Good practice in evaluating measurement uncertainty

Compendium of examples

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

27 July 2021

Good practice in evaluating measurement uncertainty

A.M.H. van der Veen[1], M.G. Cox[2], J. Greenwood[3], A. Bošnjaković[4], V. Karahodžić[4], S. Martens[5], K. Klauenberg[5], C. Elster[5], S. Demeyer[6], N. Fischer[6], J.A. Sousa[7], O. Pellegrino[7], L.L. Martins[8], A.S. Ribeiro[8], D. Loureiro[8], M.C. Almeida[8], M.A. Silva[8], R. Brito[8], A.C. Soares[8], K. Shirono[9], F. Pennecchi[10], P.M. Harris[2], S.L.R. Ellison[11], F. Rolle[10], A. Alard[6], T. Caebergs[12], B. de Boeck[12], J. Pétry[12], N. Sebaïhi[12], P. Pedone[13], F. Manta[13], M. Sega[10], P.G. Spazzini[10], I. de Krom[1], M. Singh[11], T. Gardiner[2], R. Robinson[2], T. Smith[2], T. Arnold[2], M. Reader-Harris[14], C. Forsyth[14], T. Boussouara[14], B. Mickan[5], C. Yardin[6], M. Čaušević[4], A. Arduino[10], L. Zilberti[10], U. Katscher[15], J. Neukammer[5], S. Cowen[11], A. Furtado[7], J. Pereira[7], E. Batista[7], J. Dawkins[2], J. Gillespie[16], T. Lowe[16], W. Ng[16], J. Roberts[17], M. Griepentrog[18], A. Germak[10], O. Barroso[19], A. Danion[19], B. Garrido[20], S. Westwood[21], A. Carullo[13,21], S. Corbellini[21], and A. Vallan[21]

[1]VSL, Thijsseweg 11, 2629 JA Delft, the Netherlands
[2]National Physical Laboratory (NPL), Hampton Road, Teddington, Middlesex, TW11 0LW, United Kingdom
[3]United Kingdom Accreditation Service (UKAS), 2 Pine Trees, Chertsey Lane, Staines-upon-Thames TW18 3HR, United Kingdom
[4]Institute of Metrology of Bosnia and Herzegovina (IMBIH), Sarajevo, Bosnia and Herzegovina
[5]Physikalisch-Technische Bundesanstalt (PTB), Braunschweig and Berlin, Germany
[6]Laboratoire national de métrologie et d'essais (LNE), 29 avenue Roger Hennequin 78197 Trappes Cedex, France
[7]Portuguese Institute for Quality (IPQ), Rua António Gião, 2, 2829-513 Caparica, Portugal
[8]National Laboratory for Civil Engineering (LNEC), Av. do Brasil, 101, 1700-066 Lisbon, Portugal
[9]National Metrology Institute of Japan (NMIJ), Tsukuba, Japan
[10]Istituto Nazionale di Ricerca Metrologica (INRIM), Strada delle Cacce 91, 10135 Torino, Italy
[11]Laboratory of the Government Chemist (LGC), Queens Road, Teddington, TW11 0LY, UK
[12]Federale Overheidsdienst Economie, KMO, Middenstand en Energie, Brussels, Belgium
[13]L'Ente Italiano di Accreditamento (ACCREDIA-DT), Strada delle Cacce 91, Torino, Italy
[14]National Engineering Laboratory (NEL), Scottish Enterprise Technology Park, Reynolds Ave, East Kilbride, Glasgow, G75 0QF, United Kingdom
[15]Philips Research Laboratories, Roentgenstraße 24-26, 22335 Hamburg, Germany
[16]Virginia Tech, Aerospace and Ocean Engineering, 225 Stanger St. Blacksburg, VA 24061, USA
[17]Rolls-Royce plc, Development and Experimental Engineering, Derby DE 24 8BJ, UK
[18]Bundesanstalt für Materialforschung und -prüfung (BAM), Unter den Eichen 87, 12205 Berlin, Germany
[19]World Anti-Doping Agency (WADA), Science and Medicine Department, Montreal, Canada
[20]National Institute of Metrology, Quality and Technology (Inmetro), Chemical and Thermal Metrology Division, Brazil
[21]Bureau International des Poids et Mesures (BIPM), Chemistry Department, Sèvres, France
[22]Politecnico di Torino, Corso Duca degli Abruzzi 24, Torino, Italy

Contents

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.

All examples have been peer-reviewed and assessed for internal consistency and compliance with guidance in the GUM suite of documents.

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

This project 17NRM05 "Examples of Measurement Uncertainty Evaluation" has received funding from the EMPIR programme co-financed by the Participating States and from the European Union's Horizon 2020 research and innovation programme.

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 [1–6]. 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 [6].

Many laboratories implement ISO/IEC 17025 [7] to underpin their competence. Producers of (certified) reference materials implement in many cases both ISO/IEC 17025 and ISO 17034 [8] for the same purpose. In proficiency testing, the requirements for demonstrating competence are laid down in ISO/IEC 17043 [9]. 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 [10].

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 [11], 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

Overview

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

Table 2.1 provides an overview of the examples contained in this document, their key words and, if available, a pointer to the dataset.

Table 2.1: Overview of examples

Example	Description	Key words	Dataset
E1.1	Two-point and multipoint calibration — application to pH measurement	calibration; bracketing; polynomial regression; correlation	
E1.2	Straight-line calibration in errors-in-variables models	See E4.3, E5.3 and E6.2.	
E1.3	Bayesian approach applied to the mass calibration example in JCGM 101:2008	Bayesian inference; prior; mass; calibration	[12]
E1.4	Evaluation of measurement uncertainty in SBI – Single Burning Item reaction to fire test	fire testing; single burning item	[13]
E1.5	Statistical reassessment of calibration and measurement capabilities based on key comparison results	key comparison; gauge block; CMC; Bayesian inference	[14]
E1.6	Model-based unilateral degrees of equivalence in analysis of a regional metrology organization key comparison	key comparison; degree of equivalence	
E1.7	Measurement uncertainty when using quantities that change at a linear rate — use of quartz He reference leaks to calibrate an unknown leak	calibration function; covariance; reference leak	
E1.8	Factoring effects such as calibration corrections and drift into uncertainty evaluations	poor practice; errors; conformance probability	
E2.1	Conformity assessment of an influenza medication as a multicomponent material	conformity assessment; multicomponent material; risk of false decision; correlated test results	[15]
E2.2	Measurement models involving additive or multiplicative corrections	calibration correction; errors; conformance probability; pressure	[16]
E2.3	Conformity assessment of mass concentration of total suspended particulate matter in air	conformity assessment; producer's and consumer's risk; total suspended particulates in air; mass concentration; log-normal prior distribution	[17]
E2.4	Uncertainty evaluation of nanoparticle size by AFM, by means of an optimised Design of Experiment for a hierarchical mixed model in a Bayesian framework approach	AFM; mixed model; Bayesian inference; design of experiment	[18]

Continued on next page

Table 2.1 – continued from previous page

Example	Description	Key words	Dataset
E2.5	GUM-LPU uncertainty evaluation — importing measurement traceability from a conformity statement	metrological traceability; conformity statement; OIML classification; Geometrical Product Specification (GPS)	[19]
E3.1	Evaluation of measurement uncertainty in average areal rainfall – uncertainty propagation for three methods	rain fall; GUM uncertainty framework; Monte Carlo method	
E3.2	Uncertainty evaluation for the quantification of low masses of benzo[a]pyrene	Monte Carlo method; GUM uncertainty framework; ploycyclic aromatic hydrocarbon; mass	[20]
E3.3	Calibration of an analyser for NO_x using gas mixtures prepared with mass flow controllers	Gas mixtures; nitrogen oxides; dynamic dilution; mass flow controllers; chemiluminescence analyser; calibration; correlation; Weighted Total Least-Squares	[21]
E3.4	Measurement uncertainty for routine testing of metals in soil	soil; trace elements; acid extraction; atomic emssion spectroscopy	[22]
E3.5	Comparison of methods for flow measurement in closed conduits based on measurement uncertainty	flow measurement; GUM uncertainty framework; “Monte Carlo method	
E3.6	Greenhouse gas emission inventories	greenhouse gases; inventory; correlation	
E3.7	Greenhouse gas emission inventories — emission estimates calculated by measurement of ambient mixing ratios combined with inverse modelling	greenhouse gases; inventory; emisson; inverse modelling	
E3.8	Preparation of calibration gas mixtures of NH_3 in nitrogen using permeation	GUM uncertainty framework; finite resolution; purity; validation of OLS	[23]
E4.1	Evaluation of measurement uncertainty in totalization of volume measurements in drinking water supply networks	flow measurement; water supply; totalisation	[24]
E4.2	Uncertainty of the orifice-plate discharge coefficient	flow measureement; orifice plate; discharge coefficient	[25]
E4.3	Calibration of a sonic nozzle as an example for quantifying all uncertainties involved in straight-line regression	measurement model; GUM; Monte Carlo method; straight-line regression; correlation; weighted total least-squares	[26]
E4.4	Measurement uncertainty evaluation of the load loss of power transformers	electrical power; transformer; load loss; alternating current	[27]
E4.5	Evaluation of measurement uncertainty in thermal comfort	thermal comfort; implicit model formulation; Monte Carlo method	[28]
E4.6	Bayesian evaluation of a between-bottle homogeneity study in the production of reference materials	Bayesian inference; ANOVA; between-bottle homogeneity; reference material; proficiency test	
E4.7	Flow meter calibration using the master meter method	flow measurement; calibration; master meter method	[29]
E4.8	Pressure drop due to gas leakage in a pressurized vessel	pressure drop; leak test; correlation	[30]
E5.1	2D or 3D image as a set of pixels or voxels to compute a quantity	pixellation; image metrology; SPECT imaging; adsorbed dose; nanoparticle; AFM	[31]
E5.2	Magnetic resonance-based electric properties tomography	electric properties tomography; magnetic resonance imaging; covariance matrix; shrinkage estimation; law of propagation of uncertainty	[32]
E5.3	Quantifying uncertainty when comparing measurement methods – Haemoglobin concentration as an example of correlation in straight-line regression	measurement model; GUM; straight-line regression; correlation; weighted total least squares; method comparison; haemoglobin; AHD; HiCN	[33]
E5.4	Suitability of a Monte Carlo approach for uncertainty evaluation in rheology problems	rheology; viscosity; GUM uncertainty framework; Monte Carlo method	

Continued on next page

Table 2.1 – continued from previous page

Example	Description	Key words	Dataset
E5.5	Uncertainty calculation methodologies in microflow measurements: comparison of GUM, GUM-S1 and Bayesian approach	microflow; GUM uncertainty framework; Monte Carlo method; Bayesian inference	
E5.6	Specification requirements of temperature in medical applications	conformity; decisison rule; accuracy	[34]
E6.1	Measurement uncertainty evaluation for turbofan nozzle thrust derived from non-intrusive flow measurements	inflight thrust; complex step method; finite differences method; Monte Carlo method; correlation	
E6.2	Calibration of a torque measuring system – GUM uncertainty evaluation for least-squares versus Bayesian inference	measurement model; GUM; Bayesian inference; calibration; straight-line regression; least squares estimation; torque; VDI/VDE 2600 Blatt 2	[35]
E6.3	Calibration and measurement uncertainty in hardness verification	Vickers-; Knoop-; Rockwell-; Brinell-; Instrumented Indentation- hardness test; correlated quantities; effective degrees of freedom; conformity assessment	
E6.4	Evaluation of measurement uncertainty in the calibration of a mobile optical measurement system	calibration; coordinate measurement machine	[36]
E6.5	Evaluation of measurement uncertainty associated with the quantification of ephedrine in anti-doping testing	top-down approach; CRM; proficiency testing	
E6.6	Measurement uncertainty in a multiplexed data-acquisition system	data acquisition system; electrical quantity	[37]
E6.7	Temperature measurement with a microcontroller based board	temperature; microcontroller boeard; dithering	[38]

Chapter 3

Using the Monte Carlo method

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

3.1 Preamble

One of the complicating factors in the evaluation and propagation of measurement uncertainty is the competence in mathematics and statistics required to perform the calculations. Nevertheless, standards such as ISO/IEC 17025 [7], ISO 15189 [39] and ISO 17034 [8] 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) [2] 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) [3], 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 [3] 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 [2] and further elaborated in GUM-S1 [3]. 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) [4] 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 [2]. 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 [40, 41]. 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 [11]. 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 [42]. 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.

3.2 Monte Carlo method

The heart of the Monte Carlo method of GUM-S1 can be summarised as follows [3, 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 [3], and a multivariate output quantity, see GUM-S2 [4].

GUM-S1 [3, clause 6.4] describes the selection of appropriate probability density functions for the input quantities, thereby supplementing the guidance given in the GUM [2, 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 [43, 44]. 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 [3, clause 7.9] and since then an improved method has been developed and published [45].

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 [3, 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 [46], which is also implemented in many other programming environments, including MATLAB and MicroSoft Excel (since version 2010, see [47]). Based on this random number generator, there are generators available for a number of probability distributions [11].

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.

3.3 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 [11]. It is available for Windows, MacOS and a variety of UNIX platforms (including FreeBSD and Linux) [48].

Users of Windows, MacOS, and a number of Linux distributions may also wish to download and install RStudio [49], 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)).

3.4 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 3.1). 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`)

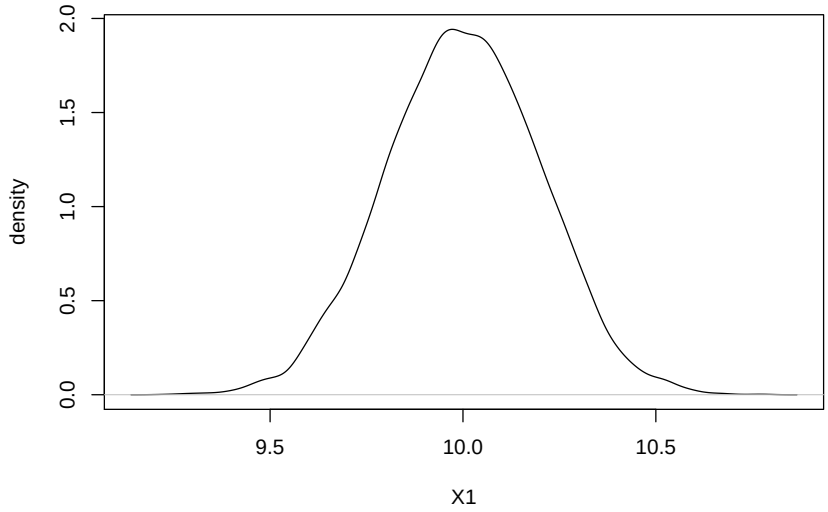


Figure 3.1: Density plot of the random variable X1 having a normal distribution with mean 10.0 and standard deviation 0.2

and gamma distribution (`rgamma`). There exists a package (extension) called "trapezoid" [50] implementing among others the trapezoidal distribution, a package called "mvtnorm" [51] implementing the multivariate normal distribution (useful when some of the input quantities are dependent [3]), and a package called "triangle" [52] 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 [3, table 1] are available in R.

The arc sine distribution can be implemented as follows in R. According to GUM-S1 [3, clause 6.4.6.1], a U-shaped random variable X on the interval $[a, b]$ can be obtained through

$$X = \frac{a+b}{2} + \frac{b-a}{2} \sin \Phi$$

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 3.2.

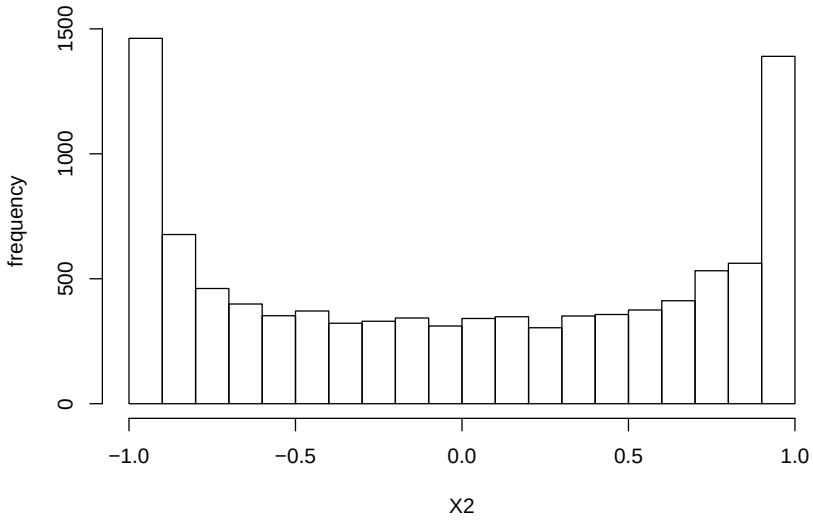


Figure 3.2: Histogram of the random variable X2 containing $M = 10\,000$ samples having an arcsine distribution between -1 and 1

3.5 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 of Pure and Applied Chemistry) [53]. The relative atomic masses that apply to "normal materials" are called standard atomic weights [53, 54]. Their interpretation is described in an IUPAC technical report [55].

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
```

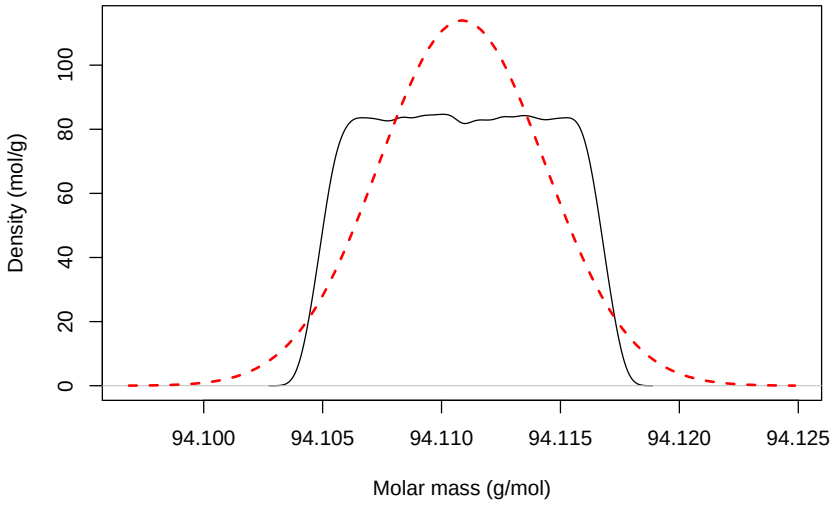


Figure 3.3: Output probability density function of the molar mass of phenol and superimposed a normal distribution with the same mean and standard deviation

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 [53]. 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) [56]. 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).

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 3.3. 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 $0.0035\,g\,\mathrm{mol^{-1}}$. The expanded uncertainty is $0.0059\,\mathrm{g\,mol^{-1}}$. The coverage factor is 1.67.

3.6 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 [57] 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 [4]. 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 [57, S2]:

$$m_X = m_S + \delta m_D + \delta m + \delta m_C + \delta B, \tag{3.1}$$

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 [3]. These probability density functions are described in the original example [57].

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 [3, 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 (3.1)) can be coded in R as follows:

```
# 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 [3]. 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 [3], 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 [56]. 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).

