing, 2019. Accessed 2019-07-05. URL: `https://www.r-project.org/`.

[151] A. Silva Ribeiro, J. Alves e Sousa, Maurice G. Cox, Alistair B. Forbes, L. Cordeiro Matias, and L. Lages Martins. Uncertainty analysis of thermal comfort parameters.

International Journal of Thermophysics, 36(8):2124–2149, jul 2015. `doi:10.1007/s10765-015-1888-1`.

[152] John Rice. *Mathematical statistics and data analysis*. Duxbury Press, Belmont, California, USA, 2nd edition, 1995.

[153] Stephen P. Robinson, Peter M. Harris, Justin Ablitt, Gary Hayman, Alex Thompson, A. Lee van Buren, Joseph F. Zalesak, Robert M. Drake, Alexander E. Isaev, Alexander M. Enyakov, Christopher Purcell, Zhu Houqing, Wang Yuebing, Zhang Yue, Pierre Botha, and Dieter Krüger. An international key comparison of free-field hydrophone calibrations in the frequency range 1 to 500 kHz. *The Journal of the Acoustical Society of America*, 120(3):1366, 2006. URL: `http://dx.doi.org/10.1121/1.2228790`, `doi:10.1121/1.2228790`.

[154] F. Rolle, V. Maurino, and M. Sega. Metrological traceability for benzo[a]pyrene quantification in airborne particulate matter. *Accredit. Qual. Assur.*, 17(2):191–197, April 2012. `doi:10.1007/s00769-011-0862-2`.

[155] RStudio Team. *RStudio: Integrated Development Environment for R*. RStudio, Inc., Boston, MA, 2015. URL: `http://www.rstudio.com/`.

[156] R. B. Schnabel and E. Eskow. A new modified Cholesky factorization. *SIAM J. Sci. and Stat. Comput.*, 11(6):1136–1158, 1990.

[157] Michela Sega, Francesca Pennecchi, Sarah Rinaldi, and Francesca Rolle. Uncertainty evaluation for the quantification of low masses of benzo[a]pyrene: Comparison between the law of propagation of uncertainty and the monte carlo method. *Analytica Chimica Acta*, 920:10–17, May 2016. URL: `http://dx.doi.org/10.1016/j.aca.2016.03.032`, `doi:10.1016/j.aca.2016.03.032`.

[158] Katsuhiro Shirono and Maurice Cox. Statistical reassessment of calibration and measurement capabilities based on key comparison results. *Metrologia*, 56(4):045001, Jun 2019. URL: `https://doi.org/10.1088%2F1681-7575%2Fab219e`, `doi:10.1088/1681-7575/ab219e`.

[159] J. A. Sousa, A. B. Forbes, A. S. Ribeiro, P. M. Harris, F. Carvalho, and L. Bacelar. The evaluation of uncertainty in mass calibration: possible approaches in a comparison study. *Journal of Physics: Conference Series*, 459:012033, sep 2013. URL: `https://doi.org/10.1088%2F1742-6596%2F459%2F1%2F012033`, `doi:10.1088/1742-6596/459/1/012033`.

[160] T. Srivastava, H. Negandhi, S.B. Neogi, J. Sharma, and R. Saxena. Methods for haemoglobin estimation: A Review of "What Works". *J. Hematol. Transfus.*, 2:1028–1034, 2014.

[161] Stan Development Team. RStan: the R interface to Stan, 2018. R package version 2.17.3. URL: `http://mc-stan.org/`.

[162] Stan Development Team. RStan: the R interface to Stan, 2018. R package version 2.18.2. URL: `http://mc-stan.org/`.

[163] A G Steele, B M Wood, and R J Douglas. Exclusive statistics: simple treatment of the unavoidable correlations from key comparison reference values. *Metrologia*, 38(6):483, 2001. URL: `http://stacks.iop.org/0026-1394/38/i=6/a=2`.

[164] Anatoliy Swishchuk. Table of basic derivatives. http://people.ucalgary.ca/ aswish/A-MAT219TABLES_W11.pdf. Accessed 2019-06-05.

[165] V Synek. Attempts to include uncorrected bias in the measurement uncertainty. *Talanta*, 65(4):829–837, 2005. `doi:10.1016/j.talanta.2004.07.038`.

[166] E. Trapet, M. Franke, F. Härtig, H. Schwenke, F. Wäldele, M. Cox, A. Forbes, F. Delbressine, P. Schellekens, M. Trenk, H. Meyer, G. Moritz, Th. Guth, and N. Wanner. Traceability of Co-ordinate Measurements According to the Method of the Virtual Measuring Machine. Final Project Report MAT1-CT94-0076, PTB-Report F-35 (Part 1 and 2), 1999.

[167] M. Usman, M. Moinuddin, and S.A. Ahmed. Role of iron deficiency anemia in the propagation of beta thalassemia gene. *Korean J. Hematol.*, 46:41–44, 2011. `doi:10.5045/kjh.2011.46.1.41`.