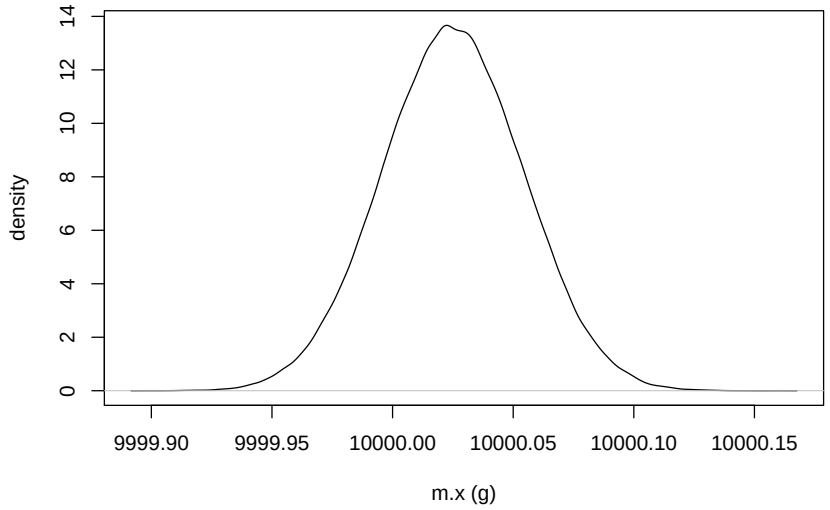


Figure 3.4: Probability density function of the output quantity m.x

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 3.4. 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)
```

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 (3.1)). That the output quantity has an (approximately) normal distribution is reflected in the coverage factor obtained from the Monte Carlo method.

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

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

In this case, one might also be willing to state that the output probability density function is a normal distribution. Whereas in this case such a statement can be made, in most cases the output probability density function cannot directly be approximated by a well-known analytic probability density function. Comparison of the three results listed above with those from the LPU would imply that for comparable data LPU would be fit for purpose in a subsequent uncertainty evaluation. In a subsequent uncertainty evaluation, with m_X as one of the input quantities, the above information suffices to apply the law of propagation of uncertainty, say [2].

3.7 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 [2], GUM-S2 [4] and EA 4/02 [57] 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 [58, 59]) is barely mentioned in such documents. Numerical approximation of the sensitivity coefficients [40, 41] 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 [57].

The R package `numDeriv` provides the function `grad` (from gradient) that returns from a function a generally good approximation, using Richardson extrapolation [60], 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 [57]. The code is also valid for measurement models with non-trivial sensitivity coefficients [41].

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 [4]) 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 4

Bayesian inference in R and RStan

A.M.H. van der Veen

4.1 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 [11] and rstan [61], 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.

4.2 Introduction

The mass example in EA 4/02 [57] was introduced in chapter 3 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) [3]. 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 [2] or the Monte Carlo method of GUM Supplement 1 (GUM-S1) [3]. 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 [2, 57] 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 [11], and the package rstan [62] that enables writing Bayesian models in a straightforward manner. This environment and the use of RStan for Bayesian inference have been introduced previously [63, 64].

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 [3]. 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 [2, 57] and obviously only makes sense if that interval is (approximately) symmetric.

4.3 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 [57] 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 (3.1). 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 [3] (see also chapter 3.

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

$$\Delta m_{\mathrm{X}} = \Delta m_{\mathrm{S}} + \delta d_{\mathrm{D}} + \delta m + \delta m_{\mathrm{C}} + \delta B \tag{4.1}$$

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 [63, 64]. This sensitivity is one of the hurdles to be taken when performing iterative calculations (as MCMC is [65]).

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 [2] are declared as model parameters in the `parameters` block. By default, Stan assigns these variables a rectangular distribution over their domain of validity [62]. 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 [65]) . 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 (4.1)). 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 [2] deals with type B evaluations of standard uncertainty is 'weakly Bayesian' [66] 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 (4.1) 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 [61, 62].

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 [57]. 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 [2, 63]).

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 (4.1). 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 [65]. More details have been given elsewhere [63, 64]. Running the model with 21 000 iterations and a warmup of 1000 iterations, using 4 chains yields the following 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 [65]. The final column, labelled `Rhat`, gives the ratio of the between-chain and within-chain variance. For convergence, it should be close to one [63, 65].

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 4.1.

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 $10\,\text{mg}/\sqrt{3} \approx 5.8\,\text{mg}$, 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 $15\,\text{mg}/\sqrt{3} \approx 8.7\,\text{mg}$, and the mean is zero [57]; the results obtained from the MCMC are 8.7 mg and 0 mg respectively.

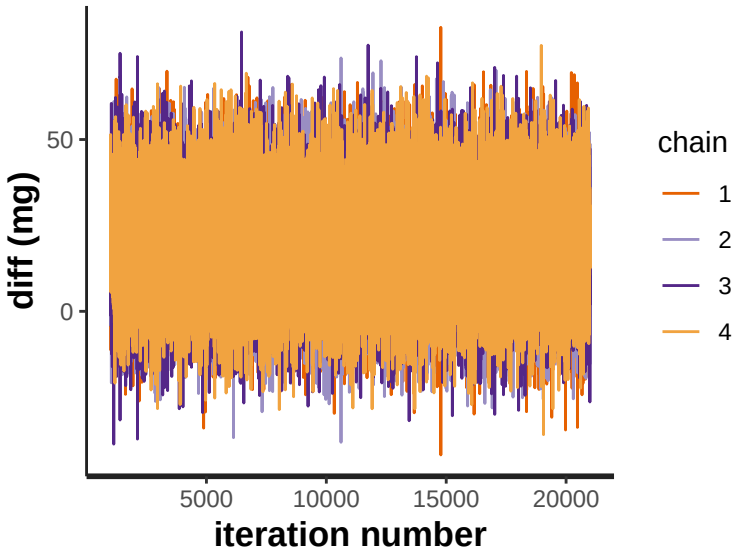


Figure 4.1: Trace plot of the model parameter `diff`

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 [57]).

The final hurdle in this example is the reproduction of the expanded uncertainty, which is stated to be 59 mg [57]. The MCMC calculation provides for all parameters the 95 % coverage intervals (see the output discussed previously). Before attempting to compute the expanded uncertainty as the half-width of an approximately symmetric coverage interval, the shape of the posterior of Δm_x should be assessed for symmetry. This posterior is shown in figure 4.2.

From figure 4.2, 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) [57]. Alternatively, the expanded uncertainty could also be computed as the half-width of the 95 % coverage interval.

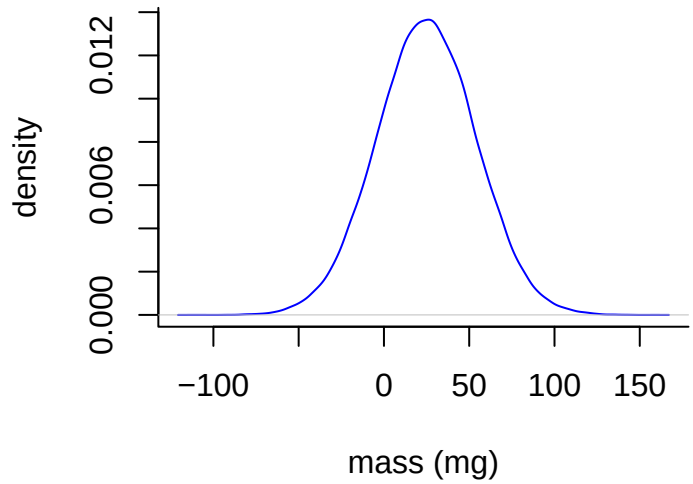


Figure 4.2: Posterior of mass difference of the weight being calibrated from its nominal mass

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 [63, 65, 67]);
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! [3, 65]).

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 $(-\infty, +\infty)$ [65]. 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 [63, 64]. 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 (4.1). This part of the model behaves in the same way as it would do when using the Monte Carlo method of GUM-S1 [2].

Chapter 5

Understanding and treating correlated quantities in measurement uncertainty evaluation

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

5.1 Preamble

The evaluation of covariances, as required by the GUM (Guide to the expression of uncertainty in measurement) [2, clause 8, step 4] is often omitted in practice when evaluating measurement uncertainty for a variety of reasons. In many instances, this habit has the consequence that it produces an *incorrect* value for the resulting uncertainty, possibly leading to wrong decisions based on the measurement result. In fact, not evaluating a covariance is equivalent to setting its value to zero, which should be justified.

Where for many experimenters the evaluation of measurement uncertainty using the law of propagation of uncertainty (LPU) of the GUM [2] or the Monte Carlo method (MCM) of GUM Supplement 1 (GUM-S1) [3] is already challenging, understanding how correlations between input variables arises and handling it is even more so. In this primer, we provide an introduction to the subject, illustrated by several examples. The purpose of these examples is to show the versatility of the GUM suite of documents [2–5] in dealing with this aspect of evaluating measurement uncertainty. In some of these examples, we also evaluate the consequences of ignoring correlations.

The GUM [2] provides two important formulae for evaluating and working with covariances due to common input effects. The first equation to be mentioned is GUM formula (13), which is the law of propagation of uncertainty for correlated input quantities. This formula is lesser known than its counterpart for independent variables [2, equation (10)], but it is the preferred choice when using a linear or linearized measurement model with interdependent input quantities. The second formula to be mentioned is given in [2, formula (F.1)], which provides the expression for the calculation of the covariance between quantities X_i and X_j, depending on a set of input quantities Q_ℓ with $\ell = 1 \dots L$:

$$u(X_i, X_j) = \sum_{\ell=1}^{L} \frac{\partial X_i}{\partial Q_\ell} \frac{\partial X_j}{\partial Q_\ell} u^2(Q_\ell), \tag{5.1}$$

where $u(x)$ denotes the standard uncertainty associated with x, the partial derivatives are evaluated at the estimates of the X_i and only those terms contribute for which $\partial X_i/\partial Q_\ell \neq 0$ *and* $\partial X_j/\partial Q_\ell \neq 0$, which implies that only those quantities Q_ℓ contribute on which both X_i and X_j depend [2].

When using the Monte Carlo method of GUM-S1 or GUM Supplement 2 (GUM-S2) [3, 4], care is automatically taken of dependencies between variables in the measurement model. If the input quantities are dependent, then samples should be drawn from a joint probability density function [4, clause 7.3] [3, clause 6.4]. The method of GUM-S2 provides a sample of the joint output probability density function for the output quantities forming the measurand, and from this sample the covariances or correlation coefficients can be readily obtained [4, clause 7.6].

5.2 Covariance and correlation

Covariance and correlation are two measures for the dependence between (estimates of) quantities. A covariance is expressed in the units of the quantities involved, whereas the correlation coefficient is dimensionless. The correlation coefficient, which always lies between -1 and 1, is defined as [2, clause 5.2.2]

$$r(X_i, X_j) = \frac{u(X_i, X_j)}{u(X_i)u(X_j)}, \tag{5.2}$$

where X_i and X_j are the quantities involved and $u(X_i, X_j)$ the covariance between them. From expression (5.2), if the covariance is zero, the correlation coefficient is also zero. Strong correlation between X_i and X_j is indicated by $|r| \approx 1$. If $r > 0$, then X_i and X_j are positively correlated, that is, an increase in X_i leads to an increase in X_j. Similarly, if the variables are negatively correlated, an increase in X_i leads to a decrease in X_j. In situations where it is difficult to compute a covariance, it is often simpler to estimate a correlation coefficient. With expression (5.2), the corresponding covariance $u(X_i, X_j)$ can then be obtained.

5.3 Correlation arising from Type A evaluation

When simultaneous observations are repeatedly made of several input quantities, it is likely that there are correlations to be associated with the estimates of these quantities.

EXAMPLE *Simultaneous observations of voltage, current and phase*

A treatment of simultaneous observations is given in GUM-S2 [4, clause 6.2] where, from a circuit element, the following quantities are concurrently measured six times:

V: amplitude of a sinusoidally-alternating potential difference across the terminals,
I: amplitude of alternating current passing through it,
ϕ: phase angle of the alternating potential difference relative to the alternating current.

Any systematic error present in V, I and ϕ is considered negligible. The $n = 6$ indications are given in table 5.1.

Estimates of V, I and ϕ are the averages $\bar{V}$, $\bar{I}$ and $\bar{\phi}$ of the observations. The associated standard uncertainties $u(\bar{V})$, $u(\bar{I})$ and $u(\bar{\phi})$ are calculated in the usual way, for example,

$$u^2(\bar{V}) = \frac{1}{n(n-1)} \sum_{j=1}^{n} (V_j - \bar{V})^2 = \frac{1}{30} \sum_{j=1}^{6} (V_j - \bar{V})^2$$

Table 5.1: Indications for voltage, current and phase angle of a circuit element [4]

Set i	V_i/V	I_i/mA	ϕ_i/rad
1	5.007	19.663	1.0456
2	4.994	19.639	1.0438
3	5.005	19.640	1.0468
4	4.990	19.685	1.0428
5	4.999	19.678	1.0433
6	4.999	19.661	1.0445

and covariances evaluated using GUM formula (17) [2, clause 5.2.3], for example,

$$u(\bar{V},\bar{I}) = \frac{1}{n(n-1)}\sum_{j=1}^{n}(V_j-\bar{V})(I_j-\bar{I}) = \frac{1}{30}\sum_{j=1}^{6}(V_j-\bar{V})(I_j-\bar{I}).$$

From these standard uncertainties and covariances, the correlation coefficients can be computed using formula (5.2). The estimates and standard uncertainties are summarized in table 5.2. The information concerning the correlations is summarized in table 5.3. In this matrix, the off-diagonal elements contain the value of the correlation coefficient for the corresponding pair of variables. (The correlation coefficient between a variable and itself is unity by default. Furthermore, since $r(X_1,X_2) = r(X_2,X_1)$, the elements below the main diagonal in table 5.3 are not shown, as they are the mirror image of the upper triangle of the matrix.)

Table 5.2: Estimates of V, I and ϕ and associated standard uncertainties [4]

	V/V	I/mA	ϕ/rad
Estimate	4.9990	19.6610	1.044 46
Std. unc.	0.0026	0.0077	0.000 61

Table 5.3: Correlation coefficients between voltage, current and phase angle of a circuit element

	V	I	ϕ
V	1	−0.355	0.858
I		1	−0.645
ϕ			1

This approach also finds application in the post-processing of data obtained using the Monte Carlo method from GUM-S1 and GUM-S2 [3, 4] to obtain, for example, standard uncertainties, covariances, correlation coefficients or a covariance matrix.

5.4 Correlation relating to common input effects

Correlation relating to common input effects arises frequently in metrology. Such an effect is due, for instance, to the same measuring system, physical measurement standard or reference datum. Consider a simple model for two measurands:

$$X_1 = Q_0 + Q_1,$$
$$X_2 = Q_0 + Q_2,$$

where Q_0, Q_1 and Q_2 denote the input quantities and X_1 and X_2 the output quantities. Further assume that Q_0, Q_1, Q_2 are mutually independent. Using equations (F.1) and (F.2) from the GUM [2],

$$\begin{aligned} u^2(X_1) &= u^2(Q_0) + u^2(Q_1), \\ u^2(X_2) &= u^2(Q_0) + u^2(Q_2), \\ u(X_1, X_2) &= u^2(Q_0). \end{aligned} \tag{5.3}$$

EXAMPLE *Calibration of a liquid-in-glass thermometer*

Two platinum resistance thermometers (PRTs) are used to calibrate a liquid-in-glass thermometer. The measurement model for the temperature, using the two PRTs takes the form

$$T = \frac{1}{2}(T_1 + T_2),$$

where T_1 and T_2 denote the quantities representing temperature obtained using the two PRTs, T denotes the aggregated temperature and $u(T)$ is given by using the law of propagation of uncertainty (LPU) of the GUM [2]:

$$u^2(T) = \frac{1}{4}[u^2(T_1) + u^2(T_2) + 2u(T_1, T_2)]. \tag{5.4}$$

Uncertainty budgets for the two PRTs are given in table 5.4. The reference thermometer and the uniformity of the temperature in the bath are considered to be identical for the two PRTs, hence giving rise to correlation.

Table 5.4: Uncertainty budgets for two PRTs

Source	Standard uncertainty contribution/K	
	PRT 1	PRT 2
Reference thermometer to calibrate PRTs	0.002 50	0.002 50
Uniformity of thermo-regulated bath	0.000 98	0.000 98
Drift	0.000 06	0.000 06
Repeatability	0.000 13	0.000 04
Adjustment from calibration curve	0.000 39	0.000 60
Stability	0.001 62	0.001 62
Combined standard uncertainty	0.003 16	0.003 19

Table 5.4 gives the main sources of uncertainty: the first two are systematic effects (the same for each PRT). All other effects are different for the two PRTs. Application of formula (5.4) provides $u(T) = 0.002\,94\,\mathrm{K}$, whereas without the covariance term $2u(T_1, T_2)$, $u(T) = 0.002\,25\,\mathrm{K}$. The former standard uncertainty is 30 % greater than the latter. In this case, ignoring correlation gives an optimistically small value for required standard uncertainty. Whether this difference is seen as significant depends on the context and the subsequent use of the quantity T.

An example such as that above can straightforwardly be handled using expressions of the form (5.3). With several dependent input quantities, the application of the GUM is somewhat more difficult, especially if the natural form of the expressions are to be used. For example, in studying the performance of a jet engine, the measurement model has five output quantities or measurands, $Y_1, \ldots, Y_5$ and involves a chain of calculation steps (see example E6.1). In the first step, Y_1 is specified in terms of (some of) the input quantities. In the second step, Y_2 is specified in terms of Y_1 and the input quantities, and so on. Such a situation is hard to handle for practitioners without the necessary skills in partial differentiation. Ways to treat this example that avoid the need for these skills are considered in example E6.1.

5.5 Identifying joint effects

In section 5.4, an example was shown where two PRTs were used to calibrate a liquid-in-glass thermometer. The use of multiple measurement standards occurs much more widely, and the importance of evaluating possible correlations is not always fully recognised. In this section, issues arising with the use of multiple standards are further explored, showing how the magnitude of the correlation can be evaluated.

In many areas of calibration, multiple measurement standards are used. For instance, two weights are used jointly to calibrate a balance or two resistors are used in an electrical circuit. Often, these weights or resistors are calibrated by the same laboratory. Calibration laboratories typically use a specific measurement standard for a particular calibration.

EXAMPLE *Weights calibrated against the same measurement standard*

Two 10 kg weights are submitted to the same calibration laboratory and calibrated against the same measurement standard. Consequently, the calibration results will be correlated. If in a subsequent calibration these two weights are used and their total mass needs to be computed, a basic measurement model takes the form

$$m_{\text{stds}} = m_1 + m_2, \tag{5.5}$$

where m_{stds} denotes the mass of the combined weight, and m_1 and m_2 the masses of the respective weights. Using the variant of LPU for correlated variables (see equation (13) in the GUM [2]), the variance associated with m_{stds} can be expressed as

$$u^2(m_{\text{stds}}) = u^2(m_1) + u^2(m_2) + 2u(m_1, m_2), \tag{5.6}$$

where $u(m_1, m_2)$ denotes the covariance between m_1 and m_2. The evaluation of this covariance requires knowledge about the uncertainty budgets for m_1 and m_2. In the simplest case, the calibration of the two weights takes place by comparison with the same standard with mass m_{S}. In that case, the covariance between m_1 and m_2 can be computed as

$$u(m_1, m_2) = u^2(m_{\text{S}}),$$

where m_{S} denotes the mass of the standard used in the calibration of both weights. When $u(m_1) = u(m_2) = u(m_{\text{S}})$, $u(m_{\text{stds}}) = \sqrt{2}u(m_{\text{S}})$ would be obtained under the assumption that m_1 and m_2 are independent. If they are dependent, $u(m_{\text{stds}}) = 2u(m_{\text{S}})$ is obtained using equation (5.6). These results are markedly different. The impact on the uncertainty in a subsequent measurement depends of course on how dominant $u(m_{\text{stds}})$ is in the uncertainty budget of the measurement involving the two weights.

Consider first the situation where detailed metrological information is available. Suppose now, following [5], that the basic measurement model (5.5) takes the extended form [57]:

$$m_j = m_{\text{S}} + \delta m_{\text{D}} + \delta m_j + \delta m_{\text{C},j} + \delta B_j, \tag{5.7}$$

where the symbols have the following meaning:

m_j	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_j	observed difference in mass between the unknown mass and the standard,
$\delta m_{C,j}$	correction for eccentricity and magnetic effects,
δB_j	correction for air buoyancy,

where the index j refers to either weight 1 or weight 2. The terms with index j are modelled as independent for the calibration of the two weights. For the observed mass difference δm_j, this assumption is consistent with the assumption made in the example that the readings of the balance are independent within a calibration, so it is reasonable to make the same assumption for the readings for the two calibrations. Buoyancy effects depend on, for example, the density of the weights and the air density. For the two weights being calibrated, the densities can be assumed to be independent (unless these have been determined in the same experiment), and a similar reasoning applies to magnetic effects. The effect of eccentricity relates to the placement of the weights on the balance, and if this placement is such that the weight is placed in the centre of the pan, it can also be modelled as independent if the eccentric loading of the balance is small. The conventional mass of the standard and the drift of the standard are, considering the way that they have been modelled, the same in both calibrations, hence contributing to the covariance. Using equation (F.1) in the GUM [2], the covariance can be expressed as

$$u(m_1, m_2) = u^2(m_S) + u^2(\delta m_D). \tag{5.8}$$

The simple model in equation (5.6) does not permit a refinement of the treatment of correlations. The choice is rather binary, which may be fit for purpose, but not necessarily so. For example, air buoyancy is a quantity that can be modelled as independent if in time (and by implication, weather conditions are sufficiently different), but is better modelled as fully dependent if the two weights had been calibrated on the same day (sharing the same air density). Correlation in this case increases the covariance [calculated in a similar fashion as shown in equation (5.6)].

It is not usual that the customer of a laboratory has detailed knowledge about the uncertainty budget of the calibration ordered. If requested, referring to the above example, the laboratory could provide a value for the covariance $u(m_1, m_2)$ or, equivalently, the correlation coefficient $r(m_1, m_2)$ so that the customer can use that information when using the two weights together. If such information is not available, the customer could make an attempt to guess the correlation coefficient [68], and use that estimate in subsequent calculations. Estimating the value of a correlation coefficient is often easier than the corresponding covariance since, as stated in section 5.2, the correlation coefficient lies in the interval $-1 \leq r \leq 1$. An important aspect in the estimation is the sign of the correlation coefficient, which can often also be determined by reasoning. In the case of the weights, it is likely that if m_1 is estimated high, m_2 will also be estimated high, so the correlation coefficient is positive (given that we know the same standard has been used for both calibrations, we can rule out that $r = 0$ is plausible).

5.6 Induced correlations

Correlations can also arise due to the fact that quantities are calculated from a more fundamental set of quantities and then subsequently used in combination with the quantities from which they are calculated.

EXAMPLE *Key comparisons*

An instance of induced correlation arises in key comparisons, or more generally, in interlaboratory comparisons with a consensus value.

In a key comparison, the capabilities of participating laboratories are assessed by degrees of equivalence (DoEs). A DoE is defined as the difference between the measured value from a national metrology institute (NMI) and the key comparison reference value (KCRV), together with its associated expanded uncertainty at a 95 % probability level. In proficiency testing, a similar metric is used in comparisons for calibration laboratories, the E_n-score, which is the ratio of the difference and associated expanded uncertainty of a measured value with respect to the assigned value [69].

The KCRV is usually calculated from the measured values $x_1, \ldots, x_N$ from the N NMIs participating in the key comparison, which are nominally measuring the same measurand, and $u(x_1), \ldots, u(x_N)$ the reported associated standard uncertainties. The commonest estimator of the KCRV is the weighted mean (WM), computed for independent measured values as [70]:

$$x_{\mathrm{KCRV}} = u^2(x_{\mathrm{KCRV}}) \sum_{i=1}^{N} \frac{x_i}{u^2(x_i)}, \qquad u^2(x_{\mathrm{KCRV}}) = \left[\sum_{i=1}^{N} \frac{1}{u^2(x_i)} \right]^{-1}.$$

The DoE (d_j, U_j) for Laboratory j is

$$d_j = x_j - x_{\mathrm{KCRV}}, \qquad U_j = k u(d_j), \qquad u^2(d_j) = u^2(x_j) - u^2(x_{\mathrm{KCRV}}),$$

where k denotes the coverage factor, equal to 1.96 under the assumption that the measured values are normally distributed. Note the minus sign in the expression for $u^2(d_j)$, which results from the correlation between x_j and x_{KCRV} arising from the dependence of x_{KCRV} on x_j [70].

As a simple illustration of the effect of including (or ignoring) the correlation between the KCRV and the measured values, consider the arithmetic mean as KCRV for $N = 3$ laboratories:

$$x_{\mathrm{KCRV}} = \frac{1}{3}(x_1 + x_2 + x_3),$$

and suppose that the three laboratories all report a standard uncertainty of one unit. From the law of propagation of uncertainty, it follows that

$$\begin{aligned} u^2(x_{\mathrm{KCRV}}) &= \frac{1}{9}[u^2(x_1) + u^2(x_2) + u^2(x_3)] \\ &= \frac{1}{9}[1^2 + 1^2 + 1^2] = \frac{1}{3}. \end{aligned}$$

Hence, $u(x_{\mathrm{KCRV}}) = 1/\sqrt{3}$, the familiar 'root N effect' on the standard uncertainty when computing the average of N values under the assumption of independence. The squared standard uncertainty of the DoE for laboratory 1 becomes

$$\begin{aligned} u^2(d_1) &= u^2(x_1) - u^2(x_{\mathrm{KCRV}}) \\ &= 1 - \frac{1}{3} = \frac{2}{3}. \end{aligned}$$

If correlation were (completely wrongly) ignored, that is, LPU for independent quantities were applied, the following value for the squared standard uncertainty would instead become

$$\begin{aligned} u^2(d_1) &= u^2(x_1) + u^2(x_{\mathrm{ref}}) \\ &= 1 + \frac{1}{3} = \frac{4}{3}. \end{aligned}$$

This value for the squared standard uncertainty is twice as large as the correct value. This problem is well-known in the literature, and appropriate ways of dealing with it have been developed [70], including the elimination of the correlation [71–73].

5.7 Missing or ignored correlation

Sometimes, correlations are missing or ignored, leading to poor decisions or logical absurdities.

EXAMPLE *pH measurement*

The pH, the negative decadic logarithm (base 10) of the activity of hydrogen ion in a solution is the most measured kind-of-quantity in chemistry. The measurement of pH often involves measuring an electrical potential E for standard solutions and temperatures and using reference tables to convert these to pH values. In short, the process starts with measuring the potential E_X of the cell in a test solution at 'standard temperature'. Then, tabular entries in the reference material certificate for standard solutions S_1 and S_2 that 'bracket' test solution in terms of E are identified. These entries are then linearly interpolated to provide pH_X, the pH of the test solution.

Laboratories following the IUPAC (2002) recommendations [74] will not take correlation into consideration when calculating the uncertainty associated with the measured pH_X, and so might be reporting optimistically small measurement uncertainties. Little pH literature on accounting for correlation in interpolated values is available, yet it is important to take the dependencies into account (for a fuller discussion, see example E1.1).

Nevertheless, the information necessary may already be available. Certificates of buffer solutions often give two uncertainties, one for an SI-traceable value and one, much smaller, that omits the uncertainty associated with the Bates-Guggenheim (BG) convention [75]. The effect of the BG convention is similar (if not largely the same) for all pH measurements, so the squared uncertainty arising from the BG convention could be taken as an approximate covariance, implying a correlation coefficient ≈ 1. For a typical case, pH is estimated as 7.0109 with a standard uncertainty of 0.0041 assuming independence and when evaluated with full correlation 0.0051, which matches the uncertainty in the certified values.

The logical absurdity of such a situation is now explained. Figure 5.1 shows interpolated pH values for temperatures between those listed on the reference material certificate. The uncertainty associated with the interpolated values is substantially smaller than those stated on the certificate. Repeating the process (so, taking the interpolated values as reference and calculating new values for the temperatures given on the certificate) will, with the same attitude, lead to even smaller uncertainties. Indefinite repetition of this procedure makes it evident that uncertainties evaluated in this way are not credible. Actually, when taking into account the correlations, a meaningful reduction of the uncertainty would not be seen. Such a reduction would only be credible if the reference points were truly independent, so carrying more information.

5.8 Removable and unremovable correlation

In many instances, it is possible to eliminate correlation by expressing an output quantity in terms of a different set of (independent) input quantities. This idea also underlies equation (5.1) in the GUM [2]. Correlation between (estimates of) quantities typically arises when these quantities are evaluated using the same pool of data.

EXAMPLE 1 *Straight-line calibration*

A well-known example of removable correlation is straight-line regression, where usually the calculated slope and intercept are correlated. If the values of the slope and intercept are used in a subsequent calculation, it is essential that the covariance between them is taken into consideration [76, 77]. Ignoring the covariance can lead to a gross overstatement of the uncertainty (see also example E5.3). There are also instances where the uncertainty would be understated, see example E4.3. The covariance between slope and intercept will become zero if the data set consisting of N pairs (x_i, y_i) is shifted to become $(x_i - \overline{x}, y_i)$, where $\overline{x} = \sum_i x_i/N$. An example illustrating this elimination is given in annex H.3 of the GUM [2, annex H.3.5].

EXAMPLE 2 *Key comparisons (once more)*

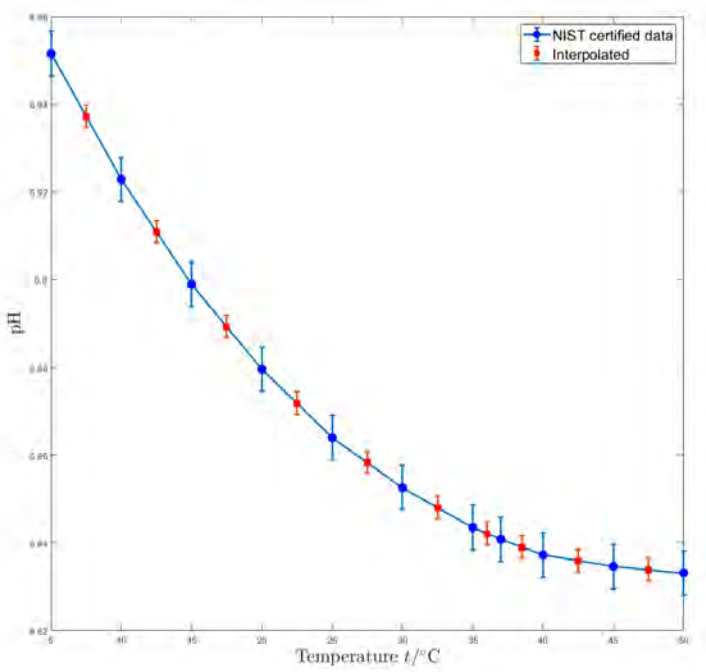


Figure 5.1: NIST certified values (filled circles, blue) and interpolated values (boxes, red) and the associated expanded uncertainties when ignoring the correlation between the certified values

In the case of the key comparison data in the example in section 5.6, working with correlations can be avoided by consistently working only with the original quantities x_1, x_2 and x_3. Instead of using the quantity x_{ref}, which is important in its own right [78], the DoE can be established by expressions that solely depend on the mutually independent input quantities as follows. Use

$$d_1 = x_1 - x_{\text{ref}} = x_1 - \frac{1}{3}(x_1 + x_2 + x_3) = \frac{1}{3}(2x_1 - x_2 - x_3).$$

Then, apply the LPU in its simplest form to give the same result:

$$u^2(d_1) = \frac{4}{9}u^2(x_1) + \frac{1}{9}u^2(x_2) + \frac{1}{9}u^2(x_3) = \frac{1}{9}(4+1+1) = \frac{6}{9} = \frac{2}{3}.$$

5.9 Multivariate measurement models

5.9.1 General

For a measurement model in which there are multiple quantities having correlations or multiple measurement equations or both, GUM-S2 [4] presents a matrix treatment of the law of propagation of uncertainty through such a model.[1] This treatment is not only useful for multivariate measurement models, but also for multi-stage measurement models (section 5.10) where it is cumbersome or even impossible to express the measurand in a set of uncorrelated input quantities. Matrix expressions for uncertainty propagation are valuable since computer systems such as MATLAB [79], Python [80] and R [11] support working with matrices. Even mainstream

[1] Some of the material here is based on that guide.

spreadsheet software contains the essential functions for implementing the law of propagation of uncertainty for multivariate measurement models, that is, the functions for matrix multiplication and matrix transpose. Often, quite involved calculations can be coded in a few lines using matrix calculus.

Although multivariate measurement models receive little treatment in the GUM [2], the same underlying GUM principles may be used to propagate estimates of the input quantities and the associated uncertainties through the measurement model to obtain estimates of the output quantities and their associated uncertainties. Mathematical expressions for the evaluation of uncertainty are stated using matrix-vector notation, rather than the subscripted summations given in the GUM, because generally such expressions are more compact and more naturally implemented within modern software packages and computer languages.

The law of propagation of uncertainty, in a more generalized form than presented in the GUM, for multivariate measurement models is a valuable tool for propagating uncertainties. It also caters for covariances associated with the input quantities and obtaining those associated with the output quantities.

For the application of the law of propagation of uncertainty, the same information concerning the input quantities as for the univariate measurement model treated in the GUM is used:

5.9.2 Explicit multivariate measurement models

1. An explicit multivariate measurement model specifying the relationship between an output quantity $\boldsymbol{Y} = (Y_1, \ldots, Y_m)^\top$ and an input quantity $\boldsymbol{X} = (X_1, \ldots, X_N)^\top$, takes the form

   $$\boldsymbol{Y} = \boldsymbol{f}(\boldsymbol{X}), \qquad \boldsymbol{f} = (f_1, \ldots, f_m)^\top,$$

   where $\boldsymbol{f}$ denotes the multivariate measurement function.

2. An estimate $\boldsymbol{x} = (x_1, \ldots, x_N)^\top$ of $\boldsymbol{X}$;

3. The covariance matrix

   $$\boldsymbol{V}_{\boldsymbol{x}} = \begin{bmatrix} u(x_1, x_1) & \cdots & u(x_1, x_N) \\ \vdots & \ddots & \vdots \\ u(x_N, x_1) & \cdots & u(x_m, x_N) \end{bmatrix},$$

   of dimension $N \times N$, associated with $\boldsymbol{x}$ containing the covariances $u(x_i, x_j)$, $i = 1, \ldots, N$, $j = 1, \ldots, N$, associated with x_i and x_j. $u(x_j, x_j) = u^2(x_j)$ denotes the variance (squared standard uncertainty) associated with x_j.

Given an estimate $\boldsymbol{x}$ of $\boldsymbol{X}$, an estimate of $\boldsymbol{Y}$ is $\boldsymbol{f}(\boldsymbol{x})$. A generic formula for propagating $\boldsymbol{V}_{\boldsymbol{x}}$ through $\boldsymbol{f}(\boldsymbol{X})$ is [4]

$$\boldsymbol{V}_{\boldsymbol{y}} = \boldsymbol{C}\boldsymbol{V}_{\boldsymbol{x}}\boldsymbol{C}^\top, \tag{5.9}$$

where $\boldsymbol{V}_{\boldsymbol{y}}$ is the output covariance matrix of dimension $m \times m$ associated with $\boldsymbol{y}$ and

$$\boldsymbol{C} = \begin{bmatrix} \dfrac{\partial f_1}{\partial X_1} & \cdots & \dfrac{\partial f_1}{\partial X_N} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial f_m}{\partial X_1} & \cdots & \dfrac{\partial f_m}{\partial X_N} \end{bmatrix}$$

is the sensitivity matrix of dimension $m \times N$ evaluated at $\boldsymbol{X} = \boldsymbol{x}$ [81, page 29].

EXAMPLE 1 *Resistance and reactance of a circuit element once more*

Continuing the example in section 5.3, a bivariate measurement model for R and X in terms of V, I and ϕ is

$$R = f_1(V, I, \phi) = \frac{V}{I}\cos\phi, \qquad X = f_2(V, I, \phi) = \frac{V}{I}\sin\phi. \tag{5.10}$$

Using the above general notation, $N = 3$, $m = 2$, $\boldsymbol{X} \equiv (V, I, \phi)^\top$ and $\boldsymbol{Y} \equiv (R, X)^\top$.

An estimate $\boldsymbol{y} \equiv (\widehat{R}, \widehat{X})^\top$ of resistance R and reactance X is obtained by evaluating expressions (5.10) at an estimate $\boldsymbol{x} \equiv (\widehat{V}, \widehat{I}, \widehat{\phi})^\top$ of the input quantity $\boldsymbol{X}$.

The covariance matrix $\boldsymbol{V}_y$ of dimension 2×2 associated with $\boldsymbol{y}$ is given by formula (5.9), where $\boldsymbol{C}_x$ is the sensitivity matrix of dimension 2×3 given by evaluating

$$\begin{bmatrix} \dfrac{\partial f_1}{\partial V} & \dfrac{\partial f_1}{\partial I} & \dfrac{\partial f_1}{\partial \phi} \\ \dfrac{\partial f_2}{\partial V} & \dfrac{\partial f_2}{\partial I} & \dfrac{\partial f_2}{\partial \phi} \end{bmatrix} = \begin{bmatrix} \dfrac{\cos\phi}{I} & -\dfrac{V\cos\phi}{I^2} & -\dfrac{V\sin\phi}{I} \\ \dfrac{\sin\phi}{I} & -\dfrac{V\sin\phi}{I^2} & \dfrac{V\cos\phi}{I} \end{bmatrix}$$

at $\boldsymbol{X} = \boldsymbol{x}$, and $\boldsymbol{V}_x$ is the covariance matrix of dimension 3×3 associated with $\boldsymbol{x}$.

EXAMPLE 2 *Calibration of mass standards*

This example constitutes an instance of a multi-stage model (section 5.10).

A set of q mass standards of unknown mass values $\boldsymbol{m} = \left(m_1, \ldots, m_q\right)^\top$ is calibrated by comparison with a reference kilogram, using a mass comparator, a sensitivity weight for determining the comparator sensitivity, and a number of ancillary instruments such as a thermometer, a barometer and a hygrometer for determining the correction due to air buoyancy. The reference kilogram and the sensitivity weight have masses m_{R} and m_{S}, respectively. The calibration is carried out by performing, according to a suitable measurement procedure, a sufficient number k of comparisons between groups of standards, yielding apparent, namely, in-air differences $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_k)^\top$. Corresponding buoyancy corrections $\boldsymbol{b} = (b_1, \ldots, b_k)^\top$ are calculated. In-vacuo mass differences $\boldsymbol{X}$ are obtained from the sub-model $\boldsymbol{X} = \boldsymbol{f}(\boldsymbol{W})$, where $\boldsymbol{W} = \left(m_{\mathrm{R}},\ m_{\mathrm{S}},\ \boldsymbol{\delta}^\top,\ \boldsymbol{b}^\top\right)^\top$.

An estimate $\boldsymbol{y} \equiv (\widehat{m}_1, \ldots, \widehat{m}_q)^\top$ of the masses $\boldsymbol{m}$ is typically given by the least-squares solution of the overdetermined system of equations $\boldsymbol{A}\boldsymbol{m} = \boldsymbol{X}$, where $\boldsymbol{A}$ is a matrix of dimensions $k \times q$ with elements equal to $+1$, -1 or zero, respecting the uncertainties involved. With this choice, the estimate $\boldsymbol{y}$ is given by

$$\boldsymbol{y} = \boldsymbol{V}_y \boldsymbol{A}^\top \boldsymbol{V}_x^{-1} \boldsymbol{x}, \tag{5.11}$$

where $\boldsymbol{x}$ is the estimate of the input quantity $\boldsymbol{X}$, and the covariance matrix $\boldsymbol{V}_y$ of dimension $q \times q$ associated with $\boldsymbol{y}$ is given by $\boldsymbol{V}_y = \left(\boldsymbol{A}^\top \boldsymbol{V}_x^{-1} \boldsymbol{A}\right)^{-1}$. $\boldsymbol{V}_x$ is the covariance matrix of dimension $k \times k$ associated with $\boldsymbol{x}$. A more detailed description of the sub-model, as well as a procedure for obtaining $\boldsymbol{V}_x$ in terms of $\boldsymbol{V}_w$, the covariance matrix associated with the estimate $\boldsymbol{w}$ of $\boldsymbol{W}$, is available [82].

The multivariate measurement model for this example can be expressed as

$$\boldsymbol{Y} = \boldsymbol{V}_y \boldsymbol{A}^\top \boldsymbol{V}_x^{-1} \boldsymbol{X},$$

where the measurement function is $\boldsymbol{U}_y \boldsymbol{A}^\top \boldsymbol{U}_x^{-1} \boldsymbol{X}$. In terms of the general notation, $N = k$, $m = q$ and $\boldsymbol{Y} \equiv \boldsymbol{m}$.

It is preferable computationally to obtain the estimate given by formula (5.11) by an algorithm based on orthogonal factorization [83], rather than use this explicit formula.

5.9.3 Implicit multivariate measurement models

An implicit multivariate measurement model specifies a relationship between an output quantity $\boldsymbol{Y} = (Y_1, \ldots, Y_m)^\top$ and an input quantity $\boldsymbol{X} = (X_1, \ldots, X_N)^\top$, and takes the form

$$\boldsymbol{h}(\boldsymbol{Y}, \boldsymbol{X}) = \boldsymbol{0}, \qquad \boldsymbol{h} = (h_1, \ldots, h_m)^\top.$$

Given an estimate $\boldsymbol{x}$ of $\boldsymbol{X}$, an estimate $\boldsymbol{y}$ of $\boldsymbol{Y}$ is given by the solution of the system of equations

$$\boldsymbol{h}(\boldsymbol{y}, \boldsymbol{x}) = \boldsymbol{0}, \tag{5.12}$$

generally to be solved numerically for $\boldsymbol{y}$, using, for example, Newton's method [84] or a variant of that method, starting from an approximation $\boldsymbol{y}^{(0)}$ to the solution.

The covariance matrix $\boldsymbol{V}_{\boldsymbol{y}}$ of dimension $m \times m$ associated with $\boldsymbol{y}$ is evaluated from the system of equations [4]

$$\boldsymbol{C}_{\boldsymbol{y}} \boldsymbol{V}_{\boldsymbol{y}} \boldsymbol{C}_{\boldsymbol{y}}^\top = \boldsymbol{C}_{\boldsymbol{x}} \boldsymbol{V}_{\boldsymbol{x}} \boldsymbol{C}_{\boldsymbol{x}}^\top, \tag{5.13}$$

where $\boldsymbol{C}_{\boldsymbol{y}}$ is the sensitivity matrix of dimension $m \times m$ containing the partial derivatives $\partial h_\ell / \partial Y_j$, $\ell = 1, \ldots, m$, $j = 1, \ldots, m$, and $\boldsymbol{C}_{\boldsymbol{x}}$ is the sensitivity matrix of dimension $m \times N$ containing the partial derivatives $\partial h_\ell / \partial X_i$, $\ell = 1, \ldots, m$, $i = 1, \ldots, N$, all derivatives being evaluated at $\boldsymbol{X} = \boldsymbol{x}$ and $\boldsymbol{Y} = \boldsymbol{y}$. The covariance matrix $\boldsymbol{V}_{\boldsymbol{y}}$ in expression (5.13) is not defined if $\boldsymbol{C}_{\boldsymbol{y}}$ is singular. 1

Formally, the covariance matrices $\boldsymbol{V}_{\boldsymbol{x}}$ and $\boldsymbol{V}_{\boldsymbol{y}}$ are related by

$$\boldsymbol{V}_{\boldsymbol{y}} = \boldsymbol{C} \boldsymbol{V}_{\boldsymbol{x}} \boldsymbol{C}^\top, \qquad \boldsymbol{C} = \boldsymbol{C}_{\boldsymbol{y}}^{-1} \boldsymbol{C}_{\boldsymbol{x}}, \tag{5.14}$$

where

$$\boldsymbol{C} = \boldsymbol{C}_{\boldsymbol{y}}^{-1} \boldsymbol{C}_{\boldsymbol{x}}, \tag{5.15}$$

a matrix of sensitivity coefficients of dimension $m \times N$.

EXAMPLE 1 *Set of pressures generated by a pressure balance*

The pressure p generated by a pressure balance is defined implicitly by the equation

$$p = \frac{m_{\mathrm{w}} (1 - \rho_{\mathrm{a}} / \rho_{\mathrm{w}}) g_\ell}{A_0 (1 + \lambda p)(1 + \alpha \delta\theta)}, \tag{5.16}$$

where m_{w} is the total applied mass, ρ_{a} and ρ_{w} are, respectively, the densities of air and the applied masses, g_ℓ is the local acceleration due to gravity, A_0 is the effective cross-sectional area of the balance at zero pressure, λ is the distortion coefficient of the piston-cylinder assembly, α is the coefficient of thermal expansion, and $\delta\theta$ is the deviation from a 20 °C reference Celsius temperature [85].

Let $p_1, \ldots, p_q$ denote the generated pressures for, respectively, applied masses $m_{\mathrm{w},1}, \ldots, m_{\mathrm{w},q}$ and temperature deviations $\delta\theta_1, \ldots, \delta\theta_q$.

In terms of the general notation, the vector $(A_0, \lambda, \alpha, \delta\theta_1, m_{\mathrm{w},1}, \ldots, \delta\theta_q, m_{\mathrm{w},q}, \rho_{\mathrm{a}}, \rho_{\mathrm{w}}, g_\ell)^\top$ is denoted by $\boldsymbol{X}$ and $(p_1, \ldots, p_q)^\top$ by $\boldsymbol{Y}$ with $N = 6 + 2q$ and $m = q$.

$\boldsymbol{X}$ and $\boldsymbol{Y}$ are related by the measurement model

$$h_j(\boldsymbol{Y}, \boldsymbol{X}) = A_0 p_j \big(1 + \lambda p_j\big)\big(1 + \alpha \delta\theta_j\big) - m_{\mathrm{w},j} (1 - \rho_{\mathrm{a}} / \rho_{\mathrm{w}}) g_\ell = 0, \qquad j = 1, \ldots, q. \tag{5.17}$$

An estimate $\widehat{p}_j$ of p_j is obtained by solving an equation of the form (5.17) given estimates of A_0, λ, α, $\delta\theta_j$, $m_{\mathrm{w},j}$, ρ_{a}, ρ_{w} and g_ℓ. The resulting estimates $\widehat{p}_1, \ldots \widehat{p}_q$ have associated covariances because they all depend on the measured quantities A_0, λ, α, ρ_{a}, ρ_{w} and g_ℓ.

The covariance matrix $\boldsymbol{V}_y$ of dimension $q \times q$ associated with $\boldsymbol{y} \equiv (\widehat{p}_1, \ldots, \widehat{p}_q)^\top$ is evaluated from expression (5.13), where $\boldsymbol{C}_y$ is the sensitivity matrix of dimension $q \times q$ containing the partial derivatives $\partial h_\ell / \partial Y_j$, $\ell = 1, \ldots, q$, $j = 1, \ldots, q$, and $\boldsymbol{C}_x$ is the matrix of dimension $q \times (6 + 2q)$ containing the partial derivatives $\partial h_\ell / \partial X_i$, $\ell = 1, \ldots, q$, $i = 1, \ldots, 6 + 2q$, both evaluated at $\boldsymbol{X} = \boldsymbol{x}$ and $\boldsymbol{Y} = \boldsymbol{y}$, and $\boldsymbol{U}_x$ is the covariance matrix of dimension $(6 + 2q) \times (6 + 2q)$ associated with $\boldsymbol{x}$.

A measurement function [giving Y_j ($\equiv p_j$) explicitly as a function of $\boldsymbol{X}$] can be determined in this case as the solution of a quadratic equation. Such a form is not necessarily numerically stable. Moreover, measurement models involving additional, higher-order powers of p are sometimes used [86]. Determination of an explicit expression is not generally possible in such a case.

EXAMPLE 2 *Use of the Soave-Redlich-Kwong equation of state*

Equations of state are key to calculating the pVT (pressure, volume, temperature)-behaviour of fluids (gases, liquids, vapours). Often, with very limited information such as the critical properties (temperature, pressure) of the components, a reasonable prediction of the phase behaviour, a pressure, specific volume or density can be obtained. To illustrate the calculation of the specific volume V_m of a fluid at a given temperature and pressure, we demonstrate the use of the Soave-Relich-Kwong (SRK) equation of state, not because it is arguably the best of all cubic equations of state, but because the calculation serves as a template for using more complicated (and accurate) equations of state. The equation of state can be formulated as [87]

$$V_\mathrm{m}^3 + \alpha_1 V_\mathrm{m}^2 + \alpha_2 V_\mathrm{m} + \alpha_3 = 0, \tag{5.18}$$

where

$$\alpha_1 = -\frac{RT}{p}, \qquad \alpha_2 = -b^2 - \frac{RT}{p} b + \frac{a}{p}, \qquad \alpha_3 = -\frac{ab}{p} \tag{5.19}$$

where a and b are coefficients of the SRK that depend on the critical properties of the fluid and the saturated vapour pressure at a reduced temperature of 0.7 [87], and the temperature. In equation (5.19), R denotes the ideal gas constant, T the thermodynamic temperature and p the pressure. Depending on the values of p and T, equation (5.18) has one or three real roots. The latter occurs if the saturation pressure is used for p, corresponding to the temperature T. Such a vapour-liquid equilibrium calculation, where p is also part of the measurand has been discussed elsewhere [88].

In this context, it is important to note that the coefficients $\boldsymbol{\alpha}$ in equation (5.18) depend on a set of variables that can be presumed to be independent [see equation (5.19)] [88].

Propagation of uncertainty can be performed readily using expression (5.15) for implicit multivariate measurement models, enabling inclusion of correlations between the α_i, that is,

$$u^2(V_\mathrm{m}) = \left(3V_\mathrm{m}^2 + 2\alpha_1 V_\mathrm{m} + \alpha_2\right)^{-2} \boldsymbol{C}_{\boldsymbol{\alpha}} \boldsymbol{U}_{\boldsymbol{\alpha}} \boldsymbol{C}_{\boldsymbol{\alpha}}^\mathrm{T}.$$

5.10 Multi-stage measurement models

A measurement model with several stages is known as a multi-stage model: the output from one stage becomes the input to the next stage (which may have further inputs). For instance, calibration is a two-step process.

EXAMPLE 1 *Straight-line calibration*

Given data comprising a set of pairs of stimulus and response values, following the VIM [89], the process for straight-line calibration constitutes two stages:

1. Determine calibration parameter values (intercept a and slope b for a straight line $y = a + bx$ expressing response y in terms of stimulus x) from the data, and

2. Use a and b to produce the stimulus $x_0 = (y_0 - a)/b$ (the measurand) corresponding to a measured response y_0.

Given the data and associated uncertainties (and covariances if present), a and b and their covariance matrix $\boldsymbol{V}_{[a,b]}$ are produced in stage 1. Then, in stage 2, a and b and their covariance matrix $\boldsymbol{V}_{[a,b]}$ and y_0 are used to provide x_0 and $u(x_0)$ using

$$x_0 = \frac{1}{b}(y_0 - a), \qquad u^2(x_0) = \frac{1}{b^2}[1 \;\; x_0]\boldsymbol{V}_{[a,b]}[1 \;\; x_0]^\top + \frac{1}{b^2}u^2(y_0). \tag{5.20}$$

Full details of the computation are contained in [90] and are based on the expressions in section 5.9.1. The process naturally consists of two stages. Note that the covariance matrix determined in stage 1 must be made available to stage 2 to evaluate the standard uncertainty $u(x_0)$.

When the recipient of the calibration results is the calibration laboratory itself, the two stages can be combined. The covariance matrix does not even have to be reported. The key difference from the two-stage process is that the computational scheme for evaluating $\boldsymbol{V}_{[a,b]}$ is 'plugged' into formula (5.20).

The above contrasting scenarios constitute illustrations of removable and unremovable correlation in section 5.8 (with a further example — peak area determined from spectral data — below).

EXAMPLE *Peak area determined from spectral data*

A requirement in spectroscopy is the detection of peaks in a signal and the determination of peak parameters such as area and location. Peak area determination nominally constitutes a two-stage model, in which (1) raw data are filtered to reduce the effects of noise and (2) peak area is calculated using the filtered data. Covariances would be passed from the first stage to the second.

(a) Two-stage model

In the first stage data are filtered using a windowing function, each item of raw data being superseded by a filtered data item. The result is a set of filtered values, with associated standard uncertainties and covariances. Covariances arise since each filtered value is a combination of raw values and any specific raw value contributes to several filtered quantities (common input as in section 5.4).

In the second stage a linear function of the filtered data is obtained such as peak area

The peak area is estimated using the filtered data. To evaluate the standard uncertainty associated with estimated peak area, use is made of the standard uncertainties and covariances associated with the filtered data established in the first stage.

(b) Single-stage model

The peak area is expressed directly as a function of the unfiltered data by combining explicitly the above two stages. The explicit provision of covariances generated in the above first stage is not needed.

Mathematical expressions for the calculation are given in [5, clause 8.4.2].

5.11 Concluding remarks

The evaluation of covariances is usually an essential part of an evaluation of measurement uncertainty. There are several ways to incorporate these dependencies in the calculations, ranging from simply working with the forms of the law of propagation of uncertainty and the Monte Carlo method that consider dependencies between input quantities to the elimination of dependencies by re-expressing the measurement model so that the measurand depends on a set of independent or at least uncorrelated input quantities.

In practice, especially for uncertainty evaluation problems involving many variables, it is often preferable to work with matrix-vector forms of expressions for dealing with standard uncertainties and covariances. Doing so requires some knowledge above the basic skills needed to apply the law of propagation of uncertainty as given in the GUM.

Messages made in the treatment here include the following:

- When properly established, a measurement model articulates the relationships between variables.

- Accounting for covariances is an essential part of an uncertainty evaluation: their proper use can increase or decrease the obtained uncertainty over an uncertainty evaluation that disregards them.

- The propagation of a covariance matrix is not required if the stages in a multi-stage model can be combined into a single-stage model (see section 5.8).

- It is not always necessary to (re-)express the measurand(s) as a set of uncorrelated input quantities. If the covariances are evaluated, the law of propagation of uncertainty for dependent quantities [2, equation (13)] can be applied directly, or when using the Monte Carlo method, samples can be drawn from a multivariate probability density function.

- Working with multivariate methods [4] is often the easier choice, but requires some familiarity with matrix calculus.

- The Monte Carlo method [3, 4] and Bayesian inference using Markov Chain Monte Carlo (see, for example, section 5.3) also provide means to extract information about covariances between output quantities.

Chapter 6

Reporting measurement results

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

6.1 Introduction

The manner in which measurement results are reported is fundamental since in many cases they are the primary outcome of a measurement, including the evaluation of the measurement data and the uncertainty evaluation. There are various styles of reporting, most of which are 'conventional', apply to a single measurand, and are given in standards from International Organization for Standardization (ISO) and International Electrotechnical Commission (IEC) [7,8,39] and recognized guidance documents such as JCGM [2,3], European Accreditation [57], Eurachem [42], International Laboratory Accreditation Cooperation (ILAC) [91] and UKAS [92]. In addition to considering this type of reporting of measurement results, vector measurands [4] are increasingly important since there is an upsurge in measuring systems with many output quantities such as pixel images obtained in fields like surface metrology and medical physics. In these areas, the vector output quantity is often used in further stages of data processing in which case correlations that exist between these quantities must be taken into consideration to provide valid results.

Further, we consider measurement results represented by probability distributions [3, 4, 89] because of the growing interest in regarding a probability distribution as a complete statement of uncertainty [93], their importance in conformance assessment [6, 94] and the availability of relevant computational facilities [95].

6.2 Measurement result

To report a measurement result, it is important to appreciate what a measurement result comprises. According to the International Vocabulary of Metrology — Basic and General Concepts and Associated Terms (JCGM 200:2012; VIM 3rd edition) [89], it is defined as

> 2.9 (3.1) **measurement result**
> result of measurement
> set of quantity values being attributed to a measurand together with any other available relevant information

NOTE 1 A measurement result generally contains "relevant information" about the set of quantity values, such that some may be more representative of the measurand than others. This may be expressed in the form of a probability density function (PDF).

NOTE 2 A measurement result is generally expressed as a single measured quantity value and a measurement uncertainty. If the measurement uncertainty is considered to be negligible for some purpose, the measurement result may be expressed as a single measured quantity value. In many fields, this is the common way of expressing a measurement result.

6.3 Measurement result reporting

At its simplest, for a scalar measurand, a measurement result may be presented by an estimate of the measurand and an associated standard uncertainty. In the case of a vector measurand, a measurement result may similarly be represented by a vector of estimates and the associated covariance matrix.

Although a probability distribution is a complete description of the metrologist's judgment regarding the measurand, it alone will not generally meet reasonable requirements for reporting because, unless the recipient is well versed in statistics, it does not readily convey usable information about the measurand. Therefore, various summaries of the distribution should be provided to convey clear and meaningful information for the recipient [96]. Summary statistics might also include a coverage interval or, in the multivariate case, a description of a coverage region that contains the measurand for a given coverage probability, and possibly some other material. A coverage region, a multivariate counterpart of a coverage interval in the univariate case, is not unique, one instance being the coverage region of smallest volume.

As regards the reporting of a PDF itself, there are several options available. In a univariate case, if the PDF can be reported to be sufficiently well represented by a normal distribution, the mean and standard deviation, taken as the estimate and associated standard uncertainty of the measurand, may be all that is needed. (The GUM [2] assumes, by appealing to the central limit theorem, the PDF for the measurand is approximately normal.) If the PDF for the measurand is (or close to) a t distribution, resulting from the presence of some sample-based input quantities (Type A evaluations), the reporting is as for a normal distribution with the addition of an (effective) degrees of freedom [2]. If another PDF is assumed for the measurand, then this PDF and its parameters can be provided as the measurement result [57, 97, 98].

The use of Monte Carlo (MC) or Monte Carlo Markov Chain (MCMC) methods provides the distribution for the measurand in the form of a large sample from that distribution. The distribution may be reported in this form, as an electronic data file, or as a suitable standard statistical distribution that is a good approximation fitted to the sample. Summary statistics such as the mean, the standard deviation and a coverage interval may be computed directly from the sample [4] and could be provided additionally to the sample of the distribution.

If that PDF is to be used as input to a further evaluation, the electronic data file can be used directly for that purpose [4, 96].

The provision of suitable graphics, such as a plot of the PDF, is frequently desirable in the univariate case or even for bivariate measurement models [96]. One- or two-dimensional sections of a multivariate PDF can also be useful for visualisation.

The measurement result applies to the measurement model in hand, which can range from real, univariate and explicit to complex, multivariate and implicit. Guide JCGM 102 [4, clause 6] categorises the measurement model according to their mathematical form. The current primer does not cover the complex case; the reader is referred to [4, 99].

So that measurement results are metrologically reproducible [100], it is recommended that the reporting of results includes all data necessary for the measurement result to be reproduced by others.

It is recognised that the metadata accompanying a measurement result depends on the method used to produce it. Although it is beyond the scope of this document to comment in detail on the relative merits of methods, we state the following. Each method uses information particular to that method in specifying the evaluation problem to be solved. An instance is in the calibration of a thermometer using straight-line regression in terms of the original variables or transformed variables [2, clause H.3]. The use of a centred variable, simply obtained by shifting the origin of the independent variable, transforms the problem into a form that not only gives more straightforward reporting, since the correlation between the line parameters (section 6.3.4) is eliminated, but also simplifies the calculation.

It would be wise to alert or remind the recipient of the measurement result that any replacement of a non-linear model by a linear model has been judged to be acceptable.

There are also options within the three methods considered for any specific evaluation problem, namely, the GUM uncertainty framework (GUF) [2,4], Monte Carlo (MC) [3,4,101] and Bayesian inference [102, 103].

For the GUF, there is the choice of the law of propagation of uncertainty (LPU) based on first or higher-order terms in the Taylor expansion of the measurement function. For MC, there is choice in the number of MC trials, the random number generators (RNGs) used and the seeds chosen for those generators. For MCMC methods used in Bayesian inference, there is choice in the number of MCMC iterations, the number of iterations in the burn-in period, the RNGs and the seeds.

This primer gives advice on the items to be reported.

6.3.1 Univariate models

Preamble

A measurement model taking the form

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

relating a single output quantity Y to the input quantities $X_1, \ldots, X_N$, is termed univariate and explicit.

A measurement model taking the form

$$h(Y, X_1, \ldots, X_N) = 0,$$

where Y is a scalar output quantity, and h denotes a function of Y and the input quantities $X_1, \ldots, X_N$, is termed univariate and implicit.

The GUM uncertainty framework

For the GUM uncertainty framework [2], report

1. A specification of the measurand;

2. The measurement model, either analytically or in terms of (or making explicit reference to) an algorithm or item of software (a data repository such as Zenodo, which can also include accompanying metadata, is valuable for this purpose);

3. For implicit models only, the manner in which the output quantity is evaluated given estimates of the input quantities, by stating or referring to a formula, algorithm or software;

4. For each input quantity, the estimate and associated standard uncertainty and (for a Type A evaluation) the degrees of freedom;

5. When appropriate, for each pair of input quantities, the associated covariance or correlation coefficient;

6. As an alternative to the previous two items, the covariance matrix associated with the set of estimates of the input quantities;

7. For the output quantity, the estimate and associated standard uncertainty and, if appropriate, the effective degrees of freedom and the expanded uncertainty or (the endpoints of) a coverage interval or both for a stated coverage probability;

8. The assumed probability density function of the measurand used to obtain the coverage factor and expanded uncertainty;

9. Whether the first- or higher-order variant of LPU has been used.

EXAMPLE *Example of items 1 and 7 for the mass m_S of a measurement standard*

The measurand is the mass m_S in vacuum of a nominally 100 g weight at a stipulated time.

A measurement result for m_S is reported as '$m_S = 100.021\,47\,\mathrm{g}$ with associated standard uncertainty 0.35 mg'.

The Monte Carlo method

For the Monte Carlo method [3], report

1. Items 1, 2 and 3 in section 6.3.1;

2. The PDF for each input quantity;

3. As an alternative to 2 when input quantities are correlated, the joint PDF for those quantities;

4. For the output quantity, the estimate of that quantity and the associated standard uncertainty and, if required, a coverage interval, stating whether it is probabilistically symmetric or shortest or otherwise, and the coverage probability;

5. As a possible addition to the previous item, the PDF for the output quantity, which can be used in subsequent evaluations that make use of input PDFs;

6. The number of MC trials taken, the RNG used and the RNG seeds selected (the latter two are important if the results are to be reproduced exactly).

EXAMPLE *Example of items 1 (part), 2, 4 and 6 for the cross-sectional area A of a pipe*

The measurand is the cross-sectional area A of a specific pipe at a stipulated time and location on the pipe.

The PDFs selected for the input quantities were normal distributions with means and standard deviations equal to the estimates and associated standard uncertainties for those quantities.

A measurement result for A is an estimate of 7.92×10^3 mm^2 with an associated standard uncertainty of 0.16×10^3 mm^2.

The probabilistically symmetric coverage interval is $[7.60 \times 10^3, 8.24 \times 10^3]$ mm^2 for 95 % coverage probability.

The number of Monte Carlo trials was 1×10^7, the random number generator was the Mersenne Twister and the random number seed was 9790.

Bayesian inference

When the measurement result originates from Bayesian inference, the measurand is part of the posterior PDF, or can be computed from it. If the Bayesian model used contains multiple parameters, one of them can be the measurand. Alternatively, if the measurand is a function of these model parameters, it can be computed from these parameters. If this posterior PDF has a well-known form, such as the normal distribution or t distribution, it can be reported as such with the parameter values obtained from the Bayesian evaluation. In many cases, the posterior PDF does not have a simple well-known form, and then a similar reporting format can be chosen as for the Monte Carlo method (see section 6.3.1):

1. Items 1, 2 and 3 in section 6.3.1;
2. prior PDF for each input quantity;
3. As an alternative to 2 when input quantities are correlated, the joint prior PDF for those quantities;
4. The likelihood function(s) used for the data;
5. For the output quantity, the estimate of that quantity and the associated standard uncertainty and, if required, a coverage interval, stating whether it is probabilistically symmetric or shortest or otherwise, and the coverage probability;
6. As a possible addition to the previous item, the posterior PDF for the output quantity, which can be used in subsequent evaluations that make use of input PDFs;
7. The number of MCMC trials taken (chain length, length of the warm-up phase, number of chains), the RNG used and the RNG seeds selected (the latter two are important if the results are to be reproduced exactly).
8. The algorithm used to perform the MCMC.

6.3.2 Multivariate models

The GUM uncertainty framework in the multivariate case

For the GUM uncertainty framework in the multivariate case [4], report as for the GUM uncertainty framework in the univariate case in section 6.3.1 except that the measurand and measurement model are to be interpreted in terms of their vector or multivariate counterparts. Moreover,

no specific advice is given here on the coverage region, the multivariate counterpart of a coverage interval for the output quantity for a stipulated coverage probability: the reader is invited to consult [4] and the relevant examples therein. Item 8 has no documented counterpart in the multivariate case.

EXAMPLE *Simple bivariate measurement model [4, clause 9.2.3]*

The bivariate measurand is $\boldsymbol{Y} = (Y_1, Y_2)^\top$ is defined by the measurement model $Y_1 = X_1 + X_3$, $Y_2 = X_2 + X_3$, where the input quantities X_1, X_2 and X_3 are independent and have zero estimates and unit standard deviations. Giving additional decimal places for purposes of comparison with those for the example in section 6.3.2, the estimate $\boldsymbol{y} = (y_1, y_2)^\top$ of the bivariate output quantity $\boldsymbol{Y}$ is $y_1 = 0.000$ and $y_2 = 0.000$, the associated standard uncertainties are $u(y_1) = 1.414$ and $u(y_2) = 1.414$ and the associated covariance is $u(y_1, y_2) = 1.000$, that is,

$$\boldsymbol{y} = \begin{bmatrix} 0.000 \\ 0.000 \end{bmatrix}, \qquad \boldsymbol{u}(\boldsymbol{y}) = \begin{bmatrix} 1.414 \\ 1.414 \end{bmatrix}, \qquad \boldsymbol{V}_y = \begin{bmatrix} 2.000 & 1.000 \\ 1.000 & 2.000 \end{bmatrix}. \tag{6.1}$$

In terms of a bivariate vector $\boldsymbol{\eta}$, a 95 % elliptical coverage region for $\boldsymbol{Y}$ can be reported as

$$(\boldsymbol{\eta} - \boldsymbol{y})^\top \boldsymbol{V}_y^{-1} (\boldsymbol{\eta} - \boldsymbol{y}) = k_p^2, \tag{6.2}$$

$\boldsymbol{y}$ specifying its location, $\boldsymbol{V}_y$ its shape, and $k_p = 2.45$ its size, determined according to the provisions of [4].

The Monte Carlo method in the multivariate case

For the Monte Carlo method in the multivariate case [4], report as for the Monte Carlo method (in the univariate case) in section 6.3.1 except that the measurand and measurement model are to be interpreted in terms of their vector or multivariate counterparts. Moreover, for the coverage region, state whether the region is hyper-ellipsoidal, hyper-rectangular, of smallest volume or otherwise.

EXAMPLE *Simple bivariate measurement model [4, clause 9.2.3] once more*

The bivariate measurand, model and estimates of the input quantities and their associated standard uncertainties are as for the example in section 6.3.2 where X_1 and X_2 are characterized by normal distributions and X_3 by a rectangular distribution so they all have estimates of zero and standard deviations of unity. The output information regarding estimate and uncertainties is the same as that in expressions (6.1).

A 95 % ellipsoidal coverage region for $\boldsymbol{Y}$ is

$$(\boldsymbol{\eta} - \boldsymbol{y})^\top \boldsymbol{V}_y^{-1} (\boldsymbol{\eta} - \boldsymbol{y}) = k_p^2, \qquad k_p = 2.15, \tag{6.3}$$

somewhat different from that in the example in section 6.3.1, which assumes underlying normality.

Bayesian inference in the multivariate case

Similarly to the extension of the reporting in the multivariate case of the Monte Carlo method (see section 6.3.2), the reporting from Bayesian inference in the univariate case (see section 6.3.1) can be extended. Rather than the posterior PDF for the measurand, now the joint posterior of the measurand is provided, and as summary data a vector with the estimates, a covariance matrix, and where relevant a coverage region, for which the same deliberations apply as outlined in the previous section for the Monte Carlo method.

6.3.3 Rounding and number of digits

Absolute uncertainty

The number of digits reported in the measurement results is often dictated by requirements of the application. If not stipulated, in the univariate case it is recommended that two significant digits be reported in a standard uncertainty or an expanded uncertainty, and the corresponding estimate or the endpoints of a coverage interval be terminated at the position of the least significant decimal digit of the reported uncertainty. If a comparison is being made of methods for providing measurement results, additional digits may be reported when considered appropriate.

NOTE If the uncertainty is rounded and stated to two significant decimal digits, the magnitude of the largest possible relative deviation from the unrounded value is 5 %.

The above advice does not apply to the multivariate case where correlation is almost invariably present in the measurement results. See section 6.3.4.

EXAMPLE 1 For a mass measurement, the estimate is 10.004 53 g and the associated standard uncertainty is 0.000 74 g (each to five decimal places).

EXAMPLE 2 For a distance measurement, the estimate is 126.3×10^3 km and the associated standard uncertainty is 2.4×10^3 km (each to one decimal place).

If the measurement result is to be used in a subsequent calculation, as many digits as required for that calculation should be reported. For information passed electronically, the computer-held numbers should be used with no rounding.

The normal rules of rounding according to ISO 80000-1:2009 [104] should be applied, with rounding to the nearest even last digit in cases of ambiguity, unless there are valid technical reasons for doing otherwise.

EXAMPLE *The two stages of calibration*

Consider the two stages of calibration [89, clause 2.39]. The first stage establishes a relation between (stimulus) values provided by measurement standards and corresponding instrument response values. The second stage uses this relation to obtain stimulus values from further instrument response values (inverse evaluation). The relation also allows a stimulus value to be obtained given a further response value (direct evaluation). When the two stages are under the control of a single party, there is little problem in moving from the first stage to the second if all calculations are carried out using a single item of software or results from the first stage are passed electronically to full machine precision to the second stage. If there are departures from this way of working or the stages are under the control of two parties, especially if a calibration certificate contains results to limited numerical precision, there may be issues, for which the reader is directed to section 6.3.4 (particularly the second example).

Relative uncertainty

In many areas of measurement, it is customary to communicate measurement uncertainty in relative terms, such as a percentage or parts-per-million of the reported estimate of the measurand. This practice is widely applied in analytical chemistry and the life sciences, but also in physical calibration (pressure, for example), the use of relative rather than absolute uncertainty is common practice. The rationale for preferring relative uncertainty is that it is a better representation of the measurement uncertainty over a wide(r) interval of values of the measurand. In many calibration and measurement capabilities, as in many instrument specifications, a combination of absolute and relative uncertainty is used to communicate the measurement uncertainty or specification.

Especially when a relative expanded uncertainty has been agreed with the recipient of the measurement result, using an absolute uncertainty with the commonly applied rounding can lead to issues. It has been proposed to carry an extra digit in the absolute uncertainty [105] over what is commonly recommended (see section 6.3.3) to express properly the measurement uncertainty.

6.3.4 Rounding correlation and covariance

Rather than reporting a covariance matrix associated with an estimate of the measurand, we recommend instead, both for human interpretation and analysis, the use of the correlation matrix $\boldsymbol{R}_y$ associated with the vector estimate $\boldsymbol{y}$ together with the vector $\boldsymbol{u}(\boldsymbol{y})$ of standard uncertainties $u(y_i)$. The covariance matrix $\boldsymbol{V}_y$ associated with $\boldsymbol{y}$ is related to $\boldsymbol{R}_y$ by

$$\boldsymbol{V}_y = \boldsymbol{D}_y \boldsymbol{R}_y \boldsymbol{D}_y,$$

where $\boldsymbol{D}_y$ is the diagonal matrix of dimension $m \times m$ with diagonal elements $u(y_1), \ldots, u(y_m)$. Element (i, j) of $\boldsymbol{V}_y$ is the correlation coefficient associated with the estimates y_i and y_j:

$$u(y_i, y_j) = r(y_i, y_j) u(y_i) u(y_j).$$

EXAMPLE *Natural gas analysis*

ISO/TS 28038 [106, clause 9.4.2] is concerned with constructing and using polynomial calibration curves with the polynomial represented in Chebyshev-series form [107]. One of the examples considers natural gas data relating amount fractions and corresponding instrument responses. Polynomial models of several degrees were considered to represent this data. One of these polynomials was of degree 2 (quadratic) for which the covariance matrix associated with the computed Chebyshev coefficients was obtained:

$$\boldsymbol{V}_y = 10^{-6} \times \begin{bmatrix} 0.61 & 0.72 & 0.53 \\ & 1.08 & 0.49 \\ \text{sym.} & & 0.74 \end{bmatrix},$$

which can be compared with the representation as a standard-uncertainty vector and the correlation matrix:

$$\boldsymbol{u}(\boldsymbol{y}) = \begin{bmatrix} 0.0008 \\ 0.0010 \\ 0.0009 \end{bmatrix}, \qquad \boldsymbol{R}_y = \begin{bmatrix} 1 & 0.89 & 0.79 \\ & 1 & 0.54 \\ \text{sym.} & & 1 \end{bmatrix}.$$

The second representation is arguably easier to interpret. For instance, the correlation coefficient (0.89) between the first and second Chebyshev coefficients is appreciable (compared with unity). It is difficult to make this interpretation of the first representation (0.72 compared with the other elements of $1 \times 10^6 \boldsymbol{V}_y$).

In the presence of non-zero covariances or correlation coefficients, reporting needs very careful consideration, especially if some of the input quantities are output quantities from a previous evaluation or are to be used subsequently. Since a covariance corresponding to a correlation coefficient having magnitude close to unity might cause numerical difficulties in subsequent evaluations, such as related to least squares' applications, the number of digits to be held should depend on that magnitude. Considering quantities X_1 and X_2, unless required otherwise for particular technical reasons, it is recommended that the correlation coefficient r associated with X_1 and X_2 is reported such that $1 - |r|$ has at least two significant decimal digits. The standard uncertainties $u(x_1)$ and $u(x_2)$ and the covariance $u(x_1, x_2)$ should be reported to the same number of decimal places. In any cases of doubt, all computer-held digits should be reported.

NOTE The above paragraph includes a rule of thumb that is not foolproof. The number of digits to be reported for a correlation coefficient or covariance depends on the application that uses such information. A full analysis of the number of digits to be reported needs details of the application and may involve advanced numerical computations such as involving the eigenvalues of the correlation matrix. The following information is based on that given in [4, clause 3.2.1]:

> If the correlation matrix is close to being singular, additional decimal digits need to be retained in order to avoid numerical difficulties when using the correlation matrix as input to an uncertainty evaluation. The number of decimal digits to be retained depends on the nature of the subsequent calculation, but as a guide can be taken as the number of decimal digits needed to represent the smallest eigenvalue of the correlation matrix with two significant decimal digits. For a correlation matrix of dimension 2×2, the eigenvalues are $1 \pm |r|$, the smaller being $1 - |r|$, where r is the off-diagonal element of the matrix. If a correlation matrix is known to be singular prior to rounding, rounding towards zero reduces the risk that the rounded matrix is not positive semi-definite.

EXAMPLE *Highly correlated quantities*

For a particular evaluation problem, the quantities X_1 and X_2, corresponding to output quantities in a previous evaluation, are very highly correlated. To seven significant decimal digits, the standard uncertainties associated with their best estimates are $u(x_1) = 0.152\,748\,2$, $u(x_2) = 0.603\,536\,4$ and the associated covariance is $u(x_1, x_2) = -0.303\,407\,2$, all in appropriate units. This covariance corresponds to a correlation coefficient r of $-0.999\,277\,4$. The value of $1 - |r|$ is $0.000\,722\,6$, which, when rounded to two significant decimal digits, is $0.000\,72$. Accordingly, these results should be reported as, $u(x_1) = 0.152\,75$, $u(x_2) = 0.603\,54$, $u(x_1, x_2) = -0.303\,41$ and $r = -0.999\,28$.

A covariance should be reported to the same number of significant digits as used when reporting a correlation coefficient.

Regarding compatibility with standards, ISO/IEC 17025 [7, clause 7.8.4.1] states

> '... calibration certificates shall include ... the measurement uncertainty of the measurement result presented in the same unit as that of the measurand or in a term relative to the measurand (e.g. percent) ...'

Thus, to conform with ISO/IEC 17025 the measurement uncertainty must be reported. The important aspect of the quoted clause is the word 'include': the possibility of providing further information (such as a probability density function (PDF)) is not excluded.

6.4 Use of LaTeX and Microsoft Word for reporting uncertainty

6.4.1 General

A value of a physical quantity and its associated uncertainty should each be expressed as the product of a number and a unit. The SI brochure [108] states that there should be a space between the number part and the unit part. This guidance still leaves several options for typesetting. Here we recommend the use of a thin space (en-space – the width of a lower case 'n') for this purpose; thus '10 m' rather than '10 m'. We recommend ways of reporting estimates of quantities and associated uncertainties using the systems LaTeX and Microsoft Word.

6.4.2 LaTeX

LaTeX package `siunitx` is invaluable for typesetting quantities, for which the manual is regularly updated:
`http://anorien.csc.warwick.ac.uk/mirrors/CTAN/macros/latex/contrib/siunitx/siunitx.pdf`.

For reporting measured values and their units, `siunitx` inserts a thin space between a number and the unit:

`\SI{997}{\kg\per\m^3}`	997 kg/m^3
`\SI{997}{\kg\per\cubic\m}`	997 kg m^{-3}
`\SI{95}{\percent}`	95 %

Package `siunitx` has many other valuable facilities for working with physical quantities, such as an excellent capability for working with tabular material such as uncertainty budgets.

6.4.3 Microsoft Word

There does not appear to be a Microsoft Word template that provides facilities that are comparable to those of `siunitx`. However, to create 95 % (with a thin space), for example, type in Microsoft Word:

`95>space>2009>ALT+X>%`

Then delete the left (the wider) of the two space characters.

Alternatively, the pre-defined special characters for an en and em space can be assigned a shortcut key in the 'Symbols' dialog box to make these better accessible when typing. Finally, holding down the Alt key as 8194 (for an en space) or 8195 (for an em space) is typed produces the special white spaces.

Part E1

Calibration, measurement and testing

Example E1.1

Two-point and multipoint calibration

M.G. Cox, J. Greenwood, A. Bošnjaković, V. Karahodžić

E1.1.1 Summary

A generic treatment of two-point and multi-point interpolation of calibration data is given with uncertainties associated with the data propagated using the law of propagation of uncertainty and its generalization to vector measurands. The approach is applied to the measurement of hydrogen ion activity (pH). Such measurement is one of the most common in chemistry, although correlations associated with the input quantities in the measurement model are rarely taken into account. The treatment given follows common practice, which tends to give an optimistically small evaluation of the uncertainty associated with an estimated pH value. A way of taking correlation into account in one typical instance is given but its implementation is problematical because of the difficulty in quantifying the correlation.

E1.1.2 Introduction of the application

E1.1.2.1 General

A generic treatment of two-point and multi-point interpolation of calibration data is first given. We stay consistent with the VIM's concept of calibration [89, definition 2.39] as constituting two stages. Here the first stage involves fitting to measured data a function that describes the relationship of a response (dependent) variable y to a stimulus (independent) variable x. The second stage involves using this relationship to determine the value of one variable given a value of the other. Uncertainties in both the stimulus and response variables are handled in both stages and propagated using the law of propagation of uncertainty (LPU) in JCGM 100:2008 (GUM) [2] and its generalization to vector measurands in GUM Supplement 2 (GUM-S2) [4].

Two scenarios are considered. One, a single party accesses the calibration data set and provides the required interpolated value. In doing so, the party may or may not determine the calibration parameters explicitly. Two, one party has access to the calibration data set, delivering the calibration parameters to a second party, which in turn provides the interpolated value.

Although the measurement models involved are simple, they are used to illustrate a number of aspects that can be carried over to examples in other areas.

The approach is applied to the measurement of hydrogen ion activity (pH) in which up to three two-point interpolations are required and uncertainties are tracked through the calculation.

E1.1.2.2 Specific: pH of a test solution using two-point calibration

pH, the negative logarithm to base 10 of the activity of hydrogen ion in a solution is probably the most measured quantity in chemistry [109]. The electric potential of a suitable cell, for example, a glass electrode and reference electrode, is proportional to pH and forms the basis of pH measurement.

In 2002 IUPAC, the International Union for Pure and Applied Chemistry, issued a recommendation for revision of the pH scale based on the concept of a primary reference measurement procedure for pH [74]. The use of an electrochemical (Harned) cell fulfils the criteria for a primary reference measurement procedure so that a pH value thus obtained is traceable to the International System of Units, here the SI measurement unit 1 (one). A solution, the pH of which is measured by such a cell at the highest metrological level, may be classified as a primary measurement standard and can be used to assign pH values to other solutions. These solutions are sold as certified reference materials to calibrate pH meters for routine use.

There are several approaches to pH measurement involving the use of 1-point, 2-point and multipoint calibration, least-squares regression, and with or without temperature correction. Here we use the 2-point calibration approach, with and without temperature correction.

The methods in the generic parts of this document apply (a) when the temperature of the test solution matches that of the standard (reference) solutions and (b) when this is not the case.

E1.1.3 Specification of the measurand(s)

In this specific example, the measurand is the pH of a solution being calibrated. More generally, the measurand is the interpolated independent or dependent variable obtained from a relationship between those variables derived from data representing values of the variables. Intermediate measurands, when required, are the parameters describing the relationship.

E1.1.4 Measurement model

E1.1.4.1 General

There are two stages involved in calibration [89]: (i) determine a calibration curve from calibration data and (ii) use that calibration curve. Because of the relative simplicity of two-point and multi-point interpolation as considered here, it may be preferable when circumstances permit to combine the stages into a single-stage model. Such a model avoids having to deal with intermediate correlation associated with the calibration curve parameters that are estimated in the first stage and used in the second. Operating in two stages corresponds to the use of a multi-stage model [5, clause 8.4] and is necessary when the construction and use of the calibration model are carried out by different parties.

Generic approach to two-point calibration

Two calibration points (x_1, y_1) and (x_2, y_2) are given that bracket x_0, an x-value for which y_0, the corresponding y-value, is required under the assumption that the y-value lies on the straight line joining the calibration points (see figure E1.1.1).

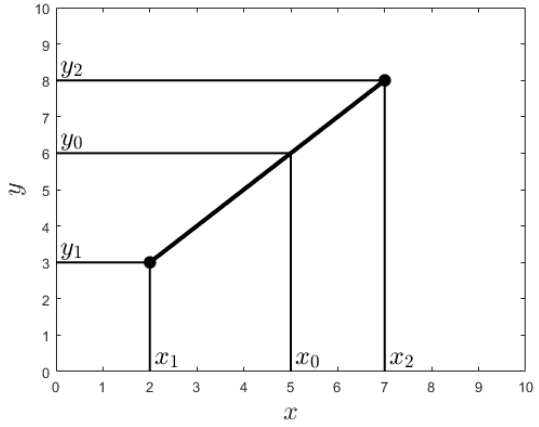


Figure E1.1.1: Two-point calibration

By similar triangles, with $\delta x = x_2 - x_1$ and $\delta y = y_2 - y_1$,

$$\frac{y - y_1}{x - x_1} = \frac{y_2 - y_1}{x_2 - x_1} = \frac{\delta y}{\delta x}, \tag{E1.1.1}$$

A common representation of a straight-line calibration function, which is used here, is

$$y = a + bx, \tag{E1.1.2}$$

where a is the intercept on the y-axis and b is the gradient [77].

NOTE The form (E1.1.2) is used in the straight-line calibration standard ISO/TS 28037 [77] and will be familiar to many end-users.

The process

The process defining the measurement model has one or two stages (section E1.1.2.1).

Single-stage model. A single party has access to (x_1, y_1) and (x_2, y_2), and also x_0, and provides y_0, the y-value on the line corresponding to x_0 (figure E1.1.1). In doing so, the party may or may not determine a and b explicitly. The measurement model is specified by the description of the provision of y_0.

Two-stage model. One party has access to (x_1, y_1) and (x_2, y_2), and provides a and b (intermediate measurands) to the second party using the straightforwardly verified

$$b = \frac{y_2 - y_1}{x_2 - x_1}. \tag{E1.1.3}$$

$$a = y_1 - bx_1. \tag{E1.1.4}$$

The second party (possibly identical to the first party) has access to a and b, and also x_0, and provides y_0 using the expression

$$y_0 = a + bx_0. \tag{E1.1.5}$$

The measurement model is again described by the process to provide y_0.

NOTE When x_1 and x_2 are far from the origin, that is, $|x_2 - x_1| \ll |x_1|$, an alternative form may be numerically more stable. One such form is given by working with a transformed x-variable

$$\widetilde{x} = x - x_1.$$

Using (E1.1.2) and (E1.1.4) the calibration function can be expressed as

$$y = y_1 + b\widetilde{x}, \tag{E1.1.6}$$

which is evaluated at the value x_0 of the independent variable. The resulting expression

$$y_0 = y_1 + b(x_0 - x_1) = y_1 + \frac{y_2 - y_1}{x_2 - x_1}(x_0 - x_1), \tag{E1.1.7}$$

and formula (E1.1.3) constitute the measurement model with y_0 as the measurand. It is accepted that such a transformation is not always appropriate.

The form of interpolation considered here is forward interpolation. Inverse interpolation, when the stimulus value x_0 corresponding to a response value y_0 is required, can also be carried out (for treatments see [77, 110]) but is not required here. The roles of x and y can be interchanged when permitted by the context.

Generic approach to multi-point calibration

Multi-point calibration is the treatment in sections above extended to an arbitrary number of points. In these sections, a straight-line segment joining two of the calibration points serves as the calibration function. When there are m calibration points ($m \geq 2$), with strictly increasing stimulus values, the points are joined pairwise by successive straight-line segments, the overall construction being a piecewise-linear function or first-degree spline [111], acting as the calibration function. For each interval between pairs of successive points, the treatment of sections above can be applied directly to the appropriate segment of the piecewise-linear function.

NOTE When $m = 2$ the calibration function is a single straight-line segment so it is naturally monotonic, a necessary condition. For $m > 2$, the ordered points may not form a monotonic sequence, a situation not considered here [110].

Alternatively, straight-line fitting by least squares can be used taking reported uncertainties associated with the calibration data into consideration [77]. Polynomial interpolation or polynomial fitting can also be used [110].

Metrological extension

The measurement model implied by two-point calibration is the algorithm to provide y_0 given x_0. The data will generally have associated uncertainties arising from a Type A evaluation of uncertainty [2, clause 4.2] especially following an analysis of repeated observations. Often there will also be uncertainties obtained from a Type B evaluation and associated covariances arising from common measurement effects [2, clause 4.3]. Such covariances should also be handled to avoid producing invalid statements of uncertainty associated with predicted y-values.

The calibration data considered here are assumed to have independent errors.

E1.1.4.2 pH measured at a specific temperature

An approach to providing the pH of a test solution [112] is a correction approach in which $\mathrm{pH_X}$, the pH of a test solution X, is given by using a cell twice to measure potential E_X in X and potential E_S in a standard solution S:

$$\mathrm{pH_X} = \mathrm{pH_S} + \frac{E_\mathrm{X} - E_\mathrm{S}}{k}. \tag{E1.1.8}$$

In expression (E1.1.8), $\mathrm{pH_S}$ is the pH of S, and

$$k = \frac{RT \ln 10}{F},$$

where R is the gas constant, T the temperature in K and F the Faraday constant.

Two other approaches are bracketing methods, which are generally more accurate and used here. Use is made of the reference material certificates for the standard solutions, which give pH values and associated standard uncertainties at stipulated temperatures.

Measurement is made at temperature T_X, one of these stipulated temperatures, and a bracketing procedure adopted [113]. The potential E_X of the test solution X is measured. Likewise, the potentials E_{S_1} and E_{S_2} are measured of two cells with standard solutions S_1 and S_2 such that the E_{S_1}- and E_{S_2}-values bracket E_X and are as near as possible to it. The pH of S_1 and S_2 at temperature T_X, namely, $\mathrm{pH}_{S_1}^{T_\mathrm{X}}$, $\mathrm{pH}_{S_2}^{T_\mathrm{X}}$, are given on certificates such as issued by NIST [114, 115].

By assuming linearity between pH and E, that is, linear interpolation is valid between the points $(E_{\mathrm{S}_1}, \mathrm{pH}_{S_1}^{T_\mathrm{X}})$ and $(E_{\mathrm{S}_2}, \mathrm{pH}_{S_2}^{T_\mathrm{X}})$, the pH value $\mathrm{pH_X}$ corresponding to potential E_X is obtained.

The output quantity, the measurand, generically y_0, is pH_X, the pH of the test solution.

The input quantities in the measurement model are E_{S_1}, E_{S_2}, $\mathrm{pH}_{S_1}^{T_\mathrm{X}}$, $\mathrm{pH}_{S_2}^{T_\mathrm{X}}$ and E_X, corresponding respectively to x_1, x_2, y_1, y_2 and x_0 in the generic approach.

In an extended model [5, clause 9], account is taken of further influences. In this case main effects are pH instrument calibration, instrument resolution and interpolated pH. The uncertainties associated with the first two effects are provided by the instrument manual and inspection of the output display. Incorporating correction terms to account for these effects,

$$\begin{aligned} E_\mathrm{X,corr} &= E_\mathrm{X} + \delta E_\mathrm{resX} + \delta E_\mathrm{cal}, \\ E_\mathrm{S1,corr} &= E_\mathrm{S1} + \delta E_\mathrm{resS1} + \delta E_\mathrm{cal}, \\ E_\mathrm{S2,corr} &= E_\mathrm{S2} + \delta E_\mathrm{resS2} + \delta E_\mathrm{cal}. \end{aligned} \tag{E1.1.9}$$

Interpolated pH and correction quantities in expressions (E1.1.9) relating to cell potential are assumed independent.

Section E1.1.7.1 contains a discussion of the validation of results.

E1.1.4.3 pH measurement accounting for temperature

The temperature T_X of the test solution is measured and the certificate of one of the standard solutions is used to identify the closest bracketing temperatures T_1 and T_2. Potential measurement gives E_{S_1}, E_{S_2}, E_X as before.

The pH values $\mathrm{pH}_{S_1}^{T_1}$ and $\mathrm{pH}_{S_2}^{T_1}$ for standard solutions S_1 and S_2 at temperature T_1 are obtained from the certificate. Linear interpolation is used between the points $(E_{\mathrm{S}_1}, \mathrm{pH}_{S_1}^{T_1})$ and $(E_{\mathrm{S}_2}, \mathrm{pH}_{S_2}^{T_1})$ to give the pH value $\mathrm{pH}_{\mathrm{X}}^{T_1}$ at temperature T_1 corresponding to potential E_{X}.

This step is repeated for temperature T_2. The certificate is used to give the pH values $\mathrm{pH}_{S_1}^{T_2}$ and $\mathrm{pH}_{S_2}^{T_2}$ for standard solutions S_1 and S_2 at temperature T_2. Linear interpolation is used between the points $(E_{\mathrm{S}_1}, \mathrm{pH}_{S_1}^{T_2})$ and $(E_{\mathrm{S}_2}, \mathrm{pH}_{S_2}^{T_2})$ to give the pH value $\mathrm{pH}_{\mathrm{X}}^{T_2}$ at temperature T_2 corresponding to potential E_{X}.

Finally, linear interpolation is applied to the points $(E_{\mathrm{S}_1}, \mathrm{pH}_{\mathrm{X}}^{T_1})$ and $(E_{\mathrm{S}_2}, \mathrm{pH}_{\mathrm{X}}^{T_2})$ to give the pH value $\mathrm{pH}_{\mathrm{X}}^{T_{\mathrm{X}}}$ at temperature T_{X} corresponding to potential E_{X}.

The generic treatment in section E1.1.4.1 is thus applied three times to implement these three stages of two-point interpolation.

The measurement model is given by the above algorithmic description where the input quantities in the model are E_{X}, E_{S_1}, E_{S_2}, T_{X}, $\mathrm{pH}_{S_1}^{T_1}$, $\mathrm{pH}_{S_1}^{T_2}$, $\mathrm{pH}_{S_2}^{T_1}$, $\mathrm{pH}_{S_2}^{T_2}$.

E1.1.5 Uncertainty propagation

E1.1.5.1 Assumption

Uncertainty propagation in this section is based on the assumption that the input quantities — the measured potentials and the pH values for the standard solutions — are independent. This assumption is consistent with IUPAC recommendations for pH measurement [74]. There procedures are given for accounting for input standard uncertainties based on the variant of the law of propagation of uncertainty in [2, section 5.1]. That variant does not account for correlations among the input quantities.

In practice, input quantities are likely to be correlated and account should be taken of that fact. See the important discussion in section E1.1.7.2.

E1.1.5.2 General two-stage model

The notation established here for the two-stage model is also used in the single-stage model.

First stage. The inputs are the calibration data x_1, y_1, x_2 and y_2 and their associated standard uncertainties. The outputs are the calibration parameters a and b and their associated covariance matrix $\boldsymbol{V}_{[a,b]}$. The model is bivariate (two output quantities):

$$b = \frac{y_2 - y_1}{x_2 - x_1} = \frac{\delta y}{\delta x}, \qquad a = y_1 - b x_1. \tag{E1.1.10}$$

For the uncertainty propagation, [4, formula (3)] is applied to obtain the output covariance matrix

$$\boldsymbol{V}_{[a,b]} = \begin{bmatrix} u^2(a) & u(a,b) \\ u(a,b) & u^2(b) \end{bmatrix} = \boldsymbol{C}^{(1)} \boldsymbol{V}_{\mathrm{in}} \boldsymbol{C}^{(1)\top}. \tag{E1.1.11}$$

In formula (E1.1.11), $u(a,b)$ denotes the covariance between a and b, $\boldsymbol{V}_{\text{in}}$ denotes the input covariance matrix, the 4×4 diagonal matrix with diagonal entries $u^2(x_1)$, $u^2(y_1)$, $u^2(x_2)$ and $u^2(y_2)$, and

$$\boldsymbol{C}^{(1)} = \frac{1}{\delta x}\begin{bmatrix} -bx_2 & x_2 & bx_1 & -x_1 \\ b & -1 & -b & 1 \end{bmatrix} \tag{E1.1.12}$$

is the sensitivity matrix containing the first partial derivatives of x_1, y_1, x_2 and y_2 with respect to a and b.

Second stage. The inputs are the outputs from the first stage together with x_0 and $u(x_0)$.

The model is

$$y_0 = a + bx_0.$$

Applying uncertainty propagation [4, formula (3)] once more, the output standard uncertainty $u(y_0)$ is given by

$$u^2(y_0) = \boldsymbol{C}^{(2)}\boldsymbol{V}_{[a,b]}\boldsymbol{C}^{(2)\top} + b^2u^2(x_0), \tag{E1.1.13}$$

where $\boldsymbol{C}^{(2)}$ is the 1×2 sensitivity matrix

$$\boldsymbol{C}^{(2)} = \begin{bmatrix} 1 & x_0 \end{bmatrix}. \tag{E1.1.14}$$

E1.1.5.3 General single-stage model

By combining the two stages above, the substitution of formula (E1.1.11) into expression (E1.1.13) yields

$$u^2(y_0) = \boldsymbol{C}^{(2)}\boldsymbol{C}^{(1)}\boldsymbol{V}_{\text{in}}\boldsymbol{C}^{(1)\top}\boldsymbol{C}^{(2)\top} + b^2u^2(x_0).$$

Setting

$$q = \frac{x_0 - x_1}{\delta x}, \tag{E1.1.15}$$

the use of expressions (E1.1.12) and (E1.1.14) gives

$$\begin{aligned}\boldsymbol{C}^{(2)}\boldsymbol{C}^{(1)} &= \frac{1}{\delta x}\begin{bmatrix} -b(x_2-x_0) & x_2-x_0 & -b(x_0-x_1) & x_0-x_1 \end{bmatrix} \\ &= \begin{bmatrix} -b(1-q) & 1-q & -bq & q \end{bmatrix}.\end{aligned}$$

Hence, using

$$\boldsymbol{V}_{\text{in}} = \begin{bmatrix} u^2(x_1) & & & \\ & u^2(y_1) & & \\ & & u^2(x_2) & \\ & & & u^2(y_2) \end{bmatrix},$$

$$\boldsymbol{C}^{(2)}\boldsymbol{C}^{(1)}\boldsymbol{V}_{\text{in}} = \begin{bmatrix} -b(1-q)u^2(x_1) & (1-q)u^2(y_1) & -bqu^2(x_2) & qu^2(y_2) \end{bmatrix},$$

and so

$$\boldsymbol{C}^{(2)}\boldsymbol{C}^{(1)}\boldsymbol{V}_{\text{in}}\boldsymbol{C}^{(1)\top} = \frac{1}{\delta x}\begin{bmatrix} -b(1-q)u^2(x_1)(-bx_2) + (1-q)u^2(y_1)x_2 - bqu^2(x_2)bx_1 - qu^2(y_2)x_1 \\ -b^2(1-q)u^2(x_1) - (1-q)u^2(y_1) + b^2qu^2(x_2) + qu^2(y_2) \end{bmatrix}.$$

Therefore,

$$\boldsymbol{C}^{(2)}\boldsymbol{C}^{(1)}\boldsymbol{V}_{\text{in}}\boldsymbol{C}^{(1)\top}\boldsymbol{C}^{(2)\top} = \frac{1}{\delta x}[b^2(1-q)u^2(x_1)x_2+(1-q)u^2(y_1)x_2-b^2qu^2(x_2)x_1-qu^2(y_2)x_1$$
$$-b^2x_0(1-q)u^2(x_1)-x_0(1-q)u^2(y_1)+b^2qx_0u^2(x_2)+qx_0u^2(y_2)]$$
$$=\frac{1}{\delta x}\big[b^2(1-q)u^2(x_1)(x_2-x_0)+(1-q)u^2(y_1)(x_2-x_0)+b^2qu^2(x_2)(x_0-x_1)+qu^2(y_2)(x_0-x_1)\big].$$

Thus expression (E1.1.13) can be written as

$$u^2(y_0) = \boldsymbol{C}^{(2)}\boldsymbol{C}^{(1)}\boldsymbol{V}_{\text{in}}\boldsymbol{C}^{(1)\top}\boldsymbol{C}^{(2)\top} + b^2u^2(x_0)$$
$$= b^2(1-q)^2u^2(x_1)+(1-q)^2u^2(y_1)+b^2q^2u^2(x_2)+q^2u^2(y_2)+b^2u^2(x_0),$$

that is,

$$u^2(y_0) = b^2(1-q)^2u^2(x_1)+(1-q)^2u^2(y_1)+b^2q^2u^2(x_2)+q^2u^2(y_2)+b^2u^2(x_0). \quad \text{(E1.1.16)}$$

The result (E1.1.16) can also be confirmed from first principles.

E1.1.5.4 pH estimation at a specific temperature and associated uncertainty evaluation

Values of potential in the test and standard solutions S_1 and S_2 were

$$E_{\text{X}} = -1.875\,\text{mV}, \qquad E_{\text{S1}} = 6.15\,\text{mV}, \qquad E_{\text{S2}} = -26.35\,\text{mV},$$

each of which was the average of 4 repeated observations. pH values at 25 °C for S_1 and S_2 from [114] are

$$\text{pH}_{\text{S1}}^{25\,^\circ\text{C}} = 6.8640, \qquad \text{pH}_{\text{S2}}^{25\,^\circ\text{C}} = 7.4157.$$

From formulæ (E1.1.3) and (E1.1.7) the resulting estimate of $\text{pH}_{\text{X}} = 7.0002$.

Associated standard uncertainties were

$$u(E_{\text{X}}) = 0.0250\,\text{mV}, \qquad u(E_{\text{S1}}) = 0.0289\,\text{mV}, \qquad u(E_{\text{S2}}) = 0.0289\,\text{mV},$$
$$u(\text{pH}_{\text{S1}}^{25\,^\circ\text{C}}) = 0.0051, \qquad u(\text{pH}_{\text{S2}}^{25\,^\circ\text{C}}) = 0.0051.$$

The above standard uncertainties associated with standard pH solutions are given in [114].

The propagation of uncertainty carried out in accordance with expressions (E1.1.15) to (E1.1.16) yields $u(\text{pH}_{\text{X}}) = 0.0041$.

These results relate to the basic measurement model for pH. The extended model would work with the corrected quantities in (E1.1.9) rather than the uncorrected quantities. The estimates of all correction terms in the extended model are taken as zero. δE_{cal} appears in three of expressions (E1.1.9), so seemingly inducing correlation. However, this quantity is eliminated when the corrected quantities are used rather than the original. This effect can be seen mathematically by substituting $E_{\text{X,corr}}$, $E_{\text{S1,corr}}$ and $E_{\text{S2,corr}}$ as the 'new' x_0, x_1 and x_2, respectively, from expressions (E1.1.9) into expressions (E1.1.3) and (E1.1.7).

The instrument display gave results in volts with 3 significant decimal places. Assume a rounding error in the last digit, that is, in the interval ±0.0005 V. Characterizing resolution by a rectangular distribution over this interval, the consequent resolution standard uncertainty E_{res} applying to all potential readings is $0.0005\,\text{V}/\sqrt{3} = 0.000\,29\,\text{V}$. This standard uncertainty is some one hundredth of the above potential standard uncertainties and so is negligible.

The extended model would deliver the same estimate and standard uncertainty as the basic model to the number of digits reported.

E1.1.5.5 pH estimation accounting for temperature and associated uncertainty evaluation

The measured temperature of the test solution was $T_X = 23.7\,°C$. The pH values of the measurement standards are given at a temperature interval of 5 °C. The bracketing pair $T_1 = 25\,°C$ and $T_2 = 20\,°C$ is therefore appropriate.

Potential measurement gives

$$E_X = -1.875\,mV, \qquad E_{S1} = 6.15\,mV, \qquad E_{S2} = -26.35\,mV,$$

each of which was the average of 4 repeated observations. pH values at 25 °C and 20 °C for S_1 and S_2 from [114] are

$$pH_{S1}^{25\,°C} = 6.8640, \qquad pH_{S2}^{25\,°C} = 7.4157, \qquad pH_{S1}^{20\,°C} = 6.8796, \qquad pH_{S2}^{20\,°C} = 7.4323.$$

Linear interpolation between the points

$$(E_{S_1}, pH_{S_1}^{T_1}) \equiv (6.15\,mV, 6.8640) \quad \text{and} \quad (E_{S_2}, pH_{S_1}^{T_1}) \equiv (-26.35\,mV, 7.4157)$$

gives the pH value $pH_X^{T_1} = 6.8681$ at temperature T_1 of 23.7 °C corresponding to potential $E_X = -1.875\,mV$.

Likewise, linear interpolation between

$$(E_{S_1}, pH_{S_1}^{T_2}) \equiv (6.15\,mV, 6.8640) \quad \text{and} \quad (E_{S_2}, pH_{S_2}^{T_2}) \equiv (-26.35\,mV, 7.4323)$$

gives the pH value $pH_X^{T_2} = 7.4200$ at temperature T_2 of 23.7 °C corresponding to potential $E_X = -26.35\,mV$.

Finally, linear interpolation between the points $(E_{S_1}, pH_X^{T_1})$ and $(E_{S_2}, pH_X^{T_2})$ is used to give the pH value $pH_X^{T_2}$ at temperature $pH_X^{T_X}$ corresponding to potential E_X.

Associated standard uncertainties were

$$u(E_X) = 0.0250\,mV, \qquad u(E_{S1}) = 0.0289\,mV, \qquad u(E_{S2}) = 0.0289\,mV,$$

$$u(pH_{S1}^{T_1}) = 0.0051, \qquad u(pH_{S2}^{T_1}) = 0.0051, \qquad u(pH_{S1}^{T_2}) = 0.0051, \qquad u(pH_{S2}^{T_2}) = 0.0051.$$

The above standard uncertainties associated with standard pH solutions are given in [114].

The application of the method of section E1.1.4.3 gives $pH_X = 7.0109$ and $u(pH_X) = 0.0041$.

E1.1.6 Reporting the result

The estimate y_0 of the measurand and the associated standard uncertainty $u(y_0)$ are directly reported in the conventional manner according to the GUM [2].

E1.1.7 Interpretation of results

E1.1.7.1 Validation of results

A check on the accuracy of linear interpolation was made for the example in sub-section E1.1.4.3. NIST certificate [114] gives pH values for the standard solutions considered at temperature values from 5 °C to 50 °C in steps of 5 °C. Cubic interpolation based in [110] was carried out using temperature values 15 °C, 20 °C, 25 °C and 30 °C (two values on either side of 23.7 °C), and the corresponding pH values for S_1 given in the certificate. The interpolated value at 23.7 °C was 6.8677 compared with 6.8681 from linear interpolation. The magnitude of the difference between these values is almost a factor of ten smaller than the standard uncertainty associated with the obtained pH value. A comparable result was obtained for S_2 and for the other linear interpolations carried out. Thus, linear interpolation is adequate in this example.

In a study by Damasco et al. [116] it was reported that a Monte Carlo method applied to primary pH measurement gave similar results to the 'GUM approach' [2]. The work of Wiora and Wiora [117] came to the same conclusion. As a simple trial, the Monte Carlo method of GUM Supplement 1 [3] was applied to the example in section E1.1.5.4. The input quantities were modelled by normal distributions with means equal to the input estimates and standard deviations equal to the associated standard uncertainties. For 10^6 Monte Carlo trials, exactly the same result was delivered as in that section to the number of decimal places stated.

The standard uncertainties associated with the interpolated value pH_X are scarcely influenced by the uncertainties associated with the pH values of the standard solutions. As an instance, if the latter standard uncertainties are replaced by zero in the example in sub-section E1.1.5.4, $u(pH_X)$ becomes 0.0040 (originally 0.0041), implying that the further repeated observations of the three potentials would do much to reduce $u(pH_X)$, assuming the repeated observations are genuinely independent.

E1.1.7.2 Correlation issues

It must be emphasized that the treatment given regards all input quantities as independent. Independence is a common assumption in general in pH uncertainty evaluation. This assumption is often made implicitly (see [118–121], for instance), but has adverse consequences in that evaluated pH uncertainties can be optimistically small. To obtain more valid results covariance effects need to be quantified and incorporated.

Laboratories that follow IUPAC recommendations [74] will not take correlation into consideration and so might be reporting optimistically small measurement uncertainties. There seems to be little relevant literature available on pH measurement on obtaining correlations associated with input quantities. If such correlations were available, they could be accounted for by applying the provisions in the GUM [2, section 5.2].

It is noted in passing that correlation issues are discussed in [122], but they relate to correlations induced by the choice of parametrization rather than being associated with input quantities.

In terms of pH certificates used here, it would appear that covariance between pH at different temperatures, and probably between pH values for different materials, can be deduced. The certificates give two uncertainties, one (as in sections E1.1.5.4 and E1.1.5.5) for an SI-traceable value and one, much smaller, that omits the uncertainty associated with the Bates-Guggenheim convention [75]. The Bates-Guggenheim conventional uncertainty, as given on the certificates, could hence reasonably be taken as at least an approximate covariance when using the SI-traceable

values. Doing so gives the correlation very close to unity, and indeed that (or any respectably high value) could be suggested as a generally conservative treatment (going somewhat against GUM conventional wisdom, which advises realism).

To indicate the effect of ignoring correlation, we carry out an exercise in which full correlation is present between the pH values for the standards in section E1.1.5.4. Assume that all quantities are independent apart from these two pH values, which are accorded a correlation of unity. For this case of perfect correlation [123–125], the standard uncertainties associated with these quantities must be identical, which indeed they are, being equal to 0.0051. In terms of the generic notation of section E1.1.5.2, the input covariance matrix $\boldsymbol{V}_{\text{in}}$ is no longer diagonal but has covariance $u^2(y_1) = u^2(y_2)$ in off-diagonal positions $(2,4)$ and $(4,2)$:

$$\boldsymbol{V}_{\text{in}} = \begin{bmatrix} u^2(x_1) & & & \\ & u^2(y_1) & & u^2(y_1) \\ & & u^2(x_2) & \\ & u^2(y_1) & & u^2(y_1) \end{bmatrix}. \tag{E1.1.17}$$

Noting that $u^2(y_1) = u^2(y_2) = u(y_1, y_2)$ in the fully correlated case, where $u(y_1, y_2)$ is the covariance associated with y_1 and y_2, by applying a similar treatment to that in section E1.1.5.3 but using the covariance matrix (E1.1.17) gives

$$u^2(y_0) = b^2(1-q)^2u^2(x_1)+(1-q)^2u^2(y_1)+2q(1-q)u(y_1,y_2)+b^2q^2u^2(x_2)+q^2u^2(y_2)+b^2u^2(x_0). \tag{E1.1.18}$$

The only difference is that the terms

$$(1-q)^2u^2(y_1) + q^2u^2(y_2) = [(1-q)^2 + q^2]u^2(y_1) \tag{E1.1.19}$$

in the uncorrelated treatment [expression (E1.1.16)] are replaced by

$$(1-q)^2u^2(y_1) + 2q(1-q)u(y_1,y_2) + q^2u^2(y_2) = [(1-q)^2 + 2q(1-q) + q^2]u^2(y_1) \tag{E1.1.20}$$

in the fully correlated case [expression (E1.1.18)]. Since expression (E1.1.20) simplifies (exactly) to $u^2(y_1)$, expression (E1.1.18) becomes

$$u^2(y_0) = b^2(1-q)^2u^2(x_1) + u^2(y_1) + b^2q^2u^2(x_2) + b^2u^2(x_0) \tag{E1.1.21}$$

in the fully correlated case.

In the application of expression (E1.1.21) to the data in section E1.1.5.4, the standard uncertainty associated with $\text{pH}_\text{X} = 7.0109$ becomes $u(\text{pH}_\text{X}) = 0.0051$ (compared with 0.0041 when correlation is disregarded). Unsurprisingly, this value of $u(\text{pH}_\text{X})$ is the same as the (identical) values of the 'input' standard uncertainties $u(\text{pH}_{\text{S1}}^{25\,^\circ\text{C}})$ and $u(\text{pH}_{\text{S2}}^{25\,^\circ\text{C}})$. Thus there is no reduction in uncertainty in the fully correlated case.

It is observed that the standard uncertainties $u(\text{pH}_{\text{S1}}^{25\,^\circ\text{C}})$ and $u(\text{pH}_{\text{S2}}^{25\,^\circ\text{C}})$ make comparatively large contributions compared with those for the measured potential values. The five standard uncertainty contributions [the square roots of the successive terms on the right side of expression (E1.1.16)] in the case where correlation is ignored are

$$-0.0004, \qquad 0.0001, \qquad 0.0038, \qquad 0.0013, \qquad -0.0004$$

to four decimal places. The values are to be compared with the corresponding four values from expression (E1.1.21), namely,

$$-0.0004, \qquad 0.0001, \qquad 0.0051, \qquad -0.0004,$$

in the correlated case, confirming that perfectly correlated standard uncertainty components are combined additively (see [2, clause 5.2.2, note 1]):

$$0.0038 + 0.0013 = 0.0051.$$

The situation is compounded in section E1.1.5.5 where four (rather than two) pH values and two temperature values are involved.

Assuming independence of the values of the pH standards in either case is not a valid assumption. A treatment such as given in [123, section 4.1] is suggested, that is, to work with a common correlation coefficient ρ associated with the input pH values. The basic change would be that the off-diagonal terms of $\boldsymbol{V}_{\text{in}}$ in formula (E1.1.17) would become $\rho u^2(y_1)$. (The case $\rho = 0$ yields the uncorrelated case and $\rho = 1$ the case of perfect correlation.) A value for ρ might be obtained on technical grounds by examining uncertainty budgets (to see the relative contribution from the Bates-Guggenheim convention, for instance) or some other means such as employing expert judgment.

A further, chemical, issue is that the NIST standards are not solutions; they are solids that have to be weighed, mixed and dissolved fully in high purity water. Buffer solutions are not particularly sensitive to minor dilution problems, but the preparation just mentioned will add further variation. Atmospheric CO_2, for example, can shift measured pH values, especially in a neutral pH test sample (pH = 7), unless that is also buffered or air excluded. Although a measurement laboratory would exercise care in measuring secondary solutions, the analysis given here omits potentially important handling effects.

Acknowledgment

Brynn Hibbert provided valuable input to the pH application. Steve Ellison provided useful comments on the effect of correlation among certified standards and other input quantities.

Example E1.2

Straight-line calibration in errors-in-variables models

S. Martens, K. Klauenberg, C. Elster

E1.2.1 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 [2, 4] or propagating distributions [3, 4]. 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 E4.3),
— "Quantifying uncertainty when comparing measurement methods – Haemoglobin concentration as an example of correlation in straight-line regression" (see E5.3), and
— "Calibration of a torque measuring system – GUM uncertainty evaluation for least-squares versus Bayesian inference" (see E6.2).

Example E1.3

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

E1.3.1 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 LPU in JCGM 100:2008 in that it combines prior knowledge about the measurand with the data obtained during calibration. From the joint posterior probability density function which is obtained from this combination, a value and a coverage interval for the measurand are obtained.

E1.3.2 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 [3], the available information is a best estimate and its associated uncertainty. A comparison of results between LPU, MCM and the Bayesian approach is given in this example. 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 choosing various prior parameter values for Gaussian prior distributions.

The data and sources of this example are available electronically [12].

E1.3.3 Specification of the measurand

As described in JCGM 101:2008 [3], 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.

It is usual to work in terms of conventional masses. The conventional mass $m_{\mathrm{W,c}}$ of W is the mass of a (hypothetical) weight of density $\rho_0 = 8 \times 10^3\,\mathrm{kg\,m^{-3}}$ that balances W in air at density $\rho_{\mathrm{a}_0} = 1.2\,\mathrm{kg\,m^{-3}}$.

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

E1.3.4 Measurement model

According to JCGM 101:2008 [3], in terms of conventional masses $m_{\mathrm{W,c}}$, $m_{\mathrm{R,c}}$ and $\delta m_{\mathrm{R,c}}$, an approximation adequate for most purposes is

$$m_{\mathrm{W,c}} = (m_{\mathrm{R,c}} + \delta m_{\mathrm{R,c}})\left[1 + (\rho_{\mathrm{a}} - \rho_{\mathrm{a}_0})\left(\frac{1}{\rho_{\mathrm{W}}} - \frac{1}{\rho_{\mathrm{R}}}\right)\right]. \tag{E1.3.1}$$

The measurement model used in the mass calibration example of [3] is

$$\delta m = (m_{\mathrm{R,c}} + \delta m_{\mathrm{R,c}})\left[1 + (\rho_{\mathrm{a}} - \rho_{\mathrm{a}_0})\left(\frac{1}{\rho_{\mathrm{W}}} - \frac{1}{\rho_{\mathrm{R}}}\right)\right] - m_{\mathrm{nom}}. \tag{E1.3.2}$$

E1.3.5 Input quantities of the measurement model

Table E1.3.1 summarizes the input quantities $m_{\mathrm{R,c}}$, $\delta m_{\mathrm{R,c}}$, ρ_{a}, ρ_{W} and ρ_{R}, and the PDFs assigned from [3]. In the table, a Gaussian distribution $\mathrm{N}(\mu, \sigma^2)$ is described in terms of expectation μ and standard deviation σ, and a rectangular distribution $\mathrm{R}(a, b)$ with endpoints a and b $(a < b)$ in terms of expectation $(a + b)/2$ and semi-width $(b - a)/2$.

Table E1.3.1: The input quantities and PDFs assigned to them for the mass calibration model (E1.3.2), from JCGM 101:2008 [3].

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

Note that the input quantity $\delta m_{\mathrm{R,c}}$ is usually associated with fresh calibration results but that in the JCGM 101:2008 [3] treatment of mass calibration, a Type B uncertainty evaluation of $\delta m_{\mathrm{R,c}}$ is performed resulting in a Gaussian distribution $\delta m_{\mathrm{R,c}} \sim \mathrm{N}(d, u^2(d))$ where d is a best estimate with associated uncertainty $u(d)$.

E1.3.6 Uncertainty propagation

E1.3.6.1 Bayesian analysis: generalities

To set up a Bayesian framework [65, 126], a statistical model is needed for which we choose to revise notation, as in [127], so that random variables are now represented by Greek letters. In this document, we consider statistical models of the form

$$D|\eta,\boldsymbol{\theta} \sim \mathrm{N}\left([\eta - G(\boldsymbol{\theta})]/C(\boldsymbol{\theta}), u^2(d)\right) \tag{E1.3.3}$$

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 (E1.3.3) is equivalent to the measurement model (E1.3.2)

$$\eta = G(\boldsymbol{\theta}) + C(\boldsymbol{\theta})\zeta \tag{E1.3.4}$$

with

$$\zeta = \delta m_{\mathrm{R,c}}, \tag{E1.3.5}$$

$$\boldsymbol{\theta} = \left(\rho_{\mathrm{a}}, \rho_{\mathrm{W}}, \rho_{\mathrm{R}}, m_{\mathrm{R,c}}\right), \tag{E1.3.6}$$

$$C(\boldsymbol{\theta}) = 1 + (\rho_{\mathrm{a}} - \rho_{\mathrm{a_0}})\left(\frac{1}{\rho_{\mathrm{W}}} - \frac{1}{\rho_{\mathrm{R}}}\right), \tag{E1.3.7}$$

$$G(\boldsymbol{\theta}) = C(\boldsymbol{\theta})m_{\mathrm{R,c}} - m_{\mathrm{nom}}. \tag{E1.3.8}$$

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

$$\pi(\eta|d) = \int_{\boldsymbol{\theta}} \pi(\eta,\boldsymbol{\theta}|d)\mathrm{d}\boldsymbol{\theta}, \tag{E1.3.9}$$

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

In this document, point estimates are derived from equation (E1.3.9) 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 [3], similar to highest posterior density (HPD) intervals in Bayesian statistics.

E1.3.6.2 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.).

E1.3.6.3 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 $\pi(\eta, \boldsymbol{\theta})$:

$$\pi(\eta, \boldsymbol{\theta}|d) = \frac{l(d|\eta, \boldsymbol{\theta})\pi(\eta, \boldsymbol{\theta})}{m(d)}, \tag{E1.3.10}$$

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, (E1.3.10) can be translated into the proportionality relation as follows

$$\pi(\eta, \boldsymbol{\theta}|d) \propto l(d|\eta, \boldsymbol{\theta})\pi(\eta, \boldsymbol{\theta}). \tag{E1.3.11}$$

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

$$l(d|\eta, \boldsymbol{\theta}) \propto \frac{1}{([C(\boldsymbol{\theta})]^2 s^2)^{\frac{1}{2}}} \exp\left\{-\frac{(\eta - m(\boldsymbol{\theta}))^2}{2[C(\boldsymbol{\theta})]^2 s^2}\right\}, \tag{E1.3.12}$$

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

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

$$\pi(\eta, \boldsymbol{\theta}|d) \sim \mathrm{N}\left(m(\boldsymbol{\theta}), [C(\boldsymbol{\theta})]^2 s^2\right)\pi(\boldsymbol{\theta}). \tag{E1.3.13}$$

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

$$\pi(\eta, \boldsymbol{\theta}|d) \sim \mathrm{N}\left(m_{\mathrm{p}}(\boldsymbol{\theta}), \sigma_{\mathrm{p}}^2(\boldsymbol{\theta})\right)\pi(\boldsymbol{\theta}), \tag{E1.3.14}$$

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 (E1.3.9) 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 (E1.3.13) or (E1.3.14) giving a total of $n_{\mathrm{MC}} \times n_{\mathrm{post}}$ simulations.

E1.3.7 Reporting the result

E1.3.7.1 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 E1.3.2 (LPU_1 and LPU_2 denote respectively the first and second order Taylor approximations) and plotted in Figure E1.3.1. The comparison shows a good agreement between methods[1].

[1] For the so-called non informative prior, [127] showed that Bayesian marginal posterior uncertainty coincides with the MCM uncertainty estimate when the model is linear.

Table E1.3.2: 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 [3].

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]

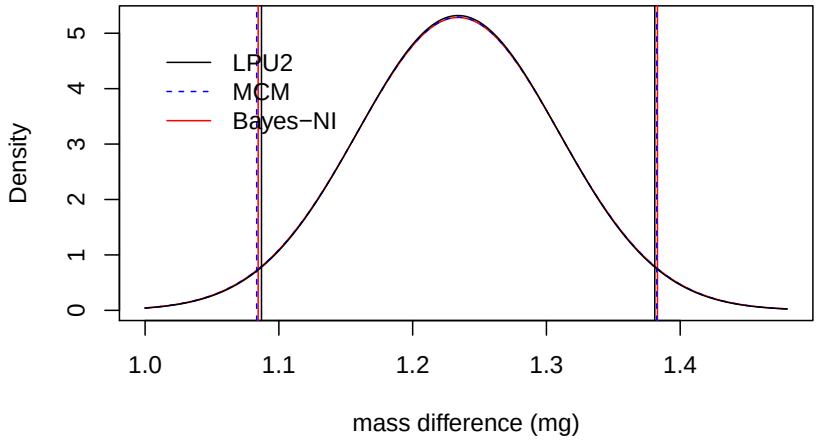


Figure E1.3.1: Distributions of mass difference δm obtained under Gaussian approximation with LPU2, MCM and Bayes-NI from the values in Table E1.3.2.

Results obtained with a Gaussian prior distribution are displayed in Table E1.3.3 and plotted in Figure E1.3.2. 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 E1.3.3: 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 ($n_{\mathrm{MC}} = 20\,000$, $n_{\mathrm{post}} = 1000$).

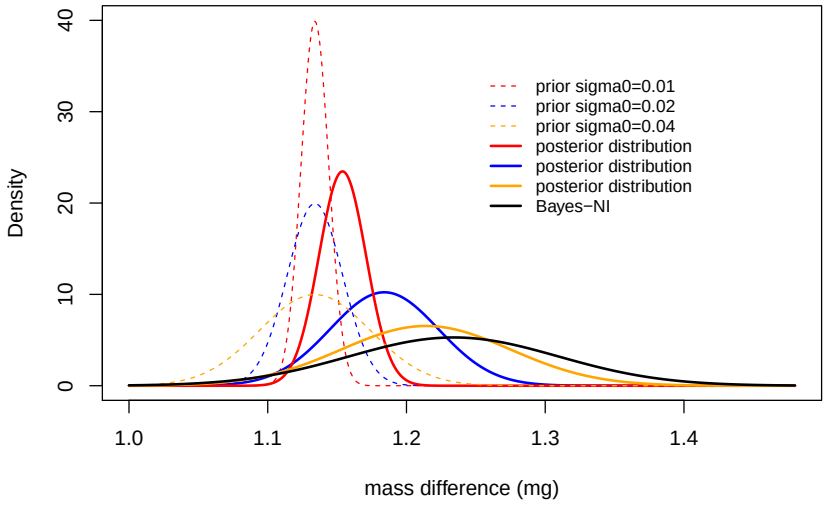


Figure E1.3.2: Posterior distributions of δm obtained under Gaussian prior distributions from the values in Table E1.3.3.

E1.3.8 Conclusion

This document shows the main features of a Bayesian approach of uncertainty evaluation applied to the mass calibration example in JCGM 101:2008 [3]. 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 [3] and by the Gaussian distribution from LPU [2].

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 MCM determines a sample for linear measurement models, see for instance [127] and [128] 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 [129], usually used to sample from high-dimensional integrals, as in [128] and [130], which can be helpful for any practitioner already familiar with MCM willing to perform a Bayesian uncertainty analysis.

Example E1.4

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

E1.4.1 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 [131]. 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) [2] described in [132].

E1.4.2 Introduction of the application

The objective of the SBI standard test [131] 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° 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.

E1.4.3 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 [131], the definitions of these quantities are related to different time periods of the SBI test, namely:

$$\mathrm{HRR}(t) = \begin{cases} 0, & t \leq 300\,\mathrm{s} \\ \max\left[0, \mathrm{HRR}_{\mathrm{total}}(t) - \mathrm{HRR}_{\mathrm{burner}}\right], & 300\,\mathrm{s} \leq 312\,\mathrm{s} \\ \mathrm{HRR}_{\mathrm{total}}(t) - \mathrm{HRR}_{\mathrm{burner}}, & 312\,\mathrm{s} < t \end{cases} \tag{E1.4.1}$$

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

$$\mathrm{SPR}(t) = \begin{cases} 0, & t \leq 300\,\mathrm{s} \\ \max\left[0, \mathrm{SPR}_{\mathrm{total}}(t) - \mathrm{SPR}_{\mathrm{burner}}\right], & 300\,\mathrm{s} < t \leq 312\,\mathrm{s} \\ \mathrm{SPR}_{\mathrm{total}}(t) - \mathrm{SPR}_{\mathrm{burner}}, & 312\,\mathrm{s} < t \end{cases} \tag{E1.4.2}$$

where $\mathrm{SPR}_{\mathrm{total}}(t)$ is the total smoke production rate of the specimen and the main burner in the time instant t, while $\mathrm{SPR}_{\mathrm{burner}}$ 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 $\mathrm{HRR}_{\mathrm{burner}}$ and $\mathrm{SPR}_{\mathrm{burner}}$, based on average values obtained when combustion occurs only in an auxiliary burner (identical to the main burner) installed in the experimental apparatus.

The heat release rate is a key intermediate quantity in the determination of two main output quantities of the SBI test – THR, the total heat release (usually expressed in MJ) from the specimen in a certain time exposure to the main burner flames (namely, in the first 600 s), and FIGRA, the fire growth rate (expressed in $W\,s^{-1}$), 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.

E1.4.4 Measurement model

The heat release rate measurement model is derived from the studies performed by [133] 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 $\mathrm{MJ\,m^{-3}}$) is considered constant regardless of the combustion material, which can be expressed by

$$\mathrm{HRR} = E' \cdot x^0_{\mathrm{O_2}} \cdot \varphi \cdot q_{V_{s'}} \tag{E1.4.3}$$

where $x^0_{\mathrm{O_2}}$ is the oxygen amount 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 $\mathrm{m^3\,s^{-1}}$).

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

$$E' = \rho_{\mathrm{O_2}} \cdot E \tag{E1.4.4}$$

The amount fraction oxygen in the ambient is given by

$$x^0_{\mathrm{O_2}} = x^{\mathrm{initial}}_{\mathrm{O_2}} \cdot \left(1 - x^{\mathrm{o}}_{\mathrm{H_2O}}\right) \tag{E1.4.5}$$

$x^{\mathrm{initial}}_{\mathrm{O_2}}$ 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 $x^0_{\mathrm{H_2O}}$, 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)

$$x^0_{\mathrm{H_2O}} = \frac{\mathrm{rh}}{100 \cdot p_{\mathrm{atm}}} \exp\left[23.2 - \frac{3816}{T_{\mathrm{initial}} - 46}\right] \tag{E1.4.6}$$

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

$$\varphi = \frac{x^{\mathrm{initial}}_{\mathrm{O_2}}(1 - x_{\mathrm{CO_2}}) - x_{\mathrm{O_2}}(1 - x^{\mathrm{initial}}_{\mathrm{CO_2}})}{x^{\mathrm{initial}}_{\mathrm{O_2}}(1 - x_{\mathrm{O_2}} - x_{\mathrm{CO_2}})} \tag{E1.4.7}$$

where $x^{\mathrm{initial}}_{\mathrm{O_2}}$ and $x^{\mathrm{initial}}_{\mathrm{CO_2}}$ 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 $x_{\mathrm{CO_2}}$ 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

$$q_{V_S} = \frac{q_{V_c}}{1 + (\alpha - 1) \cdot \varphi} \tag{E1.4.8}$$

[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.

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

$$\alpha = 1 + x^0_{O_2}(\beta - 1) \tag{E1.4.9}$$

where $x^0_{O_2}$ is obtained from expression (E1.4.5) 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 ($C_aH_bO_cN_dX_e$) 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

$$C_aH_bO_cN_dX_e + \left(a - \frac{g}{2} + \frac{b-e}{4} - \frac{c}{2}\right)O_2 \rightarrow (a-c)CO_2 + \frac{b-e}{2}H_2O + eHX + gCO + \frac{d}{2}N_2 \tag{E1.4.10}$$

where particular constants a to g apply in any specific instance. Therefore, based on expression (E1.4.10) and by definition, the β ratio is given by

$$\beta = \frac{4a + 2b + 2e + 2d}{4a + b - e - 2c - 2g} \tag{E1.4.11}$$

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

$$q_{V_c} = \frac{k_t}{k_p} \cdot \sqrt{\frac{2\Delta p}{\rho_{T_0}} \cdot \frac{T_0}{T}} \cdot A \tag{E1.4.12}$$

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

$$A = \frac{\pi}{4} \cdot d^2 \tag{E1.4.13}$$

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 [134] is used:

$$\rho_{T_0} = \frac{0.34848 \cdot p_{Atm} - 0.009024 rh \exp[0.0612 \cdot (T_0 - 273.15)]}{T_n} \tag{E1.4.14}$$

[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 differential pressure correction factor is considered in expression (E1.4.12) due to the use of a bidirectional sensor [135] instead of a conventional Pitot tube (vulnerable to solid particles in the flow). This quantity is defined by

$$k_p = \frac{\sqrt{\frac{\Delta p}{\rho_{T_{\text{amb}}}}}}{v_{\text{c}}} \tag{E1.4.15}$$

where v_{c} is the linear flow velocity in the centre of the exhaust duct cross-section (in $\text{m}\,\text{s}^{-1}$) and $\rho_{T_{\text{amb}}}$ is the moist air density [134] for ambient temperature, T_{amb} (in K) given by

$$\rho_{T_{\text{amb}}} = \frac{0.34848 p_{\text{atm}} - 0.009024 \text{rh} \exp\left[0.0612\left(T_{\text{amb}} - 273.15\right)\right]}{T_{\text{amb}}} \tag{E1.4.16}$$

The global correction factor, k_{t}, corresponds to the average of three individual corrections, $k_{t,v}$ $k_{\text{t propane}}, k_{\text{t heptane}}$ 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 $k_{\text{t,v}}$ correction, its quantification is supported by

$$k_{\text{t,v}} = \frac{\sum_{i=1}^{5} v_i}{5 \cdot v_{\text{c}}} \tag{E1.4.17}$$

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 $\text{m}\,\text{s}^{-1}$.

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.

$$k_{\text{t,propane}} = k_t' \cdot \frac{\sum_i \text{HRR}_i^{\text{theoretical}}}{\sum_i \text{HRR}_i^{\text{experimental}}} \tag{E1.4.18}$$

considering the several testing steps indexed by i of this normalised test [131], 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

$$\text{HRR}_i^{\text{theoretical}} = q_{m_i} |\Delta h_{\text{c}}^{\text{l}}|_{\text{propane}'} \tag{E1.4.19}$$

where q_{m_i} is the propane mass flow in the i th testing step (expressed in $\text{kg}\,\text{s}^{-1}$), and $\left|\Delta h_{\text{c}}^{\text{l}}\right|_{\text{propane}}$ is the low enthalpy of propane combustion per unit of mass (in $\text{kJ}\,\text{kg}^{-1}$). It should be noted that, in the calculation of $\text{HRR}_i^{\text{experimental}}$ by expression (E1.4.3), 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 [132] and used for construction materials in the SBI test. Regarding the k_{theptane} correction, this quantity is obtained from the expression

$$k_{\text{t heptane}} = k_t' \cdot \frac{|\Delta h_{\text{c}}^{\text{l}}|_{\text{heptane}} \cdot m_{\text{heptane}}}{\text{THR}} \tag{E1.4.20}$$

[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 (E1.4.18).

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 $\text{kJ}\,\text{kg}^{-1}$) 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 (E1.4.3).

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

$$\text{SPR} = \frac{q_{V_c}}{L} \cdot \frac{T}{T_0} \cdot \ln\left(\frac{1}{\tau}\right) \tag{E1.4.21}$$

where q_{V_c} is the volumetric flow rate of gases in the exhaust duct (in $\text{m}^3\,\text{s}^{-1}$), obtained from expression (E1.4.12); 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 E1.4.1 shows a schematic representation of the heat release rate calculation process, while figure E1.4.2 refers to the smoke production rate. Particular attention is given to the global correction factor and to its calculation process, schematically represented in figure E1.4.3.

E1.4.5 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 [3,4], 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 [46] was used to obtain numerical sequences with a typical dimension (number of trials) of 1×10^6, 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 E1.4.1 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.

Table E1.4.1: 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	$1.308\,\text{kg}\,\text{m}^{-3}$	$0.003\,\text{kg}\,\text{m}^{-3}$
Heat released per unit of consumed oxygen mass	E	Gaussian	$13.1\,\text{MJ}\,\text{kg}^{-1}$	$0.3\,\text{MJ}\,\text{kg}^{-1}$
Initial amount fraction of oxygen	$x_{O_2}^{\text{initial}}$	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	$x_{CO_2}^{\text{initial}}$	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	$9.6\,\text{m}\,\text{s}^{-1}$	$0.7\,\text{m}\,\text{s}^{-1}$
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 E1.4.2 presents the adopted probabilistic formulation of the input quantities which supported the MCM simulations.

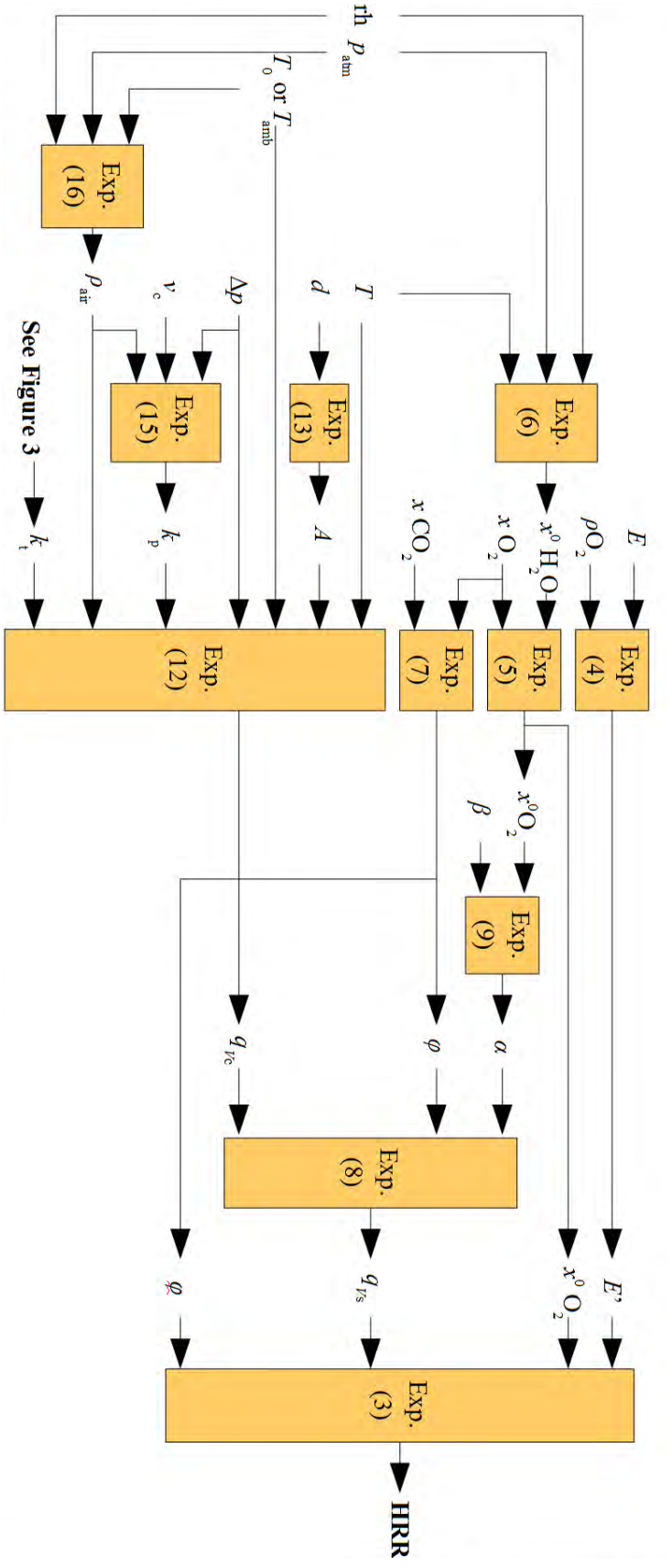


Figure E1.4.1: Functional diagram of the heat release rate quantity

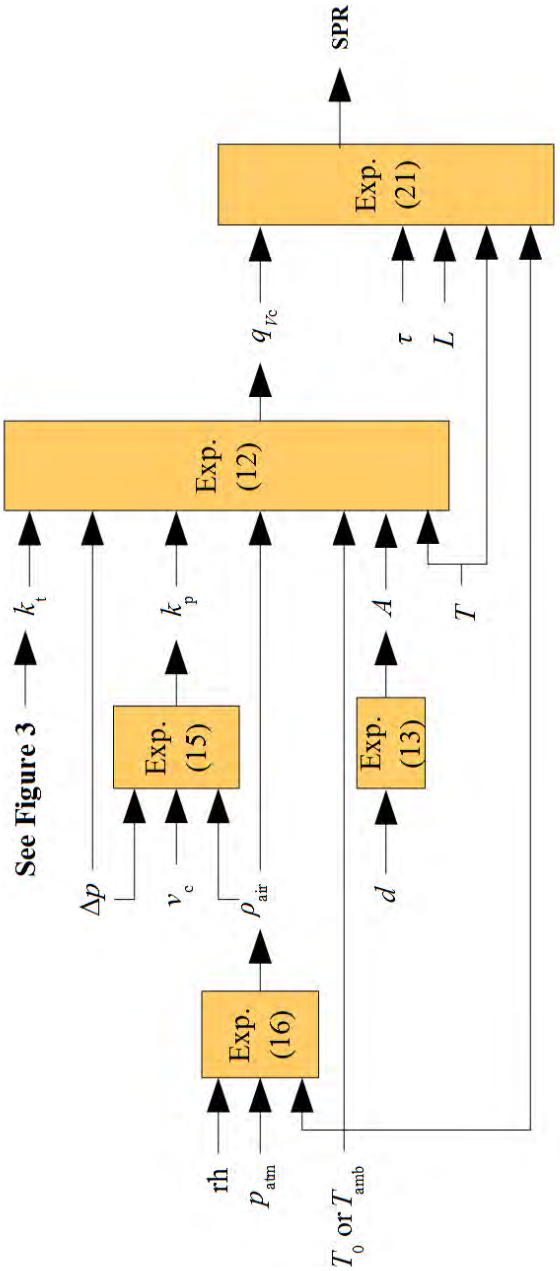


Figure E1.4.2: Functional diagram of the smoke production rate

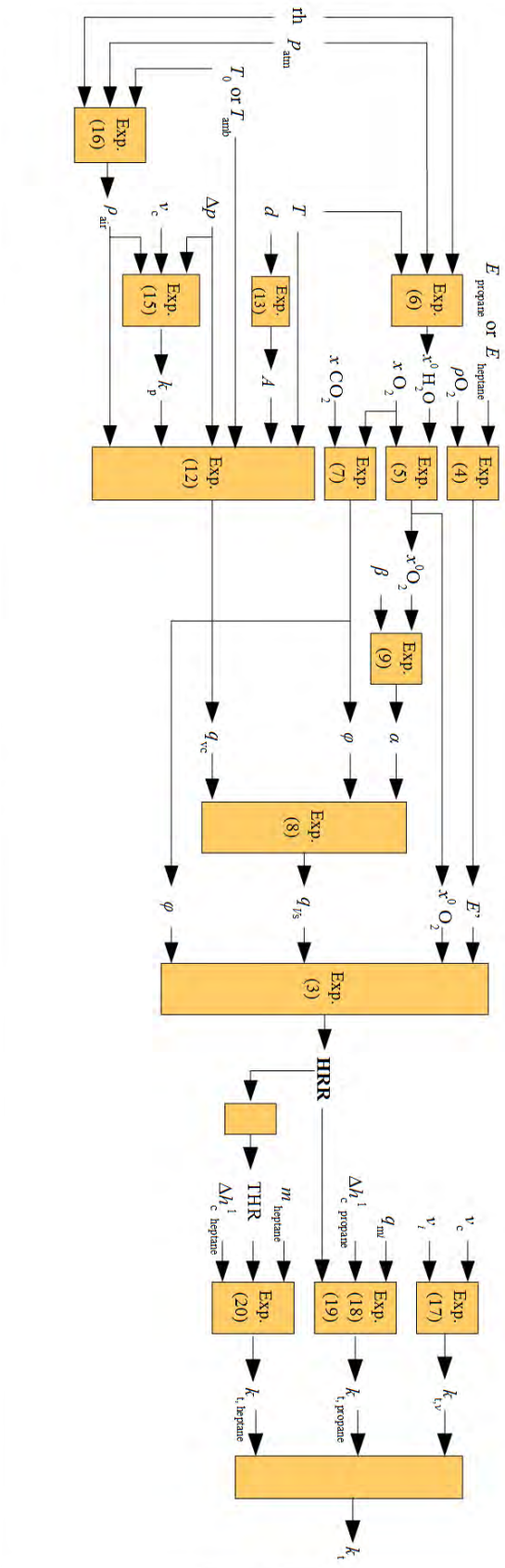


Figure E1.4.3: Functional diagram of the global correction factor quantity rate quantity

Table E1.4.2: 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 E1.4.3: MCM simulation results for intermediate quantities in the calculation of the heat release rate

Quantity	Symbol	PDF	Estimate	Standard uncertainty
Water vapour amount fraction	$x^0_{\mathrm{H_2O}}$	Gaussian	0.0146	0.0004
Moist air density for ambient temperature	$\rho_{T_{\mathrm{amb}}}$	Gaussian	1.180 kg m^{-3}	0.005 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	$q_{V_{\mathrm{c}}}$	Gaussian	0.55 m^3 s^{-1}	0.04 m^3 s^{-1}
Heat released per unit of consumed oxygen volume	E'	Gaussian	17.1 MJ m^{-3}	0.4 MJ m^{-3}
Ambient oxygen molar fraction	$x^0_{\mathrm{O_2}}$	Gaussian	0.2074	0.000 05
Oxygen depletion factor	φ	Gaussian	0.015	0.001
Ambient volumetric flow rate	$q_{V_{\mathrm{s}}}$	Gaussian	0.55 m^3 s^{-1}	0.04 m^3 s^{-1}

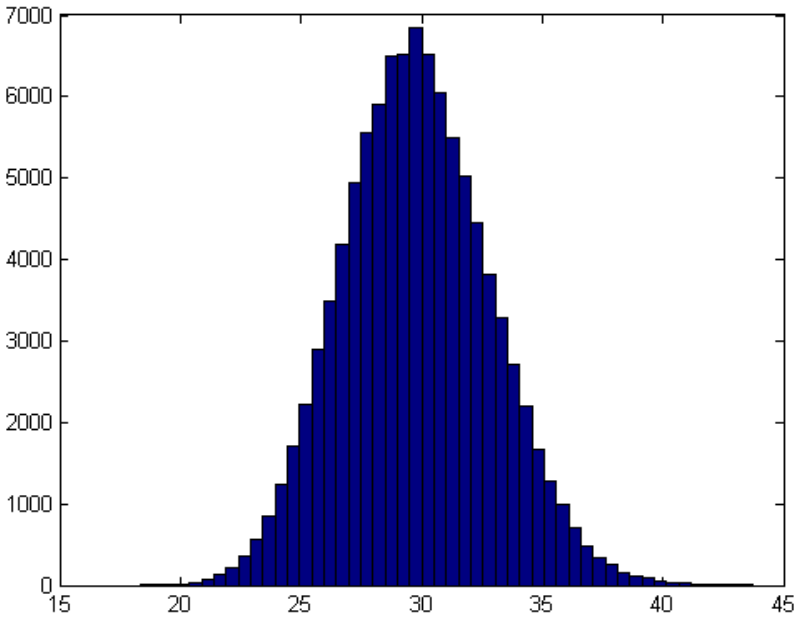


Figure E1.4.4: Output PDF of the heat release rate quantity

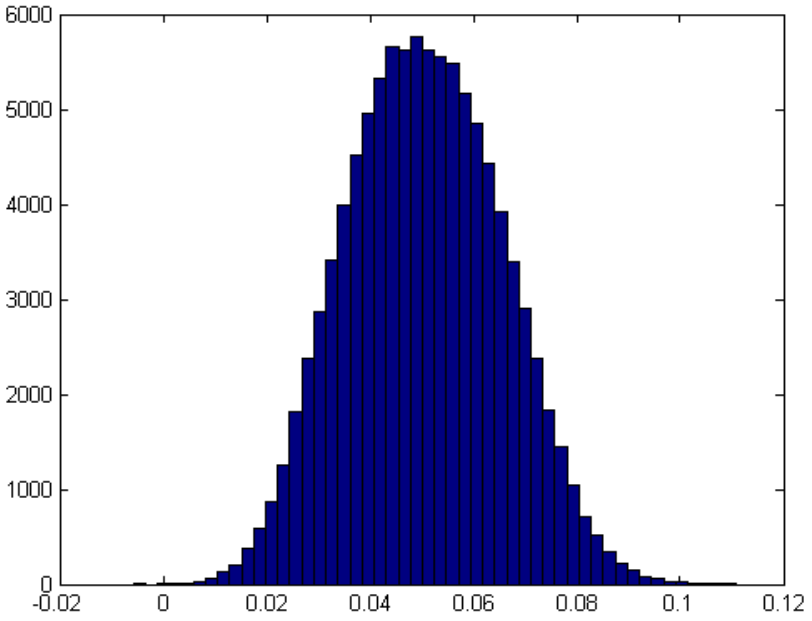


Figure E1.4.5: Output PDF of the smoke production rate quantity

Table E1.4.4: 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 E1.4.5: 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	$\rho_{T_{\mathrm{amb}}}$	Gaussian	$1.220\,\mathrm{kg\,m^{-3}}$	$0.005\,\mathrm{kg\,m^{-3}}$
Differential pressure correction factor	k_{p}	Gaussian	1.11	0.09
Volumetric flow rate of gases in the exhaust duct	$q_{V_{\mathrm{c}}}$	Gaussian	$0.55\,\mathrm{m^3\,s^{-1}}$	$0.04\,\mathrm{m^3\,s^{-1}}$

Table E1.4.6: MCM simulation results for the smoke production rate ($0.05\,\mathrm{m^2\,s^{-1}}$ level)

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

E1.4.6 Reporting the result

The measurement uncertainties of the input quantities shown in table E1.4.1 were propagated by the MCM to the intermediate quantities (results shown in table E1.4.3) and, posteriorly, to the total heat release rate quantity (see table E1.4.4 for an example of a 30 kW thermal power level). Figure E1.4.4 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 E1.4.5 (intermediate quantities) and table E1.4.6 (output quantity). Figure E1.4.5 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 %.

E1.4.7 Interpretation of results

As seen in figures E1.4.4 and E1.4.5, 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 E1.4.1) were taken as Gaussian. In terms

of validation of results, tables E1.4.4 and E1.4.6 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 GUF approach, relies on its greater simplicity and accuracy when dealing with a large number of input quantities.

Example E1.5

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

K. Shirono, M.G. Cox

E1.5.1 Summary

This example illustrates the minimal adjustment of calibration and measurement capability (CMC) uncertainty claims so they are supported by the results of a key comparison (KC). According to the CIPM Mutual Recognition Arrangement (MRA) [78, 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 (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 [136].

E1.5.2 Introduction of the application

CMCs must be consistent with results derived from KCs [78, 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 [78, 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 [137]. 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 [137].

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 national metrology institute (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 [136] relating to a KC of free-field hydrophone calibrations in the frequency interval 1 kHz to 500 kHz [138]. 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.

E1.5.3 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 E1.5.4.

E1.5.4 Measurement model

The measurement model is multi-staged[1] [5] 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 $x_i(j)$, $j = 1,\ldots,p$ [139]. Thus, the according DoEs $(d_i(j), U(d_i(j)))$, $i = 1,\ldots,N$, $j = 1,\ldots,p$, defined as follows are available.

E1.5.4.1 Degrees of equivalence

The DoE value component for laboratory i and parameter j is

$$d_i(j) = x_i(j) - x_{\mathrm{ref}}(j), \qquad i = 1,\ldots,N, \tag{E1.5.1}$$

and the corresponding uncertainty component is

$$U(d_i(j)) = k_i u(d_i(j)) = \begin{cases} k_i[u^2(x_i(j)) - u^2(x_{\mathrm{ref}}(j))]^{1/2}, & i \in I_{\mathrm{ref}}, \\ k_i[u^2(x_i(j)) + u^2(x_{\mathrm{ref}}(j))]^{1/2}, & \text{otherwise}, \end{cases} \tag{E1.5.2}$$

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 $x_{\mathrm{ref}}(j)$ [139]. 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.

[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'

$$E_n^{(i)} = \frac{d_i}{U(d_i)} = \frac{d_i}{k_i u(d_i)} \tag{E1.5.3}$$

is used: If $|E_n^{(i)}| \le 1$, laboratory i's performance is regarded as 'satisfactory'; otherwise it is 'unsatisfactory' [9].

Exclusive statistics $d_{i,\mathrm{ex}}(j)$ rather than $d_i(j)$ [71, 140], 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 [136], are used to describe the DoEs because of algebraic advantages over conventional statistics. The associated standard uncertainty $u(x_{\mathrm{ex}}(j))$ is provided.

E1.5.4.2 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.
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.

E1.5.4.3 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 E1.5.4.4.
   - Technical parameter λ_i related to the standard deviations S_i in section E1.5.4.4.

   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.

E1.5.4.4 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:

$$\boldsymbol{d}_{i,\mathrm{ex}} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_i), \tag{E1.5.4}$$

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

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

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 $\boldsymbol{\Sigma}_i$ and $\boldsymbol{P}_i$ is a correlation matrix [4]. Neither $\boldsymbol{S}_i$ nor $\boldsymbol{P}_i$ is typically available from the KC and must be estimated from reported results.

Define

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

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

$$\sigma_i^2(j) = \lambda_i[u^2(x_i(j)) + u^2(x_{\text{ex}}(j))] \approx (1 + r_{i,\min})\lambda_i u^2(x_i(j)). \tag{E1.5.5}$$

The expression in the right side of (E1.5.5) is obtained through applying the relationship

$$\tau_i + [r_i(j) - r_{i,\min}] \approx \tau_i, \qquad \tau_i = (1 + r_{i,\min})\lambda_i. \tag{E1.5.6}$$

Although we cannot say that expression (E1.5.6) is always a reasonable approximation, we can confirm the extent of its validity after estimating the parameters specified in section E1.5.4.3. More details are given in [136], 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 [136] has the form

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

where $\boldsymbol{I}$ is the identity matrix of dimension $p \times p$ and $\mathbf{1}$ 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:

$$p(\rho_i) \propto \begin{cases} 1, & 0 \le \rho_i \le 1, \\ 0, & \text{otherwise} \end{cases}, \qquad p(\lambda_i) \propto \begin{cases} {\lambda_i}^{-1}, & \lambda_i \ge 1, \\ 0, & \text{otherwise.} \end{cases} \tag{E1.5.7}$$

E1.5.4.5 Data

The data used in this example is for KC CCL.K-2 [141] 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" [141]. 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 E1.5.1 and figure E1.5.1. The reference values and their associated standard uncertainties are also given in table E1.5.1.

E1.5.5 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 [136]:

(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),

Table E1.5.1: 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 [141]

	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

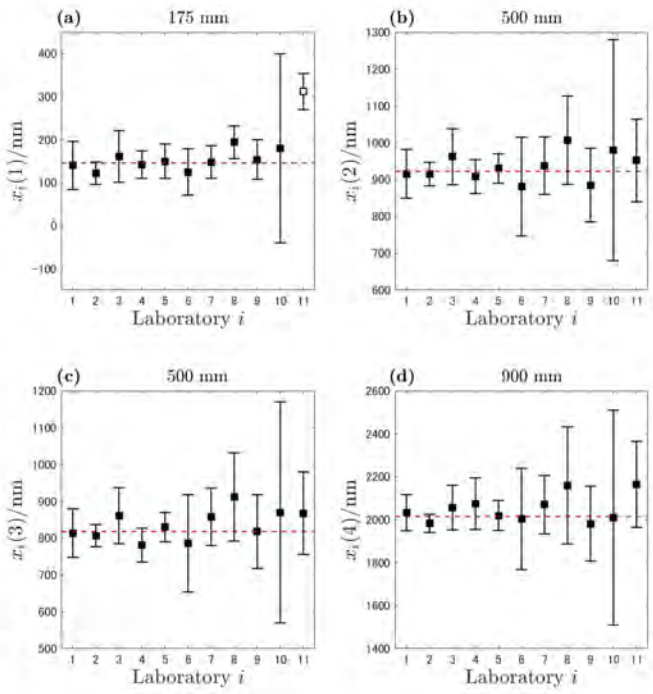


Figure E1.5.1: CCL.K-2 gauge block data for four nominal lengths and weighted means as KCRVs (broken horizontal lines). Vertical bars depict coverage intervals with coverage factor $k_i = 1.96$

(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 [136]).

E1.5.6 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 (E1.5.6) 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 E1.5.2 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 (E1.5.6) does not hold in these results. In [136], we gave the criterion $K_i/r_{i,\max} > 4$ to check the appropriateness of relationship (E1.5.6).

Table E1.5.2: 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

E1.5.7 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]

[4]Since the model (E1.5.4) expresses the variation between the reported values, no bias is shown directly.

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 [141], the laboratory found a problematic issue with its measuring system (see section E1.5.4.5). For that laboratory, whilst all the deviations are 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 E1.6

Model-based unilateral degrees of equivalence in analysis of a regional metrology organization key comparison

K. Shirono, M.G. Cox

Analysis of an RMO key comparison

E1.6.1 Summary

Measurement uncertainties associated with the unilateral degrees of equivalence (DoEs) in a KC of a regional metrology organisation (RMO) are discussed. Unilateral DoEs are obtained through assessing a linking invariant to relate the RMO results to the KC conducted by the International Committee of Weights and Measures (CIPM). A new approach to derive the unilateral DoEs is suggested, based on statistical testing. Since the mathematical model of the unilateral DoEs is given by a linear combination of the reported values and the reference value in the CIPM KC, the principle in the JCGM's document JCGM 102 [4] is applied to obtain the associated uncertainties. The proposed approach may give different DoEs from those given by some existing approaches. Since decisions are made on the basis of calculated DoEs, a conclusion is that it is valuable to have adequate knowledge of the properties of available linking methods for analyzing the results from an RMO KC.

E1.6.2 Introduction of the application

Calibration and Measurement Capabilities (CMCs) of a NMI are established through the framework of the RMO [142]. By maintaining CMCs, the progress of measurement science is delivered to diverse industries. From a technical point of view, KCs support CMCs.

It is noted in the CIPM Mutual Recognition Arrangement (MRA) that "the technical basis of this arrangement is the set of results obtained in the course of time through key comparisons carried out by the Consultative Committees of the CIPM, the Bureau International des Poids

et Mesures (BIPM) and the RMOs". Further, it is described that "participation in a CIPM key comparison is open to laboratories having the highest technical competence and experience, normally the member laboratories of the appropriate Consultative Committee". Technical and economic reasons are often barriers to participation in a CIPM KC. In both senses, the linking of an RMO KC to a CIPM KC is the only realistic way to establish global metrological traceability. Moreover, such linking is explicitly required: technical supplement T.4 of the MRA states "the results of the RMO key comparisons are linked to key comparison reference values established by CIPM key comparisons by the common participation of some institutes in both CIPM and RMO comparisons".

For the linking process, the presence of laboratories that participate in both CIPM and RMO KCs is technically required. Otherwise, statistical relations between the two KCs could not be established. The laboratories participating in both KCs are termed linking laboratories. The CMC information of these laboratories has been checked by the participation of the CIPM KC before the implementation of the RMO KC.

Although no official guidance concerning the linking procedure has been provided by the CIPM, several suggestions for the method of analysis have been made from the academic point of view [143–147]. One possible aproach is to determine a linking invariant (cf. [147]) equal to the difference between the measurands in the CIPM and the RMO KCs. The reported values in the RMO are adjusted by a estimate of the linking invariant. The value components — see section E1.6.3 — of the unilateral degrees of equivalences (DoEs) are given as the difference between the adjusted reported values and the KCRV in the CIPM KC.

In the study by Kharitonov and Chunovkina [146], we find two types of linking invariant. However, since no statistical models were given in [146], we cannot clearly specify what can be estimated by these linking invariants. We could not find possible and reasonable models that gave the mathematical forms in their study. Decker et al. [147] reported the application of one of the linking invariants proposed in [146].

There are other possible approaches for deriving a linking invariant. In this paper, we will develop a method to derive a linking invariant using a generalized least squares (GLS) method. In particular, we develop a method in which the measurand that is estimated by the KCRV of the CIPM KC is not re-estimated using a least squares method, based on the premise that the CIPM KCRV has been completed before the RMO KC is implemented. Consequently, we obtain a different linking invariant from those in some previous studies. Statistical testing using the assessed unilateral DoEs shows that the choice of linking invariant influences the performance evaluation of laboratories.

E1.6.3 Specification of the measurand(s)

The principal measurands in this study are the unilateral DoEs in the RMO KC. (A measurand is defined in the VIM as the quantity intended to be measured [89, definition 2.3].) We regard a DoE as a *measurement result*, that is, as a single measured value *and* a measurement uncertainty, in accordance with the VIM [89, definition 2.9, note 2] and for compatibility with the MRA [142, appendix B]. In the MRA, a DoE is regarded as having a value component and an uncertainty component, with the latter expressed as an expanded uncertainty at a 95 % level of confidence. Here, the *value component* of the unilateral DoE for laboratory i is a statistic, denoted by d_i, to express the difference between the reported value after adjustment by a linking invariant. The value component alone is often referred to in the literature as *the* DoE.

The linking invariant is an intermediate measurand when obtaining DoEs. The linking invariant as the population parameter is denoted by η and its estimate by h_{link}. When reporting y_i from laboratory i in the RMO KC, the relationship between d_i and h_{link} is

$$d_i = y_i + h_{\text{link}} - x_{\text{ref}}, \tag{E1.6.1}$$

where x_{ref} is the KCRV in the CIPM KC.

The developed uncertainty evaluation is applied to a completed RMO comparison of EURAMET-AUV.V-K1.1 [148], which was linked to CCAUV.V-K1 [149]. They were key comparisons in which the charge sensitivities of transducers were measured.

E1.6.4 Measurement model

Suppose m laboratories participate in a CIPM KC and n laboratories in a corresponding RMO KC, L of which participate in both comparisons and hence are linking laboratories. In this work, it is assumed that the data have already been screened (that is, no outlier remains) and the data corresponds to what the CC working group considers suitable for analysis in the RMO KC.

Let the reported values in the CIPM KC, assumed to be obtained independently, and their associated standard uncertainties be x_i and $u(x_i)$, $i = 1, \ldots, m$. When the KCRV is given as the weighted mean (WM) of the reported values, the KCRV x_{ref} and its associated standard uncertainty $u(x_{\text{ref}})$ are given by Cox [70]:

$$x_{\text{ref}} = u^2(x_{\text{ref}}) \sum_{i=1}^{m} \frac{x_i}{u^2(x_i)}, \qquad u^2(x_{\text{ref}}) = \left[\sum_{i=1}^{m} \frac{1}{u^2(x_i)} \right]^{-1}. \tag{E1.6.2}$$

Define the two sets $\{x_1, \ldots x_L\}$ and $\{x_{L+1}, \ldots x_m\}$ to be the reported values from the linking and non-linking laboratories, respectively, in the CIPM KC. Although the ordering in these two sets is not unique, the results of the analysis do not depend on it. We consider only cases where the reference value for the CIPM KC is given as the weighted mean of x_i, $i = 1, \ldots, m$, as given in formula (E1.6.2).

The data in the RMO KC are likewise classified into two sets: $\{y_1, \ldots y_L\}$ and $\{y_{L+1}, \ldots y_n\}$. The correlation between x_i and y_i for $i = 1, \ldots, L$ is denoted by ρ_i, information to determine which is typically obtained from the participants' uncertainty budgets [124, 144]. As in the CIPM KC, the identification is not unique, but the results of the analysis do not depend on it. For $i > L$ in the RMO KC, laboratory i is identified as the laboratory that reports y_i and $u(y_i)$ as its value for the measurand and the associated standard uncertainty.

It is considered that x_i for $i = 1$ to m and y_i for $i = 1$ to n are realizations of random variables having probability distributions with (unknown) expectations μ_x and μ_y, respectively. Define

$$\boldsymbol{z}_i = \begin{bmatrix} x_i \\ y_i \end{bmatrix}, \qquad \boldsymbol{V}_i = \begin{bmatrix} u^2(x_i) & \rho_i u(x_i) u(y_i) \\ \rho_i u(x_i) u(y_i) & u^2(y_i) \end{bmatrix}, \qquad i \in I_L,$$

where $\boldsymbol{V}_i$ is the covariance matrix associated with $\boldsymbol{z}_i$ and $I_L = \{1, \ldots, L\}$.

To develop a testing method, we assume the following statistical model for the data in the CIPM KC and the linking laboratories in the RMO KC:

$$\begin{aligned} &\boldsymbol{z}_i \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{V}_i), && i \in I_L, \\ &x_i \sim \mathrm{N}(\mu_x, u^2(x_i)), && i \in I_x. \end{aligned} \tag{E1.6.3}$$

where $\boldsymbol{\mu} = (\mu_x, \mu_x + \eta)^\top$, $\mathrm{N}(\boldsymbol{\mu}, \boldsymbol{V})$ denotes the bivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{V}$, and $I_x = \{L+1, \ldots, m\}$.

Furthermore, we assume the following statistical model for the data from the non-linking laboratories in the RMO KC:

$$y_i \sim \mathrm{N}(\mu_x - \eta, u^2(y_i)), \qquad i \in I_y = \{L+1, \ldots, n\}. \tag{E1.6.4}$$

E1.6.4.1 Estimation of the linking invariant

We yield the estimators of μ_x and η as the generalized least squares (GLS) solution x_{link} and h_{link} for μ_x and η, using the data in model (E1.6.3). First we recall that x_{link} must take the value x_{ref} since the CIPM KCRV is to be preserved. Accordingly, only h_{link} is to be determined. The solution vector for the according GLS problem is thus

$$h_{\mathrm{link}} = \arg\min_h f(h), \qquad f(h) = \sum_{i \in I_L} \boldsymbol{e}_i^\top \boldsymbol{V}_i^{-1} \boldsymbol{e}_i, \tag{E1.6.5}$$

where

$$\boldsymbol{e}_i = \boldsymbol{e}_i(x, y) = [x_i - x_{\mathrm{ref}},\ y_i + h - x_{\mathrm{ref}}]^\top. \tag{E1.6.6}$$

At the minimum of f,

$$\sum_{i \in I_L} \boldsymbol{e}_i^\top \boldsymbol{V}_i^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 0,$$

yielding, after some algebra whilst making use of expression (E1.6.6),

$$h_{\mathrm{link}} = -\frac{1}{Q} \sum_{i \in I_L} [p_i(x_i - x_{\mathrm{ref}}) + q_i(y_i - x_{\mathrm{ref}})], \tag{E1.6.7}$$

where

$$\begin{bmatrix} p_i \\ q_i \end{bmatrix} = \boldsymbol{V}_i^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \frac{1}{(1-\rho_i^2)u(x_i)u(y_i)} \begin{bmatrix} -\rho_i \\ u(x_i)/u(y_i) \end{bmatrix} \tag{E1.6.8}$$

and (also defining P, which is used later)

$$P = \sum_{i \in I_L} p_i, \qquad Q = \sum_{i \in I_L} q_i. \tag{E1.6.9}$$

For purposes (in section E1.6.5) of evaluating the standard uncertainty associated with the degree of equivalence d_i for a non-linking laboratory in the RMO key comparison (that is, $i \in I_y$), we express d_i explicitly and exactly as a linear combination of the values x_1, y_1, …, x_L, y_L, x_{ref} and y_i on which it depends. Using expressions (E1.6.1), (E1.6.7), (E1.6.8) and (E1.6.9),

$$d_i = \boldsymbol{c}_{\mathrm{DoE}}^\top \boldsymbol{s}_{i,\mathrm{DoE}},$$

where

$$\boldsymbol{c}_{\mathrm{DoE}}^\top = -\frac{1}{Q} \begin{bmatrix} p_1 & q_1 & \cdots & p_L & q_L & -P & -Q \end{bmatrix}, \tag{E1.6.10}$$

$$\boldsymbol{s}_{i,\mathrm{DoE}}^\top = \begin{bmatrix} x_1 & y_1 & \cdots & x_L & y_L & x_{\mathrm{ref}} & y_i \end{bmatrix}. \tag{E1.6.11}$$

Note that we consider the unilateral DoEs only for non-linking laboratories in the RMO KC, since the unilateral DoEs for the linking laboratories have already been evaluated in the CIPM KC.

E1.6.4.2 Data

An example is EURAMET-AUV.V-K1.1 [148], which is an RMO comparison in the area of vibration and shock. The results in the RMO KC were linked to CCAUV.V-K1 [149]. This key comparison is the first in that field, where measurements of sinusoidal linear accelerations were compared over a wide range of frequencies. It was the task of the comparison to measure the charge sensitivity of back-to-back and single-ended accelerometer standards. We focus on the reported values for 160 Hz and the back-to-back transducer.

There was one linking laboratory. It was noted in the final report that "the covariance of the different results of the linking lab is considered negligible" [148]. The data is given in table E1.6.1 and figure E1.6.1.

Table E1.6.1: EURAMET-AUV.V-K1.1 [148] and CCAUV.V-K1 [149] as an example of linking. The only linking laboratory is laboratory 1. No correlation is assumed between the two reported values from laboratory 1.

CCAUV.V-K1			EURAMET-AUV.V-K1.1		
Laboratory	x_i/ pC/(m/s^2)	$u(x_i)/x_i$ $\times 10^2$	Laboratory	y_i/ pC/(m/s^2)	$u(y_i)/y_i$ $\times 10^2$
1	0.126 64	0.05	1	0.125 21	0.05
2	0.126 70	0.25	2	0.125 49	0.16
3	0.126 60	0.15	3	0.124 72	0.28
4	0.126 60	0.23	4	0.125 26	0.30
5	0.126 60	0.25	5	0.125 19	0.10
6	0.126 60	0.25			
7	0.126 75	0.15			
8	0.126 49	0.18			
9	0.126 60	0.22			
10	0.126 82	0.20			
11	0.126 50	0.15			
12	0.126 60	0.17			
x_{ref}	0.126 63	0.04			

E1.6.5 Uncertainty analysis

E1.6.5.1 Uncertainty evaluation for DoE estimation with linking laboratory data

Define $\boldsymbol{V}_{i,\text{DoE}}$ as the covariance matrix of $\boldsymbol{s}_{i,\text{DoE}}$ (with zero elements not displayed),

$$\boldsymbol{V}_{i,\text{DoE}} = \begin{bmatrix} \boldsymbol{V}_1 & & & \boldsymbol{v}_{1,\text{ref}} & \\ & \ddots & & \vdots & \\ & & \boldsymbol{V}_L & \boldsymbol{v}_{L,\text{ref}} & \\ \boldsymbol{v}_{1,\text{ref}}^\top & \cdots & \boldsymbol{v}_{L,\text{ref}}^\top & u^2(x_{\text{ref}}) & \\ & & & & u^2(y_i) \end{bmatrix}.$$

All elements in $\boldsymbol{V}_{i,\text{DoE}}$ are known except for those in the vectors $\boldsymbol{v}_{i,\text{ref}}$, $i \in I_L$, which we now establish. For this purpose, we use basic concepts of exclusive statistics as in [136].

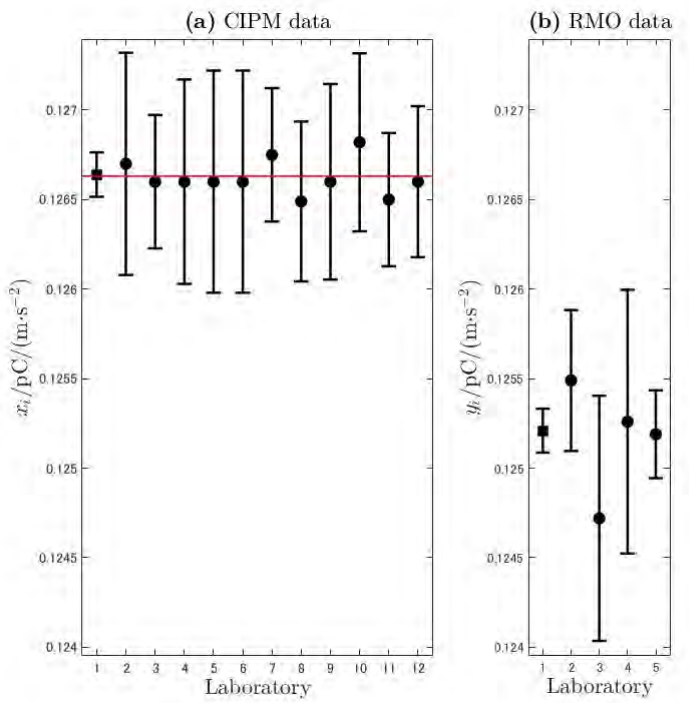


Figure E1.6.1: Reported data in (a) CCAUV.V-K1 (the CIPM KC) [149] and (b) EURAMET-AUV.V-K1.1 [148] (the RMO KC), which are given in Table E1.6.1. The horizontal red line in (a) shows x_{ref}

The exclusive WM $x_{\text{ex},j}$ for laboratory j, that is, the WM after excluding laboratory j's data, and the associated standard uncertainty $u(x_{\text{ex},j})$, are given by

$$x_{\text{ex},j} = u^2(x_{\text{ex},j}) \sum_{i \in I_L \cup I_x \setminus j} \frac{x_i}{u^2(x_i)}, \qquad u^2(x_{\text{ex},j}) = \sum_{i \in I_L \cup I_x \setminus j} \frac{1}{u^2(x_i)}.$$

As a consequence, using expressions (E1.6.2), we can write

$$\begin{aligned} x_{\text{ref}} &= u^2(x_{\text{ref}}) \sum_{i \in I_L \cup I_x \setminus j} \frac{x_i}{u^2(x_i)} + u^2(x_{\text{ref}}) \frac{x_j}{u^2(x_j)}, \\ &= u^2(x_{\text{ref}}) \frac{x_{\text{ex},j}}{u^2(x_{\text{ex},j})} + u^2(x_{\text{ref}}) \frac{x_j}{u^2(x_j)}, \end{aligned}$$

Using this result, by defining

$$\boldsymbol{z}_{j,\text{ref}} = \begin{bmatrix} x_{\text{ref}} \\ x_j \\ y_j \end{bmatrix}, \qquad \boldsymbol{C}_{j,\text{ref}} = \begin{bmatrix} u^2(x_{\text{ref}})/u^2(x_{\text{ex},j}) & & \\ u^2(x_{\text{ref}})/u^2(x_j) & 1 & \\ & & 1 \end{bmatrix},$$

we have

$$\boldsymbol{z}_{j,\text{ref}} = \boldsymbol{C}_{j,\text{ref}}^{\top} \begin{bmatrix} x_{\text{ex},j} \\ x_j \\ y_j \end{bmatrix}.$$

Applying the rule for propagating covariances [4, clause 6.2], the covariance matrix associated with $\boldsymbol{z}_{j,\mathrm{ref}}$ is

$$\boldsymbol{V}_{j,\mathrm{ref}} = \begin{bmatrix} u^2(x_{\mathrm{ref}}) & \boldsymbol{v}_{j,\mathrm{ref}}^{\top} \\ \boldsymbol{v}_{j,\mathrm{ref}} & \boldsymbol{V}_j \end{bmatrix} = \boldsymbol{C}_{j,\mathrm{ref}}^{\top} \begin{bmatrix} u^2(x_{\mathrm{ex},j}) & \\ & \boldsymbol{V}_j \end{bmatrix} \boldsymbol{C}_{j,\mathrm{ref}}.$$

In particular, after explicitly evaluating this product,

$$\boldsymbol{v}_{j,\mathrm{ref}} = \begin{bmatrix} u(x_j, x_{\mathrm{ref}}) \\ u(y_j, x_{\mathrm{ref}}) \end{bmatrix} = \begin{bmatrix} 1 \\ \rho_j \dfrac{u(y_j)}{u(x_j)} \end{bmatrix} u^2(x_{\mathrm{ref}}). \tag{E1.6.12}$$

Using expressions (E1.6.10) and (E1.6.11) and once again the rule for propagating covariances, the standard uncertainty $u(d_i)$ associated with d_i is given by

$$u^2(d_i) = \boldsymbol{c}_{\mathrm{DoE}}^{\top} \boldsymbol{V}_{i,\mathrm{DoE}} \boldsymbol{c}_{\mathrm{DoE}}.$$

Following some algebra, paying regard to symmetry we obtain

$$u^2(d_i) = u^2(y_i) + \frac{1}{Q^2} \sum_{j \in I_L} [p_j \;\; q_j] \boldsymbol{V}_j \begin{bmatrix} p_j \\ q_j \end{bmatrix} - 2\frac{P}{Q^2} \sum_{j \in I_L} \boldsymbol{v}_{j,\mathrm{ref}}^{\top} \begin{bmatrix} p_j \\ q_j \end{bmatrix} + \frac{P^2}{Q^2} u^2(x_{\mathrm{ref}}).$$

Now, using expressions (E1.6.8), (E1.6.9) and (E1.6.12),

$$[p_j \;\; q_j] \boldsymbol{V}_j \begin{bmatrix} p_j \\ q_j \end{bmatrix} = q_j, \qquad \boldsymbol{v}_{j,\mathrm{ref}}^{\top} \begin{bmatrix} p_j \\ q_j \end{bmatrix} = 0.$$

Hence,

$$u^2(d_i) = u^2(y_i) + \frac{1}{Q} + \frac{P^2}{Q^2} u^2(x_{\mathrm{ref}}).$$

The uncertainty component of the unilateral DoE is

$$U_i = k u(d_i)$$

where k is the coverage factor equal to 1.96 in this study as a consequence of the assumed normality.

E1.6.6 Reporting the result

The unilateral DoEs are shown in table E1.6.2 computed using the proposed method together with the reported values in the final report of EURAMET-AUV.V-K1.1 [148]. Since the ratios between the d_i and the U_i are the test statistics to be employed to check the statistical model (E1.6.4), the values of d_i/U_i are also shown in the table. Since the magnitudes of those values are all smaller than unity, the statistical model (E1.6.4) cannot be rejected. These test results support the equivalence of measurement that has been technically confirmed through the reviewing processes.

Table E1.6.2: Computed value and uncertainty components of unilateral DoEs and their ratios for EURAMET-AUV.V-K1.1 [148] linked to the reference value in CCAUV.V-K1 [149] using the proposed method together with a reference analysis proposed by Decker et al. [147].

Laboratory	The proposed method			Reference analysis		
	d_i /fC m^{-1} s^2	U_i /fC m^{-1} s^2	d_i/U_i	d_i /fC m^{-1} s^2	U_i /fC m^{-1} s^2	d_i/U_i
2	0.28	0.41	0.68	0.29	0.42	0.69
3	−0.49	0.70	−0.70	−0.48	0.70	−0.69
4	0.05	0.75	0.07	0.06	0.75	0.08
5	−0.02	0.27	−0.07	−0.01	0.29	−0.04

E1.6.7 Interpretation of results

In the analysis implemented in EUROMET-AUV.V-K1.1, the following linking invariant suggested in the paper reported by Decker et al. [147] was employed:

$$h_{\text{link}} = x_1 - y_1.$$

The unilateral DoEs are hence computed as follows:

$$d_i = y_i + h_{\text{link}} - x_{\text{ref}} = y_i + x_1 - y_1 - x_{\text{ref}} \qquad \text{for } i \in I_y.$$

The variance associated with d_i is

$$u^2(d_i) = \begin{bmatrix} 1 & -1 & -1 & 1 \end{bmatrix} \boldsymbol{V}_{i,\text{DoE}} \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix} = u^2(y_i) + u^2(x_1) + u^2(y_1) - u^2(x_{\text{ref}}),$$

since the covariance between x_1 and y_1 is zero. The identical mathematical method was reported by Kharitonov and Chunovkina [146]. Moreover, for the case of a single linking laboratory, the method proposed by Elster et al. [145] gives the same values as the value components of the unilateral DoEs. (When there are two or more linking laboratories, the method in reference [145] is not identical to that in reference [147].) The computed values for U_i in table E1.6.2 are slightly different from those in the final report [148], because $k = 1.96$ is used as the coverage factor instead of $k = 2$ in [148], and some rounding of numbers in the computations in the final reports. However, these minor differences do not influence the evaluation of the performance.

In the analysis applied in this work to the AUV.V-K1 key comparison, there is compensation for the difference between x_1 and x_{ref}. This bias is evaluated to be insignificant in the CIPM KC, and so no correction to the CMC information of the laboratory is applied. In general, consideration should be given to the possibility of compensation for biases even if they are insignificant.

In the proposed method, since there is only one linking laboratory, the value components of the unilateral DoEs for the AUV.V-K1 key comparison are computed as

$$d_i = y_i - y_1.$$

In fact, since no correlation is assumed for the two reported values from the linking laboratory, the results in the CIPM KC do not affect the analyses for the RMO KC. In other words, the insignificant bias in the CIPM KC is fixed to zero in this approach. Considering that the CIPM KC was conducted

to check the equivalence between the reported values and no significant bias was found for laboratory 1, the zero bias may naturally reflect the qualitative conclusion obtained through the CIPM KC.

No serious difference is found between the values with the proposed method and the actually assessed values in the case of EUROMET-AUV.V-K1.1, because x_1 is close to x_{ref}. In general, these two approaches may differ appreciably depending on the data.

To show this possibility, a dummy example is given in table E1.6.3 and figure E1.6.2. As in the actual example, Laboratory 1 is the only linking laboratory, which claims no correlation between the two reported values. In this case, the proposed method gives

$$d_2 = d_3 = d_4 = 1.9, \qquad U_2 = U_3 = U_4 = 2.2, \qquad d_2/U_2 = d_3/U_3 = d_4/U_4 = 0.9.$$

The method proposed by Decker et al. gives

$$d_2 = d_3 = d_4 = 2.6, \qquad U_2 = U_3 = U_4 = 2.3, \qquad d_2/U_2 = d_3/U_3 = d_4/U_4 = 1.1.$$

That $|d_i|/U_i$ is less than unity for the proposed method and greater than unity for Decker's method may have significance in terms of any decisions made.

Table E1.6.3: Dummy example. It is assumed that x_i and y_i are dimensionless. Laboratory 1 is the sole linking laboratory, and claims no correlation between x_1 and y_1

CIPM KC			RMO KC		
Laboratory	x_i	$u(x_i)$	Laboratory	y_i	$u(y_i)$
1	0.00	0.50	1	0.00	0.50
2	−1.30	1.00	2	1.90	1.00
3	−1.30	1.00	3	1.90	1.00
4	−1.30	1.00	4	1.90	1.00
5	−1.30	1.00			
x_{ref}	−0.65	0.35			

It should be noted that the unilateral DoEs given by the two methods are identical when $\rho_1 = 1$, because formula (E1.6.7) gives

$$h_{\mathrm{link}} \to x_1 - y_1,$$

when $\rho_1 \to 1$. The value of h_{link} is obtained using $p_1/q_1 = -\rho_1 u(y_1)/u(x_1) = -\rho_1 \to -1$ for $\rho_1 \to 1$. This result shows that the compensation implemented for the reference analysis would be similarly applied for our proposed method if the bias would be implied through the correlation information. However, when no correlation is suggested, no compensation is given. Our proposed method is based on the reliability of the uncertainty and correlation information given by the linking laboratories.

Moreover, since h_{link} has no uncertainty when $\rho_1 = 1$, the unilateral DoEs are

$$[d_i, U_i] = \left[y_i + x_1 - y_1 - x_{\mathrm{ref}},\ k\sqrt{u^2(y_i) + u^2(x_{\mathrm{ref}})} \right].$$

The uncertainty component of the unilateral DoE is determined to be smaller than that for the case of no correlation. In general, since using the correlation information can make the analysis more precise, the correlation information must be specified reliably.

The advantages of the proposed method and the method by Decker et al. can be summarized as follows:

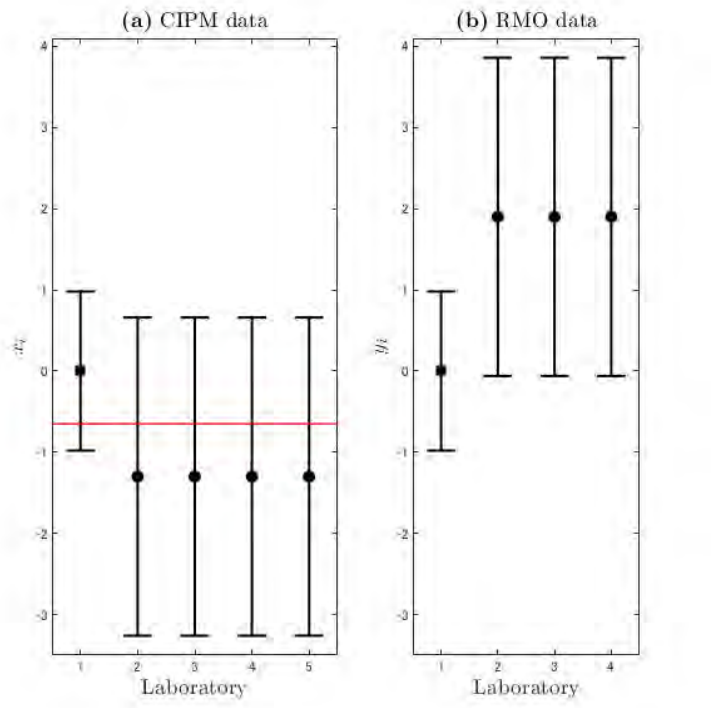


Figure E1.6.2: Dummy example in table E1.6.3. The horizontal red line in (a) shows the CIPM KCRV x_{ref}

1. The presently proposed method compensates insignificant biases in a CIPM KC only when the possible biases are implied in the correlation information.

2. The method by Decker et al. compensates insignificant biases in a CIPM KC even when the possible biases are not implied in the given information.

Moreover, since the method proposed here is based on statistical testing with a specific model, the interpretation of the analysis result is statistically clear. Since the methods have different features, the CC should choose an analysis method in accordance with its intention to implement an RMO KC. Furthermore, when linking laboratories report smaller uncertainties than non-linking laboratories in an RMO or the number of linking laboratories is large, the difference between these two methods can be marginal as shown in EURAMET-AUV.V-K1.1.

Acknowledgment

The work of the first author was partly supported by JSPS KAKENHI Grant Number 17K18411.

Example E1.7

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

E1.7.1 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-S2 [4]), and in the perhaps more familiar notation of subscripted summations (as in the GUM [2]). The matrix notation can be ignored with no loss of completeness in the examples.

E1.7.2 Introduction of the application

E1.7.2.1 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 E1.7.1.

The leaks can be calibrated using gas flow meters of the type usually found in NMIs. 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 E1.7.8). 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 [150], which in addition recommends that uncertainty in calibrated leak rate should be evaluated according to GUM principles [2], but does not provide details of the evaluation process. A more general description of leaks and leak detectors can be found in [151] and the references therein.

Figure E1.7.1: Reference leaks externally mounted on a leak detector (photograph courtesy of Vaseco Ltd.)

E1.7.2.2 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 E1.1; 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 [77, clause 6].

Case 2 has correlations present. It is addressed by following the process described in the GUM [2, annex F.1.2] and in GUM-S2 [4] 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 [77, 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 LPU-type evaluation.

E1.7.3 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.

E1.7.4 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 [77]. Case 1 uses clause 6, and Case 2 uses clause 9 of that Technical Specification. Case 3 makes use of both clauses.

E1.7.5 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 [77] 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 [77], the GUM [2] and GUM-S2 [4]. In particular, it follows the guidance on treatment of correlations elaborated in GUM Annex F.1.2 and in GUM-S2 clause 6.2.

E1.7.6 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 E1.7.2 and table E1.7.1.

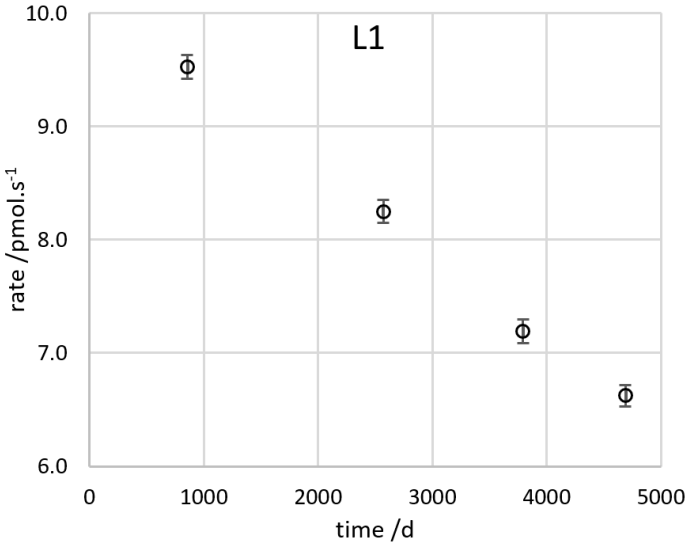


Figure E1.7.2: Calibration results for a reference leak L_1. Data points represent the reference value with error bars corresponding to ±1 standard uncertainty

Table E1.7.1: 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:

$$Q = a_1 + b_1 t, \tag{E1.7.1}$$

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 [77] 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 of [77, clause 7] could instead be followed. This approach is also relatively straightforward to implement in a spreadsheet calculation.

The results are found to be:

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

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:

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

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

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

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 (E1.7.1).

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

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

E1.7.7 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.

E1.7.7.1 Measurement model

In this situation (following example in Annex D of [77]) 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:

$$Q_j = Q_{oj} + e_\text{s}, \tag{E1.7.2}$$

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 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

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

which when combined give the result

$$\begin{aligned} u^2(Q_j) &= u^2(Q_{oj}) + u^2(e_s), \\ &= u^2(e_{1j}) + u^2(e_{2j}) + u^2(r_j) + u^2(s). \end{aligned} \tag{E1.7.3}$$

To illustrate this scenario, consider the calibration results in table E1.7.2.

Table E1.7.2: Uncertainty contributions for reference leak L_1

t/d	Q_j/pmol s^{-1}	$u(Q_{oj})$/pmol s^{-1}	$u(e_s)$/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:

$$Q = a + bt, \tag{E1.7.4}$$

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 [77] is adopted to establish values for the coefficients.

Firstly though, a covariance matrix $\boldsymbol{V_Q}$ 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-S2, 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-S2, clause 6.2)

$$\boldsymbol{Y} = \boldsymbol{Q} = \boldsymbol{f}(\boldsymbol{X}),$$

where

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

The covariance matrix $\boldsymbol{V_Q}$ is given by

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

where

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

and

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

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 $\boldsymbol{V_Q}$ are instead given by

$$u(Q_j, Q_k) = \sum_{i=1}^{N} \sum_{\ell=1}^{N} \frac{\partial f_j}{\partial x_i} \frac{\partial f_k}{\partial x_\ell} u(x_i, x_\ell). \tag{E1.7.5}$$

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

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

and

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

E1.7.7.2 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

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

In this example we have

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

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

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

$$\boldsymbol{R}_y = \begin{bmatrix} 1 & 0.279 & 0.279 & 0.306 \\ & 1 & 0.286 & 0.313 \\ & & 1 & 0.313 \\ \text{sym.} & & & 1 \end{bmatrix}. \tag{E1.7.6}$$

The results of the fitting are

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

ISO/TS 28037 [77] 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 [77, Annex F] describes software and source code that is provided free by NPL.

The standard uncertainty $u(Q)$ for a forward evaluation using equation (E1.7.4) is evaluated by a standard GUM-LPU approach. This can be expressed in matrix format as in ISO/TS 28037, clause 11.2:

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

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

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

which equates to

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

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

For example, a forward evaluation using equation (E1.7.4) 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:

$$Q_j = Q_{oj}, \tag{E1.7.7}$$

$$u^2(Q_j) = u^2(Q_{oj}) = u^2(e_{1j}) + u^2(e_{2j}) + u^2(c_j), \tag{E1.7.8}$$

which corresponds to the data model in Case 1. For the data in table E1.7.3, the process described in clause 6 of [77] can be used to calculate the fitting results in this case, giving the following:

$$\begin{aligned}
a &= 10.185\,\mathrm{pmol\,s^{-1}} &&= a_1,\\
b &= -7.678\times10^{-4}\,\mathrm{pmol\,s^{-1}\,d^{-1}} &&= b_1,\\
u(a) &= 0.119\,\mathrm{pmol\,s^{-1}} &&= u(a_1),\\
u(b) &= 3.506\times10^{-5}\,\mathrm{pmol\,s^{-1}\,d^{-1}} &&= u(b_1),\\
u(a,b) &= -3.785\times10^{-6}\,\mathrm{pmol^2s^{-2}d^{-1}} &&= u(a_1,b_1),
\end{aligned} \tag{E1.7.9}$$

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

Table E1.7.3: Calibration results for reference leak L_1

t/d	Q_j/pmol s^{-1}	$u(Q_j)$/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 E1.7.8), consider a second leak L_2 for which the data in table E1.7.4 is available, where the data model for L_2 is as described above for L_1 in equations (E1.7.2) and (E1.7.3).

Table E1.7.4: Data for reference leak L_2

t/d	Q_j/pmol s^{-1}	$u(Q_{oj})$/pmol s^{-1}	$u(e_s)$/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

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

with

$$\boldsymbol{R}_y = \begin{bmatrix} 1 & 0.594 & 0.599 & 0.615 & 0.620 & 0.636 \\ & 1 & 0.604 & 0.621 & 0.626 & 0.642 \\ & & 1 & 0.626 & 0.632 & 0.648 \\ & & & 1 & 0.648 & 0.665 \\ & & & & 1 & 0.671 \\ \text{sym.} & & & & & 1 \end{bmatrix}. \tag{E1.7.10}$$

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 (E1.7.6). Such a statement could not easily be made by examining covariance matrices.

Forward evaluation using equation (E1.7.4), again at $t = 5000\,\text{d}$ with $u(t) = 1\,\text{d}$ gives the estimate $Q = 3.284\,\text{pmol}\,\text{s}^{-1}$ and standard uncertainty $u(Q) = 0.063\,\text{pmol}\,\text{s}^{-1}$.

The corresponding results of fitting a straight line can again be evaluated for the situation where there is no correlation, as detailed in equations (E1.7.7) and (E1.7.8). In this case the data is given in table E1.7.5.

Table E1.7.5: Calibration results for reference leak L_2

t/d	Q_j/pmol s^{-1}	$u(Q_j)$/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 [77] can again be used to calculate the fitting results in this case, giving

$$\begin{aligned} a_2 &= 4.380\,\text{pmol}\,\text{s}^{-1}, \\ b_2 &= -2.200 \times 10^{-4}\,\text{pmol}\,\text{s}^{-1}\,\text{d}^{-1}, \\ u(a_2) &= 0.045\,\text{pmol}\,\text{s}^{-1}, \\ u(b_2) &= 1.620 \times 10^{-5}\,\text{pmol}\,\text{s}^{-1}\,\text{d}^{-1}, \\ u(a_2, b_2) &= -5.713 \times 10^{-7}\,\text{pmol}^2\text{s}^{-2}\text{d}^{-1}. \end{aligned} \tag{E1.7.11}$$

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

E1.7.8 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.

E1.7.8.1 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.

E1.7.8.2 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:

$$Q_1 = a_1 + b_1 t, \qquad Q_2 = a_2 + b_2 t. \tag{E1.7.12}$$

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:

$$\begin{aligned} q_1 &= Q_1[1 + \alpha(\Delta T_1 + \delta T)] = (a_1 + b_1 t)[1 + \alpha(\Delta T_1 + \delta T)], \\ q_2 &= Q_2[1 + \alpha(\Delta T_2 + \delta T)] = (a_2 + b_2 t)[1 + \alpha(\Delta T_2 + \delta T)], \end{aligned} \tag{E1.7.13}$$

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-6 [5, 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

$$q = a_M + b_M p, \tag{E1.7.14}$$

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

$$q_x = a_M + b_M p_x \tag{E1.7.15}$$

and

$$Q_x = \frac{q_x}{[1 + \alpha_x(\Delta T_x + \delta T)]}, \tag{E1.7.16}$$

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.

E1.7.8.3 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 [77]. 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 (E1.7.9) and (E1.7.11), will be used.

Measured values

Since the quantities t, α and δT are common to both leak rate expressions [equations (E1.7.13)], 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-S2 [4] 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 (E1.7.13) 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-S2 [4] clause 6.2, the quantities are

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

The covariance matrix $\boldsymbol{V}_y$ is given by

$$\boldsymbol{V}_y = \boldsymbol{C}_x \boldsymbol{V}_x \boldsymbol{C}_x^\top = \begin{bmatrix} u^2(q_1) & u(q_1, q_2) \\ u(q_2, q_2) & u^2(q_2) \end{bmatrix},$$

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

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

and

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

Alternatively, the covariance matrix can be calculated in terms of subscripted summations in line with annex F.1.2 of the GUM [2], albeit using equation (E1.7.5) 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),

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

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

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 [77]) correspond to

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

Solving the model establishes estimates for the coefficients a_M and b_M [for equation (E1.7.14)] and the elements of the covariance matrix

$$
\boldsymbol{V}_{a,b} = \begin{bmatrix} u^2(a_\mathrm{M}) & u(a_\mathrm{M}, b_\mathrm{M}) \\ u(a_\mathrm{M}, b_\mathrm{M}) & u^2(b_\mathrm{M}) \end{bmatrix}.
$$

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 $\boldsymbol{V}$ such as $\boldsymbol{V}_{q_1,q_2}$ established in Case 3 above, this involves calculating the components of a lower left matrix $\boldsymbol{L}$ such that

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

and

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

which is satisfied when

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

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 (E1.7.15):

$$q_x = a_M + b_M p_x,$$

and again follows the process described in clause 11 of ISO/TS 28037 [77], giving in matrix form

$$u^2(q_x) = \boldsymbol{C}_x^\top \boldsymbol{V}_x \boldsymbol{C}_x,$$

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

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

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 (E1.7.15)

Calculation of reference value

The reference value Q_x calculated using (E1.7.16), 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

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

where

$$\boldsymbol{C}_{\boldsymbol{Q}_\mathrm{x}} = \begin{bmatrix} C_{q_\mathrm{x}} \\ C_{\alpha_\mathrm{x}} \\ C_{\Delta T_\mathrm{x}} \\ C_{\delta T} \end{bmatrix} = \begin{bmatrix} \frac{Q_\mathrm{x}}{q_\mathrm{x}}, \\ -\frac{Q_\mathrm{x}^2}{q_\mathrm{x}}(\Delta T_\mathrm{x} + \delta T), \\ -\frac{Q_\mathrm{x}^2}{q_\mathrm{x}}\alpha, \\ -\frac{Q_\mathrm{x}^2}{q_\mathrm{x}}\alpha, \end{bmatrix}; \qquad \boldsymbol{V}_{\boldsymbol{Q}_\mathrm{x}} = \begin{bmatrix} u^2(q_\mathrm{x}) & 0 & 0 & 0 \\ 0 & u^2(\alpha_\mathrm{x}) & 0 & 0 \\ 0 & 0 & u^2(\Delta T_\mathrm{x}) & 0 \\ 0 & 0 & 0 & u^2(\delta T) \end{bmatrix},$$

which is equivalent to

$$u^2(Q_\mathrm{x}) = C_{q_\mathrm{x}}^2 u^2(q_\mathrm{x}) + C_{\alpha_\mathrm{x}}^2 u^2(\alpha) + C_{\Delta T_\mathrm{x}}^2 u^2(\Delta T_\mathrm{x}) + C_{\delta T}^2 u^2(\delta T),$$

as is found by applying GUM equation (13) to equation (E1.7.16)

E1.7.8.4 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 E1.7.3 and tables E1.7.6 and E1.7.7.

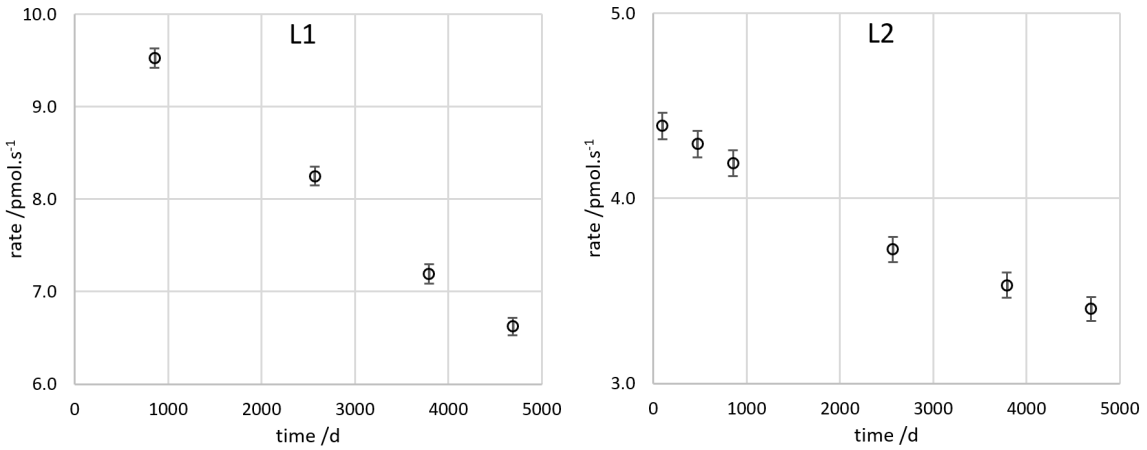


Figure E1.7.3: Calibration data for the two reference leaks. Data points represent the reference value with error bars corresponding to ±1 standard uncertainty

Table E1.7.6: Calibrated reference values for reference leak L_1

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

Table E1.7.7: Calibrated reference values for reference leak L_2

t/d	Q_j/pmol s^{-1}	$u(Q_j)$/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 [77]. The results of these operations are

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

and

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

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 (E1.7.12) above are found to be:

$$\begin{aligned} Q_1 &= 6.346\,\text{pmol}\,\text{s}^{-1}, \qquad u(Q_1) = 0.084\,\text{pmol}\,\text{s}^{-1}, \\ Q_2 &= 3.280\,\text{pmol}\,\text{s}^{-1}, \qquad u(Q_2) = 0.054\,\text{pmol}\,\text{s}^{-1} \end{aligned}$$

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

Measured values

Suppose that the leak is to be used at $t = 5000\,\text{d}$, $u(t) = 1\,\text{d}$ and the calibration conditions are as in table E1.7.8 for which it is calculated [equation (E1.7.13)] that

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

Table E1.7.8: 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

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

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 (E1.7.14) are then found to be

$$\begin{aligned} a_\text{M} &= 1.722\,\text{pmol}\,\text{s}^{-1}, \\ b_\text{M} &= 0.034\,\text{pmol}\,\text{s}^{-1}\,\text{du}^{-1}, \\ u(a_\text{M}) &= 0.139\,\text{pmol}\,\text{s}^{-1}, \\ u(b_\text{M}) &= 0.0017\,\text{pmol}\,\text{s}^{-1}\,\text{du}^{-1}, \\ u(a_\text{M}, b_\text{M}) &= -1.701 \times 10^{-4}\,\text{pmol}^2\,\text{s}^{-2}\,\text{du}^{-1}. \end{aligned}$$

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 $q_\text{x} = 5.754\,\text{pmol}\,\text{s}^{-1}$ and an associated standard uncertainty of $u(q_\text{x}) = 0.135\,\text{pmol}\,\text{s}^{-1}$.

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 E1.7.9.

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

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

Table E1.7.9: 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}

E1.7.9 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 [2] 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.

E1.7.10 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 (E1.7.10) with those in the matrix (E1.7.6). 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 E1.7.3, 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 [110] 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 E1.8

Factoring effects such as calibration corrections and drift into uncertainty evaluations

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

E1.8.1 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.

E1.8.2 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.

E1.8.3 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

$$Z_{\text{uncor}} = Y \tag{E1.8.1}$$

for an uncorrected model as in section E1.8.4. For a corrected model it is

$$Z_{\text{cor}} = Y + X, \tag{E1.8.2}$$

where X is the quantity regarded as a correction.

E1.8.4 Measurement model

A 'best estimate' of Y is the arithmetic mean

$$y = \frac{1}{n}\sum_{i=1}^{n} y_i, \qquad i = 1, \ldots, n, \tag{E1.8.3}$$

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

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

$$z_{\text{uncor}} = y. \tag{E1.8.4}$$

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

$$z_{\text{cor}} = y + x. \tag{E1.8.5}$$

x is the correction for a known bias or systematic error in the measuring system. A further term can be included in model (E1.8.4) or (E1.8.5) relating to the resolution of the measuring instrument. Its inclusion is straightforward and is treated by Lira and Wöger [152]. 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.

E1.8.5 Uncertainty propagation

Knowledge concerning Y is in terms of a set of repeated observations made under repeatability conditions (section E1.8.4). The arithmetic mean y of the observations is taken in (E1.8.3) as an estimate of Y. The associated standard uncertainty $u(y)$ is given by [2, 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 (E1.8.4) and (E1.8.5), y can be considered as a realized value of a random variable with that variable typically modelled by the normal distribution $\mathrm{N}(y, u^2(y))$, which is strictly valid only for large n. In cases where n is small, the t-distribution should be used [2, annex G.3].

E1.8.5.1 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 (E1.8.5), 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 (E1.8.5) is given by

$$u^2(z_{\text{cor}}) = u^2(y) + u^2(x). \tag{E1.8.6}$$

There is a correspondence between the right-hand sides of expressions (E1.8.5) and (E1.8.6) 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 (E1.8.6) can be regarded as giving the squared standard uncertainty associated with the corrected measured value z_{cor} in (E1.8.5). Said another way, expression (E1.8.5) is a realisation of measurement model (E1.8.2) and expression (E1.8.6) 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 $u(z_{\text{cor}})$. When n is large and x has a normal distribution, the distribution relating to z_{cor} can be taken as normal: $\mathrm{N}(z_{\text{cor}}, u^2(z_{\text{cor}}))$. If that is not the case, a Monte Carlo method [3] can be used to obtain its probability distribution given probability distributions for Y (for example, Student's t) and X (for example, rectangular).

E1.8.5.2 Case 2: Poor practice

Common practice [152–156] 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

$$z_{\text{uncor}} = y \tag{E1.8.7}$$

A variety of approaches have been adopted for evaluating the associated uncertainty [154]. For the purposes of demonstration let us suppose that the associated 'standard uncertainty' $u(z_{\text{uncor}})$ is given by one of the commonly used approaches, for which it is assumed that

$$u^2(z_{\text{uncor}}) = u^2(y) + u^2(x) + x^2/3. \tag{E1.8.8}$$

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

'Standard uncertainty' is given in quotation marks since, as stated in [157], it is *not* a standard uncertainty as defined in the GUM [2]. There is no one-to-one correspondence between the terms in expressions (E1.8.7) and (E1.8.8) as in the expressions for the corrected value (E1.8.5) and the associated variance (squared standard uncertainty) (E1.8.6). It does not possess the properties of internal consistency, transferability and universality: see GUM [2, clause 0.4] and the strong comments in [157].

The use of expression (E1.8.8) to obtain a 'standard uncertainty' for an uncorrected value constitutes *poor practice*.

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

Not correcting for a systematic effect can be serious. Molinar et al. [158] 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 $u \approx 0.2\,\text{mK}$ 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. [159] 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.'

E1.8.6 Reporting the result

E1.8.6.1 Impact on tests against specification or tolerance limits

General

Conformance probability [6, 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 [6, 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 constructing a coverage interval with a desired coverage probability given the probability distribution for Z is contained in JCGM 101:2008 [3].

Using the recommended practice in section E1.8.5.1, when g is normal, resulting from Y and X being normal, the integral can straightforwardly be computed. Otherwise, a Monte Carlo calculation [3] 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 E1.8.5.2 can be demonstrated by example. Suppose the primary length of a product is tested using an appropriate measuring instrument and there is an upper tolerance limit $T_{\text{U}} = 100\,\text{cm}$ on the length. Suppose the

known correction and its associated standard uncertainty are

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

A measured length value and its associated standard uncertainty are

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

Both x and y have Gaussian distributions.

Using expressions (E1.8.5) and (E1.8.6), the corrected value and associated standard uncertainty are

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

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

On the other hand, working with uncorrected values, expressions (E1.8.7) and (E1.8.8) 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 $T_{\text{U}} = 100\,\text{cm}$ 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 E1.8.1 depicts these two situations.

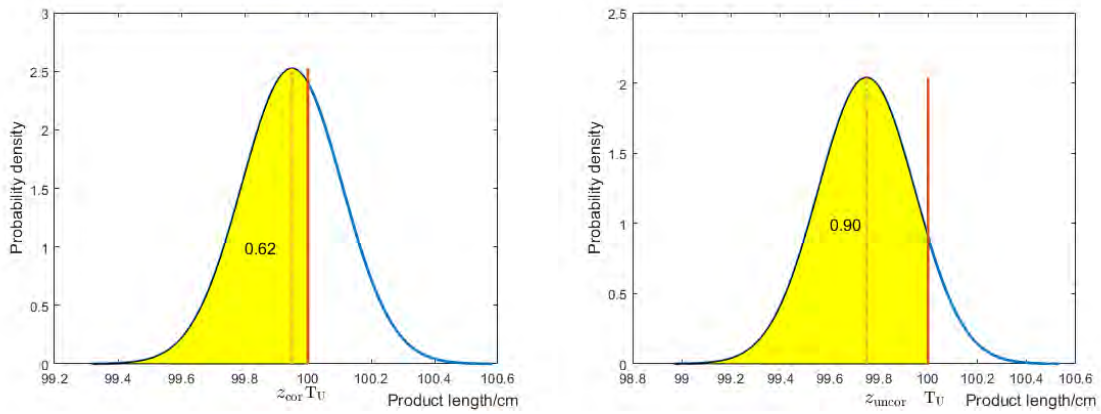


Figure E1.8.1: Using a corrected value, the conformance probability (0.62) is the shaded area in the left figure, whereas using the uncorrected value, the conformance probability (0.90) is the shaded area in the right figure; the latter (poor practice) approach allows a greater proportion of non-conforming items to be accepted

Figure E1.8.2 shows the conformance probability for a range of values y.

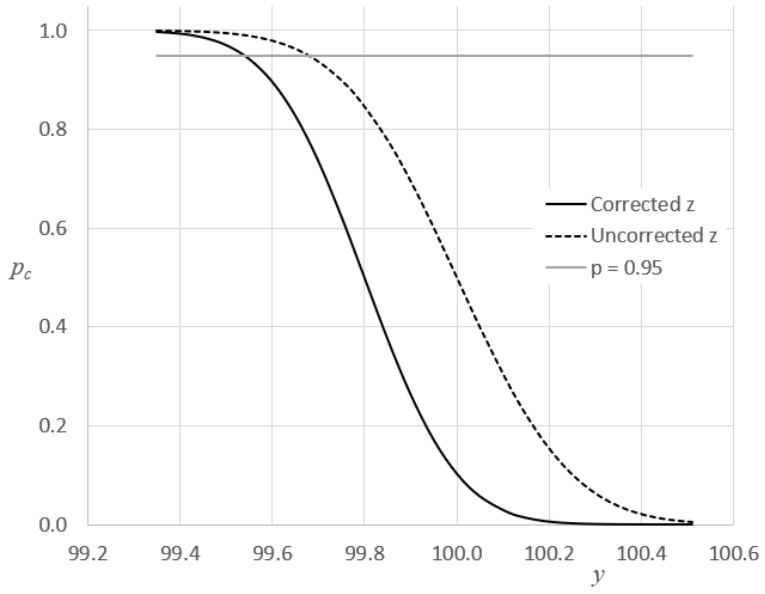


Figure E1.8.2: Conformance probability for a range of measured values y with corrected and uncorrected bias in the example described in section E1.8.6.1

E1.8.7 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 among the 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 [77] and [110]). The drift could be modelled in terms of a time dependent relationship for the fitted coefficients, perhaps resulting in a linear function

$$Z(t) = a(t) + b(t)Y, \tag{E1.8.9}$$

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

$$Z = Y + C. \tag{E1.8.10}$$

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

$$c(t) = c_n + d(t), \tag{E1.8.11}$$

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 (E1.8.11). 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 [160] 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

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

E1.8.7.1 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 (E1.8.10) 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 E1.8.1. Let $u(c_i) = 0.15$ and the evaluation be conducted for some later time $t > t_n$, say $t = 42$.

Table E1.8.1: 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 E1.8.3 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 [77]; hence

$$\begin{aligned} z(t) &= y + c(t), \\ u^2(z) &= u^2(y) + u^2(c_t). \end{aligned}$$

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

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

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.

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

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

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

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.)

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

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 [2]).

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 E1.8.3 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.

E1.8.8 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 [161], '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

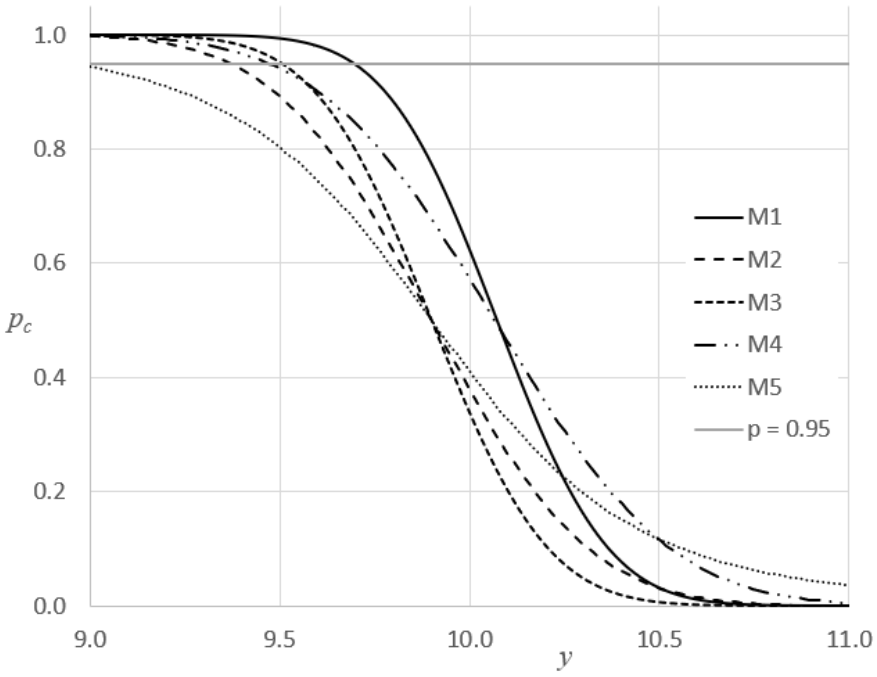


Figure E1.8.3: conformance probability for a range of observed values y, evaluated at $t = 42$ using the various models, M1 to M5 and the data in Table E1.8.1

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 E1.8.6.1 and in section E1.8.7.1.

Unfortunately, however (as explained in section E1.8.5.2), there is no way to state the uncertainty associated with an uncorrected value that is consistent with the GUM [2].

On the basis of the explanation in section E1.8.5.2 and supported by these simulations it is strongly recommended that whenever possible a corrected value and the associated uncertainty is reported as in section E1.8.5.1.

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 [154] 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 E1.8.6.1 and in section E1.8.7.1.

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.

Part E2

Conformity assessment

Example E2.1

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

E2.1.1 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.

E2.1.2 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.

An IUPAC Project [162] 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 [6] for conformity assessment of a single item. For this reason, the notation used here is consistent with the notation in [6] and [163], the latter being a relevant paper in which the reader can find more details on this case study.

E2.1.3 Specification of the measurands

This case study concerns test results for NyQuil tablets [164], a cold medication containing four active components:

- acetaminophen (APAP), as a pain reliever and fever reducer;
- dextromethorphan hydrobromide (DEX), as a cough suppressant;
- 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) $c_{i\mathrm{m}}$, obtained according to the test method described in [163], 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).

E2.1.4 Measurement uncertainty and correlations

A full uncertainty budget for the test results of the components' content is available in [163]. Relative measurement uncertainty is evaluated as 2.8 % of $c_{i\mathrm{m}}$.

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} [2, sec. C.3.6], $i < j$. Only APAP test results are not significantly correlated with the other components' 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) [163]. Correlation coefficients are reported in table E2.1.1.

Table E2.1.1: Correlation coefficients between components' content values

Component	Index i/j	APAP 1	DEX 2	DOX 3	PE 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

E2.1.5 Specification or tolerance limits

The lower and upper tolerance limits, $T_{\mathrm{L}i}$ and $T_{\mathrm{U}i}$, 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.

E2.1.6 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 [6, sec. 8.2.1], i.e., acceptance limits of test results coincide with tolerance limits ($A_{\mathrm{L}i} = T_{\mathrm{L}i}$ and $A_{\mathrm{U}i} = T_{\mathrm{U}i}$).

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.

E2.1.6.1 Bayesian framework

In the framework of the IUPAC project [162], 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 [6, 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 $h(\boldsymbol{c}_\mathrm{m}\,|\boldsymbol{c}\,)$ for the corresponding test results, where $\boldsymbol{c}_\mathrm{m} = [c_{1\mathrm{m}}, \ldots, c_{4\mathrm{m}}]$.

As discussed in [163], 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 E2.1.2), 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 E2.1.1 and s_i is the ith experimental standard deviation (see table E2.1.2). For each fixed vector value $\boldsymbol{c}$, the multivariate normal PDF modelling the likelihood function $h(\boldsymbol{c}_\mathrm{m}\,|\boldsymbol{c}\,)$ has vector mean $\boldsymbol{c}$ and covariance matrix $\boldsymbol{S}_{c\mathrm{m}}$ made by terms $S_{\mathrm{cm}_{ij}} = r_{ij}\, u_i\, u_j$, where $u_i = 0.028\, c_{i\mathrm{m}}$, % 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 E2.1.2: Experimental mean and standard deviation of the components' content values of 105 lots of the medication

Component	Index	Mean	Standard deviation
	i	m_i, % of labelled amount	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

E2.1.6.2 Total specific risk

For a vector of test results $\boldsymbol{c}_\mathrm{m}$ of a specific multicomponent item, when all the $c_{i\mathrm{m}}$ 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' contents being outside its tolerance interval. Therefore, it is calculated as

[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).

one minus the probability that all the "true" values c_i are inside their tolerance interval. Such a probability is provided by the posterior PDF 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 means $\boldsymbol{c}_{post}$ respectively equal to [65, eq. 3.13]:

$$\boldsymbol{S}_{post} = \left(\boldsymbol{S}_c^{-1} + n_{rep}\boldsymbol{S}_{cm}^{-1}\right)^{-1} \tag{E2.1.1}$$

$$\boldsymbol{c}_{post} = \boldsymbol{S}_{post}\left(\boldsymbol{S}_c^{-1}\boldsymbol{m} + n_{rep}\boldsymbol{S}_{cm}^{-1}\bar{\boldsymbol{c}}_m\right), \tag{E2.1.2}$$

where $\boldsymbol{m}$ is the vector of the prior mean values, $\bar{\boldsymbol{c}}_m$ 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, $n_{rep} = 1$ and $\bar{\boldsymbol{c}}_m = [c_{1m}, \ldots, c_{4m}]$).

Considering, for example, the special case in which all the test results c_{im} are exactly equal to the corresponding prior mean values m_i, the total specific consumer's risk is $R^*_{tot} = 0.0029$. When $c_{im} = T_{Li}$ for each i, hence $R^*_{tot} = 0.0117$; when $c_{im} = T_{Ui}$, $R^*_{tot} = 0.0002$. 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" [51] is used for the calculation of the posterior cumulative distribution.

E2.1.6.3 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 $g_0(\boldsymbol{c})\, h(\boldsymbol{c}_m|\boldsymbol{c})$. 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 corresponding vector $\boldsymbol{c}_m$ is drawn from $h(\boldsymbol{c}_m|\boldsymbol{c})$. Hence, the total risk is approximated by the frequency of cases in which, within randomly generated vectors $[\boldsymbol{c}_m, \boldsymbol{c}]$, 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 $R_{tot} = 0.0018$, 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 [163] ($R_{tot} = 0.0019$), 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".

E2.1.7 Interpretation of results

The above-reported values of total specific risk are for illustrative purposes. Value $R^*_{tot} = 0.0029$ 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

[2] The absolute error of the reported values, provided as an output of the function, is about 10^{-5} for $R^*_{tot} = 0.0029$ and for $R^*_{tot} = 0.0117$, and 10^{-6} for $R^*_{tot} = 0.0002$.

"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 $c_{i\mathrm{m}}$ values are fixed and equal to the prior mean values) is depicted in [163], showing that the risk behaviour is not easily predictable.

The obtained total global risk $R_{\mathrm{tot}} = 0.0018$ 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).

Example E2.2

Measurement models involving additive or multiplicative corrections

A. Bošnjaković, V. Karahodžić, J. Greenwood, M.G. Cox

E2.2.1 Summary

A common form of presentation for calibration results involves expressing the result as an additive or multiplicative correction. This is the case for vacuum gauges and is illustrated with data using the models described in [165]. The examples and conclusions do, however, have much wider applicability.

This example demonstrates the effect of model assumptions concerning errors in the reference value. In addition it demonstrates how conformance probability can be affected by these assumptions. The example concludes by demonstrating how correlation can be handled for calibration corrections.

E2.2.2 Introduction

Calibration measurements are reported in a wide variety of forms. A particularly popular form involves presenting measurement error as a calibration correction.

Often a limit or tolerance is defined for this correction and a conformity test is required. To make such a conformity decision requires knowledge of the measurement uncertainty associated with the correction.

Measurement uncertainty plays a crucial role, both here and in the decision processes found in most activities concerned with product or process conformity assessment. Without some account for measurement uncertainty the risk associated with a decision is *undefinable*.

The evaluation of measurement uncertainty and conformance probability are illustrated here for the calibration of a vacuum gauge; however, the analysis and methods described have more general applicability.

The calibration of vacuum (pressure) gauges is achieved by using a reference standard to establish the calibration pressure value at the inlet port of the unit under calibration (UUC). Often [166], this reference pressure is measured directly by a reference gauge and is obtained from

a corrected reading from that gauge. A model for the evaluation of measurement uncertainty in vacuum calibration is described in some detail in ISO 27893 [165], whose principles can readily be transferred to other measurement applications.

Using the same set of calibration data, this example considers an additive ('sum') model, in which the calibration pressure value can be used to determine a reading error Δp for the UUC; it considers a multiplicative 'quotient model' (applicable when the calibration value is used to determine a correction factor, sensitivity coefficient, accommodation coefficient or gauge constant), and a 'combined model'.

Three scenarios are considered, representing different practices, illustrating how these practices can affect the associated conformance probability.

In the first scenario, following best metrological practice, a reference pressure correction (that is, a known systematic bias, due for example to the calibration method, thermal transpiration, height correction, etc.) is applied, and its associated uncertainty is incorporated in the uncertainty evaluation.

In the second scenario, the reference pressure correction is not applied, and instead it is combined with the associated uncertainty to establish a larger 'correction uncertainty'. In metrological terms this way of working represents poor (albeit common) practice when it is adopted for a *known* bias, and can have significant consequences for conformity decisions [167]. In situations where the correction is *not* known, but is perhaps considered to be in a defined range then this approach is more justified [168].

In the final scenario, the correction and its associated uncertainty are simply neglected, representing what might be termed 'bad practice'.

In each case the output PDF is assumed to be normal and conformance probability is calculated using a standard normal cumulative distribution function.

Data for the analyses are taken from a calibration certificate IMT-LMT-80-2019, produced by the Laboratory of Pressure Metrology, Institute of Metals and Technology, Ljubljana, Slovenia. The UUC in this example is a capacitance diaphragm vacuum gauge with full scale range of 11 kPa.

The specification adopted for the UUC in our example requires that calibration errors should be no larger than 0.5 % of the reading.

Finally, a physically different 'sum model with correlation' is presented, demonstrating how correlation might be treated in that case.

E2.2.3 Measurands

Adopting the nomenclature of [165] the measurand is defined for the various classes of model as one of the following:

Δp – pressure difference (sum model) having standard uncertainty $u(\Delta p)$,

f – correction error (quotient model) having standard uncertainty $u(f)$,

e – error of reading (combined model) having standard uncertainty $u(e)$.

Other useful nomenclature:

$p_{\mathrm{UUC}}, u(p_{\mathrm{UUC}})$ – pressure for unit under calibration and its associated standard uncertainty,

$p_{\text{std}}, u(p_{\text{std}})$ – reference standard pressure and its associated standard uncertainty,

$\Delta p_{\text{m}}, u(\Delta p_{\text{m}})$ – reference pressure correction term and its associated standard uncertainty.

E2.2.4 Measurement model

The measurement models are of the explicit, univariate type [4]:

In such models, a single real output quantity Y is related to a number of input quantities $\boldsymbol{X} = (X_1, \ldots, X_N)$ by a functional relationship f in the form $Y = \boldsymbol{f}(\boldsymbol{X})$ as stated in the GUM [2]. The estimate of the output quantity is taken as $y = \boldsymbol{f}(\boldsymbol{x})$. The standard uncertainty $u(y)$ is associated with y is evaluated from

$$u^2(y) = \sum_{i=1}^{N} \sum_{j=1}^{N} c_i u(x_{\text{i}}, x_j) c_j,$$

where c_i is the partial derivative $\partial f / \partial X_i$ evaluated at $\boldsymbol{X} = \boldsymbol{x}$ and is known as the ith 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. For independent input quantities, we would obtain the better-known simplified expression

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

where

$$u_i(y) = |c_i| u(x_i).$$

E2.2.4.1 Sum model

In the sum model, the measurand Δp is defined as the difference between the reading of the unit under calibration (UUC) and the reference value, which is given by the pressure indication of the reference standard corrected by a (possible) reference pressure correction term Δp_{m}:

$$\Delta p = p_{\text{UUC}} - (p_{\text{std}} + \Delta p_{\text{m}}). \tag{E2.2.1}$$

E2.2.4.2 Quotient model

The standard ISO 27893 [165] describes how a model can be established in general situations where the UUC output and the reference standard output are not necessarily given in the same units of measurement. For example the UUC output may be measured as a current, voltage or frequency that is relatable to pressure through some functional relationship.

$$r_{\text{UUC}} = \frac{x_{\text{UUC}}}{p_{\text{std}}}.$$

In this example we are only concerned with a simple case in which $x_{\text{UUC}} = p_{\text{UUC}}$ and where $r_{\text{UUC}} = 1/f$ defines a correction factor f as the measurand; hence

$$f = \frac{p_{\text{std}} + \Delta p_{\text{m}}}{p_{\text{UUC}}}, \tag{E2.2.2}$$

Note that equation (E2.2.2) is not a pure quotient but is a simple example of a 'combined' model.

E2.2.4.3 Combined model

In practice, realistic measurement models are seldom a pure sum or product of quantities and a combined model is required. For example, the measurand e, the relative error of reading, can be defined by

$$e = \frac{p_{\mathrm{UUC}} - (p_{\mathrm{std}} + \Delta p_{\mathrm{m}})}{p_{\mathrm{std}} + \Delta p_{\mathrm{m}}} = \frac{p_{\mathrm{UUC}}}{p_{\mathrm{std}} + \Delta p_{\mathrm{m}}} - 1. \tag{E2.2.3}$$

E2.2.4.4 Sum model with correlation

An example that demonstrates how to evaluate a sum model with correlation is presented in section E2.2.7.

E2.2.5 Uncertainty propagation

The GUM's law of propagation of uncertainty (LPU) [2, eqn. (10)] is applied to establish the standard uncertainty associated with an estimate of the measurand for each of the three measurement models.

E2.2.5.1 Sum model

The standard uncertainty in the sum model is

$$u(\Delta p) = \left[u^2(p_{\mathrm{UUC}}) + u^2(p_{\mathrm{std}}) + u^2(\Delta p_{\mathrm{m}})\right]^{1/2}. \tag{E2.2.4}$$

E2.2.5.2 Quotient model

The standard uncertainty in the quotient model is

$$u(f) = \frac{p_{\mathrm{std}} + \Delta p_{\mathrm{m}}}{p_{\mathrm{UUC}}} \left[\frac{u^2(p_{\mathrm{UUC}})}{p_{\mathrm{UUC}}^2} + \frac{u^2(p_{\mathrm{std}})}{(p_{\mathrm{std}} + \Delta p_{\mathrm{m}})^2} + \frac{u^2(\Delta p_{\mathrm{m}})}{(p_{\mathrm{std}} + \Delta p_{\mathrm{m}})^2}\right]^{1/2}. \tag{E2.2.5}$$

E2.2.5.3 Combined model

The standard uncertainty in the combined model is

$$u(e) = \frac{p_{\mathrm{UUC}}}{p_{\mathrm{std}} + \Delta p_{\mathrm{m}}} \left[\frac{u^2(p_{\mathrm{UUC}})}{p_{\mathrm{UUC}}^2} + \frac{u^2(p_{\mathrm{std}})}{(p_{\mathrm{std}} + \Delta p_{\mathrm{m}})^2} + \frac{u^2(\Delta p_{\mathrm{m}})}{(p_{\mathrm{std}} + \Delta p_{\mathrm{m}})^2}\right]^{1/2}. \tag{E2.2.6}$$

E2.2.6 Measurand expanded uncertainty and conformance probability for three scenarios

In this section the measurand expanded uncertainty at the 95 % level of confidence ($k = 2$) and the conformance probability are evaluated for three different scenarios, described below.

For the purposes of these examples the reference pressure correction term Δp_m is taken to be 0.05 % of p_{std} and the associated expanded uncertainty ($k = 2$) is assumed to be 1 Pa for all pressure values.

The calibration data, calibration corrections and associated expanded measurement uncertainties are summarised in Table E2.2.1. This data applies to all scenarios and models.

Table E2.2.1: Calibration data used for all scenarios and models (data is given and used as provided rather than in line with common reporting principles [169]).

Point, n	p_{std}/Pa	$U(p_{std})$/Pa	p_{UUC}/Pa	$U(p_{UUC})$/Pa
1	10.89	0.050	10.7	0.23
2	17.02	0.090	17.0	0.23
3	26.03	0.13	25.4	0.23
4	40.28	0.20	39.5	0.23
5	63.10	0.32	63.0	0.23
6	96.89	0.48	97.1	0.23
7	161.3	1.0	161.2	0.23
8	256.4	1.5	255.9	0.23
9	403.3	2.4	403.7	0.23
10	647.5	3.9	647.8	0.23
11	978.7	5.9	980.4	0.23
12	1610.6	5.0	1613.2	0.23
13	2505.4	5.0	2509.2	0.23
14	4075.9	0.50	4079.5	0.23
15	6278.6	0.70	6282.7	0.23
16	9069.2	1.0	9072.9	0.23
17	10 932.8	1.1	10 937.1	0.23

E2.2.6.1 Scenario 1

In the first scenario, following best metrological practice, the reference pressure correction Δp_m is applied to the measured reference pressure p_{std}.

The calibration corrections and expanded measurement uncertainty are summarised in Table E2.2.2 where the corrections Δp, f and e are evaluated using equations (E2.2.1), (E2.2.2) and (E2.2.3) respectively and standard uncertainties are evaluated using the corresponding equations (E2.2.4), (E2.2.5) and (E2.2.6).

Figures E2.2.1 and E2.2.2, respectively, depict the correction factor f and the pressure difference Δp as a function of UUC pressure indication.

The red broken lines on these and later figures represent the specification limits defined by the UUC owner. These often, but not necessarily, correspond to limits defined by the equipment manufacturer.

Conformance probability can be calculated regarding Δp, f and e. Since all three quantities are linearly related their PDFs describe the same physical situation and the same conformance probability will be established whichever is evaluated. We therefore arbitrarily choose f for the purpose of these examples.

Table E2.2.2: Calibration corrections for Scenario 1. All uncertainties are expanded ($k = 2$)

Point	Δp/Pa	$U(\Delta p)$/Pa	e	$U(e)$	f	$U(f)$
1	−0.20	1.0	−0.0179	0.093	1.0183	0.096
2	−0.03	1.0	−0.0017	0.061	1.0017	0.061
3	−0.64	1.0	−0.0247	0.039	1.0253	0.041
4	−0.80	1.1	−0.0199	0.0025	1.0203	0.026
5	−0.13	1.1	−0.0021	0.0017	1.0021	0.017
6	0.16	1.1	0.0017	0.0011	0.9983	0.011
7	−0.18	1.4	−0.0011	0.0088	1.0011	0.0089
8	−0.63	1.8	−0.0024	0.0071	1.0025	0.0071
9	0.20	2.6	0.0005	0.0065	0.9995	0.0065
10	−0.02	4.0	0.0000	0.0062	1.0000	0.0062
11	1.21	6.0	0.0012	0.0061	0.9988	0.0061
12	1.79	5.1	0.0011	0.0032	0.9989	0.0032
13	2.55	5.1	0.0010	0.0020	0.9990	0.0020
14	1.56	1.1	0.0004	0.000 28	0.9996	0.000 28
15	0.96	1.2	0.0002	0.000 20	0.9998	0.000 20
16	−0.83	1.4	−0.0001	0.000 16	1.0001	0.000 16
17	−1.17	1.5	−0.0001	0.000 14	1.0001	0.000 14

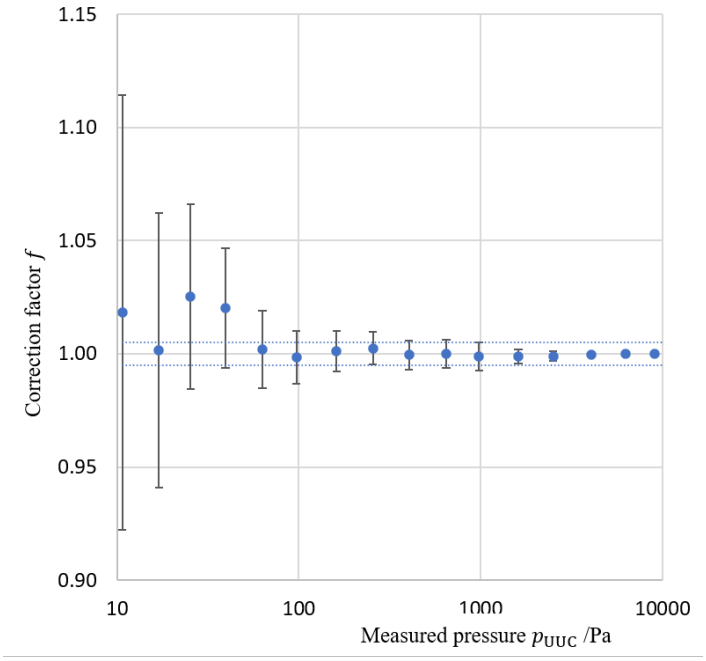


Figure E2.2.1: Correction factor as a function of UUC pressure indication (logarithmic scale), scenario 1

To calculate the conformance probability we further assume that the PDF for f is normal:

$$p(x) = \frac{1}{\sigma(2\pi)^{1/2}} e^{-[(x-m)^2/(2\sigma^2)]},$$

where m is the mean and σ is the standard deviation.

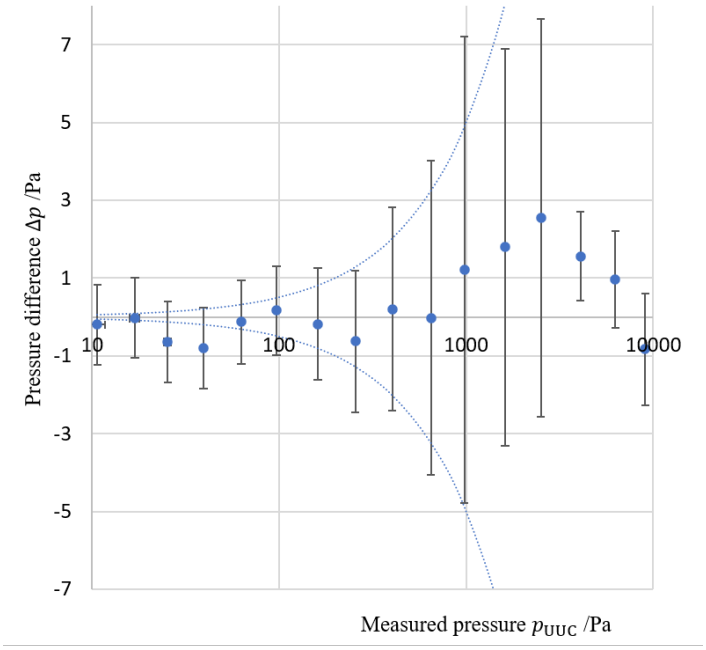


Figure E2.2.2: Pressure difference as a function of UUC pressure indication, scenario 1

To demonstrate evaluation of the conformance probability, consider a point 13 where say $m = f = 0.999$ and $\sigma = u(f) = 0.002$. The conformance probability is given by integrating the PDF over the limits of interest, say $0.995 \le x \le 1.005$:

$$p_c = \int_{0.995}^{1.005} \frac{1}{0.002\,(2\pi)^{1/2}}\, e^{-[(x-0.999)^2/(2\times 0.002^2)]}\,\mathrm{d}x \qquad \text{(E2.2.7)}$$

This integral is not analytically solvable, but it can be expressed through tabulated functions such as Q, ϕ or erfc. In this example the use of the Q function will be demonstrated where

$$Q(x) = \int_x^{+\infty} \frac{1}{2\pi^{1/2}} e^{-t^2/2}\,\mathrm{d}t.$$

Letting $t = (x-m)/\sigma$, we find new limits for the integral

$$(1.005 - 0.999)/0.002 = 3,$$
$$(0.995 - 0.999)/0.002 = -2,$$

and (E2.2.7) becomes

$$\begin{aligned} p_c &= \int_{-2}^{3} \frac{1}{2\pi^{1/2}} e^{-t^2/2}\,\mathrm{d}t \\ &= \int_{-2}^{+\infty} \frac{1}{2\pi^{1/2}} e^{-t^2/2}\mathrm{d}t - \int_{3}^{+\infty} \frac{1}{2\pi^{1/2}} e^{-t^2/2}\,\mathrm{d}t \\ &= Q(-2) - Q(3) \\ &= 0.977 - 0.001 \\ &= 97.6\,\%. \end{aligned}$$

Note: To implement the calculations in Microsoft Excel, the $Q(x)$ function can be evaluated using the relation $Q(x) = 0.5\,\mathrm{erfc}(x/\sqrt{2})$ and the Excel function ERFC.PRECISE().

Table E2.2.3: Conformance probability – Scenario 1

Point	p_{UUC}/Pa	Δp/Pa	e	f	p_{c}/%
7	161.30	−0.18	−0.001 12	1.0011	72.4
8	256.40	−0.63	−0.002 45	1.0025	74.5
9	403.30	0.20	0.000 492	0.9995	87.3
10	647.50	−0.02	−0.000 037	1.0000	89.2
11	978.70	1.21	0.001 236	0.9988	87.1
12	1610.60	1.79	0.001 114	0.9989	99.3
13	2505.40	2.55	0.001 016	0.9990	100.0
14	4075.90	1.56	0.000 383	0.9996	100.0
15	6278.60	0.96	0.000 153	0.9998	100.0
16	9069.20	−0.83	−0.000 092	1.0001	100.0

The conformance probability for this scenario using the given data and the stated specification is summarised in Table E2.2.3 (restricted to data in the top two full decades for sake of clarity).

No *general* conclusions should be drawn from the values in this table. The results are however informative for this specific calibration where, as might be expected for a well behaved instrument of this type, conformance probability tends to be highest at higher pressures. The acceptability or otherwise of the result can be based on a straightforward consideration of conformance probability.

E2.2.6.2 Scenario 2

In the second scenario no reference pressure correction Δp_{m} is applied, that is, the model equations are

$$\begin{aligned} f &= \frac{p_{\mathrm{std}}}{p_{\mathrm{UUC}}}, \\ \Delta p &= p_{\mathrm{UUC}} - p_{\mathrm{std}}, \\ e &= \frac{p_{\mathrm{UUC}}}{p_{\mathrm{std}}} - 1. \end{aligned}$$

This situation might (correctly) arise because little is known about Δp_{m}; hence the best estimate of its value is taken to be zero, albeit the uncertainty remains finite. It might also be the case that the value of the correction *is* known but following common (albeit poor) practice it is instead somehow combined with its uncertainty to establish a bigger uncertainty estimate which, it is argued, accounts for the failure to apply the correction. (See [167] and [157] for explanations of why this is considered to be poor practice.) This latter situation is evaluated in this scenario.

In this case, the standard uncertainty $u(\Delta p_{\mathrm{m}})$ associated with the unused correction term is calculated using

$$u(\Delta p_{\mathrm{m}}) = \left[0.5^2 + \frac{(0.05\,\% \, p_{\mathrm{std}})^2}{3} \right]^{1/2}$$

in which $\Delta p_{\mathrm{m}} = 0.05\,\% \, p_{\mathrm{std}}$ is assumed to be the semi-range of a rectangular distribution.

The quantities Δp, f and e and their uncertainties are again recalculated and given in Table E2.2.4

Figures E2.2.3 and E2.2.4, respectively, depict the correction factor f and the pressure difference Δp as a function of UUC pressure indication.

Table E2.2.4: Data for scenario 2. All uncertainties are expanded ($k = 2$)

Point	Δp/Pa	$U(\Delta p)$/Pa	e	$U(e)$	f	$U(f)$
1	−0.19	1.0	−0.0174	0.093	1.0178	0.096
2	−0.02	1.0	−0.0012	0.061	1.0012	0.061
3	−0.63	1.0	−0.0242	0.039	1.0248	0.041
4	−0.78	1.1	−0.0194	0.026	1.0197	0.027
5	−0.10	1.1	−0.0016	0.017	1.0016	0.017
6	0.21	1.1	0.0022	0.012	0.9978	0.012
7	−0.10	1.4	−0.0006	0.0089	1.0006	0.0089
8	−0.50	1.8	−0.0020	0.0071	1.0020	0.0071
9	0.40	2.6	0.0010	0.0065	0.9990	0.0065
10	0.30	4.0	0.0005	0.0062	0.9995	0.0062
11	1.70	6.0	0.0017	0.0061	0.9983	0.0061
12	2.60	5.1	0.0016	0.0032	0.9984	0.0032
13	3.80	5.3	0.0015	0.0021	0.9985	0.0021
14	3.60	2.6	0.0009	0.000 64	0.9991	0.000 64
15	4.10	3.8	0.0007	0.000 61	0.9993	0.000 61
16	3.70	5.4	0.0004	0.000 60	0.9996	0.000 60
17	4.30	6.5	0.0004	0.000 59	0.9996	0.000 59

As would be expected from equations (E2.2.1), (E2.2.2) and (E2.2.3) the uncertainty for scenario 2 is always larger than the corresponding uncertainty for scenario 1. In this case the difference is not large, but this is dictated by the data and may be more (or less) significant for other data.

The conformance probability for this scenario using the given data and the stated specification is summarised in Table E2.2.5 (restricted to data in the top two full decades for sake of clarity).

Table E2.2.5: Conformance probability – scenario 2

Point	p_{UUC}/Pa	Δp/Pa	e	f	p_{c}/%
7	161.30	−0.10	−0.0006	1.0006	73.4
8	256.40	−0.50	−0.0020	1.0020	77.8
9	403.30	0.40	0.0010	0.9990	85.9
10	647.50	0.30	0.0005	0.9995	88.6
11	978.70	1.70	0.0017	0.9983	84.2
12	1610.60	2.60	0.0016	0.9984	98.3
13	2505.40	3.80	0.0015	0.9985	100.0
14	4075.90	3.60	0.0009	0.9991	100.0
15	6278.60	4.10	0.0007	0.9993	100.0
16	9069.20	3.70	0.0004	0.9996	100.0

The behaviour for conformance probability is generally more complex than is the case for measurement uncertainty. In this example the conformance probability for scenario 2 is generally lower than for scenario 1, which can be explained by one or both of the measurand (based upon uncorrected reference pressure) being closer to a tolerance limit, and the uncertainty being larger, and hence the PDF being 'wider' and extending more beyond the tolerance limits. It is however quite possible for a value to be closer to the centre of the tolerance interval when no correction is applied, which may have a larger influence on the calculation of conformance probability than the increase in uncertainty, as is seen for example for our point no. 8.

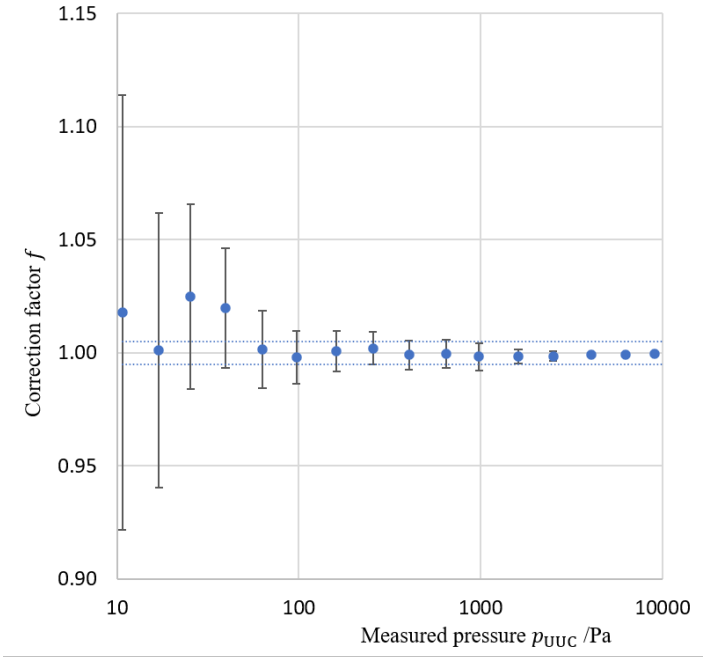


Figure E2.2.3: Correction factor as a function of UUC pressure indication (logarithmic scale), scenario 2

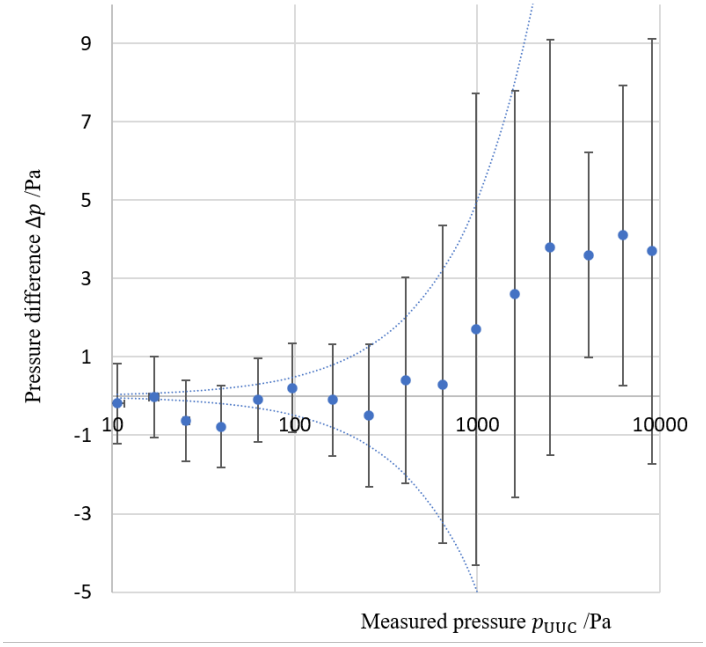


Figure E2.2.4: Pressure difference as a function of UUC pressure indication, scenario 2

E2.2.6.3 Scenario 3

The third scenario is obtained when the correction term and associated uncertainty are excluded, which is equivalent to setting $\Delta p_m = u(\Delta p_m) = 0$ in our three models, yielding the results shown in Table E2.2.6 in which the conformance probability p_c is again given for all points in the second and third decade.

Table E2.2.6: Data for scenario 3. All uncertainties are expanded ($k = 2$)

Point	Δp/Pa	$U(\Delta p)$/Pa	e	$U(e)$	f	$U(f)$
1	−0.19	0.24	−0.0174	0.022	1.0178	0.022
2	−0.02	0.25	−0.0012	0.015	1.0012	0.015
3	−0.63	0.26	−0.0242	0.010	1.0248	0.011
4	−0.78	0.30	−0.0194	0.0075	1.0197	0.0078
5	−0.10	0.39	−0.0016	0.0062	1.0016	0.0063
6	0.21	0.53	0.0022	0.0055	0.9978	0.0055
7	−0.10	1.0	−0.0006	0.0064	1.0006	0.0064
8	−0.50	1.5	−0.0020	0.0059	1.0020	0.0059
9	0.40	2.4	0.0010	0.0060	0.9990	0.0060
10	0.30	3.9	0.0005	0.0060	0.9995	0.0060
11	1.70	5.9	0.0017	0.0060	0.9983	0.0060
12	2.60	5.0	0.0016	0.0031	0.9984	0.0031
13	3.80	5.0	0.0015	0.0020	0.9985	0.0020
14	3.60	0.55	0.0009	0.000 14	0.9991	0.000 13
15	4.10	0.74	0.0007	0.000 12	0.9993	0.000 12
16	3.70	1.0	0.0004	0.000 11	0.9996	0.000 11
17	4.30	1.1	0.0004	0.000 10	0.9996	0.000 10

Figures E2.2.5 and E2.2.6, respectively, depict the correction factor f and the pressure difference Δp as a function of UUC pressure indication.

The conformance probability for this scenario using the given data and the stated specification is summarised in Table E2.2.7 (restricted to data in the top two full decades for sake of clarity).

Table E2.2.7: Conformance probability – scenario 3

Point	p_{UUC}/Pa	Δp/Pa	e	f	p_{c}/%
7	161.30	−0.10	−0.0006	1.0006	87.7
8	256.40	−0.50	−0.0020	1.0020	83.8
9	403.30	0.40	0.0010	0.9990	88.8
10	647.50	0.30	0.0005	0.9995	89.9
11	978.70	1.70	0.0017	0.9983	84.8
12	1610.60	2.60	0.0016	0.9984	98.6
13	2505.40	3.80	0.0015	0.9985	100.0
14	4075.90	3.60	0.0009	0.9991	100.0
15	6278.60	4.10	0.0007	0.9993	100.0
16	9069.20	3.70	0.0004	0.9996	100.0

As was the case for scenario 2, the behaviour for conformance probability is complex. In *this* case the conformance probability is generally higher when compared to scenario 1. The difference is entirely due to the nature of the data and the unused reference pressure correction and will vary depending upon the data and corrections in question. Conformity decisions based upon these (scenario 3) conformance probabilities would therefore likely be unreliable.

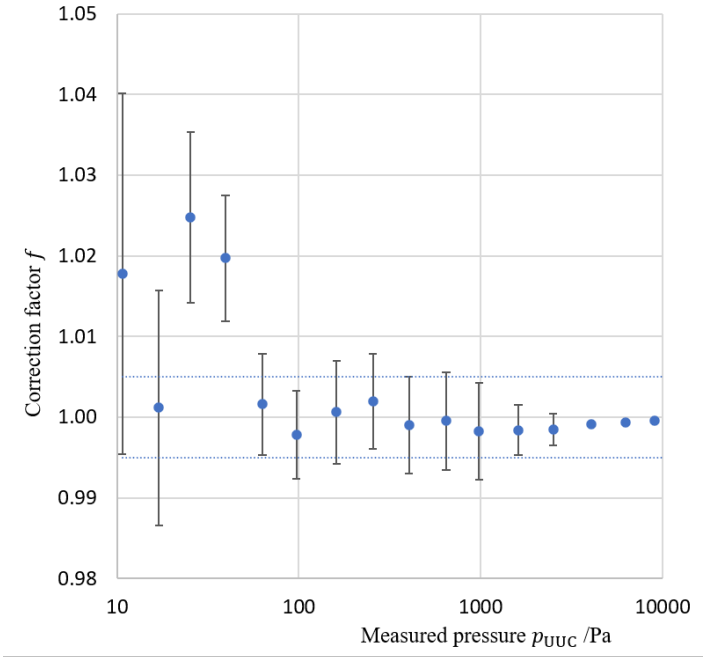


Figure E2.2.5: Correction factor as a function of UUC pressure indication (logarithmic scale), scenario 3

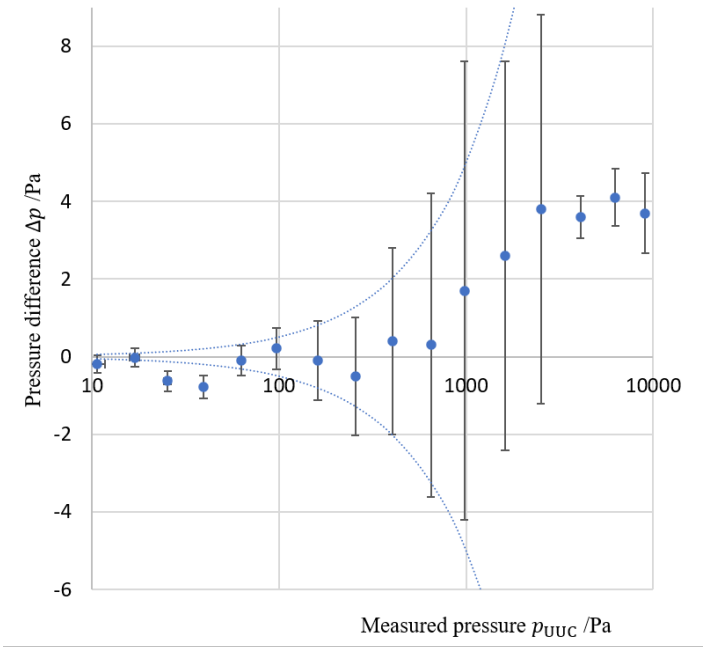


Figure E2.2.6: Pressure difference as a function of UUC pressure indication, scenario 3

E2.2.7 Measurement model: sum model with correlation

Suppose that by some mechanism not already accounted for, both p_{std} and p_{UUC} are both dependent upon another common quantity, say for example a potential systematic error ΔT in measuring gas temperature due to placement of temperature probes. This situation might be

modelled as:

$$\begin{aligned} p_{\text{std}} &= p'_{\text{std}}(1 + \alpha_{\text{std}}\ \Delta T), \\ p_{\text{UUC}} &= p'_{\text{UUC}}(1 + \alpha_{\text{UUC}}\ \Delta T), \end{aligned} \tag{E2.2.8}$$

where, for example, p'_{std} and p'_{UUC} now represent the observed values and α_{std} and α_{UUC} are corresponding temperature coefficients. Suppose also that a reliable estimate of ΔT is not available; hence the 'best' estimate is taken to be $\Delta T = 0\,\text{K}$ with an associated standard uncertainty $u(\Delta T)$.

For this example we will also assume that the reference pressure correction Δp_{m} is fully independent and is not affected by the possible temperature error; hence the measurands are to be calculated using equations (E2.2.1), (E2.2.2) and (E2.2.3).

The standard uncertainties associated with the quantities p_{std} and p_{UUC} in equation (E2.2.8) are therefore

$$\begin{aligned} u^2(p_{\text{std}}) &= u^2(p'_{\text{std}})(1 + \alpha_{\text{std}}\ \Delta T)^2 + u^2(\alpha_{\text{std}})(p'_{\text{std}}\Delta T)^2 + u^2(\Delta T)(p'_{\text{std}}\alpha_{\text{std}})^2, \\ u^2(p_{\text{UUC}}) &= u^2(p'_{\text{UUC}})(1 + \alpha_{\text{UUC}}\ \Delta T)^2 + u^2(\alpha_{\text{UUC}})(p'_{\text{UUC}}\Delta T)^2 + u^2(\Delta T)(p'_{\text{UUC}}\alpha_{\text{UUC}})^2. \end{aligned}$$

To illustrate the situation, let $u(\Delta T) = 0.57\,\text{K}$, $\alpha_{\text{std}} = \alpha_{\text{UUC}} = 1/300\,\text{K}^{-1}$ and $u(\alpha_{\text{std}}) = u(\alpha_{\text{UUC}}) = 1/3000\,\text{K}^{-1}$; hence, for example, at calibration point 8, $p'_{\text{std}} = 256.4\,\text{Pa}$ and $p'_{\text{UUC}} = 255.9\,\text{Pa}$. We find from equation (E2.2.1) that $\Delta p = -0.63\,\text{Pa}$ and

$$\begin{aligned} u(p_{\text{std}}) &= 0.90\,\text{Pa}, \\ u(p_{\text{UUC}}) &= 0.51\,\text{Pa}. \end{aligned}$$

Combining these *without* taking account of the correlation between p_{std} and p_{UUC} gives, using equation (E2.2.4), a standard uncertainty of $u(\Delta p) = 1.14\,\text{Pa}$.

In this case the estimate is about 25 % larger than is obtained by taking account of the correlation, as can be achieved by following the process described in matrix form in clause 6.2 of GUM-S2 [4] (demonstrated below) and in subscripted summation form in the GUM [2, Annex F.1.2.3].

Equations (E2.2.1), (E2.2.2) and (E2.2.3) are real univariate measurement functions of the form $Y = f(\boldsymbol{X})$ where in the case of equation (E2.2.1) we have

$$\begin{aligned} Y &= \Delta p, \\ \boldsymbol{X} &= (p_{\text{std}}, p_{\text{UUC}}, \Delta p_{\text{m}})^{\top}. \end{aligned}$$

Applying the LPU, the variance in this case is given by

$$u^2(\Delta p) = \boldsymbol{V}_{\boldsymbol{\Delta p}} = \boldsymbol{C}_{\boldsymbol{\Delta p}}^{\top} \boldsymbol{V}_{\boldsymbol{in}} \boldsymbol{C}_{\boldsymbol{\Delta p}}, \tag{E2.2.9}$$

where $\boldsymbol{C}_{\boldsymbol{\Delta p}}$ is an array containing sensitivity coefficients, and $\boldsymbol{V}_{\boldsymbol{in}}$ is the corresponding covariance matrix for the input quantities:

$$\boldsymbol{C}_{\boldsymbol{\Delta p}} = \begin{bmatrix} \frac{\partial \Delta p}{\partial p_{\text{std}}} \\ \frac{\partial \Delta p}{\partial p_{\text{UUC}}} \\ \frac{\partial \Delta p}{\partial \Delta p_{\text{m}}} \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \\ -1 \end{bmatrix},$$

$$\boldsymbol{V}_{\mathbf{in}} = \begin{bmatrix} u^2 p_{\mathrm{std}}) & u(p_{\mathrm{std}}, p_{\mathrm{UUC}}) & u(p_{\mathrm{std}}, \Delta p_{\mathrm{m}}) \\ u(p_{\mathrm{UUC}}, p_{\mathrm{std}}) & u^2(p_{\mathrm{UUC}}) & u(p_{\mathrm{UUC}}, \Delta p_{\mathrm{m}}) \\ u(\Delta p_{\mathrm{m}}, p_{\mathrm{std}}) & u(\Delta p_{\mathrm{m}}, p_{\mathrm{UUC}}) & u^2(\Delta p_{\mathrm{m}}) \end{bmatrix}. \tag{E2.2.10}$$

Equation (E2.2.9) is the matrix representation of the equation obtained by applying GUM equation (13).

The covariance matrix (E2.2.10) is obtained from

$$\boldsymbol{V}_{\mathbf{in}} = \boldsymbol{C}_X \boldsymbol{V}_X \boldsymbol{C}_X^{\mathsf{T}},$$

where, in this example we have

$$\boldsymbol{C}_X^{\mathsf{T}} = \begin{bmatrix} \frac{\partial p_{\mathrm{std}}}{\partial p'_{\mathrm{std}}} & \frac{\partial p_{\mathrm{UUC}}}{\partial p'_{\mathrm{std}}} & \frac{\partial \Delta p_{\mathrm{m}}}{\partial p'_{\mathrm{std}}} \\ \frac{\partial p_{\mathrm{std}}}{\partial \alpha_{\mathrm{std}}} & \frac{\partial p_{\mathrm{UUC}}}{\partial \alpha_{\mathrm{std}}} & \frac{\partial \Delta p_{\mathrm{m}}}{\partial \alpha_{\mathrm{std}}} \\ \frac{\partial p_{\mathrm{std}}}{\partial p'_{\mathrm{UUC}}} & \frac{\partial p_{\mathrm{UUC}}}{\partial p'_{\mathrm{UUC}}} & \frac{\partial \Delta p_{\mathrm{m}}}{\partial p'_{\mathrm{UUC}}} \\ \frac{\partial p_{\mathrm{std}}}{\partial \alpha_{\mathrm{UUC}}} & \frac{\partial p_{\mathrm{UUC}}}{\partial \alpha_{\mathrm{UUC}}} & \frac{\partial \Delta p_{\mathrm{m}}}{\partial \alpha_{\mathrm{UUC}}} \\ \frac{\partial p_{\mathrm{std}}}{\partial \Delta T} & \frac{\partial p_{\mathrm{UUC}}}{\partial \Delta T} & \frac{\partial \Delta p_{\mathrm{m}}}{\partial \Delta T} \\ \frac{\partial p_{\mathrm{std}}}{\partial \Delta p_{\mathrm{m}}} & \frac{\partial p_{\mathrm{UUC}}}{\partial \Delta p_{\mathrm{m}}} & \frac{\partial \Delta p_{\mathrm{m}}}{\partial \Delta p_{\mathrm{m}}} \end{bmatrix} = \begin{bmatrix} 1 + \alpha_{\mathrm{std}}\, \Delta T & 0 & 0 \\ p'_{\mathrm{std}}\, \Delta T & 0 & 0 \\ 0 & 1 + \alpha_{\mathrm{UUC}}\, \Delta T & 0 \\ 0 & p'_{\mathrm{UUC}}\, \Delta T & 0 \\ p'_{\mathrm{std}}\, \alpha_{\mathrm{std}} & p'_{\mathrm{UUC}}\, \alpha_{\mathrm{UUC}} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

and

$$\boldsymbol{V}_X = \begin{bmatrix} u^2(p'_{\mathrm{std}}) & 0 & 0 & 0 & 0 & 0 \\ 0 & u^2(\alpha_{\mathrm{std}}) & 0 & 0 & 0 & 0 \\ 0 & 0 & u^2(p'_{\mathrm{UUC}}) & 0 & 0 & 0 \\ 0 & 0 & 0 & u^2(\alpha_{\mathrm{UUC}}) & 0 & 0 \\ 0 & 0 & 0 & 0 & u^2(\Delta T) & 0 \\ 0 & 0 & 0 & 0 & 0 & u^2(\Delta p_{\mathrm{m}}) \end{bmatrix}.$$

Alternatively, the elements of (E2.2.10) can be calculated in terms of subscripted summations in line with annex F.1.2 of the GUM [2] using GUM equations (F.1) and (F.2). As we have assumed that the reference pressure correction Δp_{m} is fully independent and is not affected by the possible temperature error, off-diagonal covariances involving Δp_{m} are zero. The remaining covariance $u(p_{\mathrm{std}}, p_{\mathrm{UUC}}) = u(p_{\mathrm{UUC}}, p_{\mathrm{std}})$ is non-zero since both p_{std} and p_{UUC} depend upon ΔT. Its value is given by

$$u(p_{\mathrm{std}}, p_{\mathrm{UUC}}) = (p'_{\mathrm{std}}\, \alpha_{\mathrm{std}})(p'_{\mathrm{UUC}}\, \alpha_{\mathrm{UUC}}) u^2(\Delta T).$$

Evaluating the uncertainty by either equation (E2.2.9) or GUM equation (13) yields a value of $u(\Delta p) = 0.91\,\mathrm{Pa}$ and a conformance probability of $p_{\mathrm{c}} = 0.75$ rather than a probability of $p_{\mathrm{c}} = 0.67$ that is obtained when correlation is neglected. The impact of such a difference in conformance probability is dependent upon the particular application of interest, but clearly any such differences have the potential to affect conformity decisions.

E2.2.8 Interpretation of results

This example has demonstrated the evaluation of measurement uncertainty and conformance probability for various related calibration models under several common scenarios. It has also demonstrated how correlation between quantities (arising from dependency on a common effect) is treated within the GUF. In each case the consequences for conformity decisions are complex, depending as they do upon the particular data, model, and assumptions. No general rule can easily be established. For some data points the conformance probability decreases when simplifying assumptions are made, in others it increases. Caution is therefore needed unless best practice (scenario 1) is followed to avoid the risk of making poor decisions.

Acknowledgement

We would like to express our gratitude to the Laboratory of Pressure Metrology, Institute of Metals and Technology, Ljubljana, Slovenia for providing us with data.

Example E2.3

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

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

E2.3.1 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 [6], 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.

E2.3.2 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 [170]. 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 [171] and in [172].

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 [173], but are out of the scope of the present example.

E2.3.3 Specification of the measurand

For characterization of TSPM, the EPA method IO-2.1 [170] 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 m^3 to 2400 m^3, 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.

E2.3.4 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 [171] it was demonstrated, by means of 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 [17]), and depicted in figure E2.3.1.

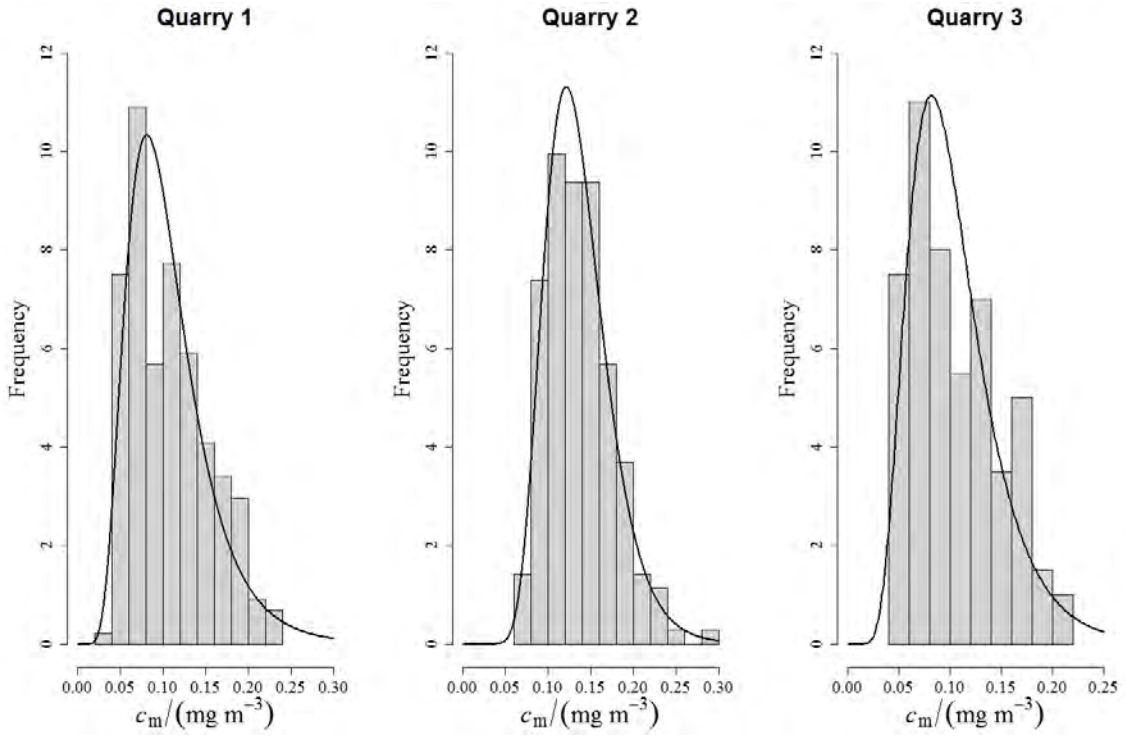


Figure E2.3.1: Histograms of the measured TSPM mass concentration values for each quarry and corresponding lognormal probability density functions smoothing the data.

A full uncertainty budget for the considered test results is available in [171], 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.

E2.3.5 Tolerance limits

The Israeli national regulations of ambient air quality prescribe an upper tolerance (regulation) limit $T_{\mathrm{U}} = 0.2\,\mathrm{mg\,m^{-3}}$ 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, $T_{\mathrm{U}i} = 0.2\,\mathrm{mg\,m^{-3}}$, for $i = 1, 2, 3$.

E2.3.6 Decision rule and conformity assessment

Regulations require direct comparison of measured values $c_{i\mathrm{m}}$ with $T_{\mathrm{U}i}$. In the present example, acceptance limits $A_{\mathrm{U}i}$ 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, $A_{\mathrm{U}i} = T_{\mathrm{U}i}$), a "shared risk" rule is considered as the decision rule for conformity assessment [6, 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 [6, 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.

E2.3.6.1 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 [6, 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 [171]. 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:

$$g_0(c_i) = \frac{1}{c_i \sigma_i \sqrt{2\pi}} \exp\left[-\frac{(\ln(c_i) - \mu_i)^2}{2\sigma_i^2} \right], \tag{E2.3.1}$$

whose distributional parameters are reported (on the logarithmic scale) in table E2.3.1. 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 E2.3.1.

Table E2.3.1: 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 $c_{i\mathrm{m}}$ 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 $u_i = 0.07c_{i\mathrm{m}}$ [171]. The corresponding likelihood for each quarry is hence a normal distribution:

$$h(c_{i\mathrm{m}}|c_i) = \frac{1}{u_i\sqrt{2\pi}} \exp\left[-\frac{(c_{i\mathrm{m}} - c_i)^2}{2u_i^2}\right]. \tag{E2.3.2}$$

When both the prior PDF and the likelihood are normal distributions, the posterior PDF [6, Eq. (1)] is also normal [6, Sec. 7.2.1][1]. In such a case, the evaluation of specific and global risks is straightforward, as detailed in [6]. 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.

E2.3.6.2 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 (E2.3.1) and the likelihood (E2.3.2), according to [6, 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 [17]), where the `R`-function `dlnorm` [11] was used for evaluating the density of the considered lognormal distributions, whose logarithms have the mean and the standard deviation, reported in table E2.3.1 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 E2.3.2, for A_{U} values varying in the interval $[T_{\mathrm{U}} - 0.05, T_{\mathrm{U}} + 0.05]$ mg m^{-3}. Considering, for example, the special case in which $A_{\mathrm{U}} = T_{\mathrm{U}}$, 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

[1] If the prior information is meagre and the likelihood function is characterised by a normal distribution, then the posterior PDF is approximately normal [6, Sec. 7.2.2].

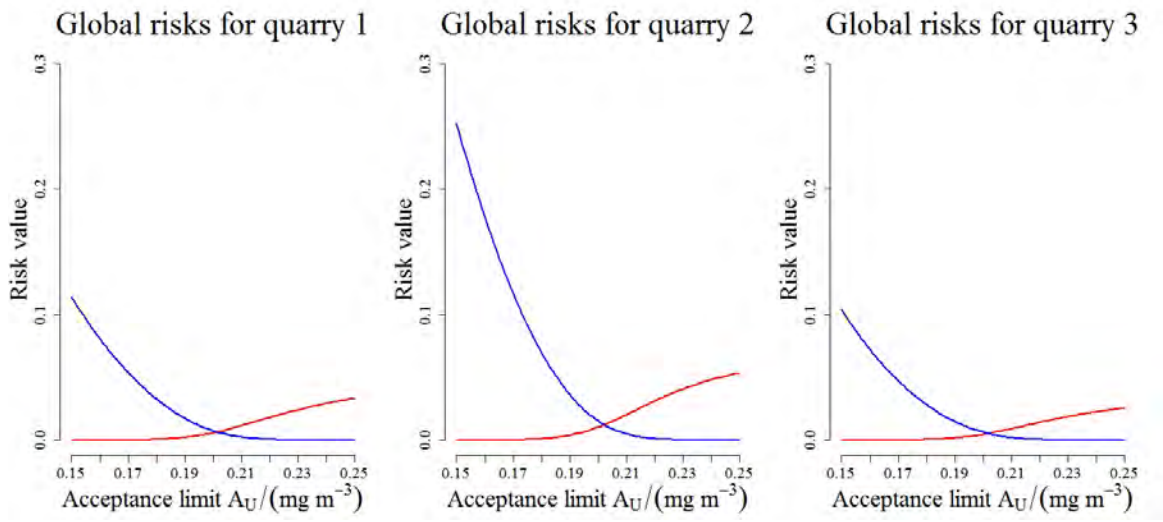


Figure E2.3.2: Consumer's (red line) and producer's (blue line) global risks versus acceptance limit values.

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 %.

E2.3.6.3 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 [6, 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 $h(c_i|c_{im})$ [6, 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 [17]), 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 E2.3.3 (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.

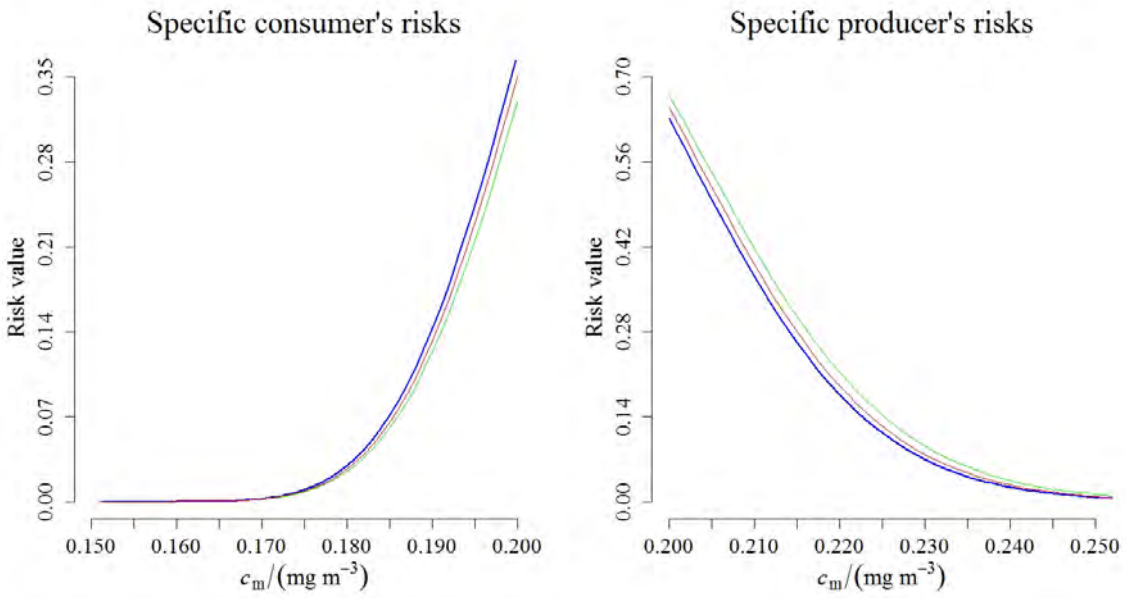


Figure E2.3.3: Consumer's and producer's specific risks versus test results, for quarry 1 (blue line), 2 (green line) and 3 (red line).

These results have been validated against those from CASoft [174], 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.

E2.3.7 Interpretation of results

Studies on global risks, such as that conducted in section E2.3.6.2, 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 E2.3.6.3 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 $c_{1m} = 0.225\,\text{mg m}^{-3}$, 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 E2.4

Uncertainty evaluation of nanoparticle size by AFM, by means of an optimised Design of Experiment for a hierarchical mixed model in a Bayesian framework approach

T. Caebergs, B. de Boeck, J. Pétry, N. Sebaïhi, M.G. Cox, N. Fischer, J. Greenwood

E2.4.1 Summary

This example presents a comprehensive framework for uncertainty evaluation for measurement of the mean size of nanoparticles in dispersion samples by Atomic Force Microscopy (AFM). No comprehensive measurement model exists for this measurement and a statistical model is built up for measurement uncertainty evaluation. Random effects and fixed effects are simultaneously considered and interactions cannot be neglected, a priori. This is realised by a Design of Experiment (DoE) for a hierarchical model|indexhierarchical model, and following a Bayesian approach. An optimised DoE is used instead of a full DoE to speed up the acquisition of data by reducing the number of AFM images. A calibration curve obtained by measurements of step-height of certified topography standards provides the traceability of the measurements to the metre definition of the Système International d'unités (SI).

This example is intended to be adaptable as a template example for commercial calibration service application: balancing the need for limited manpower while keeping the uncertainty evaluation to the best accuracy. It is based on a scientific paper already published [175], giving here more importance to the explanation of the methodology. It has also been successfully applied to ISO/IEC 17025 [7] accredited activities in Belgium.

E2.4.2 Introduction of the application

The properties of a chemical substance are usually associated with its bulk form. For example, gold is a yellow metal as ingots or in jewellery (solid gold or coating), being a good electrical conductor. The properties of a material can drastically change when the size of the functional element is of the order of 1 nm to 100 nm – the nanometre scale. Size is thus the first-line measurement in the physico-chemical characterisation of the product.

Gold is yellow in its bulk form and red in its nanoparticle form. Upon agglomeration of the nanoparticles, some blue tints additionally appear. Other than just colour, the functional properties are also different. For example, silicon dioxide acts as an anti-caking agent for powders (whether food or not). By reducing the physical size of items in a sample of the elements while keeping the same mass (quantity of matter of bulk material), the contact surface increases. For chemical reagents or catalysts, kinetics therefore increase, leading to explosive behaviour in extreme cases.

The discovery of the creation of these properties has led to recent technological development and new industrial applications have rapidly been discovered. The market has already experienced a boom in recent times. Unfortunately, adverse effects on health are not yet known and understood in detail. Following ideas of precautionary principles, several EU regulations have been established [176–178]. Two examples are to be cited: regulation on labelling of consumer products [177] and the more recent regulation on medical devices [179]. In addition, some countries have set up registers to monitor their nano-material market [180–182]. The common point of these regulations is the sizing, the EU has agreed on a formal definition [183]. These reasons motivate the work on a metrological approach for nanoparticle sizing: through traceability to the SI and rigorous uncertainty evaluation. The field of metrology at the nanoscale is nowadays facing challenges. Numerous techniques have been developed to size nanoparticles but they provide different measured sizes, with results sometimes incompatible between techniques. A deeper metrological understanding is thus needed.

This example presents an instance of uncertainty evaluation when no full measurement model is available: a statistical modelling (top-down) approach is thus followed. This application is related to example [184], which deals with one source of uncertainty in detail: the pixel size of the raster image. In the following example, this source of uncertainty is kept under control (that is, giving negligible uncertainty) by appropriate choice for the purpose of this study, in order to avoid unnecessary complication of the problem.

AFM belongs to the class of methods termed Scanning Probe Microscopy (SPM) and produces a topography of the specimen. In its classical use, the tapping mode, the topography is measured by scanning the surface by keeping a small tip probe (see figure E2.4.1) in intermittent contact with the specimen. In tapping mode, the probe is kept in intermittent contact with the sample. The specimen is less affected by the measurement in this mode. This form of contacting is carried out by oscillating the probe near its resonance frequency and this oscillating probe is brought nearer to the sample by a piezoelectric actuator. When contact is made, some absorption occurs and the oscillation amplitude (measured by mirroring a laser beam onto the back of the probe) diminishes. The ratio of the amplitude when in contact to the amplitude when not in contact is the amplitude ratio. The smaller is the amplitude ratio, the greater is the interaction between the probe and the specimen. The feedback control (by a Proportional Integral Derivative (PID) controller) to keep this intermittent contact at a constant amplitude ratio by moving the piezo actuator is used to make measurement of the vertical topography of the specimen.

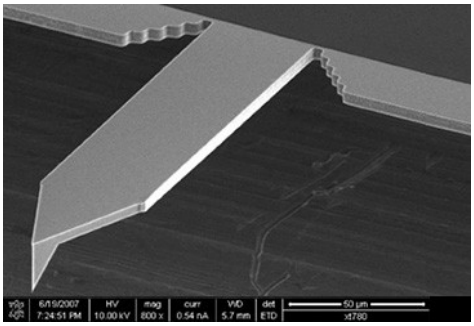
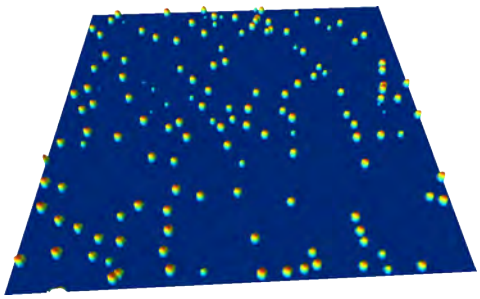


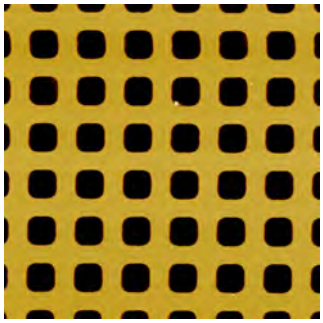
Figure E2.4.1: Electron microscopy picture of an Olympus AC160TS-R3 AFM probe

The scanning of the surface to measure the topography and render the latter as a raster image is carried out line by line in a back-and-forth movement of the tip, at a given speed (the scan speed). Two raster images of the topography are thus obtained: the "trace" image with all the lines scanned in the forward direction, and the "retrace" image made from all the lines scanned in the backward direction. The "retrace" image is usually used as the measurement. This back-and-forth scanning is illustrated by a figure in example [184].

Assuming a spherical shape for nanoparticles, the height at the top of the nanoparticle with reference to the flat substrate – on which the nanoparticles are deposited – is equal to the diameter of the spherical nanoparticle, and acts as a measurand for it. An example of topography is presented in figure E2.4.2a.



(a) 3D view of AFM topography of a polystyrene 100 nm nanoparticle sample deposited on flat mica substrate



(b) AFM topography of a grid surface topography standard. The standard height is the height difference between the black and orange areas (the colour stands for the z-scale)

Figure E2.4.2: AFM topographies of the two types of samples considered in this example

Some influence factors have been identified by the experimenters, and some of them have been studied individually, like in [185]. The feedback control parameters are adjusted manually by the operator, based on image quality (by eye) and independently for each sample. In addition, the tip probes are consumables subject to wear, which can also be dirtied during measurement and are often to be replaced. Unfortunately, no full measurement model exists to relate the measurement parameters, as would be needed for the GUM LPU [2], or a Monte Carlo approach as in [3] for uncertainty evaluation. Furthermore, correlations between effects are unknown but cannot be neglected, a priori. The effects can be categorised into fixed effects (feedback parameter, imaging parameter, type of tip probe), on which the experimenter has some control, and random effects, time dependent at different time scales (ambient conditions, tip wear, tip dirtied), which are very difficult to control. Human factors are a priori also present: as operator for image collection and for image analysis, with their own prior experience, skills and thus their potential biases.

The statistical model – an optimised DoE for a hierarchical mixed model, in a Bayesian framework – is built upon several features here briefly explained:

hierarchical mixed model : presence of uncertainty sources from fixed effects and hierarchical model for random effects to extract intermediate precision conditions with several levels of nesting – an ANOVA test would fit that purpose if only random effects were considered;

Design of Experiment (DoE) : in order to appreciate the influence of the fixed effects (individually and also jointly, by the interaction terms) and to structure the random effects (intermediate precision conditions);

optimised : it is an optimised DoE: not all the combinations of effects are considered to reduce the time needed for measurements, but this subset is optimally selected in order to minimise the induced bias;

Bayesian : parameters are estimated using the Bayesian approach instead of the frequentist approach, using a MCMC method (with the `rstan` library [186]).

In order to simplify the model and limit the number of images to be acquired, only significant effects are kept after a first analysis with topography standards (in addition to providing calibration data). This analysis results in a reduction of data and experimental time and is a second "optimisation" feature of the model, important for commercial applications.

Two main criteria are to be met to claim metrological quality for a measurement: the measurement should be traceable to SI units and an evaluation of the uncertainty on that measurement should accompany it. Surface Topography Standards are used as calibration artefacts, externally calibrated: it is a calibration by comparison. This z-scale (vertical) calibration provides the traceability of the measurement. The raw measurement data are the raster images of the topography, on which image analysis is performed to obtain the measured heights. The uncertainty contributions are evaluated at the same time as the central "mean" value for the measurement. For fixed effects, the central value is an averaged opinion over the different factors, for each uncertainty source. Each factor represents a reasonably valid choice (from expert knowledge), also for continuous variables, and can thus reasonably be represented as categorical variables, each value ("level") being considered as equally valid as the others.

The experiments have been performed on an Asylum Research MFP-3D Infinity AFM (Oxford Instruments, USA), in tapping mode. Only "retrace" images were considered for measurement. The Surface Topography Standards (VLSI Standards Inc., Milpitas, CA, USA) are grid standards for x-y lateral and z vertical dimension calibration (see figure E2.4.2b; the black parts represents the reference holes) and were externally calibrated by means of a metrological AFM. Their specifications are listed in table E2.4.1. Gold nanoparticles of this work are RM8012 (NIST, Gaithersburg, USA) reference material, with a $(24.9 \pm 1.1)\,\mathrm{nm}$ ($k = 2$, 95 % coverage probability) certified size [187]. Deposition on substrate is performed following the procedure explained in [188]. The tip probes are the consumables part of the uncertainty evaluation and they will be detailed together with other uncertainty sources later in the text.

E2.4.3 Specification of the measurand(s)

The height of the nanoparticle is the measurand acting for its size, under the assumption of the nanoparticles being of spherical shape. Figure E2.4.3 illustrates the measurement of the topography along one scanning line, and how the size h of the nanoparticle would ideally appear

Table E2.4.1: Properties of the VLSI Surface Topography Standards used for AFM calibration

Grating	Product name	Nominal step height μ_n/nm	Calibrated step height ($k = 2$) μ_c/nm
$S_{[1]}$	STS 3 180p	18	15.6 ± 1.0
$S_{[2]}$	STS 2 440p	44	42.3 ± 1.2
$S_{[3]}$	STS 3 1000p	100	99.0 ± 1.2
$S_{[4]}$	STS 3 1800p	180	177.4 ± 1.3

in it. The height of a nanoparticle is extracted as the height of the pixel of maximum height within the set of pixels identified as belonging to the nanoparticle, in the 2-dimensional raster image of the topography. For the topography standards (grids or gratings), the characteristic height is the step height, which is obtained by comparing the height between the valleys and the plateau in the pattern (see figure E2.4.4), and averaging the height at their centre with removal of the edges to avoid some instrumental effects (ISO 5436-1 [189]) by only using the greyed parts of figure E2.4.4. This averaging is done by a modified least square fit, taking into account the step shape.

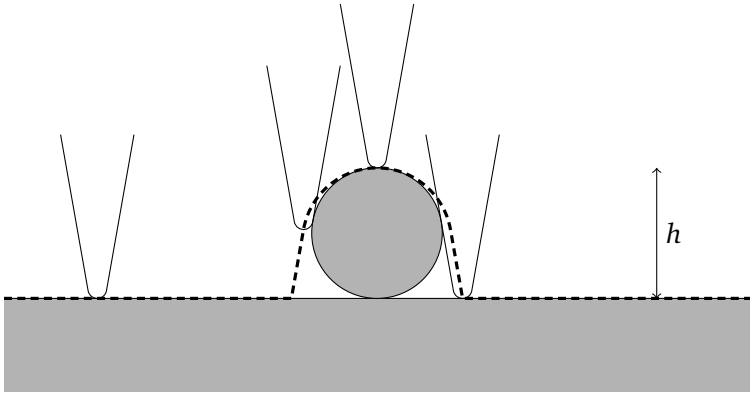


Figure E2.4.3: Measurement of nanoparticle size by AFM. The tip probe is represented at several positions along its path, and the broken line represents the measured topography

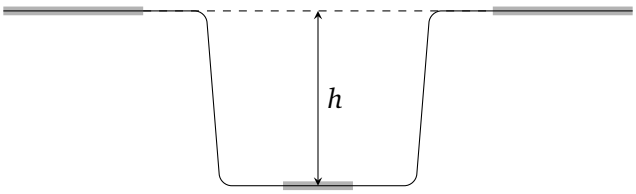


Figure E2.4.4: Height measurement for topography standards along an AFM scanning line. Only the regions in grey (i.e. regions away from the steps) are considered for measurement

The mean size of the sample is extracted, together with uncertainty evaluation, simultaneously. Each nanoparticle under measurement is assigned the same weight; it is the number-based mean size, like other microscopy techniques but unlike other measurement techniques such as light scattering or light-absorption based.

E2.4.4 Measurement model

The measurement data are collected according to a D-optimised DoE computed by the SAS JMP software [190], in order to limit the experiment time: 150 images for nanoparticles and 69, 24, 105 and 75 images for grids of 18 nm, 44 nm, 100 nm and 180 nm nominal step height (further labelled $S_{[1...4]}$ in the text), respectively. Not all the combinations of effects are considered, but only a limited number of them without introducing much bias; this is the purpose of the optimised DoEs. It took about one month of full-time measurement, with the optimisation already applied. Common procedures are used but some parameter adjustment is still needed on the instrument, dependent on the sample at hand and operator skills, which could introduce some uncertainty. The measurements are extracted from the processed raster images. After subtraction of the background (substrate), the analysis of the images is performed by an operator to identify the relevant pixels for the measurements: sets of contiguous pixels for nanoparticles and sets of lines passing through the central areas of patterns for step height standards. Again, some skills are needed and the image analyst could potentially introduce some bias in the results.

No physical measurement model can be obtained; a statistical modelling of the measurement is used instead, taking into account the different aspects listed above. Its global equation is

$$h_{ijkl} = \underbrace{\beta_0}_{\text{grand mean}} + \underbrace{\sum_{f=1}^{m_f} \mathrm{X}_{f,ijk}\,\vec{\beta}_f + \sum_{f\neq g} \mathrm{Y}_{fg,ijk}\,\vec{\beta}_{fg}}_{\text{fixed effects}} + \underbrace{a_i + b_{ij} + c_{ijk}}_{\text{random effects}} + \epsilon_{ijkl}, \tag{E2.4.1}$$

where h_{ijkl} represents the measured height and where indices i, j, k and l are for different days, positions, images and nanoparticle size measurements respectively. It contains three parts: the grand mean β_0, which is the *bare* mean measured size in the absence of uncertainty sources (said another way, the average opinion among the fixed effects, and when *removing* the random effects), the latter being in fixed and random effect parts. The unknown m_f fixed effects (labelled by f) are considered in $\vec{\beta}_f$, with their DoE matrix $\mathrm{X}_{f,ijk}$ coded for each effect and their 2-way interactions $\vec{\beta}_{fg}$. $\mathrm{Y}_{fg,ijk}$ is the combined DoE with coded effects matrix expressed as the tensor product of the matrices for single effects. Random effects form the third part, where the hierarchical structure with 3 *nested blocks* (hierarchical layers, see figure E2.4.5) can be understood from the indices. In this model $a_i \sim \mathcal{N}(0, \sigma_{a_i}^2)$, where σ_{a_i} is the actual estimated parameter; and similarly for b_{ij} (with parameter $\sigma_{b_{ij}}$) and c_{ij} (with parameter $\sigma_{c_{ijk}}$), where $\mathcal{N}(\mu, \sigma^2)$ denotes the normal distribution with mean μ and standard deviation σ. ϵ_{ijkl} is the residual fitting term and has a particular significance for the measurement that will be later explained. The DoE is defined as the set of values i,j,k together with the $\vec{\beta}_f$ coded in it and its blocking structure. The blocking structure is the roadmap of measurements in timely order and grouping, that is, the intermediate precision conditions (day, position, image). Priors are chosen to be non-informative.

Each opinion of the categorical variables is considered equally valid, and thus should be modelled as balanced. In the coding, a contrast between each categorical value ("level") is desired; this is the aim of orthogonality. The effect-type coding matches these purposes [191, 192]. Effect-type coding codes n effects into $n-1$ variables by using an identity matrix for the first $n-1$ effects and a vector with its $n-1$ elements set to -1 for the last effect. This choice brings $n-1$ degrees of freedom for the uncertainty evaluation of the considered fixed effect, and 1 degree of freedom to the estimation of the grand mean value: estimation of the mean size, in our example. The fixed effects considered here are:

tip probe type 3 levels: Olympus AC160TS (illustrated in figure E2.4.1), Olympus AC240TS, NanoSensors PPP-NCHR (classically used probes with different geometric tip shape, stiffness constant and resonance frequencies),

tapping force 4 levels: 65 %, 70 %, 75 % and 80 % (set-point amplitude ratios for the feedback control in tapping mode; the ratio is relative to the free oscillation of the tip: the smaller is the ratio, the more the tip pushes on the specimen),

scan speed 3 levels: $1.8\,\mu m\,s^{-1}$, $3.6\,\mu m\,s^{-1}$ and $5.4\,\mu m\,s^{-1}$ (lateral scan speeds along the scanning line),

operator, image analysts 3 levels each, all operators/analysts were trained but with different experience with AFM.

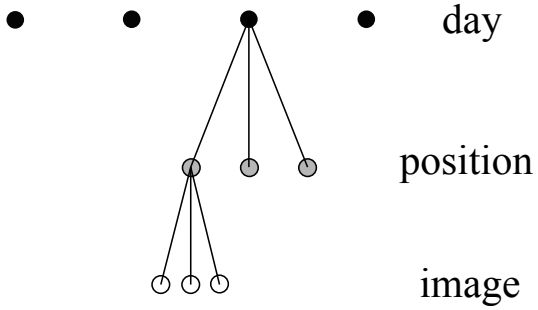


Figure E2.4.5: Nested design schema for the random variables, with its 3 stages of blocks

Considering all these possible effects would lead to a large amount of data and a huge time needed for experiment (recall that the AFM is considered as a slow measurement technique). A first DoE is thus followed for step height standards, which are somewhat faster to measure, and data are then processed to identify significant effects and interactions (with the `mixed` procedure in the SAS software [193], with 95 % confidence level criterion). The operator and image analyst effects turned out to be not significant, for both step height standards and nanoparticles. For step height standards, only the probe effect was significant. Only interaction terms between the probe type and the scan speed, and between the amplitude ratio and the scan speed were found to be significant. The set of considered parameters is then reduced to those that are significant and the new restricted optimised DoE is applied on nanoparticle measurements.

E2.4.4.1 Calibration

Measurement data are calibrated according to height measurement performed on Surface Topography Standards, based on measurement reported in section E2.4.6. These step height standards were purchased from VLSI Standards Inc. and externally calibrated by a metrological AFM, resulting in directly traceable measurement and small uncertainties.

The proposed calibration model is

$$q_c = \alpha + \gamma q_m + \delta q_m^2 + \epsilon, \tag{E2.4.2}$$

where q_c is a certified value, q_m a measured value, and α, γ, δ and ϵ are adjustable parameters. In the Bayesian approach, all these terms are to be understood as random variables rather than simple real numbers. The parameter ϵ is a normal deviation from the quadratic model and α, γ and δ are calibration curve parameters.

Calibration parameter estimation

Information at hand are certified values for the step height standards together with their uncertainties. In the model, they were assigned normal distribution with relevant information from certificate (this would form the prior). A thorough analysis of the step height standard was carried out by following the full procedure of this DoE (see results in other sections of this text). For each standard, the information was summarised in the form of a mean value with an associated uncertainty. A normal distribution with parameters adopted from summary information was used for q_{m}, for each standard. The validity of the approximation by a normal distribution was checked by eye.

The approach of establishing estimates of the calibration curve coefficients is similar to MCMC that would be applied in `rstan` [186], but without the merit function optimisation. The use of ordinary least squares regression is the single optimisation step present. The joint density for $(\alpha, \gamma, \delta, \epsilon)$ is approximated by sampling N times from its distribution through the following procedure:

- take N samples $(\mu_{\mathrm{m}j}, \mu_{\mathrm{c}j})_{[r]}$ (for $j = 1, .., N$) from the 4 calibration points ($r = 1, 2, 3, 4$);
- calculate for $j = 1, .., N$ the estimated coefficients α_j, γ_j and δ_j, and the mean squared error s_j^2 by performing N times an ordinary least-squares method (OLS) quadratic regression with regression data $(\mu_{\mathrm{m}j}, \mu_{\mathrm{c}j})_{[r]}$ for $r = 1, 2, 3, 4$;
- sample a random value ϵ_j from $\mathrm{N}(0, s_j^2)$ (for $j = 1, .., N$);
- collect the sample $(\alpha_j, \gamma_j, \delta_j, \epsilon_j)$.

The joint distribution of $(\alpha, \gamma, \delta, \epsilon)$ is approximated by merging the N samples $(\alpha_j, \gamma_j, \delta_j, \epsilon_j)$ for $j = 1, .., N$. The table in figure E2.4.6 shows the results obtained for the parameters with $N = 10^6$. Given the value of δ, the calibration curve can be defined as linear. The 95 % confidence interval of the calibration curve and the mean calibration curve are also illustrated in figure E2.4.6. The random trials of all 4 parameters $(\alpha, \gamma, \delta, \epsilon)$ are saved as tuples and form the calibration curve information, as a joint PDF. Full correlations inferred from the Bayesian fit are thus kept.

Parameter	$\mathbb{E}[.]$	$s[.]$
α/nm	−0.2059	0.7724
γ	1.0025	0.0252
δ/nm^{-1}	0.000 003	0.000 135
ϵ/nm	−0.0007	0.6919

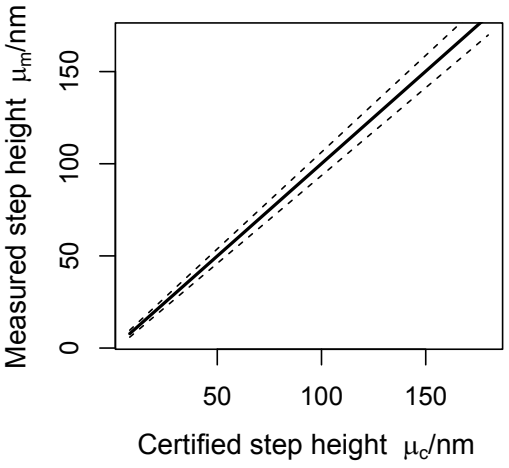


Figure E2.4.6: Expected values and standard uncertainties for the calibration curve parameters – graphical representation of the 95 % confidence interval of the calibration curve

Calibrating measurements