[168] J. Vallet and C. Windenberger. Improvement of thermodynamic calculations used for the flow rate of sonic nozzles. 10th International Conference on Flow Measurement FLOMEKO, 2000.

[169] A. M. H. van der Veen, T. P. Linsinger, and J. Pauwels. Uncertainty calculations in the certification of reference materials. 2. Homogeneity study. *Accreditation and Quality Assurance*, 6(1):26–30, jan 2001. URL: `http://dx.doi.org/10.1007/s007690000238`, `doi:10.1007/s007690000238`.

[170] Adriaan M. H. van der Veen. Bayesian analysis of homogeneity studies in the production of reference materials. *Accreditation and Quality Assurance*, 22(6):307–319, nov 2017. `doi:10.1007/s00769-017-1292-6`.

[171] Adriaan M H van der Veen and Katarina Hafner. Atomic weights in gas analysis. *Metrologia*, 51(1):80, 2014. URL: `http://stacks.iop.org/0026-1394/51/i=1/a=80`.

[172] Adriaan M H van der Veen, Gerard Nieuwenkamp, Rob M Wessel, Masaaki Maruyama, Gwi Suk Heo, Yong doo Kim, Dong Min Moon, Bernhard Niederhauser, Manuela Quintilii, Martin J T Milton, Maurice G Cox, Peter M Harris, Franklin R Guenther, George C Rhoderick, L A Konopelko, Y A Kustikov, V V Pankratov, D N Selukov, V A Petrov, and E V Gromova. International comparison CCQM-K46: Ammonia in nitrogen. *Metrologia*, 47(1A):08023, 2010. URL: `http://stacks.iop.org/0026-1394/47/i=1A/a=08023`.

[173] Adriaan M. H. van der Veen and J. Pauwels. Uncertainty calculations in the certification of reference materials. 1. Principles of analysis of variance. *Accreditation and Quality Assurance*, 5(12):464–469, dec 2000. URL: `http://dx.doi.org/10.1007/s007690000237`, `doi:10.1007/s007690000237`.

[174] Adriaan M.H. van der Veen. Bayesian methods for type A evaluation of standard uncertainty. *Metrologia*, 55(5):670–684, jul 2018. `doi:10.1088/1681-7575/aad103`.

[175] Adriaan M.H. van der Veen and Gerard Nieuwenkamp. Revision of ISO 19229 to support the certification of calibration gases for purity. *Accreditation and Quality Assurance*, 24(5):375–380, aug 2019. `doi:10.1007/s00769-019-01402-x`.

[176] Joële Viallon, Edgar Flores, Faraz Idrees, Philippe Moussay, Robert Ian Wielgosz, D Kim, Y D Kim, S Lee, S Persijn, L A Konopelko, Y A Kustikov, A V Malginov, I K Chubchenko, A Y Klimov, O V Efremova, Z Zhou, A Possolo, T Shimosaka, P Brewer, and T Macé. CCQM-K90, Formaldehyde in nitrogen, 2 μmol mol^{-1} Final report. *Metrologia*, 54(1A):08029–08029, jan 2017. `doi:10.1088/0026-1394/54/1a/08029`.

[177] Gerd Wübbeler, Peter M Harris, Maurice G Cox, and Clemens Elster. A two-stage procedure for determining the number of trials in the application of a monte carlo method for uncertainty evaluation. *Metrologia*, 47(3):317–324, may 2010. `doi:10.1088/0026-1394/47/3/023`.

[178] R. C. Weast. *CRC Handbook of Chemistry and Physics: A Ready-Reference Book of Chemical and Physical Data*. CRC Press Inc, 64th edition, 1984.

[179] Sten Westgard and David Armbruster. *Risk, Error and Uncertainty: Laboratory Quality Management in the Age of Metrology*. ELSEVIER, 2017. URL: `https://www.ebook.de/de/product/28452346/sten_westgard_david_armbruster_risk_error_uncertainty_lab_q.html`.

[180] N. J. White. Anaemia and malaria. *Malaria Journal*, 17:371–387, 2018. `doi:10.1186/s12936-018-2509-9`.

[181] K. Witt, H. U. Wolf, C. Heuck, M. Kammel, A. Kummrow, and J. Neukammer. Establishing traceability of photometric absorbance values for accurate measurements of the haemoglobin concentration in blood. *Metrologia*, 50(5):539–548, 2013. URL: `https://doi.org/10.1088%2F0026-1394%2F50%2F5%2F539`, `doi:10.1088/0026-1394/50/5/539`.

[182] R. Zander, W. Lang, and H. U. Wolf. Alkaline haematin D-575, a new tool for the determination of haemoglobin as an alternative to the cyanhaemiglobin method. *Clinica chimica acta*, 136(1):83–93, 1984.

Index

Acronyms

EA European co-operation for Accreditation. 17

GUF GUM Uncertainty Framework. 33, 203

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

LPU law of propagation of uncertainty. 27, 29

MCM Monte Carlo method. 27, 29, 31–33, 38, 203

MCMC Markov Chain Monte Carlo. 17, 18

SBI Single Burning Item. 33–35, 38