To apply the calibration on a real measurement (a nanoparticle height, for example), each data value is applied to the calibration curve, with all the saved trial information. For each original data point, a set of certified values is thus obtained.

E2.4.5 Uncertainty propagation

The uncertainty evaluation is performed by MCMC in `rstan`. It is important to keep in mind that the evaluation is not a full Bayesian inversion. The main estimated parameter is the mean height (i.e. mean nanoparticle size or mean step height, depending on the case of application), and the other parameters are uncertainty contributions, which were unknown a priori. In parallel, a frequentist methodology has also been followed in the SAS software (version 9.4) and its `mixed` procedure [193]. Measurement uncertainties derived in the Bayesian approach are generally slightly smaller, although not significantly, for the fixed effects and comparable for the random effects.

E2.4.6 Reporting the result

The results to be reported are extracted using the distributions obtained from MCMC. A combination of variables can be constructed by using the random trials. For example, if only fixed effects are considered, the measured mean size is:

$$\mu_{\text{fixed}} = \beta_0 + \sum_{f=1}^{m_f} \mathrm{X}_{f,ijk}\,\vec{\beta}_f + \sum_{f\neq g} \mathrm{Y}_{fg,ijk}\,\vec{\beta}_{fg}. \tag{E2.4.3}$$

If random errors are also considered:

$$\mu_{\text{fix, rnd}} = \beta_0 + \sum_{f=1}^{m_f} \mathrm{X}_{f,ijk}\,\vec{\beta}_f + \sum_{f\neq g} \mathrm{Y}_{fg,ijk}\,\vec{\beta}_{fg} + a_i + b_{ij} + c_{ijk}. \tag{E2.4.4}$$

From the distributions provided by MCMC, the expected value ($\mathbb{E}$) and the standard deviation (s) are extracted to report the mean parameters. For the random effects, the mean of the MCMC estimation of the σ^2 parameters is reported. For the fixed effects, the standard deviation will be reported to evaluate the uncertainty coming from this effect, after merging the MCMC PDFs of all effect levels into a single PDF. It is important to note that these are only summary information and the real information is instead the distribution underlying the estimated values, i.e. the full set of MCMC trials (less those associated with the MCMC warm-up stage). The full sets are presented in several figures of the PDFs, for nanoparticles and for illustration purposes, not only for the main fixed effects, but also their interactions.

E2.4.6.1 Step height standards

The results of parameter estimation for the step height standards are summarised in table E2.4.2 for the estimated mean size. The intermediate precision uncertainties estimated from this statistical model are presented in table E2.4.3. Day and position (very approximately chosen position

before starting any measurement) effects are coarsely of similar magnitude, for each step height standard considered apart. One possible explanation for this effect is the difficulty in arriving at exactly the same position with an AFM. This means that in attempting to image the same position, a nearby position is imaged instead. Variability is thus present, which behaves in a similar way to the day-to-day variability. A further consequential uncertainty arises that increases with increasing nominal step height. One possible explanation would come from fluctuations (of different spatial frequencies) in the lithography process, which would scale with nominal step height.

Table E2.4.2: Expected values and standard uncertainties for the measured mean step height μ_m for each reference standard grating

Grating	$\mathbb{E}[\mu_m]$/nm	$s[\mu_m]$/nm
$S_{[1]}$	15.91	0.13
$S_{[2]}$	42.15	0.01
$S_{[3]}$	99.06	0.54
$S_{[4]}$	177.04	0.70

Table E2.4.3: Variance of the random effects influencing the step height measurements for each grating

Effect	σ^2/nm^2 for $S_{[1]}$	σ^2/nm^2 for $S_{[2]}$	σ^2/nm^2 for $S_{[3]}$	σ^2/nm^2 for $S_{[4]}$
day	0.0075	0.0001	0.5439	1.0923
position	0.0149	0.0016	0.1623	0.6170
image repeat.	0.0000	0.0003	0.0002	0.0018
within image var. (ϵ_{res})	0.0309	0.0455	0.0105	1.4797

It is also observed that image repeatability has almost null contribution; and this effect is thus not considered for nanoparticles, again considering the experiment time issue. The within-image variability (residual) is the main contributor to observed variability of results.

Table E2.4.4: Standard uncertainty of the probe fixed effect for each grating

Grating	$s[\beta_{probe}]$/nm
$S_{[1]}$	0.11
$S_{[2]}$	not significant
$S_{[3]}$	0.25
$S_{[4]}$	not significant

Details for the probe fixed effects are given in table E2.4.4, for which the effect is significant for standards $S_{[1]}$ and $S_{[3]}$ and not for the other two standards. No explanation could be found for this peculiar effect: the standards come from the same manufacturer and are produced in

a similar way. This effect possibly results from the different number of images and in relation to the difference enhanced by the optimisation of the DoE. The observed effect is nevertheless taken into account in further calculations.

E2.4.6.2 Nanoparticles

Main effects

Figure E2.4.7 presents the general estimated distributions for the means β_0, μ_{fixed} and $\mu_{\text{fix, rnd}}$ from the mixed model, showing that the main contributor is from the random effects. Summary values are provided in table E2.4.5, where h_{m} is the measured gold nanoparticle height (raw data, from equation (E2.4.1)) and μ_{m} is the estimated mean measured height (μ_{fixed} in equation (E2.4.3)). The estimated mean value is $\mathbb{E}[\mu_{\text{m}}] = 23.40\,\text{nm}$, $s[\mu_{\text{m}}] = 1.19\,\text{nm}$.

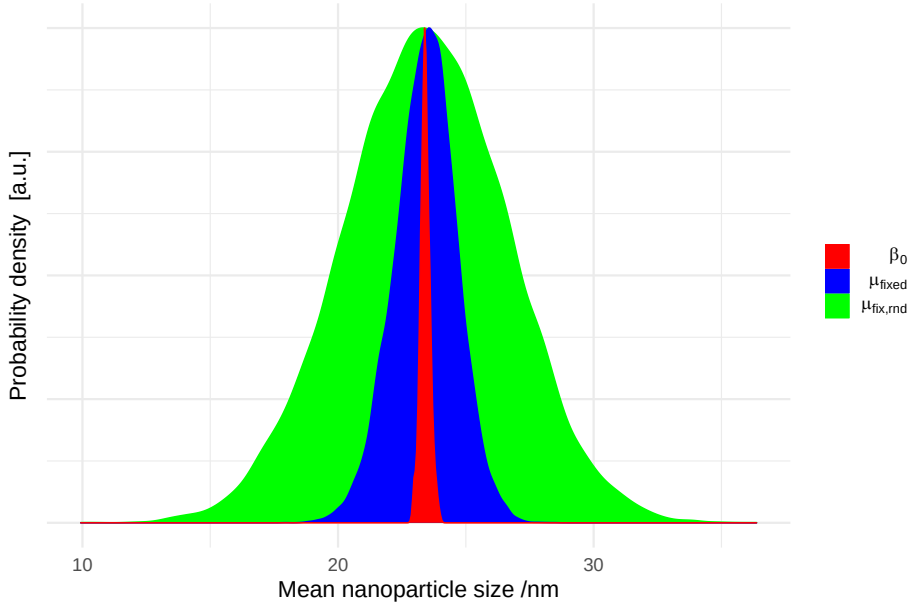


Figure E2.4.7: MCMC PDFs of the means, scaled to their maximum peak heights. The grand mean β_0 (intrinsic mean) is in red, the mean when taking into account from fixed effects μ_{fixed} (eq. (E2.4.3)) in blue and when considering all uncertainty factors, $\mu_{\text{fix, rnd}}$ (eq. (E2.4.4)), in green.

Table E2.4.5: Expected values and standard uncertainties for the measured particle height h_{m} and the mean measured height μ_{m} of the gold nanoparticle sample

$\mathbb{E}[h_{\text{m}}]$/nm	$s[h_{\text{m}}]$/nm	$\mathbb{E}[\mu_{\text{m}}]$/nm	$s[\mu_{\text{m}}]$/nm
23.39	3.18	23.40	1.19

The contributions of intermediate precision conditions (hierarchical random effects) are presented in table E2.4.6, where parameters are summarised blocking level per blocking level. As for the step height standards, the main source of variability of the measurement is the within-image variability (residual ϵ_{res}). The position random effect seems to bring a little more variability than the day of measurement.

Table E2.4.6: Evaluated variance of the random effects influencing the nanoparticle height measurements

effect	$u^2[.]/\mathrm{nm}^2$
day	$\mathbb{E}[\sigma^2_{\mathrm{day}}] = 0.16$
position	$\mathbb{E}[\sigma^2_{\mathrm{pos}}] = 0.88$
within image variability (ϵ_{res})	$\mathbb{E}[\sigma^2_{\mathrm{res}}] = 7.61$

More detailed investigation of the fixed effects is presented in figures E2.4.8, E2.4.9 and E2.4.10. An amplitude ratio of 80 % shows higher estimation of the size, compared to other amplitude ratio parameter values. No clear effect of the scan speed nor the probe could be observed. Note that a null average is expected over the whole effect (if merging all effects levels into a single PDF), because of the choice of effect-type coding. If a bias were present in the data for a given fixed effect, its contribution would be propagated and included in the grand mean β_0. The fixed effect contributions to uncertainty are summarised in table E2.4.7, by taking the estimated standard deviations (after merging the MCMC PDFs from all levels of the fixed effect) as fixed effect contributions. The different fixed effects lead to contributions of similar magnitude, about 0.5 nm.

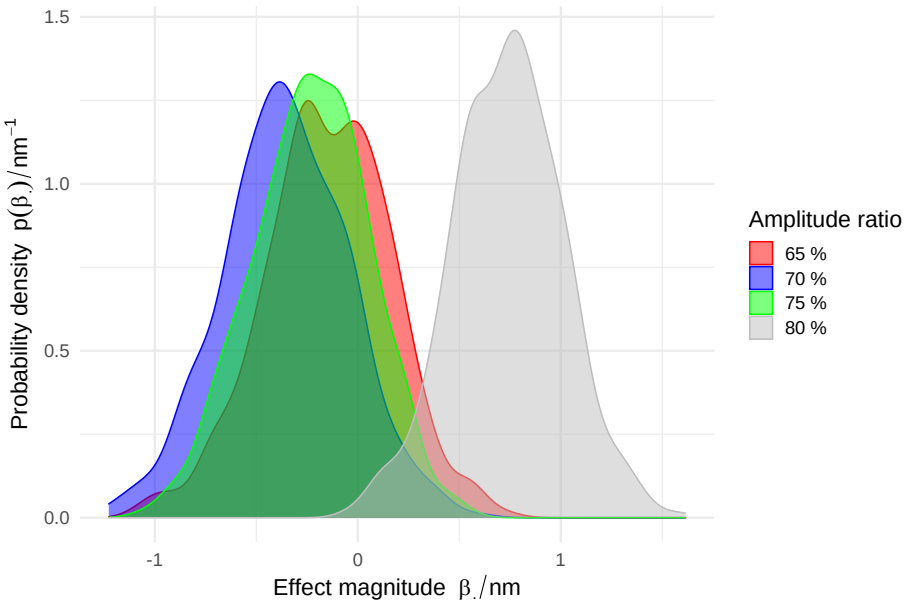


Figure E2.4.8: MCMC fixed effect coefficients PDFs for the amplitude ratio (related to tapping force). The amplitude ratio of 80 % has a tendency to increase the size; this value has a smaller tapping force (free and contact amplitudes are closer). The other values have very similar behaviour

Interactions

Interrelations of the various quantities are investigated via the estimated interaction terms. A trend with the scan speed is observed from figure E2.4.11 at high amplitude ratio (80 %). Its magnitude is as high as 1 nm, but only visible via the interaction terms.

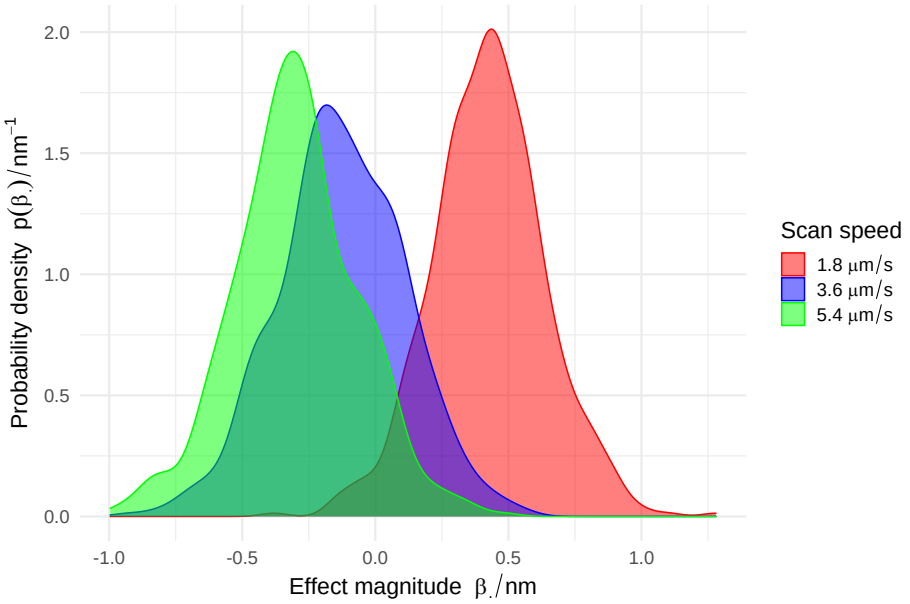


Figure E2.4.9: MCMC fixed effect coefficients PDFs for scan speed

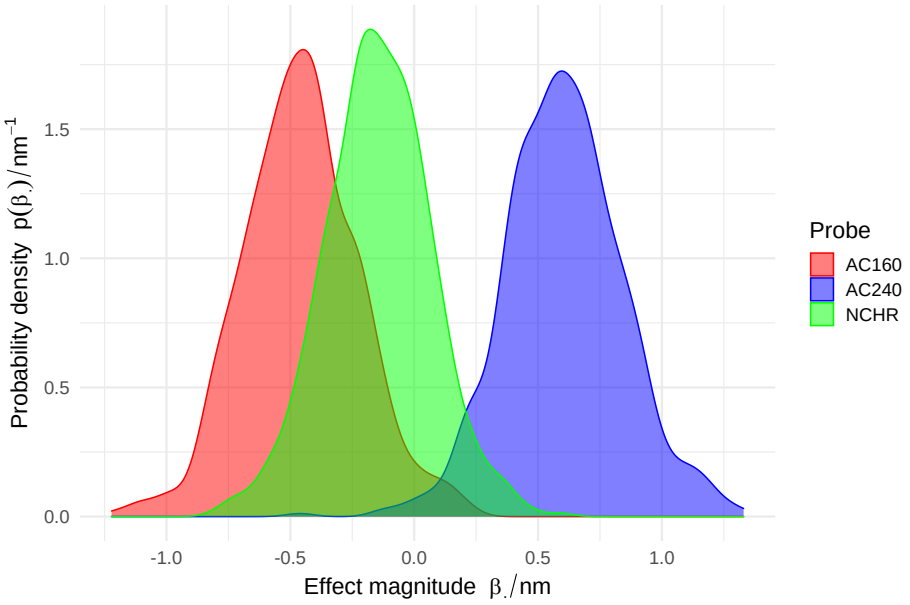


Figure E2.4.10: MCMC fixed effect coefficients PDFs for the probe model

The PPP-NCHR probe displays different behaviour from that of the other probe types, and seems to measure higher nanoparticles at high speed, as can be seen from figure E2.4.12. The observed effect would be of the order of 1 nm in magnitude.

Table E2.4.7: Estimated standard deviation of the respective fixed effects significantly influencing the nanoparticle height measurements

Effect	$u[.]$/nm
probe	$s[\beta_{\text{probe}}] = 0.49$
amplitude ratio	$s[\beta_{\text{tapping force}}] = 0.52$
scan speed	$s[\beta_{\text{speed}}] = 0.39$

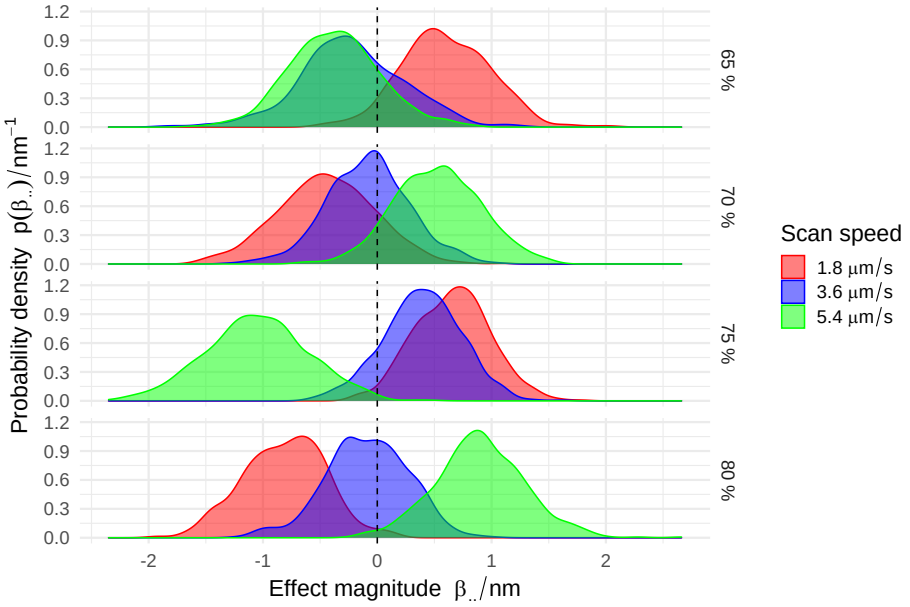


Figure E2.4.11: Interaction PDFs between the amplitude ratio and the scan speed parameters, grouped by amplitude ratio

E2.4.7 Interpretation of results

Only the interpretation of nanoparticle results is reported here, because the interpretation of results on step height standards was already carried out in the previous section, and the purpose of step height measurement is more to derive a calibration curve than to be the subject of the main measurement, itself.

β_0 is the "central-intrinsic" mean value of the sample. This is the average response over the different fixed effects, with their levels all considered on an equal footing. To consider the uncertainty sources from fixed and random effects, μ_{fixed} and $\mu_{\text{fix,rnd}}$ are computed as MCMC distributions. ϵ_{ijk} carries the remaining variability in the model: for nanoparticles, it is the variability within the image and thus relates to the spread of the nanoparticle size distribution (intrinsic, physical). As the interest here is in measuring the *mean* size of the nanoparticle sample by parametric estimation, ϵ_{ijk} (being equal to ϵ_{res} on average) is left aside. For characterisation of the object under measurement and comparison with other measurement techniques, the value with fixed effects, μ_{fixed}, is to be considered. If biases are present for some fixed effects values, they should be covered as an uncertainty, and averaged out because of the effect-type coding, and this average part propagated to the grand mean. The following result for the gold nanoparticle sample is obtained from the MCMC distribution: $\mathbb{E}[\mu_{\text{m}}] = 23.40\,\text{nm}$, $s[\mu_{\text{m}}] = 1.19\,\text{nm}$ (table E2.4.5). Under normality assumption, these values can be used, after calibration, to compare with other results

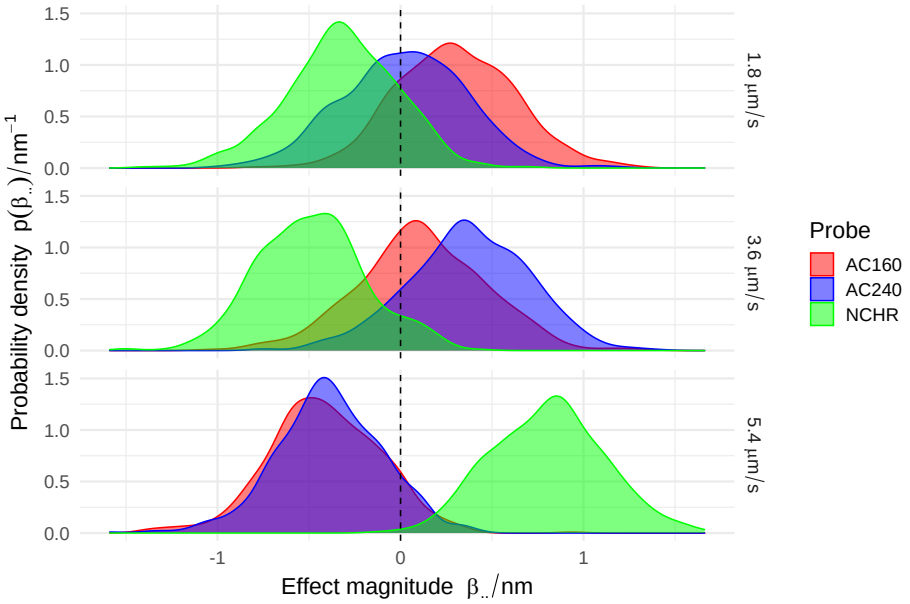


Figure E2.4.12: Interaction PDFs between the probe and the scan speed parameters, grouped by scan speed

Table E2.4.8: Expected values and standard uncertainties for the certified particle height h_c and the certified mean height μ_c of the gold nanoparticle sample

$\mathbb{E}[h_c]$ /nm	$s[h_c]$ /nm	$\mathbb{E}[\mu_c]$ /nm	$s[\mu_c]$ /nm
23.24	3.28	23.25	1.44

obtained via a classical GUM-LPU method. The final result $\mathbb{E}[\mu_c] = 23.25\,\text{nm}$, $s[\mu_c] = 1.44\,\text{nm}$ (table E2.4.8) is compatible with the certified value of the sample: $(24.9 \pm 1.1)\,\text{nm}$ ($k = 2$, 95 % coverage probability), validating the approach.

With this approach, μ_{fixed} and $\mu_{\text{fix,rnd}}$ cannot be obtained by splitting contributions from each contribution and summing them up quadratically, as in a GUM-LPU approach. This is, again, related to the fact that numbers are only summary information and not the full distribution information. Furthermore, correlations and non-linearities of the model are better accounted for by the Bayesian approach. In the present case of uncertainty evaluation by means of an optimised DoE, correlations appear via estimated interaction terms[1] but also in simultaneous trials among the estimated variables (Bayesian approach), and this, non-linearly. It would only be linearly in the GUM-LPU approach.

More precisely, interactions and correlations are different in several respects. Correlations relate to simultaneous fluctuations of variables, while interactions, defined in the paradigm of continuous variables, more relate to non-linearities of the physics here statistically modelled. In this example, the continuous variables (i.e. scan speed and amplitude ratio) are encoded as categorical, to be implemented in the same way as naturally categorical variables (i.e. probe type). The continuity aspect is lost, which could be damaging in case of modelling, but not in this case, as occurrences of these continuous variables are all equally valid, experimentally speaking, and

[1] In [191] (p.221), a note is inserted to make clear the distinction between *interaction* and *correlation*. However, in the present context of uncertainty evaluation and potential comparison with a classical approach, the distinction is not so straightforward

should thus be considered on an equal footing (not necessarily in their sequence order, 65 %, 70 %, 75 % and 80 % amplitude ratio, for example). In an ideal non-optimised DoE, all probed combinations of parameters would be equally populated : no correlation is present among the parameter input quantities, only the response variable (the measured height h, here) fluctuates. The estimated interactions thus reflect properties of the underlying physics, which is statistically modelled.

We now continue with the amplitude ratio example (a similar discussion holds for the scan speed). Experimentally, its continuous values are manually typed into the machine, and any fluctuation in the parameter adjustment by the machine will be reflected in the result. The coverage of the valid range is made through this choice of the four values. As a reminder, the amplitude ratio is the ratio of the amplitude set point (when in contact) to the free amplitude (when not in contact). A bigger amplitude ratio means less contact with the specimen, which can cause bad tracking of the specimen, resulting in a parachuting effect. With this effect, the measured height (hence nanoparticle size) can be greater, as can be clearly seen in figure E2.4.8. This is also one interpretation of the behaviour seen of interactions for the 80 % amplitude ratio with scanning speed variation, as can be seen on figure E2.4.11: the faster is the scan, the more pronounced is the parachuting effect and the higher is the measured size. Note that this effect might be related to pixel size effect described in example [184]: the higher size coming from averaging higher points in the vicinity of the apex over the pixel, its being measured higher because of the parachuting effect. From expert knowledge, no parachuting effect is expected from the lower amplitude ratio values, the topography being more closely tracked. The other fluctuations present in the estimated effect magnitudes could be attributed to non-ideality of the optimised DoE and fluctuations in data, enhanced by the chosen effect-type coding which makes the global average null and forces some shuffling of the estimated effects around zero. This effect should be further investigated with more data but it is not visible without looking at interactions, because it is averaged out by the methodology of effect-type coding for the main effect. If looking at correlation among variables without use of categorical variables, this effect would be much less pronounced and could have gone unnoticed because of dilution of the information with data at the three other amplitude ratios.

Unfortunately, no sensible explanation could be provided for the different behaviour of the PPP-NCHR probe at high scan speed, compared to the other probes, as seen from figure E2.4.12. Some speculation can however be drawn about its origin. The PPP-NCHR has a different tip shape than the other probes, and maybe some discrepancies could be observed in non-ideal conditions, like it is for the high scan speed. In addition, it is also possible that the different spring constants play some role, through the interplay of several adjusted factors. Despite the lack of explanation, this effect is nevertheless taken into account in the calculations.

Both of these effects at high amplitude ratio and for the PPP-NCHR probe were not visible from the analysis of the main effects. If not considering these significant interaction terms, the fixed effect estimations could be biased and some correlation be present among the estimated fixed effects (if not averaged out). In an ideal case, a DoE should be balanced among all the levels, meaning that input quantities (experimental parameters of the fixed effects) do not present correlations. By reducing the number of combinations in the DoE, some correlations between estimates can appear by imbalance among the effects (at their levels). It is the case here because an optimised DoE is used, although the D-optimality criterion was applied, mitigating the problem. Some enhancement of this issue can also potentially occur because the number of nanoparticles in an image can vary. It is also to be noticed that the magnitude of the observed effects is rather small, being less than 1 nm.

E2.4.8 Conclusions

This example illustrates uncertainty evaluation when no full measurement model is available. For the present example of nanoparticle size measurement, some models to describe behaviour of single effects can be found in the literature, but they unfortunately do not permit to take into account their interrelations into a single full measurement model. A statistical model is thus adopted (in the spirit of [5, clause 11]), and the uncertainty evaluation is carried out within a common framework for intermediate precision conditions and fixed effects. When several parameter values are to be considered on an equal footing for a given effect, categorical variables is used instead of continuous variables. These variables are also effect-type coded for equally valid experimental parameter values. An optimised DoE is followed to limit measurement time, and significant parameters are identified to reduce further this time. The mean of the size distribution, with size being the height measured by the AFM, is the measurand. A Bayesian approach is followed for this parameter estimation, using `rstan`.

This method allows for a better handling of these interrelations, at several steps of the process: for the calibration curve and by using categorical variables. For the calibration curve estimation, full correlations among the calibration parameters are kept by retaining the vector of parameters (and not their individual PDFs) for each trial. By using categorical variables, some insight is lost on modelling the physics of the effects, but one has access to interrelations under the form of interactions, value-per-value (= level-per-level) of each effect and not effect-per-effect as in the conventional methodologies. This phenomenon was highlighted for the case of amplitude ratio of 80 %, together with the scan speed variation. The main goal of the methodology developed here is to provide a measurement and its uncertainty evaluation in more affordable time, and not a to provide a detailed investigation of the various effects. Some interesting features could nevertheless be observed and discussed.

The relevant variables and relevant interactions were first identified by a frequentist estimation with the SAS software and the final parameter evaluation by MCMC (`rstan`). The two methods however yield similar results for this parameter estimation, with slightly smaller uncertainties for fixed effects with the Bayesian approach.

This approach has already been successfully applied to ISO/IEC 17025-accredited activities by the authors in Belgium, and could be adapted to other commercial application where no full measurement model exists and time savings are important.

Example E2.5

GUM-LPU uncertainty evaluation — importing measurement traceability from a conformity statement

J. Greenwood, A. Bošnjaković, V. Karahodžić, P. Pedone, F. Manta, M.G. Cox

E2.5.1 Summary

Measurement traceability is commonly obtained from calibration measurements that provide a result in terms of a single value and its associated uncertainty. However, there are circumstances where instead, the result may consist of a range of possible values. Such circumstances might arise when a result is provided in the form of the output from a conformity decision process, for example as a conformity statement in which a range of acceptable values rather than a specific value is reported. In terms of metrological traceability this style of result provides less information than a specific value, but it may be sufficient to obtain an acceptable target measurement uncertainty for a given application. The standard ISO/IEC 17025 acknowledges the provision of such information in informative annex A. This example describes how such information might be used to propagate traceability.

E2.5.2 Introduction

Under typical circumstances, evaluation of measurement uncertainty following the GUM [2] law of propagation of uncertainty (LPU), involves assigning a probability density function (PDF) to the measurand that usually has a normal distribution (or sometimes a t distribution). This 'output' PDF corresponds to a combination of the PDFs for all the inputs to the measurement. It is characterised by a location parameter — the mean value corresponding to the estimate y of the value of the measurand; and a dispersion parameter — the corresponding variance $u^2(y)$ associated with that estimate. If this result subsequently becomes an input to a further measurement, the variance is 'imported' into the corresponding uncertainty evaluation.

However, suppose instead that whilst we still obtain information that allows us to establish the variance $u^2 = u^2(y)$, we are *not* given a specific value for y. Instead we receive only information about the interval A, e.g. $[-a, a]$, in which the estimate y is located. In other words, we know the dispersion of y, but we do not have a value for the location of y, only a range of possible values.

In this case, the information can be still be brought into a subsequent uncertainty budget, but now since *two* independent PDFs, say for example a normal distribution $\mathrm{N}(0, u)$ characterising the dispersion of values around any given value of y, and a rectangular distribution $\mathrm{R}(-a, a)$ characterising the available information about the location parameter. For ease of explanation we will usually assume here that intervals A for y are centred on zero, but this is not a necessary requirement.

This situation is of potential interest to those concerned with meeting the requirements of ISO/IEC 17025:2017 since this standard [7, Informative Annex A.2.3] accepts that metrological traceability could be provided by statements of conformity.

Ideally, a statement of conformity will include (i) the specification or tolerance interval C for the measurand Y (such that $-c \leq Y \leq c$), (ii) an acceptance interval A for the estimate y (such that $-a \leq y \leq a$) defined by a decision rule that takes direct or indirect account of measurement uncertainty, and (iii) a conformance probability p_c, which is the basis for (or a consequence of) how the acceptance interval is defined. In fact, in many practical situations a so-called 'Simple Acceptance' criterion is used to define the limits for deciding conformity, in which case $A = C$. In this case, in order to meet the requirements for a decision rule appropriate for ISO/IEC 17025:2017, uncertainty is taken into account *indirectly*, usually by specifying an upper limit $u_{\max}$ that, as a *prerequisite*, must not be exceeded for the Simple Acceptance criteria to be applied.

The aim here is to provide examples with various forms for the statement of results and to show whether they allow the results to be traceably propagated. We begin by describing some likely scenarios and then provide two extended examples.

E2.5.3 Examples

In all the following examples it is assumed that the estimate $y \in A$, that is, the outcome is accepted as conforming, and that intervals are centred on zero. For this (two-distribution) model to be applied it is therefore necessary to identify A and u in each case.

E2.5.3.1 Information given: Acceptance interval and measurement uncertainty for any specific value

For purposes of metrological traceability, it makes no difference *how* the interval A has been established ($A \neq C$ or $A = C$), only that it *is* somehow defined. Given A and u the approach is straightforward; the information can be brought into a subsequent uncertainty budget as two distinct distributions e.g. $\mathrm{R}(-a, a)$ and $\mathrm{N}(0, u)$ respectively.

In this situation the information might be obtained from a statement such as

> "The measured value y has a standard uncertainty $u = 1.3$ and is within the range $-10.0 \leq y \leq 10.0$."

Note that such a statement is not a conformity statement, as no specification or tolerance interval is given, nor is there an associated decision rule.

E2.5.3.2 Information given: Acceptance interval is the same as the tolerance interval

This scenario, in which we *only* know that $y \in A$ and $A = C$, corresponds to 'unconstrained' simple acceptance, as there is no account of measurement uncertainty either 'directly' or 'indirectly'. (For this reason alone it would *not* meet the decision rule requirements of ISO/IEC 17025:2017.)

For example, suppose that a result is stated as:

> "The specified tolerance interval is from −2.0 to +2.0; the measured result is 'conforming' as it is within the tolerance interval"

In this case there is insufficient information to establish a PDF for the outcome. The 'unconstrained' simple acceptance conformity statement is therefore insufficient to provide metrological traceability. It could not be 'imported' into an uncertainty evaluation, nor could any statement of risk be made on the basis of this information.

To make use of such a statement it would be necessary to establish uncertainty by other (external) means, e.g., to request the value of u from the information provider.

E2.5.3.3 Information given: Tolerance and acceptance intervals and a statement about limits of probability or risk of acceptance

In this case, as well as stating C and A, a statement may include the minimum conformance probability $p_{c_{\min}}$ or the related quantity, maximum probability of false acceptance $R^*_{C_{\max}}$ (in the notation of [6]), which for the usual *specific* risk scenario is given by $R^*_{C_{\max}} = 1 - p_{c_{\min}}$.

The information might be found in a statement of conformity, for example, a statement such as

> "... the specified tolerance interval is from −2.0 to +2.0; the measured value is conforming as it is within the acceptance interval −1.5 to 1.5. The minimum conformance probability is 0.97"

An acceptance interval A has been provided for which we see that $-1.5 \leq y \leq 1.5$, that is, $a = 1.5$. The standard uncertainty u is not given, but can be calculated from the information provided since $p_{c_{\min}}$ occurs when $y = \pm a$; hence, for a normal distribution, measurement uncertainty u is calculable from

$$u = (c - a)/r, \tag{E2.5.1}$$

where r is the guard band multiplier [6] (sometimes called the guard band factor) by which the standard uncertainty has been scaled to obtain the particular conformance probability,

$$r = G^{-1}(p_{c_{\min}}), \tag{E2.5.2}$$

and $G^{-1}(p)$ is the inverse of the cumulative standard normal distribution G.

In Microsoft Excel, r can be evaluated using $r = \text{NORM.S.INV}(p_{c_{\min}})$ for situations where a significant proportion of the PDF lies beyond *only one or other* of the tolerance limits. Otherwise, in situations where the PDF is broad with respect to the tolerance interval, r must be established by other means (for example, UKAS LAB-48 ed 2, appendix D [194])

In this example we find that $c = 2$, $a = 1.5$ and $r = 1.88$; hence $u = 0.266$. As above, this information can be brought into a subsequent uncertainty budget as two distinct distributions $\text{N}(0, u) = \text{N}(0, 0.266)$ and $\text{R}(-a, a) = \text{R}(-1.5, 1.5)$.

Special case 1: interval not centred at zero

For a tolerance interval that is not centred at zero, say $[c_1, c_2]$, with corresponding (co-centred) acceptance interval $[a_1, a_2]$, the uncertainty is instead

$$u = [(c_2 - c_1) - (a_2 - a_1)]/(2r). \tag{E2.5.3}$$

Special case 2: Simple acceptance

Consider the special case when $A = C$, which corresponds to the so-called Simple Acceptance criteria. In this scenario it is usually reported that $p_{c_{\min}} = 50\,\%$. For such a case (where $A = C$, $p_{c_{\min}} = 50\,\%$) we find that u is undefined since $(c - a)/r = 0/0$, that is, there is insufficient information to calculate u; therefore the information is not sufficient to provide metrological traceability.

Note that this simple acceptance scenario ($A = C$) is sometimes misleadingly referred to as 'shared risk', referring to the situation when an accepted value corresponds to the tolerance limit ($y = \pm a$). In fact, this equality of risk is only true for single-sided specifications, or situations where $u \ll c$. In other situations where u is sufficiently large that *both* tails of the PDF have a significant portion outside C, then $p_{c_{\min}} < 50\,\%$ and the risk is no longer 'shared' equally. Fortunately, in those cases (where $A = C$ and $p_{c_{\min}} < 50\,\%$), it is possible to calculate u for a normal PDF from

$$u = \frac{2c}{G^{-1}(p_{c_{\min}} + 0.5)}, \tag{E2.5.4}$$

Alternatively, for a tolerance interval that is not centred on zero, say $[c_1, c_2]$, the uncertainty is instead

$$u = \frac{c_2 - c_1}{G^{-1}(p_{c_{\min}} + 0.5)}. \tag{E2.5.5}$$

In Microsoft Excel $G^{-1}(p_{c_{\min}} + 0.5)$ is given by the cell function NORM.INV($[p_{c_{\min}} + 0.5], 0, 1$).

E2.5.3.4 Information given: Acceptance intervals and a statement about limits of probability or risk of acceptance

This case corresponds to that described in section E2.5.3.3 but without information concerning the tolerance interval C. There is now insufficient information to establish a PDF for the outcome as u is not provided and cannot be calculated from the information given. The information is therefore not sufficient to provide metrological traceability (as it could not be 'imported' into an uncertainty evaluation).

Note that, for *accredited* conformity decisions under ISO/IEC 17025:2017 it is a requirement to define and report the specification (or standard), which usually corresponds to providing C.

E2.5.3.5 Information given: Tolerance and acceptance intervals and a statement about limits of global conformance probability or global risk of acceptance

In certain situations, it is possible that the conformance probability p_c may be presented in terms of *global* risk [6]. Global risk is a measure of the risk associated with future measurements, i.e. measurements that have not yet taken place. Although it is an important quantity in the

evaluation of risk in general quality processes, it is arguably not consistent with the definitions of calibration [89, clause 2.39] or of a metrological traceability chain [89, clause 2.42], being a "sequence of measurement standards and calibrations that is used to relate a measurement result to a reference".

The information needed to implement the approach described in this example is therefore not generally available in such a conformity statement.

E2.5.3.6 Traceability from a statement of conformance to an OIML weight classification

In this example we demonstrate how traceability might be propagated when the available information consists only of an OIML R111-1 [195] weight classification. This example corresponds to the case in section E2.5.3.3 above and is depicted graphically in Figure E2.5.1.

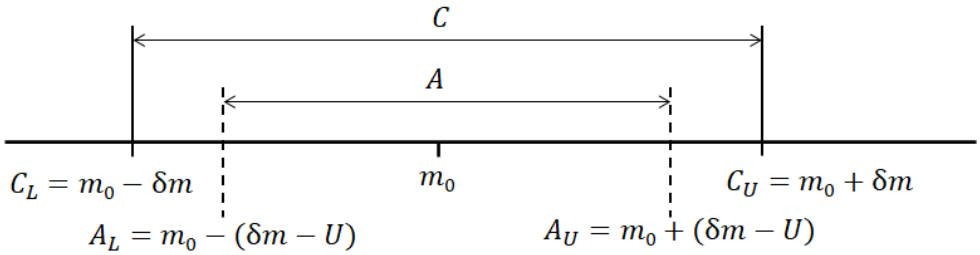


Figure E2.5.1: OIML Guard Band criteria [195]

From OIML R111-1, for each weight, the expanded uncertainty U of the conventional mass shall be less than or equal to one-third of the maximum permissible error: $U \leq \delta m/3$, where U relates to a coverage interval with a 95 % coverage probability.

Also, for each weight, the conventional mass, m_c shall not differ from the nominal value of the weight m_0 by more than the maximum permissible error (δm) minus the expanded uncertainty. The acceptance interval A is defined such that

$$m_0 - (\delta m - U) \leq m_c \leq m_0 + (\delta m - U) \tag{E2.5.6}$$

and the tolerance interval C is defined by $[m_0 - \delta m, m_0 + \delta m]$.

Following the approach described above, the standard uncertainty associated with a classified weight value can be evaluated by combining the standard uncertainties of the PDFs describing the acceptance interval (information about location) and the standard uncertainty associated with dispersion.

For example, for an E_2 class weight of nominal value 2 kg, OIML R111 defines the maximum permissible error δm as [195]

$$\delta m = 3\,\text{mg}. \tag{E2.5.7}$$

The corresponding maximum expanded uncertainty (95 % coverage, assumed normal distribution) is defined as

$$U = \frac{\delta m}{3} = 1\,\text{mg} \tag{E2.5.8}$$

with the related standard uncertainty

$$u_1 = \frac{U}{1.96} = 0.51\,\text{mg}. \tag{E2.5.9}$$

Note that the standard uncertainty characterising dispersion in this and similar scenarios is based upon an upper limit of possible values. In situations where this standard uncertainty is likely to represent a significant contribution to the overall uncertainty it may be appropriate to seek further information.

The limits of the acceptance interval for δm are $\pm a$, where

$$a = \delta m - U = 2\,\text{mg}. \tag{E2.5.10}$$

If A is represented by a rectangular PDF, then the corresponding standard uncertainty is

$$u_2 = \frac{\delta m - U}{\sqrt{3}} = 1.15\,\text{mg}. \tag{E2.5.11}$$

The standard uncertainty u_c associated with the nominal mass value can therefore be evaluated by combining these uncertainties:

$$u_\text{c} = \sqrt{u_1^2 + u_2^2} = 1.3\,\text{mg}. \tag{E2.5.12}$$

More generally, if the expanded uncertainty U is required to be some factor D less than a maximum permissible error δm ($D = 3$ for the example above), and if the coverage probability p is obtained using a coverage factor k_p, then

$$u_1 = \frac{\delta m}{k_\text{p} D} \tag{E2.5.13}$$

and

$$u_2 = \frac{\delta m(1 - 1/D)}{\sqrt{3}}. \tag{E2.5.14}$$

Figure E2.5.2 shows how the standard uncertainty varies with factor D for a maximum permissible error $\delta m = 3\,\text{mg}$.

Note that as the standard uncertainty u_1 *decreases* with increasing D, the overall uncertainty u_c *increases* (due to the proportionately greater contribution corresponding to a). In this situation (where, in use, the value assigned to a weight will be the nominal value) we might perhaps conclude that it is not in the interest of a purchaser for U to be low when the weight is classified, whereas it *is* in the interest of a supplier of weights, as fewer potentially conforming products will be rejected.

Note also that the PDF associated with u_c is not normal since it results from the convolution of normal and rectangular distributions. However, provided that u_2 is not a dominant quantity in the budget into which it is subsequently imported, the shape of the corresponding output PDF would be approximately normal.

Comment on ISO/IEC 17025:2017 Annex A.2.3

Those readers familiar with ISO/IEC 17025:2017 [7] and in particular Annex A.2.3 might perhaps interpret that (informative) Annex to suggest that metrological traceability can be obtained from a rectangular PDF with limits corresponding to the tolerance interval $C = [-c, c]$. Annex A.2.3 cites "The use of OIML R 111 class weights to calibrate a balance", which might be interpreted as an example of that practice.

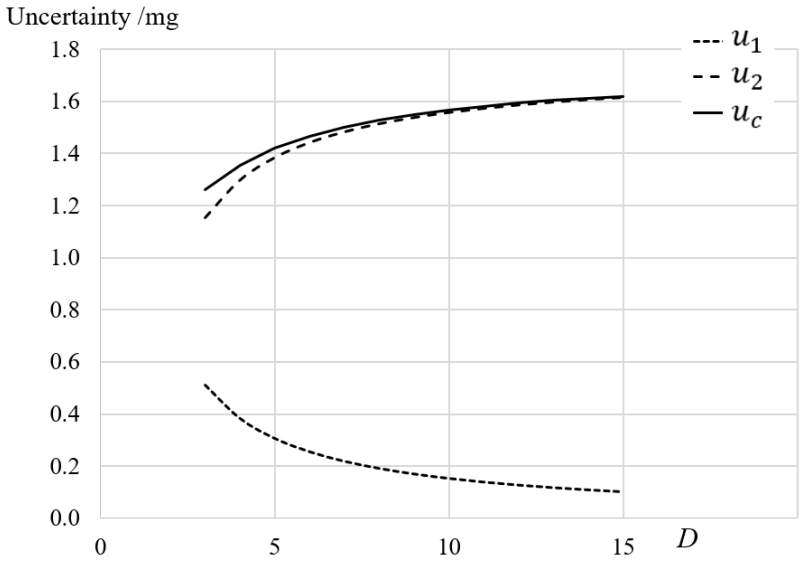


Figure E2.5.2: Maximum permitted standard uncertainty u_1 (E2.5.13), standard uncertainty u_2 corresponding to the acceptance interval (E2.5.14), and the combined standard uncertainty u_c (E2.5.12) for a range of values of $U = \delta m/D$

However, such an approach does not make best use of the information available: whereas the approach described in this example correctly converges to the appropriate 'combined' PDF in the limits of small and large u_1/c, the A.2.3 interpretation as stated above employs a rectangular PDF throughout, even though the normal distribution becomes proportionately more significant as u_1/c increases. As a consequence, that particular interpretation of A.2.3. can significantly overstate the uncertainty. This overstatement is demonstrated in Table E2.5.1 where, for the particular scenario given, we see that the difference between standard uncertainty evaluations can exceed 50 %.

Table E2.5.1: Comparison between standard uncertainty evaluations obtained for the interpretation of ISO/IEC 17025 A.2.3 described above, identified as "u_a", and evaluations based upon the two-PDF approach described in this example, identified as "u_b". PDFs are centred at zero. Estimates are for model data over a range of tolerance intervals $[-c, c]$ and corresponding acceptance intervals $[-a, a]$ with $u_1 = 1$ and guard band $w = c - a = 2u_1$.

c/mg	a/mg	u_a/mg	u_b/mg
1000	998	577	576
10	8	5.8	4.7
9	7	5.2	4.2
8	6	4.6	3.6
7	5	4.0	3.1
6	4	3.5	2.5
5	3	2.9	2.0
4	2	2.3	1.5
3	1	1.7	1.2
2	0	1.2	1.0

For the OIML E2 – 2kg weight discussed in this example, $c = \delta m = 3\,\text{mg}$ and $w = U = 1\,\text{mg}$; hence $u_1 = 0.51\,\text{mg}$, yielding $u_a = 1.7\,\text{mg}$ and $u_b = 1.3\,\text{mg}$, a difference of nearly 40 % (based on unrounded data).

E2.5.3.7 Calibration and verification of a caliper according to Geometrical Product Specification (GPS) standard ISO 13385-1:2019

In this example, we demonstrate how traceability can be obtained from a verification statement for an instrument certified under a Geometrical Product Specifications (GPS) standard.

In general, laboratories accredited for the calibration of calipers adopt the GPS standard ISO 13385-1 [196]. According to this standard — which includes requirements for test methods, default values for maximum permissible errors (maximum permissible errors (MPEs)) and related decision rules — laboratories are variously required to provide two different uncertainty evaluations: one for the measured calibration values of the instrument, u_{cal}, and the other for its verification 'test uncertainty' u_{test} as defined in ISO 14253-5:2015) [197].

Certificates that meet the requirements of ISO/IEC 17025:2017 [7] concerning the reporting of calibration results (variously described in clauses 7.8.4, 7.8.6 and A.2.3 of that standard) could present the information in various forms, as considered in the following scenarios:

1. Calibration certificate containing indication errors with the associated calibration standard uncertainty u_{cal},
2. Calibration certificate containing statement of conformity with a specification (MPE), associated test standard uncertainty u_{test} and decision rule, without indication errors (consistent with paragraph E2.5.3.3);
3. Calibration certificate containing only a statement of conformity with a specification (expressed in terms of MPE) with no reported indication errors, calibration measurement uncertainty or test verification uncertainty, but with a GPS decision rule that has somehow accounted for these quantities (consistent with paragraph E2.5.3.3).

Note that the purpose of *calibration* is to establish a traceable link to the SI, whereas the purpose of *verification* is only to decide conformity with a specification. In the case of specifications such as those represented by the GPS standards, the calibration standard measurement uncertainty u_{cal} (scenario 1) is therefore different from the test verification uncertainty u_{test}.

In the case of GPS, a specification is defined in terms of limits (MPE) that somehow already account for various influence quantities such as repeatability and resolution that would normally be incorporated into a calibration uncertainty evaluation. The evaluation of test verification uncertainty therefore does not include these quantities and is therefore less than the calibration measurement uncertainty (for further details see ISO 14978:2018 [198, Annex D]). In principle however, all relevant influence quantities are present and, if combined correctly, the test verification uncertainty and information represented by the specification can be used to provide an evaluation of calibration measurement uncertainty required for dissemination of measurement traceability.

Scenario 1

In this straightforward scenario, the calibration of the caliper produces indication errors with associated calibration measurement uncertainty. Following best practice, the errors can be corrected and the calibration measurement uncertainty can be propagated through the measurement chain.

[167], the GUM suggests a method to achieve this [2, clause F.2.4.5].

Scenario 2

In this case no quantitative information on the indication errors is available other than their values being within specification limits. In the absence of any other information, the best estimate of the error is therefore zero.

The uncertainty when in use by the customer, can be evaluated from the PDF resulting from the convolution of a normal probability distribution $\mathrm{N}(0, u_{\mathrm{test}})$ and a rectangular distribution $\mathrm{R}(-a, a)$, where

$$a = \mathrm{MPE} - k u_{\mathrm{test}} \tag{E2.5.15}$$

and, for a two-sided specification, k can be calculated iteratively by applying equation (11) of JCGM 106:2012 [6], given the values of p_{c} and u_{test}.

For example, figures E2.5.3 and E2.5.4 present the PDFs for conformance probability values p_{c} equal to 50 % and 95 % respectively, for a range of values for measurement capability index C_{m}, where

$$C_{\mathrm{m}} = \frac{2 \times \mathrm{MPE}}{4 u_{\mathrm{test}}}. \tag{E2.5.16}$$

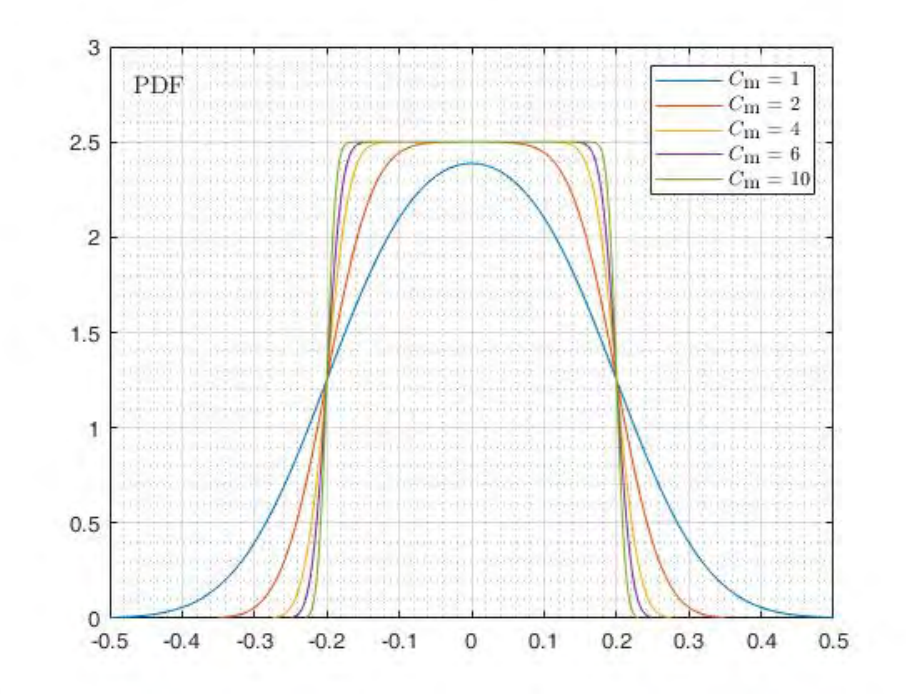


Figure E2.5.3: PDFs for $p_{\mathrm{c}} = 50\,\%$, MPE $= 0.2$ and various C_{m} values

Once the PDF is established, its standard deviation u can be determined, for example by combining variances:

$$u^2 = \frac{(\mathrm{MPE} - k u_{\mathrm{test}})^2}{3} + u_{\mathrm{test}}^2. \tag{E2.5.17}$$

Assuming that the specification limits and u_{test} account correctly for all influence quantities that contribute to the calibration of the caliper (as is the premise of the GPS standard), then u corresponds to the calibration standard measurement uncertainty (that is, $u_{\mathrm{cal}} = u$).

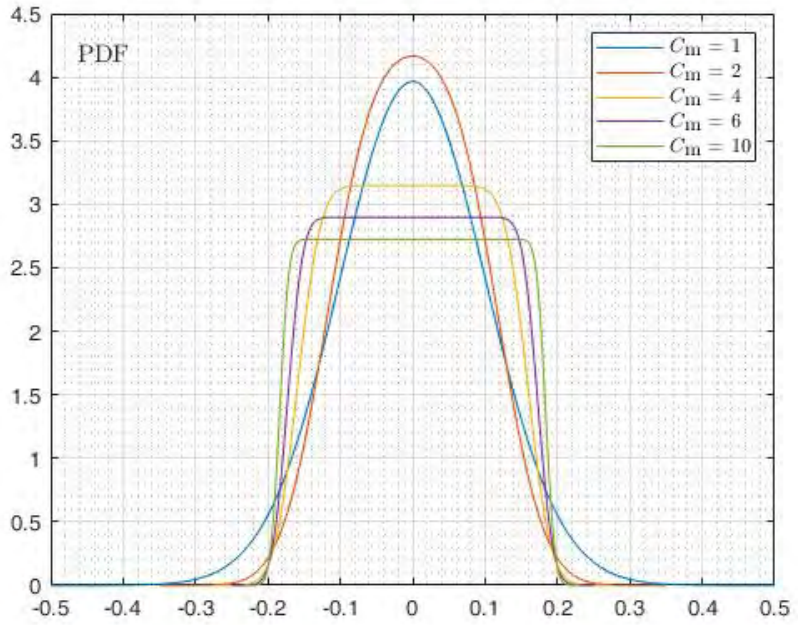


Figure E2.5.4: PDFs for $p_c = 95\,\%$, MPE $= 0.2$ and various C_m values

Note that in general the standard uncertainty alone provides insufficient information for propagation of measurement results; knowledge of the PDF is needed, for example, whether it can be described by a known distribution such as a normal distribution. In cases where the shape of the PDF is dominated by the specification it will be more 'rectangular' than normal. In that case some other means of conveying information about the PDF is needed, such as in figures E2.5.3 and E2.5.4 or as might be provided by using a numerical approach for evaluating the uncertainty [3].

A standard uncertainty u_{MPE} associated with the specification can also be established by taking it to be the standard deviation of a rectangular PDF with limits ±MPE:

$$u_{\mathrm{MPE}} = \frac{\mathrm{MPE}}{\sqrt{3}}. \tag{E2.5.18}$$

In some situations, u_{MPE} is used as an estimate of the calibration standard measurement uncertainty.

Figure E2.5.5 shows in the ordinate the standard deviation u of the PDF divided by u_{MPE} and in the abscissa C_m for various p_c values.

Note that for p_c values greater than 85 %, u_{MPE} provides a conservative overestimate of u_{cal} for all C_m values considered.

It is also clear that it is not possible to ensure values of p_c higher than 95 % for low values of C_m. As an example, considering the curve corresponding to the probability of 99 %, the smallest value of C_m that allows this probability is about 1.3.

For p_c values lower than 85 %, the use of the u_{MPE} can lead to an understatement of measurement uncertainty for low C_m values, which happens for example in the case of p_c equal to 70 % and for C_m values lower than about 1.6.

For p_c values of 50 %, in order not to understate uncertainty, it is possible to multiply u_{MPE} by a 'safety factor' which can be determined by the graph for C_m less than 4. For higher C_m values the underestimation of the uncertainty is less than 1 % and therefore may not be significant.

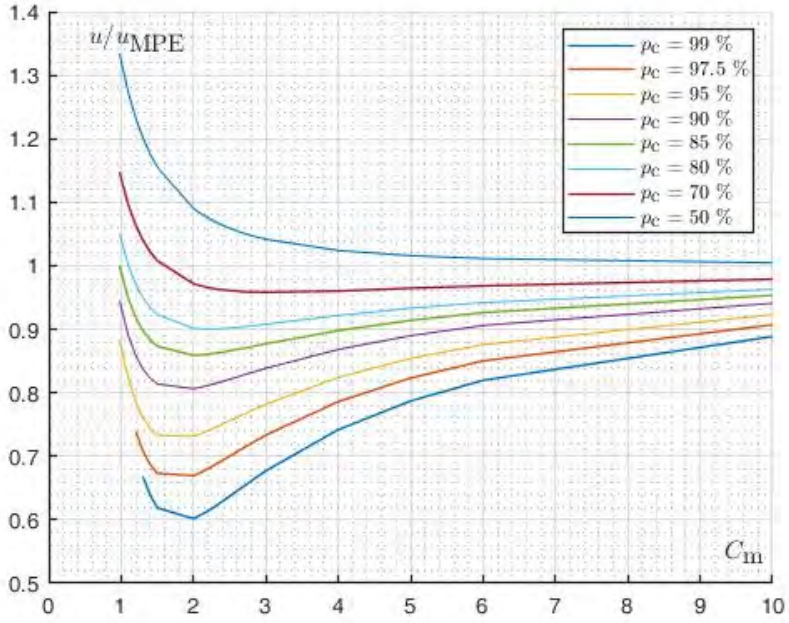


Figure E2.5.5: u/u_{MPE} as a function of C_{m} for various conformance probabilities p_{c}.

Scenario 3

In this case, no quantitative information on the indication errors and test uncertainty is available. The decision rule comes from ISO 13385-1:2019 [196]. However, the standard provides two different rules depending on the agreement with the customer:

Decision rule A If no decision rule is stated with the specifications, and no special agreement is made between supplier and customer, then the default rule of ISO 14253-1 [199] applies (ISO 13385-1 [196, clause 6.3]). In this case the default conformance probability limit is $p_{\mathrm{c}} = 95\,\%$, which corresponds to a false acceptance probability less than or equal to 5 %. This information, combined with the MPE, can be used to evaluate the uncertainty to be attributed to the instrument when used by the customer, ensuring traceability. From this information (MPE and p_{c}), assuming that the distribution associated with the test uncertainty is normal, it is possible to provide an upper limit of standard uncertainty u_{max}. This value can be calculated from ISO 14253-1 [199, annex A, figure A.3], considering the most conservative condition with the ratio of the specification and the test uncertainty equal to 3.92, which corresponds to $C_{\mathrm{m}} = 0.98$:

$$\frac{2 \times \mathrm{MPE}}{u_{\mathrm{max}}} = 3.92. \tag{E2.5.19}$$

This case is equivalent to scenario 2 with $p_{\mathrm{c}} = 95\,\%$ and $C_{\mathrm{m}} = 0.98$ (see figure E2.5.5).

Decision rule B If there is an agreement with the customer to verify the caliper with respect to the MPE values reported in [196, table B.1], the decision rule that applies shall be simple acceptance, with the measurement capability index C_{m} being four or larger (ref. annex B of ISO 13385-1) [196]. Although the test uncertainty is not reported in the calibration certificate, it is possible to calculate a limit value of uncertainty u_{max} from the limit value of $C_{\mathrm{m}} = 4$:

$$u_{\mathrm{max}} = \frac{\mathrm{MPE}}{2C_{\mathrm{m}}} = \frac{\mathrm{MPE}}{8}. \tag{E2.5.20}$$

This case is equivalent to scenario 2 with $p_c = 95\,\%$ and $C_m = 4$ (see figure E2.5.5).

E2.5.4 Additional notes and comments

E2.5.4.1 Notes on risk in relation to uncertainty

The GUM is concerned with the propagation of PDFs (or in the case of the LPU their variances).

Estimates of risk are given by integrals over certain ranges of a PDF (or a joint PDF) as described in [6]. They are not in a form that is directly propagated using GUM methodology.

Risk is usefully evaluated at times when a decision is needed concerning the acceptability of a result, normally at the *end* of a measurement chain. It may be of *passing* interest at intermediate points in the chain, but for propagation of traceability it is the underlying PDF that is of interest.

Therefore a statement of conformity and risk is generally *not* a useful alternative to a description (or summary) of the PDF for the measurand. An accredited laboratory, for example, would be expected to ensure that customers are aware of this lack of utility when their customers request a statement of conformity.

E2.5.4.2 Notes on Simple Acceptance

It is worthwhile re-iterating the point that assertions of conformity based upon Simple Acceptance criteria *on their own*, with no account for measurement uncertainty whether it be direct or indirect, are *not* sufficient to provide traceability (or to define a meaningful decision rule, as the associated risk is *undefinable*).

Further, it is not possible to take indirect account of uncertainty by simply stating the value of the uncertainty *after* the decision is made, which corresponds to a situation in which the decision is made regardless of uncertainty or risk at the time the decision is being made.

E2.5.4.3 Single sided specifications

The examples presented here have all been presented in terms of two-sided specifications that define an upper and a lower limit for the measurand. A key point is that such specifications allow a rectangular PDF to be established to describe the location of the quantity of interest, which would not be possible with a truly single-sided specification for which no such PDF can be established, there being only one defined limit.

E2.5.5 Conclusions

The examples presented here have demonstrated various situations in which there is no explicit statement of a measurement result (in terms of a specific value and associated uncertainty), yet metrological traceability can be obtained from a statement of conformity together with a suitable specification and decision rule. (Such a situation is anticipated in [7].) Making optimum use of available information to establish metrological traceability is demonstrated for several general scenarios and is illustrated with two extended examples. The process recommended involves identifying two or more independent PDFs to represent the information that has been

provided. Typically this approach results in a PDF that characterises the location of possible quantity values for the measurand, and a PDF that characterises the dispersion of possible values around any given value of the measurand. In practice, these PDFs are likely to have the well-known rectangular and normal distributions respectively and can be individually 'imported' into uncertainty evaluations based upon the LPU or Monte Carlo Simulation as independent input quantities.

Part E3

Environment

Example E3.1

Evaluation of measurement uncertainty in average areal rainfall – uncertainty propagation for three methods

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

E3.1.1 Summary

Precipitation measurement has diverse applications in contexts such as hydrology, meteorology and climatology, and is of increasing importance for the assessment of climate change, both as an indicator and as a parameter applied in modelling aiming at the interpretation of climatological phenomena and forecasting. To obtain the quantities of concern (precipitation and rainfall intensity) there are several methods, for which it is relevant to determine the measurement uncertainty associated with an estimate of the quantity as a comparative element in order to relate its magnitude to the intended evaluation. For a long-term analysis, several methods are available for calculating the accumulated values of the precipitation quantity and the average values observed in given catchment areas and in certain time intervals. However, it is not common to promote information about the impact that the selection of method has on the results. This selection is one of the main objectives of the comparative analysis proposed in this study, that is, the difference that results from this selection regarding the estimate of the quantity and its associated uncertainty. In addition, the examples given are used to illustrate the adequacy of the approaches recommended by the Guide to the expression of Uncertainty in Measurement (GUM) and GUM Supplement 1 (GUM-S1) .

E3.1.2 Introduction to the application

The growing awareness of the impact of climate change and the United Nations' sustainable development goals [200] show the need to have reliable measurements of quantities to support urban and water resources management. Precipitation and rate of rainfall (or rainfall intensity) are quantities widely measured, being applied in hydrology, climatology and meteorology, namely, in modelling, studying patterns and in forecasting. The definition of precipitation according to the World Meteoroligical Organization (WMO) [201] is the following:

> Precipitation is a quantity defined as the liquid or solid products of the condensation of water vapour falling from clouds, in the form of rain, drizzle, snow, snow grains, snow pellets, hail and ice pellets; or falling from clear air in the form of diamond dust.

From this definition (being the quantity interpreted as the mass or volume of the liquid or solid products), precipitation intensity is a quantity defined as the amount of precipitation collected per unit time interval. The unit of precipitation is linear depth in mm (corresponding to a volume per area) and for liquid precipitation, $\mathrm{kg\,m^{-2}}$ (corresponding to a relation of mass per unit area). The difference between rainfall and precipitation is that rainfall is related to water in its liquid state in the form of precipitated condensed droplets from atmospheric water vapour, while precipitation is related to the product of the condensation of atmospheric water vapour that falls under gravity. The measurement unit of rainfall intensity is linear depth per hour ($\mathrm{mm\,h^{-1}}$). Rainfall intensity is usually obtained at a time intervals of 1 min, being lower in case of extreme events or systems with high variability or intensity.

In practice, rainfall and precipitation are measured in different geographical locations, in order to model the behaviour of meteorological and climatic phenomena in a spatial and temporal dynamic regime. Observation of these quantities can be strongly affected by influence conditions, namely, exposure, wind and topography, thus limiting the accuracy of measurement. Wind effect is critical for the performance of instrumentation, leading to different shapes of gauges as seen in figure E3.1.1 [201], illustrating how streamlines of wind deformation are expected to affect the trajectory of precipitation particles, promoting a relevant error contribution to the measurement.

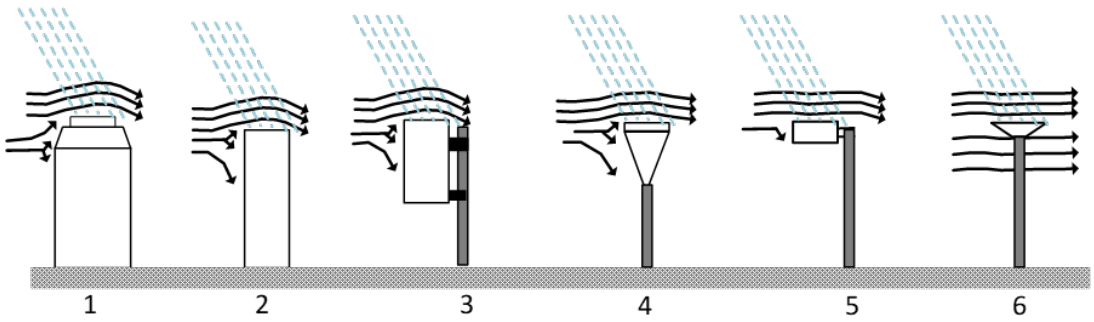


Figure E3.1.1: Different shapes of gauges induce the way that streamlines of wind deformation affect the trajectory of precipitation particles [201]

The recommended use of recording precipitation gauges is to have sufficient information related to the time scale and time resolution to cater for the high variability of the precipitation intensity. Such information is used in the technical process of reducing evaporation and wetting losses as sources of error and uncertainty that can affect significantly non recording devices. Three types of automatic precipitation recorders to measure rainfall are commonly used: weighing-recording type; tipping-bucket type; and floating type. The study carried out considered the use of the weighing and tipping-bucket types, for which data were available. There are many types of instruments and methods to take observations of rainfall intensity [202]. The study carried out considered the use of the weighing and tipping-bucket types, both of which collect precipitation using an orifice and a funnel directed into receptacles allowing the volume to be weighed, in the first case, or multiples of volumes collected in a pair of buckets each having a reference volume quantity per second.

The WMO establishes reference conditions for the installation and use of these gauges [203], namely, the orifice height above the ground (commonly between 0.5 m and 1.5 m), the conditions of the surrounding ground to avoid in-splashing from the ground and specific geometries adopted for the orifice and the gauge. Wind field in the surroundings can be a major influence on the measurement. Special care is highly recommended by including in the setup of windshields, by establishing the type of surrounding surface (for example, short grass instead of hard and flat surfaces to avoid in-splashing), by adopting the suggested relations with the vertical angle obstacles in the surrounding of the site, and by choosing an appropriate gauge size and shape in order to minimize the wind effect [201].

To obtain the data needed for meteorology, climatology and hydrology predictive models, networks of stations are distributed in areas of interest in a way that properly represents the distribution of rainfall, being required to obtain rainfall intensity measurements at single points and combine them in order to calculate the volume of precipitation that falls over a given catchment area [204]. To achieve this aim, there are several methods that use a set of single point measurements to obtain the average areal rainfall.

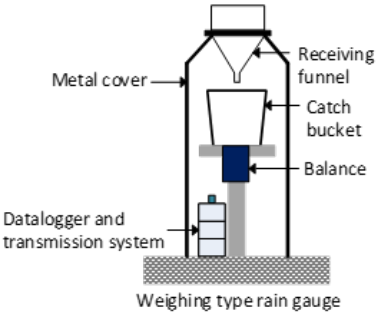
For this study the following were considered: arithmetic mean method, Thiessen polygons' method and isohyetal method. These methods use distinctive interpretations of the physical quantity in relation to the geometric context. For this reason, uncertainty in the common output plays a relevant role, namely, because that allow the comparison of the accuracy of the methods. A brief description of each method for estimating the average areal rainfall is presented:

a) Arithmetic mean method: evaluates the arithmetic mean of considered single point observations for a certain area;

b) Thiessen polygons' method: applies a graphical approach that defines relative polygonal areas related to each single station observations, providing a weighted sum of the observations;

c) Isohyetal method: applies a graphical approach based on the drawing of isohyet lines of equal rainfall, combining the observations weighted by the coverage areas between these isohyetal lines.

In this framework, two steps are required to obtain the average areal rainfall: first, the evaluation of the measurand (rainfall intensity) at each single point (gauge station); and second, the propagation of these uncertainties to the combined output (average areal rainfall).

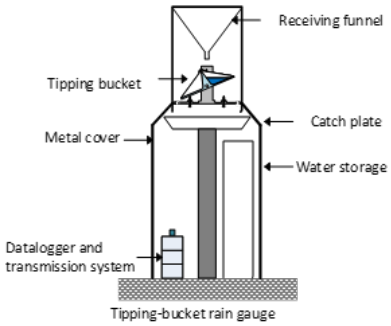


(a) Elements of a weighing gauge

(b) Weighing gauge installed in field

Figure E3.1.2: Weighing gauge

Weighing-recording gauges effectively are balances that record the weight of accumulated precipitation volume over time. The container should have a large capacity considering that this system does not empty itself. Solutions for its use in harsh climate conditions are also required (for example, use of oil to avoid large evaporation effects and antifreeze solutions).



(a) Elements of a tipping-bucket gauge

(b) Tipping-bucket gauge installed in field

Figure E3.1.3: Tipping-bucket gauge

The operation of tipping-bucket gauges allows water collection and guidance to a twin bucket balance with both parts having an equal weight and reference volume. Every time a bucket is full the balance changes position within a pivot axis, and the other bucket moves into position to collect water while the first one will empty the collected water. In this process, the time between each change (tipping) can be measured, allowing the calculation of the rate of rainfall quantity. Tipping-bucket gauges employ a contact closure (reed switch or relay contact), such that each tip produces an electrical impulse as a signal output. This output must be recorded by a data logger or an ADC (data acquisition system equipped with reed switch reading ports). This mechanism provides a continuous measurement without manual interaction.

The rainfall intensity is an intermediate measurand that depends on the height, h in cm (the observation is usually expressed in mm), being related with the volume, V in cm^3. Although not used explicitely in this work, the collected water volume V is given by expression (1) being related with the ratio of the mass content, m in g, of the bucket by the density of water, ρ of $1\,\mathrm{g\,cm}^{-3}$, or by the ratio of the height, h, and the area of the surface of the collector, S in cm^2.

E3.1.3 Specification of the measurands

This study employs a two-step approach. The first step is the observation of rainfall intensity, P_i, at several locations, i. The second step is the combination of observations from those locations to obtain the average areal rainfall.

Considering the use of the weighing recording gauges (figures E3.1.2a and E3.1.2b) and tipping-bucket gauges (figures E3.1.3a and E3.1.3b), two types of transduction of the quantity are applied. In the first case the collected amount of water generates weighing observations (m) over time while, in the second case, impulses are generated for a fixed volume (V).

$$V = \frac{m}{\rho} = h \cdot S. \tag{E3.1.1}$$

The rainfall intensity for non-corrected tipping-bucket gauges is based on the number of tips, n in a periodic sampling rate Δt (typically 6 s or 10 s) and averaged over a chosen time interval (for example, 1 min). The estimate of the rainfall intensity is

$$P_i = \frac{(n \cdot h)}{\Delta t}. \tag{E3.1.2}$$

The primary measurand, P_{av}, is the average areal rainfall as a function of the rainfall intensity values or averages obtained at the k locations of measurement stations, being the function a weighted linear combination of the P_i:

$$P_{\text{av}} = f\,(P_1, \ldots, P_k) = \sum_{i=1}^{m} w_i P_i. \tag{E3.1.3}$$

This function is applied to all the methods studied, the arithmetic mean method, the Thiessen-polygons' method and the isohyetal method.

E3.1.4 Measurement models

The nature of the measurement of precipitation, being affected by many natural conditions, implies the need to account for corrections and to evaluate the effect of errors in the methods [205, 206]. Reports issued by WMO point out the need to use models to adjust the measured precipitation [207], based on corrections obtained from statistical data. Regarding errors (systematic effects), WMO also collected information provided by research, being able to state [203]: "The amount of precipitation measured by commonly used gauges may be less than the actual precipitation reaching the ground by up to 30 % or more.". Considering the interest of this study in the rainfall intensity measurements obtained using tipping-bucket gauges and weighing gauges (other types like floating gauges and optical gauges were not considered for this purpose), data provided in the WMO reports [202, 208] were taken into account.

The assessment of errors in precipitation measurement usually relates its origin to the effects of wind, wetting and evaporation losses [207]. A general description of these sources is given in [203], including estimates based on [207], is the following:

(a) Error due to systematic wind field deformation above the gauge orifice: typically 2 % to 10 % for rain and 10 % to 50 % for snow;

(b) Error due to the wetting loss on the internal walls of the collector;

(c) Error due to the wetting loss in the container when it is emptied: typically 2 % to 15 % in summer and 1 % to 8 percent in winter, for (b) and (c) together;

(d) Error due to evaporation from the container (most important in hot climates): 0 % to 4 %;

(e) Error due to blowing and drifting snow;

(f) Error due to the in- and out-splashing of water: 1 % to 2 %;

(g) Systematic mechanical and sampling errors, and dynamic effects errors (i.e. systematic delay due to instrument response time): typically 5 % to 15 % for rainfall intensity, or even more in high-rate events (see [208]);

(h) Random observational and instrumental errors, including incorrect gauge reading times.

Considering these sources of error and uncertainty, a functional relation for precipitation (rain and snow contributions) was proposed by WMO [203], adapted in 1990 by Legates and Willmott [209] as

$$P_k = k_\mathrm{r} P_\mathrm{cr} + k_\mathrm{s} P_\mathrm{cs},$$

where

$$\begin{aligned} P_\mathrm{cr} &= P_\mathrm{gr} + \Delta P_{1\mathrm{r}} + \Delta P_{2\mathrm{r}} + \Delta P_{3\mathrm{r}} + \Delta P_{4\mathrm{r}}, \\ P_\mathrm{cs} &= P_\mathrm{gs} + \Delta P_{1\mathrm{s}} + \Delta P_{2\mathrm{s}} + \Delta P_{3\mathrm{s}} + \Delta P_{4\mathrm{s}}. \end{aligned} \tag{E3.1.4}$$

The quantities in expression (E3.1.4) are as follows:

subscript r – relates to liquid precipitation "rain";
subscript s – relates to "solid" precipitation;
P_k – adjusted precipitation;
k – adjustment factor for the effects of wind field deformation;
P_c – amount of precipitation caught by the gauge collector;
P_g – measured amount of precipitation in the gauge;
ΔP_1 – adjustment for the wetting loss on the internal walls of the collector;
ΔP_2 – adjustment for wetting loss in the container after emptying;
ΔP_3 – adjustment for evaporation from the container; and
ΔP_4 – adjustment for systematic mechanical errors.

The adjustment factor k is a variable obtained from studies developed by Nešpor and Sevruk [210], in which the ratio of correct to measured precipitation or rain and snow was studied using two unshielded gauges in different weather conditions of wind speed and intensity. The measurement of rainfall intensity, in $\mathrm{mm\,h^{-1}}$ units, using weighing-recording gauges or tipping-bucket gauges, starts respectively with the measurement of mass or volume in units of time. The measurand is affected by sources of uncertainty according to the relational function presented in equation (E3.1.4), considering only the liquid contributions,

$$P_k = k_\mathrm{r} P_\mathrm{cr} = k_\mathrm{r} \left(P_\mathrm{gr} + \Delta P_{1\mathrm{r}} + \Delta P_{2\mathrm{r}} + \Delta P_{3\mathrm{r}} + \Delta P_{4\mathrm{r}} \right). \tag{E3.1.5}$$

The measurement uncertainty related to the output quantity of this function can be evaluated using the conventional LPU [2] or using a MCM as described in GUM-S1 [3].

As mentioned above, single location measurements are a first step to obtain information about precipitation in a certain area, thus requiring the estimation of the average areal rainfall, calculated using the mentioned common methods to achieve this purpose.

The primary measurand, P_av, is the average areal rainfall as a function of the rainfall intensity values or averages obtained at the m locations of measurement stations, being a function of a weighted linear combination of the P_i:

$$P_\mathrm{av} = f\left(P_1, \ldots, P_m\right) = \sum_{i=1}^{m} w_i P_i. \tag{E3.1.6}$$

The function (E3.1.6) is considered generic, being applied to all three methods studied. In a general approach the methods consider that there are m measurement stations able to obtain rainfall values $P_i (i = 1, \ldots, m)$ distributed across a basin. To illustrate the approach adopted in each method, as a starting point consider $m = 4$ measurements at rainfall stations located in a basin, as in figure E3.1.4.

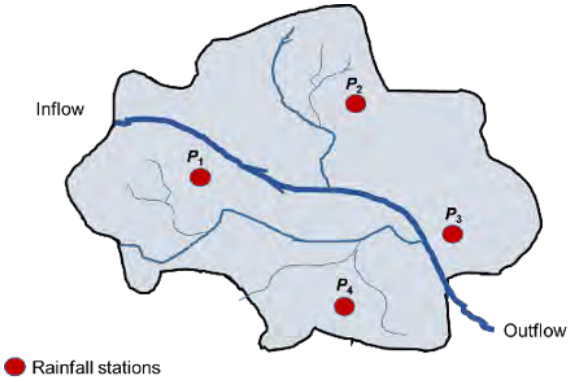
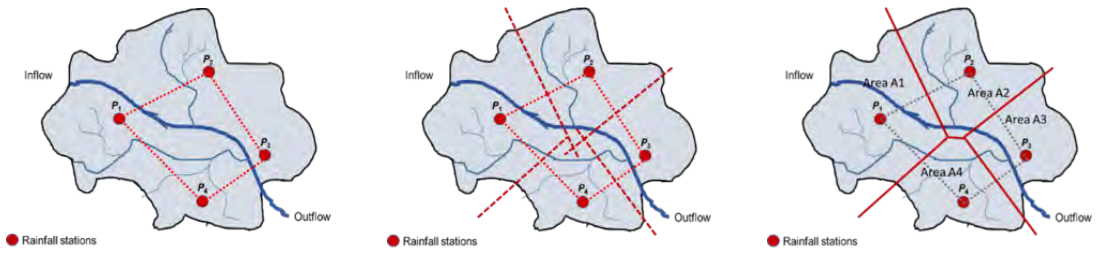


Figure E3.1.4: Location of four rainfall stations in a basin

The arithmetic-mean method evaluates the average of the estimates obtained at each location without establishing a relation between the position of the stations and the geometry of the area of observation. The average areal rainfall (E3.1.6) considers that all weights are equal to $1/m$

$$P_{\text{av}} = \sum_{i=1}^{m} w_i P_i = \frac{1}{m} \sum_{i=1}^{m} P_i = \frac{1}{m} (P_1 + \cdots + P_m). \tag{E3.1.7}$$

The second method studied is the method of Thiessen polygons, which uses a given set of locations in the plane to make a partition of the plane into convex polygons (Voronoi tessellation), each of which comprises the points closest to one of the given locations. It uses a geometric division of the space that can be explained in three steps: (1) connect the rainfall measurement locations by straight line segments (figure E3.1.5a), (2) draw perpendicular bisectors to these segments (figure E3.1.5b), and (3) divide the area using polygons (figure E3.1.5c).



(a) Step 1 (b) Step 2 (c) Step 3

Figure E3.1.5: Steps 1, 2 and 3 of the geometric approach in Thiessen polygons' method

The average areal rainfall (equation (E3.1.8) (8)) using the Thiessen polygons' method, P_{axty} is given by a weighted approach to the arithmetic mean:

$$P_{\text{av,Tp}} = \sum_{i=1}^{m} w_i P_i = \sum_{i=1}^{m} \frac{A_i}{A} P_i, \tag{E3.1.8}$$

where w_i denotes the weights given by the relative areas of the polygons obtained, A_i is the area of the polygon related to station i, and A is the total area of the basin. Any change of rainfall at the stations does not affect the geometry of the polygons.

The third method studied uses weights proportional to contour map areas according to the location of isohyets (lines on a map or chart connecting areas of equal rainfall). The isohyetal method (figure E3.1.6) uses the single-point station information to establish contour map areas [211, 212], with weights obtained by multiplying each contour area by the average rainfall in the area.

In the example (where $m = 4$), using the same basin and considering locations, $m+1$ isohyets $(P_1, \ldots, P_{m+1})$ are defined (P_a, P_b, P_c, P_d, P_e in figures E3.1.4–E3.1.6) dividing the basin into m contour areas $(A_1, \ldots, A_m)$, allowing the determination of the areal rainfall average, $P_{\text{av,isoh}}$, using

$$P_{\text{av,isoh}} = \sum_{i=1}^{m} w_i P_i = \sum_{i=1}^{m} A_i \bar{P}_l = \frac{1}{A} \sum_{i=1}^{m} A_i \left(\frac{P_i + P_{i+1}}{2} \right). \tag{E3.1.9}$$

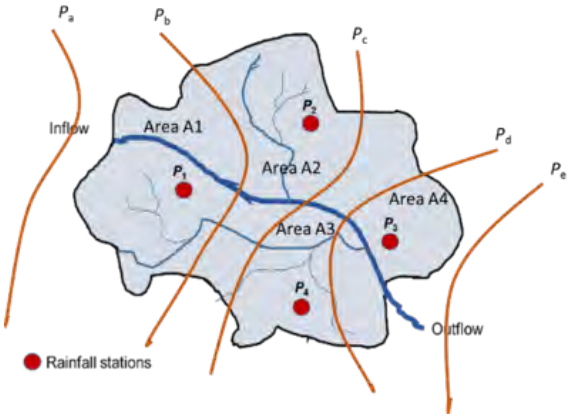


Figure E3.1.6: Defining isohyets and contour areas to apply the isohyetal method

E3.1.5 Uncertainty propagation

The propagation of uncertainty in this study of the average areal rainfall intensity has two stages. The first stage requires the evaluation of the uncertainty related to the measurement of rainfall intensity in the individual locations, using either weighing-recording gauges or tipping-bucket gauges. The second stage requires the evaluation of the uncertainty of average areal rainfall intensity using one of the three methods mentioned.

Given the nature of precipitation phenomena and the variability inherent in the main sources of measurement error, the quantification of the resulting effects is usually difficult to establish. For the purpose of obtaining estimates of these quantities, references [213–218] were consulted. In the first stage, for both type of gauges, the input quantities are described in expression (E3.1.2). A probability density function (PDF) was assigned to each quantity based on knowledge of that quantity. The mean of that PDF was used as an estimate of the quantity and the standard deviation as the associated standard uncertainty. The input quantities are shown in table E3.1.1 together with the assigned PDFs and their relative standard uncertainties (in the table the index "r" was suppressed).

Since the PDFs for wetting loss are not symmetrical about zero, a correction was made to the estimate equivalent to the half width of the PDF and a zero-centred PDF was used in the uncertainty evaluation.

The second stage accounts for the uncertainty associated with the measurement of rainfall intensity at each location, for both types of gauges considered, having as input the combined average areal rainfall uncertainty obtained in stage 1. In this case, the evaluation of measurement uncertainty does not account for possible correlation between measurements at the different locations.

Table E3.1.1: Input quantities, relative standard uncertainties and assigned PDFs related to rainfall intensity measurement using weighing gauges and tipping-bucket gauges

Quantity	Description	Standard uncertainty/%		PDF
		Weighing gauge	Tipping-bucket gauge	
k	Error due to systematic wind field deformation above the gauge orifice	5	5	Gaussian
P_{g1}	Error due to the in- and out-splashing of water	2	2	Uniform
P_{g2}	Random observational and instrumental errors	2	2	Uniform
ΔP_1	Error due to the wetting loss on the internal walls of the collector	n/a	5	Uniform*
ΔP_2	Error due to the wetting loss in the container when it is emptied	1	5	Uniform*
ΔP_3	Error due to evaporation from the container	2	1	Uniform
ΔP_4	Systematic mechanical and sampling errors, and dynamic effects errors	5	10	Uniform

* The uniform PDFs adopted have intervals from 0 % to 5 % (asymmetric with respect to 0 %), considering that loss quantity would increase the estimate and negative values are not achievable.

E3.1.6 Reporting the results

The evaluation of the uncertainty for a measurement of rainfall precipitation of $10\,\mathrm{mm\,h^{-1}}$, was made using R/RStudio programming [11,49], with 1×10^6 Monte Carlo trials for each calculation. The evaluation allowed, for both gauges, the PDF for the output quantity, P_k, and the relative expanded uncertainty, $U_{0.95}(P_k)$, for a confidence interval of 95 % by applying GUM [2] and GUM-S1 [3], to be provided. The values obtained and the related PDFs are shown in table E3.1.2 and in figures E3.1.7 and E3.1.8 for the weighing gauge and tipping-bucket gauge, respectively. These figures also show the scaled histograms produced using GUM-S1 and used as a basis for the (continuous) PDFs shown in blue and obtained using the mean and standard deviation values as parameters of the normal distributions.

The results show consistency with the normal distribution in the case of the weighing gauge and a small deviation from normality in the case of tipping-bucket gauge, identified by skewness and kurtosis parameter values that differ slightly from normal reference values of 0 and 3, respectively.

For the second stage, a comparison of the measurement uncertainty of the average areal rainfall is made for the three methods considering that local measurement was made using either weighing gauges or tipping-bucket gauges.

For this study, figure E3.1.4 was adopted as being representative of a certain territory having four measurement stations with the following daily average rainfall: $P_1 = 12\,\mathrm{mm}$, $P_2 = 18\,\mathrm{mm}$, $P_3 = 36\,\mathrm{mm}$ and $P_4 = 28\,\mathrm{mm}$. The relative expanded measurement uncertainty considered for each estimate was based in the evaluation obtained at stage 1, respectively, 12 % for weighing gauges and 16 % for tipping-bucket gauges.

The first approach to calculate the daily average areal rainfall used the arithmetic mean method, applying equation (E3.1.7). To evaluate the measurement uncertainty for this linear model, the GUM approach gives an exact solution, making the assumptions that considering that measure-

Table E3.1.2: Parameters and expanded measurement uncertainties obtained for the weighing gauge and tipping-bucket gauge using GUM and GUM-S1

Weighing gauge / ($mm\,h^{-1}$)				Tipping-bucket gauge / ($mm\,h^{-1}$)			
GUM		GUM-S1		GUM		GUM-S1	
Mean (P)	$U_{95}(P)$	Mean (P)	$U_{95}(P)$	Mean (P)	$U_{95}(P)$	Mean (P)	$U_{95}(P)$
10.05	1.2	10.05	[8.9, 11.3]	10.5	1.6	10.5	[9.0, 12.1]
Skewness (GUM-S1)		0.08		Skewness (GUM-S1)		0.11	
Kurtosis (GUM-S1)		2.96		Kurtosis (GUM-S1)		2.73	

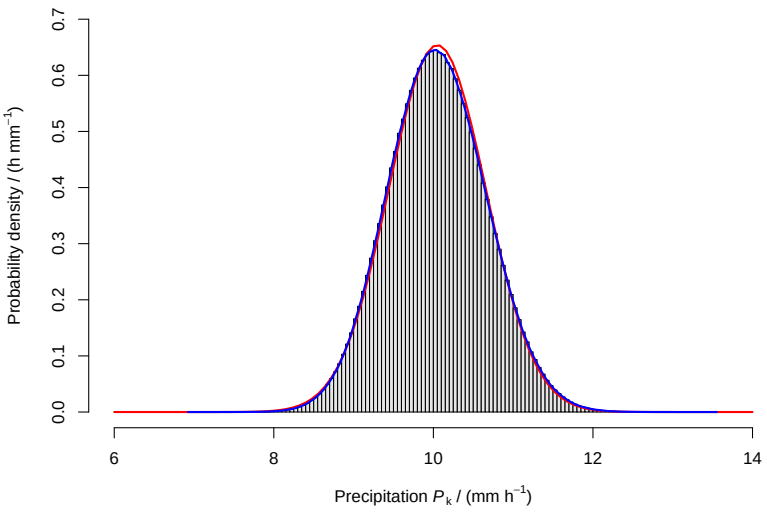


Figure E3.1.7: Weighing gauge PDFs obtained using GUM (red line) and GUM-S1 (blue line) and scaled histogram of output numerical sequence

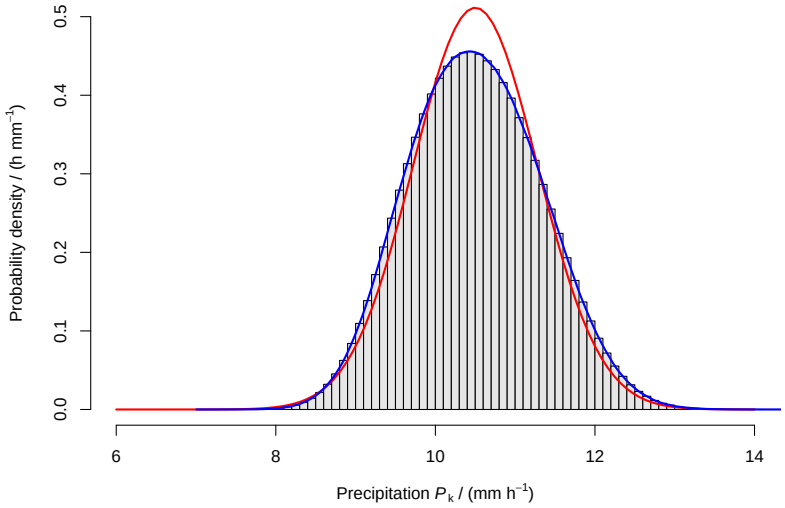


Figure E3.1.8: Tipping-bucket gauge PDFs obtained using GUM (red line) and GUM-S1 (blue line) and scaled histogram of output numerical sequence

ments at the different stations are not correlated,

$$u^2(P_{\text{av}}) = \sum_{i=1}^{4}\left(\frac{\partial P_{\text{av}}}{\partial P_i}\right)^2 u^2(P_i) = \frac{1}{16}\left[u^2(P_1) + u^2(P_2) + u^2(P_3) + u^2\left(P_4\right)\right]. \qquad \text{(E3.1.10)}$$

In the given catchment area, using the values given above for P_1 to P_4, the estimate of P_{av} (daily average areal rainfall) is 23.5 mm h^{-1}. Considering that $u(P_i) = 0.06P_i$ for weighing gauges and $u(P_i) = 0.08P_i$ for tipping-bucket gauges, expression (E3.1.10) is used to obtain the standard uncertainties for both types of gauges considered in stage 1:

$u(P_{\text{av}}) = 0.76$ mm, using weighing gauges' uncertainty

$u(P_{\text{av}}) = 1.0$ mm, using tipping-bucket gauges' uncertainty

The second method, the Thiessen polygons method, was applied using the same P_i values but it requires to evaluate the areas of the polygons that gives the weights considered in expression (E3.1.8). The areas related to the polygons shown in figure (E3.1.7) were obtained using a planimeter technique, giving approximate values of

$$\frac{A_1}{A} = 0.37; \frac{A_2}{A} = 0.24; \frac{A_3}{A} = 0.20; \frac{A_4}{A} = 0.19.$$

To evaluate measurement uncertainty using the Thiessen polygons method, uncertainty contributions for daily average areal rainfall were the same as considered in the previous example, being $u(P_i) = 0.06P_i$ for weighing gauges and $u(P_i) = 0.08P_i$ for tipping-bucket gauges. The combined uncertainty also takes account of the uncertainty contributions related to the area weight of each polygon, $u(A_i/A)$, estimated to be 0.01. In this case, the evaluation of the measurement uncertainty associated with the annual average areal rainfall intensity used a MC approach according to GUM-S1. The numerical evaluation was developed for both types of gauges, using RStudio programming, with 1×10^6 MC trials. Using expression (E3.1.8) the daily average areal rainfall, $P_{\text{ax.Tp}}$ is 21.3 mm and,

$u\left(P_{\text{av.Tp}}\right) \approx 0.8$ mm, using weighing gauges uncertainty,

$u\left(P_{\text{av.Tp}}\right) \approx 1.0$ mm, using tipping-bucket gauges uncertainty.

The third approach applies the isohyetal method to the same area, requiring to obtain the values for the isohyets presented in figure E3.1.8 in order to make the computation of the annual average areal rainfall intensity according with expression (E3.1.9). In this case, the values considered for the isohyets, considering the average estimates of P_1 to P_4 were the following:

$$P_a = 6\,\text{mm}; P_b = 15\,\text{mm}; P_c = 24\,\text{mm}; P_d = 32\,\text{mm}; P_e = 44\,\text{mm}.$$

Standard measurement uncertainties considered for the isohyets were taken as those used previously. As in the second method, the relative areas between adjacent isohyets need to be evaluated, which allowed to obtain:

$$\frac{A_1}{A} = 0.31; \frac{A_2}{A} = 0.28; \frac{A_3}{A} = 0.23; \frac{A_4}{A} = 0.18.$$

The standard uncertainty related to the area weight of each subarea, $u\left(A_i/A\right)$, were taken to be 0.01.

The evaluation of the measurement uncertainty associated with the daily average areal rainfall intensity used again an MCM approach according to GUM-S1. The numerical evaluation was developed for both types of gauges, using RStudio [49] programming, with 1×10^6 runs for each calculation. Using expression (E3.1.9) the estimate of, $P_{\text{av.Isoh}}$, is 21.9 mm and the standard uncertainties are

$$u\left(P_{\text{av.Isoh}}\right) \approx 1.9\,\text{mm}, \text{using weighing gauges uncertainty,}$$
$$u\left(P_{\text{av.Tp}}\right) \approx 2.4\,\text{mm}, \text{using tipping-bucket gauges uncertainty.}$$

A summary of the results is given in tables E3.1.3 and E3.1.4 and the output PDFs in figures E3.1.9 and E3.1.10, respectively, considering the use of weighing gauges and tipping-bucket gauges as measurement instruments.

Table E3.1.3: Comparison results obtained using weighing gauge data input

Daily average areal rainfall	Method/mm		
	Arithmetic mean	Thiessen polygons	Isohyetal
Estimate	23.5	21,3	21.9
95 % uncertainty (GUM)	1.5	1.6	—
95 % uncertainty (GUM-S1)	—	[19.7 , 22.9]	[20.0 , 23.8]
95 % uncertainty half-width (GUM-S1)	—	1.6	1.9

Table E3.1.4: Comparison results obtained using tipping-bucket gauge data input

Daily average areal rainfall	Method/mm		
	Arithmetic mean	Thiessen polygons	Isohyetal
Estimate	23.5	21.3	21.9
95 % uncertainty (GUM)	2	2	—
95 % uncertainty (GUM-S1)	—	[19.3 , 23.3]	[19.5 , 24.3]
95 % expanded uncertainty (GUM-S1)	—	2	2.4

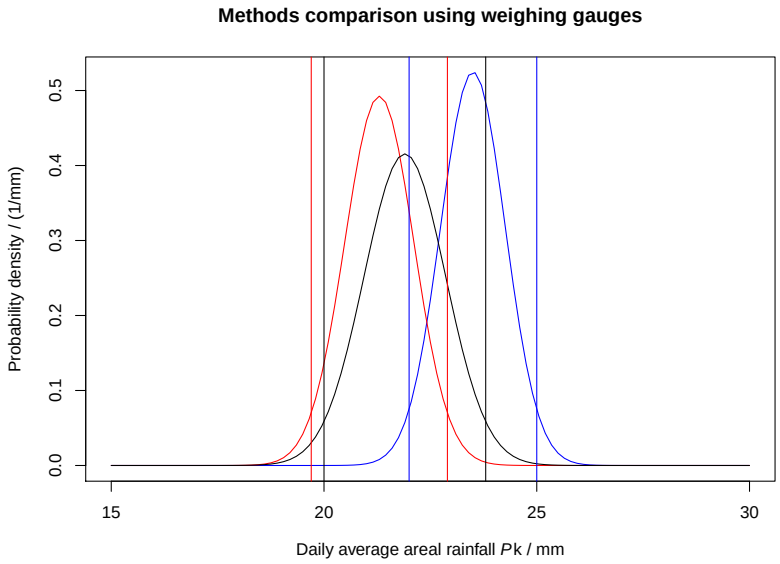


Figure E3.1.9: Comparison of results and 95 % expanded uncertainty, using as input the uncertainty rainfall intensity measurement of weighing gauges, for the arithmetic mean method (blue line), Thiessen polygon method (red line) and isohyetal method (black line)

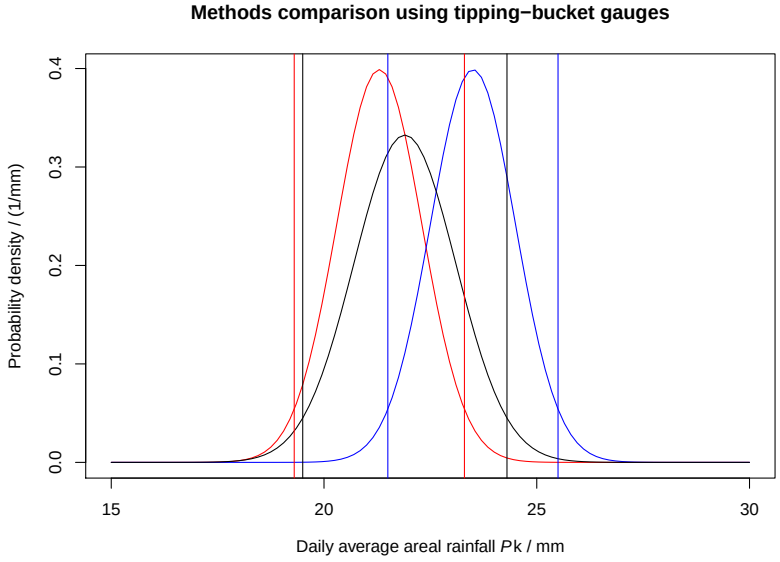


Figure E3.1.10: As figure E3.1.9 but for tipping-bucket gauges

E3.1.7 Interpretation of results

The diversity of measurement instruments has impact on the estimates and the associated uncertainties and the decisions taken should consider the effects due to the uncertainty contributions. In this study, two different techniques (weighing and tipping-bucket) for the same measurements (precipitation and rainfall) were adopted. In both cases the results obtained using GUF and MCM were consistent. In the case of the tipping-bucket gauge, a higher degree of flatness of GUM-S1

histogram was found showing a deviation from normality. In both cases, average estimates were corrected, in order to consider the contribution of the wetting loss bias, being the correction to add a positive quantity not measured in the process.

Regarding the three methods considered, commonly applied to evaluate daily, monthly or annual average areal rainfall, the interest of the studies carried out were related to performing a comparison of the results based on the measurement uncertainties. The results obtained and presented in tables E3.1.3 and E3.1.4 and in figures E3.1.9 and E3.1.10 show differences in the estimates of the average areal rainfall, from 21.3 mm to 23.5 mm. Regarding the 95 % expanded measurement uncertainty obtained (tables E3.1.3 and E3.1.4), agreement was found between arithmetic mean method and Thiessen polygons method being the measurement uncertainty around 20 % higher for the isohyetal method. The comparison between weighing gauges and tipping-bucket gauges showed a difference of 30 %, enhancing the conclusions that the impact of the type of gauge used and the method adopted are high.

The studies carried out were able to show some interesting features of the models and the way they affect the measurand of interest. Further studies should be made to include the effect of correlation that was not considered in this simple analysis, and the effect of conditions related with the dynamics of the measurement process and the use of corrective algorithms related to post-processing of data [219–221].

Example E3.2

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

E3.2.1 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 (GUF) [2] and the Monte Carlo method (MCM) for the propagation of distributions [3, 4], applied to both real and simulated data sets, are shown and discussed.

E3.2.2 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 [222] 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) [2] and the MCM for propagation of distributions [3] 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 [223], whereas details on the uncertainty evaluation process, not explicitly reported in the present example, can be found in [224].

E3.2.3 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 [223]. 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 µL 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 [223], i.e., 2.5×10^{-3} ng.

E3.2.4 Measurement model

Quantification of the mass of BaP contained in 1 µL of the second extract was performed according to the Internal Standard method described in EN 15549 [225]. 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 µg mL^{-1}, to be used as the internal standard. Then, an aliquot of 1 µL 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:

$$m_E = \left(f\,\overline{A_E}\,m_{ISE}\right) / \left(\overline{A_{ISE}}\right), \tag{E3.2.1}$$

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 [224]. 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 certified reference material (CRM) used for preparing the three necessary reference solutions. For the same reason, f and m_{ISE}, as input quantities of measurement model (E3.2.1), 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 [2, Sec. 4.2.3] (hence, having two degrees of freedom). A strong lin-

ear 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 [2, Sec. 5.2.3].

Estimates, uncertainties $u(x)$ and covariances $u(x, y)$ of the input quantities in model (E3.2.1) are reported in table (E3.2.1), together with other parameters relevant to the uncertainty evaluation performed by MCM. The corresponding estimate for the measurand is $m_{\mathrm{E}} = 0.014\,\mathrm{ng}$.

Table E3.2.1: Estimate, uncertainty, covariance and distributional parameters of the input quantities in model (E3.2.1).

Parameter	Value
f	0.616
$u(f)$	0.017
m_{ISE}	0.2455 ng
$u(m_{\mathrm{ISE}})$	0.0036 ng
$u(f, m_{\mathrm{ISE}})$	$-3.3 \cdot 10^{-5}$ ng
$\overline{A_{\mathrm{E}}}$	85 114 a.u.
$u(\overline{A_{\mathrm{E}}})$	9564.35 a.u.
$\overline{A_{\mathrm{ISE}}}$	917 545.67 a.u.
$u(\overline{A_{\mathrm{ISE}}})$	44 492.21 a.u.
$u(\overline{A_{\mathrm{E}}}, \overline{A_{\mathrm{ISE}}})$	$-203\,436\,959.5$ (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 (E3.2.1) are still valid for the simulated case[1], except for the $\overline{A_{\mathrm{E}}}$ value which becomes equal to 18 114 a u . The corresponding estimate for the (simulated) measurand is $m_{\mathrm{E}} = 0.003$ ng.

E3.2.5 Uncertainty propagation

For calculating the uncertainty associated with the estimates of the measurands (i.e., $m_{\mathrm{E}} = 0.014\,\mathrm{ng}$ and $m_{\mathrm{E}} = 0.003\,\mathrm{ng}$, respectively), both the LPU [2] and the MCM for the propagation of probability distributions [3, 4] 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" [20].

[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.

E3.2.5.1 GUM uncertainty framework

Applying the LPU to model (E3.2.1), taking into account the uncertainty and covariance contributions of the input quantities reported in table (E3.2.1), the resulting uncertainty $u(m_{\mathrm{E}})$ was 0.0020 ng and 0.0017 ng for the experimental and the simulated case, respectively. Table (E3.2.2) reports the uncertainty budget for the mass $m_{\mathrm{E}} = 0.014\,\mathrm{ng}$ of BaP of the sample obtained with the second extraction (the first two columns repeat part of the information already available in table (E3.2.1)). Effective degrees of freedom ν_{eff} were calculated, according to the Welch-

Table E3.2.2: Uncertainty budget for the mass of BaP (in 1 μL) of the sample obtained with the second extraction: associated combined uncertainty was $u(m_{\mathrm{E}}) = 0.002$ ng.

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

Satterthwaite formula [2, eqn. (G2.b)] applied to the input uncertainties and the corresponding degrees of freedom discussed in Sec. E3.2.4. They were equal to 3.07 and 2.05, giving coverage factors k of 3.1 and 4.2, respectively, 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 [2, 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$^{\mathrm{th}}$ percentile of the distribution is used for calculating the expanded uncertainty according to [2, eqn. (G.1d)]. In the present case, the expanded uncertainties $U = k\,u(m_{\mathrm{E}})$ 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 (E3.2.3). 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 E3.2.3: Estimate of the measurand, with associated standard and expanded uncertainty, for the experimental and the simulated case.

	m_{E}/ng	$u(m_{\mathrm{E}})$/ng	$U(m_{\mathrm{E}})$/ng
Experimental case	0.014	0.0020	0.006
Simulated case	0.003	0.0017	0.007

E3.2.5.2 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 (E3.2.1), according to the criteria prescribed in [3, 4].

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 [3, Sec. 6.4.8]. Hence, the bivariate normal distribution had a (vector) expectation equal to $[f, m_{\mathrm{ISE}}]$ and a covariance matrix $\mathbf{\Sigma}$ equal to

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

whose components are available in table (E3.2.1).

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 [4, Sec. 5.3.2]. Hence, the bivariate t-distribution had a (vector) expectation equal to $[\overline{A_{\mathrm{E}}}, \overline{A_{\mathrm{ISE}}}]$ and the scale matrix $\mathbf{S}/n$ with $\mathbf{S}$ defined by:

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

whose components S_{ij} are available in table E3.2.1, 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 [4, Section 5.5.2, Note 1].

The numerical simulation of the input probability distributions and their propagation through measurement model (E3.2.1) were implemented in R environment [11] by applying R functions “rmvnorm” and “rmvt” available in the “mvtnorm” package [51]. 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 [3, 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 9×10^6 and 7×10^6 for the real and the simulated case, respectively, hence still providing a reliable numerical approximation for the measurand distribution.

From the MCM distribution, the shortest 95 % coverage interval was obtained, for both real and simulated data, and reported in table E3.2.4.

E3.2.6 Reporting the result

Figures E3.2.1 and E3.2.2 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 GUF. The relevant interval limits are also reported in table E3.2.4.

In the present example, the MCM 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

Table E3.2.4: 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]	$[5.9 \cdot 10^{-10}, 0.020]$

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 E3.2.4, proving themselves as robust and sensible estimates for the measurand.

E3.2.7 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 [3, eqs. (16) and (17)]. Nonetheless, NOTE 2 in [3, 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) in GUM-S1 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 GUF, as reported in table E3.2.3, whereas feasible coverage intervals are those provided by MCM, as discussed shortly.

From both figures E3.2.1 and E3.2.2, it is evident that the two approaches give quite different results in terms of coverage intervals. Although the assumed output distribution in the GUF 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 GUF, 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 (figure E3.2.2), the GUF would lead to a coverage interval stretching into a region of negative (unfeasible) values. The MCM, instead, can provide a realistic asymmetric interval. It is worth mentioning that the EURACHEM/CITAC guidelines on uncertainty evaluation for analytical measurements [42] recommend truncating the expanded uncertainty interval at zero whenever a negative lower limit is found for a non-negative quantity.

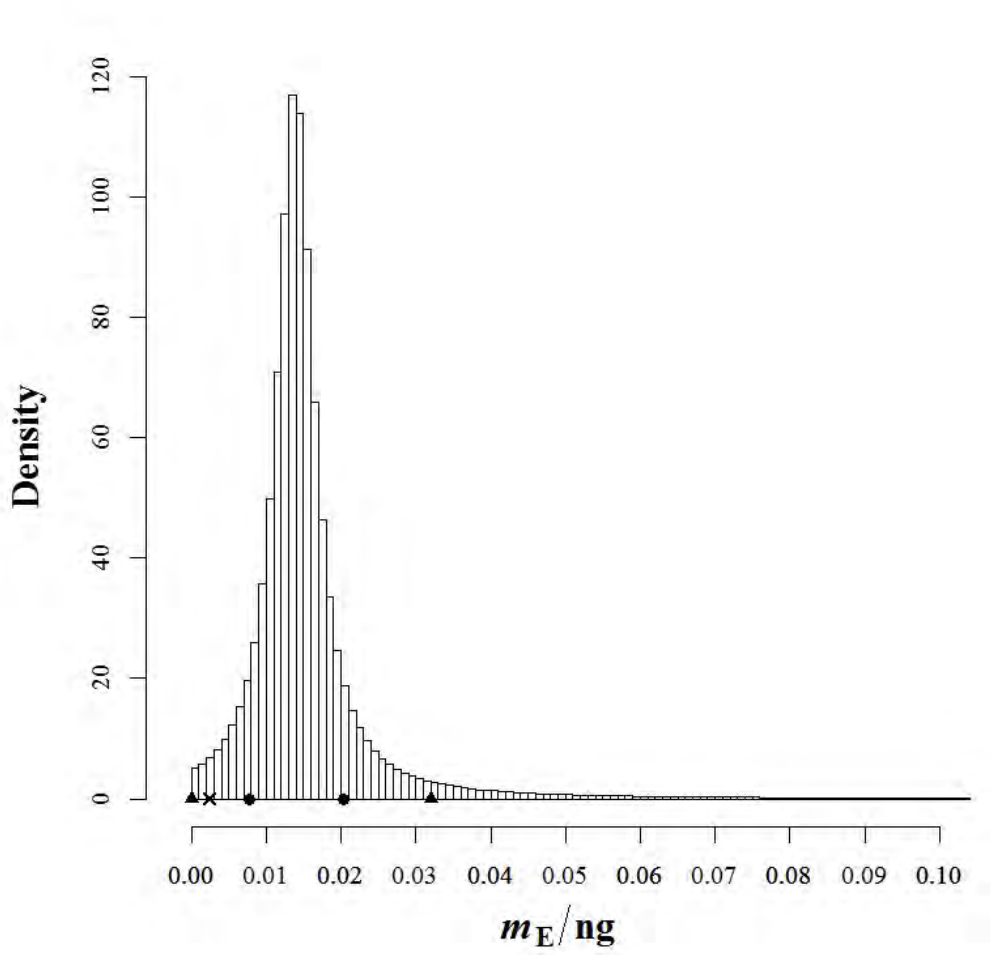


Figure E3.2.1: Numerical representation of the pdf associated with the mass of BaP in the nominal volume of 1 µL of the sample obtained with the second extraction ($m_E = 0.014$ ng). Circle and triangle symbols indicate the limits of the 95 % coverage interval obtained according to the GUM uncertainty framework and by MCM, respectively. Symbol x indicates the minimum detectable mass of the analytical method.

Summarizing, this example is a clear case in which blind adherence either to the approach in the GUM [2] or to the MCM in [3,4] would be dangerous. Careful considerations on estimates, standard uncertainties and coverage intervals are always needed, according to the specific problem under study.

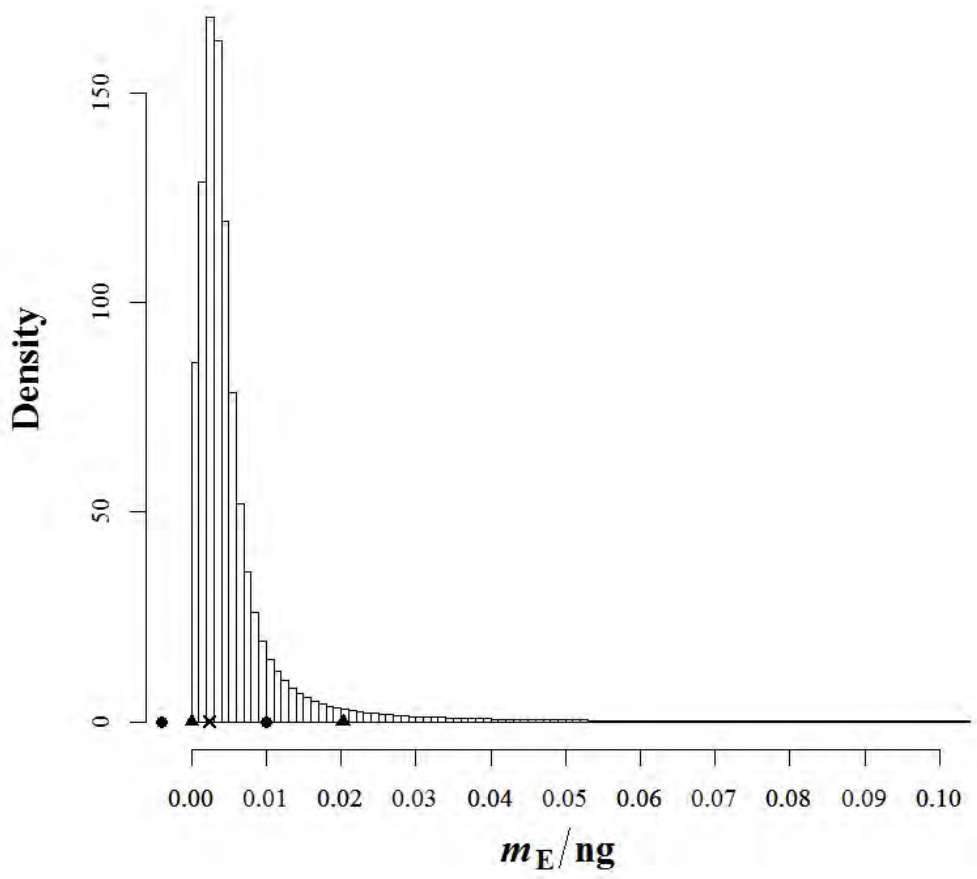


Figure E3.2.2: Numerical representation of the pdf associated with a simulated very small mass value of BaP ($m_E = 0.003$ ng). Circle and triangle symbols indicate the limits of the 95 % coverage interval obtained according to the GUM uncertainty framework and by MCM, respectively. Symbol x indicates the minimum detectable mass of the analytical method.

Example E3.3

Calibration of an analyser for NO_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

E3.3.1 Summary

The present example shows the uncertainty evaluation of the calibration of a chemiluminescence analyser for nitrogen oxides (NO_x) using a multi-point calibration as described in ISO 6143 [226] with dynamically prepared calibration gas mixtures obtained by dynamic dilution of standard gas mixtures performed by means of calibrated mass flow controllers (MFCs) [227].

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.

E3.3.2 Introduction of the application

The European Directive on ambient air quality [228] prescribes the monitoring of NO_x by means of chemiluminescence as the reference method [229], 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 mixtures 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.

E3.3.3 Specification of the measurand

E3.3.3.1 Use of two calibrated MFCs

In the present example, two MFCs from MKS with full scale range (FSR) of 500 $cm^3\,min^{-1}$ and 2000 $cm^3\,min^{-1}$ 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 $cm^3\,min^{-1}$ the following model was found appropriate

$$C_1 = \alpha_1/q_{VN} + \beta_1 + \gamma_1\, q_{VN} + \delta_1\, q_{VN}^2\,, \tag{E3.3.1}$$

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

$$C_2 = \alpha_2/q_{VN} + \beta_2/\sqrt{q_{VN}} + \gamma_2 + \delta_2\sqrt{q_{VN}} + \epsilon_2\, q_{VN}\,, \tag{E3.3.2}$$

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 [110].

Weighted Least-Squares regression was employed for fitting eqs. (E3.3.1) and (E3.3.2) 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 E3.3.1, whereas associated (squared) uncertainties and covariances are shown in covariance matrices (E3.3.3) and (E3.3.4), respectively.

$$V_{C_1} = \begin{pmatrix} 9.2\times10^{-4} & -1.84\times10^{-5} & 9.5\times10^{-8} & -1.3\times10^{-10} \\ -1.84\times10^{-5} & 4.4\times10^{-7} & -2.4\times10^{-9} & 3.4\times10^{-12} \\ 9.5\times10^{-8} & -2.4\times10^{-9} & 1.4\times10^{-10} & -2.1\times10^{-14} \\ -1.3\times10^{-10} & 3.4\times10^{-12} & -2.1\times10^{-14} & 3.3\times10^{-17} \end{pmatrix} \tag{E3.3.3}$$

Table E3.3.1: Calibration curve parameter estimates for the two MFCs according to eqs. (E3.3.1) and (E3.3.2) (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}

$$V_{C_2} = \begin{pmatrix} 18.60 & -4.0 & 3.0 \times 10^{-1} & -9.3 \times 10^{-3} & 9.9 \times 10^{-5} \\ -4.0 & 8.9 \times 10^{-1} & -6.7 \times 10^{-2} & 2.1 \times 10^{-3} & -2.2 \times 10^{-5} \\ 3.0 \times 10^{-1} & -6.7 \times 10^{-2} & 5.1 \times 10^{-3} & -1.6 \times 10^{-4} & 1.7 \times 10^{-6} \\ -9.3 \times 10^{-3} & 2.1 \times 10^{-3} & -1.6 \times 10^{-4} & 4.9 \times 10^{-6} & -5.3 \times 10^{-8} \\ 9.9 \times 10^{-5} & -2.2 \times 10^{-5} & 1.7 \times 10^{-6} & -5.3 \times 10^{-8} & 5.8 \times 10^{-10} \end{pmatrix} \quad \text{(E3.3.4)}$$

As an example, the calibration curve for the MFC1 with FSR of $500\,\mathrm{cm^3\,min^{-1}}$ is shown in figure E3.3.1 (all volumes here and in the following are referred to temperature and pressure standard conditions, i.e., 0 °C and 1013.25 mbar).

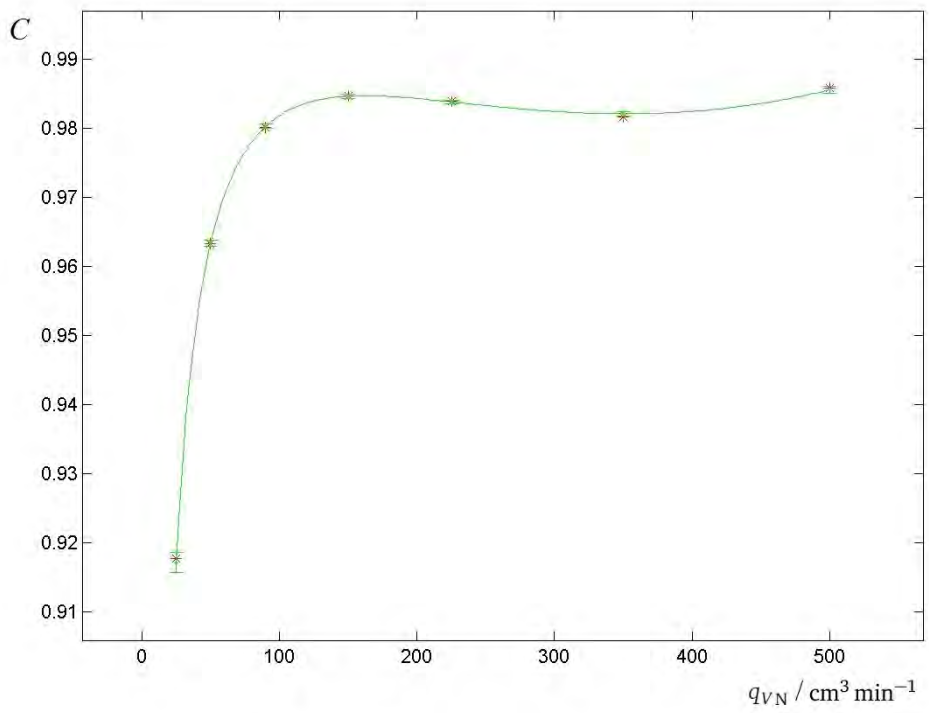


Figure E3.3.1: Calibration curve of the MFC1 with FSR of $500\,\mathrm{cm^3\,min^{-1}}$

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.

E3.3.3.2 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 [227].

E3.3.3.3 Calibration of a chemiluminescence analyser

A Thermo Fisher Scientific 42*i* chemiluminescence analyser is calibrated for NO_2/SA, in the range from 700 nmol mol^{-1} to 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 [226] 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.

E3.3.4 Measurement model

E3.3.4.1 Use of two calibrated MFCs

After the MFCs are calibrated, the flow $q_{V\,i}$ they provide at a nominal flow $q_{V\,i\mathrm{N}}$ is given by the following model:

$$q_{V\,i} = q_{V\,i\mathrm{N}}\, C_i, \tag{E3.3.5}$$

where C_i is the calibration coefficient calculated according to eq. (E3.3.1) or (E3.3.2).

E3.3.4.2 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:

$$x_{\mathrm{a}} = \frac{x_1 q_{V\,1\mathrm{a}} + x_2 q_{V\,2\mathrm{a}}}{q_{V\,1\mathrm{a}} + q_{V\,2\mathrm{a}}}\,, \tag{E3.3.6}$$

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), $q_{V\,1\mathrm{a}}$ is the flow (cm^3 min^{-1}) of the parent mixture supplied by MFC1, $q_{V\,2\mathrm{a}}$ 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

[1]Equation (E3.3.6) 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.

and that this is certified with zero uncertainty (as in the present case study), equation (E3.3.6) can be simplified into the following:

$$x_{\mathrm{a}} = \frac{x_1 q_{V\,1\mathrm{a}}}{q_{V\,1\mathrm{a}} + q_{V\,2\mathrm{a}}} \,. \tag{E3.3.7}$$

Three different mixtures having amount fractions x_{a}, x_{b} and x_{c}, respectively, are generated.

E3.3.4.3 Calibration of a chemiluminescence analyser

As the analysis curve of the instrument, a straight line

$$y = A + Bw, \tag{E3.3.8}$$

is fitted to the data, which are the three different amount fractions x_a, x_b, x_c (the y values in equation (E3.3.8)), 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 [230], 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. (E3.3.14) and (E3.3.16), 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.

E3.3.5 Uncertainty propagation

E3.3.5.1 Use of two calibrated MFCs

Using to the law of propagation of uncertainty (LPU) from [2], the uncertainty associated with a flow $q_{V\,i}$ (E3.3.5) produced by the i-th calibrated MFC is given by

$$u(q_{V\,i}) = q_{V\,i\mathrm{N}}\, u(C_i). \tag{E3.3.9}$$

Uncertainty $u(C_i)$ can be expressed as the sum in quadrature of a systematic contribution, $u(C_i)_{\mathrm{cal}}$, due to the MCF calibration (and calculated by applying the LPU to eq. (E3.3.1) or (E3.3.2), respectively, taking into account uncertainties of and covariances between the curve parameters), and a repeatability contribution, $u(C_i)_{\mathrm{rep}}$, 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,

$$u(C_i) = \sqrt{u^2(C_i)_{\mathrm{cal}} + u^2(C_i)_{\mathrm{rep}}} \,. \tag{E3.3.10}$$

Concerning covariances, covariance term $u(q_{V1a}, q_{V2a})$ between flow rates provided by two MFCs at the same nominal value $q_{V\mathrm{Na}}$ 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 $q_{V\mathrm{Na}}$ and $q_{V\mathrm{Nb}}$ are not negligible since flow estimates by the same MFC are recovered by the application of the very same calibration curve. By considering expression (E3.3.5), one has

$$u(q_{Vi\mathrm{a}}, q_{Vi\mathrm{b}}) = u(q_{Vi\mathrm{Na}} C_{i\mathrm{a}}, q_{Vi\mathrm{Nb}} C_{i\mathrm{b}}) = q_{Vi\mathrm{Na}}\, q_{Vi\mathrm{Nb}}\, u(C_{i\mathrm{a}}, C_{i\mathrm{b}})\,, \tag{E3.3.11}$$

where, when $i = 1$, for example, and hence applying eq. (E3.3.1),

$$u(C_{1\mathrm{a}}, C_{1\mathrm{b}}) = u(\alpha_1/q_{V1\mathrm{Na}} + \beta_1 + \gamma_1\, q_{V1\mathrm{Na}} + \delta_1\, q_{V1\mathrm{Na}}^2,\ \alpha_1/q_{V1\mathrm{Nb}} + \beta_1 + \gamma_1\, q_{V1\mathrm{Nb}} + \delta_1\, q_{V1\mathrm{Nb}}^2). \tag{E3.3.12}$$

Employing the covariance property for linear combinations of variables, eq. (E3.3.12) becomes

$$u(C_{1\mathrm{a}}, C_{1\mathrm{b}}) = 1/(q_{V1\mathrm{Na}} q_{V1\mathrm{Nb}})\, u^2(\alpha_1) + 1/q_{V1\mathrm{Na}}\, u(\alpha_1, \beta_1) + \ldots + (q_{V1\mathrm{Na}} q_{V1\mathrm{Nb}})^2\, u^2(\delta_1). \tag{E3.3.13}$$

Therefore, uncertainties associated with and covariances between parameters of calibration curve (E3.3.1) influence the covariance between two different flows produced by the same MFC1. Analogous expressions are derived for MFC2 as well.

E3.3.5.2 Dynamic dilution

Using the LPU, the (squared) uncertainty associated with the amount fraction x_{a} of the analyte (E3.3.7) is given by

$$u^2(x_{\mathrm{a}}) = \left(\frac{\partial x_{\mathrm{a}}}{\partial x_1}\right)^2 u^2(x_1) + \left(\frac{\partial x_{\mathrm{a}}}{\partial q_{V1\mathrm{a}}}\right)^2 u^2(q_{V1\mathrm{a}}) + \left(\frac{\partial x_{\mathrm{a}}}{\partial q_{V2\mathrm{a}}}\right)^2 u^2(q_{V2\mathrm{a}})\,, \tag{E3.3.14}$$

where $u(x_1)$ is provided by the certificate of the reference parent mixture, while $u(q_{V1\mathrm{a}})$ and $u(q_{V2\mathrm{a}})$ are calculated according to expression (E3.3.9). Note that $u(q_{V1\mathrm{a}}, q_{V2\mathrm{a}}) = 0$. 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:

$$\begin{aligned} u(x_{\mathrm{a}}, x_{\mathrm{b}}) &= u\left(\frac{x_1 q_{V1\mathrm{a}}}{q_{V1\mathrm{a}} + q_{V2\mathrm{a}}}, \frac{x_1 q_{V1\mathrm{b}}}{q_{V1\mathrm{b}} + q_{V2\mathrm{b}}}\right) \approx \\ &\approx \frac{\partial x_{\mathrm{a}}}{\partial x_1}\frac{\partial x_{\mathrm{b}}}{\partial x_1} u(x_1, x_1) + \frac{\partial x_{\mathrm{a}}}{\partial x_1}\frac{\partial x_{\mathrm{b}}}{\partial q_{V1\mathrm{b}}} u(x_1, q_{V1\mathrm{b}}) + \\ &\quad + \frac{\partial x_{\mathrm{a}}}{\partial x_1}\frac{\partial x_{\mathrm{b}}}{\partial q_{V2\mathrm{b}}} u(x_1, q_{V2\mathrm{b}}) + \ldots + \frac{\partial x_{\mathrm{a}}}{\partial q_{V2\mathrm{a}}}\frac{\partial x_{\mathrm{b}}}{\partial q_{V2\mathrm{b}}} u(q_{V2\mathrm{a}}, q_{V2\mathrm{b}}). \end{aligned} \tag{E3.3.15}$$

Recalling that $u(q_{V1\mathrm{a}}, q_{V2\mathrm{b}}) = u(q_{V2\mathrm{a}}, q_{V1\mathrm{b}}) = 0$, and considering that there is no covariance between x_1 and any of MFC flow values, eq. (E3.3.15) reduces to:

$$\begin{aligned} u(x_{\mathrm{a}}, x_{\mathrm{b}}) &\approx \frac{q_{V1\mathrm{a}} q_{V1\mathrm{b}}}{(q_{V1\mathrm{a}} + q_{V2\mathrm{a}})(q_{V1\mathrm{b}} + q_{V2\mathrm{b}})}\, u^2(x_1) + \\ &+ \frac{x_1^2 q_{V2\mathrm{a}} q_{V2\mathrm{b}}}{(q_{V1\mathrm{a}} + q_{V2\mathrm{a}})^2 (q_{V1\mathrm{b}} + q_{V2\mathrm{b}})^2}\, u(q_{V1\mathrm{a}}, q_{V1\mathrm{b}}) + \\ &+ \frac{x_1^2 q_{V1\mathrm{a}} q_{V1\mathrm{b}}}{(q_{V1\mathrm{a}} + q_{V2\mathrm{a}})^2 (q_{V1\mathrm{b}} + q_{V2\mathrm{b}})^2}\, u(q_{V2\mathrm{a}}, q_{V2\mathrm{b}})\,, \end{aligned} \tag{E3.3.16}$$

where $u(q_{V\,i\mathrm{a}}, q_{V\,i\mathrm{b}})$ are calculated according to eq. (E3.3.11). Analogous expressions hold for $u(x_\mathrm{a}, x_\mathrm{c})$ and $u(x_\mathrm{b}, x_\mathrm{c})$, as well.

E3.3.5.3 Calibration of a chemiluminescence analyser

Estimate of coefficients A and B of analysis curve (E3.3.8) 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 [230].

E3.3.6 Reporting the result

Nominal and measured flow values of MFC1 and MFC2 are reported in table (E3.3.2) together with associated uncertainties. These are calculated by eqs. (E3.3.9) and (E3.3.10), where contribution $u(C_i)_\mathrm{rep}$ is equal to 0.013 % and 0.037 % of the measured flow $q_{V\,i}$ value for MFC1 and MFC2, respectively.

Table E3.3.2: Nominal and measured flow values of MFC1 and MFC2 with associated uncertainties ($\mathrm{cm^3\,min^{-1}}$).

	$q_{V\,1\mathrm{N}}$	$q_{V\,2\mathrm{N}}$	$q_{V\,1}$	$u(q_{V\,1})$	$q_{V\,2}$	$u(q_{V\,2})$
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. (E3.3.11) and (E3.3.12) (and corresponding ones for amount fraction values b and c, and for MFC2), and reported in table (E3.3.3).

Table E3.3.3: Covariance terms between measured flow values of the same MFC (MFC1 and MFC2) at different fraction levels a, b and c (all expressed in $(\mathrm{cm^3\,min^{-1}})^2$).

MFC	$u(q_{V\,\mathrm{a}}, q_{V\,\mathrm{b}})$	$u(q_{V\,\mathrm{a}}, q_{V\,\mathrm{c}})$	$u(q_{V\,\mathrm{b}}, q_{V\,\mathrm{c}})$
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 $x_1 = 10.252\,\mathrm{\mu mol\,mol^{-1}}$, with associated uncertainty $u(x_1) = 0.016\,\mathrm{\mu mol\,mol^{-1}}$.

Applying eq. (E3.3.7), and corresponding ones for amount fractions b and c, the three reference amount fractions are obtained as reported in table (E3.3.4).

Relevant squared uncertainties and covariances (in $\mathrm{nmol^2\,mol^{-2}}$), reported in the covariance matrix E3.3.17, are calculated according to eqs. (E3.3.14) and (E3.3.16), exploiting uncertainty and covariance values associated with $q_{V\,i}$ reported in tables E3.3.2 and E3.3.3.

$$V_{x_{\mathrm{a,b,c}}} = \begin{pmatrix} 39.5 & 1.7 & 2.2 \\ 1.7 & 37.0 & 3.0 \\ 2.2 & 3.0 & 41.2 \end{pmatrix}. \qquad \text{(E3.3.17)}$$

Table E3.3.4: Amount fractions of the three calibration mixtures obtained by dynamic dilution and associated uncertainties (nmol mol^{-1}).

amount fraction	x	$u(x)$
a	708.1	6.3
b	974.1	6.1
c	1290.1	6.4

Each calibration mixture is analyzed three times: the sample mean and the associated uncertainty values are reported in table E3.3.5. A straight line is fitted to the data by means of WTLS

Table E3.3.5: Sample mean of repeated readings and associated uncertainties (a u).

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

regression. Parameter estimates of model (E3.3.8) 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 E3.3.18.

$$V_{A,B} = \begin{pmatrix} 2.85 \cdot 10^{2} & -1.30 \\ -1.30 & 6.36 \cdot 10^{-3} \end{pmatrix} . \qquad \text{(E3.3.18)}$$

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 (E3.3.8) 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 (E3.3.8) propagating uncertainty $u(w)$ and terms of the covariance matrix E3.3.18 through the model, that is applying the following equation:

$$u^2(y) = u^2(A) + u^2(B)\,w^2 + u^2(w)\,B^2 + 2\,u(A,B)\,w \; . \qquad \text{(E3.3.19)}$$

The validation result is reported in figure E3.3.2.

E3.3.7 Interpretation of results

Covariances in matrix (E3.3.17) between amount fractions of the three calibration mixtures are mainly due to the term proportional to $u(x_1)$ (first term of eq. (E3.3.16)), 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 $u(q_{V\,\mathrm{a}}, q_{V\,\mathrm{b}})$, $u(q_{V\,\mathrm{a}}, q_{V\,\mathrm{c}})$ and $u(q_{V\,\mathrm{b}}, q_{V\,\mathrm{c}})$ in table E3.3.3 (second and third terms of eq. (E3.3.16)) are smaller by one or two orders of magnitude.

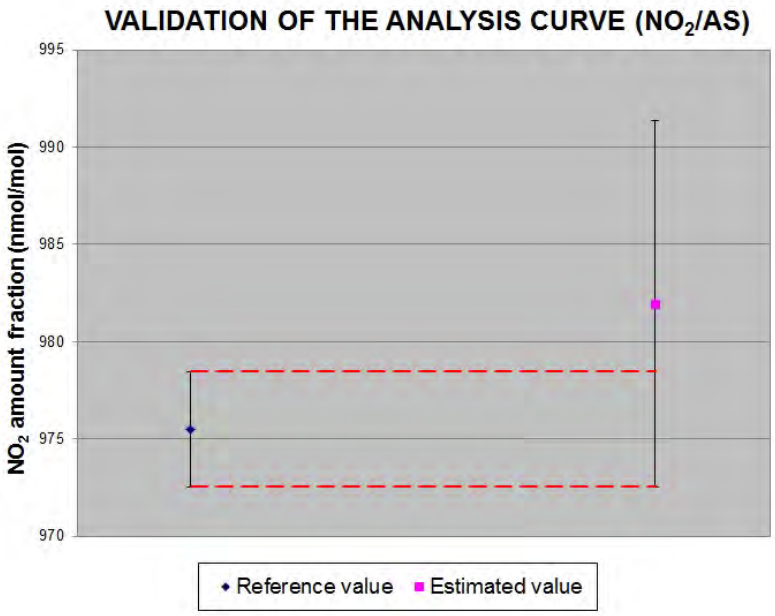


Figure E3.3.2: Validation of the analysis curve for the quantification of NO_2 in SA. Estimate of a gas mixture of 975.5 nmol mol^{-1} of NO_2/SA as provided by the calibrated instrument (Estimated value) in comparison with the actual value of the mixture as reported in its calibration certificate (Reference value).

If covariances between the three amount fractions were ignored in the calibration of the chemiluminescence analyser, i.e., if a diagonal version of matrix E3.3.17 were used in the WTLS regression, slightly different estimates of A and B for the analysis curve (E3.3.8) would be obtained, with a different associated covariance matrix. The corresponding results of the validation process would be $y = 982.0\,\text{nmol}\,\text{mol}^{-1}$ and $u(y) = 4.5\,\text{nmol}\,\text{mol}^{-1}$, showing an undervaluation of the uncertainty up to 3.4 %.

Example E3.4

Measurement uncertainty for routine testing of metals in soil

S.L.R. Ellison, M. Singh, M.G. Cox

E3.4.1 Summary

The example describes the evaluation of measurement uncertainty for the routine determination of acid-extractable toxic metals in soil using a combination of acid extraction and atomic emission spectrometry. The example illustrates the general approach to measurement uncertainty evaluation taken by ISO 21748, which uses information on precision and trueness of a routine test procedure to provide an indication of the measurement uncertainty to be expected from the procedure. The example further illustrates the experimental determination of sensitivity coefficients that cannot readily be derived from a mathematical model, examines the evaluation of uncertainties arising from calibration using straight-line regression with zero intercept, and discusses the issues arising in the event of an appreciable bias which is not corrected for when within permitted limits.

E3.4.2 Introduction to the application

The example describes the evaluation of measurement uncertainty for the routine determination of acid-extractable toxic metals in soil using a combination of acid extraction and inductively coupled plasma optical emission spectrometry (ICP-OES). The procedure is based on that of ISO 11466:1995 [231], a method for the extraction of trace elements from soils and similar materials containing less than about 20 % (m/m) organic carbon. ISO 11466:1995 is widely used to determine the levels of toxic metals in soil. This is, in turn, important for determining permitted land use, the need for soil remediation, and in some cases for enforcing effluent or disposal regulations.

The acid extraction step uses *aqua regia*, a mixture of concentrated hydrochloric and nitric acids that takes its Latin name from its ability to dissolve the metals gold and platinum. The process does not dissolve all metal; it does, however, solubilise a large proportion of many metals present in soil as contaminants. Use of a standard method is required to harmonise extraction conditions, which include, for example, reagent concentrations and times.

Following extraction, the extract is made up to a known volume and the dissolved metal content determined by spectrometry. In this example, the spectrometric method is ICP-OES, used for speed, acceptable detection capability and for its ability to determine many elements at the same time. Accurate determination relies on calibration of the instrument using standard solutions; in this example, the calibration procedure uses a simple zero-intercept linear model after correcting instrument response for baseline offset.

The available data in this application are from method validation experiments, which provide information on overall performance of a procedure rather than on individual uncertainty sources. The overall approach to measurement uncertainty evaluation accordingly follows the principles of ISO 21748 [232]. The example illustrates the use of analysis of variance applied to validation data, together with summary information on bias derived from available reference materials, in order to determine major contributions to measurement uncertainty. The example also includes consideration of the uncertainties associated with the use of a zero-intercept model for calibration, and provides an example of experimental determination of sensitivity coefficients for effects for which no useful predictive model exists.

> NOTE: While this example of measurement uncertainty evaluation does not discuss safety aspects, *aqua regia* and its components are highly corrosive and the procedure uses a number of other corrosive or toxic reagents. Attention to safety is accordingly the first priority in practical application of the measurement method.

The data for this example is available elsewhere [22].

E3.4.3 Scope

This example considers the determination of *aqua regia* extractable metals from a test sample prepared according to ISO 11464 [233]. Uncertainties arising from sampling prior to sample preparation and from the sample preparation steps are not considered.

E3.4.4 Specification of the measurand(s)

The measurand for this example is the mass fraction of *aqua regia* extractable metals in soil according to ISO 11466:1995. This is an operationally defined measurand; use of different extraction or sample preparation conditions would not generally be expected to return the same mass fractions for all metals. For operational reasons, the laboratory in this example uses a variant of ISO 11466:1995, after verification of equivalence.

Current soil testing guidelines in the UK provide performance requirements for soil testing; for metals, procedures should provide within-laboratory reproducibility not greater than 5 % and bias (measured against suitable reference materials) that is not significantly greater than 10 %, after allowing for uncertainties in bias assessment. No correction to results is required for procedures meeting the bias criteria.

E3.4.5 Measurement model

E3.4.5.1 Measurement principle and basic model

An outline of the measurement procedure is shown in figure E3.4.1, which also includes the principal parameters of interest for THE uncertainty evaluation. An accurately weighed mass (nominally 3 g) of the prepared test material is placed in a suitable container, moistened with water ((2.0 ± 0.1) ml, not shown in figure E3.4.1) and known volumes ((21.0 ± 0.5) ml and (7.0 ± 0.5) ml respectively) of concentrated hydrochloric and nitric acid are added. (In this section, the ranges indicated by "±" are as stated in the documented procedure and indicate a permitted range). After standing at ambient temperature overnight, the mixture is heated for 2 h ± 5 min, then allowed to cool. For the standard method, the heating is at gentle reflux, that is, the boiling point of the liquid; for the present laboratory implementation, the mixture is held at a nominal 65 °C. The extract is then filtered, any solid residue is washed with dilute nitric acid to remove associated liquid, and the filtered extract, with washings, is standardised by transferring to a 100 ml volumetric flask (class A) and making up to 100 ml.

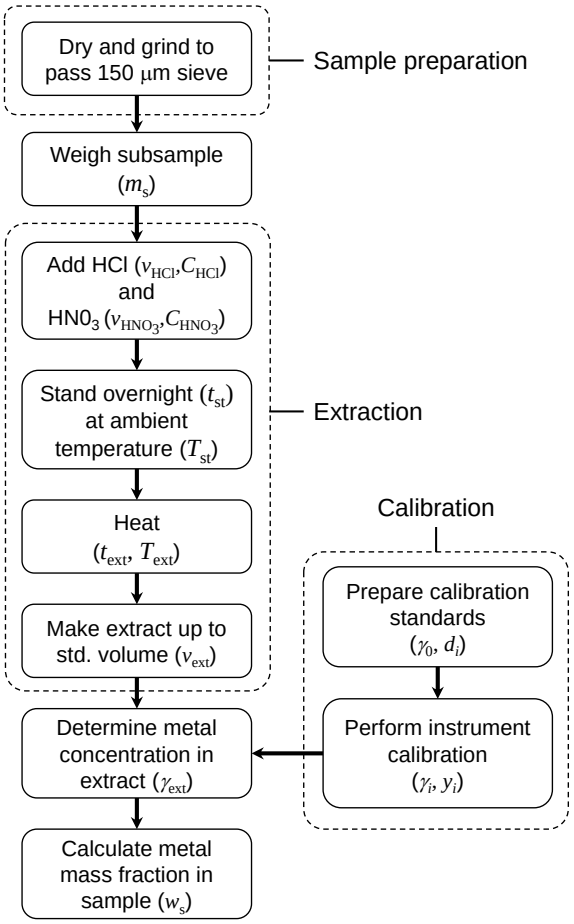


Figure E3.4.1: Overview of *aqua regia* extraction and analysis procedure

The dissolved metal content of the resulting extract is then determined by ICP-OES. The ICP-OES instrument is calibrated by use of a series of solutions containing a mixture of dissolved metals at known concentrations, prepared (in this case) from certified single-element stock solutions. The

instrument calibration and the associated uncertainties are discussed further in section E3.4.6.4. Given a measured concentration γ_{ext}, expressed in mg l^{-1}, for the standardised extract, the mass fraction w_s of metal in the original test sample is given by

$$w_s = \frac{\gamma_{ext} v_{ext}}{m_s}, \tag{E3.4.1}$$

where v_{ext} is the standardised volume of the extract (nominally 100 ml) and m_s the mass of the test portion. The result, w_s, is conventionally expressed in mg kg^{-1}).

E3.4.5.2 The ISO 21748 model

Comparison of figure E3.4.1 with equation (E3.4.1) shows that the equation omits many of quantities that might reasonably affect the measured value. The most important systematic effects are those associated with the extraction conditions, in particular the initial acid concentrations and volumes (C_{HCl}, C_{HNO_3}, v_{HCl} and v_{HNO_3}) and the time t_{ext} and temperature T_{ext} for the heating phase. In addition, equation (E3.4.1) does not make explicit the dependence of measured value on the values associated with calibration standards, the corresponding observed signals y_i and the line fitting method.

For the purpose of uncertainty evaluation, the influence of the calibration step is most simply considered in terms of uncertainty associated with the intermediate measured value γ_{ext}. This is considered further in section E3.4.6.4.

Effects of extraction conditions are much more difficult to include in a measurement model. First, the precise form of the dependence is not usually known. Extraction can in principle be expected to follow a diffusion model. However, diffusion models depend on particle size distribution and shape, neither of which is known here. While semi-empirical diffusion models have given useful results when measured values are monitored over time [234], routine soil extraction methods use a single fixed extraction time, and no generally applicable diffusion model is available for *aqua regia* extraction. Most of the effects also depend on the particular test material and element, making it impractical to develop a quantitative model that applies to every test material.

> NOTE: Sample preparation effects – particularly where they affect particle size – could also have important effects on measured values. These effects are outside the scope of the present example.

For these reasons, routine test procedures usually provide for sufficiently close control of such conditions to make sure that the effects on measured value are negligible compared to other, less controllable, effects. This is the reason for the comparative simplicity of equation (E3.4.1). In addition, routine test procedures are typically characterised by precision and trueness studies of the procedure as a whole. ISO 21748, which provides guidance on the use of precision and trueness data for uncertainty evaluation, accordingly offers a simplified model, equation (E3.4.2), that uses the data from such studies.

$$y = \mu + \delta + B + \sum c_i x_i' + e, \tag{E3.4.2}$$

where

y is the measurement result, assumed to be calculated from an appropriate function;
μ is the (unknown) expectation of ideal results;
δ is a term representing bias intrinsic to the measurement method;

B is the laboratory (or, for intermediate precision, run) component of bias;
x_i' is the deviation from the nominal value of x_i;
c_i is the sensitivity coefficient, equal to $\partial y / \partial x_i$;
e is the random error term under repeatability conditions.

B and e are usually assumed to be normally distributed, with variances of σ_L^2 and σ_r^2.

ISO 21748 additionally explains that, since the estimate observed standard deviations for B, and e are measures of dispersion under the conditions of the study, the summation $\sum c_i x_i'$ is over those effects subject to deviations *other* than those incorporated in B, and e. The summation term accordingly represents the effects of quantities that do not vary appreciably during the corresponding precision or trueness studies. Annex A of ISO 21748 provides further details of the rationale for the model [232].

> NOTE: When precision is expressed in relative terms, it can be important to distinguish between models that relate precision, or other uncertainties, to the unknown true value of the measurand from those that relate effects to the observed value. For small uncertainties there is no practical difference, but as relative standard uncertainties increase beyond approximately 0.2, the different models have appreciably different implications for calculated uncertainty intervals and hence for conformity assessment [235].

E3.4.6 Uncertainty evaluation

E3.4.6.1 Model for uncertainty evaluation

Application of the LPU to equation (E3.4.2), and noting that the theoretical true value μ does not have associated uncertainty, leads to equation (E3.4.3)

$$\begin{aligned} u^2(y) &= u^2(\hat{\delta}) + s_L^2 + \sum c_i^2 u^2(x_i) + s_r^2 \\ &= u^2(\hat{\delta}) + s_I^2 + \sum c_i^2 u^2(x_i), \end{aligned} \tag{E3.4.3}$$

where $u(\bullet)$ denotes the standard uncertainty of the term in parentheses, $\hat{\delta}$ is an estimate of bias δ from a trueness study, s_L^2 and s_r^2 are estimates of the variances σ_L^2 and σ_r^2, and $s_I^2 = s_L^2 + s_r^2$ is the intermediate (within-laboratory) precision standard deviation estimated from the precision study.

Uncertainty evaluation then proceeds by, first, obtaining data from a suitable precision and trueness study to obtain $u(\hat{\delta})$ and s_I, and then identifying effects that are insufficiently represented in the precision study and quantifying their uncertainties.

E3.4.6.2 Random variation – evaluation of s_I

Estimates of within-laboratory reproducibility over time (sometimes called intermediate precision) were obtained from a validation study following the general principles of the MCERTS performance standard for laboratories undertaking chemical analysis of soil [236]. This requires a minimum of 10 degrees of freedom for the intermediate precision standard deviation; the simplest guarantee of this is to run 11 analytical batches including duplicate extraction and measurement – a so-called "11 × 2" design. Four different test materials were included in the study; data are provided in the file `Soil_metals_v1.txt` in the compilation [22].

Repeatability and intermediate precision RSD were estimated for this example using classical analysis of variance, setting the between-run term to zero when the difference in mean squares was negative as in, for example, ISO 5725-2 [237]. The results are summarised in table E3.4.1.

Table E3.4.1: Within-laboratory reproducibility data for *aqua regia* extractable metals in soil

Material	Metal	Mean	s_w	DF (w)	s_b	DF (b)	s_I	RSD (I) (%)	DF (I)
C279a	Cr	875.80	21.57	11	25.22	10	33.18	3.8	10
	Cu	123.77	4.01	11	4.08	10	5.72	4.6	10
	Ni	16.28	0.34	11	0.00	10	0.34	2.1	11
	Pb	27.29	0.41	11	0.19	10	0.45	1.7	10
	Zn	58.70	8.19	11	0.00	10	8.19	13.9	11
C282a	Cr	304.53	7.85	11	5.98	10	9.87	3.2	10
	Cu	46.80	1.53	11	0.68	10	1.67	3.6	10
	Ni	35.79	0.97	11	0.52	10	1.10	3.1	10
	Pb	124.56	6.78	11	2.44	10	7.21	5.8	10
	Zn	107.02	1.82	11	2.77	10	3.32	3.1	10
LGC6145	Cr	41.50	0.80	11	0.91	10	1.22	2.9	10
	Cu	49.83	2.04	11	0.00	10	2.04	4.1	11
	Ni	20.24	1.36	11	0.44	10	1.43	7.1	10
	Pb	37.55	0.88	11	0.00	10	0.88	2.3	11
	Zn	136.00	3.74	11	3.36	10	5.03	3.7	10
LGC6187	Cr	75.66	1.45	11	2.53	10	2.92	3.9	10
	Cu	73.08	1.46	11	1.94	10	2.43	3.3	10
	Ni	30.21	0.44	11	0.37	10	0.57	1.9	10
	Pb	67.14	0.65	11	0.73	10	0.98	1.5	10
	Zn	402.62	6.47	11	5.46	10	8.46	2.1	10

A common feature of random variation in chemical measurement is a strong dependence on level; for many routine measurements, the *relative* standard deviation is almost constant above detection limits. This is the case here: Figure E3.4.2, which plots the intermediate reproducibility standard deviation s_I against mean measured mass fraction for all four materials and six metals, shows a clear and approximately linear dependence (though note the log scales). Regression on log-transformed data confirms that a linear dependence is both strongly significant ($p = 7.8 \times 10^{-8}$) and sufficient, returning $p = 0.47$ for an added curvature term. The relative standard deviation is fairly constant between 30 mg kg^{-1} and 1000 mg kg^{-1} with the exception of an unusually high value for zinc in material C279a. The high value arises from an unusual and apparently material-specific systematic effect on replicate measurements during the experiment; excluding the anomalous value for zinc in C279a, the median intermediate precision RSD is 3.2 %.

Note: The median is cited here as a summary as it is not generally safe to average standard deviations or relative standard deviations. For standard deviations, the root mean square is more appropriate. With this in mind, a plausible and somewhat more conservative alternative is the root mean square RSD, 3.6 %.

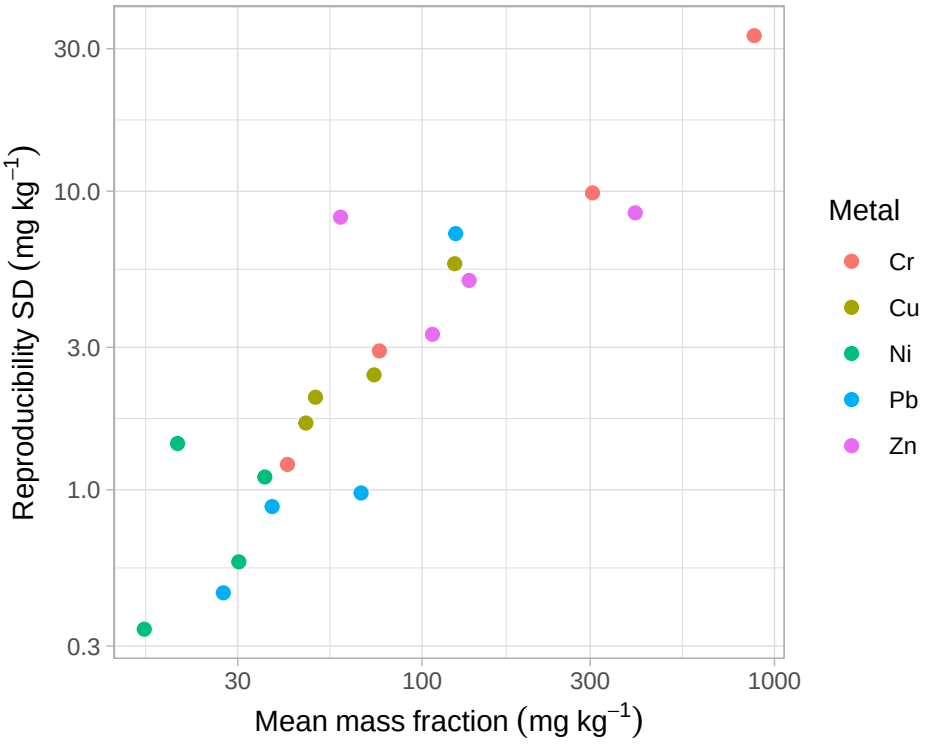


Figure E3.4.2: Dependence of reproducibility SD on mean mass fraction

E3.4.6.3 Measurement bias – evaluation of $\hat{\delta}$ and $u(\hat{\delta})$

Bias and uncertainty associated with bias can be assessed from the precision experiment, given reference values for the materials studied. Two of the materials studied (LGC6145 and LGC6187) were certified reference materials; C279a and C282a were proficiency test (PT) materials and the reports provide assigned values with uncertainties based on the interlaboratory consensus value. In all four cases, the reference values are values obtained by laboratories applying ISO 11466 or demonstrably equivalent procedures, allowing an estimate of the present procedure's bias against reference values for the correct measurand. Bias estimates $\hat{\delta}$ are just the difference $\bar{x}_{\text{obs}} - x_{\text{ref}}$ between the reference and observed values x_{ref} and $\bar{x}_{\text{obs}}$ respectively.

The uncertainty for the bias estimates is a combination of uncertainties for the observed and reference values. The statistical uncertainty for the observed mean value in a nested design needs to take account of the fact that replicates within a group are not independent. This can be done in a number of ways. ISO 5725-4 [238] and ISO 21748 [232] give formula (E3.4.4) for the standard error of the mean, $s_{\hat{\delta}}$, in such a design:

$$s_{\hat{\delta}} = \sqrt{\frac{s_{\text{I}}^2 - (1 - 1/n)s_{\text{r}}^2}{p}}, \tag{E3.4.4}$$

where p and n and are the number of groups and the number of replicates per group, respectively, and s_{r} and s_{I} are the repeatability and intermediate precision standard deviations in table E3.4.1. In equation (E3.4.4), s_{I} has been substituted for the reproducibility standard deviation s_{R} used in ISO 21748.

The standard error from the precision study can be combined with the standard uncertainty u_{ref} for the reference value using

$$u^2(\hat{\delta}) = s_{\hat{\delta}}^2 + u_{\text{ref}}^2, \tag{E3.4.5}$$

to give the standard uncertainty $u(\hat{\delta})$ for the estimated bias. Application of equations (E3.4.4) and (E3.4.5) gives the results in table E3.4.2. A pictorial representation is also given in figure E3.4.3.

Inspection of the bias estimates shows immediately that, with the exception of LGC6145, which appears to need separate consideration, the modified procedure shows a consistent low bias near 10 % in this experiment. (An analytical chemist might consider this in terms of "recovery" of the analyte; a negative bias of 10 % measured on reference materials could be reported as an apparent recovery of 90 %). Although the bias estimates meet the regulatory performance requirements and therefore would not normally give rise to an applied correction, this presents a problem for uncertainty evaluation; the GUM offers no consistent approach to the treatment of a known but uncorrected bias. This is considered further in sections E3.4.6.5 and E3.4.7.

LGC6145 is a contaminated clay loam soil. Clays can show poor analyte recovery in extraction procedures; given the poorer recovery obtained for this material in this study, the laboratory may choose to use the more aggressive extraction conditions of the standard test method for regulatory samples with higher clay content. Alternatively, the modified procedure – which shows excellent precision – may be reserved for applications which prioritise precision over bias, such as homogeneity testing.

For the purpose of equation (E3.4.3) applied to future test materials, it is useful to retain a single summary estimate for $u(\hat{\delta})$. Inspection of table E3.4.2 shows that individual uncertainties again depend strongly on mean mass fraction, whereas the relative standard uncertainties $u(\hat{\delta})/x_{\text{ref}}$ are much more consistent. A reasonable summary is, again, the root mean square relative standard uncertainty, 3.3 %; for comparison the median and mean are 2.9 % and 3.3 %.

Table E3.4.2: Bias estimates and uncertainties for *aqua regia* extractable metals in soil. All values in mg kg^{-1}, unless stated otherwise

Material	Metal	x_{ref}	u_{ref}	x_{obs}	u_{obs}	Bias $\hat{\delta}$	$u(\hat{\delta})$	$\hat{\delta}/x_{\text{ref}}$ (%)	$u(\hat{\delta})/x_{\text{ref}}$ (%)
C279a	Cr	1007.50	14.55	875.80	8.89	–131.7	17.0	–13.1	1.7
	Cu	136.69	2.87	123.77	1.50	–12.9	3.2	–9.5	2.4
	Ni	18.16	0.51	16.28	0.07	–1.9	0.5	–10.3	2.8
	Pb	30.00	0.83	27.29	0.10	–2.7	0.8	–9.0	2.8
	Zn	64.40	1.63	58.70	1.75	–5.7	2.4	–8.8	3.7
C282a	Cr	321.00	13.88	304.53	2.46	–16.5	14.1	–5.1	4.4
	Cu	55.30	1.84	46.80	0.39	–8.5	1.9	–15.4	3.4
	Ni	38.60	2.17	35.79	0.26	–2.8	2.2	–7.3	5.7
	Pb	136.91	8.27	124.56	1.62	–12.4	8.4	–9.0	6.2
	Zn	115.90	2.75	107.02	0.92	–8.9	2.9	–7.7	2.5
LGC6145	Cr	47.60	0.90	41.50	0.32	–6.1	1.0	–12.8	2.0
	Cu	62.20	1.80	49.83	0.44	–12.4	1.9	–19.9	3.0
	Ni	39.00	1.25	20.24	0.32	–18.8	1.3	–48.1	3.3
	Pb	45.10	1.15	37.55	0.19	–7.6	1.2	–16.8	2.6
	Zn	137.00	3.00	136.00	1.29	–1.0	3.3	–0.7	2.4
LGC6187	Cr	84.00	4.70	75.66	0.82	–8.3	4.8	–9.9	5.7
	Cu	83.60	2.05	73.08	0.66	–10.5	2.2	–12.6	2.6
	Ni	34.70	0.85	30.21	0.15	–4.5	0.9	–12.9	2.5
	Pb	77.20	2.25	67.14	0.26	–10.1	2.3	–13.0	2.9
	Zn	439.00	13.00	402.62	2.15	–36.4	13.2	–8.3	3.0

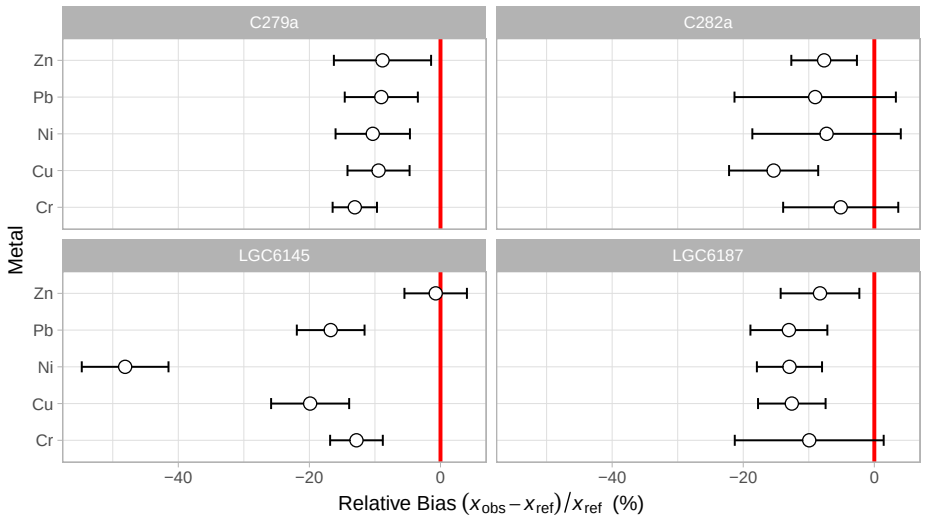


Figure E3.4.3: Relative bias for *aqua regia* extractable metals in soil. The vertical red line is at zero bias. Error bars show expanded uncertainty with $k = 2$.

E3.4.6.4 Identification and evaluation of additional uncertainty sources

Identifying additional contributions

Most of the steps in figure E3.4.1 are repeated independently in the different measurement runs contributing to the precision and bias assessment. Reagents were made up with different volumetric glassware; the ICP-OES instrument calibration is repeated for each run, and each replicate uses a different subsample of the test materials. Two potentially important features did, however, remain constant throughout the exercise. First, the extraction temperature for the high-temperature portion of the extraction was largely constant owing to the adoption of a consistent heat setting within a potentially wider range. Second, the calibration standards used to prepare stock solutions were constant for the complete exercise. Uncertainties arising from these two sources therefore merit further attention and, possibly, inclusion in equation (E3.4.3). The following subsections accordingly consider these two sources.

Extraction condition effects

As noted above, there are no simple, general models that can predict the effect of changes in extraction conditions such as times, temperatures and reagent concentrations on measured mass fraction. In the absence of applicable theoretical models, the most useful way to evaluate uncertainties in the output quantity (measured mass fraction, in this case) that arise from uncertainties in the specified extraction conditions is to estimate sensitivity coefficients empirically, by experiment. This is illustrated here using a short experimental study.

The most important factors in this extraction process are expected to be the time t_{ext} and temperature T_{ext} for the period of heating. The nominal time is two hours (120 min); the nominal temperature, following the laboratory's standard operating procedure, is 65 °C. The operating tolerance for the extraction time, given in the laboratory operating procedure, is 5 min; this is a permitted range rather than a measurement uncertainty, as the time is measured by standard laboratory times with uncertainty of a second or better. No exact tolerance is given for temperature as it is the liquid temperature that is specified; however, checks on extraction temperatures at stable settings indicate that the liquid temperature is maintained well within 5 °C of the nominal temperature in normal use. To estimate the sensitivity of measured values to change in the

extraction conditions, the laboratory accordingly performed a short study in which the relevant time and temperature were varied by, respectively, ±15 min and approximately ±10 °C from their nominal values. The study was conducted as a two-factor, two-level full factorial in duplicate, giving two observations for all four combinations of time and temperature, and the different treatments were applied to two test materials. All the extractions were run in parallel using two separate heating blocks for the different temperatures. After the extraction, solutions were diluted to the requisite standardised volume and run in randomised order in a single run to give the measured mass fraction of *aqua regia* extractable metal. The results (omitting those for a blank control material) are shown in figure E3.4.4, in which the panel headings identify the test material and metal of interest, and observations are colour coded to indicate the extraction time.

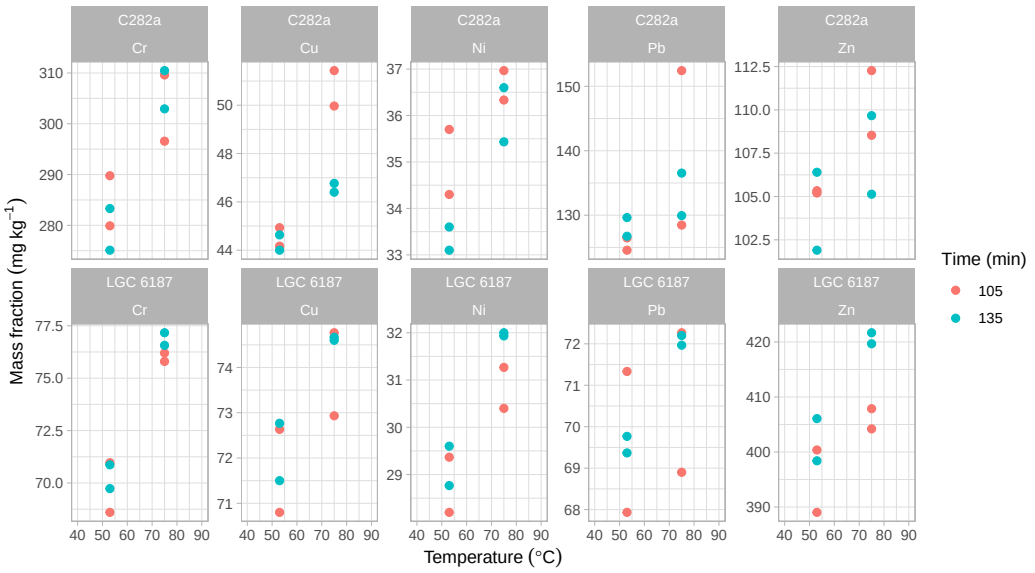


Figure E3.4.4: Results from an experimental study of time and temperature effects on *aqua regia* extraction. See the text for details.

There are, as might be expected in soil analysis, some possible anomalies. The higher temperature results for lead in material C282a appear unusually imprecise. Reference to the laboratory indicated that the highest value for lead for C282a was considered invalid for technical reasons and the observation was accordingly omitted from further analysis. Observations at 105 min for lead in material LGC 6187 also appear comparatively imprecise; in the absence of specific reasons to remove any, however, these were retained for further analysis.

All of the sets of results, however, appear to show a clear increase in measured mass fraction over the 20 °C change in temperature, with little evidence of a consistent change with time.

Using such information to support measurement uncertainty evaluation requires some further analysis. In principle, the results in each panel in figure E3.4.4 provide an approximate individual rate of change of measured mass fraction by time and temperature; these are the sensitivity coefficients needed for a basic uncertainty evaluation following the GUM. The data also provide information about interactions between the two conditions. The coefficients and any interactions might be specific to each material and to each element, or might be sufficiently consistent to generalise over modest ranges.

One simple way of proceeding, which requires only basic statistical tools, is to examine each metal/material subset individually, checking for any strong time/temperature interactions before estimating rates of change, and for this illustration we adopt this approach.

A two-way ANOVA (not shown here) for each of the subsets in figure E3.4.4 indicates that the effect of temperature change is significant ($p < 0.05$) for all but zinc in material C282a and lead in both materials, while the time/temperature interaction is statistically significant ($p = 0.013$) only for copper in C282a. Considered as one of ten hypothesis tests, and after reviewing the plot for this case, this is not strong evidence for an important interaction term. Similar conclusions are drawn by comparison of alternative linear models using the small-sample corrected Akaike's information criterion (AICc) [239,240], which give a consistently lower AICc for models without a time/temperature interaction term, indicating that the interaction term is not needed.

Most of the data sets also showed no significant effect for change in extraction time. There is, however, reason to expect a modest effect of time on extraction, and an estimated gradient – even one that is small compared to residual variation - is of interest in evaluating the uncertainty in measured mass fraction that arises from uncertainties in extraction time. Table E3.4.3 accordingly provides estimated coefficients for both temperature and time, together with standard errors, derived from a simple linear model including both time and temperature.

Table E3.4.3: Time and temperature effects on *aqua regia* extractable metals in soil. Means are in mg kg^{-1}, sensitivity coefficients and their standard deviation in mg kg^{-1} °C^{-1} and mg kg^{-1} min^{-1} for temperature and time effects respectively, and relative sensitivity coefficients in °C^{-1} and min^{-1}

Material	Metal	Mean $\bar{y}$	$\frac{\partial y}{\partial T_{\text{ext}}}$	$s\left(\frac{\partial y}{\partial T_{\text{ext}}}\right)$	$\frac{\partial y}{\partial t_{\text{ext}}}$	$s\left(\frac{\partial y}{\partial t_{\text{ext}}}\right)$	$\frac{\partial y}{\partial T_{\text{ext}}}\Big/\bar{y}$	$\frac{\partial y}{\partial t_{\text{ext}}}\Big/\bar{y}$
C282a	Cr	293.5	1.040	0.222	-0.034	0.163	0.0035	-0.0001
	Cu	46.5	0.191	0.044	-0.072	0.032	0.0041	-0.0016
	Ni	35.3	0.098	0.023	-0.038	0.017	0.0028	-0.0011
	Pb	128.9	0.191	0.096	0.117	0.070	0.0015	0.0009
	Zn	106.8	0.191	0.078	-0.069	0.057	0.0018	-0.0006
LGC 6187	Cr	73.2	0.291	0.028	0.023	0.020	0.0040	0.0003
	Cu	73.1	0.105	0.030	0.020	0.022	0.0014	0.0003
	Ni	30.2	0.110	0.019	0.026	0.014	0.0036	0.0008
	Pb	70.5	0.079	0.051	0.024	0.038	0.0011	0.0003
	Zn	405.9	0.677	0.162	0.369	0.119	0.0017	0.0009

The coefficients (denoted $\partial y/\partial T_{\text{ext}}$ and $\partial y/\partial t_{\text{ext}}$ in the table) could, in principle, be used directly as estimates of coefficients c_i for T_{ext} and t_{ext} in equation (E3.4.3). In practice, these coefficients are specific to the test materials in the study, and for routine application of the procedure they are best used to give a reasonable estimate of the typical size of uncertainties arising from the extraction time and temperature. Inspection suggests that, as is often the case, the effects increase with measured value; for generalisation, relative rates of change are useful; the relative rates of change in the final two columns of table E3.4.3 show that the temperature effect is in the range $(0.001 - 0.004)\,°C^{-1}$, while the (unsigned) effects of time are in the range $(10^{-4} - 0.002)$ min^{-1}. A reasonably conservative estimate of the typical relative effect for routine estimation (rounded to one significant figure) might be $0.004\,°C^{-1}$ and 9×10^{-4} min^{-1} for temperature and time respectively, or approximately $0.4\,\%\,°C^{-1}$ and $0.09\,\%$ min^{-1}.

Use of the coefficients above also requires standard uncertainties for the time and temperature. The extraction time is controlled to within ±5 min by the procedure; this can be treated as a rectangular distribution and a standard uncertainty obtained by division of the half-range by $\sqrt{3}$. The temperature range is less clearly defined; the standard operating procedure instructs the

user to maintain the temperature at "about 65 °C". In practice, laboratory checks show that the temperature is maintained to within 5 °C. This can also be treated as a rectangular distribution, giving a standard uncertainty of 2.89 °C.

ICP-OES Instrument calibration

Uncertainties from zero-intercept regression The instrument calibration step is simplified in figure E3.4.1. Calibration involves observation of the emission signal from a number of working standards of known concentration, which are prepared in the laboratory from a stock solution, followed by regression to determine a calibration line. Uncertainties include those for the concentration of the stock solution, uncertainties in dilution to the working standards, and uncertainties arising from the regression (which include the random variations in observed signal). Because uncertainty evaluation for zero-intercept calibration is rarely treated in the general literature, the complete process is considered before returning to the sources needed in applying equation (E3.4.3).

In this case, calibration used four standard solutions including one at zero concentration (that is, a blank solution) and three prepared from a certified stock solution of concentration γ_0, diluted by mass. For simplicity the degree of dilution is denoted by a dilution factor d_i, as in figure E3.4.1; a given working standard concentration γ_i is just $d_i\gamma_0$. In practice, dilution is stepwise so that the d_i are in turn products of two or more larger dilution factors, calculated as mass ratios when diluting by mass. Uncertainties for typical working standard solutions, as used in this example, are given in are given in the file `Soil_cal_stds_v1.txt` in [22]. Dilution by mass, rather than using volumetric equipment, leads to relatively small dilution uncertainties. However, because the two intermediate concentration solutions were prepared by dilution from the highest concentration working solution (in turn prepared direct from certified stock), uncertainties in the concentrations are highly correlated ($r > 0.99$), calculated as in [241]). Combined uncertainties are small; the largest relative standard uncertainty for a calibration solution concentration (for Ni) is approximately 0.0025 and most are below 0.0013. Since the uncertainty for dilution stages is small and uncertainty evaluation can be intricate, detailed discussion of the uncertainty evaluation for the dilutions is omitted here. A detailed presentation of uncertainties arising in the preparation of calibration solutions by dilution is provided in a Eurachem guide [42].

Uncertainties associated with the regression are determined by the ordinary statistical methods. In this instance, rather than a four-point linear regression with slope and intercept, the laboratory uses a still simpler model, setting the intercept to zero after subtracting the mean blank value. This partially offsets the effect of heteroscedasticity, which (because the higher observations are more variable) would lead to large relative variation for low interpolated observations. (Weighted regression would also be effective in avoiding this, but calibration points are not replicated in routine use and there is then insufficient information to generate reliable weights). Two typical calibrations are shown in figure E3.4.5.

> NOTE: In practice, the laboratory collects emission intensities for two separate emission lines for each element to provide spectral interference checks or an alternative calibration in case of interference. Both signals are included in the data set associated with this example. The analysis here assumes the higher intensity emission line is used for reported results.

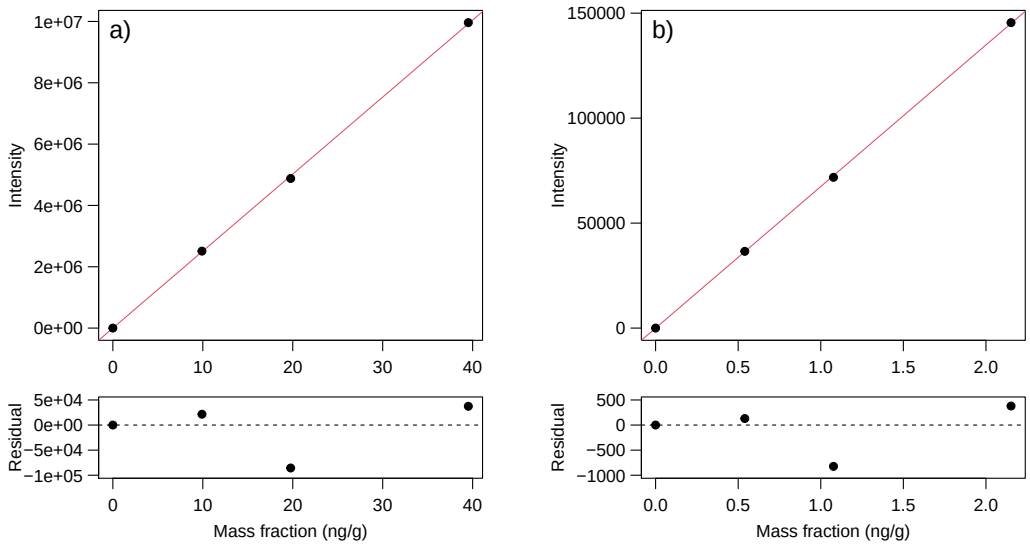


Figure E3.4.5: ICP-OES calibration examples. a) chromium; b) nickel. The upper panel shows the fitted line; the lower panel shows the residuals. The intensity is corrected for the blank signal, omitted from the fit but included in the plot at [0,0]. Correlation coefficients r^2 were 0.999 93 and 0.999 97, respectively.

Estimation of an extract concentration γ_{ext} from the fitted slope b_1 and an observed intensity y_{obs} uses equation (E3.4.6):

$$\gamma_{\text{ext}} = \frac{y_{\text{obs}} - y_0}{b_1}, \tag{E3.4.6}$$

where y_0 is the intensity for a blank sample. Usually, this will be the same blank signal as that subtracted prior to fitting the calibration line. Unfortunately, this complicates an otherwise simple uncertainty evaluation, as changes in blank reading will affect the estimated slope, generating a non-zero covariance between the estimated slope and the blank sample intensity. Further, the resulting covariance is not available from the fitted regression model because the blank subtraction step is applied before the regression fitting.

There are two ways of handling this covariance problem. The first is a simple approximation that relies on the fact that the standard deviation of emission near zero is typically much smaller than that of the signals for other working standards. Where this is the case, the covariance can be taken as negligible and the standard error for the slope is (together with a standard deviation for y_{obs}) sufficient for uncertainty evaluation.

Where the blank signal y_0 has a non-negligible standard deviation, it is possible to estimate the covariance using the fact that the slope b_1 for a zero-intercept model (without variance weighting) is given by

$$b_1 = \frac{\sum_{i=1}^{n}(y_i - y_0)\gamma_i}{\sum_{i=1}^{n}\gamma_i^2}, \tag{E3.4.7}$$

in which y_i is the observed intensity for a working standard with concentration γ_i and n is the number of observations in the regression. The covariance $s^2(b_1, y_0)$ is then given by

$$\begin{aligned} s^2(b_1, y_0) &= \frac{\partial b_1}{\partial y_0} s^2(y_0) \\ &= -\frac{\sum_{i=1}^{n}\gamma_i}{\sum_{i=1}^{n}\gamma_i^2} s^2(y_0), \end{aligned} \tag{E3.4.8}$$

where $s(y_0)$ is the standard deviation (or, if y_0 is an average, standard error) for y_0.

Either approach needs an estimate of the standard deviation for blank samples. Fortunately, while the calibration standards are not replicated in routine application of the measurement procedure, a separate reagent blank is measured at intervals in each measurement run to provide QC information. For the present example, eight such blank observations are available. Omitting one blank sample with anomalous readings, which followed the observation of the highest working calibration standard, blank standard deviations for each element are given in table E3.4.4, for the emission line with highest intensity in each case. The table also gives the residual standard deviation for the corresponding calibrations. In all cases, the blank sample standard deviation is at least an order of magnitude less than the residual standard deviation for the calibration. With such a large difference, it is clearly safe to omit the contribution of the variability of the blank signal, and the small covariance between the blank sample value y_0 and the estimated gradient b_1, in estimating the uncertainty of a predicted value γ_{ext}.

Table E3.4.4: Blank and calibration residual standard deviations (SD) for soil analysis elements

Element	Lambda	Blank SD	Residual SD
Cr	283.563	188.1	67788
Cu	324.752	135.8	14021
Ni	231.604	29.1	647
Pb	220.353	24.3	413
Zn	202.548	137.1	13736

Omitting the small blank value uncertainty leads to a relatively simple uncertainty evaluation for γ_{ext} (equation (E3.4.6)). The statistical uncertainty in the gradient b_1 is taken as the standard error of the slope coefficient in the relevant model. The uncertainty associated with y_{obs} is taken as the residual standard deviation for the calibration (unless otherwise available). Dilution uncertainties are considered negligible (see above). The uncertainty associated with the certified stock solution concentration γ_0 does, however, need to be considered.

Since the working standard solution concentrations are simple multiples of the stock solution concentration, which in turn directly affect the gradient estimate, one simple way of incorporating the uncertainty in γ_0 is to include the relative standard uncertainty for the stock solution concentration as an additional term in a simple combination of relative standard uncertainties associated with b_1 and $y_{\text{obs}} - y_0$ [42]. This gives

$$\begin{aligned}\frac{u_{\text{cal}}(\gamma_{\text{ext}})}{\gamma_{\text{ext}}} &= \sqrt{\left(\frac{u(y_{\text{obs}})}{y_{\text{obs}}-y_0}\right)^2+\left(\frac{u(b_1)}{b_1}\right)^2+\left(\frac{u(\gamma_0)}{\gamma_0}\right)^2} \\ &\approx \sqrt{\left(\frac{u(y_{\text{obs}})}{y_{\text{obs}}}\right)^2+\left(\frac{u(b_1)}{b_1}\right)^2+\left(\frac{u(\gamma_0)}{\gamma_0}\right)^2},\end{aligned} \tag{E3.4.9}$$

where $u_{\text{cal}}(\gamma_{\text{ext}})$ is the standard uncertainty arising from instrument calibration and determination of the signal intensity for an extract solution. The approximation in the second line holds when the blank correction y_0 is small compared to y_{ext} (for example, $y_0 < 0.1 y_{\text{ext}}$ would not materially affect the uncertainty).

As an example, consider the calibration for chromium in figure E3.4.5 a), and assume an observed signal intensity at 3.7×10^6, close to the median of the calibration range. Table E3.4.5 gives the relevant values and standard uncertainties.

Table E3.4.5: Values and uncertainties for simplified calibration uncertainty. Observed values are given as recorded instrument intensity (arbitrary units).

	x	u	u/x	Remarks
y_{obs}	3700000.0	68000.0	0.0180	From residual SD.
y_0	888.7	190.0	0.2100	From QC blanks.
b_1	251100.0	1500.0	0.0060	From fitted calibration model.
γ_0 /(mg kg^{-1})	980.4	1.2	0.0012	From certificate.

Equation (E3.4.6) gives a value for γ_{ext} of 14.7 mg l^{-1}. Applying the approximate form of equation (E3.4.9) then gives

$$\begin{aligned}\frac{u_{\text{cal}}(\gamma_{\text{ext}})}{\gamma_{\text{ext}}} &= \sqrt{\left(\frac{6.8\times10^4}{3.7\times10^6}\right)^2+\left(\frac{1500}{2.511\times10^5}\right)^2+\left(\frac{1.2}{980.4}\right)^2}\\ &= \sqrt{0.018^2+0.006^2+0.0012^2}\\ &= 0.019,\end{aligned} \tag{E3.4.10}$$

that is, a relative standard uncertainty of about 2 %, largely driven by the uncertainty associated with y_{obs}, in turn from the residual standard deviation in the calibration. If required, multiplying by γ_{ext} gives a standard uncertainty for γ_{ext} of 0.3 mg l^{-1}, to one significant digit.

Equation (E3.4.9) is often sufficient in a chemical testing environment. It gives a relatively simple indication of the uncertainty associated with use of a zero-intercept calibration with prior blank correction, and can be implemented easily in spreadsheet software. It can provide a relative standard uncertainty for γ_{ext} without the need for an estimated value of γ_{ext}, using only relative uncertainties as input; this is often sufficient when the majority of measurement results fall in a region in which the relative standard uncertainty is approximately constant. Finally, noting the simple multiplicative form of equation (E3.4.1), the relative standard uncertainty can be compared directly with other relative uncertainties to determine whether it is important.

For an explicit model-based evaluation of uncertainty, the stock solution concentration γ_0 must appear explicitly in a calculation for γ_{ext}. This could be, for example, in a prior stage of a multi-stage model, contributing to the uncertainty in b_0; in an extension of equation (E3.4.6); or in a software function whose output can be used to derive numerical sensitivity coefficients or generate a Monte Carlo simulation. One possible approach, which uses the fact that all of the working standard concentrations γ_i are just multiples of the stock solution concentration by a dilution factor d_i, is shown below.

Writing γ_i as $d_i\gamma_0$ in equation (E3.4.7) gives

$$\begin{aligned}b_1 &= \frac{\sum_{i=1}^{n}(y_i-y_0)\,d_i\gamma_0}{\sum_{i=1}^{n}(d_i\gamma_0)^2}\\ &= \frac{\sum_{i=1}^{n}(y_i-y_0)\,d_i}{\gamma_0\sum_{i=1}^{n}(d_i)^2}.\end{aligned} \tag{E3.4.11}$$

Substituting in (E3.4.6) in turn gives

$$\begin{aligned}\gamma_{\text{ext}} &= \frac{y_{\text{obs}}-y_0}{\frac{\sum_{i=1}^{n}(y_i-y_0)d_i}{\gamma_0\sum_{i=1}^{n}(d_i)^2}}\\ &= \gamma_0\frac{y_{\text{obs}}-y_0}{\frac{\sum_{i=1}^{n}(y_i-y_0)d_i}{\sum_{i=1}^{n}(d_i)^2}}.\end{aligned} \tag{E3.4.12}$$

Differentiating:

$$\frac{\partial \gamma_{\text{ext}}}{\partial \gamma_0} = \frac{y_{\text{obs}} - y_0}{\frac{\sum_{i=1}^{n}(y_i - y_0)d_i}{\sum_{i=1}^{n}(d_i)^2}} = \frac{\gamma_{\text{ext}}}{\gamma_0}, \tag{E3.4.13}$$

so that a model for γ_{ext} that includes the effect of deviations $\delta_{\gamma_0} = \gamma_0 - \hat{\gamma}_0$ from the nominal stock concentration $\hat{\gamma}_0$ can be written

$$\gamma_{\text{ext}} = \frac{y_{\text{obs}} - y_0}{b_1} + \frac{\hat{\gamma}_{\text{ext}}}{\hat{\gamma}_0}\delta_{\gamma_0}, \tag{E3.4.14}$$

where $\hat{\gamma}_{\text{ext}}$ is the point estimate of γ_{ext}.

A rather simpler model, noting that γ_{ext} is directly proportional to γ_0, is simply

$$\gamma_{\text{ext}} = \frac{\hat{\gamma}_0}{\gamma_0}\frac{y_{\text{obs}} - y_0}{b_1}, \tag{E3.4.15}$$

which is a proportional correction for deviations from the nominal stock concentration $\hat{\gamma}_0$.

Contribution to combined uncertainty Section E3.4.6.4 evaluates uncertainties arising from zero-intercept regression, including random variation. However, equation (E3.4.3) requires only those contributions which do not already contribute to the precision estimates and bias uncertainty. The intermediate precision standard deviations in section E3.4.6.2 were obtained from multiple independent measurement runs, each with its own set of dilutions and its own calibration run. The random variation in (for example) y_{obs} that dominates equation (E3.4.9) has therefore already contributed to the intermediate precision estimates in table E3.4.1. It follows that, useful as equation (E3.4.9) is for evaluating the uncertainty sources in this part of the procedure, using the resulting uncertainty in γ_{ext} in equation (E3.4.3) would have the effect of "double-counting" the contribution of random variation in the calibration step. Since the mass calibration uncertainties in the dilutions are negligible (see above), the only uncertainty source that needs to be considered in the summation term of equation (E3.4.3) is the uncertainty associated with γ_0, the calibration standard value. Equation (E3.4.15) demonstrates that this can be treated as a simple proportional effect on the measured value w_s; it follows that the coefficient c_{γ_0} will just be w_s/γ_0. Where uncertainties are combined in the form of relative standard uncertainties, the relative standard uncertainty $u(\gamma_0)/\gamma_0$ can be used directly. The relevant stock concentration values (coverted to a mass fraction basis as the laboratory dilutes by mass) and the associated uncertainties are summarised in table E3.4.6.

Table E3.4.6: Calibration standards – Concentration x and standard uncertainty u

Element	$x/(\text{mg kg}^{-1})$	$u/(\text{mg kg}^{-1})$	u/x
Cr	980.40	1.18	0.0012
Cu	983.80	1.03	0.0010
Ni	983.60	1.08	0.0011
Pb	984.74	0.98	0.0010
Zn	983.86	0.98	0.0010

E3.4.6.5 Combined uncertainty

Following ISO 21748, the combined uncertainty for a measured value w_s is obtained by applying equation (E3.4.3), with relevant coefficients c_i and uncertainties $u(x_i)$ for the important additional contributions identified in section E3.4.6.4. Writing this explicitly for the particular quantities involved:

$$\begin{aligned} u^2(w_s) =& u^2(\hat{\delta}) + s_\mathrm{I}^2 + c_{T_\mathrm{ext}}^2 u^2(T_\mathrm{ext}) \\ &+ c_{t_\mathrm{ext}}^2 u^2(t_\mathrm{ext}) + c_{\gamma_0}^2 u^2(\gamma_0). \end{aligned} \tag{E3.4.16}$$

Since so many of the estimated effects in this example are approximately proportional to measured value and have been estimated as relative effects, it is useful to recast equation (E3.4.16) in terms of an estimated relative standard uncertainty $u(w_s)/w_s$, giving

$$\begin{aligned} \left(\frac{u(w_s)}{w_s}\right)^2 =& \left(\frac{u(\hat{\delta})}{w_s}\right)^2 + \left(\frac{s_\mathrm{I}}{w_s}\right)^2 + \left(\frac{c_{T_\mathrm{ext}}}{w_s} u(T_\mathrm{ext})\right)^2 \\ &+ \left(\frac{c_{t_\mathrm{ext}}}{w_s} u(t_\mathrm{ext})\right)^2 + \left(\frac{c_{\gamma_0}}{w_s} u(\gamma_0)\right)^2, \end{aligned} \tag{E3.4.17}$$

which is just a combination of relative uncertainties in measured value.

The object of the uncertainty evaluation exercise is to obtain a reasonable indication of the measurement uncertainty that can be expected from future application of the measurement procedure to typical test materials. For this purpose, the individual estimates in tables E3.4.1–E3.4.3 are not used directly; rather, the approximate summary figures given in the relevant sections are used. The relevant information is collected in table E3.4.7. Information for the calibration stock solutions is specific to each element and is given in table E3.4.6; for illustration, the values for lead (Pb) are included in table E3.4.7.

Table E3.4.7: Uncertainty information for *aqua regia* extraction of soils

Term	Value	Unit	Remarks
$u(\hat{\delta})/w_s$	0.033	–	Estimated as relative standard uncertainty
s_I/w_s	0.036	–	Estimated as relative standard deviation
c_{T_ext}/w_s	0.004	$^\circ\mathrm{C}^{-1}$	Relative effect on w_s
$u(T_\mathrm{ext})$	2.89	°C	Rectangular, with limits ±5
c_{t_ext}/w_s	9×10^{-4}	min^{-1}	Relative effect on w_s
$u(t_\mathrm{ext})$	2.89	min	Rectangular, with limits ±5
$c_{\gamma_0}(\mathrm{Pb})/w_s$	1/984.74	–	See section E3.4.6.4
$u(\gamma_0(\mathrm{Pb}))$	0.985	$\mathrm{mg\ kg}^{-1}$	

Using equation (E3.4.17) and the values in table E3.4.7, the relative standard uncertainty $u'(\mathrm{Pb})$ for *aqua regia* extractable lead in the range studied is given by

$$\begin{aligned}\left(u'(\mathrm{Pb})\right)^2 &= 0.033^2 + 0.036^2 + (0.004 \times 2.89)^2 \\ &\quad + (0.0009 \times 2.89)^2 + \left(\frac{1}{984.74} \times 0.985\right)^2 \\ &= 0.033^2 + 0.036^2 + 0.01156^2 \\ &\quad + 0.002601^2 + 0.0010003^2 \\ &= 0.002\,526,\end{aligned} \tag{E3.4.18}$$

giving a relative standard uncertainty $u'(\mathrm{Pb})$ of $\sqrt{0.002\,526} = 0.05$, or about 5 %. Noting the very small contribution of the calibration standard uncertainty $u(\gamma_0)$ and the similarity in calibration solution relative uncertainties (all near 0.001), this would serve sufficiently well for all of the elements in the present study.

E3.4.7 Reporting the result

Given a measurement result such as that for lead in C279a, the measurement result with expanded uncertainty could be reported as

> "(27.8 ± 2.8) mg kg^{-1}, where the uncertainty is an expanded uncertainty using a coverage factor $k = 2$, giving a level of confidence of approximately 95 %."

In addition, it is important to consider the presence of an appreciable bias, of the order of 10 %, in the measured values. Neither the GUM nor ISO 21748 give guidance on treatment of this issue. Eurachem, however, recommend that, where a known significant bias is not corrected or eliminated by improvements to the procedure, the bias should be reported along with the measurement result and its uncertainty [42]. This is the approach taken here. In the absence of established reporting arrangements, the above report of the measured concentration should be accompanied by an advisory statement along the lines of

> "The procedure in use shows apparent recovery of approximately 90 %. Since the recovery is within permitted limits, no correction for recovery has been applied."

E3.4.8 Conclusion

Use of ISO 21748 provides a indication of measurement uncertainty based on the results of precision and trueness studies, supplemented by further contributions that are not adequately represented in a given precision study. The present example has applied this approach, using data acquired during within-laboratory validation of a modified test procedure. The data included measurements of five elements in four different reference materials, allowing simultaneous assessment of precision and bias.

Some important features of the approach have been explored. First, the raw data generate a range of individual precision and bias estimates from each material/element pair. These cannot be applied directly to future materials; rather, it is important to extract a representative estimate from the collected data. This requires some judgement as to the most appropriate estimate.

Second, the measurement procedure involves a number of important steps which have considerable influence on the measured value, but are hard to model mathematically. This made it impossible to obtain suitable sensitivity coefficients by analytical differentiation or computational methods. Experimental studies were accordingly used to determine the applicable coefficients for the most important step in the procedure.

Finally, the procedure in use showed a significant bias which was, nonetheless, within permitted performance limits for the intended use. This presents problems for measurement uncertainty evaluation, as there is no straightforward method for incorporating a bias in a measurement uncertainty statement. In this example, Eurachem guidance [42] was followed and the bias is reported as additional information with the measurement result and uncertainty.

Example E3.5

Comparison of methods for flow measurement in closed conduits based on measurement uncertainty

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

E3.5.1 Summary

In some fields of science, there is a wide diversity of technical principles and methods that can be used to obtain measurands. This is the case in the measurement of flow rate, in closed conduits in water supply facilities. Measurement of flow related to water supply infrastructures has become in recent years a major topic for research for several reasons:

1. the growing impact of climate change;
2. the need to develop sustainable management of resources;
3. the need for robust management of networks of increasing complexity;
4. the criticality of these infrastructures for urban management with high social and economic impact; and
5. the increasing perception of the need for measurement accuracy and consistency in developing modelling for interpretation and forecasting in distinct fields, namely, climatology, hydrology and urban hydraulics.

Flow rate is a quantity often measured to evaluate instantaneous values and to obtain cumulative values of volume by integration, usually measured by a diversity of instruments using distinct principles of operation and techniques able to provide an indirect measurement of the quantity. Instruments most used for flow rate measurement are based either on the relation between the electromagnetic field and flow velocity or on the relation between the propagation of the velocity of sound in a fluid and the flow velocity. In these cases, the comparison of methods for the selection of the measuring system should consider the advantages and disadvantages of each solution, considering the nature and specific characteristics of the fluid, the behaviour of the flow, and the performance of the instrumentation under those circumstances. For the last-mentioned, measurement uncertainty and sensitivity analysis should be considered as a major source of information for decision making, this being the main motivation for this example of method comparison.

E3.5.2 Introduction of the application

Measurement of flow rate and total flow of fluids are the two commonest approaches found in flow measurement instrumentation. For this purpose, different methods and measurement principles are applied, with their own characteristics and specific requirements for installation and operation. These instruments are distinguishable at a first level considering the application to closed pipes with fluid under pressure or to open channels and free surface pipelines (see figure E3.5.1). In the case of flow measurements in closed pipes, the second level of classification is usually made considering the measurand output, namely, total flow (totalizers) or flow rate.

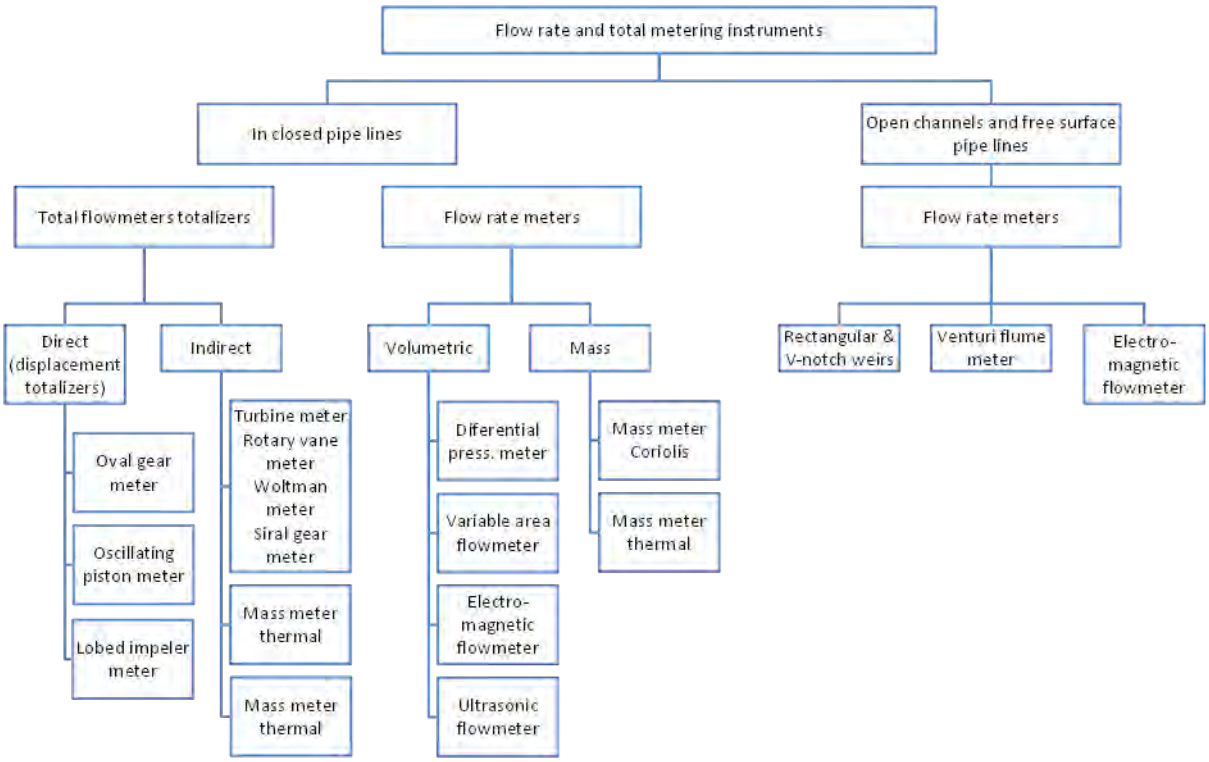


Figure E3.5.1: The different types of instruments for the measurement of total flow and flow rate used in closed pipes, open channels, and free surface pipes (redrawn from [242])

Total flowmeters, also called volume totalizers, are devices with two distinct approaches to the measurement:

1. direct volume totalizers, usually devices having a chamber with a defined volume and a system able to transfer the amount of fluid from the inlet to the outlet of the measuring instrument and being able to integrate the measured volumes to obtain the total flow volume; and
2. indirect volume totalizers, without chambers, employing mechanical or electromechanical systems usually using rotary vanes, being able to transport specific volumes and to count electrical pulses generated by the rotary angular displacement.

The selection of measuring instruments is made considering the quantity to be measured and the nature of the installation where measurement is intended to take place. In the context of water service providers, the volume of water is one of the most common, the accuracy requirements often being obtained using electromagnetic flowmeters and ultrasonic flowmeters. However, these two types of measuring instruments have significant differences of operation due to the

choice of measurands and measurement principles, with an impact on the sources of uncertainty that are required to be considered to make robust decisions. For this reason, a brief description of the quantities involved in both methods is given below.

E3.5.3 Specification of the measurand(s)

The measurands are the volumetric flow rate Q_V passing through a conduit at a given instant, the velocity v of the fluid at a specific instant and the volume V of fluid collected during a specified time interval.

E3.5.4 Measurement models

A measurement model in the case of electromagnetic flowmeters installed in pipes, the volumetric flow rate Q_V at a specific instant is obtained as the product of the cross-sectional area A of the pipe and the velocity v of the fluid:

$$Q_V = Av. \tag{E3.5.1}$$

Expression (E3.5.1) is regarded as a basic model describing the measurement principle that holds in ideal conditions. Effects arising in the practical implementation of the measurement are incorporated to obtain a measurement model adequate for the purpose [5]. Accordingly, an extended measurement model for the application is

$$Q_V = Av + \varepsilon_{Q_V} + \sum_{i=1}^{N_Q} \delta Q_{Vi}, \tag{E3.5.2}$$

where ε_{Q_V} is a calibration correction, and the δQ_{Vi} are N_Q other influence quantities that affect the measurement.

Ultrasonic flowmeters use the same expression (E3.5.1) but measure the average flow velocity of fluids. In this case, to calculate the flow rate, the diameter of the pipe at the measurement location is used. Like the approach used for electromagnetic flowmeters, an extended measurement model for the flow velocity v at a specific instant is

$$v = v_0 + \varepsilon_v + \sum_{i=1}^{N_v} \delta v_i, \tag{E3.5.3}$$

where v_0 is the uncorrected flow velocity at that instant, ε_v an error correction due to calibration, and the δv_i are N_v influence quantities that affect the measurement.

Electromagnetic flowmeters are widely disseminated as a common solution for measuring flow rate, using Faraday's Law of Induction, which relates the voltage induced across an electrical conductor to the velocity of a conductive fluid moving orthogonally to a magnetic field [243]. This principle is particularly applied to conductive fluids in motion in pipes.

The common technical approach (figure E3.5.2) applies a magnetic field B using magnetic coils and measures the output electrical voltage U_o using a pair of electrodes at a distance D (in the case of cylindrical pipes, equal to the diameter of a cross-section), in a plane orthogonal to the

magnetic field plane. The electrical output voltage U_o is proportional to the velocity of the fluid (in the axis direction). In these conditions, U_o is related to the input quantities by

$$U_o = BvD. \tag{E3.5.4}$$

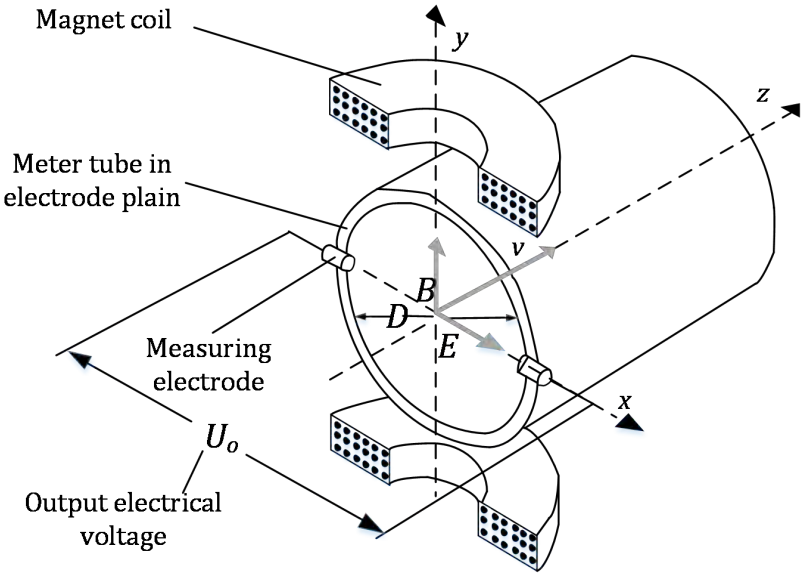


Figure E3.5.2: Electromagnetic flowmeter principle of operation and related quantities

The volumetric flow rate Q_V crossing a section of the pipe with area A is

$$Q_V = Av = \frac{\pi D^2}{4} v. \tag{E3.5.5}$$

Combining equations (E3.5.4) and (E3.5.5),

$$U_o = \frac{4BQ_V}{\pi D}, \tag{E3.5.6}$$

showing that, assuming good measurement and influence conditions, the induced output voltage has a linear proportional relation with the volumetric flow rate crossing the section of the pipe:

$$U_o \propto Q_V. \tag{E3.5.7}$$

The proportional relation (E3.5.7) can be expressed in terms of a constant calibration parameter K allowing the measurand Q_V to be obtained from the induced electrical voltage:

$$Q_v = \frac{\pi D}{4B} U_o = KU_o. \tag{E3.5.8}$$

The volume V of fluid collected during a time interval Δt_{total} is usually obtained as the sum of n measured volumetric flow rates Q_i at a set of consecutive fixed time intervals Δt_i:

$$\Delta t_{\text{total}} = \sum_{i=1}^{n} \Delta t_i, \tag{E3.5.9}$$

whence

$$V = \sum_{i=1}^{n} Q_{Vi} \Delta t_i. \tag{E3.5.10}$$

Ultrasonic flowmeters are also widely used to measure volumetric flow rate and volume of fluids. For this study, the type of ultrasonic flowmeter considered is that based on the transit time, with a single path and direct transmission of the signal. Other common systems make use of doppler methods [242].

The operation of the ultrasonic flowmeter is based on the principle that the sound velocity c in a fluid in motion results from the combination of the velocity of sound in the fluid at rest combined with the velocity of the fluid itself. This type of device measures the time taken for an ultrasonic pulse to travel in a pipe section both with and against the flow (see Figure 3) to calculate the flow velocity and, knowing the diameter of the pipe, to obtain the volumetric flow rate.

Regarding the principle of operation, it should be mentioned that c depends on the characteristics of the fluid, namely, the propagation velocity of a sound wave affected by conditions like temperature and pressure [244, 245].

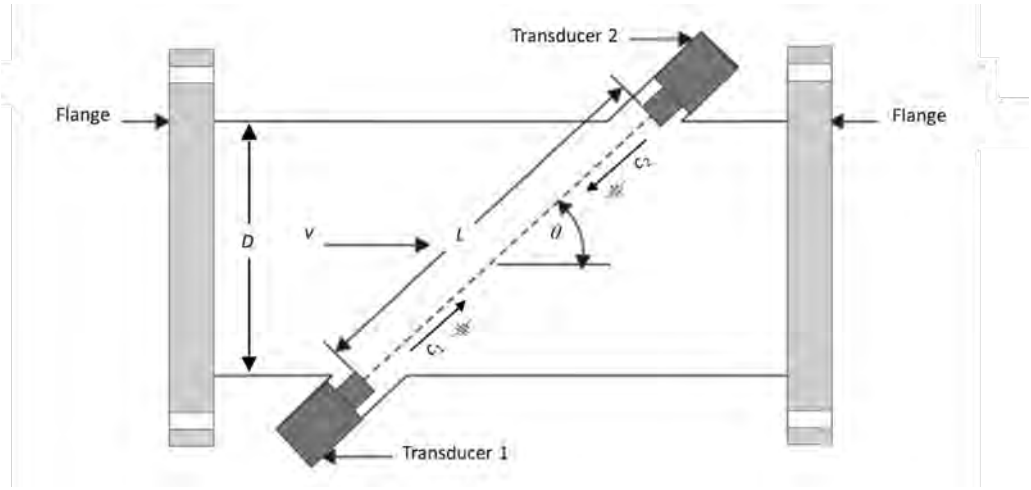


Figure E3.5.3: Transit time ultrasonic flowmeter (single path)

In figure E3.5.3, v is the average flow velocity, L is the acoustic path length in the fluid (travelled by the pulse between the two transducers), and θ the angle of the path to the pipe axis. The acoustic path length is given by

$$L = \frac{D}{\sin\theta}. \tag{E3.5.11}$$

Referring to figure E3.5.3 and letting c_1 denote the velocity of propagation of the pulse waves from transducer 1 to transducer 2, and c_2 that in the reverse direction, the combined effects of the sound velocity and the average flow velocity are

$$c_1 = c + v\cos\theta, \tag{E3.5.12}$$

$$c_2 = c - v\cos\theta. \tag{E3.5.13}$$

The corresponding times of transit are, respectively, expressed by

$$t_1 = \frac{L}{c + v\cos\theta}, \tag{E3.5.14}$$

$$t_2 = \frac{L}{c - v\cos\theta}. \tag{E3.5.15}$$

Combining these two equations, using $\Delta t = t_2 - t_1$, the average flow velocity becomes

$$v = \frac{L}{2\cos\theta}\left(\frac{1}{t_1} - \frac{1}{t_2}\right) = \frac{L}{2\cos\theta}\frac{\Delta t}{t_1 t_2}, \tag{E3.5.16}$$

allowing the volumetric flow rate Q crossing the section of the pipe with area A as presented in equation (E3.5.2) to be expressed in terms not involving v.

The description of the two methods above is useful for the identification of the sources of error that affect both instruments.

E3.5.5 Uncertainty propagation

E3.5.5.1 General

The generic measurement model $y = f(x_1, \ldots, x_n)$ is considered, where y represents the output quantity obtained from n input quantities x_i.

For an input quantity x_i for which a sample of observations is made, the sample average is taken as the estimate of the quantity, the standard deviation of the mean as the associated standard uncertainty $u(x_i)$ and the sample size minus one as the degrees of freedom ν_i. A t distribution would be assigned to the quantity for which the mean is an estimate. For a quantity for which non-statistical information is available, that information would be used to assign an appropriate distribution (for example, strict limits would yield a rectangular distribution).

To evaluate the measurement uncertainty $u(y)$ associated with an estimate of y, the law of propagation of uncertainty (LPU) [2] is used, assuming a linearization of the measurement model about the estimates of the input quantities gives sufficient validity to the results. Further assuming independence of the x_i, the application of LPU gives

$$u^2(y) = \sum_{i=1}^{n} c_i^2 u^2(x_i), \tag{E3.5.17}$$

the c_i being the sensitivity coefficients (first-order partial derivatives of f with respect to the input quantities evaluated at the x_i. The expanded measurement uncertainty for a confidence level of 95 % is

$$U_{95}(y) = ku(y). \tag{E3.5.18}$$

In turn, a 95 % confidence interval $y \pm U_{0.95}(y)$ is provided for the measurand. In equation (E3.5.18), k denotes a coverage factor [2] that converts a standard uncertainty to an expanded uncertainty. Under the assumption of normality for the measurand,

$$k = \mathrm{t}_{0.95}(\nu_{\mathrm{eff}}),$$

where $\mathrm{t}_p(\nu)$ denotes the value for the t distribution for $100p$ % confidence (two-sided) with ν degrees of freedom and the "effective degrees of freedom" ν_{eff} is provided by the Welch-Satterthwaite formula [2]:

$$\nu_{\mathrm{eff}} = \frac{u^4(y)}{\sum_{i=1}^{n} \frac{u^4(x_i)}{\nu_i}}. \tag{E3.5.19}$$

The evaluation of the standard measurement uncertainty $u(y)$ can be developed in terms of an uncertainty budget [42, 246]. A two-step approach to the information required for the uncertainty budget is (i) identify the sources of uncertainty and (ii) assign PDFs to the input quantities.

A typical approach for the identification of sources of uncertainty accounts for the contributions related to the phenomena that generate the conditions for the observation of the quantity, namely, the measurement method, the measuring instrument's metrological characteristics, the user influence (if applicable) and the data processing. Regarding the quantification of the contributions, Type A (statistical) and Type B (non-statistical) evaluations are considered as in the GUM [2].

E3.5.5.2 Electromagnetic flow meters

For the electromagnetic flowmeter, the volumetric flow rate Q_V is the measurand. Applying LPU, the standard measurement uncertainty $u(Q_V)$ associated with an estimate of Q_V [equation (E3.5.2)] is obtained from

$$u^2(Q_V) = u^2(Av) + u^2(\varepsilon_Q) + \sum_{i=1}^{N_Q} u^2(\delta Q_i). \tag{E3.5.20}$$

A possible set of uncertainty sources [247–251] is given in table E3.5.1.

An instance of an uncertainty budget for a specific measurement location, considering only non-negligible uncertainty contributions identified in table E3.5.1 is given in table E3.5.2. The contributions $|c_i|\,u(x_i)$ are those for a linearized version of the measurement model $y = f(x_1, \ldots, x_n)$. In this table, the degrees of freedom used in many of the Type B evaluations considered is 50, which is obtained using GUM clause G.4.2, being the relative uncertainty in $u(x_i)$, in square brackets below, considered reliable to 90 %:

$$v_i = \frac{1}{2}\left[\frac{\Delta u(x_i)}{u(x_i)}\right]^{-2}. \tag{E3.5.21}$$

E3.5.5.3 Ultrasonic flow meters

Applying the LPU [2] to the ultrasonic flowmeter model (E3.5.3), again assuming independent input quantities, the standard uncertainty $u(v)$ is given by

$$u^2(v) = u^2(v_0) + u^2(\varepsilon_v) + \sum_{i=1}^{N_v} u^2(\delta v_i). \tag{E3.5.22}$$

Again, the evaluation of the measurement uncertainty can be facilitated by an uncertainty budget table, using the above two-step approach. A set of uncertainty sources [244, 252–254] is given in Table 3. Having evaluated the flow velocity measurement uncertainty $u(v)$ in a similar way to that given in table E3.5.2, the next step is to evaluate the standard uncertainty $u(Q_V)$ associated with the flow rate Q_V.

Applying the LPU to the model (E3.5.5),

$$u^2(Q_V) = v^2u^2(A) + A^2u^2(v). \tag{E3.5.23}$$

Table E3.5.1: Identification of sources of uncertainty and data sources related to flow rate measurements using electromagnetic flowmeters

Uncertainty contribution	Data source
Calibration measurement	Calibration certificate
Calibration curve	Calibration correction model
Correction residual errors	Calibration correction model
Repeatability	Calibration certificate
Flow regime (laminar, turbulent, other)	Assembly and installation
Equipotential of pipes	Assembly and installation
Flow velocity range	Assembly and installation
Power supply influence	Assembly and installation
Straight pipe levelling	Assembly and installation
Grounding and shielding	Assembly and installation
Wiring, cables, and connectors	Assembly and installation
Resolution	Equipment metrological characteristic
Linearity	Equipment metrological characteristic
Pipe diameter	Equipment metrological characteristic
Long term stability	Equipment metrological characteristic
Reproducibility	Equipment metrological characteristic
Symmetry of the magnetic excitation coil	Equipment metrological characteristic
Ageing	Equipment metrological characteristic
Vibration	Flow conditions
Electrode coating (deposits)	Flow conditions
Fouling (crystallization, deposition, chemical reactions, corrosion, biofilms)	Flow conditions
Air pockets and bubbles	Flow conditions
Fluid compressibility and pulsation	Flow conditions
Density of the fluid	Liquid properties
Conductivity	Liquid properties
Viscosity	Liquid properties
Current interference	Performance with influence quantities
Electromagnetic interference	Performance with influence quantities
Temperature	Performance with influence quantities
Pressure	Performance with influence quantities
Sampling rate	Signal and data processing
Time resolution	Signal and data processing
ADC resolution	Signal and data processing
Zero offset	Signal and data processing
Signal gain	Signal and data processing
Signal filtering	Signal and data processing
Signal noise	Signal and data processing
Statistical calculations	Signal and data processing

Table E3.5.2: Example of an uncertainty budget related to the measurement of flow rate using an electromagnetic flowmeter (DoF = degrees of freedom; N = normal, R = rectangular, T = triangular distribution)

Quantity	Parameters or limits x_i/%	PDF	Sensitivity coefficient c_i	Uncertainty contribution $u(x_i)$/%	DoF ν_i
Calibration measurement	0.05	R	1	0.029	200
Calibration curve	0.01	N	1	0.01	6
Correction residual errors	0.01	N	1	0.01	6
Repeatability	0.01	N	1	0.01	4
Flow regime (e.g., laminar, turbulent)	0.005	R	1	0.0029	50
Straight pipe levelling	0.002	R	1	0.0012	50
Grounding and shielding	0.002	R	1	0.0012	50
Resolution	0.05	R	1	0.029	50
Linearity	0.002	T	1	0.00082	50
Long term stability	0.002	R	1	0.0012	50
Vibration	0.002	R	1	0.0012	50
Electrodes coating (deposits)	0.005	R	1	0.0029	50
Fouling (crystallization, deposition, chemical reactions, corrosion, biofilms)	0.01	R	1	0.0058	50
Air pockets and bubbles	0.005	R	1	0.0029	50
Density of the fluid	0.01	R	1	0.0058	50
Conductivity	0.01	R	1	0.0058	50
Viscosity	0.005	R	1	0.0029	50
Temperature	0.01	R	1	0.0058	50
Pressure	0.001	R	1	0.00058	50
ADC resolution	0.05	R	1	0.029	50
Signal gain	0.01	R	1	0.0058	50
Signal filtering	0.01	R	1	0.0058	50
Statistical calculations	0.02	R	1	0.012	50
			Standard-uncertainty $u(Q_v)$/%:	0.056	
			Effective degrees of freedom ν_{eff}:	272	
			Coverage factor k:	2.01	
			Relative expanded uncertainty $U_{0.95}(Q_v)$/%:	0.11	

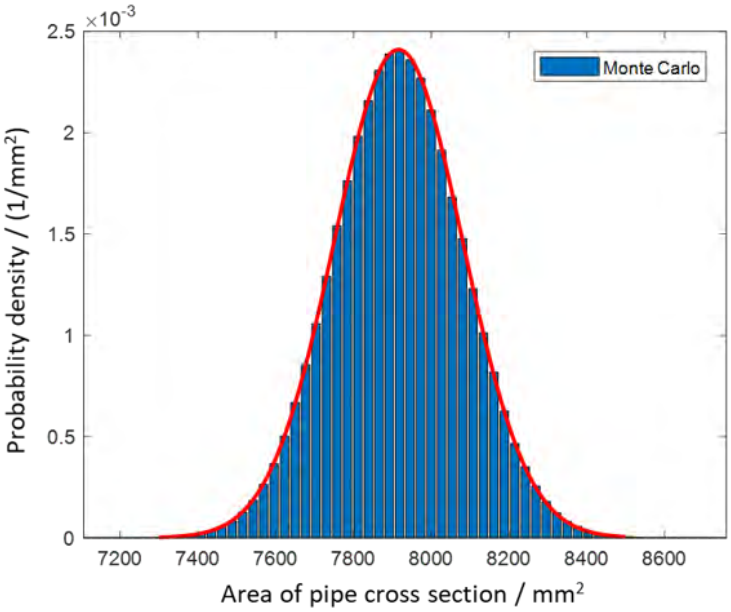


Figure E3.5.4: PDF for the area of a cross-section of a pipe

A typical experimental approach to estimate A is to estimate the internal diameter D, which is itself an indirect measurement involving the external perimeter $p_{\varnothing}$ of the pipe (used to calculate the external diameter D_0) and the pipe thickness l_t. Thus,

$$A = \frac{\pi}{4}D^2 = \frac{\pi}{4}(D_o - 2l_t)^2 = \frac{\pi}{4}\left(\frac{p_{\varnothing}}{\pi} - 2l_t\right)^2 = \frac{1}{4\pi}\left(p_{\varnothing} - 2l_t\pi\right)^2. \tag{E3.5.24}$$

The evaluation of the standard uncertainty $u(A)$ associated with A was made, in a first approach, using the LPU, taking experimental results for a flowmeter installed in a water facility pipe of nominal diameter 100 mm. Data used was, for the external perimeter $p_{\varnothing}$, 439.8 mm, with a standard uncertainty of 1.0 mm, and an average thickness I_t of the pipe of 19.8 mm with a standard uncertainty of 0.5 mm. A normal PDF was assumed for both. The estimated value of A was 7.916×10^3 mm^2 (leading to an estimate of D of 100.4 mm), with an expanded uncertainty (95 %) of 0.330×10^3 mm^2.

Usually for this type of analysis and despite some non-linearity in the model, the use of LPU suffices to provide acceptable results. In this case, however, the imbalance of the terms in equation (E3.5.24) was considered a reason to apply a MCM [3] to provide a more robust uncertainty evaluation. The evaluation used RStudio [49] programming to generate 1×10^6 samples from the PDFs for A and v and, for each sample of this pair of quantities, Q_V was evaluated from equation (E3.5.5). By ordering the values of Q_V so determined, the 2.5 % and the 97.5 % quantiles were used as the endpoints of a 95 % coverage interval for Q_V.

The MCM-estimated value of A was 7.916×10^3 mm^2, identical, to the number of digits quoted, to that provided by the LPU (leading to an estimate of D of 100.4 mm), with a 95 % coverage interval of $[7.595, 8.243] \times 10^3$ mm^2. An expanded uncertainty for 95 % confidence of 0.324×10^3 mm^2 can be taken as the half-width of this confidence interval. In comparison, the LPU approach gives an expanded uncertainty for 95 % confidence of 0.330×10^3 mm^2, almost identical to the MC value. Figure E3.5.4 presents a scaled histogram approximating the PDF for A obtained by the MCM and the normal PDF obtained using LPU. The process to obtain the volumetric flow rate is given in figure E3.5.5.

Table E3.5.3: Identification of common sources of uncertainty and data sources that can be related to flow rate measurements using transit time ultrasonic flow meters

Uncertainty contribution	Data source
Calibration measurement uncertainty	Calibration certificate
Calibration curve	Calibration correction model
Correction residual errors	Calibration correction model
Repeatability	Calibration certificate
Pipe diameter accuracy	Measurement certificate
Flow regime (laminar, turbulent, other)	Assembly and installation
Flow velocity range	Assembly and installation
Distance between the electrodes	Assembly and installation
Power supply influence	Assembly and installation
Straight pipe levelling	Assembly and installation
Grounding and shielding	Assembly and installation
Wiring, cables, and connectors	Assembly and installation
Resolution	Equipment metrological characteristic
Linearity	Equipment metrological characteristic
Velocity profile	Equipment metrological characteristic
Pipe diameter	Equipment metrological characteristic
Long term stability	Equipment metrological characteristic
Reproducibility	Equipment metrological characteristic
Aging	Equipment metrological characteristic
Vibration	Flow conditions
Equipotential of pipes	Assembly and installation
Fouling (crystallization, deposition, chemical reactions, corrosion, biofilms)	Flow conditions
Air pockets and bubbles	Flow conditions
Fluid compressibility and pulsation	Flow conditions
Density of the fluid	Liquid properties
Conductivity	Liquid properties
Viscosity	Liquid properties
Current interference	Performance with influence quantities
Temperature	Performance with influence quantities
Pressure	Performance with influence quantities
Sampling rate	Signal and data processing
Time resolution	Signal and data processing
ADC resolution	Signal and data processing
Integration of time of transit	Signal and data processing
Remote access to signals and cables	Signal and data processing
Zero offset	Signal and data processing
Signal gain	Signal and data processing
Signal filtering	Signal and data processing
Signal noise	Signal and data processing
Statistical calculations	Signal and data processing

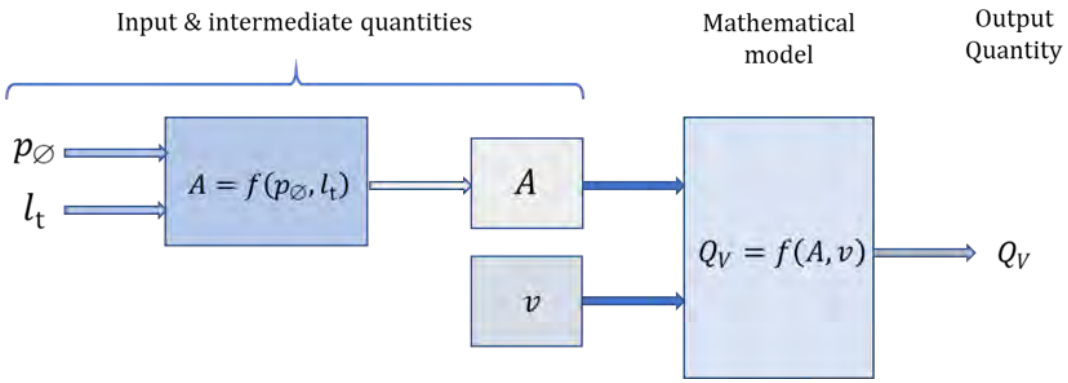


Figure E3.5.5: Process for the measurement of volumetric flow rate using ultrasonic flowmeters

The Monte Carlo method has validated the use of LPU in this case. However, considering the larger magnitude of the term involving the cross sectional area, if the PDF for A obtained by MC were different, for example, rectangular, a different output PDF and related parameters would be obtained, with the result that the difference between the LPU and MC results would be likely to be greater.

E3.5.6 Interpretation of results

The flow engineer faces the difficult task of selecting the technically best and most cost effective measuring device for his application. For this purpose, account should be taken of the required accuracy related to the management process including the target uncertainty, the uncertainty of the measuring instruments, the characteristics of the fluid and the conditions of the flow, the installation set up, the metrological management needs, and the data management. In this analysis, a comparison of the uncertainty achieved by each method can play a major role in the decision making.

The identification of the sources of uncertainty and its impact by applying a sensitivity analysis can be very important to control the impact of errors that, in this field, can be of high magnitude, to develop best practices at the operational level and to achieve the growing demands of accuracy from clients and regulatory and normative requirements.

The uncertainty related to the area of the cross-section of pipes is a major difference between both the use of electromagnetic flowmeters and ultrasonic flowmeters. In the former case, its accuracy is highly controlled by the manufacturing process. In the latter, the accuracy can be highly variable because of the need to perform measurements locally, with different pipe materials and coatings (sometimes painted), using different instrumentation to obtain the length of the perimeter and thickness through sampling, promulgating an increase in uncertainty.

Finally, the growing concern with climate change leading to the need for improvement of water resources management makes the quality of measurement of water service providers a key issue, uncertainty being a major parameter that allows comparisons to be made and to show the competence of measurement activities.

Example E3.6

Greenhouse gas emission inventories

M.G. Cox, T. Gardiner, R. Robinson, T. Smith, S.L.R. Ellison, A.M.H. van der Veen

E3.6.1 Summary

The compilation of country-scale greenhouse gas emissions inventories involves combining standard emissions' factors and activity data to provide sector emissions' estimates. This work constitutes a small study of this huge, impactful topic, concentrating on agriculture, where large uncertainties arise, and the effect of correlation on the results. It also gives a brief comparison of the attitudes taken to uncertainty evaluation by the Intergovernmental Panel on Climate Change (IPCC) and the Joint Committee on Guides in Metrology (JCGM).

E3.6.2 Introduction of the application

Greenhouse gas (GHG) emissions inventories are usually obtained by combining standard emissions' factors and activity data to provide sub-sector emissions' estimates. These sub-sectors are then combined to give sector estimates and further combined to estimate total emissions.

A specific example is given relating to part of the inventory in the agriculture, land use and waste sector to provide an estimate of total emissions and the associated uncertainty.

Current practice provided by the Intergovernmental Panel on Climate Change (IPCC) regarding uncertainty propagation in the area is considered in the context of IPCC Guidelines for National Greenhouse Gas Inventories [255]. The IPCC practice is contrasted with that in the Guide to the expression of Uncertainty in Measurement (GUM) [2] and other related JCGM guides. In particular, the possible effects of correlation, often ignored in practice, on the estimate are considered.

E3.6.3 Background

E3.6.3.1 Greenhouse gas inventories

Calculation of GHG inventories

The UK's and other countries' GHG inventories are compiled according to IPCC 2006 Guidelines [255]. Each year the inventory is updated to include the latest data available. The UK submits a report to the UNFCCC annually via a consolidated report which contains all EU countries.

The inventory includes the seven direct GHGs under the Kyoto Protocol:

- Carbon dioxide (CO_2)
- Methane (CH_4)
- Nitrous oxide (N_2O)
- Hydrofluorocarbons (HFCs)
- Perfluorocarbons (PFCs)
- Sulfur hexafluoride (SF_6)
- Nitrogen trifluoride (NF_3)

A bottom-up calculation of emissions is based on contributions of the form

$$\text{Emissions} = (\text{Activity data}) \times (\text{Emission factor}). \tag{E3.6.1}$$

In all there are 743 emission factors and more than 1700 sources of emissions. There are 17 key data sources together with involvement from industry, government and academics.

In terms of emissions from the three largest GHG contributors (i.e., CO_2, CH_4 and N_2O), the majority of the uncertainty in the GHG inventory comes from the Agriculture, Land Use and Waste sectors [256], with these three sectors contributing 86 % of the uncertainty in the total inventory emissions.

European countries, especially the UK, are moving towards metrological assessment of current uncertainty quantification including accounting for the effect of correlated quantities. They also want to extend the current top-down validation activity E3.7 to cover the complete GHG inventory.

Figure E3.6.1 shows the breakdown of the inventory contributions across sectors. The fifth biggest contribution is agriculture. As reductions in other contributions are made, somewhat more readily, agriculture will in future make a larger contribution.

An emission inventory can be expressed as a linear combination of terms of the form (E3.6.1):

$$E = \sum_i F_i A_i, \tag{E3.6.2}$$

where

E total GHG emission for a given sector, geographic area and time period,
F_i emissions factor for the emissions of a given pollutant from source category i, and
A_i activity for source category i.

In practice, the model (E3.6.2) is often enlarged to incorporate scaling factors to allow total GHG emission to be given in terms of CO_2e or 'CO_2 equivalent'. That is, each GHG in the table has a conversion or scaling factor associated with it, giving

$$E = \sum_i s_i F_i A_i, \tag{E3.6.3}$$

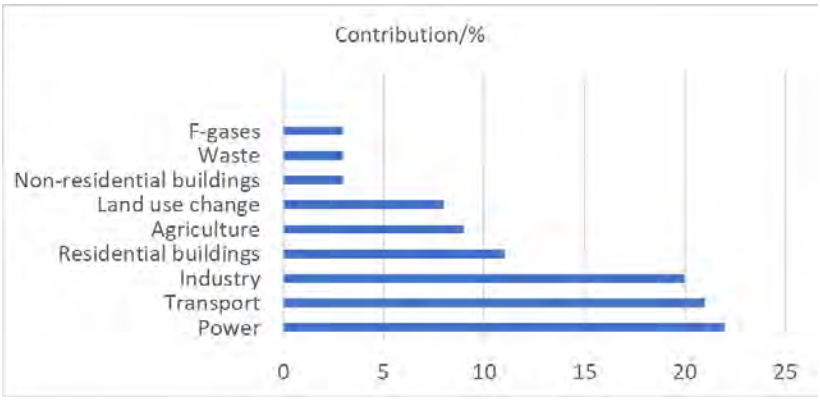


Figure E3.6.1: Inventory contributions across sectors

An example incorporating such scaling factors is given in section E3.6.6.

Combination and correlation

Inventory uncertainties are generally combined in quadrature (root-sum-squares), assuming no correlation between the input quantities involved, using the variant of the law of propagation of uncertainty (LPU) applicable to such cases [2, clause 5.1.2, formula (10)]. The implication is that the influence of larger uncertainties is magnified, more so than when cases of full or partial correlation, which would apply to the variant of LPU that applies in that case [2, clause 5.2.2, formula (13)], are considered. Ways of assessing the validity of this assumption both within and across sectors are required.

Figure E3.6.2 shows the expanded uncertainty contributions across sectors. The biggest contribution is agriculture. Figure E3.6.3 shows the squares of these expanded uncertainty contributions. It emphasizes the greater influence of the dominant uncertainty contributions to the combined uncertainty.

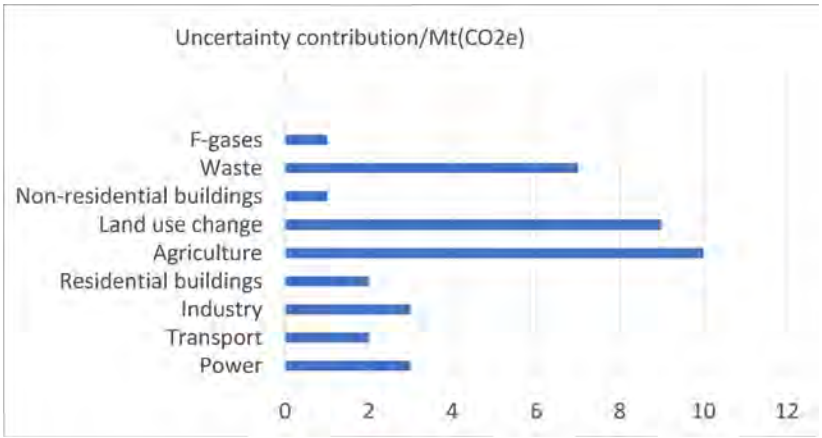


Figure E3.6.2: Expanded uncertainty contributions (at the 95 % level) across sectors

According to [257], emissions inventories are critical to many environmental decision-making processes. Typical questions that decision makers may ask that motivate the need to deal with uncertainties in emissions' inventories relate to precision and bias in the estimates, whether the estimates are based upon measurements, modelling or expert judgment, the main sources of

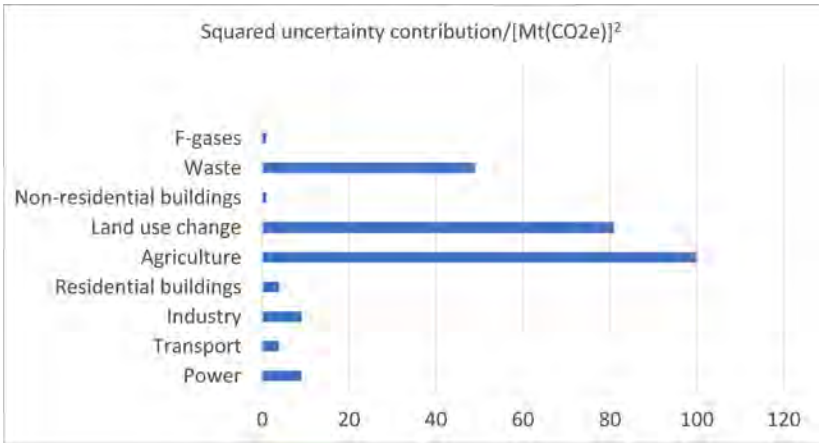


Figure E3.6.3: Squared expanded uncertainty contributions across sectors

uncertainty in these estimates, and how uncertainty can be reduced. To answer such questions, a reasonable indication of the uncertainty associated with the estimates of the quantities F_i, A_i and E in expression (E3.6.3) that comprise an emission inventory is needed.

E3.6.3.2 IPCC

The United Kingdom in 2018 submitted its National Inventory Report [258] to the United Nations' Framework Convention on Climate Change (UNFCCC). It contains national greenhouse gas emission estimates for the period 1990–2016, and descriptions of the methods used to produce the estimates. The report was prepared in accordance with "UNFCCC reporting guidelines on annual greenhouse gas inventories" (decision 24/CP.19) and includes elements required under the Kyoto Protocol.

The reporting guidelines refer to the 'error propagation approach' and the 'Monte Carlo approach' for uncertainty propagation. An extract from [258], which uses these approaches, is:

> 'Comparing the results of the error propagation approach, and the Monte Carlo estimation of uncertainty by simulation, is a useful quality control check on the behaviour of the Monte Carlo model.
>
> 'The reason that the error propagation approach is used as a reference is because the approach to the error propagation approach has been defined and checked by the IPCC, and is clearly set out in the IPCC 2000 Good Practice Guidance and the 2006 Guidelines. The UK has implemented the IPCC error propagation approach as set out in this guidance. The implementation of an uncertainty estimation by simulation cannot be prescriptive, and will depend on how the country constructs its model, and the correlations included within that model. Therefore, there is a greater likelihood of errors being introduced in the model used to estimate uncertainty by Monte Carlo simulation.'

We return to this extract in section E3.6.8.

E3.6.3.3 JCGM

JCGM 100:2008, the GUM [2], has for a long time been the authoritative document concerned with the evaluation and expression of measurement uncertainty that attempts to meet this objective:

> 'This Guide establishes general rules for evaluating and expressing uncertainty in measurement that can be followed at various levels of accuracy and in many fields — from the shop floor to fundamental research. Therefore, the principles of this Guide are intended to be applicable to a broad spectrum of measurements ...'

The JCGM is responsible for maintaining and promoting the use of the GUM and the *International vocabulary of basic and general terms in metrology* (VIM) [89]. Starting with the 2008 publication of GUM Supplement 1 (GUM-S1) about the Monte Carlo method (MCM) for the propagation of probability distributions [3], several documents [1, 4–6] have been made available by the JCGM that describe methods for evaluating measurement uncertainty for various classes of problems and the manner in which the measurement model, employed as the basis for the evaluation, is established and used.

All eight JCGM member (international) organizations and many national metrology institutes (including the major ones) review and approve JCGM documents before publication.

E3.6.3.4 This document

In this document we consider the IPCC approach to uncertainty evaluation in the context of greenhouse gas emissions. We make statements concerning the extent of alignment with guidance provided by the JCGM.

Wherever measurement uncertainty is used, it is important that it is evaluated consistently by different practitioners so that users' interpretation is consistent and reliable. Universally agreed methods for uncertainty evaluation are hence required.

E3.6.4 Specification of the measurand(s)

'Measurand' is a term little used in IPCC documents. A meaningful interpretation, used here, is 'total emissions for a given pollutant, geographic area and time period' (see section E3.6.2).

E3.6.5 Measurement model

The IPCC does not seem explicitly to use the concept of 'measurement model' in their documents. However, expression (E3.6.3) can be regarded as a basic measurement model central to IPCC considerations.

E3.6.6 Uncertainty propagation

We consider the use of the model specified as expression (E3.6.3), in which we take the sector as mobile machinery in agriculture (IPCC category A4ci), the region as the UK and the time period as 2018.

Uncertainties are propagated using the law of propagation of uncertainty in the GUM [2], comparable to the IPCC 'error propagation' method.

First, the quantities concerned are considered as independent. Then, perceived correlation is taken into account.

Data for this sector from the UK National Atmospheric Emissions Inventory site (2018) is given as table E3.6.1 where the number of significant decimal digits is stated as reported. This total comprises contributions from the use of two fuels, 'Gas Oil' and Petrol, reported in terms of the 2018 Activity Data (AD) for those fuels. The table also specifies the Emissions' Factors (EF) for each of the key GHGs (CO_2, CH_4 and N_2O) for those fuels. The reported expanded uncertainties (at the 95 % confidence level) on the individual activity data and emission factors are given.

This table incorporates the scaling mentioned in section E3.6.3.1. The scaling factors to give 'CO_2 equivalent' equivalent emissions are $11/3 = 3.667$ for CO_2 as C, 25 for methane (CH_4) and 298 for nitrous oxide (N_2O).

Table E3.6.1: Reported activity data and emission factors with associated expanded uncertainties U (for 95 % confidence)

Fuel	Gas	AD/(TJ)	U(AD)/%	EF/(kt TJ^{-1})	U(EF)/%
Gas oil	CO_2	0.020 438	38.6	2.0438×10^{-2}	2.7
Gas oil	CH_4	0.020 438	1.6	3.5368×10^{-6}	80.0
Gas oil	N_2O	0.020 438	1.6	3.0984×10^{-6}	216.3
Petrol	CO_2	67.19	50.7	1.9127×10^{-2}	4.8654×10^{-5}
Petrol	CH_4	67.19	1.6	4.8654×10^{-5}	80.0
Petrol	N_2O	67.19	1.6	3.3578×10^{-7}	216.3

Let $A_1, \ldots, A_6$ denote the activity data values AD and $F_1, \ldots, F_6$ the corresponding emissions' factors EF in table E3.6.1 in the order given there. Then, comparing with expression (E3.6.3), we write

$$E = \sum_{i=1}^{6} E_i, \qquad E_i = s_i A_i F_i, \tag{E3.6.4}$$

where the s_i are the scaling factors, assumed to have no associated uncertainty.

Total emissions for this sector were reported to be 4282 kt CO_2e. The reported expanded uncertainty (at the 95 % level of confidence) associated with this total was 1635 kt CO_2e (relative expanded uncertainty of 38 %).

Table E3.6.2 summarizes the emissions' calculations for this sub-sector, namely, the emissions as CO_2 equivalent for each contribution and the total CO_2 equivalent, obtained from the UK National Atmospheric Emissions Inventory site (2018).

E3.6.6.1 Assuming independence

The application of the law of propagation of uncertainty (LPU) [2, eqn. (10)] for independent quantities gives the associated standard uncertainty $u_{\text{ind}}(E)$:

$$u_{\text{ind}}^2(E) = \sum_{i=1}^{6} s_i^2 \left[A_i^2 u^2(F_i) + F_i^2 u^2(A_i) \right]. \tag{E3.6.5}$$

Table E3.6.2: Emissions as CO_2 equivalent for each contribution and the total CO_2 equivalent

GHG	Fuel	AD/ TJ	EF/ (kt/TJ)	Reported emissions/kt	Scaling factor	CO_2e/ kt
CO_2 as C	Gas oil	56 317.0920	2.0438×10^{-2}	1150.9806	3.667	4220.2623
CH_4	Gas oil	56 317.0920	3.5368×10^{-6}	0.1992	25	4.9795
N_2O	Gas oil	56 317.0920	3.0984×10^{-6}	0.1745	298	51.9987
CO_2 as C	Petrol	67.1856	1.9127×10^{-2}	1.2851	3.667	4.7120
CH_4	Petrol	67.1856	4.8654×10^{-5}	0.0033	25	0.0817
N_2O	Petrol	67.1856	3.3578×10^{-7}	0.0000	298	0.0067
Total						4282.0409

The use of the data in tables E3.6.1 and E3.6.2 gives E, as already reported, and by applying the LPU [2, formula (10)] for independent quantities:

$$E = 4282.04\,\text{kt}, \qquad u_{\text{ind}}(E) = 835.16\,\text{kt}. \tag{E3.6.6}$$

An uncertainty budget is given in table E3.6.3.

Table E3.6.3: Uncertainty budget

GHG	Fuel	AD/TJ	EF/(kt/TJ)	CO_2e std. unc./kt
CO_2 as C	Gas oil	56 317.0920	2.0438×10^{-2}	833.1844
CH_4	Gas oil	56 317.0920	3.5368×10^{-6}	2.0329
N_2O	Gas oil	56 317.0920	3.0984×10^{-6}	57.3860
CO_2 as C	Petrol	67.1856	1.9127×10^{-2}	1.2198
CH_4	Petrol	67.1856	4.8654×10^{-5}	0.0334
N_2O	Petrol	67.1856	3.3578×10^{-7}	0.0074
			Total CO_2e std. unc./kt =	835.16

E3.6.6.2 Accounting for perceived correlation

We now consider perceived correlations associated with the various quantities, which seems not to be strongly considered in the IPCC documents. It appears that correlation is not entertained with the 'error propagation' approach (broadly equivalent to LPU in the GUM), although it is stated, without giving detail, that the Monte Carlo method may be used for this purpose [259, section 6.3].

Examining the activity contributions A_i, we see that A_1, A_2 and A_3 are identical numerically as are A_4, A_5 and A_6. This observation raises the possibility that A_1, A_2 and A_3 have a common origin and might even be the same quantity, and similarly for A_4, A_5 and A_6.

This perception is considered by the authors to be more than reasonable. Moreover, there may even be a more subtle correlation between the emission factors if similar methods were used to derived these for the different fuels, but that consideration goes beyond the scope of this study. Assuming this perception to be correct, some adjustment to the above uncertainty calculation is needed. When several input quantities are associated via a common origin or effect, the resulting correlation can typically be handled either by constructing and using an appropriate covariance

matrix, or by re-parametrizing to isolate the common effect. In this case, if the (numerically identical) activities A_1 to A_3 and A_4 to A_6 are either very highly correlated or are simply repeated use of the same information, it is simplest to re-parametrize by replacing the nominally separate instances by a single term.

Following this approach implies that in expression (E3.6.4), A_1, A_2 and A_3 can be replaced by A_{GO} (GO = Gas Oil), say, and A_4, A_5 and A_6 by A_{P} (P = Petrol), say. As a result, this expression becomes

$$E = A_{\mathrm{GO}}F_{\mathrm{GO}} + A_{\mathrm{P}}F_{\mathrm{P}}, \tag{E3.6.7}$$

where

$$F_{\mathrm{GO}} = \sum_{i=1}^{3} s_i F_i, \qquad F_{\mathrm{P}} = \sum_{i=4}^{6} s_i F_i,$$

yielding $u_{\mathrm{cor}}(E)$, signifying the standard uncertainty associated with E obtained by taking perceived correlations into consideration:

$$u_{\mathrm{cor}}^2(E) = \underbrace{A_{\mathrm{GO}}^2 u^2(F_{\mathrm{GO}}) + F_{\mathrm{GO}}^2 u^2(A_{\mathrm{GO}})}_{\text{Term A}} + \underbrace{A_{\mathrm{P}}^2 u^2(F_{\mathrm{P}}) + F_{\mathrm{P}}^2 u^2(A_{\mathrm{P}})}_{\text{Term B}}, \tag{E3.6.8}$$

where

$$u^2(F_{\mathrm{GO}}) = \sum_{i=1}^{3} s_i^2 u^2(F_i), \qquad u^2(A_{\mathrm{P}}) = \sum_{i=4}^{6} s_i^2 u^2(F_i).$$

For the current data, E is identical to the previously calculated value and

$$u_{\mathrm{cor}}(E) = 846.33\,\mathrm{kt}, \tag{E3.6.9}$$

an increase of 1.3 % over the standard uncertainty assuming independence.

The reason for the increase in the standard uncertainty associated with total emissions being so small is that one particular contribution to the budget, CO_2 as C for Gas oil, the first in table E3.6.2, is by far the dominant.

In terms of the results, although the 1.3 % difference in the standard uncertainties is small, it is considered more than negligible. Also note that if a simple scaling factor of 2 were used between standard and expanded uncertainties, as would be the case if normality were assumed for the quantities involved, then the reported expanded uncertainty of 1635 kt would indicate a standard uncertainty of 818 kt, which is smaller than that of either of our correlated or uncorrelated results.

The principle stands that such correlations should generally be taken into account since they might make a significant difference where one uncertainty contribution is less dominant. It would be good to identify an instance where the contribution of the different sub-sectors is more even and the effect of correlation therefore more pronounced. Such activity will take place following the conclusion of this project EMUE (see the last paragraph of section E3.6.8.1).

E3.6.6.3 Insight into the degree of dependence

It would be possible to handle the uncertainty propagation using [2, formula (13)] taking full correlation into account for the quantities concerned, which would yield the same expression (E3.6.8). The use of variable substitution as above reparametrizes the problem in such a way that the need to take correlation into account is avoided. The assumption is made that the remaining quantities are independent, which should be verified in practice to the extent possible.

Insight into the difference in the standard uncertainty in the measurand under the assumptions of independence and dependence (in the above respect) is given by taking an extreme case.

Consider the following idealized scenario. For all i, take $A_i = A$, $F_i = F$, $u(A_i) = u_A$, $u(F_i) = u_F$ and $s_i = s$. Then, expression (E3.6.5) becomes

$$u^2_{\text{ind}}(E) = 6s^2A^2u_F^2 + 6F^2s^2u_A^2 = s^2A^2F^2\left[6u^2_{\text{rel}}(F) + 6u^2_{\text{rel}}(A)\right] \tag{E3.6.10}$$

with relative standard uncertainties

$$u_{\text{rel}}(F) = \frac{u_F}{F}, \qquad u_{\text{rel}}(A) = \frac{u_F}{A}.$$

On the other hand, from expression (E3.6.8),

$$u^2_{\text{cor}}(E) = 6s^2A^2u_F^2 + 18F^2s^2u_A^2 = s^2A^2F^2\left[6u^2_{\text{rel}}(F) + 18u^2_{\text{rel}}(A)\right]. \tag{E3.6.11}$$

The ratio of the terms in square brackets in expressions (E3.6.10) and (E3.6.11), namely,

$$\lambda = \frac{u^2_{\text{cor}}(E)}{u^2_{\text{ind}}(E)} = \frac{6u^2_{\text{rel}}(F) + 18u^2_{\text{rel}}(A)}{6u^2_{\text{rel}}(F) + 6u^2_{\text{rel}}(A)} = \frac{u^2_{\text{rel}}(F) + 3u^2_{\text{rel}}(A)}{u^2_{\text{rel}}(F) + u^2_{\text{rel}}(A)}, \tag{E3.6.12}$$

is informative regarding the respective contributions from the emissions factor and the activity data under the above correlation and independence assumptions. Examining expression (E3.6.12), for an activity data relative uncertainty that is small compared with that for the emissions factor, it is seen that λ is close to unity and in that case the assumption of independence is reasonable. For the converse, λ is close to 3 with the consequence that accounting for correlation, $u_{\text{cor}}(E)$ is $\sqrt{3} \approx 1.7$ times $u_{\text{ind}}(E)$, a 70 % increase.

E3.6.7 Reporting the result

The reporting of results has been integrated into section E3.6.6 dealing with uncertainty propagation.

E3.6.8 Interpretation of results

E3.6.8.1 Sources having large uncertainty

Relative expanded uncertainties of 100 % or more are not uncommon in emissions' inventories, which is why it is important these large uncertainty sources are carefully taken into consideration. The UK introduction to GHG inventories [260], shows a highly asymmetric interval for inventory NO_2 figures, in that emissions' estimates for N_2O are far from the centre of their 95 % confidence intervals, in its figure 5.2. The implication is that the probability distributions for N_2O are strongly asymmetric, but in the previous section (5.1) of that document uncertainties are expressed as 95 % confidence intervals (for example, table 5.1), with no indication that they may be asymmetric.

Some areas within emissions' inventories are so uncertain that it is not known with confidence whether they are sources or sinks. However, the referenced figure implies that N_2O from agriculture is definitely a source so the probability distribution must be asymmetric. However, the report does not provide an explanation for the figure.

After some investigation, these aspects seem to be based on methodology described in Milne et al. [261]. Figure 2 in that paper shows the distribution given by Monte Carlo modelling and the asymmetry is attributed to skew in the emissions factors. Log-normal probability distributions are used to represent knowledge of such quantities.

However, reference [261] is focused on agricultural emissions (that is, manure, enteric fermentation and field burning), but table E3.6.1 in this document is based on non-road vehicle emissions (that is, from tractors, combine harvesters, etc.), which burn gas-oil and diesel. No evidence has been found by the authors of this document on the probability distributions for the quantities under consideration here.

In this work, we focused on an agricultural sub-sector where there was sufficient data publicly available to carry out this type of assessment. We are confident that the correlation issue is more significant within other agricultural sub-sectors but the raw data needed to test this seemed not to be available. We understand that such data and more extensive information concerning the numbers appearing in the UK inventory, from which correlations can be quantified, will subsequently be made available, but not before the completion of the EMUE project.

E3.6.8.2 Uncertainty guidance promoted by the IPCC and the JCGM

Aspects of the methods used by IPCC and in the UK and some comments on them where appropriate are as follows:

1. In terms of regarding expression (E3.6.3) as the measurement model, the GUM [2] refers to the quantities F_i and A_i as 'input quantities' and E as the 'output quantity' or 'measurand'.

2. IPCC documents seem not to emphasize that propagation of uncertainty should be carried out in the context of a *measurement model*. In the model (E3.6.3), the expression on the right-hand side constitutes the *measurement function*, with sensitivity coefficients given by the first-order partial derivatives of that function.

3. It is said in section E3.6.2 that the reason the error propagation approach is used as a reference is that it has been defined and checked by the IPCC and is clearly set out in the IPCC 2000 Good Practice Guidance [262] and the 2006 Guidelines [255]. The 'error propagation approach' referred to is the LPU in the GUM [2] and occurs specifically as [259, section 6.3, formula (6.3)].

   Comment 1: That it is 'clearly set out' in [262] and [255] is not in itself a sufficient reason for using the error propagation approach. That it has been defined and checked by the IPCC also seems to be not a good reason unless the definition and checking comprises some high degree of validation. If that is the case, it is unclear what evidence has been used to support the statement.

   Comment 2: The statement should be compared with the attitude of the JCGM. The MCM as prescribed by the GUM-S1 [3,4] can handle situations where the uncertainties are large or the measurement model is (perhaps highly) non-linear, whereas the error propagation approach (the LPU in the GUM [2] can produce invalid uncertainty statements. Thus, MCM can be used as a reference against which other methods such as LPU, which require linearization of the model, can be compared. This way of thinking is very different from that of the IPCC, which regards the error approach as a reference. MCMs are seen, at least by Bayesian statisticians, as a 'gold standard' for uncertainty propagation: Huggins et al. [263] state:

> 'Classical Monte Carlo methods ... remain the gold standard for approximate Bayesian inference because they have a robust finite-sample theory and reliable convergence diagnostics.'

There are many other references (such as [96, 263–266]) to the Monte Carlo method as a 'gold standard' for uncertainty propagation.

Non-pathological examples exist quantifying the disagreement of results produced by the two approaches for various case studies. One instance arose in mass calibration [3] where the law of propagation of uncertainty gave a standard uncertainty for the measurand of 0.054 mg whereas that produced by the Monte Carlo method was 0.075 mg, some 40 % larger. There was good agreement (to the two significant decimals reported) between the standard uncertainty determined by Monte Carlo and that (0.075 mg) provided by the law of propagation based on higher-order terms, that is, employing a better approximation to the non-linear measurement function than that given by linearisation as in the basic GUM approach.

4. Within the context of its GHG emissions' inventory, the UK has implemented the error propagation approach as set out in its guidance. It is stated:

   > 'Uncertainty estimation by simulation (MCM) cannot be prescriptive, and will depend on how the country constructs its model and the correlations considered. Therefore, 'there is a greater likelihood of errors being introduced in the model used to estimate uncertainty by Monte Carlo simulation'. It is further said that 'if all the distributions in the Monte Carlo model were normal, and the assumed correlations were identical, the estimated errors on the trend from the Monte Carlo model should approach those estimated by the error propagation approach if enough iterations are done'.

   Comment: We disagree that 'uncertainty estimation by simulation (MCM)' cannot be prescriptive. It is said in [3]:

   > 'Whereas there are some limitations to the GUM uncertainty framework, the propagation of distributions [implemented by Monte Carlo] will always provide a PDF for the output quantity that is consistent with the model of the measurement and the PDFs for the input quantities.'

   Comment: A comparison of MCM and LPU is possible if the means and standard deviations of the probability distributions used by MCM are taken as the estimates and associated standard uncertainties used by LPU. There is no choice in the model itself. Both approaches use exactly the same model but, as stated, LPU linearizes it. Rather than errors being introduced by MCM, they will always be introduced by LPU unless the model is linear. Importantly, the main model used in GHG inventory work is (E3.6.3), which is non-linear. The statement about the conditions for the two methods to converge is sound.

5. *Comment:* It would seem that in discussing correlations, the IPCC mainly considers 'perfect correlations', that is, a variable is totally dependent on other variables and hence can be eliminated from the analysis [5, 267].

6. The IPCC states [268, page 36]:

'Based on the Guidance Note for Lead Authors of the IPCC Fifth Assessment Report [AR5] on Consistent Treatment of Uncertainties, this WGI Technical Summary and the WGI Summary for Policymakers rely on two metrics for communicating the degree of certainty in key findings, which is based on author teams' evaluations of underlying scientific understanding:

- Confidence in the validity of a finding, based on the type, amount, quality and consistency of evidence (e.g., mechanistic understanding, theory, data, models, expert judgement) and the degree of agreement. Confidence is expressed qualitatively.
- Quantified measures of uncertainty in a finding expressed probabilistically (based on statistical analysis of observations or model results, or expert judgement).'

Comment:

- The use of qualitative statements of uncertainty is reasonable, particularly where such statements correspond to clearly stated probability 'bands'. The fact that almost any model used as a basis of calculation and the expression of uncertainty is not unique is recognized by the IPCC. The JCGM also recognizes the importance of this point. It released in 2020 a guide on modelling that embodies this principle [5].
- The JCGM is very much concerned with deriving and expressing uncertainties probabilistically. Three of its guides [3,4,6] (with more under development) use probabilistic considerations in the use of Monte Carlo methods for uncertainty propagation and for conformance assessment. The JCGM promotes the use of a probability distribution as the primary measurement result, certainly for the higher end of the measurement hierarchy. It recognizes the relevance of summary location and scale statistics, particularly for the lower end of the measurement hierarchy. The international vocabulary of basic and general terms in metrology [89] also recognizes that a measurement result may usefully be specified as a probability distribution but also that summaries of the distribution are often useful in practice.

7. IPCC states [268, page 121]:

'**Treatment of Uncertainties**

For AR5, the three IPCC Working Groups use two metrics to communicate the degree of certainty in key findings: (1) Confidence is a qualitative measure of the validity of a finding, based on the type, amount, quality and consistency of evidence (e.g., data, mechanistic understanding, theory, models, expert judgment) and the degree of agreement; and (2) Likelihood provides a quantified measure of uncertainty in a finding expressed probabilistically (e.g., based on statistical analysis of observations or model results, or both, and expert judgement).

'**Advances in Measurement and Modelling Capabilities**

Over the last few decades, new observational systems, especially satellite-based systems, have increased the number of observations of the Earth's climate by orders of magnitude. Tools to analyse and process these data have been developed or enhanced to cope with this large increase in information, and more climate proxy data have been acquired to improve our knowledge of past changes in climate. Because the Earth's climate system is characterized on multiple spatial and temporal scales, new observations may reduce the uncertainties

surrounding the understanding of short timescale processes quite rapidly. However, processes that occur over longer timescales may require very long observational baselines before much progress can be made.'

Comment: The attitude here is consistent with that of the JCGM. There are two forms of uncertainty evaluation prescribed in the GUM and related JCGM documents: Type A (statistical) and Type B (non-statistical including expert judgement). An uncertainty evaluation almost always involves a combination of Type A and Type B evaluations. In [2], frequency-based statistics are used for the former and Bayesian statistics for the latter, whereas, in later documents [3, 4], they are both put on a Bayesian footing.

Example E3.7

Greenhouse gas emission inventories — emission estimates calculated by measurement of ambient mixing ratios combined with inverse modelling

T. Arnold, M.G. Cox

E3.7.1 Summary

A top-down Bayesian method using a weakly-informative prior for assessing emission estimates is presented. It is based on the paper 'Inverse modelling of CF_4 and NF_3 emissions in East Asia' by Arnold et al. published in Atmospheric Chemistry and Physics in 2017. The method hinges on the calculation of ambient mixing ratios combined with inverse modelling. It is compared with bottom-up methods used for compiling country-scale greenhouse gas emission inventories.

E3.7.2 Introduction of the application

The example described here is predominantly based on material given in detail in [269] from which major extracts have been made consistent with the provisions of open-access publications. Also see [270].

The major greenhouse gases (GHGs) — carbon dioxide, methane, and nitrous oxide — have natural and anthropogenic sources. The synthetic fluorinated species — chlorofluorocarbons (CFCs), hydrochlorofluorocarbons (HCFCs), hydrofluorocarbons (HFCs), and perfluorocarbons (PFCs), sulfur hexafluoride (SF_6) and nitrogen trifluoride (NF_3) — are almost or entirely anthropogenic and are released from industrial and domestic appliances and applications. Of the synthetic species, tetrafluoromethane (CF_4) and NF_3 are emitted nearly exclusively from point sources of specialised industries [271–273]. Although these species currently make up only a small percentage of current emissions contributing to global radiative forcing, they have potential to form large portions of specific company, sector, state, province, or even country-level GHG budgets.

Although decadal trends in the atmospheric abundances of CF_4 and NF_3 have been well characterized and have provided a time series of global total emissions, information on locations of emissions contributing to the global total is currently poor. Observations made between 2008

and 2015 from Gosan station (hereafter referred to as GSN), Jeju Island, South Korea (part of the Advanced Global Atmospheric Gases Experiment network), together with an atmospheric transport model, were used to make spatially disaggregated emission estimates of these gases in East Asia (disaggregation: separation into component parts). An informative Bayesian prior was used in [269]. Due to its weakness, the resulting emission estimates are mostly influenced by the atmospheric observations, but the prior has a key stabilizing influence.

Within a quantification of emissions determination, top-down studies of CF_4 and NF_3 emissions have shown bottom-up inventories can be inaccurate. Kim et al. [274] showed that global bottom-up estimates for CF_4 are as much as 50 % lower than top-down estimates. Arnold et al. [271] showed that the best estimates of global NF_3 emissions calculated from industry information and statistical data total only about 35 % of those estimated from atmospheric observations.

Accurate emission estimates of NF_3 and CF_4 are difficult to make based on simple parameters such as integrated country-level uptake rates and leakage rates, which, for example, underpin calculations of HFC emissions. Active or passive activities to reduce emissions vary between countries, and between industries and companies within countries. The impetus to have an accurate understanding of emissions is also lacking in regions that have not been required to report emissions under the United Nations Framework Convention on Climate Change (UNFCCC).

This problem is compounded by the difficulty in making observations of these gases. After methane, CF_4 and NF_3 are the two most volatile GHGs. They have very low atmospheric abundances, making routine observations in the field to the required accuracy difficult. The Advanced Global Atmospheric Gases Experiment (AGAGE) has been monitoring the global atmospheric trace gas budget for decades [275]. Most recently, AGAGE's 'Medusa' pre-concentration GC-MS (gas chromatography–mass spectrometry) system has been able to measure a full suite of the long-lived halogenated GHGs [276, 277]. The Medusa is the only instrument demonstrated to measure NF_3 in ambient air samples and the only field-deployable instrument capable of measuring CF_4. The utility of the Medusa on Jeju Island has already been demonstrated in numerous previous studies to understand emissions of many GHGs from East Asia [278–280].

Observations of CF_4 from 2008 and NF_3 from 2013 were used in an inversion framework — coupling each observation with an air history map computed using a particle dispersion model [281]. These observations were used to find emission hot-spots in this unique region with weak prior information, showing that East Asia is a major source of these species. Focused mitigation efforts, based on these results, could have a significant impact on reducing GHG emissions from specific areas.

Air reaches GSN from the most heavily developed areas of East Asia, making the observations and their interpretation an informative source for top-down emission estimates in the region. Ambient air observations are made every 130 min and are bracketed with measurement of a standard before and after the air sample to correct for instrumental drift in calibration. Further details on the methodology for the calibration of these gases are given in [272, 275, 276, 282].

E3.7.3 Specification of the measurand(s)

The measurands are the spatially disaggregated gas emissions in East Asia, especially NF_3 and CF_4 in South Korea.

E3.7.4 Measurement model

The measurement model is described by the process used to obtain, ultimately, the required measurands, given in the following subsections.

E3.7.4.1 Atmospheric model

Lagrangian particle dispersion models [283] are well suited to determine emissions of trace gases on this spatial scale as they can be run backwards, allowing for the source–receptor relationship to be calculated efficiently. A transport model, the Numerical Atmospheric dispersion Modelling Environment (NAME III), henceforth called NAME, developed by the UK Met Office [284, 285], was used. A grid is laid down over the region of interest and inert particles are advected backwards in time with a mass associated with each trajectory. Hence, output is provided as the time-integrated near-surface (0 m to 40 m) air concentration ($\mathrm{g\,s\,m^{-3}}$) in each grid cell — the surface influence resulting from a conceptual release at a specific rate ($\mathrm{g\,s^{-1}}$) from the site. 'Offline', this surface influence is divided by the total mass emitted during the 1 h release time and multiplied by the geographical area of each grid cell to form a new array with each component representative of how $1\,\mathrm{g\,m^{-2}\,s^{-1}}$ of continuous emissions from a grid cell would result in a measured concentration at the model's release point (the measurement site). Multiplication of each grid component by an emission rate then results in a contribution to the concentration.

The meteorological parameter inputs to NAME are from the Met Office's operational global Numerical Weather Prediction (NWP) model, the Unified Model (UM) [286]. The number of vertical levels in the UM has increased over this period, with NAME using the lowest 31 levels in 2009 and the lowest 59 levels in 2015. The NAME model was run to estimate the 30 d history of the air on the route to GSN. The time-integrated air concentration (dosage) was calculated in each grid cell.

The computational domain covers 391 grid cells east to west and 340 grid cells south to north and extends to more than 19 km vertically. Despite the improvement in the resolution of the UM from some 40 km to 17 km over the time period covered, the resolution of the NAME output was kept constant throughout. For each 1 h period, 5000 inert model particles were used to describe the dispersion of air. By dividing the dosage in $\mathrm{g\,s\,m^{-3}}$ by the total mass emitted ($3600\,\mathrm{s\,h^{-1}} \times 1\,\mathrm{h} \times 1\,\mathrm{g\,s^{-1}}$) and multiplying by the geographical area of each grid cell ($\mathrm{m^2}$), the model output was converted into a dilution matrix $\boldsymbol{H}$, each element ($\mathrm{s\,m^{-1}}$) of which dilutes a continuous emission of $1\,\mathrm{g\,m^{-2}\,s^{-1}}$ from a given grid cell over the previous 30 d to simulate an average concentration ($\mathrm{g\,m^{-3}}$) at the receptor (measurement point) during a 1 h period. $\boldsymbol{H}$ contains additional contributions based on the sensitivity of changes in domain boundary conditions on measured mixing ratio.

E3.7.4.2 Inversion framework

Let $\boldsymbol{y}$ denote the observed concentrations, which comprise two distinct components: (a) the Northern Hemisphere (NH) background concentration, referred to as the baseline, which changes slowly over time, and (b) rapidly varying perturbations above the baseline. These deviations above background are assumed to be caused by emissions on a regional scale that have yet to be fully mixed on the hemisphere scale. The magnitude of these deviations and, crucially, how they change as the air arriving at the stations travels over different areas, is the key to understanding where the emissions have occurred.

For most long-lived trace gases (lifetimes of years or longer), the assumption that atmospheric mole fractions respond linearly to changes in emissions holds well, that is,

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{e},$$

where $\boldsymbol{x}$ is a vector of emissions and domain boundary conditions, $\boldsymbol{y}$ is a vector of observations and $\boldsymbol{e}$ is a vector of residuals. A Bayesian framework is typically used in trace gas inversions and thus, incorporating a priori information, gives rise to the cost function

$$\boldsymbol{C} \equiv (\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x})^{\top}\boldsymbol{R}^{-1}(\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}) + (\boldsymbol{x} - \boldsymbol{x}_{\mathrm{p}})^{\top}\boldsymbol{B}^{-1}(\boldsymbol{x} - \boldsymbol{x}_{\mathrm{p}}), \tag{E3.7.1}$$

where $\boldsymbol{R}$ is a diagonal covariance matrix of combined model and observation squared standard uncertainties, $\boldsymbol{x}_{\mathrm{p}}$ is a vector of prior estimates of emissions and domain boundary conditions, and $\boldsymbol{B}$ is a diagonal covariance matrix associated with $\boldsymbol{x}_{\mathrm{p}}$.

The first term on the right side of expression (E3.7.1) is a measure of the mismatch between the modelled and observed time series at the observation stations. The second term describes the mismatch between the emissions and domain boundary conditions $\boldsymbol{x}$ and prior estimated emissions and domain boundary conditions $\boldsymbol{x}_{\mathrm{p}}$ considering the uncertainties embodied in $\boldsymbol{B}$ (see section E3.7.5.1).

The basis for expression (E3.7.1) is as follows. For an arbitrary prior probability distribution for $\boldsymbol{x}$, there is in general no analytical solution for the posterior distribution. However, assuming underlying normality, independence and homogeneity for the elements of $\boldsymbol{x}$, by using a so-called conjugate prior the posterior distribution for $\boldsymbol{x}$ can be derived analytically [65]. Under such an assumption, the formulation (E3.7.1) is obtained.

The cost function is minimized with respect to $\boldsymbol{x}$ using a high-quality non-negative least squares algorithm (NNLS) [287], which determines the least-squares solution under the condition that the emissions are non-negative.

The aim of the inversion is to estimate the spatial distribution of emissions across a defined geographical area. The emissions are assumed to be constant over the inversion time period (in this case, one calendar year, as is typically reported in inventories). Assuming the emissions are invariant over such periods is a simplification deemed necessary given the limited number of observations. In order to compare the observations and the model time series, the latter are converted from air concentration in $\mathrm{g\,m^{-3}}$ to mole fraction, for instance, in parts per trillion, using the modelled temperature and pressure at the observation point.

E3.7.4.3 Baseline calculation and domain boundary conditions

For each observation at GSN, it is valuable to have an accurate understanding of the portion of the total mixing ratio arriving from outside the inversion domain and the portion from emission sources within the domain; otherwise, emissions from specific areas could be biased (over- or under-estimated). GSN is uniquely situated, receiving air masses from all directions over the course of the year, which can have distinct compositions of trace gases, driven mainly by the different emission rates between the two hemispheres and slow inter-hemispheric mixing.

In addition to the time-integrated air concentration produced by NAME (section E3.7.4.1), the 3-D coordinate where each particle left the computational domain was also recorded. This information was then post-processed to produce the percentage contributions from the boundaries of the 3-D domain. Thus, the influence of air arriving at GSN from outside the domain was simplified as a combination of air masses arriving from discrete directions.

Observations from the Mace Head observatory (hereafter termed MHD) on the west coast of Ireland — a key AGAGE site providing long-term in situ atmospheric observations — were used to act as a starting point for an estimate of the composition of air from northern hemisphere (NH) mid-latitudes entering the East Asian domain.

The composition of air arriving from any of the discrete directions was calculated using corresponding multiplying factors applied to the MHD baseline, which were included as part of the state vector $\boldsymbol{x}$ and taken as constant factors for a given inversion year. The prior baseline was therefore perturbed as part of the inversion based on the relative contribution of air arriving from different boundaries of the 3-D domain and the multiplying factors that are included within the cost function (E3.7.1).

E3.7.4.4 Domains and inversion grids

The domain used in the inversion is smaller than the computational NAME transport model domain. GSN is situated within a region surrounded by countries with major developed industries, and therefore the site is relatively insensitive to emissions from further away that are diluted en route to the site. NAME is run on a larger domain to ensure that on the occasion when air circulates out of the inversion domain and then back, its full 30-day history in the inversion domain is included.

An initial computational inversion grid (termed here the 'coarse grid') was created based on (a) aggregated information from the NAME footprints over the period of the inversion (in this case, 1 year), aggregating fewer grid cells in areas that are 'seen' the most by GSN, and (b) the prior emissions flux, that is, areas known to have low emissions (for example, ocean), with higher aggregation. Coarse grid cells were not aggregated over countries or regions, a total of $n \approx 100$ coarse grid cells being created. After the initial inversion, an individual coarse grid cell was chosen to divide in two by area. The decision on which single coarse grid cell was split was based on the posterior emission density $(\mathrm{g\,a^{-1}\,m^{-2}})$[1] of the coarse grids and the ability of the posterior emissions to have an impact on the observations at GSN (using information from the NAME output). A new inversion was run using identical inputs except the number of grid cells was increased by one by the above process. This exercise of adding one new grid cell was repeated 50 times, creating ≈ 150 coarse grid cells within the inversion domain for the final inversion. The results from the inversions with the maximum disaggregation are presented in [269].

E3.7.5 Uncertainty propagation

E3.7.5.1 Prior emission information

Global emission estimates of CF_4 and NF_3 using atmospheric observations have demonstrated that bottom-up accounting methods for one or more sectors, or one or more regions, are highly inaccurate [271,272]. This study makes no effort to improve such inventory methods but instead focuses on minimizing the reliance of prior information on the adopted Bayesian-based posterior emission estimates.

Data sets from the Emissions Database for Global Atmospheric Research (EDGAR) v4.2 emission grid maps [288] were used as the basis of the prior information used. Since these data sets cover the years 2000 to 2010, the prior based on 2010, the latest available year, was applied

[1] The unit a (annum) is used for 'year'. Note that 'year' is not an SI unit or derived SI unit, but is used here for consistency with the source material [269].

for each year from 2011 to 2015. The EDGAR emission maps were first re-gridded based on the lower resolution of the inversion grid. In order to remove the influence of the within-country prior spatial emission distribution, each country's emissions were averaged across their entire landmass.

Five inversion experiments having increasingly weaker prior information were carried out. The first experiment involved using $\boldsymbol{B}$ based directly on the information provided by the EDGAR emission maps. The following four experiments involved making $\boldsymbol{B}$ successively larger by a factor of 10. The sensitivity of the solution to the prior emission uncertainty was then tested and thus evidence provided for the low influence of prior information on the emission estimates in the posterior.

E3.7.5.2 Input (prior) uncertainties

In addition to questionable prior information, another appreciable source of uncertainty in estimating emissions stems from the model, from both the input meteorology and the atmospheric transport model itself. The covariance matrix $\boldsymbol{R}$ is a crucial part of expression (E3.7.1) that allows uncertainties assigned to the observations to be adjusted depending on how well it is thought the model is performing at that time. It describes, per hour period, a combined uncertainty of the model and the observation at each time. (The method of assigning model uncertainties remains under development.) One method that has been applied to the modelling of GSN observations is as follows. All elements of the modelled meteorology (wind speed and direction, boundary layer height (BLH), temperature, pressure, etc.) are important in understanding the dilution and uncertainty in modelling from source to receptor. However, quantifying the impact of each element that each model particle experiences in order to quantify fully the model uncertainty at each measurement time is beyond what is available from numerical weather prediction models. So in order to attempt to quantify a model or observation uncertainty a pragmatic approach was taken and BLH modelled at the receptor as a proxy.

Emissions are primarily diluted by transport and mixing within the planetary boundary layer (PBL), and hence modelling of the PBL height (BLH) is crucial for accurate modelling of the mixing ratios. Changes in BLH at or surrounding the measurement location can cause appreciable changes to the measured mixing ratio. A low BLH (causing a larger model uncertainty) has two implications for observations at the Gosan site. The first implication is a greater possibility of air from above the PBL being sampled in reality but not in the model. Subtle changes in the BLH at the exact measurement location are not well modelled and the difference between sampling above or within the PBL can have considerable influence on the amount of pollutant assigned to a back trajectory. The second implication is the greater influence of emissions from sources very near GSN. A lower BLH means that a lower rate of dilution of local emissions will occur, in turn increasing the signal of the local pollutant above the baseline. A relatively small change in a low BLH will have considerable influence on this dilution compared to the same change on a high BLH. Thus, any error in the BLH at low levels can considerably amplify the uncertainty in the pollutant dilution. This effect is coupled with the fact that the modelled BLH has considerable uncertainty especially when BLH is low. To assign empirically a model standard uncertainty u_{model} to each hourly window of observations, BLH model information is used:

$$u_{\text{model}} = u_{\text{baseline}} \times f_{\text{BLH}},$$

where u_{baseline} is the variability (standard deviation) associated with the baseline calculation (see section E3.7.4.3), and f_{BLH} is a multiplying factor that increases or decreases the relative uncertainty assigned to each model time period. f_{BLH} is based on modelled BLH magnitude and

variability over a 3 h period and is calculated thus:

$$f_{\mathrm{BLH}} = \frac{\max(\text{BL-inlet})}{\min(\text{BL-inlet})} \times \frac{\text{Threshold}}{\min(\text{BLH})}.$$

Here max(BL-inlet) is the larger of 100 m and the maximum distance, calculated hourly, between the inlet and the modelled BLH within a period of 3 h around the measurement time, min(BL-inlet) is the smallest of the distances calculated between the inlet and the BLH over the same 3 h period, 'Threshold' is an arbitrary value set at 500 m, and min(BLH) is the smallest BLH recorded over the 3 h period. Thus, the relative assigned uncertainty considers the proximity of the varying BLH to the inlet height and a recognition that observations taken when the BLH is varying at higher altitudes (more than 500 m above ground level) is likely to have less impact and therefore have lower uncertainty compared to those taken when the BLH is varying at lower altitudes.

E3.7.5.3 Propagating uncertainties through the cost function

The method used for propagating uncertainties through the cost function was closely based on the treatment [289] and is not repeated here.

E3.7.6 Reporting the result

E3.7.6.1 Country total emission estimates

Table 1 in the open-access publication [269] provides a summary of emission estimates from the five major emitting countries or regions within the East Asian domain. These posterior emission estimates use a prior emission uncertainty in each fine grid cell of 100 times the emission magnitude (see section E3.7.5.1).

E3.7.6.2 HFC-23

Fang et al. [278] conducted a very thorough bottom-up study within their work on HFC-23, constraining an inversion model using both prior information and atmospheric observations. They employed an inverse method based on the FLEXible PARTicle dispersion model (FLEXPART) using observations from three sites in East Asia, one of which was GSN. The estimation process uses a completely independent inverse method and only data from GSN, yet the results are very close to those of [278], being about 10 % different in 2008 and 20 % different in 2012, and to [290]. The posterior uncertainties in these two studies mainly reflect the difference in the uncertainty assumed for the prior information. Because a very high level of uncertainty is assumed for the prior emissions, the posterior uncertainties are significantly higher. However, these inversion result estimates are lower than estimates based on inter-species correlation analysis [280] where calculated emissions of HFC-23 from China in 2008 in the range 7.2 Gg year^{-1} to 13 Gg year^{-1} are reported. Using a CO tracer-ratio method, Yao et al. [291] estimated particularly low emissions of 2.1 Gg a^{-1} ± 4.6 Gg a^{-1} for 2011-2012. (The use of '±' as here is not recommended in the GUM [2]; we nevertheless retain it in this context since the numbers given are to be interpreted in spirit.) The estimates derived from atmospheric inversions do not rely on any correlations with other species or known emissions for certain species and, given two separate inversion studies, have produced very similar results. These results suggest a more reliable top-down emission

estimate of HFC-23 is provided. As well as providing an independent validation of the previous work on HFC-23 [278,290], the alignment of the obtained HFC-23 emission estimates with those previous studies provides confidence in the inversion methodology adopted for the CF_4 and NF_3 emission estimates.

E3.7.6.3 CF_4

Generally, the understanding of emissions of CF_4 and NF_3 is very poor, which is highlighted in global studies based on atmospheric observations that show bottom-up estimates of emissions are significantly underestimated [271,272]. With such a poor prior understanding of emissions, the effect of prior uncertainty on the posterior emissions is assessed. With assignment of uncertainty on the prior of each fine grid cell at 10 times the prior emission value, the posterior is still significantly constrained by the prior for both China and South Korea. When larger uncertainties are applied to the prior (100 times to 10 000 times), the posterior estimates are very consistent, indicating that when greater than 100 × standard uncertainty is applied, emission estimates are most significantly constrained by the atmospheric observations.

The estimates obtained in this work are significantly higher than those provided by inter-species correlation: [279] Kim et al. estimated CF_4 emissions in the range of only 1.7 Gg a^{-1} to 3.1 Gg a^{-1} in 2008 and Li et al. [280] only 1.4 Gg a^{-1} to 2.9 Gg a^{-1} over the same period. The inter-species correlation approach inherently requires that the sources of the different gases that are compared are coincident in time and space. HCFC-22 was used as the tracer compound for China in [279, 280] with a calculated emission field from an inverse model. Most emissions of this gas originate from fugitive release from air conditioners and refrigerators. However, CF_4 is emitted mostly from point sources in the semiconductor and aluminium production industries with different spatial emission distribution within countries, and likely different temporal characteristics compared to HCFC-22.

Emission estimates from South Korea and Japan are an order of magnitude lower than those from China. For 2008, Li et al. [280] estimate emissions of CF_4 from the combination of South and North Korea of 0.19 Gg a^{-1} to 0.26 Gg a^{-1} and from Japan of 0.2 Gg a^{-1} to 0.3 Gg a^{-1}, which although mutually consistent are on the low end of the range of the estimates obtained in the current study for that year: see table 1 in [269] As one of the largest, if not the largest, countries for semiconductor wafer production, Taiwan is also an emitter of CF_4. However, observations at GSN provide only poor sensitivity to detection of emissions from Taiwan, and the results obtained can only suggest that emissions are likely to be less than 0.5 Gg a^{-1}. North Korean emissions were small and no annual estimate was above 0.1 Gg a^{-1}.

E3.7.6.4 NF_3

The general understanding of NF_3 emissions from inventory and industry data is even poorer than for CF_4. On a global scale, the emission estimates from industry are underestimated [271]. This study suggests that at least some emissions of NF_3 stem from China; however, gaining meaningful quantitative estimates has been difficult due to large uncertainties. In contrast, the posterior estimates of emissions from South Korea have relatively small uncertainties. Emissions from China travel a greater distance to the measurement site compared to emissions from South Korea. Thus, the magnitudes of NF_3 pollution events from China, in terms of the mixing ratio detected at GSN, are smaller than for pollution arriving from neighbouring South Korea. Also, the poorer measurement precision for NF_3 compared to CF_4 leads to a larger baseline uncertainty, which in turn

affects the uncertainty on the pollution episode, especially for more dilute signals. Emission estimates for Japan are difficult to make without improved prior information and more atmospheric observations in other locations. Other large changes in the obtained emission estimates from 2014 to 2015 could arguably be real. For example, Japan's National Inventory Report for NF_3 shows a reduction in emissions of 63 % between 2013 and 2015 (Ministry of the Environment Japan et al., 2018), which is within the uncertainty of the observed relative rate of decrease.

As for CF_4, emission estimates of NF_3 from Taiwan and North Korea are highly uncertain. However, results from this study indicate that emissions of NF_3 from Taiwan might be lower than from South Korea despite very similar-sized semiconductor production industries. Focusing on the more meaningful estimates from South Korea, emissions of NF_3 in 2015 are estimated to be $(0.60 \pm 0.07)\,\mathrm{Gg\,a^{-1}}$, which equates to $(9660 \pm 1127)\,\mathrm{Gg\,a^{-1}}$ CO_2-equivalent emissions based on a GWP_{100} of 16 100. GWP_{100} is an important policy metric. It means Global Warming Potential and GWP_{100} is specific for a 100 a time horizon. It is defined as the global warming impact over 100 years for a unit mass of emissions relative to the same unit mass of emissions of CO_2, that is, 1 g of NF_3 emitted has 16 100 times the impact on global warming as 1 g of CO_2. In this way, emissions are weighted in trading schemes, etc..

These NF_3 emissions constitute some 1.6 % of the country's CO_2 emissions [292], thus making an appreciable impact on its total GHG budget. Further, given that the sources of NF_3 are relatively few, these emissions can be assigned to a small number of industries, potentially making NF_3 an easy target for focused mitigation policy. Rigby et al. [293] updated the global emission estimates from Arnold et al. [271], and calculated an annual emission estimate of $1.61\,\mathrm{Gg\,a^{-1}}$ for 2012, with an average annual growth rate over the previous 5 years of $0.18\,\mathrm{Gg\,a^{-1}}$. Linearly extrapolating this growth to 2014 and 2015 leads to projected global emissions of $1.97\,\mathrm{Gg\,a^{-1}}$ and $2.15\,\mathrm{Gg\,a^{-1}}$ for 2014 and 2015, respectively. Thus, South Korean emissions as a percentage of these global totals equate to 20 % for 2014 and 28 % for 2015, which is around the proportion of semiconductor wafer fabrication capacity in South Korea relative to global totals (20 %) [294].

E3.7.7 Interpretation of results

The first Bayesian-based inversion estimates of CF_4 and NF_3 from the East Asian region were presented in [269]. In doing so, a class of successively weaker informative priors, largely removing the influence of available bottom-up information, were used. As a consequence, the results of the study are predominantly driven by the data for the period covered. However, the use of a weak prior has a stabilizing influence on the results.

The largest CF_4 emissions are from China, estimated at $4\,\mathrm{Gg\,a^{-1}}$ to $6\,\mathrm{Gg\,a^{-1}}$ for six of the eight years studied, appreciably larger than previous estimates. Despite significantly smaller emissions from South Korea, the spatial disaggregation of CF_4 emissions was consistent between independent inversions based on annual measurement data sets, indicating the north-west of South Korea is a hot spot for significant CF_4 release, presumably from the semiconductor industry. Emissions of NF_3 from South Korea were quantifiable with small uncertainty, and represent large emissions on a CO_2-equivalent basis (some 1.6 % of South Korea's CO_2 emissions in 2015). HFC-23 emissions were also calculated using the same inversion methodology with large uncertainty on prior information. Good agreement was found with other studies in terms of aggregated country totals and spatial emissions' patterns, providing confidence that the adopted methodology is suitable and the conclusions are justified for estimates of CF_4 and NF_3.

The results highlight an inadequacy in both the bottom-up reported estimates for CF_4 and NF_3 and the limitations of the current measurement infrastructure for top-down estimates for these gases. Adequate bottom-up estimates have been lacking due to the absence of reporting requirements for these gases from China and South Korea, and top-down estimates have been hampered by poor measurement coverage due to the technical complexities required to measure accurately these volatile, low-abundance gases. Improvements in both bottom-up information and measurement coverage, alongside refinements in transport modelling and developments in inversion methodologies, will lead to improved emissions estimates of these gases in future studies.

HFC-23 emissions were also calculated, not to add to current knowledge, but provided greater confidence in the adopted methodology.

The method of calculating emissions of CHF_3 (HFC-23) was also applied to data for the period between 2008 and 2012. Good agreement was found with other studies, which helps to support the methodology adopted for CF_4 and NF_3.

The Bayesian method adopted in the inversion framework of section E3.7.4.2 is capable of employment in other applications. Already, it is being used by BIPM and NPL in transferring the model of the BIPM ionization chamber [295], used to establish equivalence of radioactivity measurement made by the national metrology institutes, to a secondary chamber at the BIPM.

The data and software used in [269] and reported here are owned by a third party and therefore not publicly available.

Acknowledgment

Valuable comments were received from Stephen Ellison (LGC) and Adriaan van der Veen (VSL).

Example E3.8

Preparation of calibration gas mixtures using permeation

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

E3.8.1 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.

E3.8.2 Introduction of the application

Permeation [296] is one of several techniques to prepare dynamically calibration gas mixtures [297]. 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 [97] 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 [298, 299].

E3.8.3 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.

E3.8.4 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.

E3.8.4.1 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 [296]. 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 [300]. 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 [296]

$$y_k = \frac{\frac{\dot{m}_1}{\bar{M}_1} x_{1k}}{\frac{\dot{m}_1}{\bar{M}_1} + \frac{\dot{m}_2}{\bar{M}_2}}, \tag{E3.8.1}$$

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 [97] is necessary to determine the value for x_{1k} and the associated standard uncertainty. Furthermore, equation (E3.8.1) 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 [301,302], which can be used if the assumptions underlying equation (E3.8.1) are not met.

E3.8.4.2 Permeation rate

The permeation rate (mass flow rate from the permeation tube) $\dot{m}_1$ is defined as [296]

$$\dot{m}_1 = q_{m1} = \frac{\mathrm{d}m_1}{\mathrm{d}t},$$

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.

E3.8.4.3 Material purity

Material purity affects the molar masses $\bar{M}_1$ and $\bar{M}_2$ of the parent gases in equation (E3.8.1). 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 [97]

$$x_{sj} = 1 - \sum_{\substack{i=1 \\ i \neq s}}^{q} x_{ij}, \tag{E3.8.2}$$

where s denotes index of the most abundant component.

E3.8.4.4 Molar masses

The molar mass of any component can be expressed as [301]

$$M_i = \sum_{z=1}^{Z} \nu_{zi} A_z, \tag{E3.8.3}$$

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

$$\bar{M}_j = \sum_{i=1}^{q} x_{ij} M_i, \tag{E3.8.4}$$

where x_{ij} denotes the amount fraction of component i in parent gas j.

E3.8.5 Uncertainty propagation

E3.8.5.1 General

Throughout this example, the law of propagation of uncertainty in the GUM [2] 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.

E3.8.5.2 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 (E3.8.1) 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

$$\begin{aligned} u^2(y_k) = &\left(\frac{\partial y_k}{\partial x_{1k}}\right)^2 u^2(x_{1k}) + \left(\frac{\partial y_k}{\partial \dot{m}_1}\right)^2 u^2(\dot{m}_1) + \left(\frac{\partial y_k}{\partial \bar{M}_1}\right)^2 u^2(\bar{M}_1) + \left(\frac{\partial y_k}{\partial \dot{m}_2}\right)^2 u^2(\dot{m}_2) \\ &+ \left(\frac{\partial y_k}{\partial \bar{M}_2}\right)^2 u^2(\bar{M}_2) + 2\frac{\partial y_k}{\partial \bar{M}_1}\frac{\partial y_k}{\partial \bar{M}_2}u(\bar{M}_1, \bar{M}_2) + 2\frac{\partial y_k}{\partial x_{1k}}\frac{\partial y_k}{\partial \bar{M}_1}u(x_{1k}, \bar{M}_1), \end{aligned}$$

where

$$\begin{aligned} \frac{\partial y_k}{\partial x_{1k}} &= \frac{\dot{n}_1}{\dot{n}_1 + \dot{n}_2} \\ \frac{\partial y_k}{\partial \dot{m}_1} &= \frac{\dot{n_2} x_{1k}}{(\dot{n}_1 + \dot{n}_2)^2}\frac{1}{\bar{M}_1} \\ \frac{\partial y_k}{\partial \bar{M}_1} &= -\frac{\dot{n_2} x_{1k}}{(\dot{n}_1 + \dot{n}_2)^2}\frac{\dot{m}_1}{\bar{M}_1^2} \\ \frac{\partial y_k}{\partial \dot{m}_2} &= -\frac{\dot{n}_1 x_{1k}}{(\dot{n}_1 + \dot{n}_2)^2}\frac{1}{\bar{M}_2} \\ \frac{\partial y_k}{\partial \bar{M}_2} &= \frac{\dot{n}_1 x_{1k}}{(\dot{n}_1 + \dot{n}_2)^2}\frac{\dot{m}_2}{\bar{M}_2^2} \end{aligned}$$

and, assuming that the amount fraction x_{1k} is calculated by difference (see equation (E3.8.2)),

$$u(x_{1k}, \bar{M}_1) = -\sum_{i \neq k} M_i u^2(x_{1i})$$

Uncertainty due to temperature variations in the permeation rate

The temperature dependence of the permeation rate follows Arrhenius' law [303–305]; thus the temperature dependence takes the form

$$\dot{m} = k_0 \exp\left(-\frac{E_A}{RT}\right), \tag{E3.8.5}$$

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

$$\frac{\partial \dot{m}}{\partial T} = \frac{k_0 E_A}{RT^2} \exp\left(-\frac{E_A}{RT}\right). \tag{E3.8.6}$$

Values for k_0 and E_A can be obtained using the measured permeation rate determined at two different temperatures. From equation (E3.8.5), 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 [297] 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 [2].

The uncertainty due to temperature variations in the permeation rate is then calculated as

$$\frac{\partial \dot{m}}{\partial T} = \frac{k_0 E_A}{RT^2} \exp\left(-\frac{E_A}{RT}\right) u(T). \tag{E3.8.7}$$

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 $0.0022\,\mu g\,min^{-1}$.

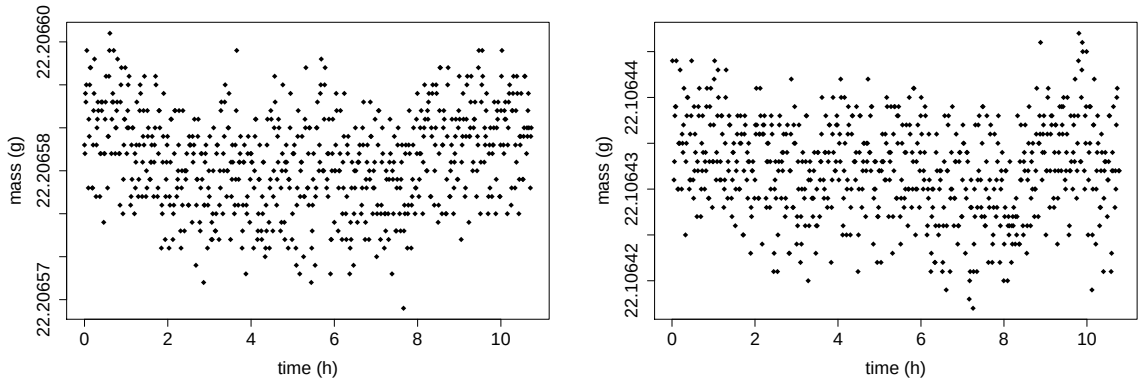
Uncertainty of weighing with the balance

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 [87] or a simplified formula such as given in ISO 6976 [306].

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 E3.8.1.



(a) Combined 20 g and 100 mg weights

(b) 20 g weight

Figure E3.8.1: Results of the weighings performed for a linearity assessment of the balance over a range of 100 mg

From the data in figure E3.8.1, 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

$$d = \mu_2 - \mu_1, \tag{E3.8.8}$$

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 [2], based on the normal distribution, or that in Supplement 1 of the GUM (GUM-S1) [3], 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 [11] package `RStan` [61]. 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 [65]

$$p(\mu, \sigma^2) \propto \sigma^{-2}.$$

The raw observations are distributed as

$$z_i \sim R(y_i, r),$$

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 [65, 307] and also revisited in the framework of evaluating measurement uncertainty [63]. The complete Bayesian model computes the difference d using equation (E3.8.8). 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 E3.8.1.

Fitting the data shown in figure E3.8.1 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).
```

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 [63] 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 E3.8.2). 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.

```
data {
    int<lower=1> N1;
    vector[N1] y1;
    int<lower=1> N2;
    vector[N2] y2;
}
parameters {
    real mu1;
    real mu2;
    real<lower=0> sigma_sq1;
    real<lower=0> sigma_sq2;
    vector<lower=-0.5, upper=0.5>[N1] y_err1;
    vector<lower=-0.5, upper=0.5>[N2] y_err2;
}
transformed parameters {
    real<lower=0> sigma1;
    real<lower=0> sigma2;
    vector[N1] z1;
    vector[N2] z2;
    sigma1 = sqrt(sigma_sq1);
    sigma2 = sqrt(sigma_sq2);
    z1 = y1 + y_err1;
    z2 = y2 + y_err2;
}
model
{
    target += -2 * log(sigma1);
    z1 ~ normal(mu1, sigma1);
    target += -2 * log(sigma2);
    z2 ~ normal(mu2, sigma2);
}
generated quantities {
    real diff;
    diff = mu1-mu2;
}
```

Listing E3.8.1: 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

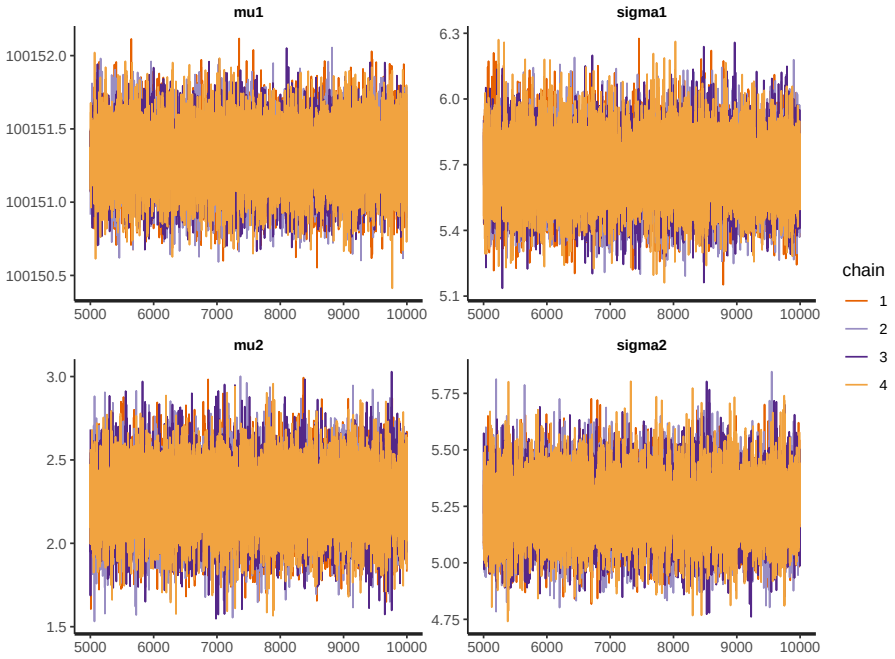


Figure E3.8.2: Trace plot of the four model parameters in the Bayesian model used to fit the weighing data

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 [306, 308]. 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 % [308], 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 E3.8.1 and E3.8.2. According to ISO 6145-10 [296], 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.

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 [226, 309]. This regression method is appropriate here, for it takes into consideration the uncertainties associated with the independent and dependent variables.

Table E3.8.1: 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

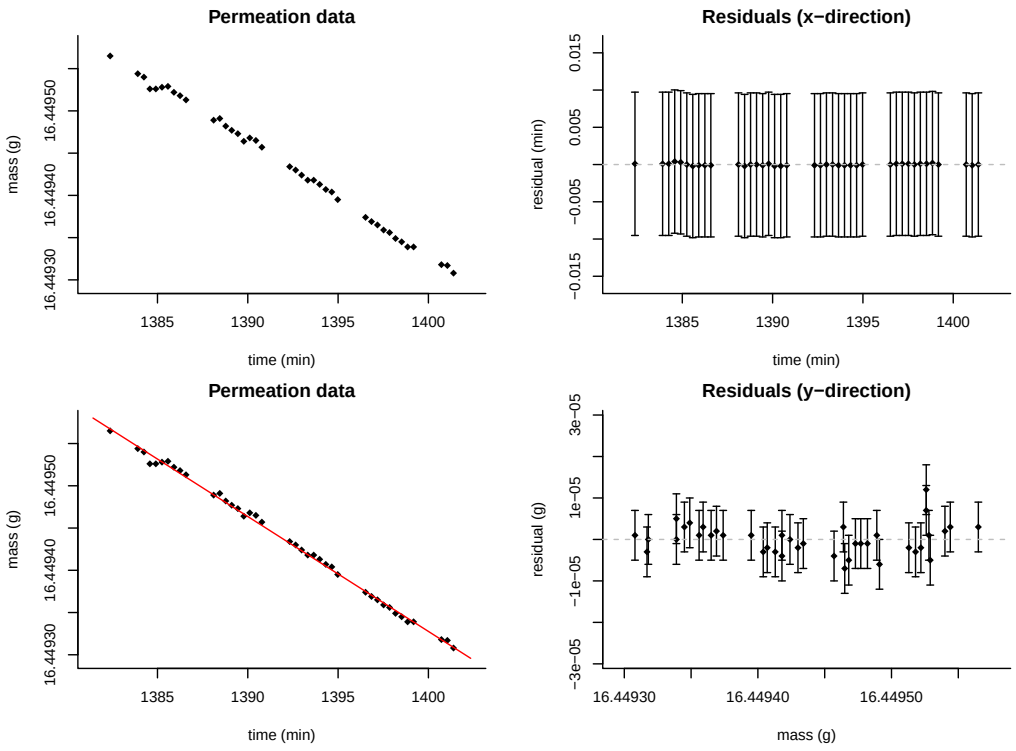


Figure E3.8.3: Results of the errors-in-variables regression of the permeation data at 62 °C from table E3.8.2

Table E3.8.2: 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

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 E3.8.3. 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 [226, 310], namely that their absolute value does not exceed the expanded uncertainty. From the EIV regression, $\dot{m} = -13.599\,\mu g\,min^{-1}$ with standard uncertainty $u(\dot{m}) = 0.088\,\mu g\,min^{-1}$. Using OLS [76], $\dot{m} = -13.599\,\mu g\,min^{-1}$ with standard uncertainty $u(\dot{m}) = 0.108\,\mu g\,min^{-1}$. To the last stated digit, the calculated permeation rates concur, which 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.

E3.8.5.3 Uncertainty due to materials purity

The squared standard uncertainty of x_{sj} is computed as

$$u^2(x_{sj}) = \sum_{\substack{i=1 \\ i \neq s}}^{q} u^2(x_{ij}). \tag{E3.8.9}$$

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 [311]

$$u(x_{sj}, x_{ij}) = -u^2(x_{ij}).$$

Both expressions follow from the applying the law of propagation of uncertainty to equation (E3.8.2). The composition of the materials used is summarised in table E3.8.3. The amount fractions ammonia (parent 1) and nitrogen have been computed using equation (E3.8.2) and the standard uncertainty has been obtained from equation (E3.8.9). The amount fractions methane, carbon monoxide, carbon dioxide and oxygen in nitrogen stem from the limit of quantification of the respective methods. The derivation of the values and standard uncertainties has been performed using the rectangular distribution in accordance with ISO 19229 [97].

Table E3.8.3: Purity information of ammonia (Parent 1) and nitrogen (Parent 2), expressed in amount fractions ($\mathrm{mol\,mol^{-1}}$)

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		

E3.8.5.4 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 [2] to equation (E3.8.3):

$$u^2(M_i) = \sum_{z=1}^{Z} \nu_{zi}^2 u^2(A_z), \tag{E3.8.10}$$

where $u(A_z)$ is obtained by using the rectangular distribution to model the uncertainty of the standard atomic weights [301].

ISO 6142-1 [301] does not take into account covariance between the molar masses of two components, yet according to [311] it can be calculated as:

$$u(M_i, M_j) = \sum_{z=1}^{L} \nu_{zi} \nu_{zj} u^2(A_z), \tag{E3.8.11}$$

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 (E3.8.3). As an illustration of equation (E3.8.11) 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 $\nu_{zi} = 2$ and in ammonia $\nu_{zj} = 3$. 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

$$\boldsymbol{V_M} \times 1 \times 10^{-8} = \begin{bmatrix} 33.3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 43.1 & 33.3 & 33.3 & 4.86 & 4.86 & 0 & 0 & 7.29 \\ 0 & 33.3 & 37.9 & 42.5 & 0 & 4.56 & 0 & 9.13 & 0 \\ 0 & 33.3 & 42.5 & 51.6 & 0 & 9.13 & 0 & 18.3 & 0 \\ 0 & 4.86 & 0 & 0 & 2.43 & 2.43 & 0 & 0 & 3.64 \\ 0 & 4.86 & 4.56 & 9.13 & 2.43 & 6.99 & 0 & 9.13 & 3.64 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2.41 & 0 & 12.0 \\ 0 & 0 & 9.13 & 18.3 & 0 & 9.13 & 0 & 18.3 & 0 \\ 0 & 7.29 & 0 & 0 & 3.64 & 3.64 & 12.0 & 0 & 11.5 \end{bmatrix}, \quad \text{(E3.8.12)}$$

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 E3.8.3. 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 E3.8.4.

Table E3.8.4: 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	28.010 00	6.16×10^{-4}
Carbon dioxide	44.0094	7.18×10^{-4}
Hydrogen	2.015 95	1.56×10^{-4}
Water	18.015 35	2.64×10^{-4}
Nitrogen	28.013 71	4.91×10^{-4}
Oxygen	31.9988	4.27×10^{-4}
Ammonia	17.030 78	3.39×10^{-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 [2, 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.

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 [2] 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 [312]

$$u(Y,Z) = \sum_i \frac{\partial f}{\partial X_i} \frac{\partial g}{\partial X_i} u^2(X_i) + \sum_i \sum_{j \neq i} \frac{\partial f}{\partial X_i} \frac{\partial g}{\partial X_j} u(X_i, X_j). \tag{E3.8.13}$$

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 [4], 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 $\bar{M}_1 = 17.040\,63\,\mathrm{g\,mol^{-1}}$ for ammonia and $\bar{M}_1 = 28.013\,77\,\mathrm{g\,mol^{-1}}$ for nitrogen. The associated squared standard uncertainties and covariance have been computed as $u^2(\bar{M}_1) = 3.99 \times 10^{-6}\,\mathrm{g^2\,mol^{-2}}$,
$u^2(\bar{M}_2) = 2.42 \times 10^{-7}\,\mathrm{g^2\,mol^{-2}}$ and $u(\bar{M}_1, \bar{M}_2) = 1.19 \times 10^{-7}\,\mathrm{g^2\,mol^{-2}}$. 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.

E3.8.6 Reporting the result

The amount fraction NH_3 is $70.8\,\mu\mathrm{mol\,mol^{-1}}$ with standard uncertainty $0.6\,\mu\mathrm{mol\,mol^{-1}}$. 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 $(70.8 \pm 1.2)\,\mu\mathrm{mol\,mol^{-1}}$ ($k = 2$) with 95 % probability.

The full composition of the calibration gas mixture is given in table E3.8.5, computed with the measurement model from ISO 6142-1 [301]. The measurement model in this example only provides the amount fraction, which is also given in table E3.8.5.

E3.8.7 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

Table E3.8.5: Composition of a calibration gas mixture of NH_3 in N_2, expressed in amount fractions ($\mathrm{mol\,mol^{-1}}$)

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	0.000 003	0.0003%
Oxygen	5×10^{-9}	3×10^{-9}	60.0 %
Ammonia	7.08×10^{-5}	5.99×10^{-7}	0.8%

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 [97, 313].

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 E3.8.1 and E3.8.2). 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 E3.8.5.2 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 [314]. 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 [301] 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 [311].

Part E4

Energy

Example E4.1

Evaluation of measurement uncertainty in totalization of volume measurements in drinking water supply networks

A.S. Ribeiro, M.G. Cox, J.A. Sousa, A.M.H. van der Veen, M. Reader-Harris, L.L. Martins, D. Loureiro, M.C. Almeida, M.A. Silva, R. Brito, A.C. Soares

E4.1.1 Summary

Management of modern urban drinking water supply utilities requires measurement of volume and flow to obtain information to inform the decision-making process, including management of resources and supply conditions, and relations with stakeholders. The networks involve net balances based on large amounts of data relating to water inflows and outflows provided by equipment in many locations, together with hidden losses. This level of complexity makes uncertainty of measurement a valuable tool to support the analysis of performance and risk related to these utilities. This example provides a simple approach that allows this information to be obtained using the data already provided by regular measurements of volume.

E4.1.2 Introduction of the application

Clean water and sanitation are one of the 17 sustainable development goals (SDG) of the United Nations' 2030 agenda [200], being directly related to several other objectives, namely, economic growth, sustainable cities and communities, responsible consumption and production, and climate action. The growing demand for water increases the risk of scarcity and the need to promote water supply efficiency and the improvement of water management in our society [315].

Water utilities rely on water supply networks based on hydrological and hydraulic elements allowing the supply of water to households, industries, facilities, services and other users. Measurement of quantities like flow, volume, level and velocity are key to providing reliable information to management and to assure compliance with requirements of regulators and trust in trade [316]. The knowledge regarding measurement data reliability and associated uncertainty is a key issue for the management of water supply networks.

Modern urban water supply systems have complex distribution structures; problems exist due to lack of knowledge about the real performance of these systems, namely, related to the impact of hidden water losses [317–320]. The development of good measurement practices and the evaluation of uncertainty are needed to improve the quality of data and to increase confidence in results.

In these water supply systems, there are usually several locations enabling the inflow and outflow of water, the management of which is concerned with the water demand and the net balance of the system. For this purpose, equipment is placed in multiple measurement locations to provide measurement data for the volume net balance calculation, commonly using a three-stage process:

1. Data acquisition of flow or volume measurements with a fixed sampling rate, generating time series;
2. Data processing, allowing the totalized volume for a time interval to be obtained;
3. Calculation of the net balance of the water supply system involving sums and differences of the totalized volumes obtained for the several measurement locations.

In this process, the balance between inflows and outflows provides values of the quantity of interest (in this case, volume) commonly used in trade relations. To perform the assessment of the compliance and to make decisions, decision rules [6,321] need to be defined, thus requiring knowledge of the associated measurement uncertainties. For many water utilities, although the need for uncertainty evaluation is recognized as important, this task is still considered difficult, requiring support to apply the provisions of the GUM. The application of the GUM to the simple mathematical models used in this context enables simplified equations to be obtained that can be used in specific conditions of measurement (for example, measurement of constant flow, totalization of volume at a single measurement point, sums and differences obtained by combining branches of a network), allowing non-expert users to be supported by more straightforward approaches.

E4.1.3 Specification of the measurand(s)

The quantity of (liquid) volume is the measurand of interest in water supply networks, being used in the management of these systems and in trade relations between suppliers and users.

Volume can typically be obtained directly or indirectly. Direct measurement uses totalizer type flow meters that count cycles related to a reference volume of fluid passing through a compartment. These meters are commonly based on electromechanical transducers such as volume totalizers, oval gear totalizers, oscillating piston totalizers, lobed impeller totalizers and turbine totalizers.

Indirect measurement uses flowmeters able to evaluate the rate of a fluid flow (volumetric flow rate) during a time interval (since observations are a time-dependent phenomenon). They allow the volume to be calculated as the product of fluid flow and the time interval. Such measurement usually involves the use of flowmeters based on various techniques: vortex, swirl, ultrasonic, differential pressure, compact orifice, pitot, variable area and mass Coriolis.

The commonest type of direct measurement uses a turbine flowmeter (developed by R. Woltman, often being called a Woltman water meter). In this case, the measurement method makes the transduction of energy from a flowing fluid to the mechanical motion of a rotor causing it to

rotate with angular velocity proportional to the flow rate of the fluid [242]. The energy transfer is made from the mechanical motion to a counter or the generation of an electrical pulse signal (usually a sine wave), its frequency being related to a volumetric unit and the speed of rotation proportional to the flow rate. The mathematical model is:

$$\frac{\omega_r R}{\bar{v}} = \tan\alpha, \tag{E4.1.1}$$

where ω_r is the angular velocity of the blade, R is the radius of the blade, $\bar{v}$ is the average velocity of the fluid and α is the angle between the blade and the vertical axis of the rotor. Letting Q be the volume flow rate and A the effective area of the conduit, a relation between the volumetric flow rate and the angular speed is obtained:

$$\omega_r = \bar{v}\frac{\tan\alpha}{R} = \frac{Q}{A}\frac{\tan\alpha}{R} = kQ, \tag{E4.1.2}$$

where

$$k = \frac{\tan\alpha}{RA}.$$

Knowing that each pulse is related to a certain volume and combining this information with the frequency of the electrical signal, which is related to the flow rate, an estimate of volume for a time interval can be obtained.

Indirect methods enable the measurement of the volumetric flow rate, Q, defined as the volume of fluid that passes in a section per unit time (sometimes referred to as volume velocity using the symbol $\dot{V}$) and having $\mathrm{m}^3\,\mathrm{s}^{-1}$ as SI unit. Volumetric flow rate can be obtained using equation (E4.1.3), which relates the fluid flow velocity, v, and the cross-sectional vector area, A, where measurement takes place:

$$Q = vA. \tag{E4.1.3}$$

Considering the case of a closed conduit with circular geometry having an internal diameter, D, and a cross section orthogonal to the velocity vector 0°, equation (E4.1.3) becomes

$$Q = vA\cos\theta = v\frac{\pi D^2}{4}. \tag{E4.1.4}$$

Theoretically, to obtain the volume, V, from flow rate measurement, an integration over a time interval Δt should be made:

$$V = \int_{\Delta t} Q\mathrm{d}t. \tag{E4.1.5}$$

At the experimental level, flow rate is usually measured by sampling at constant time intervals, Δt, creating a discrete set of n values of a time series. Thus, equation (E4.1.5) can be approximated by the sum

$$V = \sum_{i=1}^{n} Q_i \Delta t, \tag{E4.1.6}$$

allowing the volume to be calculated at a certain physical location of the water supply system where measurement is taken during the time interval considered.

E4.1.4 Measurement model

The measurement model presented in equation (E4.1.6) can be applied to the two common cases of behaviour of flow in water supply systems [322]:

- flow rate showing similar cycles of water consumption, usually related to users' demands; figure E4.1.1 typifies the consumption of water measured during a time interval, sometimes making it possible to model (predictively) the system demands; and
- flow rate having two stages, fully open at a certain stationary flow or closed (being controlled by the provider or by the user), as illustrated in figure E4.1.2, during a time interval (e.g. filling a storage tank).

Figures E4.1.1 and E4.1.2[1] [322] illustrate the behaviour of experimental time series showing the two cases mentioned.

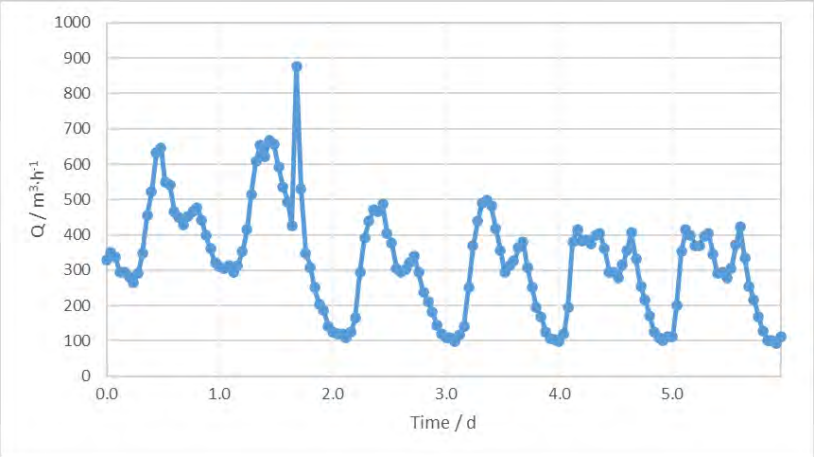


Figure E4.1.1: Input flow measurement experimental data obtained in a water distribution network with users and water losses during 6 days with time interval of sampling of 1 hour

The total volume, V_x, at a measurement location, x, obtained during a time interval, is given by the sum of the n single volume observations, $V_{x,i}$, measured in that interval:

$$V_x = \sum_{i=1}^{n} V_{x,i}. \tag{E4.1.7}$$

To estimate the net balance of a water supply system with various measurement locations for inflow and outflow having volumes $V_x^{(in)}$ and $V_x^{(out)}$, respectively, the net volume, V_{net}, is given by

$$V_{net} = \sum V_x^{(in)} - \sum V_x^{(out)}. \tag{E4.1.8}$$

Although the mathematical model is of a very simple nature, for the water industry the issue concerns the need to have tools that are easy to understand and simplify the calculation in order to reduce time spent in this process for large quantities of data obtained from multiple measurement locations.

[1] reproduced with permission

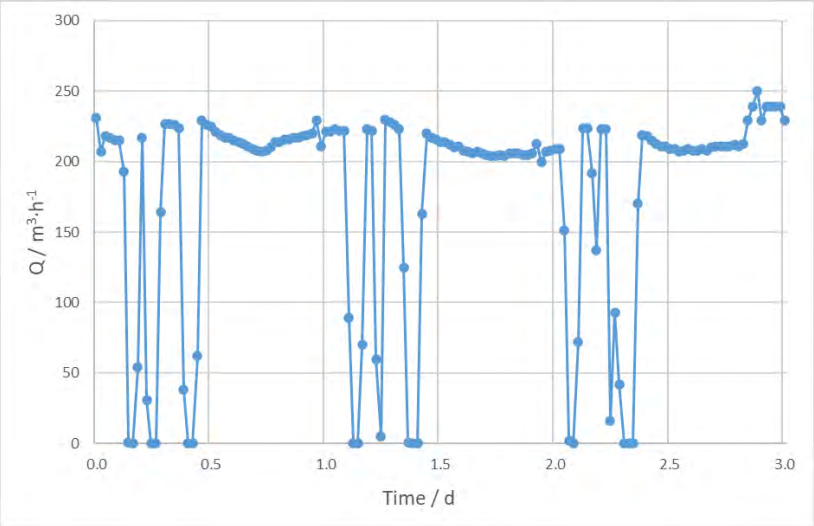


Figure E4.1.2: Flow measurement experimental data obtained at the entrance of a storage tank during 3 days with time interval of sampling of 30 minutes

E4.1.5 Uncertainty propagation

As starting point for the analysis of the problem, equation (E4.1.7) provides the functional relation used to obtain the output quantity required, the totalized volume (for a single location), V_x, given as a sum of discrete values measured during the time of acquisition. Flow measurement standard uncertainty is expressed in relative form (proportional to the measured values of the quantity), given by

$$w(V_i) = \frac{u(V_i)}{V_i}, \tag{E4.1.9}$$

requiring the relative standard uncertainty of each single volume measurement to be known. The standard measurement uncertainty of the totalized volume (at a single location) can be obtained by applying the LPU of the GUM [2] (not considering (auto)correlation):

$$u^2(V_x) = \left(\frac{\partial V_x}{\partial V_{x,1}}\right)^2 u^2\left(V_{x,1}\right) + \left(\frac{\partial V_x}{\partial V_{x,2}}\right)^2 u^2\left(V_{x,2}\right) + \cdots + \left(\frac{\partial V_x}{\partial V_{x,n}}\right)^2 u^2\left(V_{x,n}\right). \tag{E4.1.10}$$

Considering that the partial derivatives of V with respect to the V_i are in this case all equal to unity, equation (E4.1.10) becomes

$$u^2(V_x) = u^2\left(V_{x,1}\right) + u^2\left(V_{x,2}\right) + \cdots + u^2\left(V_{x,n}\right). \tag{E4.1.11}$$

Substitution of relation (E4.1.9) into (E4.1.11) yields an equation based on relative uncertainty contributions:

$$u^2(V_x) = w^2\left(V_{x,1}\right)V_{x,1}^2 + w^2\left(V_{x,2}\right)V_{x,2}^2 + \cdots + w^2\left(V_{x,n}\right)V_{x,n}^2. \tag{E4.1.12}$$

Some conditions usually met in water supply measurement systems allow the development of further simplifications of the calculations needed to obtain measurement uncertainty without requiring additional information. Three of these cases are presented:

– constant relative standard uncertainty in variable flow rate measurements,

– constant relative standard uncertainty in stationary flow rate, and

– sampling effect in stationary flow rate measurements.

For the first condition, considering the common case of having a relative standard uncertainty approximately constant for the measurement interval and variable flow rate:

$$w\left(V_{x,i}\right) = w\left(V_{x,1}\right) = w\left(V_{x,2}\right) = \cdots = w\left(V_{x,n}\right), \tag{E4.1.13}$$

which can be used in equation (E4.1.12) to obtain

$$u^2\left(V_x\right) = w^2\left(V_{x,i}\right)\left(V_{x,1}^2 + V_{x,2}^2 + \cdots + V_{x,n}^2\right). \tag{E4.1.14}$$

The use of this expression allows the evaluation of the standard uncertainty of the total volume based on the knowledge of the relative standard uncertainty common to the measurement range and the set of values measured over time. Furthermore, it facilitates the approach used to evaluate the total volume relative standard uncertainty, again using equation (E4.1.9) in the form

$$w\left(V_x\right) = \frac{u\left(V_x\right)}{V_x}. \tag{E4.1.15}$$

Consider an example with 10 measurements of volume (experimental data given in table E4.1.1), each with relative standard uncertainty of 2.0 %.

Table E4.1.1: Example of 10 measurements of volume obtained using a volumetric counter

$V_{x,1}$	$58\,m^3$	$V_{x,6}$	$61\,m^3$
$V_{x,2}$	$63\,m^3$	$V_{x,7}$	$52\,m^3$
$V_{x,3}$	$62\,m^3$	$V_{x,8}$	$57\,m^3$
$V_{x,4}$	$57\,m^3$	$V_{x,9}$	$69\,m^3$
$V_{x,5}$	$79\,m^3$	$V_{x,10}$	$76\,m^3$

With this information, using equations (E4.1.14) and (E4.1.15), the total volume standard uncertainty and relative uncertainty, respectively, would be evaluated for the total volume of $643\,m^3$:

$$u\left(V_x\right) = \sqrt{(0.02^2)\,(40\,878)}\,m^3 \approx 4.0\,m^3, \tag{E4.1.16}$$

$$w\left(V_x\right) \approx 0.64\,\%. \tag{E4.1.17}$$

The second condition yields cases having a relative standard uncertainty approximately constant for the measurement interval and stationary flow rate, which for measurements taken at the same frequency of acquisition implies the measurement of approximately the same volume for each reading. This is the case found when filling a storage tank (figure E4.1.2):

$$V_{x,i} = V_{x,1} = V_{x,2} = \cdots = V_{x,n}. \tag{E4.1.18}$$

This relation applied to equation (E4.1.12), which does not take into account autocorrelation, gives

$$u^2(V_x) = w^2\left(V_{x,i}\right)\left[nV_{x,i}^2\right], \tag{E4.1.19}$$

from which

$$u(V_x) = \sqrt{n}w\left(V_{x,i}\right)V_{x,i}, \tag{E4.1.20}$$

$$w(V_x) = \frac{\sqrt{n}w\left(V_{x,i}\right)V_{x,i}}{V_x}. \tag{E4.1.21}$$

These equations provide a relation between the value of the measurand and the number of measurement samples (n), the effect of which is explored in the next section.

The third condition yields cases having a relative standard uncertainty that is approximately constant for the measurement interval and stationary flow rate, being of interest in relating the measurement uncertainty evaluation to the decision regarding the sampling procedure.

The first rationale for the definition of a sampling interval should naturally be an interval adequate to allow observations of the expected phenomenon variability. However, in special cases like that mentioned with constant flow, intuitively one could think that it had no impact on the level of accuracy of the method.

Looking at equations (E4.1.20) and (E4.1.21), applicable to obtaining the standard uncertainty and relative standard uncertainty of the total volume, respectively, it becomes clear that results depend on the number of samples, n, and on the value obtained of each observation of the volume, Vi. For a certain total volume V that is fixed, when n grows, the single observation of volume decreases proportionally:

$$\widehat{V}_t = \frac{V}{n}. \tag{E4.1.22}$$

Using this relation in equation (E4.1.21), a simplified relation is obtained showing that the relative standard uncertainty of the total volume decreases with an increasing number of samples, n :

$$w(V) = \frac{w(V_i)}{\sqrt{n}}. \tag{E4.1.23}$$

Consider a simple example, having a relative standard uncertainty of 2 %, and 10 observations each of 100 m^3, or five observations of 200 m^3, with total volume in both cases of 1000 m^3. Applying equation (E4.1.23) the results are, for the data series of 10 values,

$$w(V) = \frac{w(V_i)}{\sqrt{n}} = 0.63\,\%, \tag{E4.1.24}$$

and, for the data series of 5 values,

$$w(V) = \frac{w(V_i)}{\sqrt{n}} = 0.89\,\%. \tag{E4.1.25}$$

Having these results for a single location, the development of an approach applied to a water supply system network in which the water volume net balance is the main output of the evaluation becomes the most relevant outcome of this example.

Most of water supply infrastructures are part of services provided for trade of this resource, in which it is required to make a net balance of inflow and outflow of water volume in the system and in a set period. The approach usually taken is based on measuring the volume at different locations and the use of sums and differences to obtain information needed in the economic process. In many cases, it also makes it possible to identify water losses and to evaluate the efficiency of the system. A functional relation to characterise the net balance is given by

$$V_{\text{net}} = \sum_{i=1}^{l} V_i - \sum_{j=1}^{m} \widetilde{V}_j + \delta V_{\text{loss}}, \tag{E4.1.26}$$

where V_i represents the n measuring locations of inflow of water into the system, $\widetilde{V}_i$ represents the m measuring locations of water outflow of water in the system, and δV_{loss} the water losses during the transfer process.

The LPU presented in the GUM for a functional relation of the type, $Y = f\left(X_1, \cdots, X_n\right)$, for linear models, provides an exact solution given by:

$$u^2(y) = \sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i} \right)^2 u^2\left(x_i\right) + 2 \sum_{j=1}^{n-1} \sum_{k=j+1}^{n} \frac{\partial f}{\partial x_j} \frac{\partial f}{\partial x_k} u(x_j) u(x_k) r(x_j, x_k). \tag{E4.1.27}$$

It includes, as the first sum, contributions related to the diagonal elements of the covariance matrix and, as the second sum, contributions related to the correlation of input quantities, where $r\left(x_j, x_k\right)$ is the correlation associated with x_j and x_k

Applying the LPU to equation (E4.1.26), neglecting the correlation contributions and noticing that the partial derivatives are all equal to one, the estimate of the net balance uncertainty (without correlation), $u\left(V_{\text{net,nc}}\right)$, is given by

$$u^2\left(V_{\text{net.nc}}\right) = \sum_{i=1}^{l} u^2\left(V_i\right) + \sum_{j=1}^{m} u^2\left(\widetilde{V}_j\right) + u^2\left(\delta V_{\text{loss}}\right). \tag{E4.1.28}$$

Considering the existence of a similar standard uncertainty, $u(V)$, for the inflow and outflow measurement locations, and neglecting the contribution related to the quantity lost, a simplified equation is obtained:

$$u\left(V_{\text{net,nc}}\right) \approx \sqrt{(l+m)} u(V). \tag{E4.1.29}$$

The nature of the net balance model, however, requires the consideration of correlation between the inflow and outflow contributions. The evaluation of the correlation of measured values obtained from time-series at different locations of the net is a difficult task because of the dynamics of the system. At any instant of time the flow at a measuring point a of the net is not equivalent to the flow at another point b. Thus, a matching algorithm is applied to the time-series that is able to evaluate time delays based on flow and distance between measurement locations in the net branches.

Given the formula to obtain the net-balance standard uncertainty without correlation, in order to estimate the limits of the uncertainty in the presence of correlation, another approach is presented considering the assumption of maximum correlation between input quantities. To implement this approach, it is needed to calculate the contribution of the term

$$2 \sum_{j=1}^{n-1} \sum_{k=2}^{n} \frac{\partial f}{\partial x_j} \frac{\partial f}{\partial x_k} u(x_j) u(x_k) r(x_j, x_k). \tag{E4.1.30}$$

Again, the same standard uncertainty, $u(V)$, for all input quantities is considered and it is assumed that the correlation coefficient is maximum, $r(x_i, x_k) = 1$, for the inflow and outflow measurement locations, allowing equation (E4.1.30) to be simplified:

$$2\sum_{j=1}^{n-1}\sum_{k=2}^{n}\frac{\partial f}{\partial x_j}\frac{\partial f}{\partial x_k}u^2(V). \tag{E4.1.31}$$

Regarding the partial derivatives, for the l inflow input quantities, $\frac{\partial f}{\partial V_i} = 1$; and for the m outflow input quantities, $\frac{\partial f}{\partial \widetilde{V}_j} = -1$. These conditions imply that the will be positive and negative unit contributions from correlated terms due to multiplying the partial derivatives. In this case, there are three types of combinations:

a) Product of partial derivatives of inflow input quantities, l, being the result of the product equal to 1

b) Product of partial derivatives of outflow input quantities, m, being the result of the product equal to 1

c) Product of partial derivatives of inflow and outflow input quantities, being the result of the product equal to -1

The number of combinations without repetition in cases a) and b) can be evaluated using the general expression for choosing r members from a set of n members, $\binom{n}{r} = \frac{n!}{r!(n-r)!}$. In these cases, r is equal to 2, and the general expression to apply is

$$\binom{n}{2} = \frac{1}{2}n(n-1), \tag{E4.1.32}$$

which gives, for combinations multiplied by 2 (factor in equation (E4.1.31)), for the case a for case b), $m(m-1)$. For case c) there are $2(lm)$ possible combinations.

Combining these results, expression (E4.1.31) becomes,

$$[l(l-1)+m(m-1)-2(lm)]u^2(V) = \left[(l-m)^2 - l - m\right]u^2(V), \tag{E4.1.33}$$

which gives a final expression for the net-balance standard uncertainty, considering that input quantities are taken as fully correlated,

$$u\left(V_{\text{net,fc}}\right) = \left[(l+m)+(l-m)^2 - l - m\right]^{\frac{1}{2}}u(V) = |l-m|u(V). \tag{E4.1.34}$$

Equations (E4.1.29) and (E4.1.34) provide the uncertainty evaluation considering the two extreme possibilities for the net balance uncertainty (input quantities are not correlated or are fully correlated, respectively). Therefore, for each net system, it is possible to obtain extreme limits for the uncertainty.

Taking a simple example of a network with $l = 6$ inflow contributions and $m = 4$ outflow contributions, using equations (E4.1.29) and (E4.1.34), the results would be:

$$u\left(V_{\text{net,nc}}\right) = \sqrt{l+m}\,u(V) = 3.2u(V), \tag{E4.1.35}$$

$$u\left(V_{\text{net,fc}}\right) = |l-m|u(V) = 2u(V), \tag{E4.1.36}$$

since, for positive l and m

$$|l-m| < \sqrt{l+m},$$

correlation contributions reduce the uncertainty, which allows $u\left(V_{\text{net,nc}}\right)$ to be viewed as an upper bound for the net-balance uncertainty. The case $l = m$ gives $u\left(V_{\text{net,fc}}\right) = 0$.

E4.1.6 Reporting the result

For the present example, three cases of behaviour of the experimental data related to the measurement of flow rate to obtain the quantity volume were considered:

- constant relative standard uncertainty in variable flow rate measurements;
- constant relative standard uncertainty in stationary flow rate; and
- sampling effect in stationary flow rate measurements.

Furthermore, it was assumed that the relative standard uncertainties of the individual measurements were approximately equal, using the knowledge of the authors of typical measuring systems.

Based on this approach to the problem, equations were obtained allowing the calculation of the standard uncertainty related to the volume measured for a time interval.

For the first case, considering a single location, x, and the condition of having a relative standard uncertainty approximately constant for the measurement interval and variable flow rate, the simplified equation (E4.1.14) was given:

$$u^2(V_x) = w^2\left(V_{x,i}\right)\left[V_{x,1}^2 + V_{x,2}^2 + \cdots + V_{x,n}^2\right].$$

For the second case, considering a relative standard uncertainty approximately constant for the measurement interval and stationary flow rate, the simplified equation (E4.1.19) was given:

$$u^2(V_x) = w^2\left(V_{x,i}\right)\left[nV_{x,i}^2\right].$$

The third case was intended to explore the impact of sampling in the calculation of uncertainty, allowing to obtain a relation (E4.1.20) between the measurand and the number of measurement samples (n)

$$u(V_{\mathrm{x}}) = \sqrt{n}w\left(V_{\mathrm{x},i}\right)V_{\mathrm{x},i}.$$

Finally, considering the water volume net balance in an infrastructure based on equation (26),

$$V_{\mathrm{net}} = \sum_{i=1}^{l} V_i - \sum_{j=1}^{m} \widetilde{V}_j + \delta V_{\mathrm{loss}},$$

where V_i represents the l measuring locations of inflow of water into the system, $\widetilde{V}_i$ represents the m measuring locations of water outflow of water from the system, and δV_{loss} the water losses during the transfer process.

Considering the existence of a similar standard uncertainty, $u(V)$, for the inflow and outflow measurement locations, and neglecting the contribution related to the quantity lost, simplified equations are obtained for the net balance uncertainty without correlation effects (nc) and with full correlation (fc) giving the bounds for expected values:

$$u\left(V_{\mathrm{net,nc}}\right) \approx \sqrt{l+m}\,u(V),$$
$$u\left(V_{\mathrm{net,fc}}\right) \approx |l-m|u(V),$$

showing that net-balance standard uncertainty not including the effect of correlation gives an upper bound, and that the increase of the number of locations used in the net balance will increase the measurement standard uncertainty of the estimated net volume.

E4.1.7 Interpretation of results

Observation of time-varying phenomena but with conditions of a cyclic or permanent nature is common in many branches of science and is known in particular in the context of hydrology.

Given the usual complexity of metering systems in infrastructures such as those promoting the management of resources such as water supply, it is very useful to find solutions that can simplify the application of uncertainty calculation considering, on one hand, the importance of integrating this component into the instrumentation metrological management process and, on the other hand, by the relevance that this information may have in the decision making process.

Considering this context, the main objective of this study is to demonstrate how it is possible to develop mechanisms that allow the calculation of measurement uncertainties in a simplified way in these processes. Moreover, the study demonstrates that under some circumstances it is possible to perform the calculation without the need to apply sophisticated operations such as the determination of partial derivatives. Even if these derivatives can be obtained using simple numerical operations, they often lead to difficulties in the industrial context.

In addition to the analysis presented within the scope of volume observation at a specific location, the infrastructure management process makes decisions based on the water volume net balance that results from the combination of information of inflow and outflow at different system locations. The approach presented here makes it possible to relate the measurement standard uncertainty to the number of measurement sites considered, establishing a relationship with complementary information of interest for the design and development component of the measuring systems that are included in the life cycle of these infrastructures.

If actual correlation information were available such as common effects relating to the flowmeters used, that information should be taken into consideration in providing more valid uncertainties associated with estimates of water volume net balance.

Example E4.2

Uncertainty of the orifice-plate discharge coefficient

M. Reader-Harris, C. Forsyth and T. Boussouara

E4.2.1 Summary

The uncertainty of the orifice-plate discharge coefficient given by the Reader-Harris/Gallagher (1998) Equation has been calculated taking account of the uncertainty of the data on which it is based and of the variability in manufacture permitted by ISO 5167-2. Although using the correct method to determine the uncertainty in ISO 5167-2 has made an insignificant difference to the value given in the standard, in other similar cases where the uncertainty for an artefact is based on data from other similar artefacts the uncertainty values obtained by the correct method may be significantly different from those by the incorrect method.

E4.2.2 Introduction to the application

It is estimated that at least 40 % of industrial flow meters in use at present are differential-pressure-based devices, with the orifice plate being the most popular for accurate measurement of fluid flow. These devices have been described in standards for many years: ISO 5167 [323,324] now provides the most wide-ranging of these standards. ISO 5167 is used in the measurement of many different fluids across a wide range of applications; nevertheless, the measurement of natural gas is highest in terms of financial importance: about £16 billion (€20 billion) of natural gas each year in the UK is measured by devices conforming to this standard [325].

A diagram of an orifice meter is provided in figure E4.2.1. Differential pressure can be measured using one of three permissible tapping pairs: corner tappings (in the corners), D and $D/2$ tappings (one diameter upstream and half a diameter downstream of the orifice plate) and flange tappings (one inch (25.4 mm) both upstream and downstream of the orifice). Flange tappings are the most convenient and most common. Data taken with Corner 2 tappings (a different method of inserting corner tappings) have not been analysed in this report.

NOTE ‘Orifice plate’ can refer just to the plate or to the whole meter; where it is important to be clear that the plate and pipework are meant, ‘orifice meter’ can be used.

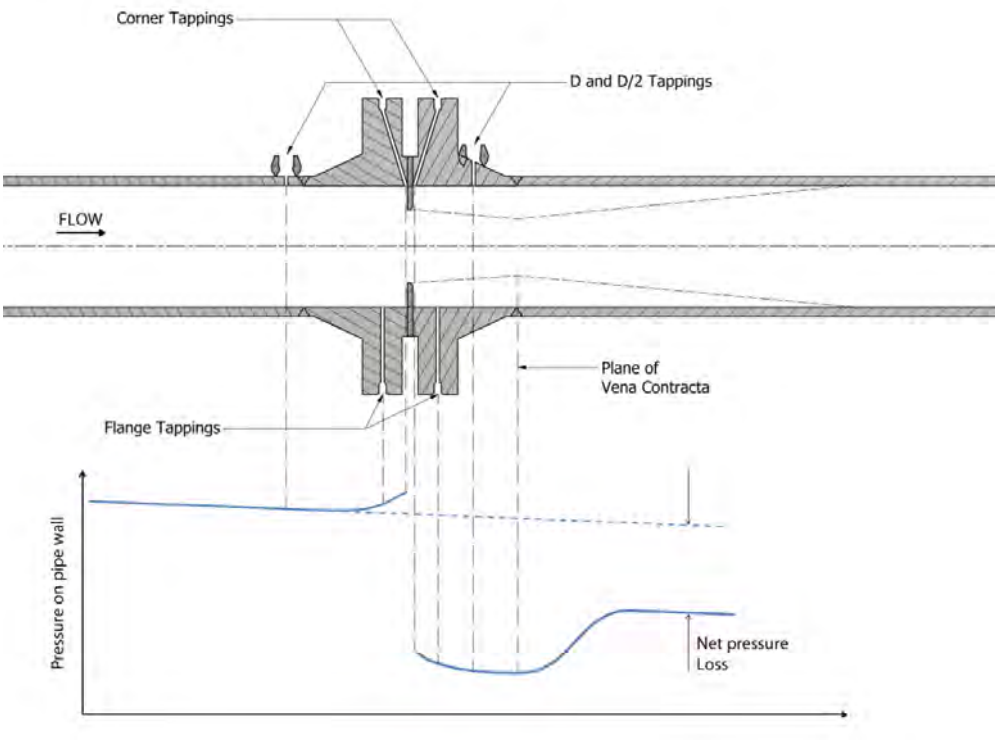


Figure E4.2.1: The approximate pressure profile through an orifice meter

E4.2.3 Specification of the measurand(s)

The measurand is the mass flow rate.

E4.2.4 Measurement model

E4.2.4.1 Fundamental equation

The attraction of using an orifice meter is that it is very rarely necessary to calibrate it in a flowing fluid. It is usual to make it in accordance with the pattern in ISO 5167 and then calculate the mass flowrate from equation (E4.2.1) using measurements made in the factory, measurements made in the field and equations in ISO 5167:

$$q_m = \frac{C\varepsilon}{\sqrt{1-\beta^4}} \frac{\pi d^2}{4} \sqrt{2\rho_1 \Delta p}, \tag{E4.2.1}$$

where q_m is the mass flow rate, C is the discharge coefficient (given in ISO 5167-2 and discussed in section E4.2.4.2), ε is the expansibility factor (given in ISO 5167-2), ρ_1 is the density at the upstream tapping, Δp is the differential pressure, and β is the diameter ratio, equal to d/D, where d and D are the orifice and pipe diameters.

E4.2.4.2 Reader-Harris/Gallagher (1998) equation

Using ISO 5167-2:2003, the discharge coefficient is given by the Reader-Harris/Gallagher (1998) Equation [326]:

$$
\begin{aligned}
C = {} & 0.5961 + 0.0261\beta^2 - 0.216\beta^8 && C_\infty \text{ term ,} \\
& + 0.000521\left(10^6\beta/Re_D\right)^{0.7} + (0.0188 + 0.0063A)\beta^{3.5}\left(10^6/Re_D\right)^{0.3} && \text{slope term ,} \\
& + \left(0.043 + 0.080e^{-10L_1} - 0.123e^{-7L_1}\right)(1 - 0.11A)\frac{\beta^4}{1-\beta^4} && \text{upstream tapping term ,} \\
& - 0.031\left(M_2' - 0.8M_2'^{1.1}\right)\beta^{1.3} && \text{downstream tapping term.}
\end{aligned}
\tag{E4.2.2a}
$$

Where $D < 71.12\,\text{mm}\ \left(2.8''\right)$ the following term should be added to equation (E4.2.2a):

$$
+0.011(0.75 - \beta)\left(2.8 - \frac{D}{25.4}\right) \qquad \text{small pipe diameter term .} \tag{E4.2.2b}
$$

where D is the pipe diameter in mm.

In equation (E4.2.2a),

$$A = \left(\frac{19000\beta}{Re_D}\right)^{0.8} \quad \text{and} \quad M_2' = \frac{2L_2'}{1-\beta}.$$

The Reynolds number, Re_D, can be obtained from

$$Re_D = \frac{4q_m}{\pi\mu_1 D},$$

where μ_1 is the dynamic viscosity at the upstream tapping. L_1 is the quotient of the distance of the upstream tapping from the upstream face of the plate and L_2' is the quotient of the distance of the downstream tapping from the downstream face of the plate.

Equations (E4.2.2a) and (E4.2.2b) are based on the EEC/API database, which is described in pages 131 to 133 of [327]. These equations represent model fits to the data, the model being a hybrid based on the physics and some empirical terms. Where flange tappings are used it is required that both $Re_D \geq 5000$ and $Re_D \geq 170\beta^2 D$, where D is expressed in millimetres.

To achieve this discharge coefficient and its associated uncertainty there are many specific requirements in ISO 5167-2 [324] on the dimensions and geometry of both the orifice plate (depicted in figure E4.2.2) and the upstream and downstream pipework. There is a picture of one of the original EEC plates in Figure E4.2.3.

The uncertainty in the discharge coefficient is often the largest component of the uncertainty in the flowrate and thus of great importance for fair taxation and custody transfer.

E4.2.4.3 Calculating the uncertainty

It is not simple to calculate the uncertainty for the orifice discharge-coefficient equation. Evaluating the uncertainty for that equation is different from most evaluations of uncertainty: a normal evaluation of uncertainty is for an instrument which is calibrated and itself subsequently used. However, the orifice discharge-coefficient equation is used for plates other than those used to collect the data.

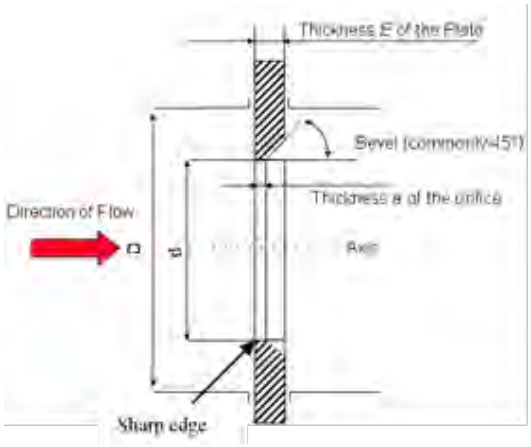


Figure E4.2.2: Diagram of an orifice plate

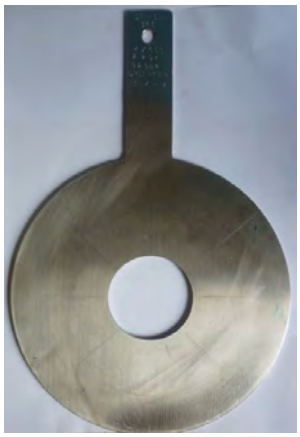


Figure E4.2.3: An EEC orifice plate

Because the standard deviation in the EEC/API database does not necessarily reflect the variability permitted by the standard it is not appropriate to evaluate the expanded uncertainty, corresponding to 95 % coverage, of the orifice discharge-coefficient equation as twice the standard deviation of the data in the EEC/API database about the equation, which is the method used by API 14.3:1990 [328].

Rather than simply taking twice the standard deviation it is desirable first to reduce as much as possible the effects of random errors in the database (see section E4.2.4.4), then to determine what the uncertainty would be for an orifice plate of average edge sharpness in a pipe of average roughness, taking into account the uncertainty of the original data as well as the spread of the data about the equation. However, there is variability within prescribed limits permitted in the manufacture of orifice meters: so the uncertainty of an orifice meter in accordance with ISO 5167-2 should have additional uncertainty incorporated. In this way the uncertainty of the Reader-Harris/Gallagher (1998) Equation can be calculated.

NOTE: It is possible to look at the deviations between orifice plates that were not used to derive the equation and the equation itself (Appendix E4.2.A), but there is less information about these plates and the uncertainty of their calibration, and the permitted variability should still be included.

E4.2.4.4 Reducing the effects of random errors in the database

For each orifice meter (considering an orifice plate with three different sets of tappings as three orifice meters) using a single point from each laboratory means that the points are independent and can be combined as in E4.2.4.6. Moreover, this has the advantage described in this section.

Twice the standard deviation of the data in the database about the Reader-Harris/Gallagher (1998) Equation includes the effects of random errors in the original sets of data collected with equipment whose random scatter may be higher than that available today. To avoid the uncertainty of the equation being increased by the uncertainty due to random errors in each set of data in the database it is possible to fit each set of data in the database as a linear function of $Re_D^{-0.5}$ and then represent each line by a single point at the middle of its range. Such fitted points have reduced uncertainty due to random errors. Figure E4.2.4 (discussed below) illustrates this well.

Since uncertainties increase for large β, small Re_d ($= Re_D/\beta$), small d, and large $\Delta p/p_1$ (where p_1 is the static pressure at the upstream tapping) the data for $0.19 < \beta < 0.67$, $Re_d > 150000\beta$, $d > 50\,\text{mm}$ and, if gas, $\Delta p/p_1 \leq 0.04$ were analysed as a core population (NIST Gaithersburg

data with Δp less than 180 mbar were excluded: see E4.2.4.6). Data taken to determine installation effects were also excluded). For this core population of 7031 points the standard deviation, s, about the equation was determined: $2s = 0.41\,\%$; there are 339 data sets making up the core population: for the 339 mid-range calculated points $2s = 0.38\,\%$. To illustrate the difference in deviations Figure E4.2.4 shows one set of 24 water points from the database. Although twice the r.m.s. (root mean square) deviation of the (24) points about the Reader-Harris/Gallagher (1998) Equation is 0.28 %, the (expanded) uncertainty due to random effects of the fitted point is 0.06 %. The deviation of the fitted point from the Reader-Harris/Gallagher (1998) Equation and the mean deviation of the 24 points from the Reader-Harris/Gallagher (1998) Equation are both equal to −0.03 %.

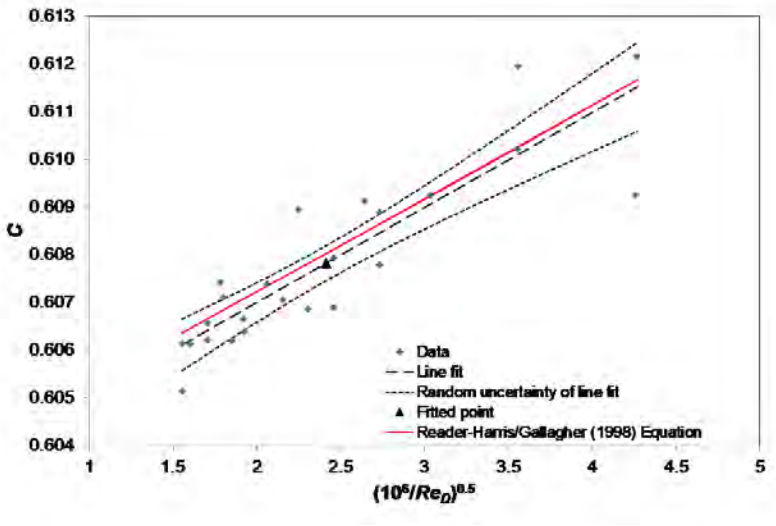


Figure E4.2.4: One set of NEL 4″ (100 mm) $\beta = 0.57$ water data from the orifice-plate discharge-coefficient database

In figure E4.2.4 the fitted straight line is almost exactly parallel to the Reader-Harris/Gallagher (1998) Equation; i.e. the errors vary very little with Reynolds number (have nearly zero gradient). At the fitted point the fitted discharge coefficient is equal to the average value of the measured discharge coefficients. The difference between taking the average of the deviations and calculating the deviation at the fitted point is less than 0.001 %. The uncertainty of the fitted line at the fitted point has not been included, since the fitted point will not have a larger uncertainty than the individual points do.

NOTE 30 years after the data were collected it would be hard to re-evaluate the uncertainties; the average deviations from different laboratories are generally consistent even when it is assumed that the deviations are not a function of Reynolds number (see E4.2.4.6).

E4.2.4.5 Errors if data had average pipe roughness and edge sharpness

The database contains points with a range of values of relative pipe roughness, R_a/D, and relative edge sharpness, r_e/d: the fitted equation has effectively been derived for a particular relative pipe roughness (as a function of Reynolds number) and relative edge sharpness (based on the relative pipe roughness and relative edge sharpness of the original data); if the actual orifice meter to

be manufactured for a user of ISO 5167-2 had the same relative pipe roughness and the same relative edge sharpness as those on which the fitted equation is based the uncertainty could be considerably reduced.

The Reader-Harris/Gallagher (1998) Equation is associated with the relative pipe roughness (here expressed as k/D) and friction factor λ shown in Table 9 of ISO/TR 12767 [329], given here as table E4.2.1.

Table E4.2.1: Values of k/D and λ associated with the Reader-Harris/Gallagher (1998) Equation

Re_D	1×10^4	3×10^4	1×10^5	3×10^5	1×10^6	3×10^6	1×10^7	3×10^7	1×10^8
$1 \times 10^4 k/D$	1.75	1.45	1.15	0.9	0.7	0.55	0.45	0.35	0.25
λ	0.031	0.024	0.0185	0.0155	0.013	0.0115	0.0105	0.01	0.0095

The relative edge sharpness associated with the Reader-Harris/Gallagher (1998) Equation can be calculated from the measurements of edge sharpness for each plate. The average reduces as the Reynolds number increases: see figure E4.2.5, on which the following equation is plotted:

$$1000\frac{r_e}{d} = 0.6343e^{-0.139 \lg Re_D}. \tag{E4.2.3}$$

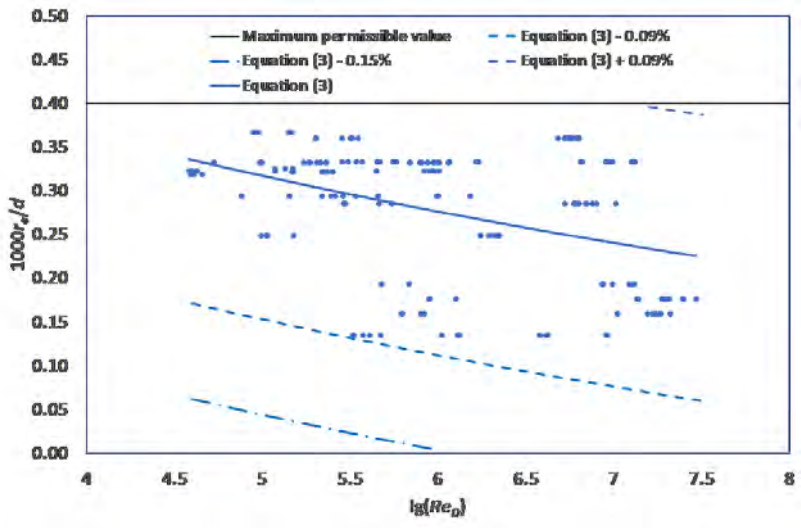


Figure E4.2.5: Edge sharpness (r_e) relative to the orifice diameter for each set of EEC data: values are plotted at the pipe Reynolds number of the fitted point for the set; equation (3) in the figure legend refers to equation (E4.2.3) in the text.

Taking the fitted points in the database, calculating (using the calculations in E4.2.4.7 and E4.2.4.7) what the value would have been with the relative pipe roughness in table E4.2.1 and the edge sharpness in equation (E4.2.3) (where there are multiple points taken in the same laboratory taking averages to include all the points) gives the values that are included in subsequent sections.

To the API data no correction was made for edge sharpness.

E4.2.4.6 Calculating the deviations taking account of the uncertainty of the laboratories

It is necessary also to consider the uncertainty contributed by the original laboratories: on the one hand the uncertainty of the average of the data from the independent laboratories should be added to the uncertainty of the Equation; on the other hand the scatter contributed by calibration in a range of independent laboratories should be removed from the uncertainty of the Equation.

Almost all the European orifice plates were calibrated in multiple laboratories. The errors for the orifice plates in 24″ (600 mm) pipe are shown in figure E4.2.6.

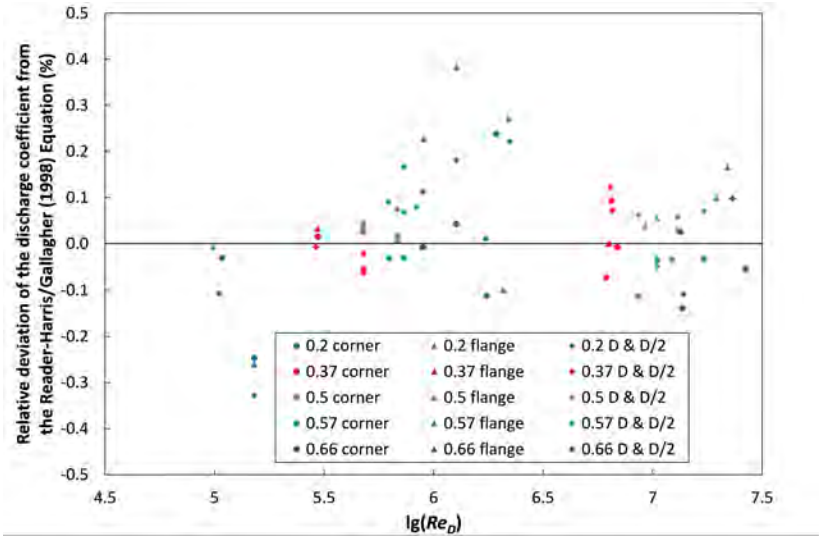


Figure E4.2.6: Discharge coefficient deviation from the Reader-Harris/Gallagher (1998) Equation for the EEC 24” (600 mm) data adjusted to average pipe roughness and edge sharpness (in this key and those for figures E4.2.7 and E4.2.8 the value of β and the pressure tappings used are given)

In figure E4.2.6 for $\beta = 0.37, 0.5$ and 0.57 the deviations are consistent and have little variation with Reynolds number. It is less obvious whether the same is true for $\beta = 0.2$ and 0.66. Testing the data with the χ^2 test in section 5.3 of Cox [70] gives $\chi^2_{\text{obs}} < 1.01$ for $\beta = 0.37, 0.5$ and 0.57, $\chi^2_{\text{obs}} < 3.34$ for $\beta = 0.66$, and $8.8 < \chi^2_{\text{obs}} < 11.4$ for $\beta = 0.2$. The consistency check in Cox is met for all β except 0.2. Removing either the highest set or the lowest one for $\beta = 0.2$ gives consistency: removing the lowest one gave slightly lower χ^2_{obs}.

A similar test is applied to the EEC 10″ (250 mm) data: see figure E4.2.7. The results were found to be consistent for all pairs of tappings (except for a plate that was subsequently replaced. Only the data from the replacement plate are included in figure E4.2.7 or in subsequent calculations). ‘0.57*’ refers to a second orifice plate with $\beta = 0.57$.

The fact that the data are consistent in the 10″(250 mm) and 24″(600 mm) sets suggests that the claimed uncertainties are not too small (at least after some of the uncertainty due to random effects has been removed).

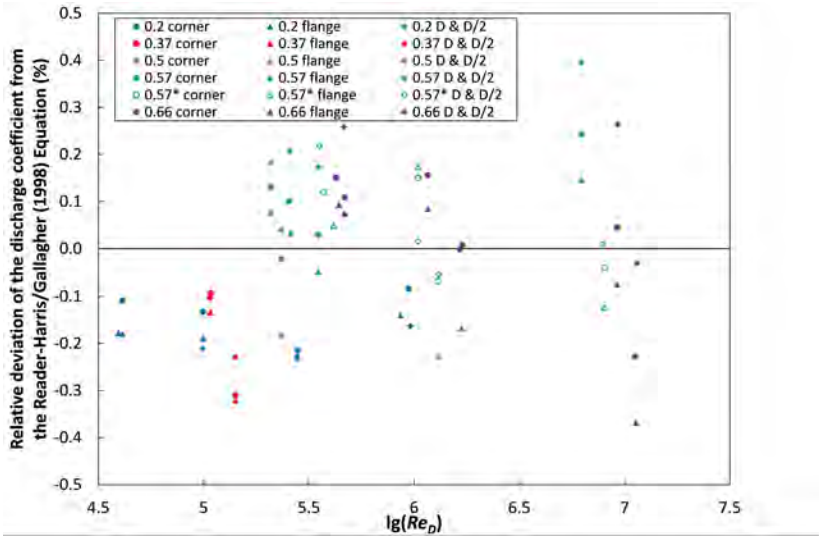


Figure E4.2.7: Discharge coefficient deviation from the Reader-Harris/Gallagher (1998) Equation for the EEC 10″ (250 mm) data adjusted to average pipe roughness and edge sharpness

A similar test is applied to the EEC 4″ (100 mm) data: see figure E4.2.8. The results are only found to be consistent for all three tappings once the second set of three points from the right was removed for both $\beta = 0.57$ and $\beta = 0.66$ and the left hand set of three points removed for $\beta = 0.57$.

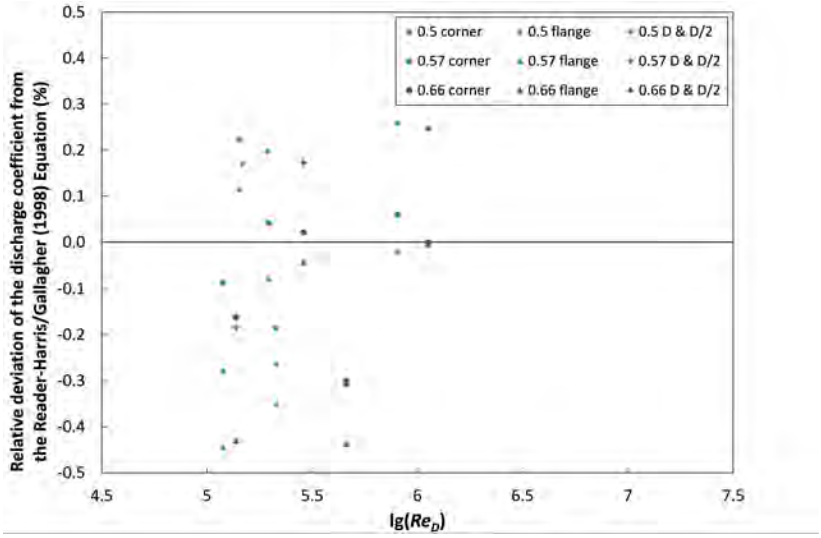


Figure E4.2.8: Discharge coefficient deviation from the Reader-Harris/Gallagher (1998) Equation for the EEC 4” (100 mm) data adjusted to average pipe roughness and edge sharpness

The 4″ (100 mm) data that were collected to study installation effects have been omitted.

The NIST Gaithersburg data have an uncertainty that depends very strongly on the differential pressure. By limiting the data to those with differential pressure above 180 mbar each of the averaged data has an uncertainty within 0.01 % of 0.135 %.

If three points or fewer for a single plate from a laboratory met the criteria in E4.2.4.4 the averaged point was not used.

The weighted average deviations for the consistent sets are shown in figure E4.2.9. Each point represents a different orifice meter (considering an orifice plate with three different sets of tappings as three orifice meters). The weighted averages were obtained using the method in section 5.1 of Cox [70].

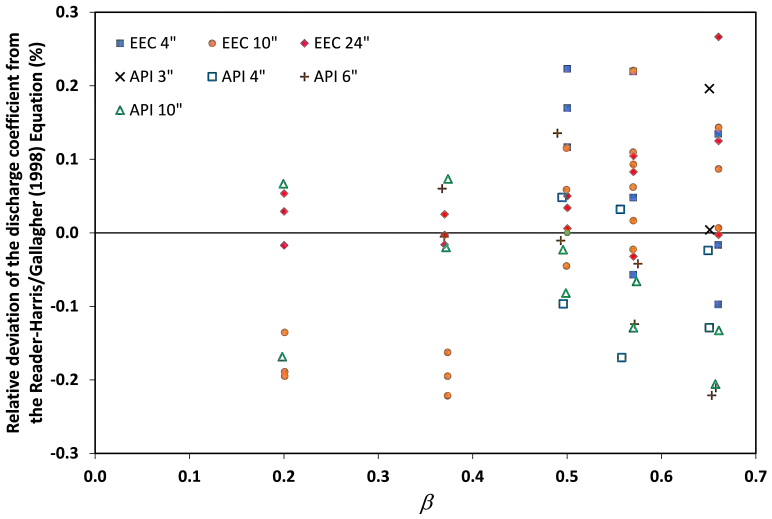


Figure E4.2.9: Weighted average deviations for the consistent sets

In support of the idea that a plate has a single value of deviation some modern NEL data are provided in Appendix E4.2.B: the data are roughly parallel to the Reader-Harris/Gallagher (1998) Equation. Moreover, in figures E4.2.6 to E4.2.8 roughly the same number of deviations have a positive slope against the Reynolds number as have a negative slope. Even if it were not accepted that the deviation can be considered constant then for each plate the weighted mean deviation would apply for an intermediate value of the Reynolds number.

The deviations of the points in figure E4.2.9 are summarized in table E4.2.2.

Table E4.2.2: Deviations of the fitted and adjusted points from the Reader-Harris/Gallagher (1998) Equation

	$\beta < 0.6$	$\beta \approx 0.66$
Mean deviation (%)	0.001	−0.004
2 x r.m.s. deviation (%)	0.220	0.288
Expanded uncertainty of points	0.141	0.130
Square root of the sums of the squares of the two lines above (%)	0.261	0.315

The expanded uncertainty of the points was obtained by evaluating the uncertainty for each point in figure E4.2.9 using the method in section 5.2 of Cox [70] and then taking the r.m.s. value of all these uncertainties. The expanded uncertainty of the individual points in figure E4.2.9 varied from 0.11 % to 0.16 %.

The API data (with no averaging between facilities) give roughly the same deviations as the EEC data when all the fitted and adjusted data points in each set are considered.

E4.2.4.7 Variability permitted by ISO 5167

Introduction

What is required as the uncertainty of the discharge-coefficient equation is the uncertainty of an orifice meter with the variability permitted by the standard.

Pipe roughness

The effect of pipe roughness on discharge coefficient is given in figure E4.2.10.

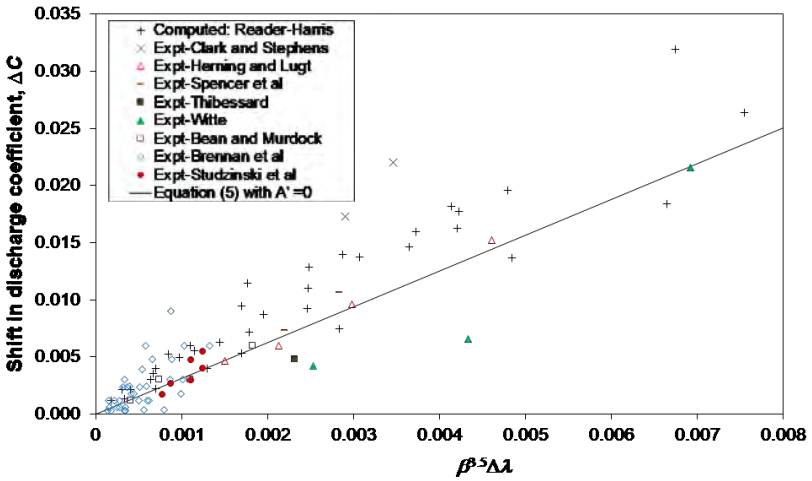


Figure E4.2.10: The effect of rough pipe on the orifice plate discharge coefficient; equation (5) in the figure legend refers to equation (E4.2.5).

Figure E4.2.10 gives measured and computed (using CFD (Computational fluid dynamics)) values of ΔC as a function of $\beta^{3.5}\Delta\lambda$, where λ is the friction factor given by

$$\Delta p = \frac{\lambda \rho \bar{u}^2 x}{2D}, \tag{E4.2.4}$$

where Δp is the difference in pressure between two tappings spaced a distance x apart in a pipe of diameter D and $\bar{u}$ is the mean velocity in the pipe.

References to the computational and experimental data are given in Reader-Harris [327], which also includes the derivation of equation (E4.2.5) :

$$\Delta C_{\text{rough}} = \left(3.134 + 4.726A'\right)\beta^{3.5}\Delta\lambda, \tag{E4.2.5}$$

where ΔC_{rough} is the change in discharge coefficient due to a change in friction factor $\Delta\lambda$, and A' is given by

$$A' = \left(\frac{2100\beta}{Re_D}\right)^{0.9}. \tag{E4.2.6}$$

P, the maximum permissible percentage shift in C, is given in table E4.2.3.

Table E4.2.3: Maximum permissible percentage shift, P, in C due to pipe roughness

β	P	U
$0.1 \leq \beta \leq 0.2$	0.5β	$0.7-\beta$
$0.2 \leq \beta \leq 0.5$	0.5β	0.5
$0.5 < \beta \leq 0.6$	0.25	0.5
$0.6 < \beta \leq 0.71$	$0.5(1.667\beta-0.5)$	$1.667\beta-0.5$
$0.71 < \beta$	$1.13\beta^{3.5}$	$1.667\beta-0.5$

U, the stated percentage uncertainty of the Reader-Harris/Gallagher (1998) Equation in ISO 5167-2, is also given in table E4.2.3, for P was determined to ensure that for $\beta \leq 0.5$, where other sources of error are dominant, $P/U \leq \beta$; for $0.5 < \beta \leq 0.71$ $P/U = 0.5$; for $0.71 < \beta$ P/U increases from 0.5 at $\beta = 0.71$ to 0.55 at $\beta \leq 0.75$.

NOTE It can be checked that this value of P is correct, since it, together with the value of λ associated with the data in the database (and thus with the Reader-Harris/Gallagher (1998) Equation) in Table E4.2.1, and equation E4.2.5 to calculate the effect of $\Delta\lambda$ on C, leads to the maximum and minimum values of pipe roughness for pipes upstream of orifice plates given in Tables E4.2.1 and E4.2.2 of ISO 5167-2 (except that the maximum value of R_a/D has been set as 1.5×10^{-3}; so in some cases the maximum deviation is smaller than P above).

Installation effects

The requirement for upstream lengths used by ISO 5167-2 was put forward by API: all the installation-effects data of sufficient quality are brought together; at least two high-quality sets of data are necessary to determine each upstream length; the permissible upstream length chosen for the standard is a location at which data were taken; moreover, the data taken at that location and all the data with longer upstream lengths must lie within the acceptable band; the acceptable band is half the (expanded) uncertainty of the Reader-Harris/Gallagher (RG) Equation at infinite Reynolds number as calculated by API (from the scatter of the original data about the Equation). The value of this (expanded) uncertainty is shown in figure 1-4 of API MPMS 14.3.1:1990 (figure 4 of API MPMS 14.3.1:2012 [330]): for example, the value is 0.51 % at $\beta = 0.2$, reducing to 0.44 % at $\beta = 0.5$ and 0.6 and then increasing to 0.47 % at $\beta = 0.67$ and 0.56 % at $\beta = 0.75$.

NOTE By using this rule the maximum shift due to upstream installation should in practice often be significantly less than the maximum permitted value (i.e., half of the (expanded) uncertainty shown in figure 1-4 of API MPMS 14.3.1:1990).

Edge sharpness

The permissible value of edge sharpness is given by $r_e \leq 0.0004d$. The effect of edge radius is given in figure E4.2.11, which shows the data of Hobbs and Humphreys [331], taken in 12″ (300 mm) pipe. The data points with the smallest value of r_e/d for each β are taken as having no shift in discharge coefficient. The slope of the fitted line is approximately 550; so the effect of a change of edge radius of $0.000\,27d$ on C is 0.15 %.

Putting a maximum effect of edge sharpness as 0.15 % gives the lower (dash-dot) line shown in figure E4.2.5. All the points in the database are between the dashed line and the maximum permissible value. For Re_D greater than 1×10^6 the edge sharpness cannot cause a deviation of

more than 0.15 % from the Reader-Harris/Gallagher (1998) Equation. No actual EEC plates gave a deviation due to edge sharpness more than 0.09 % from the Reader-Harris/Gallagher (1998) equation.

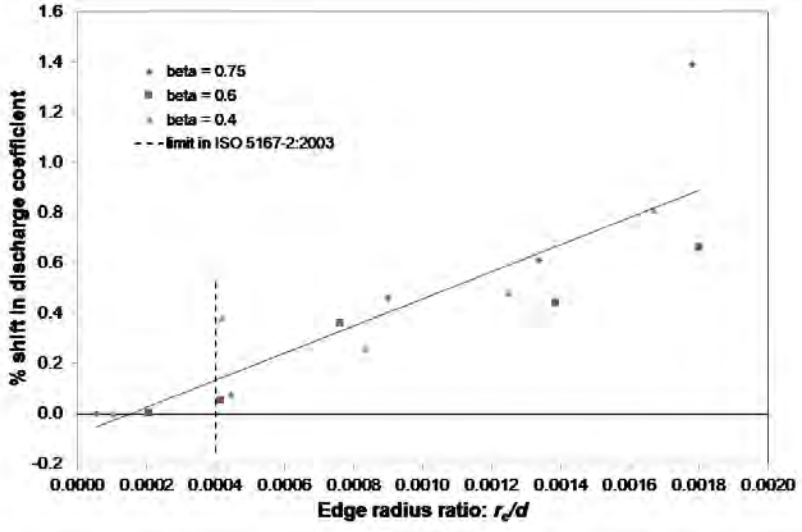


Figure E4.2.11: Effect of orifice edge radius (re) on discharge coefficient: NEL tests

Plate bending

With a slope of 1 % the theoretical model of Jepson and Chipchase [332, 333] gives a shift in discharge coefficient of a little less than 0.2 % in absolute value at maximum. In practice an orifice meter system is checked so that it is known that the plate has a slope of less than 0.5 % (i.e. a shift in discharge coefficient of less than 0.1 %) when there is no differential pressure across the plate (ISO 5167-2:2003 5.1.3.1). The plate thickness will have been chosen so that elastic deformation will not give a shift in discharge coefficient of more than 0.1 % (i.e. the plate has a slope of less than 0.5 % due to elastic bending: see ISO/TR 9464:2008 5.2.5.1.2.3 [334]).

It is assumed (reasonably) that the two types of bending, i.e. at manufacture and due to differential pressure, are independent of each other.

Eccentricity

The ISO limit according to figure 6 of ISO/TR 12767:2007 [329] gives shifts up to about 0.09 % in magnitude.

Steps

The effect of a permissible step is about 0.1 % at worst for $\beta = 0.67$: in the case depicted in figure E4.2.12 (see Reader-Harris and Brunton [335]) a Schedule 10 pipe is permitted beyond $10D$ from the orifice plate and a Schedule 80 pipe beyond $28D$ from the orifice plate (no length is specified for a Schedule 120 pipe). The requirements are in 6.4.3 of ISO 5167-2:2003. The effect is about $0.1(0.6/0.67)^{3.5}$ at worst for $\beta = 0.6$ as a dependence on β raised to a power of about 3-4 is typical for velocity profile effects on orifice plates [327, p. 249] and the minimum lengths are $10D$ and $26D$ for Schedule 10 pipe and Schedule 80 pipe, respectively.

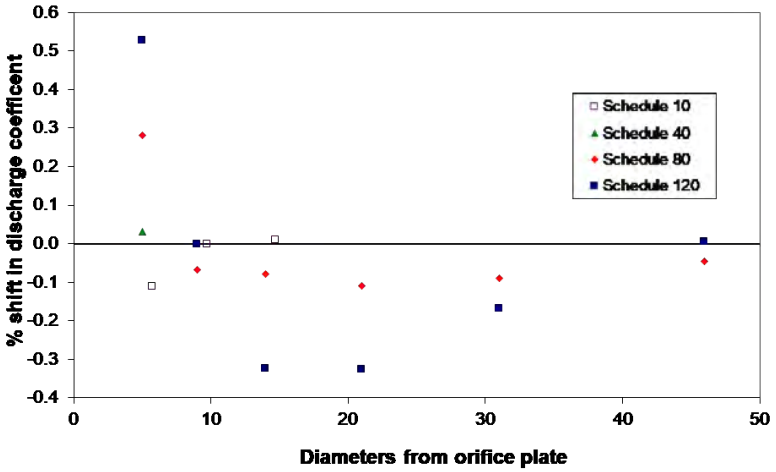


Figure E4.2.12: Upstream steps: % shift in discharge coefficient due to pipes at various distances upstream of an orifice plate ($\beta = 0.67$) in a Schedule 40 pipe

E4.2.5 Uncertainty propagation

All the above uncertainty components are listed in Table E4.2.4.

Table E4.2.4: Calculated percentage expanded uncertainty components for the Reader-Harris/Gallagher (1998) Equation

β	0.2	0.3	0.4	0.5	0.6	0.67
Base uncertainty for calculated pipe roughness and edge sharpness	0.261	0.261	0.261	0.261	0.261	0.315
Pipe roughness	0.1	0.15	0.2	0.25	0.25	0.31
Installation effects	0.255	0.24	0.23	0.22	0.22	0.235
Edge sharpness*	0.15 or 0.09	0.15 or 0.09	0.15 or 0.09	0.15 or 0.09	0.15 or 0.09	0.15 or 0.09
Plate bending at manufacture	0.1	0.1	0.1	0.1	0.1	0.1
Plate bending due to differential pressure	0.1	0.1	0.1	0.1	0.1	0.1
Eccentricity	0.09	0.09	0.09	0.09	0.09	0.09
Steps	0.003	0.007	0.018	0.039	0.068	0.1

* 0.15 % includes likely practical range; 0.09 % includes all the EEC data.

If the components in table E4.2.4 (with 0.15 % for edge sharpness) are combined by taking the root sum square of the components in table E4.2.4 the total expanded uncertainty increases monotonically from 0.44 % to 0.48(4) % as β increases from 0.2 to 0.6 and then to 0.55(8) % at $\beta = 0.67$. All these are a little smaller than the uncertainty values in ISO 5167-2:2003 (see table E4.2.3).

If, following the GUM [2], the standard uncertainty is obtained from the expanded uncertainty by dividing by 2 for the base uncertainty (since the probability distribution is assumed to be normal) and by $\sqrt{3}$ for the other components (since their probability distributions are assumed to be rectangular) the components are given in table E4.2.5.

Table E4.2.5: Calculated percentage standard uncertainty components for the Reader-Harris/Gallagher (1998) equation

β	0.2	0.3	0.4	0.5	0.6	0.67
Base uncertainty for calculated pipe roughness and edge sharpness	0.131	0.131	0.131	0.131	0.131	0.158
Pipe roughness	0.058	0.087	0.115	0.144	0.144	0.179
Installation effects	0.147	0.139	0.133	0.127	0.127	0.136
Edge sharpness*	0.087 or 0.052	0.087 or 0.052	0.087 or 0.052	0.087 or 0.052	0.087 or 0.052	0.087 or 0.052
Plate bending at manufacture	0.058	0.058	0.058	0.058	0.058	0.058
Plate bending due to differential pressure	0.058	0.058	0.058	0.058	0.058	0.058
Eccentricity	0.052	0.052	0.052	0.052	0.052	0.052
Steps	0.002	0.004	0.01	0.023	0.039	0.058

* 0.087 % includes likely practical range; 0.052 % includes all the EEC data

If the components in table E4.2.5 (with 0.052 % for edge sharpness) are combined by taking the root sum square of the components in table E4.2.5 the standard uncertainty increases monotonically from 0.233 % to 0.260 % as β increases from 0.2 to 0.6 and then to 0.301 % at $\beta = 0.67$. When these are multiplied by 1.96 the expanded uncertainty increases monotonically from 0.45(6) % to 0.51 % as β increases from 0.2 to 0.6 and then to 0.59 % at $\beta = 0.67$. Four out of six values are smaller than the uncertainty values in ISO 5167-2:2003 (see table E4.2.3); none of the values exceeds the uncertainty values in ISO 5167-2:2003 by more than 0.01 %. In some cases the higher value of edge-sharpness uncertainty might be appropriate; in many cases there will be no steps and the plate thickness will be chosen to avoid plate bending near the maximum permitted; moreover, in many cases the uncertainty due to installation effects has been overstated (see section E4.2.4.7).

E4.2.6 Reporting the result

The uncertainty of the Reader-Harris/Gallagher (1998) Equation has been calculated taking account of all the sources of uncertainty, including the uncertainty of the original data and the variability permitted by ISO 5167-2. The value is very similar to that in ISO 5167-2; the standard does not need to be changed. Confidence in such an important standard has been increased.

E4.2.7 Interpretation of results

Although no change to ISO 5167 will be proposed, a standard has been set here for other similar calculations where results may be significantly different from those by the incorrect method.

Although the uncertainty of calibration facilities has improved since the 1980s it would not be easy to achieve a very large reduction in uncertainty by collecting new data and fitting an improved discharge-coefficient equation: if the base uncertainty in table E4.2.4 were halved the total expanded uncertainty would only reduce by about 0.06 % (unless, for example, the minimum upstream straight lengths were increased or a narrower range of permitted pipe roughness achieved).

E4.2.A Comparison with other data

One way to determine the uncertainty of the Reader-Harris/Gallagher (1998) Equation is to examine how it performs when the discharge coefficients of orifice plates not used to determine the equation are compared with it. Just over 100 runs of calibration data using more than 50 orifice plates [336–345] have been compared with it: for each set of data a mean deviation has been calculated. These sets of data include both water data and high-pressure gas data; so there is a useful range of Reynolds number. The mean deviations are shown in figure E4.2.13. From this figure the quoted uncertainty in ISO 5167-2:2003 appears reasonable, but no account is taken of the uncertainty of the data itself or any analysis of the range of permissible manufactured pipes or plates or installations.

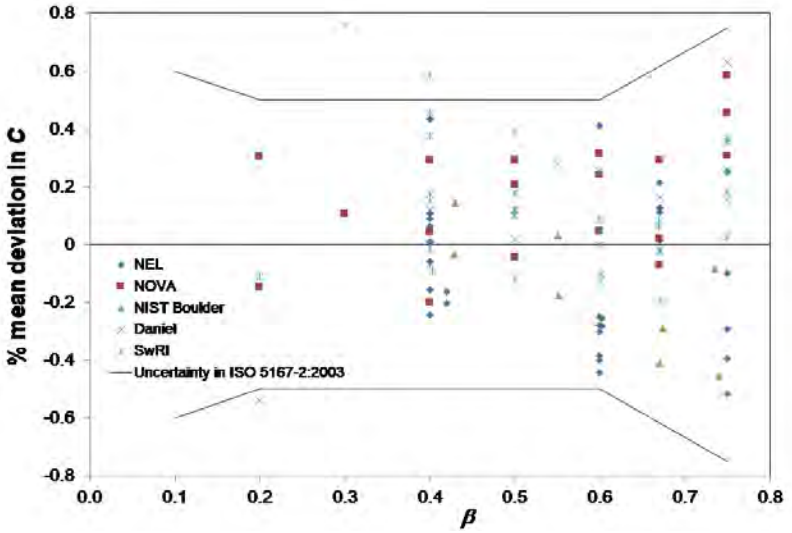


Figure E4.2.13: Percentage mean deviation in C from the Reader-Harris/Gallagher (1998) Equation from more than 50 orifice plates

E4.2.B Calibrating $8''$ (200 mm) orifice plates

Six orifice plates were calibrated in $8''$ (200 mm) pipe: one $\beta = 0.2$, one $\beta = 0.4$ and four $\beta = 0.6$. The plates had spark-eroded orifices: although no data on edge sharpness or pipe roughness were collected it has been assumed that the edge radius was 9 µm, the same as on smaller orifices from the same manufacturer (see [327, p .71]) and that the pipes had the roughness given on the drawings (upstream pipes were 0.8 µm or 1.6 µm; 1.2 µm has been used for the graphs below).

Figures E4.2.14 to E4.2.16 show the data for the three values of β: the Reader-Harris/Gallagher (1998) Equation (here called (RG(1998)) is shown on each together with its uncertainty bands; an adjusted equation taking account of the edge sharpness of the orifice and the roughness of the pipe is also shown together with an uncertainty of 0.261 %. The uncertainty of the data is 0.16 %. Moreover, all the sets of data are close to parallel to the Reader-Harris/Gallagher (1998) Equation.

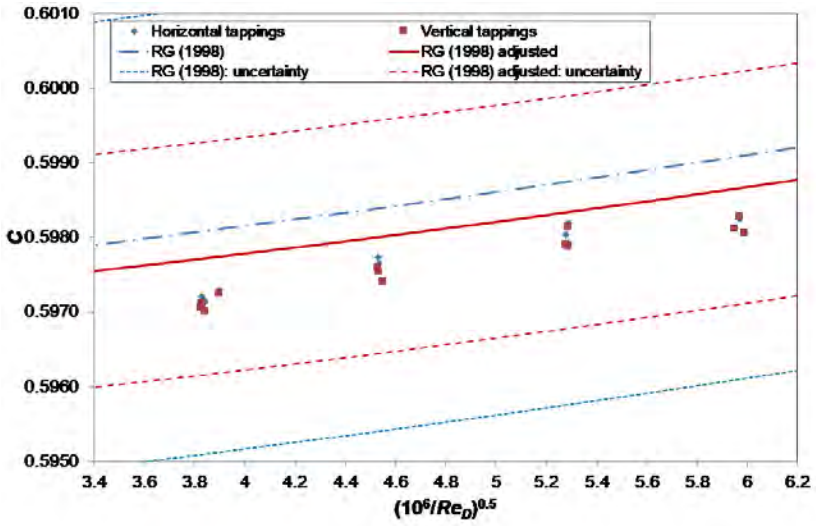


Figure E4.2.14: Data for $\beta = 0.2$

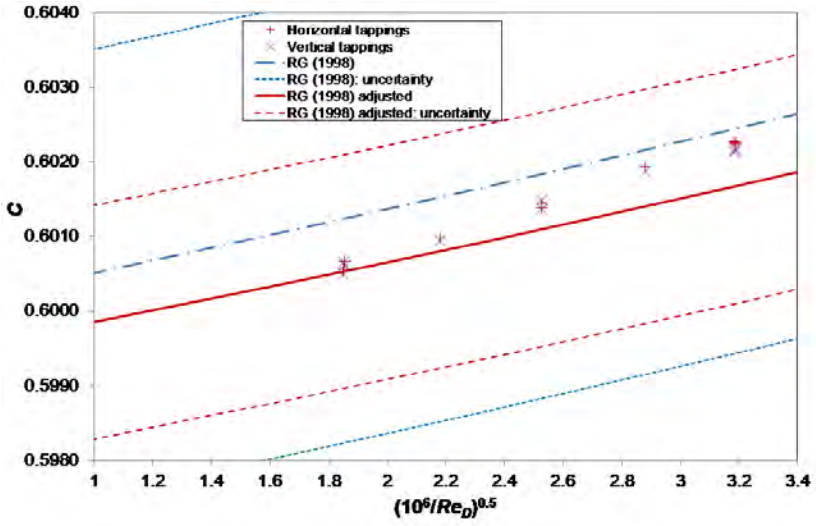


Figure E4.2.15: Data for $\beta = 0.4$

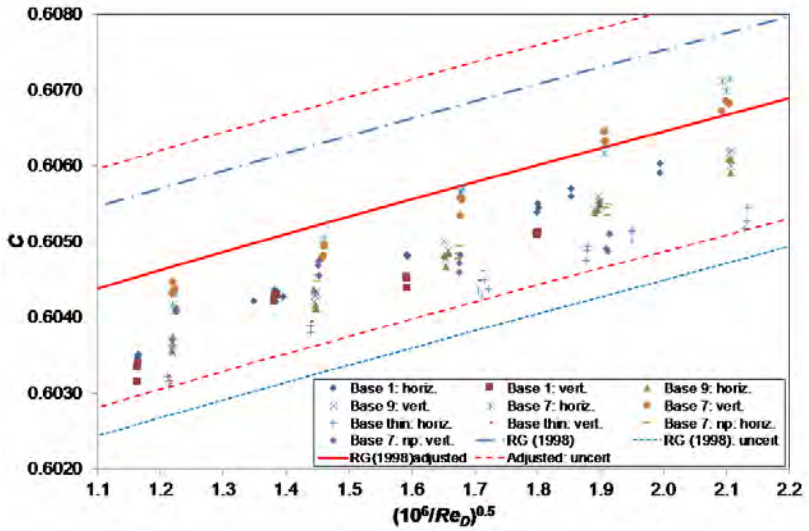


Figure E4.2.16: Data for $\beta = 0.6$

Acknowledgement

The authors are grateful to Maurice Cox (NPL, UK), Álvaro Silva Ribeiro (LNEC, Portugal), João Alves e Sousa (IPQ, Portugal) and Adriaan van der Veen (VSL, Netherlands), who provided valuable comments on this report and made it clearer to the reader.

Example E4.3

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

E4.3.1 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 method (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).

E4.3.2 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 [346], for flow limiting and overspeed protection of gas flow meters, to name a few.

For toroidal sonic nozzles, the standard ISO 9300 [347] specifies the following relation between the discharge coefficient Y and the Reynolds number Re

$$Y = \beta_0 + \frac{\beta_1}{\sqrt{Re}}, \tag{E4.3.1}$$

see Appendix E4.3.A for some background information. The two parameters β_0 and β_1 characterise a specific nozzle (depending on its inner contour curvature and surface structure [348, 349]). The use of the nozzle as transfer standard depends on the reliability of the estimates and uncertainties of these parameters.

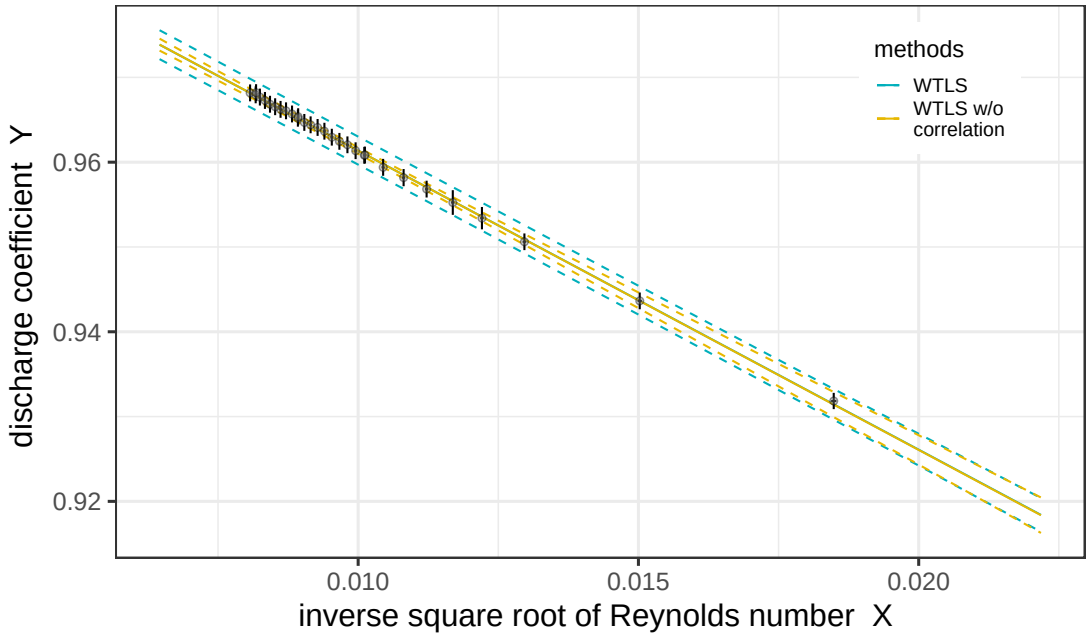


Figure E4.3.1: Estimates and uncertainties for the discharge coefficient Y_n as a function of the inverse square root of Reynolds number X_n (both dimensionless) for measurements $n = 1, \ldots, 30$. The data are available online in repository [350]. Superimposed are the estimated straight-line relationship (solid lines) and the associated 95 % coverage bands (dashed lines) obtained by WTLS (blue) and by WTLS without correlation (yellow).

In this example, a toroidal sonic nozzle shall be characterized based on $N = 30$ pairs of discharge coefficients Y_n and inverse square root of Reynolds numbers $X_n = Re_n^{-0.5}$, $n = 1, \ldots, N$. Estimates and uncertainties for these quantities, as measured at PTB and displayed in figure E4.3.1, rely on common input quantities which cause correlation (cf. [2, clause 5.2.4]). Such correlations 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 (E4.3.1).

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 E4.3.2, 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 MCM.

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 WTLS method which accounts for all uncertainties and correlations derived in stage 1.

E4.3.3 Specification of the measurands

Let X denote the inverse square root of the Reynolds number, and Y the discharge coefficient. According to equation (E4.3.1) and equivalently standard [347], Y depends linearly on X,

$$Y = \beta_0 + \beta_1 X. \tag{E4.3.2}$$

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 [4, clause 3.20]. Thus, uncertainties and correlations can be derived directly from $\boldsymbol{U}$.

E4.3.4 Measurement models

The uncertainty evaluation in this example consists of two consecutive measurement models, see figure E4.3.2. Section E4.3.4.1 first describes the measurement model for the intermediate measurand $(\boldsymbol{X}, \boldsymbol{Y})$ and characterises its input quantities. Section E4.3.4.2 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.

E4.3.4.1 Intermediate measurands X and Y (stage 1)

Referring to appendix E4.3.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

$$X_n = \sqrt{\frac{\pi M D}{4 Q_n}}, \quad Y_n = \frac{4 Q_n}{\pi D^2 \Psi_n^*} \quad \text{with } n = 1, \ldots, N, \tag{E4.3.3}$$

and denoting $\boldsymbol{Q} = (Q_1, \ldots, Q_N)^\top$ and $\boldsymbol{\Psi}^* = (\Psi_1^*, \ldots, \Psi_N^*)^\top$. Equations (E4.3.3) 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

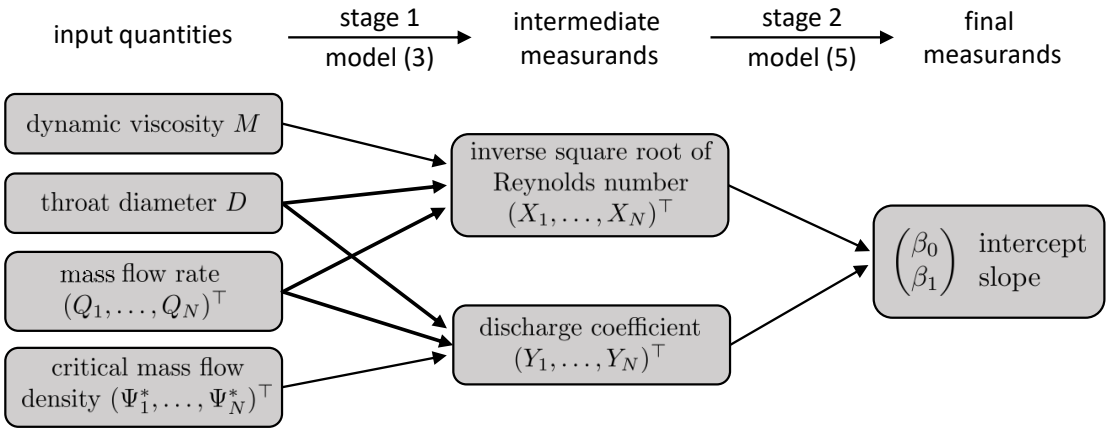


Figure E4.3.2: Illustration of the two consecutive measurement models and their input quantities. density Ψ_n^* impact on the quantity Y_n. (See also figure E4.3.2.) The intermediate measurands $\boldsymbol{X}$ and $\boldsymbol{Y}$ are correlated due to the common quantities $\boldsymbol{Q}$ and D. The 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 E4.3.1. 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 $u_\mathrm{A}(.)$ are obtained from repeated measurements and are provided online in the repository [350]. The Type B uncertainty contributions, say $u_\mathrm{B}(.)$, are given in table E4.3.1. The Type A and Type B evaluated uncertainties are added quadratically $\sqrt{u_\mathrm{A}^2(.)+u_\mathrm{B}^2(.)}$ 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 [347, clause 9] and [351] if necessary). For general guidance, how to assign distributions to input quantities (also multivariate ones), see GUM-S1 and GUM-S2 [3, 4].

Having specified all input quantities, the assigned distributions can be propagated through model (E4.3.3) 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 E4.3.5.1 and Supplement 2 to the GUM [4]. Estimates $(\boldsymbol{x},\boldsymbol{y})$ and the full covariance matrix $\boldsymbol{U}$ can then be derived from the Monte Carlo samples of this distribution.

Table E4.3.1: Characterization of the input quantities in measurement model (E4.3.3). 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	1.82×10^{-5}	$\mathrm{kg\,m^{-1}\,s^{-2}}$	5	[352]
Q_n	mass flow rate	Gaussian	–	$\mathrm{kg\,s^{-1}}$	0.55	[353]
Ψ_n^*	critical mass flow	Gaussian	–	$\mathrm{kg\,m^{-2}\,s^{-1}}$	0.25	[347, clause 8.3] [351, 354]

E4.3.4.2 Weighted total least-squares method (stage 2)

The measurement model for the straight-line regression (E4.3.2) can be constructed from the appropriate least-squares method. The frequently applied ordinary and the weighted least-squares method (WLS) may be inappropriate here, because they assume the measured values $\boldsymbol{x}$ to be exact. WTLS takes into account uncertainties associated with the estimates $\boldsymbol{x}$ and $\boldsymbol{y}$ for the intermediate measurands $\boldsymbol{X}$ and $\boldsymbol{Y}$ as well as their associated covariances. WTLS is recommended by multiple standards [77, 110] and applied here.

The WTLS is based on minimizing the generalized sum of squares

$$S = \begin{pmatrix} \boldsymbol{x} - \widetilde{\boldsymbol{\xi}} \\ \boldsymbol{y} - (\widetilde{\beta}_0 + \widetilde{\beta}_1 \widetilde{\boldsymbol{\xi}}) \end{pmatrix}^{\top} \boldsymbol{U}^{-1} \begin{pmatrix} \boldsymbol{x} - \widetilde{\boldsymbol{\xi}} \\ \boldsymbol{y} - (\widetilde{\beta}_0 + \widetilde{\beta}_1 \widetilde{\boldsymbol{\xi}}) \end{pmatrix}, \tag{E4.3.4}$$

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 $\boldsymbol{x} = (x_1, \ldots, x_N)^{\top}$ and the vectors $\boldsymbol{y}$ and $\widetilde{\boldsymbol{\xi}}$ are likewise defined. The minimizer of (E4.3.4) defines the solution $(\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\boldsymbol{\xi}}^{\top})$ of the WTLS.

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

$$(\beta_0, \beta_1, \boldsymbol{\xi}^{\top})^{\top} = \arg \min_{\widetilde{\beta}_0, \widetilde{\beta}_1, \widetilde{\boldsymbol{\xi}}} \begin{pmatrix} \boldsymbol{X} - \widetilde{\boldsymbol{\xi}} \\ \boldsymbol{Y} - (\widetilde{\beta}_0 + \widetilde{\beta}_1 \widetilde{\boldsymbol{\xi}}) \end{pmatrix}^{\top} \boldsymbol{U}^{-1} \begin{pmatrix} \boldsymbol{X} - \widetilde{\boldsymbol{\xi}} \\ \boldsymbol{Y} - (\widetilde{\beta}_0 + \widetilde{\beta}_1 \widetilde{\boldsymbol{\xi}}) \end{pmatrix}, \tag{E4.3.5}$$

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

E4.3.5 Estimation and uncertainty evaluation

Following the GUM [2, 4], estimates $\widehat{\beta}_0$ and $\widehat{\beta}_1$ of the final measurands are obtained by evaluating measurement model (E4.3.5) at the estimates $\boldsymbol{x}$ and $\boldsymbol{y}$ of the intermediate measurand $(\boldsymbol{X}, \boldsymbol{Y})$. The uncertainties associated with $(\widehat{\beta}_0, \widehat{\beta}_1)$ result from propagating the covariance matrix $\boldsymbol{U}$ associated with these intermediate estimates through the same model, which will be detailed in section E4.3.5.2. Before, section E4.3.5.1 describes how to arrive at the intermediate estimates $\boldsymbol{x}$ and $\boldsymbol{y}$, and the full covariance matrix $\boldsymbol{U}$ following the GUM [4].

E4.3.5.1 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 E4.3.4.1, shall be propagated through model (E4.3.3) to evaluate estimates, uncertainties and correlations for the intermediate measurand $(\boldsymbol{X}, \boldsymbol{Y})$. Application of the multivariate LPU [4, clause 6] is lengthy due to the full correlation structure resulting from model (E4.3.3), and it may not be adequate due to non-linearities in model (E4.3.3). Instead, the MCM [4, clause 7] is applied to approximate the $2N$-dimensional distribution for $(\boldsymbol{X}, \boldsymbol{Y})$ 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 (E4.3.3) is applied to each of these samples, generating R samples of the intermediate measurand $(\boldsymbol{X}, \boldsymbol{Y})$ – each sample representing a random

[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" [355].

realisation from the $2N$-dimensional distribution. If these output samples form a matrix of dimension $2N \times R$, an estimate $(\boldsymbol{x}, \boldsymbol{y})$ is obtained by averaging over its columns, and the covariance matrix $\boldsymbol{U}$ is obtained by calculating the covariance between all rows (cf. [4, clause 7.6]).

This summary information for $(\boldsymbol{X}, \boldsymbol{Y})$ is available online in repository [350]. The estimates and uncertainties are displayed in figure E4.3.1 and the correlation matrices for $\boldsymbol{X}$ as well as for $\boldsymbol{Y}$ are displayed in figure E4.3.3. 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 E4.3.1). The correlation coefficients associated with pairs of estimates y_n and y_m are mostly about 2/3 (see right panel in figure E4.3.3). Only the correlation associated with the components $n = 23$ and $n = 24$ of $\boldsymbol{Y}$ 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 $\boldsymbol{Y}$, while the covariance element $u(y_n, y_m)$ remains unaffected. The magnitude of the correlation between $\boldsymbol{X}$ and $\boldsymbol{Y}$ is much smaller (in the range between -0.067 and -0.045) and not displayed here.

Let us note, that model (E4.3.3) 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 [356] code, which is also available online in repository [350]. We chose to implement $R = 1 \times 10^8$ Monte Carlo trials providing a relative numerical accuracy of smaller than 1×10^{-7} for the estimates $(\boldsymbol{x}, \boldsymbol{y})$, but as high as 5×10^{-3} for the covariances $\boldsymbol{U}$ (especially for the covariance between $\boldsymbol{X}$ and $\boldsymbol{Y}$). Nevertheless, the results of the final measurand (β_0, β_1) vary little when repeating the Monte Carlo procedure (no more than 1×10^{-3} times the associated standard uncertainty).

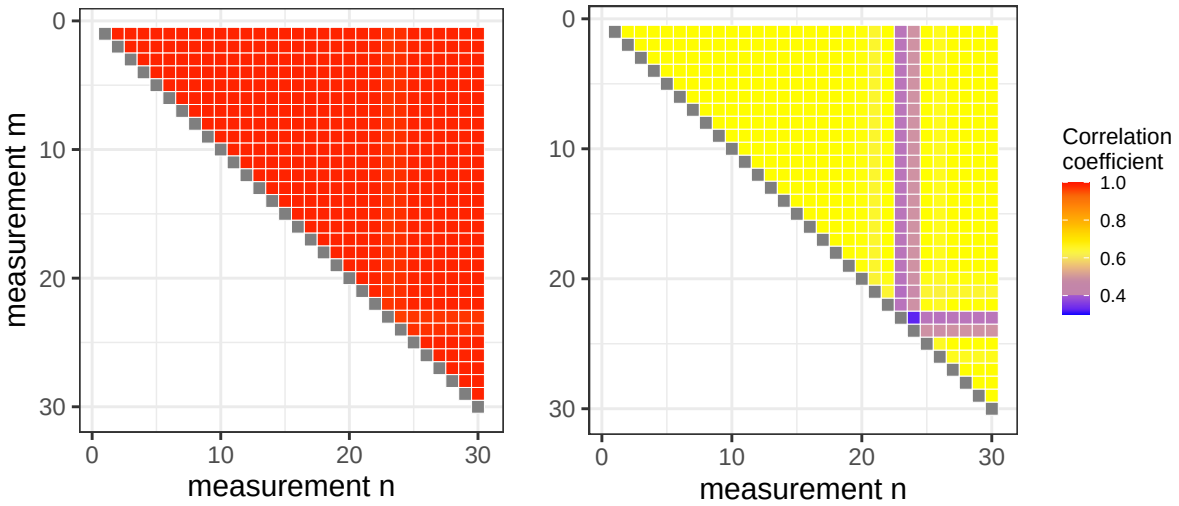


Figure E4.3.3: Upper triangle of the correlation matrices associated with $\boldsymbol{x}$ (left panel) and with $\boldsymbol{y}$ (right panel), obtained from model (E4.3.3) and by the Monte Carlo method, cf. sections E4.3.4.1 and E4.3.5.1. To increase the visibility of strongly correlated values, values with a perfect correlation of 1 are marked in grey.

E4.3.5.2 Weighted total least-squares method (stage 2)

The measurement model (E4.3.5) based on WTLS is implicit, multivariate, non-linear and usually no closed form is available for its solution. An iterative scheme for deriving estimates $\widehat{\beta}_i$ with $i = 0, 1$ and their associated uncertainties $u\left(\widehat{\beta}_i\right)$ is described in clause 10 of the standard [77] and implemented in the R Markdown code in [350]. This simple scheme also provides correlations between β_0 and β_1. The iterative algorithm requires the covariance matrix $\boldsymbol{U}$ 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 [357], or other decomposition methods.

Assuming a Gaussian distribution[2] for the measurand, a 95 % coverage interval for each measurand β_i is given by

$$\left[\widehat{\beta}_i - 1.96\, u\left(\widehat{\beta}_i\right), \widehat{\beta}_i + 1.96\, u\left(\widehat{\beta}_i\right)\right]. \tag{E4.3.6}$$

A two-dimensional, joint 95 % coverage region can be calculated following [4, clause 6.5.2].

E4.3.6 Reporting the result

Table E4.3.2 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 (E4.3.5) for the weighted total least-squares method to the estimates $\boldsymbol{x}$, $\boldsymbol{y}$ and their associated covariance matrix $\boldsymbol{U}$; which in turn are obtained by applying measurement model (E4.3.3) using the Monte Carlo method.

Table E4.3.2: 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	$\widehat{\beta}_0$	$u(\widehat{\beta}_0)$	$\widehat{\beta}_1$	$u(\widehat{\beta}_1)$	$\mathrm{cov}(\widehat{\beta}_0,\widehat{\beta}_1)$
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}

For comparison, measurement model (E4.3.5) is applied ignoring any correlation (WTLS w/o correlation) by assuming a diagonal covariance matrix $\boldsymbol{U}$ (see also [77, clause 8]). In addition, measurement model (E4.3.5) is applied omitting the uncertainty and correlation in the quantity $\boldsymbol{X}$, as well as the correlation between $\boldsymbol{X}$ and $\boldsymbol{Y}$. That is, a simple weighted least-squares (WLS) fit is applied accounting only for the uncertainty and correlation in quantity $\boldsymbol{Y}$ (see also [77, clause 9]). Figure E4.3.4 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 GUM-S2 [4] to measurement model (E4.3.5) and the algorithm in [77, clause 10]. The non-linearity of (E4.3.5) or a non-normality behind (E4.3.6) could cause differences, however, this was not observed.

[2] Cf. clause 6.3.3 in [2] and section 10.2.3 in [77] for the approximate validity of this normality assumption.

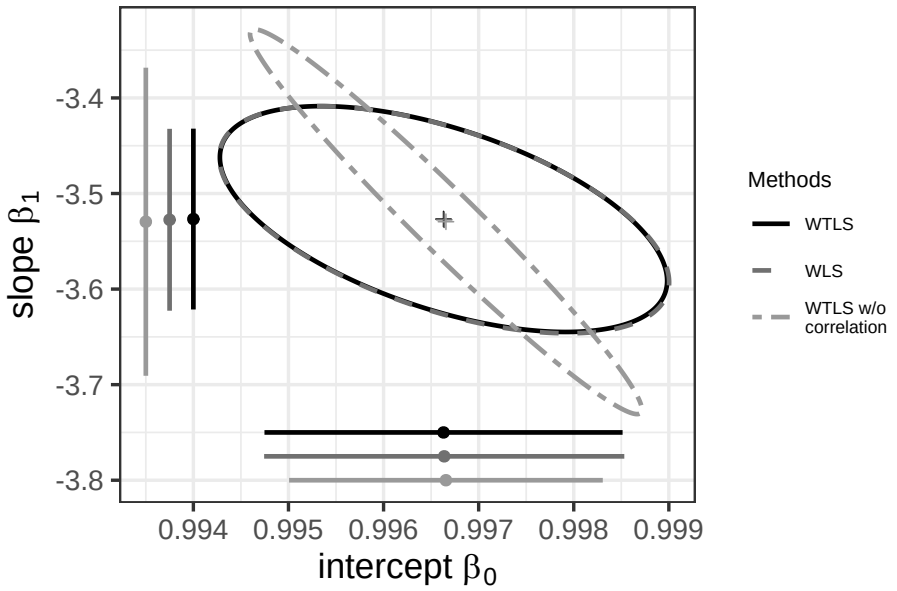


Figure E4.3.4: Displayed are the estimates $\widehat{\beta}_0$ and $\widehat{\beta}_1$ (markers), their joint 95 % coverage region (ellipses), and the associated 95 % coverage intervals (horizontal and vertical lines) for the least-squares methods listed in table E4.3.2.

We stress that OLS and WLS using standard software often involves the estimation of a multiple for the variance of the input quantity $\boldsymbol{Y}$. 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 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 E6.2.

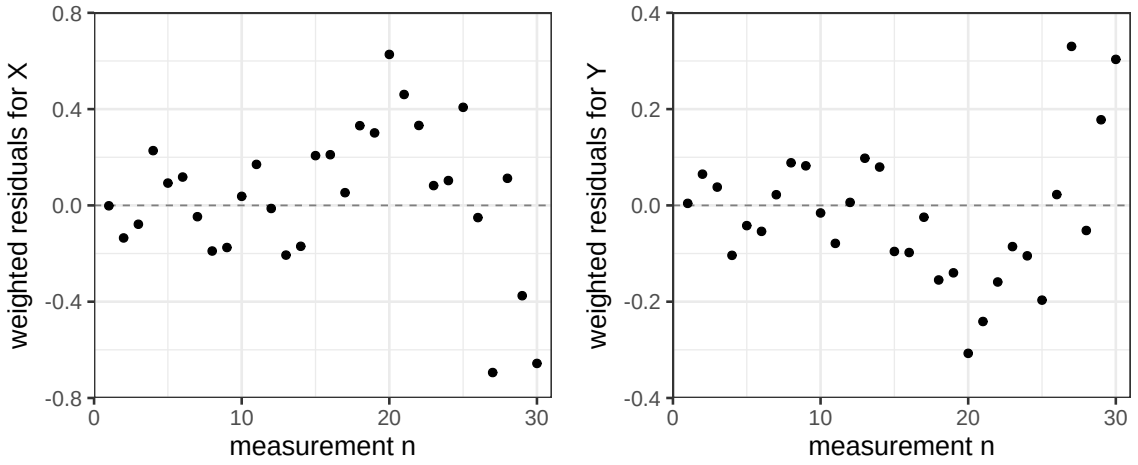


Figure E4.3.5: Displayed are the residuals $\left(\left(\boldsymbol{x}-\widehat{\boldsymbol{\xi}}\right)^{\top},\left(\boldsymbol{y}-\widehat{\beta}_0-\widehat{\beta}_1\widehat{\boldsymbol{\xi}}\right)^{\top}\right)$ weighted by the Cholesky decomposed covariance matrix U. The absence of systematic behaviour seen here indicates that there is no violation of the straight-line relation (E4.3.2).

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 (E4.3.2), see figure E4.3.5. 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 E4.3.6. 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.

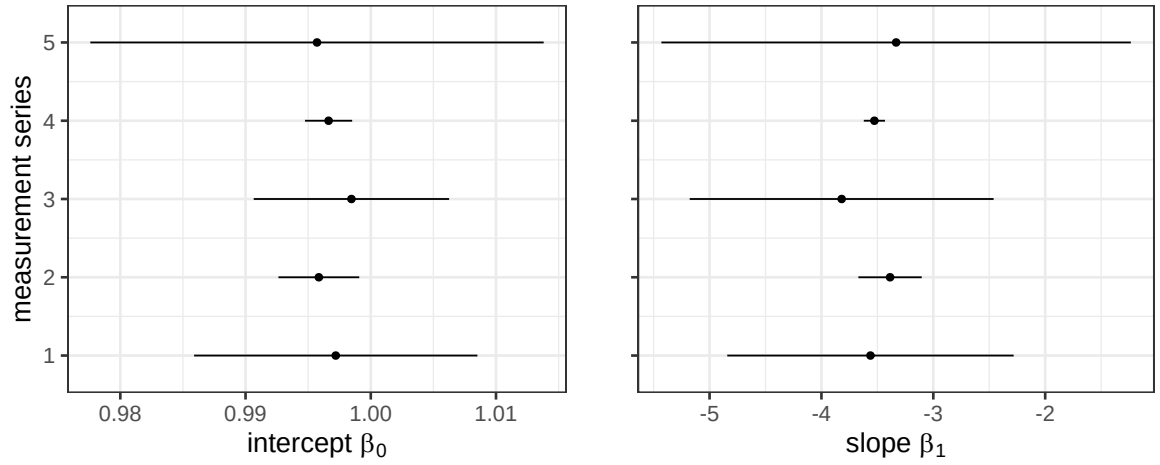


Figure E4.3.6: Displayed are the estimates $\widehat{\beta}_0$ and $\widehat{\beta}_1$ (dots) and their 95 % coverage intervals obtained by the WTLS method (E4.3.5) for each of five, temporally separate, measurement series. These were measured (1) at 06/21/2010 with $N = 3$, (2) at 06/28/2012 with $N = 6$, (3) at 10/08/2012[3] with $N = 9$, (4) at 07/15/2018 with $N = 30$, and (5) at 02/21/2019 with $N = 3$.

E4.3.7 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 E4.3.2 and figure E4.3.4). 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 estimated straight line has a different shape and is much broader over the range of the data when correlation is accounted for; see figure E4.3.1.

Ignoring the uncertainty and correlation in the quantity $\boldsymbol{X}$ yields almost identical results in this example (compare WTLS and WLS in table E4.3.2 and figure E4.3.4). The reason is twofold: First, the correlation between $\boldsymbol{X}$ and $\boldsymbol{Y}$ is small, and implicitly ignoring it in the simpler weighted

[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}.

least-squares method is inconsequential. Second, also the squared standard uncertainties and covariances 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

The authors would like to thank Adriaan van der Veen (VSL) and Michael Reader-Harris (NEL) for reviewing the example.

E4.3.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 [347]. Here, we follow the notation of the GUM (cf. Note 1–3 in [2, clause 4.1]) and denote all quantities which are used in the main part of this example by capital letter symbols.

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 [351, 354].

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 [358, p. 250] [359]

$$Y = \beta_0 \left(1 - 2\frac{\delta_1}{D}\right)^2. \tag{E4.3.7}$$

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 (E4.3.7), one obtains the functional relationship between the discharge coefficient Y and the Reynolds number Re provided by the standard ISO 9300 [347], namely,

$$Y = \beta_0 + \beta_1 Re^{-0.5}. \tag{E4.3.8}$$

In [347], two nozzle types, the cylindrical and the toroidal one, are defined with tight requirements to keep to specifications. Thereby, equation (E4.3.8) 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 E4.4

Measurement uncertainty evaluation of the load loss of power transformers

A. Bošnjaković, V. Karahodžić, M. Čaušević, A.M.H. van der Veen

E4.4.1 Summary

In this example, the evaluation of measurement uncertainty of power transformer losses are given. It is shown how this evaluation can be performed using the law of propagation of uncertainty (LPU) from the Guide to the expression of Uncertainty in Measurement (GUM) as well as the propagation of distributions using the Monte Carlo method (MCM) from GUM Supplement 1 (GUM-S1). It is shown how the approach using the Monte Carlo method can be used to validate the output from the law of propagation of uncertainty from the GUM.

E4.4.2 Introduction of the application

When it comes to electrical energy production, transformation, distribution or consumption, energy losses generated by transformers play a very important role. These losses form the second largest part of the total losses in the distribution of energy and the network. Measurement results of power transformer losses should be accurate as these are often the object of guarantee and penalty in many contracts and hence play an important role in billing. Costs of losses in power transformers are comparable to the product cost and play an important role in the evaluation of the total costs of obtaining the energy at the point of use.

Many European regulations set requirements for reliable loss values. For instance, the Ecodesign Directive [360] provides EU rules for improving environmental performance of products, among which power transformers are recognized. The directive sets out minimum mandatory requirements for the energy efficiency of these products, which helps to prevent creation of barriers to trade, improve product quality and yield environmental protection. The European Commission Regulation [361] on implementation of Directive [360] with regard to small, medium and large power transformers recognised energy in the use phase as one of the most significant environmental aspect that can be addressed through product design. This regulation provides basic measurements concepts as well as tolerances to be achieved in order to comply with this regulation.

In the case of market surveillance, if a power transformer facility exceeds the guaranteed limit for losses, a fine must be paid. Therefore, IEC 60076-19 [362] recommends that the evaluated uncertainty should be less than the tolerance limit set out in regulations. The measurement uncertainty is often not taken into account in considering the agreement between the manufacturer and costumer on who will pay the fine in case the losses exceed the guaranteed value. Furthermore, IEC 60076-19 recommends that guarantee and penalty calculations should refer to the best estimated values of the losses without considering the measurement uncertainties, based on a shared risk concept, where both parties are aware of and accept the consequences of non-negligible measurement uncertainty.

In this example, it is shown how the evaluation of the measurement uncertainty for load losses can be conducted to improve existing practice for power load loss measurements. Two approaches are used: the law of propagation of uncertainty of the GUM [2] and the propagation of distributions [3]. The latter method is applied to validate the uncertainty evaluation method proposed in IEC 60076-19 [362].

E4.4.3 Specification of the measurand(s)

The measurand in this example is power load loss, stated at the rated current at a given reference temperature. A value for the measurand is either obtained by measuring the power load loss at the rated current, or by recalculation to this current. The measurand is specified at a temperature. The conversion to that reference temperature is part of the measurement model.

E4.4.4 Measurement model

The total load loss P_{LL} of a transformer at reference temperature θ_{r}, summed over the three phases, can be expressed as (adapted from [362])[1]

$$P_{\mathrm{LL}} = \sum_{i=1}^{3} I_{\mathrm{NHV}}^2 R_{1,\mathrm{HV},i} \frac{t+\theta_{\mathrm{r}}}{t+\theta_1} + I_{\mathrm{NLV}}^2 R_{1,\mathrm{LV},i} \frac{t+\theta_{\mathrm{r}}}{t+\theta_1} + P_{\mathrm{a2},i} \frac{t+\theta_2}{t+\theta_{\mathrm{r}}}, \tag{E4.4.1}$$

where I_{NHV} denotes the rated primary phase current, I_{NLV} the rated secondary phase current, $R_{1,\mathrm{HV},i}$ the resistance of the primary winding at temperature θ_1 for phase i, $R_{1,\mathrm{LV},i}$ the resistance of the secondary winding at temperature θ_1 for phase i, t a constant for the winding material, θ_1 the temperature during the resistance measurement, θ_2 the temperature during loss measurement, θ_{r} the reference temperature and $P_{\mathrm{a2},i}$ the additional loss for phase i. The index i runs over the three phases (denoted by U, V, and W). The additional loss is computed as

$$P_{\mathrm{a2},i} = P_{2,i} - I_{\mathrm{NHV},i}^2 R_{2,\mathrm{HV},i} - I_{\mathrm{NLV},i}^2 R_{2,\mathrm{LV},i}, \tag{E4.4.2}$$

where

$$R_2 = R_1 \frac{t+\theta_2}{t+\theta_1}, \tag{E4.4.3}$$

[1] IEC 60076-19 does not provide an explicit equation for the summation over the three phases of the alternating current. In this example, we provide this expression explicitly, and substitute for the summands (the load loss $P_{\mathrm{LL},i}$ of the three phases the expression given in the said standard instead).

and the power $P_{2,i}$ measured at the load loss measurement corrected for known systematic deviations and referred to the current $I_{\mathrm{N,prim},i}$ is computed as

$$P_{2,i} = k_{\mathrm{CN}}(1+\varepsilon_{\mathrm{C}})k_{\mathrm{VN}}\frac{1}{1+\varepsilon_{\mathrm{V}}}\frac{P_{\mathrm{W},i}}{1-(\Delta_{\varphi V,i}-\Delta_{\varphi C,i})\tan\varphi_i}\left(\frac{I_{\mathrm{NHV},i}}{k_{\mathrm{CN}}I_{\mathrm{M},i}}\right)^2, \tag{E4.4.4}$$

where

$\left(\frac{I_{\mathrm{NHV}}}{k_{\mathrm{CN}}I_{\mathrm{M},i}}\right)^2$	the term related to the actual current $I_{\mathrm{M},i}$, rated to the reference current I_{NHV} for which transformer shall be tested,
k_{CN}	rated transformation ratio of the current transformer,
k_{VN}	rated transformation ratio of the voltage transformer,
ε_{C}	actual ratio error of the current transformer (% of nominal ratio),
ε_{V}	actual ratio error of the voltage transformer (% of nominal ratio),
$\frac{1}{1-(\Delta_{\varphi\mathrm{V},i}-\Delta_{\varphi\mathrm{C},i})\tan\varphi_i}$	the term related to the correction for phase displacement ($F_{\mathrm{D},i}$) of the current ($\Delta_{\varphi\mathrm{C},i}$) and voltage transformers ($\Delta_{\varphi\mathrm{V},i}$),
$P_{\mathrm{W},i}$	reading of the wattmeter and the wattmeter is considered to have unknown errors, and thus no correction term.

The actual phase angle φ_i between voltage and current under the sinusoidal conditions normally valid for load-loss measurement depends on the phase angle measured with power meter (φ_{M}) and is obtained from

$$\varphi_i = \varphi_{\mathrm{M},i} - (\Delta_{\varphi\mathrm{V},i}-\Delta_{\varphi\mathrm{C},i}) = \arccos\left(\frac{P_{\mathrm{W},i}}{I_{\mathrm{M},i}U_{\mathrm{M},i}}\right) - (\Delta_{\varphi\mathrm{V},i}-\Delta_{\varphi\mathrm{C},i}). \tag{E4.4.5}$$

The substitution of equations (E4.4.2) and (E4.4.3) into (E4.4.1) yields

$$\begin{aligned} P_{\mathrm{LL}} &= \sum_{i=1}^{3} I_{\mathrm{NHV}}^2 R_{1,\mathrm{HV},i}\frac{t+\theta_{\mathrm{r}}}{t+\theta_1} + I_{\mathrm{NLV}}^2 R_{1,\mathrm{LV},i}\frac{t+\theta_{\mathrm{r}}}{t+\theta_1} + \left(P_{2,i} - I_{\mathrm{NHV},i}^2 R_{2,\mathrm{HV},i} - I_{\mathrm{NLV},i}^2 R_{2,\mathrm{LV},i}\right)\frac{t+\theta_2}{t+\theta_{\mathrm{r}}} \\ &= \frac{t+\theta_{\mathrm{r}}}{t+\theta_1}\sum_{i=1}^{3}\left(I_{\mathrm{NHV}}^2 R_{1,\mathrm{HV},i} + I_{\mathrm{NLV}}^2 R_{1,\mathrm{LV},i}\right) + \frac{t+\theta_2}{t+\theta_{\mathrm{r}}}\sum_{i=1}^{3}\left(P_{2,i} - I_{\mathrm{NHV},i}^2 R_{2,\mathrm{HV},i} - I_{\mathrm{NLV},i}^2 R_{2,\mathrm{LV},i}\right) \\ &= \left(\frac{t+\theta_{\mathrm{r}}}{t+\theta_1} - \frac{t+\theta_2}{t+\theta_{\mathrm{r}}}\frac{t+\theta_2}{t+\theta_1}\right)\sum_{i=1}^{3}\left(I_{\mathrm{NHV}}^2 R_{1,\mathrm{HV},i} + I_{\mathrm{NLV}}^2 R_{1,\mathrm{LV},i}\right) + \frac{t+\theta_2}{t+\theta_{\mathrm{r}}}\sum_{i=1}^{3} P_{2,i} \end{aligned} \tag{E4.4.6}$$

E4.4.5 Evaluation of the input quantities

E4.4.5.1 Specifications and measurement data

The characteristics of the transformer are summarised in table E4.4.1. The rated primary and secondary currents are considered constants, that is, without uncertainty. The same applies to the reference temperature θ_{r}. As the windings are made from copper (Cu), $t = 235$ and also treated as a constant [362].

Table E4.4.1: Specification of the phase transformer – dry type transformer [362]

Component	Symbol	Value	Unit
Rated power	S_r	630	kV A
Rated primary voltage	U_{NHV}	6000	V
Rated secondary voltage	U_{NLV}	400	V
Rated primary phase current	I_{NHV}	60.62	A
Rated secondary phase current	I_{NLV}	909.33	A
Reference temperature	θ_r	120	°C
Winding material		Cu	

This example uses the specifications of the transformer as well as of the measuring instruments. These specifications are summarised in table E4.4.2.

Table E4.4.2: Specifications of measuring system current and voltage transducers for the three phases U, V and W (adapted from IEC 60076-19 [362])

Component	Symbol/Unit	U	V	W
CT accuracy class		0.2	0.2	0.2
CT ratio	k_{CN}	1	1	1
CT max. amplitude error	$e_{class,CT}$/%	±0.2	±0.2	±0.2
CT max. phase displacement	$\Delta_{\phi,C}/'$	±10	±10	±10
VT accuracy class		0.2	0.2	0.2
VT ratio	k_{VN}	1	1	1
VT max. amplitude error	$e_{class,VT}$/%	±0.2	±0.2	±0.2
VT max. phase displacement	$\Delta_{\phi,V}/'$	±10	±10	±10
Maximum of current range of Power Meter	I_{max}/A	50	50	50
Maximum of voltage range of Power Meter	U_{max}/V	240	240	240

The measured data are summarised in table E4.4.3.

Table E4.4.3: Experimental data of the load loss measurement [362]

Component	Symbol/Unit	U	V	W
Resistance HV	R_{HV}/Ω	0.0490	0.0500	0.0510
Resistance LV	R_{LV}/Ω	0.001 600	0.001 500	0.001 700
Temp. during resistance meas.	θ_1/°C	22.1	22.1	22.1
Voltage Phase-Neutral	U_M/V	191.5	195.2	194.8
Phase current	I_M/A	40.55	40.2	40.6
Measured active power	P_W/W	748	756	762
Power factor	$\cos\varphi/1$	0.096 33	0.096 34	0.096 35
Temp. during loss meas.	θ_2/°C	21.8	21.8	21.8

Starting with the input quantities in equation (E4.4.4), ε_C is modelled as having a rectangular distribution (uniform distribution) with zero mean and semi-width $e_{\text{class,CT}}$ (see table E4.4.2). The standard uncertainty is given by

$$u(\varepsilon_C) = \frac{e_{\text{class,CT}}}{\sqrt{3}}.$$

In the application of the law of propagation of uncertainty of the GUM, the fact will be exploited that equation (E4.4.4) can be considered as a product of a set of independent quantities of the kind

$$Y = cX_1^{p_1}X_2^{p_2}\ldots X_N^{p_N},$$

where c denotes a constant. If the input quantities $X_1, X_2, \ldots, X_N$ are mutually independent, then the squared relative standard uncertainty associated with the measurand Y can be expressed as follows in terms of the relative standard uncertainties associated with the input quantities [2, clause 5.1.6]:

$$u_{\text{rel}}^2(Y) = \sum_{i=1}^{N} p_i^2 u_{\text{rel}}^2(X_i)$$

Considering $1 + \varepsilon_C$ as a term in the product, the relative standard uncertainty of this term is needed and is given by

$$u_{\text{rel}}(1+\varepsilon_C) = \frac{u(\varepsilon_C)}{1+\varepsilon_C}. \tag{E4.4.7}$$

The evaluation of the standard uncertainty of ε_V is very similar to that of ε_C. ε_V is modelled using the rectangular distribution with zero mean and semi-width $e_{\text{class,VT}}$. The relative standard uncertainty of $1/(1+\varepsilon_V)$ is given by

$$u_{\text{rel}}\left(\frac{1}{1+\varepsilon_V}\right) = \frac{u(\varepsilon_V)}{1+\varepsilon_V} \tag{E4.4.8}$$

using the fact that $|u_{\text{rel}}(1/X)| = |u_{\text{rel}}(X)|$.

The standard uncertainty of the power measurement $P_{W,i}$ is based on the specification for the maximum permissible error. This error is specified as $0.015\,\% P_{W,i} + 0.010\,\% P_{\max}$ where $P_{\max} = I_{\max}U_{\max} = 12\,000\,\text{W}$. $P_{W,i}$ is modelled using the rectangular distribution with mean $P_{W,i}$ and semi-width the error thus computed.

The standard uncertainty associated with the phase displacement $F_{D,i}$ according to IEC 60076-19 [362] is evaluated as follows. The relative standard uncertainty is approximated by

$$u_{\text{rel}}(F_{D,i}) = \tan\varphi_i \, u(\Delta_{\varphi V,i} - \Delta_{\varphi C,i}), \tag{E4.4.9}$$

where [362]

$$u(\Delta_{\varphi V,i} - \Delta_{\varphi C,i}) = \sqrt{u^2(\Delta_{\varphi V,i}) + u^2(\Delta_{\varphi C,i})}.$$

The standard uncertainties $u(\Delta_{\varphi V,i})$ and $u(\Delta_{\varphi C,i})$ are obtained from the rectangular distribution with zero mean and the maximum errors specified in table E4.4.2.

In the uncertainty evaluation described in IEC 60076-19 [362], the uncertainty associated with φ_i is not considered. When applying the law of propagation of uncertainty, the sensitivity coefficient equals $\Delta_{\varphi V,i} - \Delta_{\varphi C,i}$ and is zero in this particular case (the values of $\Delta_{\varphi V,i}$ and $\Delta_{\varphi C,i}$ are both

zero). The standard uncertainty associated with φ_i can be obtained from equation (E4.4.5). The derivative of $\arccos x$ is $-1/\sqrt{1-x^2}$ [59] and that of $\tan x$ is $\cos^{-2} x$ [59]. So,

$$u^2(\tan\varphi_i) \approx \frac{1}{\cos^4\varphi_i} u^2(\varphi_i)$$

and

$$u^2(\varphi_i) \approx \frac{1}{1-[P_W/(I_{\mathrm{M},i}U_{\mathrm{M},i}))]^2} u^2\left(\frac{P_{\mathrm{W},i}}{I_{\mathrm{M},i}U_{\mathrm{M},i}}\right) + u^2(\Delta_{\varphi\mathrm{C},i}) + u^2(\Delta_{\varphi\mathrm{V},i})$$

and

$$u^2_{\mathrm{rel}}\left(\frac{P_{\mathrm{W},i}}{I_{\mathrm{M},i}U_{\mathrm{M},i}}\right) = u^2_{\mathrm{rel}}(P_{\mathrm{W},i}) + u^2_{\mathrm{rel}}(I_{\mathrm{M},i}) + u^2_{\mathrm{rel}}(U_{\mathrm{M},i}),$$

which are obtained using the law of propagation of uncertainty from the GUM [2]. Given that both $\tan x$ and $\arccos x$ are strongly non-linear functions, the results are only approximate; hence the $\approx$ sign. Strictly speaking, this applies also to all expressions for the relative standard uncertainty based on clause 5.1.6 of the GUM.

The uncertainty associated with the current $I_{\mathrm{M},i}$ is evaluated as follows. The specification of the current measurement is that the maximum permissible error equals $0.01\,\% I_{\mathrm{M},i} + 0.02\,\% I_{\max}$. The current is modelled using the rectangular distribution with centred at $I_{\mathrm{M},i}$ with semi-width equal to the maximum permissible error.

The standard uncertainty of the voltage measurement $U_{\mathrm{M},i}$ is based on the specification for the maximum permissible error. This error is specified as $0.010\,\% U_{\mathrm{M},i} + 0.018\,\% U_{\max}$. $U_{\mathrm{M},i}$ is modelled using the rectangular distribution centred at $U_{\mathrm{M},i}$ with semi-width equal to the error thus computed.

Use of the specifications provides the relative standard uncertainties in table E4.4.4. This uncertainty component can be included in the uncertainty budget by approximating the sensitivity coefficient numerically using the definition of a derivative. In that case, the sensitivity coefficient is 1 (the partial derivatives are respectively $\partial\Delta_{\varphi\mathrm{C},i}/\partial\Delta_{\varphi\mathrm{C},i}$ and $\partial\Delta_{\varphi\mathrm{V},i}/\partial\Delta_{\varphi\mathrm{V},i}$).

Table E4.4.4: Uncertainty evaluation of $\tan\varphi_i$ for the three phases based on equation (E4.4.5). Values in percentages are relative standard uncertainties; other values are absolute standard uncertainties

Source	Unit	Phase		
		U	V	W
CT max. phase displacement	(rad)	0.001 679	0.001 679	0.001 679
VT max. phase displacement	(rad)	0.001 679	0.001 679	0.001 679
Phase current unc	%	0.020	0.020	0.020
Measured active power unc	%	0.101	0.100	0.100
Voltage phase-neutral unc	%	0.019	0.019	0.019
$u(\cos\varphi)$		0.002 377	0.002 377	0.002 377
$u(\tan\varphi)$		0.26	0.26	0.26

E4.4.6 Propagation of uncertainty

E4.4.6.1 Law of propagation of uncertainty

The propagation of uncertainty for the power measured at load loss P_2 (equation (E4.4.4)) is performed using the formula for a measurement model as a pure product (see GUM clause 5.16 [2]). Substituting the values and specifications provides the following relative uncertainties (table E4.4.5) for the three phases.

Table E4.4.5: Uncertainty evaluation for the measured load loss P_2 at ambient temperature (adapted from [362])

Quantity	Component	u_{rel}/%			Exponent
		U	V	W	
CT ratio error	ε_{C}	0.115	0.115	0.115	1
VT ratio error	ε_{V}	0.115	0.115	0.115	1
Measured power	P_{W}	0.101	0.100	0.100	1
Phase displacement	F_{D}	2.454	2.454	2.454	1
Ampere meter	I_{M}	0.040	0.040	0.040	2
Load loss	P_2	2.930	2.928	2.926	

The values of the measured and additional load loss obtained according to equations (E4.4.1) and (E4.4.2) are shown in table E4.4.6.

Table E4.4.6: Values of measured load loss and additional load loss at temperature θ_2 and load losses at temperature θ_r

Quantity	Component	Value /W		
		U	V	W
Measured load loss at θ_2	P_2	1671.7	1719.1	1698.8
Additional load loss	P_{a2}	170.4	296.7	107.5
$I_{\mathrm{N}}^2 R$ loss	$I_{\mathrm{N}}^2 R$	1501.3	1422.4	1591.3
Load loss at θ_r	P_{LL}	2198.7	2181.0	2277.5

It should be noted that both the primary and secondary current and resistance of the windings are taken into account when calculating additional load loss.

The sensitivity coefficients for calculation of the uncertainty contributions of the uncertainty of resistance at temperature θ_2 are computed in accordance with equation (E4.4.3) as

$$\frac{\partial R_2}{\partial R_1} = \frac{t+\theta_2}{t+\theta_1}, \tag{E4.4.10}$$

$$\frac{\partial R_2}{\partial \theta_1} = -\frac{R_2}{t+\theta_1}, \tag{E4.4.11}$$

$$\frac{\partial R_2}{\partial \theta_2} = \frac{R_2}{t+\theta_2}. \tag{E4.4.12}$$

The sensitivity coefficient as the result from the mathematical formulation (E4.4.2) of the measured load loss at ambient temperature equals 1, whereas its standard uncertainty is obtained from the relative standard uncertainties given in table E4.4.5. The sensitivity coefficient for the resistance of the windings at temperature θ_2 is equal to the squared rated current for both the primary and secondary coil. Both quantities, temperature during resistance measurement and load loss measurement, have rectangular distributions with mean 0 K and semi-width 1 K. The uncertainty contributions of each component are given in table E4.4.7.

Table E4.4.7: Uncertainty of additional load loss at temperature θ_2

Quantity	Component	Uncertainty contribution /W (sensitivity coeff. × standard unc.)		
		U	V	W
Resistance HV at θ_1	$R_{1,\mathrm{HV}}$	2.83×10^{-5}	2.88×10^{-5}	2.94×10^{-5}
Resistance LV at θ_1	$R_{1,\mathrm{LV}}$	9.23×10^{-7}	8.65×10^{-7}	9.80×10^{-7}
Temp. resistance meas. (HV)	θ_1	-1.09×10^{-4}	-1.12×10^{-4}	-1.14×10^{-4}
Temp. resistance meas. (LV)	θ_1	-3.59×10^{-6}	-3.36×10^{-6}	-3.81×10^{-6}
Temp. load loss meas. (HV)	θ_2	1.10×10^{-4}	1.12×10^{-4}	1.14×10^{-4}
Temp. load loss meas. (LV)	θ_2	3.59×10^{-6}	3.37×10^{-6}	3.82×10^{-6}
Measured load loss at θ_2	P_2	2398.91	2534.03	2471.27
Resistance HV at θ_2	$R_{2,\mathrm{HV}}$	0.58	0.59	0.60
Resistance LV at θ_2	$R_{2,\mathrm{HV}}$	4.27	4.00	4.53
Additional load loss at θ_2	$u(P_{\mathrm{a}2})$	49.17	50.50	49.92

Using equation (E4.4.1) and the computed values given in the table E4.4.6 the uncertainty of load loss at reference temperature θ_r is calculated for all three phases. The values in table E4.4.8 are obtained by taking into account the sensitivity coefficient for each quantity in the measurement model (E4.4.1), except the reference temperature and indicated currents which are considered to be constant.

Table E4.4.8: Uncertainty of load losses reported at reference temperature θ_r

Quantity	Component	Uncertainty /W		
		U/W	V/W	W/W
Reported load loss	$u(P_{\mathrm{LL}})$	35.89	36.81	36.46
Expanded uncertainty of reported load loss ($k = 2$)	$U(P_{\mathrm{LL}})$	71.77	73.63	72.93

The total load loss, computed as the load loss of the three phases U, V, and W, is 6657 W with standard uncertainty 63 W.

E4.4.6.2 Monte Carlo method

The implementation of the propagation of distributions is carried using the measurement model as given in equations (E4.4.1), (E4.4.2), (E4.4.4) and (E4.4.5). As to the input quantities, the same probability density functions are used as for the evaluation using the law of propagation of uncertainty.

For the purpose of describing the R software used for the evaluation, the variables holding the measured values and some constants are declared below, where their names are largely self-explanatory (table E4.4.9. Finally, the measured values for resistance, power, current and voltage are declared similarly.

Table E4.4.9: Cross reference of symbols and variable names used in the R code

Variable	Symbol	Variable in R code
Offset for winding material	t	`tW`
Rated maximum current	$I_{\max}$	`I_max`
Rated maximum voltage	$U_{\max}$	`U_max`
Rated maximum power	$P_{\max}$	`P_max`
Temperature during resistance measurement	θ_1	`theta1.val`
Temperature during load loss measurement	θ_2	`theta2.val`
Reference temperature	θ_{ref}	`theta_r`
CT ratio	k_{CN}	`k_CN`
VT ratio	k_{VN}	`k_VN`
CT maximum amplitude relative error	ε_{C}	`eps_C.val`
VT maximum amplitude relative error	ε_{V}	`eps_V.val`
Half-width of the rectangular distribution of ε_{C}		`eps_C.hw`
Half-width of the rectangular distribution of ε_{V}		`eps_V.hw`
Currents in the high-voltage and low-voltage circuit	I_{NHV}	texttttI_NHV
Currents in the high-voltage and low-voltage circuit	I_{NLV}	`I_NLV`

```
t_W = 235
I_max = 50
U_max = 240
P_max = I_max*U_max

theta1.val = 22.1 # rectangular distribution, half-width = 1 K
theta2.val = 21.8 # rectangular distribution, half-width = 1 K
theta_r = 120
k_CN = 1
k_VN = 1
eps_C.val = 0.0 # rectangular distribution, half-width = 0.2% of k_CN
eps_C.hw = 0.002
eps_V.val = 0.0 # rectangular distribution, half-width = 0.2% of k_VN
eps_V.hw = 0.002
I_NHV = 60.62
I_NLV = 909.33
Delta_C = 0.0 # rectangular distribution, half-width = 10 arcmin
Delta_C.hw = 10/(60*180)*pi
Delta_V = 0.0 # rectangular distribution, half-width = 10 arcmin
Delta_V.hw = 10/(60*180)*pi

# measured resistances for the three phases
R_HV = c(0.049,0.0500,0.0510)
# at theta1 ... (rectangular, half-width 0.1% relative)
R_LV = c(0.001600,0.001500,0.001700)
# at theta1 ... (rectangular, half-width 0.1% relative)

# measured power, current and voltage for the three phases
P_W = c(748,756,762) # rectangular distribution
I_M = c(40.55,40.2,40.6) # rectangular distribution
U_M = c(191.5,195.2,194.8) # rectangular distribution
```

The three functions needed to calculate the load loss are coded as follows (see code below). The constants are declared as variables with a default value in the function heading, which enables calling the function without having to include these variables (R will then use their default values). This approach makes the code less cluttered.

```
# power P_2
P2func <- function(k_CN=1,eps_C=0,k_VN=1,eps_V=0,P_W,Delta_C=0,Delta_V=0,
I_M,U_M,I_N) {
phi = acos(P_W/(I_M*U_M)) - (Delta_V-Delta_C)
k_CN*(1/(1+eps_C))*k_VN*1/(1+eps_V)*P_W/
(1-(Delta_V-Delta_C)*tan(phi))*(I_N/(k_CN*I_M))^2
}

# temperature correction
temp.corr <- function(R,theta1,theta2,t_W = 235) {
R*(t_W+theta2)/(t_W+theta1)
}

# load loss for a single phase
PLL.func <- function(I_NHV,I_NLV,R_2HV,R_2LV,P2,theta2,theta_r = theta_r,
t_W = 235) {
I_NHV^2*temp.corr(R_2HV,theta2 = theta_r,theta1 = theta2)+I_NLV^2*
temp.corr(R_2LV,theta2 = theta_r,theta1 = theta2) +
(P2-I_NHV^2*R_2HV-I_NLV^2*R_2LV)*(t_W+theta2)/(t_W+theta_r)
}
```

To implement the Monte Carlo method, the R function `runif()`, which provides a random variate with a rectangular distribution, is recast so that it can be called with a central value and half-width as arguments. The number of trials M is set to 1×10^6. First, the variables are generated for all phases, followed by generating the data for the phases U, V, and W. The phase angle `phi` (ϕ) is calculated as described in IEC 60076-19 [362].

The measurement model in this application is quite involved in that it contains several steps. Also, it should be noted that the summation over the three phases is not necessarily a summation of three independent quantities, that is, the load losses for the phases U, V, and W are correlated. All these aspects are taken care of by the Monte Carlo method, as outlined below.

```
rrect <- function(num,middle=0,hw=1) {
runif(n=num,min=middle-hw,middle+hw)
}

# MC data
M = 1000000

# all phases
theta1 = rrect(M,theta1.val,1.0)
theta2 = rrect(M,theta2.val,1.0)
eps_C = rrect(M,eps_C.val,eps_C.hw)
eps_V = rrect(M,eps_V.val,eps_V.hw)

# phase U
Delta_C.U = rrect(M,0.0,10*pi/(60*180)) # use specified half-width
Delta_V.U = rrect(M,0.0,10*pi/(60*180))
R_HV.U = rrect(M,R_HV[1],0.001*R_HV[1])
R_LV.U = rrect(M,R_LV[1],0.001*R_LV[1])
U_M.U = rrect(M,U_M[1],0.00010*U_M[1]+0.00018*U_max)
I_M.U = rrect(M,I_M[1],0.00010*I_M[1]+0.00020*I_max)
```

```
P_W.U = rrect(M,P_W[1],0.00015*P_W[1]+0.00010*P_max)

R_2HV.U = temp.corr(R_HV.U,theta1 = theta1,theta2 = theta2)
R_2LV.U = temp.corr(R_LV.U,theta1 = theta1,theta2 = theta2)

phi = acos(P_W.U/(I_M.U*U_M.U)) - (Delta_V.U-Delta_C.U)
F_D.U = 1/(1-(Delta_V.U-Delta_C.U)*tan(phi))

P2.U = P2func(eps_C = eps_C, eps_V = eps_V,P_W = P_W.U,Delta_C = Delta_C.U,
Delta_V = Delta_V.U,I_M = I_M.U,U_M = U_M.U,I_N = I_NHV)

PLL.U = PLL.func(I_NHV = I_NHV,I_NLV = I_NLV,R_2HV = R_2HV.U,R_2LV = R_2LV.U,
P2 = P2.U,theta2 = theta2, theta_r = theta_r)

# phase V
Delta_C.V = rrect(M,0.0,10*pi/(60*180)) # use specified half-width
Delta_V.V = rrect(M,0.0,10*pi/(60*180))
R_HV.V = rrect(M,R_HV[2],0.001*R_HV[2])
R_LV.V = rrect(M,R_LV[2],0.001*R_LV[2])
U_M.V = rrect(M,U_M[2],0.00010*U_M[2]+0.00018*U_max)
I_M.V = rrect(M,I_M[2],0.00010*I_M[2]+0.00020*I_max)
P_W.V = rrect(M,P_W[2],0.00015*P_W[2]+0.00010*P_max)

phi = acos(P_W.V/(I_M.V*U_M.V)) - (Delta_V.V-Delta_C.V)
F_D.V = 1/(1-(Delta_V.V-Delta_C.V)*tan(phi))

P2.V = P2func(eps_C = eps_C, eps_V = eps_V,P_W = P_W.V,Delta_C = Delta_C.V,
Delta_V= Delta_V.V,I_M = I_M.V,U_M = U_M.V,I_N = I_NHV)

R_2HV.V = temp.corr(R_HV.V,theta1 = theta1,theta2 = theta2)
R_2LV.V = temp.corr(R_LV.V,theta1 = theta1,theta2 = theta2)

PLL.V = PLL.func(I_NHV = I_NHV,I_NLV = I_NLV,R_2HV = R_2HV.V,R_2LV = R_2LV.V,
P2 = P2.V,theta2 = theta2, theta_r = theta_r)

# phase W
Delta_C.W = rrect(M,0.0,10*pi/(60*180)) # use specified half-width
Delta_V.W = rrect(M,0.0,10*pi/(60*180))
R_HV.W = rrect(M,R_HV[3],0.001*R_HV[3])
R_LV.W = rrect(M,R_LV[3],0.001*R_LV[3])
U_M.W = rrect(M,U_M[3],0.00010*U_M[3]+0.00018*U_max)
I_M.W = rrect(M,I_M[3],0.00010*I_M[3]+0.00020*I_max)
P_W.W = rrect(M,P_W[3],0.00015*P_W[3]+0.00010*P_max)

R_2HV.W = temp.corr(R_HV.W,theta1 = theta1,theta2 = theta2)
R_2LV.W = temp.corr(R_LV.W,theta1 = theta1,theta2 = theta2)

phi = acos(P_W.W/(I_M.W*U_M.W)) - (Delta_V.W-Delta_C.W)
F_D.W = 1/(1-(Delta_V.W-Delta_C.W)*tan(phi))

P2.W = P2func(eps_C = eps_C, eps_V = eps_V,P_W = P_W.W,Delta_C = Delta_C.W,
Delta_V = Delta_V.W,I_M = I_M.W,U_M = U_M.W,I_N = I_NHV)


PLL.W = PLL.func(I_NHV = I_NHV,I_NLV = I_NLV,R_2HV = R_2HV.W,R_2LV = R_2LV.W,
P2 = P2.W,theta2 = theta2, theta_r = theta_r)

# total load loss
PLL.tot = PLL.U + PLL.V + PLL.W
```

```
#output
# c(mean(PLL.tot),sd(PLL.tot),quantile(PLL.tot,probs=c(0.025,0.975)))
PLL.tot.val = mean(PLL.tot)
PLL.tot.unc = sd(PLL.tot)
PLL.tot.Unc = 0.5*(quantile(PLL.tot,probs=c(0.975))-
quantile(PLL.tot,probs=c(0.025)))
PLL.tot.k = PLL.tot.Unc/PLL.tot.unc
```

The load loss is 6657 W with standard uncertainty 53 W, about 15 % smaller than that obtained with the GUM uncertainty framework. The expanded uncertainty is obtained using the `quantile()` function in R as shown above. The coverage factor is obtained by dividing the expanded uncertainty by the standard uncertainty. The expanded uncertainty is 104 W with a coverage factor $k = 1.95$. The relevant guidance can be found in GUM-S1 [3, clause 7.6] and GUM-S2 [4, clause 7.6].

The assessment of the strength of the dependencies between the measurement results of the three phases U, V, and W can be readily accomplished in R, using the output of the Monte Carlo method. First, the vectors `PLL.U`, `PLL.V` and `PLL.W` are combined into a matrix called `PLL.matrix`. Then, using the R-function `cor`, the correlation matrix is obtained:

```
PLL.matrix = cbind(PLL.U,PLL.V,PLL.W)
cor(PLL.matrix)

##            PLL.U      PLL.V      PLL.W
## PLL.U 1.00000000 0.01333439 0.01479030
## PLL.V 0.01333439 1.00000000 0.01474804
## PLL.W 0.01479030 0.01474804 1.00000000
```

From the correlation matrix thus obtained, it is evident that the correlation coefficients between phases are very small (all positive and less than 0.015, compared with unity for fully correlated values). Hence, summing the results of the three phases as if they were independent (uncorrelated) is justified.

E4.4.7 Reporting the result

According to the Monte Carlo method, the load loss is 6657 W with standard uncertainty 53 W. The expanded uncertainty is 104 W and the coverage factor, computed as the ratio of the expanded uncertainty and standard uncertainty is 1.95.

The GUM uncertainty framework gives the same estimate, namely 6657 W, as the Monte Carlo method and a somewhat larger standard uncertainty of 63 W. Assuming a normal distribution yields for 95 % coverage probability a coverage factor of 1.96, and thus an expanded uncertainty of 124 W.

E4.4.8 Interpretation of results

The outcomes of the GUM uncertainty framework and the Monte Carlo method agree quite well. The Monte Carlo method enables, as demonstrated, to assess the validity of the assumption made when applying the law of propagation of uncertainty that the results of the three phases can be treated as independent.

An improvement of the treatment could be the consideration that the results of the three phases are dependent (due to using the same equipment for the measurement of the electrical quantities of the three phases).

For the data used, the Monte Carlo method, regarded by many (such as [96, 263–266]) as a 'gold standard' for uncertainty propagation, indeed validates the GUM results although the latter can be regarded as moderately conservative. In this instance, the Monte Carlo method is also used to justify treating the losses over the three phases as independent (see also equation (E4.4.1)).

Example E4.5

Evaluation of measurement uncertainty in thermal comfort

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

E4.5.1 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 the field of thermal comfort and, besides the application and uncertainty evaluation, the limitations of this standard will also be discussed.

E4.5.2 Introduction of the application

The example, which is based in part on the paper [363], 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 [364], 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.

E4.5.3 Specification of the measurand(s)

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

$$\begin{aligned}\mathrm{PMV} =& [0.303\exp(-0.036M)+0.028] \\ &\times \left\{M-W-3.05\times10^{-3}\left[5733-6.99(M-W)-\rho_{\mathrm{a}}\right]\right. \\ &-0.42(M-W-58.15)-1.7\cdot10^{-5}(5867-\rho_{\mathrm{a}})M-0.0014(34-t_{\mathrm{a}})M \\ &\left.-3.96\cdot10^{-8}\left[(t_{\mathrm{cl}}+273)^4-(t_{\mathrm{r}}+273)^4\right]f_{\mathrm{cl}}-(t_{\mathrm{cl}}-t_{\mathrm{a}})f_{\mathrm{cl}}h_{\mathrm{c}}(t_{\mathrm{cl}})\right\},\end{aligned} \tag{E4.5.1}$$

where

M	metabolic rate in $\mathrm{W\,m^{-2}}$
W	effective mechanical power in $\mathrm{W\,m^{-2}}$
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 $\mathrm{W\,m^{-2}\,K^{-1}}$

with the main complication arising from the fact that the quantity t_{cl} is defined implicitly (see expression (E4.5.2) 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 $\mathrm{m^2\,K\,W^{-1}}$), and (b) quantities $t_{\mathrm{a}}, t_{\mathrm{r}}, v_{\mathrm{ar}}$ and RH obtained by measurement (v_{ar} is the relative air velocity in $\mathrm{m\,s^{-1}}$ 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.

E4.5.4 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:

$$\begin{aligned}t_{\mathrm{cl}} =& 35.7-0.028(M-W) \\ &-I_{\mathrm{cl}}\left\{3.96\cdot10^{-8}\left[(t_{\mathrm{cl}}+273)^4-(t_{\mathrm{r}}+273)^4\right]f_{\mathrm{cl}}+(t_{\mathrm{cl}}-t_{\mathrm{a}})f_{\mathrm{cl}}h_{\mathrm{c}}(t_{\mathrm{cl}})\right\},\end{aligned} \tag{E4.5.2}$$

where the convective heat transfer coefficient is

$$h_{\mathrm{c}}(t_{\mathrm{cl}}) = \max\left(2.38\,|t_{\mathrm{cl}}-t_{\mathrm{a}}|^{1/4}, 12.1\sqrt{v_{\mathrm{ar}}}\right), \tag{E4.5.3}$$

which is mathematically identical to the expression given in the standard [364], 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:

$$f_{cl} = \left\{ \begin{array}{ll} 1.00 + 1.290 I_{cl}, & I_{cl} \leq 0.0775\,\mathrm{m^2\,K\,W^{-1}} \\ 1.05 + 0.645 I_{cl}, & I_{cl} > 0.0775\,\mathrm{m^2\,K\,W^{-1}} \end{array} \right\}. \tag{E4.5.4}$$

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 $\mathrm{RH} = 100\rho_a/\rho_s(t_a)$, using

$$\rho_a = \mathrm{RH} \times \rho_s(t_a)\,10\,\mathrm{Pa}, \tag{E4.5.5}$$

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

$$\rho_s(t_a) = \exp\left(16.6536 - \frac{4030.183}{t_a + 235}\right)\mathrm{K\,Pa}. \tag{E4.5.6}$$

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

$$g(t_{cl}) = 3.96 \cdot 10^{-8}\left[(t_{cl}+273)^4 - (t_r+273)^4\right] f_{cl} + (t_{cl} - t_a) f_{cl} h_c(t_{cl}), \tag{E4.5.7}$$

is common to equations (E4.5.1) and (E4.5.2). 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 (E4.5.2) can be expressed as

$$t_{cl} = 35.7 - 0.028(M - W) - I_{cl} g(t_{cl}), \tag{E4.5.8}$$

and equation (E4.5.1) can be simplified accordingly to

$$\begin{aligned} \mathrm{PMV} = & [0.303\exp(-0.036\mathrm{M}) + 0.028] \\ & \times \left\{ M - W - 3.05 \times 10^{-3}\left[5733 - 6.99(M-W) - \rho_a\right] \right. \\ & - 0.42(M - W - 58.15) - 1.7 \cdot 10^{-5}(5867 - \rho_a) M - 0.0014(34 - t_a) M \\ & \left. - \left[35.7 - 0.028(M-W) - t_{cl}\right]/I_{cl} \right\}. \end{aligned} \tag{E4.5.9}$$

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

Concerning the associated uncertainties, the metabolic rate M, the effective power W and the clothing insulation 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.

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

In table E4.5.1, 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.

Table E4.5.1: Estimates of the input quantities

Location	$M \times 58.2$ $\mathrm{W\,m^{-2}}$	W $\mathrm{W\,m^{-2}}$	$I_{cl} \times 0.155$ $\mathrm{m^2\,K\,W^{-1}}$	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 E4.5.2, 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 E4.5.2: Standard uncertainties associated with the estimates of the measured quantities in table E4.5.1

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 E4.5.2 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-S1 [3]. In the former the sample standard deviation is used as the standard uncertainty whereas in the latter a factor of $[(n-1)/(n-3)]^{1/2}$ is included to obtain the standard deviation of the state-of-knowledge distribution assigned to the corresponding input quantity. In this case, for a sample size of $n = 40$ the effect is small (less than 3 %), not affecting the standard uncertainties in table E4.5.2 to the number of decimal digits reported.

E4.5.5 Uncertainty propagation

E4.5.5.1 Preamble

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

Both approaches depend on knowledge of the PDFs for the input quantities, but whereas the GUF 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.

E4.5.5.2 GUM Uncertainty Framework

The GUF 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 (E4.5.8). As for many complicated models, determining the required partial derivatives algebraically is not always practical and a numerical approach is recommended [365]. It is a burden not shared by the GUM-S1 approach.

In this model, special attention must be paid to the derivative $\partial(\mathrm{PMV})/\partial t_{\mathrm{cl}}$ given values for all other input quantities. Equation (E4.5.8) (with equation (E4.5.7)) is solved for t_{cl}, and then PMV is evaluated using equation (E4.5.9). The partial derivatives of PMV with respect to the input quantities, required by the 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 (E4.5.1) involves $g(t_{\mathrm{cl}})$, defined by equation (E4.5.7), and hence the derivative $g'(t_{\mathrm{cl}})$ is required. This derivative is not necessary when using equation (E4.5.8), is a simplification that poses no problem since $g(t_{\mathrm{cl}})$ is already used when solving equation (E4.5.8) 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 (E4.5.4)) and to the convective heat transfer coefficient h_{c} (equation (E4.5.3)).

The first of these model functions (the second is similar) is illustrated in Figure E4.5.1 , showing f_{cl} as a function of clothing insulation I_{cl}. Equation (E4.5.4) can be also be expressed as

$$f_{\mathrm{cl}} = 1.1 + \min\left[0.645\,(I_{\mathrm{cl}} - 0.0775)\,, 1.290\,(I_{\mathrm{cl}} - 0.0775)\right],$$

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

Figure E4.5.2 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 [366], applying the formula separately in both

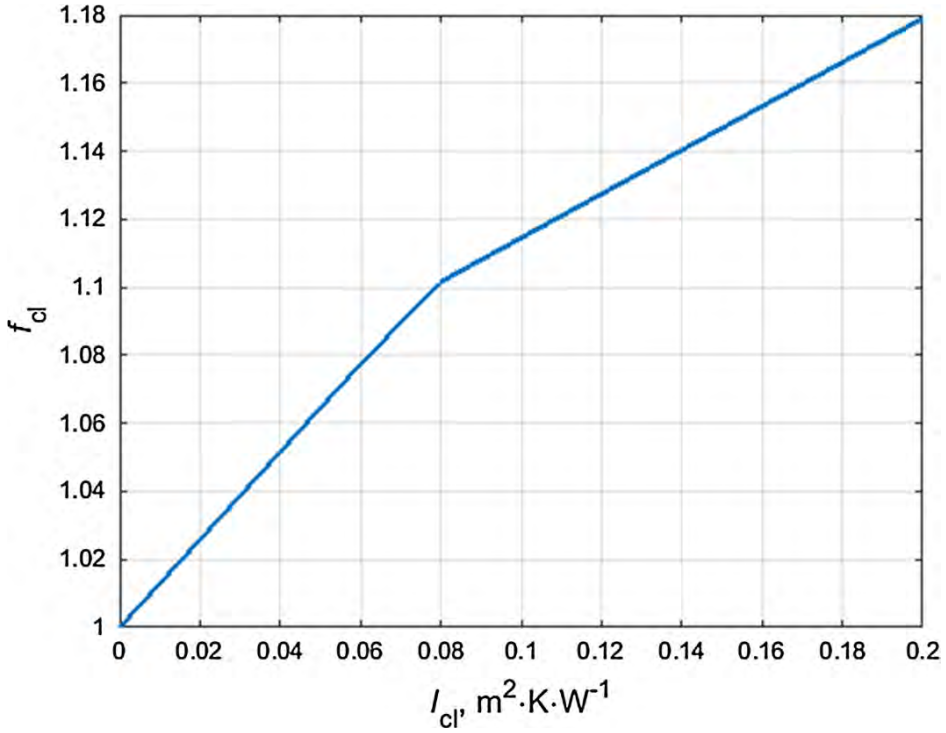


Figure E4.5.1: Clothing surface area factor f_{cl} as a function of clothing insulation I_{cl}

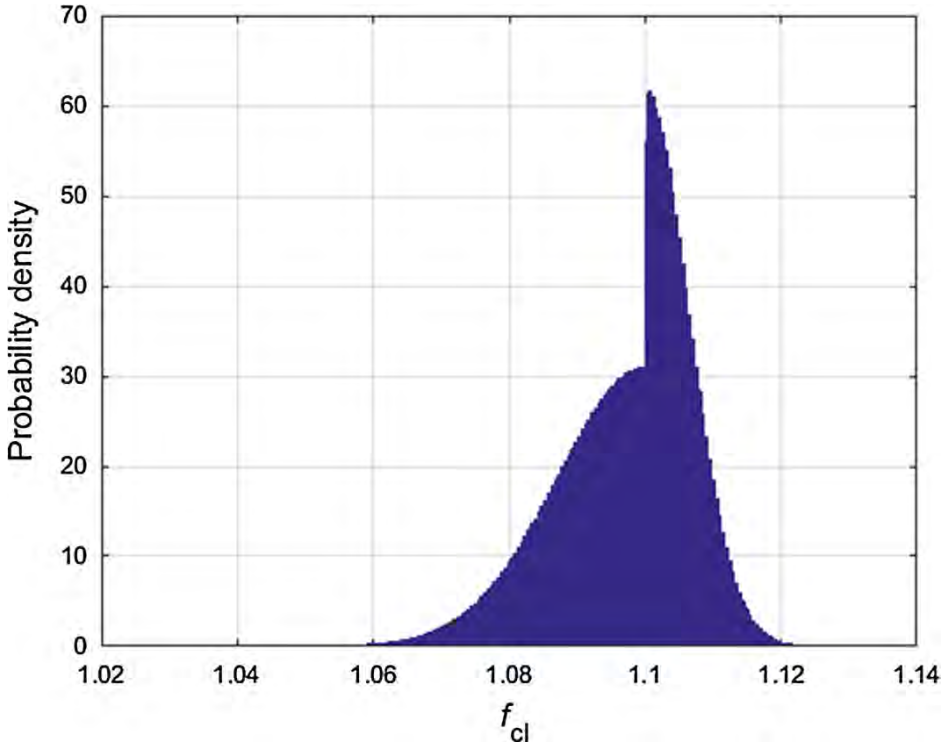


Figure E4.5.2: PDF for f_{cl} as a function of I_{cl}, as produced by a Monte Carlo calculation

branches. The PDF obtained agrees to graphical accuracy with that provided by the Monte Carlo calculation. The PDF in this case is discontinuous, in fact a mixture of two "half-Gaussians", with standard deviations in the ratio 2:1.

Even when the measurement model involves derivative discontinuities, as here, the standard deviation of the measurand (f_{cl} or subsequently PMV index) is a continuous function of the input quantities in the model. Figure 3 shows the standard deviation of f_{cl} (equal to the standard uncertainty associated with an estimate of f_{cl}) for I_{cl} ranging from 0.001 $m^2\,W\,K^{-1}$ to 0.200 $m^2\,W\,K^{-1}$. The smooth but rapid change of the standard deviation from 0.0129 to 0.0065 over this interval is apparent. Convection plays a predominant role in the thermal comfort perception, as expected.

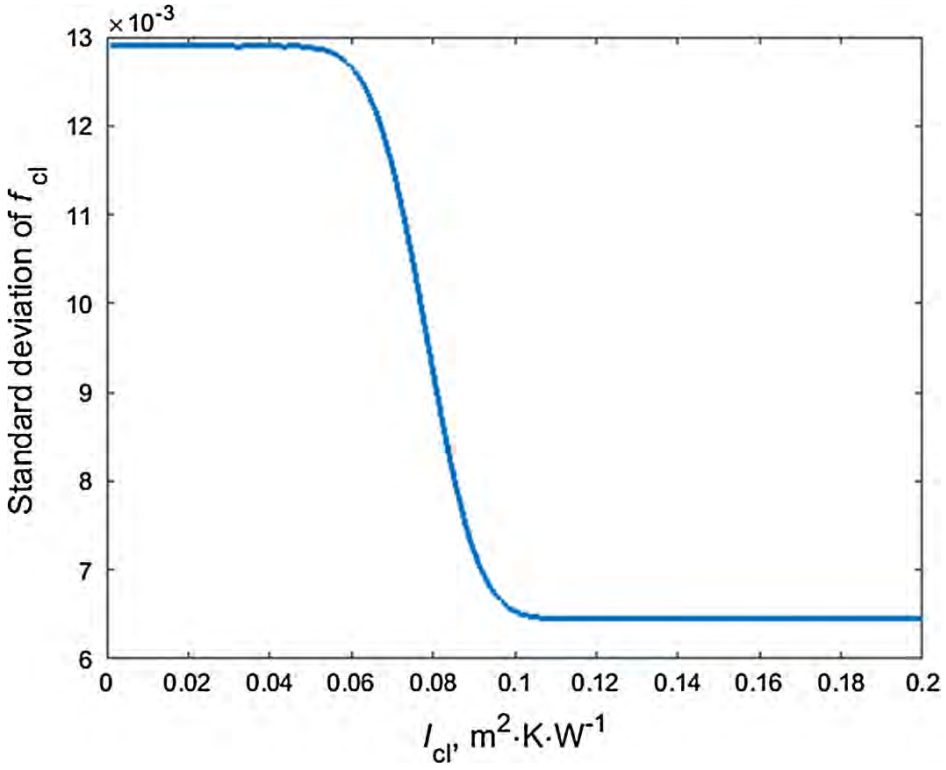


Figure E4.5.3: Standard deviation of f_{cl} as a function of I_{cl}

E4.5.6 Reporting the result

Table E4.5.3 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 experimental, while other quantities have tabulated values [364] for which values are regarded as fixed and exact. Together with the above partial derivatives evaluated at the estimates of the input quantities, LPU is applied to produce results given in the table. In the case of t_a the PDF results from the combination of data from measurement and instrument calibration.

A relevant conclusion that can be drawn from table E4.5.3 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,

Table E4.5.3: 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	$u(\mathrm{PMV}) = 0.094$		$U_{0.95}(\mathrm{PMV}) = 0.18$

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 (E4.5.1).

The Monte Carlo calculation was based on the experimental input indicated in tables E4.5.1 and E4.5.2. The number of Monte Carlo trials was taken as 1×10^5. Thus, samples of 1×10^5 drawn from the PDFs for the input quantities were used to obtain a PDF for the output quantity as described in GUM-S1 [3].

Figure E4.5.4 shows the PDF for the PMV index provided by the MCM for the same location as table E4.5.3. 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.

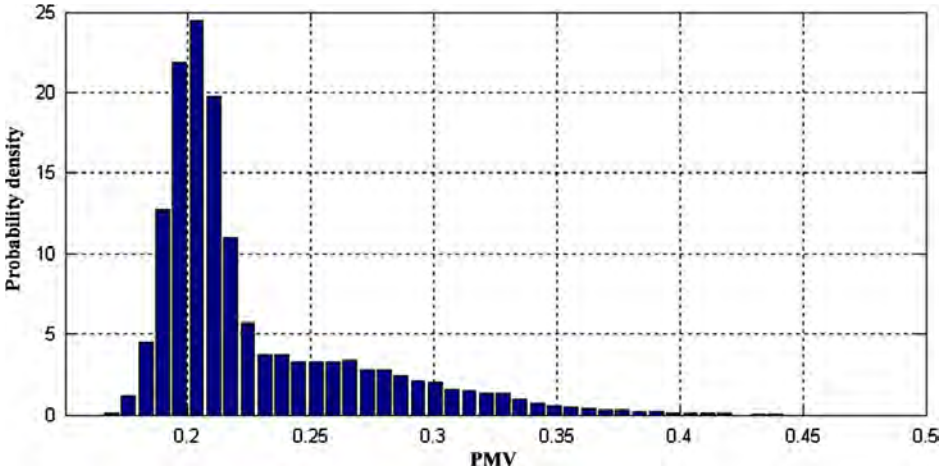


Figure E4.5.4: PDF for the PMV index from Monte Carlo calculation

The advantages of MCM over GUF 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. The MCM 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 E4.5.4 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 for-

mally give the expectation (mean) of the distribution. The mode (point at which the probability density is greatest) might be more meaningful, but it is recommended that only the expectation is used for purposes of uncertainty propagation [2].

Table E4.5.4: 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 [364]. 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 MCM. In relation to the relative influence of uncertainty contributions, the sensitivity analysis, assuming the Gaussian PDF $\mathrm{N}(1.2, (0.05)^2)$ for M and the Gaussian PDF $\mathrm{N}(0.7, (0.05)^2)$ 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 [364], 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 [364], 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 GUF and MCM, 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 [364]. 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 [6].

E4.5.7 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 [364], but also of ISO 7726 [367], 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 [368], and close collaboration should therefore be pursued with researchers from the social sciences.

With respect to ISO 7726 [367], one study has shown [369] 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.

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

E4.5.A.1 Explicit Model

In the more common explicit univariate measurement model, a single real output quantity Y is related to a number of input quantities $\boldsymbol{X} = (X_1, \ldots, X_N)^\top$ by a functional relationship f in the form $Y = f(\boldsymbol{X})$ as stated in the GUM [364]. The estimate of the output quantity is taken as $y = f(\boldsymbol{x})$ The standard uncertainty $u(y)$ associated with y is evaluated from

$$u^2(y) = \sum_{i=1}^{N} \sum_{j=1}^{N} c_i u\left(x_i, x_j\right) c_j \tag{E4.5.10}$$

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 $u\left(x_i, x_j\right)$ the covariance associated with x_i and x_j

A compact way of writing the sum in expression (E4.5.10), better suited for scientific software based on matrix formulation, e.g., MATLAB, is

$$u^2(y) = \boldsymbol{c}^{\mathrm{T}} \boldsymbol{V}_{\boldsymbol{x}} \boldsymbol{c} \tag{E4.5.11}$$

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

$$\boldsymbol{V_x} = \begin{bmatrix} u(x_1, x_1) & \cdots & u(x_1, x_N) \\ \vdots & \ddots & \vdots \\ u(x_N, x_1) & \cdots & u(x_N, x_N) \end{bmatrix} \tag{E4.5.12}$$

and the (row) vector $\boldsymbol{c}^\top = \lceil c_1, \ldots, c_N \rceil$ of dimension $1 \times N$ contains the sensitivity coefficients. Both expressions (E4.5.10) and (E4.5.11) are equivalent representations of LPU of the GUM [364].

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

$$u^2(y) = \sum_{i=1}^{N} [c_i u(x_i)]^2 = \sum_{i=1}^{N} u_i^2(y), \quad u_i(y) \equiv |c_i| u(x_i) \tag{E4.5.13}$$

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.

E4.5.A.2 Implicit Model

For an implicit univariate measurement model, however, a single output quantity Y is related to real input quantities $\boldsymbol{X}$ in a way that cannot readily or stably be represented by a direct functional relationship. Instead, a model for the measurement takes the form $h(Y, \boldsymbol{X}) = 0$, in which Y is not expressed directly as a function of $\boldsymbol{X}$, often requiring a numerical implementation to obtain a solution [3].

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 [368]. The standard uncertainty $u(y)$ associated with y is evaluated from

$$u^2(y) c_y^2 = \boldsymbol{c_x}^\top \boldsymbol{U_x} \boldsymbol{c_x} \tag{E4.5.14}$$

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 $Y = y$ [364].

E4.5.A.3 Conditions for valid application

There are a number of conditions for valid application of the GUF for non-linear models. They include [2] 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.

E4.5.A.4 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 MCM 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 GUF, 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 GUF is questionable. It should always be checked that any given target uncertainty has been attained [366], which is a further feature the GUF 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 an MCM 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 [2].

The GUF does not provide the PDF for Y, but instead assumes that Y can be described by a Gaussian PDF $\mathrm{N}\left(\mathrm{y}, u^2(y)\right)$, 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 [363].

Example E4.6

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

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

E4.6.1 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`.

E4.6.2 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 [370]. This form of (in)homogeneity accounts for the (small) differences in the property of interest between the bottles (or more generally, items [371]) 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 [370, 372]. The evaluation of between-bottle homogeneity is both a requirement in reference material production [8] as well as in proficiency testing [9, 69].

Traditionally, classical ANOVA is used [69, 371] for this purpose, which is more fully described elsewhere [370]. The parameter of prime interest is the between-group standard deviation, which in this specific case is called the between-bottle standard deviation [64, 370]. 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 [64]. Such situations should be avoided [370, 371] but that is not always possible [64].

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 [370]. The background of the model is briefly summarised here; a more elaborate treatise is available elsewhere [63, 64]. A similar model is also available in the NIST Consensus Builder [373, 374].

E4.6.3 Specification of the measurand(s)

The measurand in this example is the between-bottle standard deviation (τ) of the amount fraction of a component in a batch of gas mixtures.

E4.6.4 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 [370]

$$y_{ij} = \mu + B_i + \varepsilon_{ij}. \tag{E4.6.1}$$

The objective of the evaluation is to determine $\tau^2 = \mathrm{var}(B_i)$ and $\sigma^2 = \mathrm{var}(\varepsilon_{ij})$. If no pooling is used, then $\sigma_i^2 = \mathrm{var}(\varepsilon_{ij})$, 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 [64]. 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 PDF should be chosen for the model parameters. In the case that the parameters are assumed to be mutually independent, then this joint prior PDF 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

$$\mu \sim \mathrm{N}(\mu_0, \sigma^2_{\mathrm{target}}), \tag{E4.6.2}$$

$$\tau \sim \mathrm{Cauchy}(0, \tau_0), \tag{E4.6.3}$$

$$\sigma \sim \mathrm{Cauchy}(0, \sigma_0). \tag{E4.6.4}$$

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 [63, 64].

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 [375]. 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 [63]

$$y_{ij}|\mu,\tau,\sigma \sim \mathrm{N}(\mu,\tau^2+\sigma^2). \tag{E4.6.5}$$

The likelihood of $\bar{y}_i|\theta_i,\sigma_i$ can be described as [65]

$$\bar{y}_i|\theta_i \sim \mathrm{N}(\theta_i,\sigma^2/k),$$

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

$$\bar{y}_i|\mu,\tau \sim \mathrm{N}(\mu,\tau^2+\sigma^2/k).$$

E4.6.5 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 (E4.6.2)-(E4.6.4)), 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.

E4.6.6 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 that are 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 the user of the model to transfer the original data, and as we will see the model also

Table E4.6.1: 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

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 [61, 62].

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 [65]. These group means are declared in the block `transformed parameters`.

The data for ethane are shown in table E4.6.1. Using traditional ANOVA, 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} [64].

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 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 [63, 64] 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,
```

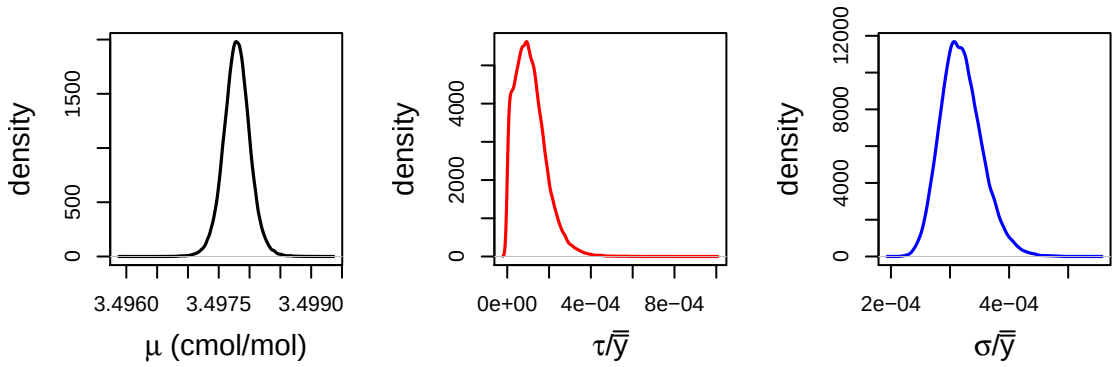


Figure E4.6.1: Posterior densities for the mean, between-group and within-group standard deviations for ethane

```
## and Rhat is the potential scale reduction factor on split chains (at
## convergence, Rhat=1).
```

The posterior PDFs for μ, τ, and σ computed from the dataset of ethane are shown in figure E4.6.1. 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 [376]. Then the function `hdi` is used to compute the highest posterior density intervals [377]. The lower (L) and upper (H) limits of the 95 % highest posterior density intervals are shown in table E4.6.2. The lower bound on the coverage interval for τ is 1.13×10^{-10} cmol mol^{-1}, which is very close to zero.

As the uncertainty evaluation in accordance with the GUM [2] and ISO Guide 35 [371] 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 E4.6.1), these three options are not equivalent. The mode is 3.22 μmol mol^{-1}, the median is 3.54 μmol mol^{-1}, and the mean is 3.89 μmol mol^{-1}.

Table E4.6.2: 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

The data for nitrogen are shown in table E4.6.3. Using traditional ANOVA for the dataset of nitrogen, the between-bottle standard deviation is 0.00 μmol mol^{-1} and the (pooled) repeatability standard deviation is 7.00 μmol mol^{-1} [64]. 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 $MS_{\text{within}} = 4.903 \times 10^{-7}$.

Table E4.6.3: 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

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 [63, 64] 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 PDFs for μ, τ, and σ computed from the dataset of nitrogen are shown in figure E4.6.2. 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 E4.6.4. The lower (L) and upper (H) limits of the 95 % highest posterior density intervals are shown in table E4.6.4. The lower bound on the coverage interval for τ is 2.18×10^{-9} cmol mol^{-1}.

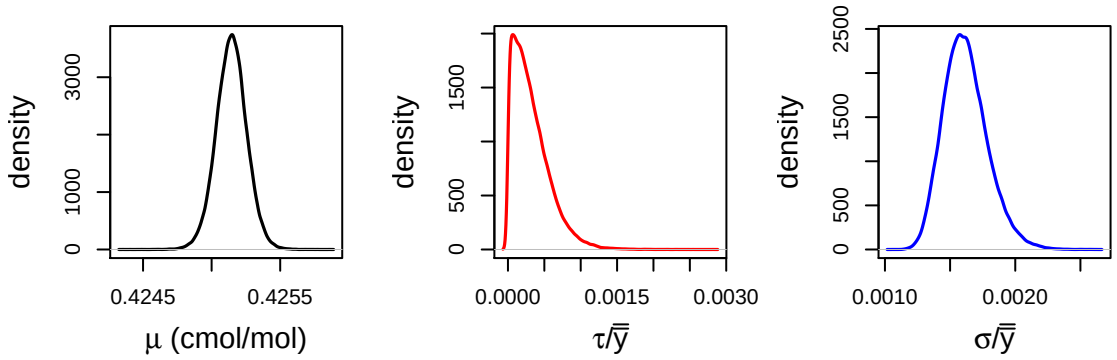


Figure E4.6.2: Posterior densities for the mean, between-group and within-group standard deviations for nitrogen

Table E4.6.4: 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 μmol mol^{-1}, the median is 1.12 μmol mol^{-1}, and the mean is 1.36 μmol mol^{-1}.

E4.6.7 Reporting the result

The prime result is the value for the between-bottle standard deviation τ. In a previous paper [64], 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 [370, 378] 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 E4.7

Flow meter calibration using the master meter method

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

E4.7.1 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.

E4.7.2 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 E4.7.1). 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

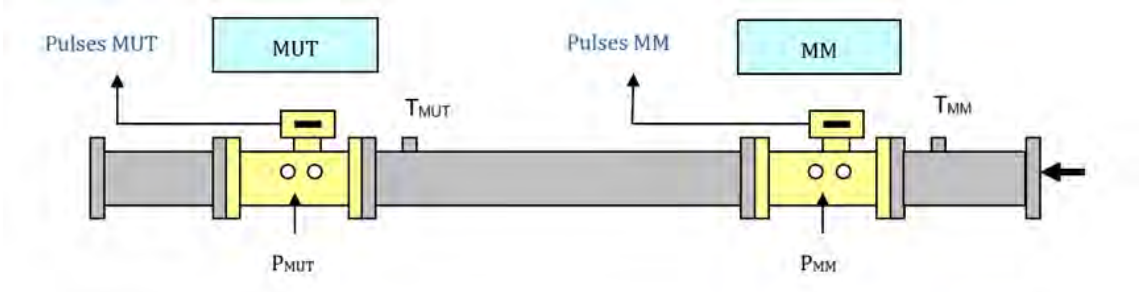


Figure E4.7.1: Set-up in the SARAJEVOGAS laboratory

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 E4.7.1 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.

E4.7.3 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 E4.7.1.

Table E4.7.1: Volume flow rate ranges of the meters involved

G40 Rotary gas meter	G250 Turbine gas meter	G1000 Turbine gas meter
$20\,m^3\,h^{-1}$–$50\,m^3\,h^{-1}$	$100\,m^3\,h^{-1}$–$350\,m^3\,h^{-1}$	$450\,m^3\,h^{-1}$–$1000\,m^3\,h^{-1}$

E4.7.4 Measurement model

E4.7.4.1 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 [379]:

$$e = \frac{V_{\mathrm{MUT}} - V_{\mathrm{REF}}}{V_{\mathrm{REF}}} \tag{E4.7.1}$$

$$= \frac{V_{\mathrm{MUT}}}{V_{\mathrm{REF}}} - 1, \tag{E4.7.2}$$

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 [380]

$$pV = ZRT, \tag{E4.7.3}$$

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 [379]:

$$\frac{p_{\mathrm{MUT}} V_{\mathrm{REF}}}{T_{\mathrm{MUT}}} = \frac{p_{\mathrm{MM}} V_{\mathrm{MM}}}{T_{\mathrm{MM}}} = ZR, \tag{E4.7.4}$$

$$V_{\mathrm{REF}} = V_{\mathrm{MM}} \frac{p_{\mathrm{MM}}}{p_{\mathrm{MUT}}} \frac{T_{\mathrm{MUT}}}{T_{\mathrm{MM}}}, \tag{E4.7.5}$$

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 (E4.7.5) into equation (E4.7.2),

$$e = \frac{V_{\mathrm{MUT}}}{V_{\mathrm{MM}}} \frac{p_{\mathrm{MUT}}}{p_{\mathrm{MM}}} \frac{T_{\mathrm{MM}}}{T_{\mathrm{MUT}}} - 1. \tag{E4.7.6}$$

By using a slightly different notation from that in [379] 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:

$$V_{\mathrm{MUT}} = \frac{I_{\mathrm{MUT}}}{K_{\mathrm{MUT}}}, \tag{E4.7.7}$$

$$V_{\mathrm{MM}} = \frac{I_{\mathrm{MM}}}{K_{\mathrm{MM}}(1 + f_{\mathrm{MM}})}, \tag{E4.7.8}$$

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 (E4.7.6) becomes

$$e = \frac{I_{\mathrm{MUT}}}{I_{\mathrm{MM}}} \frac{K_{\mathrm{MM}}(1 + f_{\mathrm{MM}})}{K_{\mathrm{MUT}}} \frac{p_{\mathrm{MUT}}}{p_{\mathrm{MM}}} \frac{T_{\mathrm{MM}}}{T_{\mathrm{MUT}}} - 1. \tag{E4.7.9}$$

E4.7.4.2 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.

METER UNDER TEST

- Unknown characteristics.

LABORATORY

- 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:

$$e_{flow} = e + Q, \tag{E4.7.10}$$

where Q denotes other influencing quantities and

$$Q = Q_{MM} + Q_{REP} + Q_{AUX}. \tag{E4.7.11}$$

E4.7.5 Uncertainty propagation

Uncertainty propagation follows the procedure described within GUM [2] (although we validate the results obtained using the propagation of distributions in section E4.7.7). The measurement model used for the uncertainty propagation is described by equation (E4.7.10), 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 (E4.7.10) can be expressed using the law of propagation of uncertainty [2] as follows:

$$u^2(e_{flow}) = \left[\left(\frac{\partial e_{flow}}{\partial e} u(e) \right)^2 + \left(\frac{\partial e_{flow}}{\partial Q} u(Q) \right)^2 \right], \tag{E4.7.12}$$

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.

E4.7.5.1 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 (E4.7.9), and can be expressed as follows:

$$\begin{aligned} u^2(e) = &\left[\frac{\partial e}{\partial I_{\mathrm{MUT}}} u(I_{\mathrm{MUT}})\right]^2 + \left[\frac{\partial e}{\partial I_{\mathrm{MM}}} u(I_{\mathrm{MM}})\right]^2 + \left[\frac{\partial e}{\partial p_{\mathrm{MUT}}} u(p_{\mathrm{MUT}})\right]^2 \\ &+ \left[\frac{\partial e}{\partial p_{\mathrm{MM}}} u(p_{\mathrm{MM}})\right]^2 + \left[\frac{\partial e}{\partial T_{\mathrm{MUT}}} u(T_{\mathrm{MUT}})\right]^2 + \left[\frac{\partial e}{\partial T_{\mathrm{MM}}} u(T_{\mathrm{MM}})\right]^2. \end{aligned} \tag{E4.7.13}$$

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 (E4.7.9), let

$$S = e + 1 = \frac{I_{\mathrm{MUT}}}{I_{\mathrm{MM}}} \frac{K_{\mathrm{MM}}(1 + f_{\mathrm{MM}})}{K_{\mathrm{MUT}}} \frac{p_{\mathrm{MUT}}}{p_{\mathrm{MM}}} \frac{T_{\mathrm{MM}}}{T_{\mathrm{MUT}}}.$$

Then, sensitivity coefficients are calculated from equation (E4.7.9) as follows:

$$\frac{\partial e}{\partial I_{\mathrm{MUT}}} = \frac{S}{I_{\mathrm{MUT}}}, \tag{E4.7.14}$$

$$\frac{\partial e}{\partial I_{\mathrm{MM}}} = -\frac{S}{I_{\mathrm{MM}}}, \tag{E4.7.15}$$

$$\frac{\partial e}{\partial p_{\mathrm{MUT}}} = \frac{S}{p_{\mathrm{MUT}}}, \tag{E4.7.16}$$

$$\frac{\partial e}{\partial p_{\mathrm{MM}}} = -\frac{S}{p_{\mathrm{MM}}}, \tag{E4.7.17}$$

$$\frac{\partial e}{\partial T_{\mathrm{MM}}} = \frac{S}{T_{\mathrm{MM}}}, \tag{E4.7.18}$$

$$\frac{\partial e}{\partial T_{\mathrm{MUT}}} = -\frac{S}{T_{\mathrm{MUT}}}. \tag{E4.7.19}$$

The standard measurement uncertainty of the following individual measurement quantities can be determined as the standard deviations of the according rectangular probability distributions:

$$u(e_{n_i}) = \frac{L_{n_i}}{\sqrt{3}}, \tag{E4.7.20}$$

$$u(e_{p_i}) = \frac{L_{p_i}}{\sqrt{3}}, \tag{E4.7.21}$$

$$u(e_{T_i}) = \frac{L_{T_i}}{\sqrt{3}}, \tag{E4.7.22}$$

where

- e_{n_i} is the presumed error due to reading the number of pulses on the measuring instrument. According to [2] 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 $L_{p_i} = 0.2\,\mathrm{mbar}$,

– 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 $L_{T_{\mathrm{MM}}} = 0.132\,\mathrm{K}$ and for the MUT $L_{T_{\mathrm{MUT}}} = 0.163\,\mathrm{K}$.

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 (E4.7.14)–(E4.7.19).

E4.7.5.2 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 [2]. The repeatability of measurement is calculated using

$$u_{\mathrm{REP}} = \frac{s}{\sqrt{n}}, \tag{E4.7.23}$$

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.

$$u_{\mathrm{MM}} = \frac{U_{\mathrm{MM}}}{k}, \tag{E4.7.24}$$

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.

$$u_{\mathrm{AUX}} = \frac{e_{\mathrm{AUX}}}{\sqrt{3}}, \tag{E4.7.25}$$

where e_{AUX} is the estimated error, which is usually bounded by $|e_{\mathrm{AUX}}| \leq 0.1\,\%$.

Since the sensitivity coefficients in equation (E4.7.12) have the value 1,

$$u(e_{\mathrm{flow}}) = \sqrt{u^2(e) + u^2_{\mathrm{REP}} + u^2_{\mathrm{MM}} + u^2_{\mathrm{AUX}}}. \tag{E4.7.26}$$

E4.7.6 Reporting the result

The measurements were carried out according to the requirements set out in standards [381] and [382] for turbine and rotary gas meters. During the calibration process the measurement data presented in tables E4.7.2 and E4.7.3 were obtained.

Table E4.7.2: 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 E4.7.3: 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 E4.7.2 and E4.7.3 '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 E4.7.2 and E4.7.3. The error of the measuring instrument for this measuring point can be calculated according to equation (E4.7.9) as follows:

$$e = \left(\frac{6413.67}{115\,029.33} \frac{8100(1+0.1171/100)}{450.238} \frac{954.96}{956.67} \frac{295.06}{295.07} - 1 \right) \times 100\,\% \tag{E4.7.27}$$

$$= 0.2441\,\%. \tag{E4.7.28}$$

The standard measurement uncertainty of the measurement error of the MUT can be calculated from equation (E4.7.13). However, it is necessary first to calculate the sensitivity coefficients of each influencing parameter, which can be carried out according to formulæ (E4.7.14) and (E4.7.15). The value of S was calculated as 100.243.

Sensitivity coefficients are calculated as follows:

$$\frac{\partial e}{\partial I_{\mathrm{MUT}}} = \frac{100.243}{6413.67} = 0.0156, \tag{E4.7.29}$$

$$\frac{\partial e}{\partial I_{\mathrm{MM}}} = -\frac{100.243}{115\,029.33} = -0.000\,871, \tag{E4.7.30}$$

$$\frac{\partial e}{\partial p_{\mathrm{MUT}}} = \frac{100.243}{954.96} = 0.149\,70, \tag{E4.7.31}$$

$$\frac{\partial e}{\partial p_{\mathrm{MM}}} = -\frac{100.243}{956.67} = -0.147\,80, \tag{E4.7.32}$$

$$\frac{\partial e}{\partial T_{\mathrm{MM}}} = \frac{100.243}{295.06} = 0.339\,74, \tag{E4.7.33}$$

$$\frac{\partial e}{\partial T_{\mathrm{MUT}}} = -\frac{100.243}{295.07} = -0.339\,729. \tag{E4.7.34}$$

From equations (E4.7.20)–(E4.7.22) and information on errors given in subsection E4.7.5.1, we obtain

$$u(I_i) = \frac{1}{\sqrt{3}} = 0.5780, \tag{E4.7.35}$$

$$u(p_i) = \frac{0.2}{\sqrt{3}} = 0.1156, \tag{E4.7.36}$$

$$u(T_{MUT}) = \frac{0.163}{\sqrt{3}} = 0.0942 \tag{E4.7.37}$$

$$u(T_{MM}) = \frac{0.132}{\sqrt{3}} = 0.0763. \tag{E4.7.38}$$

Results (E4.7.35)–(E4.7.37) combined with equation (E4.7.13) give the standard uncertainty of the measuring system measurement error:

$$u(e) = 0.045\,\%. \tag{E4.7.39}$$

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:

$$u_{\mathrm{REP}} = 0.0018\,\%. \tag{E4.7.40}$$

The standard measurement uncertainty u_{MM} of the standard used -– master meter – is calculated according to equation (E4.7.24), where the the expanded measurement uncertainty of the MM (obtained from the calibration certificate) is $U_{\mathrm{MM}} = 0.25\,\%$, giving

$$u_{\mathrm{MM}} = \frac{0.25}{2} \tag{E4.7.41}$$

$$= 0.125\,\%. \tag{E4.7.42}$$

The relative standard uncertainty u_{AUX} of other influence factors can be calculated according to equation (E4.7.25), where it is assumed the estimated error is $e_{\mathrm{AUX}} = 0.1\,\%$ and to have a rectangular probability distribution:

$$u_{\mathrm{AUX}} = \frac{0.1}{\sqrt{3}}\,\% \tag{E4.7.43}$$

$$= 0.0578\,\% \tag{E4.7.44}$$

By substituting the uncertainty contributions obtained in (E4.7.40), (E4.7.42), (E4.7.39) and (E4.7.44) into equation (E4.7.26), the combined relative standard uncertainty becomes

$$u(e_{\text{flow}}) = (0.045^2 + 0.0018^2 + 0.125^2 + 0.0578^2)^{1/2}\,\%, \tag{E4.7.45}$$

$$u(e_{\text{flow}}) = 0.1449\,\%. \tag{E4.7.46}$$

The expanded measurement uncertainty, with coverage factor $k = 2$ is

$$U = 2 \times u(e_{\text{flow}}) \tag{E4.7.47}$$

$$= 0.29\,\%. \tag{E4.7.48}$$

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 E4.7.4.

Table E4.7.4: 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

E4.7.7 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 [2], 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-S1 [3] was used for the evaluation of measurement uncertainty for the measurement model described by equation (E4.7.10). For this method the number of Monte Carlo trials was prescribed to be 1×10^8. Values for the input quantities, their associated standard uncertainties and their probability distributions remained the same as described in section E4.7.6.

The results of the applied method are shown in table E4.7.5.

Table E4.7.5: 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 E4.7.2 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) [3]. It can be noticed, both in table E4.7.5 and from figure E4.7.2 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.

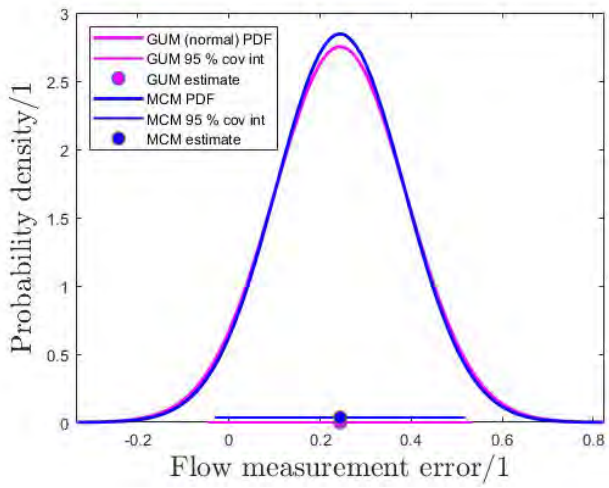


Figure E4.7.2: Probability density functions for Monte Carlo and GUM approach

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.

Acknowledgement

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.

Example E4.8

Pressure drop measurement

M. Čaušević, M.G. Cox, J. Greenwood

E4.8.1 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.

E4.8.2 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 [383].

E4.8.3 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.

E4.8.4 Measurement model

E4.8.4.1 Measurement principle

The measurement model in this example originates from the gas equation, which is the equation of state of the 'ideal' gas [380]:

$$pV = nRT, \tag{E4.8.1}$$

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 (E4.8.1) 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

$$n_1 = \frac{Vp_1}{RT_1} \tag{E4.8.2}$$

and accordingly the amount of gas at the end time (t_2) is

$$n_2 = \frac{Vp_2}{RT_2}. \tag{E4.8.3}$$

Loss of material in the form of gas escaping from the system is therefore provided by the difference

$$\Delta n = n_2 - n_1 = \frac{V}{R}\left(\frac{p_2}{T_2} - \frac{p_1}{T_1}\right). \tag{E4.8.4}$$

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:

$$\Delta p_\mathrm{s} = \Delta n \frac{R}{V} T_\mathrm{s}. \tag{E4.8.5}$$

By substituting equation (E4.8.4) into (E4.8.5) we obtain

$$\Delta p_\mathrm{s} = \left(\frac{p_2}{T_2} - \frac{p_1}{T_1}\right) T_\mathrm{s}. \tag{E4.8.6}$$

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 (E4.8.6) becomes

$$\Delta p_\mathrm{s} = \left(\frac{p_2}{T_2} - \frac{p_1}{T_1}\right) T_\mathrm{s} \frac{\Delta t_\mathrm{s}}{\Delta t}. \tag{E4.8.7}$$

For test items of a predefined volume, pressure change is often a sufficient measurand. Therefore, expression (E4.8.7) 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

$$Q_m = \frac{V}{MR}\left(\frac{p_2}{T_2} - \frac{p_1}{T_1}\right)\frac{1}{\Delta t}, \tag{E4.8.8}$$

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:

$$Q_v = V\left(\frac{p_2}{T_2} - \frac{p_1}{T_1}\right)\frac{T_s}{p_s}. \tag{E4.8.9}$$

The arguments presented shortly apply equally well to flow rate measurement equations (E4.8.8) and (E4.8.9).

E4.8.4.2 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 (E4.8.7) in terms of T_1 and the temperature *difference*

$$\Delta T = T_2 - T_1, \tag{E4.8.10}$$

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,

$$u(\Delta T) \ll u(T_1). \tag{E4.8.11}$$

Making use of (E4.8.10) and (E4.8.11), expression (E4.8.7) becomes

$$\Delta p_s = \left(\frac{T_1 p_2 - T_2 p_1}{T_2 T_1} \right) \frac{T_s}{\Delta t} \Delta t_s \tag{E4.8.12}$$

$$\approx \left(\frac{p_2 - p_1}{T_1} - \frac{\Delta T p_1}{{T_1}^2} \right) \frac{T_s}{\Delta t} \Delta t_s. \tag{E4.8.13}$$

E4.8.5 Uncertainty propagation

E4.8.5.1 Law of propagation of uncertainty

For the measurement uncertainty evaluation, the GUM [2] law of propagation of uncertainty (LPU) is applied for the two measurement equations (E4.8.7) and (E4.8.13). 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 (E4.8.13), 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 (E4.8.7) and (E4.8.13) 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.

E4.8.5.2 Ignoring temperature correlation

In this subsection correlation between temperature values is ignored, that is, it is taken as zero. For model (E4.8.7) the sensitivity coefficients are

$$\frac{\partial \Delta p_s}{\partial p_1} = -\frac{1}{T_1} \frac{T_s}{\Delta t} \Delta t_s = -\frac{T_s \Delta t_s}{T_1 \Delta t}, \tag{E4.8.14}$$

$$\frac{\partial \Delta p_s}{\partial p_2} = \frac{1}{T_2} \frac{T_s}{\Delta t} \Delta t_s = \frac{T_s \Delta t_s}{T_2 \Delta t}, \tag{E4.8.15}$$

$$\frac{\partial \Delta p_s}{\partial T_2} = -\frac{p_2}{{T_2}^2} \frac{T_s}{\Delta t} \Delta t_s = -\frac{p_2 T_s \Delta t_s}{{T_2}^2 \Delta t}, \tag{E4.8.16}$$

$$\frac{\partial \Delta p_s}{\partial T_1} = \frac{p_1}{{T_1}^2} \frac{T_s}{\Delta t} \Delta t_s = \frac{p_1 T_s \Delta t_s}{{T_1}^2 \Delta t}, \tag{E4.8.17}$$

$$\frac{\partial \Delta p_s}{\partial t_1} = \left(\frac{p_2}{T_2} - \frac{p_1}{T_1} \right) \frac{T_s}{(\Delta t)^2} \Delta t_s = \frac{\Delta p_s}{\Delta t}, \tag{E4.8.18}$$

$$\frac{\partial \Delta p_s}{\partial t_1} = \left(\frac{p_2}{T_2} - \frac{p_1}{T_1} \right) \frac{T_s}{(\Delta t)^2} \Delta t_s = -\frac{\Delta p_s}{\Delta t}. \tag{E4.8.19}$$

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

$$u(p_1) = \frac{U(p_1)}{2}, \tag{E4.8.20}$$

$$u(p_2) = \frac{U(p_2)}{2}, \tag{E4.8.21}$$

$$u(T_2) = \frac{a_T}{\sqrt{3}}, \tag{E4.8.22}$$

$$u(T_1) = \frac{a_T}{\sqrt{3}}, \tag{E4.8.23}$$

$$u(t_1) = \frac{a_t}{\sqrt{3}}, \tag{E4.8.24}$$

$$u(t_2) = \frac{a_t}{\sqrt{3}}. \tag{E4.8.25}$$

All the results obtained from equations (E4.8.14)–(E4.8.25) can be combined into a Type B measurement uncertainty evaluation for the pressure drop in the vessel:

$$\begin{aligned} u_B^2 = &\left[\frac{\partial \Delta p_s}{\partial p_1} u(p_1)\right]^2 + \left[\frac{\partial \Delta p_s}{\partial P_2} u(p_2)\right]^2 + \left[\frac{\partial \Delta p_s}{\partial T_2} u(T_2)\right]^2 \\ &+ \left[\frac{\partial \Delta p_s}{\partial T_1} u(T_1)\right]^2 + \left[\frac{\partial \Delta p_s}{\partial t_1} u(t_1)\right]^2 + \left[\frac{\partial \Delta p_s}{\partial t_2} u(t_2)\right]^2. \end{aligned} \tag{E4.8.26}$$

The combined measurement uncertainty u_c for the pressure drop in the vessel is obtained from

$$u_c^2 = u_A^2 + u_B^2. \tag{E4.8.27}$$

E4.8.5.3 Accounting for temperature correlation

In this subsection correlation between temperature values is taken into consideration through use of the model presented in equation (E4.8.13). The sensitivity coefficients are as follows:

$$\frac{\partial \Delta p_s}{\partial p_1} = -\left(\frac{1}{T_1} - \frac{\Delta T}{{T_1}^2}\right)\frac{T_s}{\Delta t}\Delta t_s = -\frac{(T_1 + \Delta T)T_s \Delta t_s}{{T_1}^2 \Delta t} \approx -\frac{T_s \Delta t_s}{T_1 \Delta t}, \tag{E4.8.28}$$

$$\frac{\partial \Delta p_s}{\partial p_2} = \frac{1}{T_1}\frac{T_s}{\Delta t}\Delta t_s = \frac{T_s \Delta t_s}{T_1 \Delta t}, \tag{E4.8.29}$$

$$\frac{\partial \Delta p_s}{\partial \Delta T} = -\frac{p_1}{{T_1}^2}\frac{T_s}{\Delta t}\Delta t_s = -\frac{p_1 T_s \Delta t_s}{{T_1}^2 \Delta t}, \tag{E4.8.30}$$

$$\frac{\partial \Delta p_s}{\partial T_1} = \left(\frac{p_1 - p_2}{{T_1}^2} + \frac{2T_1 \Delta T p_1}{{T_1}^4}\right)\frac{T_s}{\Delta t}\Delta t_s = \frac{T_s \Delta t_s [T_1(p_1 - p_2) + 2\Delta T p_1]}{{T_1}^3 \Delta t}, \tag{E4.8.31}$$

$$\frac{\partial \Delta p_s}{\partial t_1} = -\left(\frac{p_2 - p_1}{T_1} - \frac{\Delta T p_1}{{T_1}^2}\right)\frac{T_s}{(\Delta t)^2}\Delta t_s = -\frac{\Delta p_s}{\Delta t}, \tag{E4.8.32}$$

$$\frac{\partial \Delta p_s}{\partial t_2} = \left(\frac{p_2 - p_1}{T_1} - \frac{\Delta T p_1}{{T_1}^2}\right)\frac{T_s}{(\Delta t)^2}\Delta t_s = \frac{\Delta p_s}{\Delta t}. \tag{E4.8.33}$$

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,

$$u(p_1) = \frac{U(p_1)}{2}, \tag{E4.8.34}$$

$$u(p_2) = \frac{U(p_2)}{2}, \tag{E4.8.35}$$

$$u(\Delta T) = \frac{a_{\Delta T}}{\sqrt{6}}, \tag{E4.8.36}$$

$$u(T_1) = \frac{a_T}{\sqrt{3}}, \tag{E4.8.37}$$

$$u(t_1) = \frac{a_t}{\sqrt{3}}, \tag{E4.8.38}$$

$$u(t_2) = \frac{a_t}{\sqrt{3}}. \tag{E4.8.39}$$

All the values obtained from equations (E4.8.28)–(E4.8.39) can be combined into a Type B measurement uncertainty evaluation for the pressure drop in the vessel:

$$\begin{aligned} u_{\mathrm{B}}^2 = &\left[\frac{\partial \Delta p_{\mathrm{s}}}{\partial p_1} u(p_1)\right]^2 + \left[\frac{\partial \Delta p_{\mathrm{s}}}{\partial p_2} u(p_2)\right]^2 + \left[\frac{\partial \Delta p_{\mathrm{s}}}{\partial \Delta T} u(\Delta T)\right]^2 \\ &+ \left[\frac{\partial \Delta p_{\mathrm{s}}}{\partial T_1} u(T_1)\right]^2 + \left[\frac{\partial \Delta p_{\mathrm{s}}}{\partial t_1} u(t_1)\right]^2 + \left[\frac{\partial \Delta p_{\mathrm{s}}}{\partial t_2} u(t_2)\right]^2. \end{aligned} \tag{E4.8.40}$$

The combined standard uncertainty for the pressure drop in the vessel is obtained from equation (E4.8.27).

E4.8.6 Reporting the result

During the measurement the data in table E4.8.1 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 E4.8.2.

E4.8.7 Interpretation of results

It can be seen from the results in section E4.8.6 that both equations (E4.8.7) and (E4.8.13) 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

Table E4.8.1: Measured values including values of the pressure drop, assigned probability density functions and combined standard uncertainty based upon application of LPU to equation (E4.8.7)

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 E4.8.2: Measured values including values of the pressure drop, assigned probability density functions and combined standard uncertainty based upon application of LPU to equation (E4.8.13)

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

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 [2], the MCM described in GUM-S1 [3] was also used for the evaluation of the measurement uncertainty for the measurement models described by equations (E4.8.7) and (E4.8.13). 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 1×10^7. 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 E4.8.1 and E4.8.2 in section E4.8.6. The results of the GUM uncertainty framework and the Monte Carlo method are presented in tables E4.8.3 and E4.8.4. Table E4.8.3 shows results for the first approach in which the temperature correlation was ignored and table E4.8.4 shows results for the second approach.

Table E4.8.3: 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 E4.8.4: 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 E4.8.1 and E4.8.2 (note the different scales).

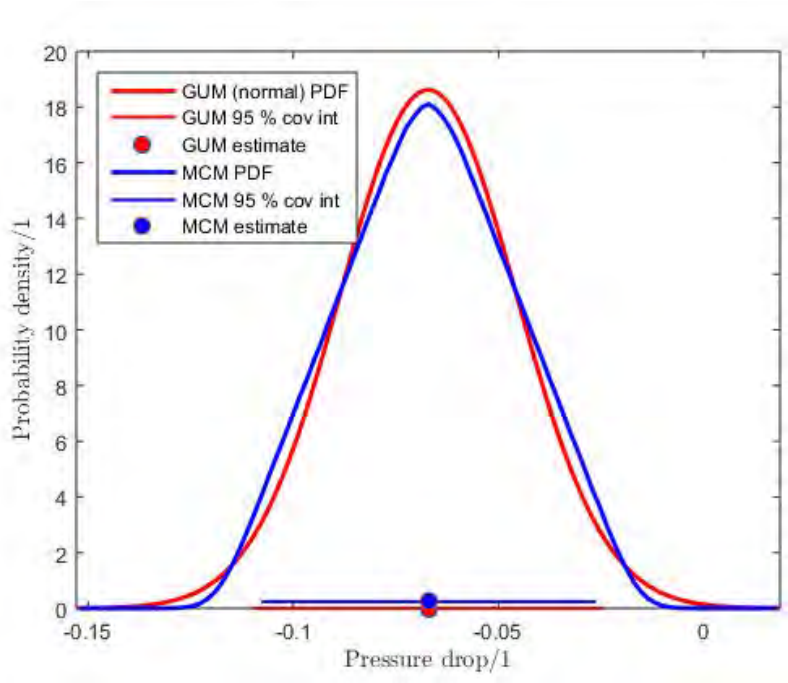


Figure E4.8.1: Probability density functions for pressure drop obtained by the GUM uncertainty framework and the Monte Carlo method not accounting for temperature correlation

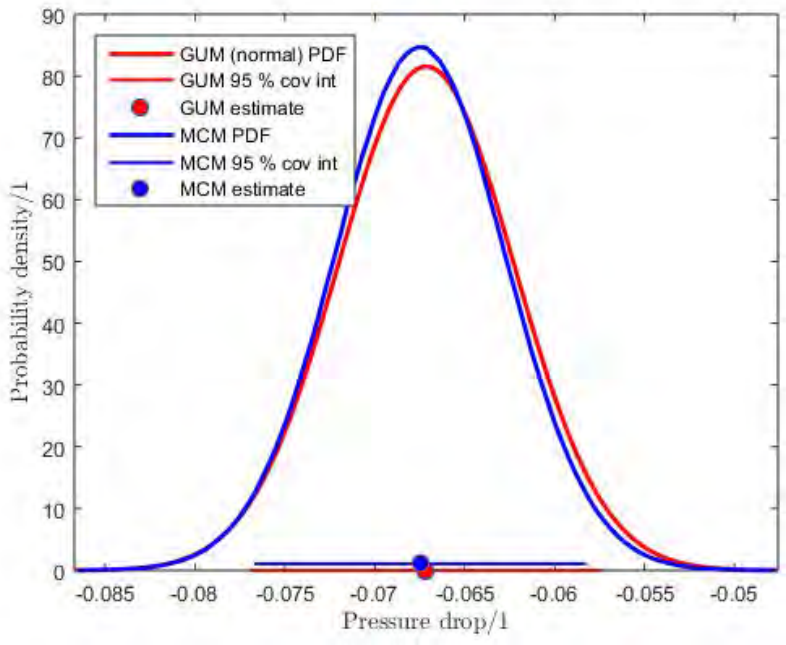


Figure E4.8.2: Probability density functions for pressure drop obtained by the GUM uncertainty framework and the Monte Carlo method accounting for temperature correlation (note the different scale from figure E4.8.1)

Figures E4.8.1 and E4.8.2 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 E4.8.1 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 E4.8.2 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 [2], 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 E4.8.3 and table E4.8.4). 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.

Part E5

Quality of life

Example E5.1

Effect of considering a 2D or 3D image as a set of pixels or voxels on a computed quantity

T. Caebergs, M.G. Cox

E5.1.1 Summary

Two examples are presented concerned with properties of a physical object extracted from a raster image. With more pixels (or voxels) to reproduce the image, a better description of the underlying physical object is generally obtained in terms of the estimate of the measurand and its associated uncertainty. However, the consequent increase in image resolution has a cost under different perspectives: time, money, susceptibility of image result to other instrumental effects. In some cases, an increase in resolution does not lead to appreciable improvement of the global result because of other uncertainty sources. A trade-off is often sought.

The first example relates to molecular radiotherapy where pixel size impacts on the area of a 2D section of an organ or tumour given by the delineation of a 2D image. Such a delineation is made by an operator or an algorithm and constitutes one step in the estimation of the volume of a 3D region (the ultimate measurand) corresponding to a reconstructed 3D image of a region of interest (ROI) such as from a set of parallel planar sections.

The second example relates to nanoparticle size in waste water, daily life products, etc., where AFM is used to measure nanoparticles deposited on a flat substrate. The size of a nanoparticle of interest is assessed as the height of the highest pixel in the pixels identified as belonging to the nanoparticle [175]. The example extends example E2.4, concerning the evaluation of measurement uncertainty for nanoparticle height measurements, and uses results from it. The example focuses on the model building and its interpretation by the practitioner for precise uncertainty evaluation or to keep uncertainty from this source under control. A frequentist approach is followed. The alternative use of a Bayesian treatment is briefly explained.

E5.1.2 Introduction of the first application: estimation of organ or tumour mass

The estimation of organ or tumour mass necessitates some form of outlining of the organ on an image. Under assumptions of homogeneity, organ mass is directly proportional to its volume and so it can be obtained using suitable imaging. Various outlining or segmenting techniques are used for this purpose [384].

The volume or mass of an organ or tumour is generally obtained from an ROI (region of interest) outlined on anatomical imaging data in the form of an array of pixels or voxels [385]. Knowing that outline, the corresponding volume of the ROI can be estimated. Approaches are considered for evaluating a measure of the outlining uncertainty, and the use of that measure to evaluate the uncertainty associated with estimated volume. The method used will depend largely on the information and resources available at the time of outlining and the method employed by the operator or operators to define the ROI. When determining volumes, an outline is typically drawn manually by a single operator across all images that comprise the data set of concern. Alternatively, automated methods are used to obtain the outlines.

In practice, a set of planar sections of the ROI, usually at a constant spacing, is considered, from which the volume is determined. Here, we pay attention to a single section and, rather than consider a volume, for simplicity of exposition consider an area as one step in the overall process.

In general, segmentation methods can be divided into the following categories: manual methods, threshold-based methods, boundary-based methods and stochastic- and machine learning-based methods. More details of these categories are given in [386] and the references therein. We here consider just boundary-based methods, such as gradient-based edge detection, which require the user to define an initial region of interest (ROI) inside which an algorithm estimates object boundaries [385].

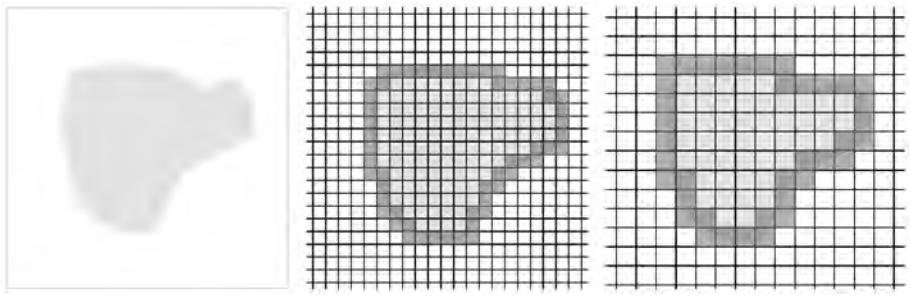
E5.1.3 Specification of the measurand(s)

The ultimate measurand is the volume of the tumour or organ. As stated, a simpler measurand is used here, the area of a section of the tumour or organ.

E5.1.4 Measurement model

Figure E5.1.1 shows an outline (a) of a tumour obtained from a pixellated image produced by scanning equipment, which we will use to provide numerical results. It also shows the effect of different pixel sizes (b) and (c). It is assumed that the actual outline of the tumour is contained within the darker-shaded pixels. These pixels would be chosen by an operator or some boundary detection algorithm. We take each pixel as of unit size (dimension 1×1).

The measurement model is defined by the algorithm used to compute the volume of the tumour or organ. Consistent with the specification of the measurand in section E5.1.3, a simpler measurement model is used, the area of a section of the tumour or organ given by a formula or an algorithm.



(a) (b) (c)

Figure E5.1.1: Uncertainty in outline definition for different voxel sizes (acknowledgment [385])

The area depends on knowledge of the boundary of the tumour, specifically a mathematical description of a curve representing the boundary for which we use a parametric representation:

$$x = F(t), \qquad y = G(t),$$

where the parameter t is ideally taken as the normalized distance along the curve from some starting point. Since the curve is unknown a priori, we use instead cumulative chord length based on the polygonal function joining successive data points as the parameter, as is common practice when working with such representations.

The area could alternatively be computed as some average (such as the arithmetic mean) of two extreme areas. One area would be obtained by taking the sum of the areas of the pixels shaded light grey in figures E5.1.1 (b) and (c). The other area would be the sum of the first area and the areas of all boundary pixels (dark grey). See section E5.1.5, which argues, with quantitative support, in favour of representing the boundary by a smooth curve.

Various empirical functions such as periodic splines and Fourier series can be used for F and G to form the profile. We select Fourier series based on our considerable experience [111, 387–389] with their use for data analysis. For this purpose, we adapt some guidance in [110] concerned with the determination and use of polynomial calibration functions. That standard gives advice on determining the degree of polynomial that is most appropriate for the data by first constructing polynomials of all degrees n from one up to some maximum $n_{\max}$ and then selecting a particular degree. Here we determine Fourier series involving an increasing number of harmonics, choosing the number we regard as suitable. By analogy with polynomials, we term the number of harmonics, as in other publications (for instance, [390]), the degree. Degree n has $2n+1$ defining parameters: a constant term and n cosine terms and n sine terms for the x and y components, respectively:

$$F(t) = b_1 + \sum_{r=1}^{n} (b_{2r} \cos 2\pi r t + b_{2r+1} \sin 2\pi r t), \tag{E5.1.1}$$

$$G(t) = c_1 + \sum_{r=1}^{n} (c_{2r} \cos 2\pi r t + c_{2r+1} \sin 2\pi r t).$$

The data comprises a set of m ordered x, y points (x_r, y_r), $r = 1, \ldots, m$. These points are regarded as values of the dependent variables x and y corresponding to an increasing set of values of the independent variable t. The values of t used as data are the midpoints of successive pixels containing the boundary. Fourier series in x and y, each having $n = 1, 2, \ldots, n_{\max}$ harmonics, are fitted to the data, the root-mean-square residual (RMSR) for each n being formed, where

$$\mathrm{RMSR}(n) = \left[\frac{\chi^2_{\mathrm{obs}}(n)}{m-n-1} \right]^{1/2}, \tag{E5.1.2}$$

which applies for $n < m-1$. Here $\chi^2_{\text{obs}}(n)$, also referred to as $S(n)$, is the residual sum of squares, taken over all x- and y-components for degree n. We consider separate RMSR values for the x- and y-components since we wish to determine whether there is any appreciable difference in the ability of Fourier series to represent these components.

Typical behaviour is that RMSR values initially show a tendency to decrease with increasing degree, saturating at some level indicative of a reasonable degree. As well as RMSR, various model-selection criteria, specifically, Akaike's Information Criterion (AIC), corrected AIC (AICc) and the Bayesian Information Criterion (BIC) [391] can be useful in providing a balance between goodness of fit and smoothness. For m data points and a polynomial model with $n+1$ parameters, these criteria are

$$\text{AIC}(n) = S(n)+2(n+1), \quad \text{AICc}(n) = \text{AIC}(n)+\frac{2(n+1)(n+2)}{m-n-2}, \quad \text{BIC}(n) = S(n)+(n+1)\ln m.$$

All three criteria are designed to balance goodness of fit and simplicity of model in terms of the number of parameters. Given a number of candidate models, here Fourier series of degrees $n = 1, \ldots, n_{\max}$, the model having the smallest value of AIC (or AICc or BIC) would usually be selected. According to [110], experience with polynomial calibration problems indicates that the same degree of polynomial is often selected by all three criteria, although there are some exceptions such as when the data set is small. On the other hand, it is difficult to automate the process of selecting from the sequence of RMSR values an appropriate degree. Traditionally, a degree is chosen, often by eye, when the above saturation takes place.

Account should of course be taken of uncertainties in the data. We know that the required curve passes through the boundary pixels. As such, an evaluation can be made of the standard uncertainties in the x- and y-coordinates of the centres of the pixels. Looking at pixels independently, since the pixels are taken as unit squares, the errors in these coordinates lie between $-1/2$ and $1/2$. If no other knowledge is available, it can be assumed that these errors follow rectangular distributions $\text{R}(-1/2, 1/2)$ having a standard deviation of $1/(2\sqrt{3}) \approx 0.29$, which is used as the according standard uncertainty. For the least-squares solution, weights all equal to the reciprocals of these standard uncertainties could be applied. Alternatively, an unweighted least-squares solution could be sought, as here, and a comparison made of the resulting RMSR values to see whether, for a suitable number of harmonics, the value of 0.29 is matched.

The area enclosed by the curve is computed using the expression derived in annex E5.1.A:

$$A = \pi \sum_{r=1}^{n} r(b_{2r+1}c_{2r} - b_{2r}c_{2r+1}) \tag{E5.1.3}$$

E5.1.5 Uncertainty propagation

We consider, as an instance, the profile of the section in figure E5.1.1(c), using x and y to denote horizontal and vertical axes, respectively, with origin at the centre of the bottom-left pixel. The midpoints of the boundary pixels, taken of unit size, are then given by the $m = 38$ coordinate pairs

$$(4,2), (5,2), (6,2), (7,2), (7,3), (8,3), (8,4), \ldots, (3,3), (4,3).$$

We consider approximations to this set of discrete data points by continuous profiles in the form of curves represented by the parametric Fourier functions (E5.1.1). The values of t are taken as cumulative chord length as in section E5.1.4. These t-values will have uncertainties associated

with them due to the x- and y-values on which they based being uncertain. These uncertainties are ignored at this stage. Their effect is taken into account when the Monte Carlo method is applied

Fourier series (E5.1.3) with $n = 1, 2, \ldots, 15$ harmonics were determined (see figure E5.1.2), the corresponding sequence of values of area, RMSR, AIC, AICc and BIC being calculated (table E5.1.1).

We wish to select an appropriate number of harmonics that gives a balance between accounting for the data uncertainties and yielding an acceptably smooth representation of the profile.

Table E5.1.1: Tumour section area estimates, root-mean-square residuals and information criteria for various numbers of harmonics in the case of a coarse pixelated image

		x-variable				y-variable			
n	Area	RMSR	AIC	AICc	BIC	RMSR	AIC	AICc	BIC
1	65.11	0.64	20	21	28	0.72	24	25	32
2	63.88	0.45	17	19	28	0.44	16	19	28
3	64.21	0.41	19	24	34	0.43	20	24	35
4	65.73	0.28	20	28	39	0.32	21	29	39
5	65.67	0.28	24	37	46	0.33	25	37	46
6	65.25	0.26	28	46	53	0.29	28	46	53
7	65.21	0.27	32	57	60	0.29	32	58	60
8	65.38	0.26	35	71	67	0.29	36	72	67
9	65.39	0.25	39	89	74	0.30	40	89	74
10	65.36	0.27	43	111	81	0.32	44	111	82
11	65.28	0.28	47	139	88	0.33	48	140	89
12	65.30	0.30	51	179	96	0.33	51	179	96
13	65.32	0.30	55	235	103	0.33	55	236	103
14	65.24	0.15	58	324	109	0.37	59	325	110
15	65.42	0.11	62	484	116	0.35	63	485	117

For the coarse pixelated image, the sequences of RMSR values for the x- and the y-variable initially show a tendency to decrease to a minimum value of approximately 0.30 when $n = 6$ (six harmonics). Then, the sequences saturate at about that level for a few degrees, subsequently tending to behave erratically, particularly for the x-variable, for larger n due in part to the errors in the data being followed too closely. The Fourier series of degree 6 was selected although a case could be made for choosing degree 4 or 5 since to one significant digit the RMSR values were identical to that for degree 6. Curiously, the three information criteria AIC, AICc and BIC all exhibited minimum values for $n = 2$ (indicated by underlining in table E5.1.1) for which the RMSR values are 0.45 and 0.44, whereas the choice of $n = 6$ gives RMS values of 0.26 and 0.29, some 40 % lower. Moreover, the RMS values for $n = 6$ agree much more closely with the prior standard uncertainty of 0.29.

This observation about the information criteria is not consistent with our experience of selecting an appropriate degree in polynomial calibration studies. This choice of degree does not seem particularly consistent with the data on which it was based: see the '2 harmonics' sub-figure in figure E5.1.2. In particular, the re-entrant features in the shape of the contour in the 'north-east' and 'south-east' regions of figure E5.1.1(a) are not reproduced. These features are increasingly

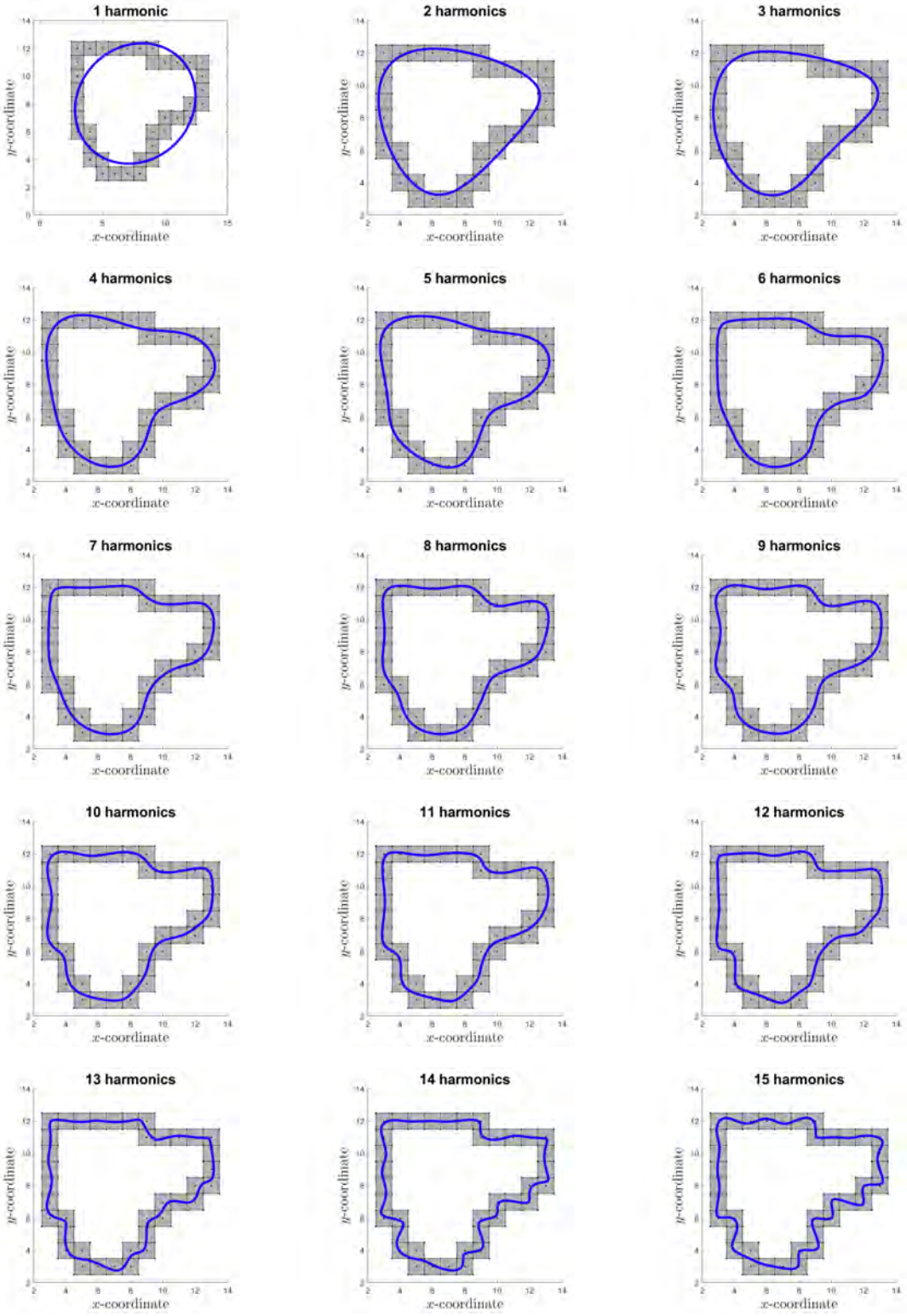


Figure E5.1.2: Fourier representations with 1, 2, . . . , 15 harmonics of the profile of a section of a tumour showing the according root-mean-square residuals in (non-dimensional) pixel units

reproduced by Fourier series of degree four upwards. However, by the time degree 11 is reached spurious effects are introduced in the curves of the Fourier series that are not apparent in the contour of the tumour.

In general, the progress in capturing more closely the outline of the tumour in figure E5.1.1(a) with successive degrees is shown in the sequence of sub-figures in figure E5.1.2. It would be unwise to select too low a degree such as corresponding to one, two or three harmonics in figure E5.1.2 since these Fourier series over-smooth the tumour outline. At the other extreme, the choice of a large number of harmonics such as 13, 14 or 15 is dominated by the Fourier representations following too closely the sequence of pixel midpoints specified by the data. An argument could be made purely on visual grounds that any number of harmonics from four to 12 might be suitable. Degree six was selected for the reasons above.

The corresponding area contained within the modelled curve is 65.2 pixel areas. This value should be compared with the minimum and maximum values given by the inner and outer profile of the boundary pixels, namely, 48 and 86, which have a mean of 67, tolerably close to the above value of 65.2.

The RMSR value of 0.30 does not relate at all directly to the uncertainty associated with the estimate of the area of the section of the tumour. To evaluate that uncertainty involves a complicated calculation. However, a Monte Carlo calculation is relatively straightforward. For such a calculation, we assume that the x- and y-coordinates of the midpoint of each boundary pixel have rectangular distributions $\mathrm{R}(-1/2, 1/2)$ relative to that midpoint. Accordingly, by sampling from these distributions, a new set of xy data can be created, a Fourier series with six harmonics computed and the according area determined. Repetition of this calculation a large number of times yields a distribution of values for the area from which the standard deviation, to be used as the standard uncertainty associated with the area, can be deduced.

There is a further aspect that must be explained. The sequence of data t-values (cumulative chord length) so generated will also have serial correlation associated with them since each t-value (apart from the first) is calculated from the previous value and the distance between the current point and the previous point. That aspect is automatically taken into consideration when applying the Monte Carlo method. See section E5.1.7.2.

Carrying out this calculation for 10^5 Monte Carlo trials gave the probability density function for the area in figure E5.1.3. This figure was obtained by sorting the 1×10^5 Monte Carlo samples into 40 bins of equal width between the smallest and largest sample area value, obtaining the number of samples in each bin, assigning those numbers to bin centres, and joining the resulting points by straight-line segments after normalizing the area to be unity. From the results, the area estimate of 65.5 and its associated standard uncertainty of 1.6 were obtained. These values were repeatable using different random number generator seeds to the number of decimal places quoted. This estimate is close to that above of 65.2, indicating little bias in the process. This standard uncertainty compares favourably with the value of 11.0 obtained by regarding 48 and 86 above as the limits of a rectangular distribution.

E5.1.6 Reporting the result

For the example presented in section E5.1.5, two estimates of the area a of the considered section of the tumour and the associated standard uncertainties were provided. One result was

$$\widehat{a} = 67, \qquad u(\widehat{a}) = 11.0$$

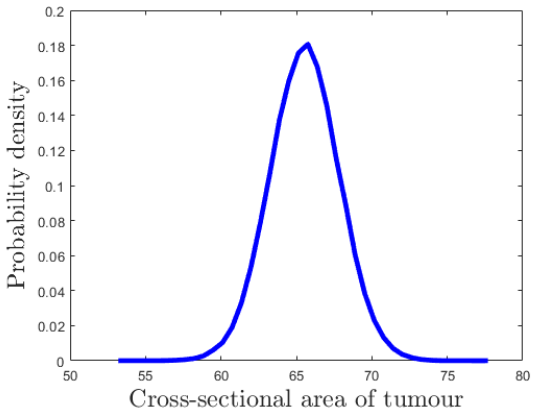


Figure E5.1.3: Probability density function for the cross-sectional area of the tumour provided by Monte Carlo sampling

using the simple approach of taking the mean of the areas defined by the inner and outer perimeters of the shaded pixels in figure E5.1.1(c) and using a rectangularity assumption for the distribution of possible values between those extremes.

The second result was

$$\widehat{a} = 65.2, \qquad u(\widehat{a}) = 1.6.$$

The estimate was obtained by modelling the centre points of the boundary pixels by a Fourier series and selecting an appropriate degree for the series. Because of the complexity of propagating uncertainties through the measurement model, a Monte Carlo method (MCM) was applied to determine the standard uncertainty of 1.6. The MCM delivered the estimate of 65.5.

To demonstrate stability of the results with respect to the model used, MCM results were also obtained for neighbouring degrees. For $n = 5$, the estimate 65.8 and standard uncertainty 1.5 were obtained. For $n = 7$, the corresponding values were 65.4 and 1.6.

These various estimates are consistent with respect to their associated uncertainties.

E5.1.7 Interpretation of results

E5.1.7.1 General

The volume or mass of an organ or tumour is generally obtained from a region of interest (ROI) outlined on anatomical imaging data [385]. It is therefore possible to estimate the outlining accuracy from consideration of factors that affect delineation. The method used will depend largely on the information and resources available at the time of outlining and the method employed by the operator or algorithm to define the ROI.

The process in sections E5.1.4 and E5.1.5 is indicative of one approach, namely, the use of regression techniques to fit the mid-points of an outline taking into account the knowledge that the midpoints are a crude guide to points on the boundary.

For the given data, a Fourier series of degree six (six harmonics or 13 parameters) appears to capture reasonably well the given outline bearing in mind that only data points representing the midpoint of the boundary pixels are used.

The importance of the principles used to select an appropriate degree cannot be over-emphasized. RMSR or information criteria can be used but have their failings. In the NPL author's considerable experience with regression using polynomials, splines and other functions [111, 295, 392–396], it has been found that identification of the point where the RMSR values start to saturate with respect to degree is generally reliable. However, such identification is difficult to automate and there is much value in having the user 'in the loop'. Examination of RMSR values against degree and assessing candidate fits to the data such as in figure E5.1.2 are valuable.

E5.1.7.2 Correlation

We would anticipate that the sequence of x-coordinates would have some correlation associated with them because there is pattern in the arrangement of the corresponding pixels as a consequence of expected underlying smoothness in the tumour boundary. A similar remark applies to the y-residuals. Accordingly, we would expect to see pattern in the x-residuals $d_i = x_i - F(t_i)$ and in the y-residuals $e_i = x_i - G(t_i)$. To investigate, we examine residual plots for the chosen degree: see figure E5.1.4. We observe what appears to be some oscillatory behaviour in the residuals.

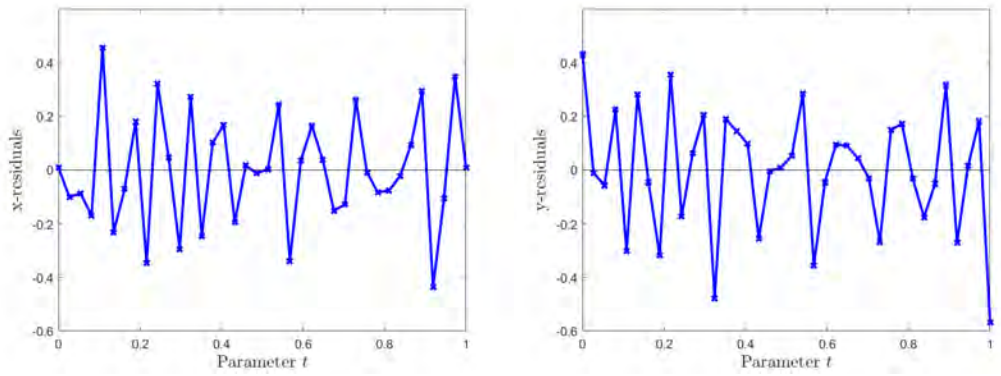


Figure E5.1.4: x- and (right) y-residuals for the Fourier-series representation of degree 6 (residuals are joined by straight-line segments purely for visualization purposes)

Let $g_0 = \sum_{t=2}^{m} d_t^2$ and $g_1 = \sum_{t=2}^{m} d_t d_{t-1}$. The estimate of the first serial correlation coefficient r_x for the x-data is $r_x = g_1/g_0$. For the example, $r_x = -0.440$. Similarly, for the y-data, $r_y = -0.430$. Since these estimates are so similar, we take their average $r = -0.435$ as applying to both the x- and y-data.

The calculation is then repeated as one of generalized least squares with a covariance matrix based on this observed correlation. The Fourier coefficients so calculated were close to those obtained when no account was taken of correlation: the norm of the difference between the two sets of Fourier coefficients relative to the norm of the coefficients computed by generalized least squares was approximately 2×10^{-3}.

Applying the MCM as in section E5.1.5, but now to the generalized least squares formulation, gave an area estimate of 65.5 with an associated standard uncertainty of 1.6, agreeing to the one decimal place quoted with those when ordinary rather than generalized least squares was used.

Some experiments with taking higher-order serial correlation into consideration did not alter the above results to the number of digits reported.

E5.1.7.3 Analysis of the fine pixelated image

The whole exercise was repeated for the ($m = 56$) darkly shaded pixels in figure E5.1.1(b) using again pixels of unit size. See table E5.1.2.

Table E5.1.2: Tumour section area estimates, root-mean-square residuals and information criteria for various numbers of harmonics in the case of a fine pixelated image

		x-variable				y-variable			
n	Area	RMSR	AIC	AICc	BIC	RMSR	AIC	AICc	BIC
1	142.51	0.90	48	49	58	0.91	49	50	59
2	140.70	0.56	26	27	40	0.51	23	25	37
3	141.67	0.48	25	28	43	0.46	24	27	43
4	143.62	0.28	22	27	44	0.32	23	28	45
5	143.60	0.28	25	33	52	0.32	27	34	53
6	143.54	0.27	29	39	59	0.30	30	40	60
7	143.46	0.27	33	47	67	0.29	33	47	68
8	143.30	0.27	37	55	75	0.29	37	56	76
9	143.33	0.27	41	65	83	0.29	41	65	84
10	143.37	0.27	45	75	91	0.30	45	76	92
11	143.45	0.27	48	87	99	0.31	49	88	100
12	143.35	0.27	52	101	107	0.30	53	101	107
13	143.34	0.28	56	116	115	0.32	57	117	115
14	143.15	0.29	60	135	123	0.28	60	134	123
15	143.35	0.28	64	156	131	0.29	64	156	131

The RMSR residuals saturated essentially to the same level of 0.3 for $n = 6$ and again the degrees that would have been selected by the three information criteria differed. The corresponding area estimate, corrected by the factor $(38/56)^2$, to take account of the numbers of pixels involved [see sub-figures (b) and (c) in figure E5.1.1] was 66.1 (Monte Carlo gave 66.2) with an associated standard uncertainty of 1.8.

E5.1.8 Introduction of the second application: nanoparticle sizing by AFM

In the form of nanoparticles, materials can exhibit different properties than in their common bulk form. These properties open the way to new technological applications, but also potential hazards. In their characterization, the first parameter to be measured is thus their size, or, more precisely, their size distribution. An explanation of the need for metrology in nanoparticle sizing can be found in example in E2.4.

Size of nanoparticles can be measured by AFM, after deposition from a liquid dispersion onto a flat substrate. AFM is a microscopy technique, and like other microscopy techniques, suffers from the drawback of being time-consuming for the measurement of size distribution of the nanoparticles. This application presents the influence of the choice of the pixel size (that is, how refined the picture is) on the measurement of the size of a nanoparticle. This choice can

in turn serve to optimize the imaging time (a coarser image can be acquired more quickly), by minimizing it and keeping the resulting uncertainty from pixel size under control. This approach can thus be useful in a laboratory.

A physical model of the measurement is first set up. Its application in the classical GUM framework is explained, as well as a possible extension to a Bayesian framework. Real data on gold nanoparticles illustrate some of the effects, which might seem counter-intuitive, but support the model here presented.

E5.1.9 Specification of the measurand(s)

AFM belong to the class of methods termed Scanning Probe Microscopy (SPM) and produces a topography of the specimen. In its classical use, the topography is measured by scanning the surface by keeping a small tip in intermittent contact with the specimen. The feedback control to keep this intermittent contact is used to make measurement of the vertical topography of the substance. Assuming a spherical shape for small nanoparticles, the height at the top of the nanoparticle with reference to the substrate is equal to the diameter of the spherical nanoparticle, and acts as a measurand for it. An example of topography is presented in figure E5.1.5. More details about the measurement and investigation of other influence factors can be found in example in E2.4.

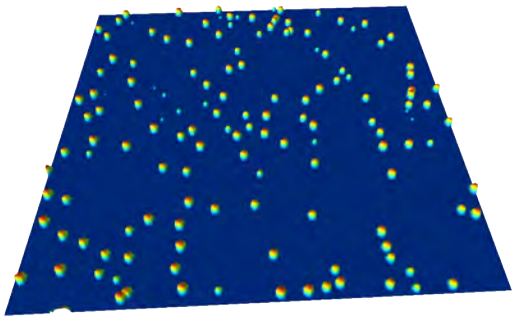


Figure E5.1.5: 3D view of AFM topography of a polystyrene 100 nm nanoparticle sample deposited on flat mica substrate

E5.1.10 Measurement model

For the height measurement, the topography is scanned in successive lines, as illustrated in figure E5.1.6. It is a continuous sampling, and the pixel value (height) is extracted from electronics readout in continuous time. It is the result of the average over the pixel by integration of the signal representing the height (voltage applied to a piezoelectric crystal to move the tip vertically). The flexture movement of the tip is monitored by movement of the spot of light reflected by the tip on photodetectors. To keep the same intermittent contact of the tip probe, a feedback control – PID – is applied continuously, with adjustable parameters. The signal is collected in two passes, the "trace" and "retrace" signals, of which the retrace signal is used for measurement, as illustrated in figure E5.1.6. With correct adjustment of feedback parameters, the "trace" and "retrace" signals are almost identical. Other parameters also influence the measurement: the scan speed, set point

amplitude, the type of tip, environmental conditions, and so on. Uncertainty evaluation example E2.4 gives a more detailed presentation of statistical modelling of such measurement and the associated uncertainty evaluation.

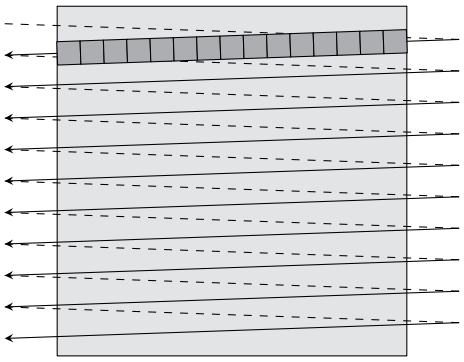


Figure E5.1.6: Scanning pattern of AFM measurement. The dashed line represents the acquisition line of "Trace signal" and the plain line represents the "Retrace signal". Measurement are usually performed on the "Retrace signal". Pixels are illustrated on the first line.

The real tip shape and size are neglected here, though a more realistic example would need to consider these features. These intricacies of the problem are left for further consideration. References [397,398], provide information on the description of the tip shape and effect on AFM measurement. For application to the present case, interested readers are invited to contact the authors of reference [399].

No full measurement model exists for such nanoparticle size measurement by an AFM, because of the complex acquisition electronics and manual tuning. A global statistical model [175] and, in parallel, an effect-per-effect investigation and modelling as building blocks can be introduced, as in [400]. We adopt here the latter for the modelling of the pixel-size effect on the measurement of the size of nanoparticles by an AFM.

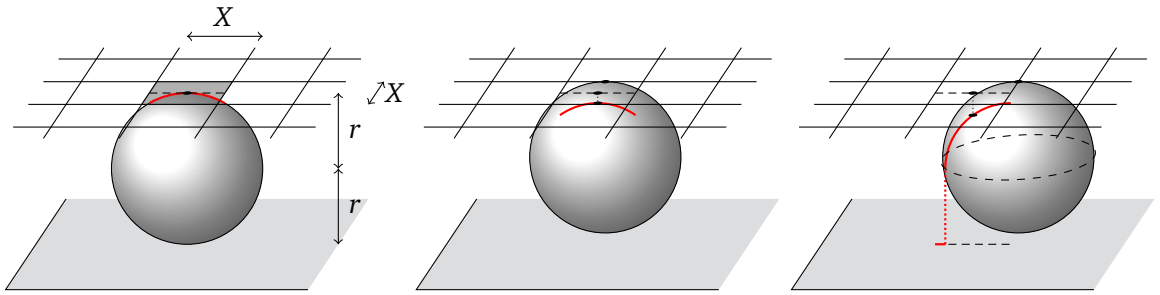
To recapitulate, the effect of choosing a given pixel size is to change the average height recorded by the system for the pixel of maximum height, which is the measurand for the nanoparticle height. It is the pixel that contains the nanoparticle apex. By denoting the nanoparticle radius by r and the size of one pixel edge by X, the measured height h will be

$$h = \frac{1}{X}\int_{-X/2}^{X/2} z_{\text{tip}}(x)\,\mathrm{d}x \tag{E5.1.4}$$

with $z_{\text{tip}}(x)$ being the continuously read measured height for the tip along the tip path (x coordinate; retrace signal in figure E5.1.6). In the ideal case, the tip will pass through the apex of the spherical nanoparticle along its path, as illustrated in figure E5.1.7a, which corresponds to the configuration with maximal yielded size, where

$$z_{\text{tip}}(x) = \begin{cases} r + \sqrt{r^2 - x^2}, & \text{if } x \le r, \\ 0, & \text{elsewhere,} \end{cases} \tag{E5.1.5}$$

and which corresponds to the red line in the figure. It can directly be noticed that the integration along the x-axis (i.e., under the red curve) will always induce an underestimation of the real size ($2r$, at the apex). Moreover, the smaller the nanoparticle is, the more its size is underestimated.



(a) Centered (b) Off on one side (c) Most disadvantageous relative position

Figure E5.1.7: Examples of different relative position between sampling grid (pixels) and the nanoparticle. The line in red corresponds to the line of measurement

In a realistic case, the tip acquisition path is not necessarily aligned with the nanoparticle. A nanoparticle can be completely missed if it is too far from the tip path, that is, the pixels are too big (coarse). Figures E5.1.7a, E5.1.7b and E5.1.7c illustrate gradual underestimation by the only fact of misalignement of the nanoparticle and the pixel grid, for the same nanoparticle size and the same pixel size. If $(\Delta x, \Delta y)$ is the position of the apex of the nanoparticle in the pixel, with y orthogonal to x,

$$z_{\text{tip}}(x) = \begin{cases} r + \sqrt{r_{\text{eff}}^2 - (x - \Delta x)^2}, & \text{if } x - \Delta x \leq r_{\text{eff}} \text{ and } \Delta y \leq r, \\ 0, & \text{elsewhere} \end{cases} \tag{E5.1.6}$$

with $r_{\text{eff}} = \sqrt{r^2 - (\Delta y)^2}$ being the effective radius along the tip path (see figure E5.1.8). The problem is now specified.

This problem is geometrical, and it can be quickly noticed that it is the relative scale of the nanoparticle radius r and the pixel size X that is relevant. In other words, the problem can be reformulated, independently of the unit system, by non-dimensionalizing to the nanoparticle radius r. $\widetilde{r} = r/X$ and $h/(2r)$ are thus introduced in the calculation. By scaling variables, computations are now universal to all classes of the problem rather than just a specific case. The measurement problem now involves the following equation:

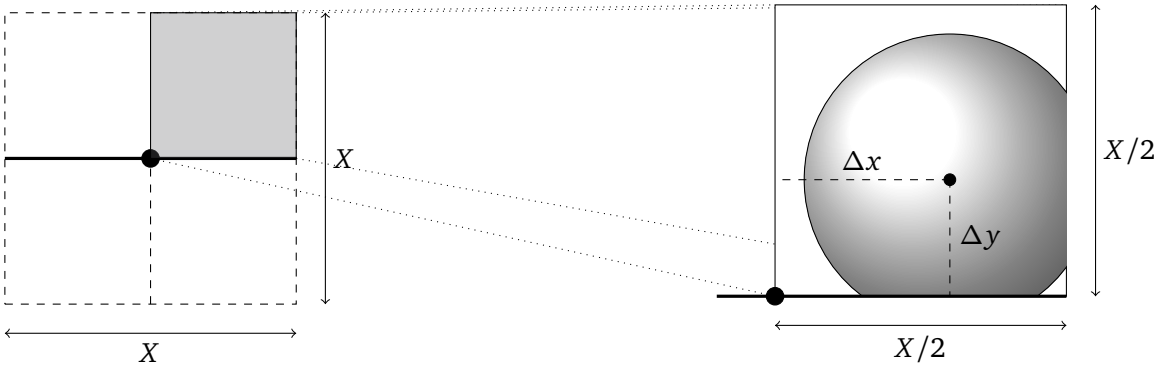
$$\frac{h}{2r} = f(\widetilde{r}). \tag{E5.1.7}$$

The ideal case — with perfect pixel size resolution — would be to measure $h/(2r)$ as unity, which is not achievable. Moreover equation (E5.1.7) is to be considered as a statistical relation and not strictly as a function, as explained in section E5.1.10.1.

Several cases of integration exist, depending on the condition of equation (E5.1.4) met at the integration borders of the pixel (at $-X/2$ and $X/2$): the pixel is fully inside the nanoparticle (figure E5.1.7a is an example), only a part of the scanning path is touching the nanoparticle (figure E5.1.7c, for example) or the nanoparticle is fully inside the pixel. In some extreme cases, the nanoparticle is small with its apex in a corner of the pixel, and the tip is not touching the nanoparticle at all along its path. The nanoparticle is missed. This corresponds to one part of the probability, as a Dirac delta, for the $h/(2r)$ value.

E5.1.10.1 Law of probability

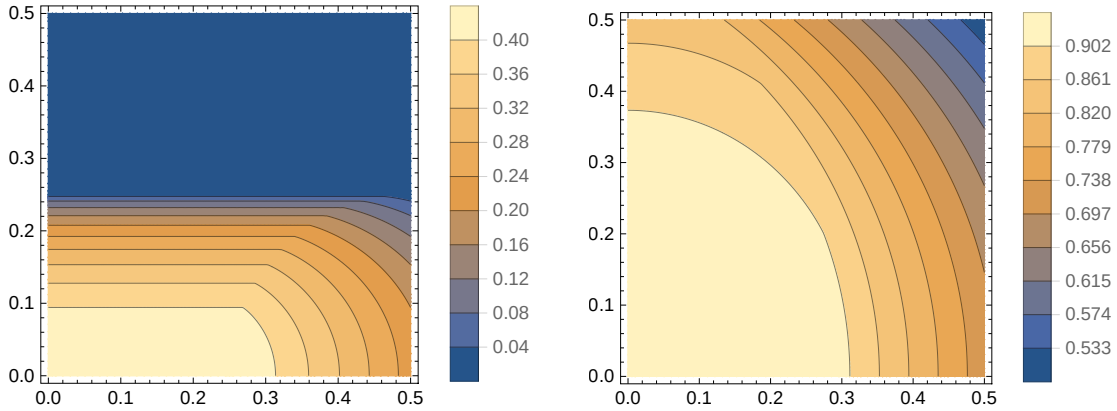
The probability of a given relative measured height $h/(2r)$ for a given relative pixel size $\widetilde{r} = r/X$ is derived from the probability measure of the positions $(\Delta x, \Delta y)$. Because the nanoparticles are randomly distributed over the substrate and the pixel grid is randomly set relatively to the



(a) View of a full pixel, and internal symmetries of the problem

(b) View of a nanoparticle and representation of variables, in a quarter of a pixel

Figure E5.1.8: Symmetry reduction for the position of the nanoparticle, illustrated as viewed from the top. The thick line represents the integration line (scanning path between $x = -X/2$ and $X/2$)



(a) $\widetilde{r} = 0.2$. Nanoparticle is missed (dark blue part on top) if too far from scanning line $\Delta y = 0$

(b) $\widetilde{r} = 0.7$.

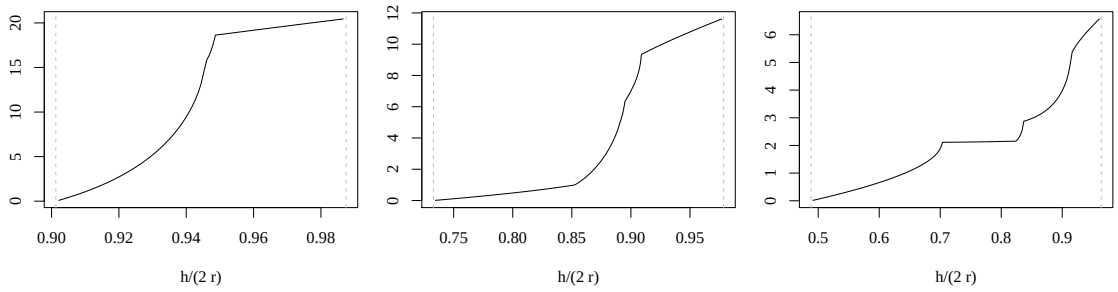
Figure E5.1.9: Iso-height contour plot of estimation of the relative height $h/(2r)$ as a function of the position $(\Delta x, \Delta y)$ of the nanoparticle apex

substrate, $(\Delta x, \Delta y)$ are taken as uniformly distributed over the pixel. By symmetry along the $\Delta x = 0$ and $\Delta y = 0$ axes, the integration domain can be reduced to a quarter of a pixel for numerical computations, as shown in figure E5.1.8. In a more practical way, the probability of $h/(2r)$ given $\widetilde{r}$ is thus the length of a iso-height line contour of the graph of height as a function of $(\Delta x, \Delta y)$ (with parameter $\widetilde{r}$ fixed; see figure E5.1.9 with layout from figure E5.1.8b), and taking into account the Jacobian factor. It can also be seen, from a probabilistic point of view, as a change of variables:

$$P(h|\widetilde{r})\, dh = P(h(\Delta x, \Delta y)|\widetilde{r})\, \mathrm{d}(\Delta x)\, \mathrm{d}(\Delta y). \tag{E5.1.8}$$

The probability density is computed for several fixed values of $\widetilde{r}$, as profiles, and as illustrated in figure E5.1.10. Including the probability of missing the nanoparticle by the mean of a Dirac delta (for null $h/2r$) and expressing in terms of reduced variables, the PDF takes the following mathematical form

$$P\left(\frac{h}{2r}\middle|\widetilde{r}\right) + (1-2\widetilde{r})\, \delta\left(\frac{h}{2r}\right). \tag{E5.1.9}$$



(a) $\widetilde{r} = 1.305$ (b) $\widetilde{r} = 1.005$ (c) $\widetilde{r} = 0.787$

Figure E5.1.10: Probability density function of height measurement $(h/(2r))$ at fixed $\widetilde{r} = r/X$

Collating all the computed profiles results in the two-dimensional graphical representation of the parametric PDF in figure E5.1.11.

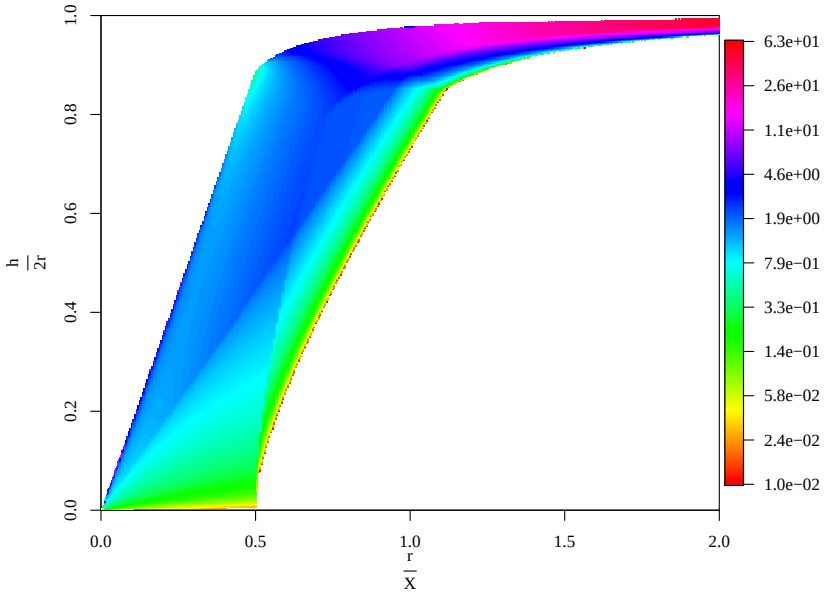


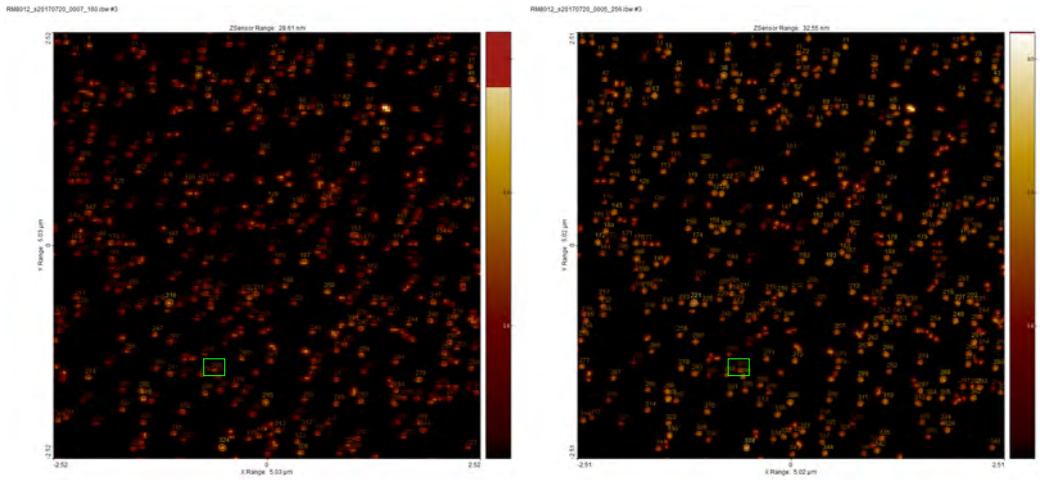
Figure E5.1.11: Probability of height measurement $h/(2r)$, as a function of the $\widetilde{r} = r/X$ parameter

E5.1.10.2 Confronting the model with real data

After setting up a model, it is valuable to verify its predictions or features in typical conditions that are already well known by other means or previous experience. The reference material RM 8012 [187] was used as a nanoparticle sample for confronting the model to the pixel size effect for real data. It is a suspension of gold nanoparticles in ultrapure water, with citrate salt for stabilizing the dispersion, of a certified mean size of 24.9±1.1 nm ($k = 2$, 95 % coverage probability). As a consequence of the production process of nanoparticle samples, different nanoparticles within the sample have different sizes, but only the central size value is certified. The actual size of each individual nanoparticle is thus not precisely known. On the other hand, confronting AFM measurement with another technique for each nanoparticle will likely not improve the present analysis, if possible at all (by another SPM technique; such an example is presented in [401]). Difficulties arising from the comparison of techniques will affect the comparison by instrumental effect from the other technique. The pixel-size effect can be observed relatively by comparing the measurements of the same nanoparticle substrate deposition (same area), by the same AFM

instrument and by identifying the same nanoparticles on two images. It is not possible to scan exactly the same area because of limitations of the instrument. A slight shift in (x, y) is present, from image to image and a matching of nanoparticles had to be carried out by a dedicated algorithm.

Figure E5.1.12 present several images, applying the above methodology. Nanoparticle identification numbers do not necessarily match between the pictures. An animation is provided in [31]. It is noticed that the pictures with lower resolution are generally darker, meaning that the mean measured height is smaller. This is in agreement with the model.



(a) 160 px (b) 256 px

(c) 512 px (d) 1024 px

Figure E5.1.12: Images of the same $5\,\mu\text{m} \times 5\,\mu\text{m}$ area with the same nanoparticles. The height scale is common. The colour of the number identifying the nanoparticle corresponds to the measurand, the maximum height pixel

The probabilistic effect is visible in figure E5.1.12. Nanoparticles number 269, 281 and 325 on respectively, figures E5.1.12a, E5.1.12b, E5.1.12d illustrate this peculiar effect for the same real nanoparticle (identified by green boxes). The height is greater for nanoparticle 269 (14.8 nm on figure E5.1.12a) than nanoparticle 281 (14.1 nm on figure E5.1.12b). This can be seen from the colour of the label, which is a slightly darker for the smaller nanoparticle. This is an observation going in the opposite direction of the global tendency of lower measured height with lower

resolution, but for individual nanoparticles, showing the probabilistic effect. This effect is also visible on the animation, as scintillation (changes to higher and lower measured size, randomly) of the individual nanoparticles when increasing resolution. The same nanoparticle is measured 21.4 nm on figure E5.1.12d. This value is already closer to the certified mean size of the sample (24.9 ± 1.1 nm [187]).

Interested readers are invited to contact the authors for more information about the validation of the model. Comparison of data across several resolutions has been carried out, using the same nanoparticles, identified on all the images. Unfortunately, a thorough validation of the model PDF would imply the collection of an enormous quantity of data, which is unrealistic with current means. Natural limits (edges of the coloured zone on figures E5.1.11 and E5.1.13) of the values of the model induce extreme bands that delimit measurements from both resolutions. The probabilistic fluctuations illustrated above are allowed within these bands. All data points actually lie within these bands. It is important to keep in mind that the shape of the tip also has an indirect effect on the measurement of validation data, an aspect that was neglected here for readability.

E5.1.11 Interpretation of results

E5.1.11.1 Correction factor and uncertainty evaluation in the GUM framework

The obtained PDF can be exploited by reducing it to a correction value and a given standard uncertainty, as the corresponding mean and standard deviation, therefore making a normal approximation of the PDF, obviously not exact from figure E5.1.10. The size correction would be estimated a posteriori from the mean measured size, because of dependence of the model on the $\widetilde{r} = r/X$ parameter (where $2r$ is the true size). This information can then be used when following the GUM approach, as one of the contributions of an uncertainty budget. This approximate approach can be considered as a first step of the analysis of the problem. In a second step, a Monte Carlo approach (GUM-S1) would yield the correction and uncertainty evaluation at the same time, and provide more valid results. This can easily be understood from the obviously non-Gaussian PDF, as shown in figure E5.1.10.

Mean and one-standard-deviation bands are drawn on figure E5.1.13, as a function of $\widetilde{r} = r/X$. The black curve considers only the non-null measured heights, while the brown curves take into account undetected (missed) nanoparticles: the Dirac delta part in equation (E5.1.9). Having two means certainly seem unexpected. If a single particle is measured and one would like the effect of the pixel size to be considered, the black curve should be used: the nanoparticle was measured and its own true size is to be estimated. However, when considering the population as a whole, the brown curve is to be used for correction, in order to take into account the missed (invisible) nanoparticles.

E5.1.11.2 Experimental optimization in the GUM framework

The present model can also be used as a basis for obtaining a first approximation to the uncertainty associated with imaging pixel size, sufficient for deciding whether this source of uncertainty is negligible or not. For example, choosing a resolution of $\widetilde{r} = r/X \approx 1.4$ should be sufficient to underestimate the size by at most 5 % (grey line) with a probability of at most 16 % (with one-sided $1\,\sigma$ – one standard deviation – band), as illustrated by the intersection of the grey line and brown (or black) band in figure E5.1.13. This choice of resolution leads to advantages

in the imaging processing. By reducing the resolution and keeping the same imaging area, the imaging time (and the associated cost) is reduced. Imaging time is often cited as one of the main disadvantages of AFM, and could thus be optimized for this application by this technique. Seen another way, the imaged area can be increased in such a way to match the minimal $\widetilde{r} = r/X$ allowed by the target uncertainty on size underestimation. With this approach, the number of nanoparticles on a single image is optimized. The greater is the number of nanoparticles the better is the PDF description of the sample (its size distribution). Only a subset of the sample can be realisticly imaged. Representativeness of the nanoparticle sample is another common criticism about AFM, also optimized by this method. The precision of the measurement here comes from the combined effects of the number of nanoparticles measured (scanned area) and the precision on the measurement of the size of a single nanoparticle). The choice of optimization of one compared to the other is a strategy choice of the laboratory.

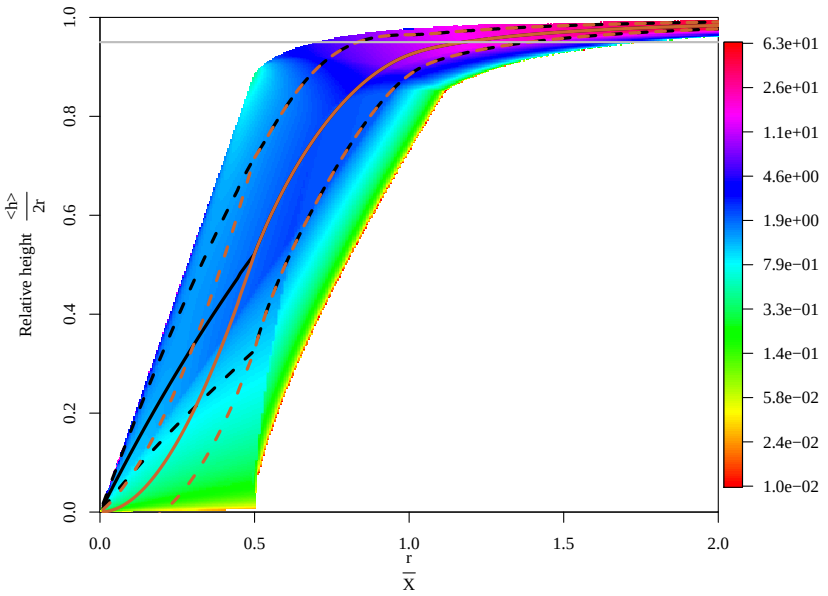


Figure E5.1.13: Probability of height measurement ($h/(2r)$). The brown $1\,\sigma$ band considers all existing nanoparticles, the black only considers nanoparticle for which a (non-null) measurement exists. The solid curves are the mean value as a function of $\widetilde{r}$, surrounded by their $1\,\sigma$ bands (dashed).

A good strategy for a calibration service on nanoparticle samples could be to perform a screening measurement to have an indication of the order of magnitude of the sample size and then to design the analysis (for example, choice of pixel size) to reach the optimal trade-off to the calibration laboratory: precision vs. measurement time. This is a different approach from the simple uncertainty evaluation approach.

This strategy is easily applicable to mono-disperse samples. A mono-disperse sample contains a population of nanoparticles all of similar size. On the contrary, a poly-disperse sample is made of a population covering a wide range of sizes. Conditions that are adapted to a given size, by the optimization explained here above, are only suitable for that size and not necessarily for other sizes. With this model approach, one can evaluate uncertainties for smaller and bigger sizes, quantitatively.

E5.1.11.3 Scope for further work: interpretation in Bayesian framework

Another exploitation of the model can be to extract an estimate of the true radius from the experimental data by inversion. Up to now, we have given a description of the measurements h, where the true size of the nanoparticle was a parameter (r). The data of the experiment modelled here can be synthetically summarized by: h is a measured value, X is an experimental parameter and r is the "true value". h/X is thus data of sole experimental nature (no a priori knowledge of the sample, r, is included in it), of which the experimenters would like to extract the true value of the size r.

In the model, a constant h/X corresponds to constant $h/(2r) \times 2 \times r/X$, that is, a hyperbolic line in the graph of figure E5.1.11. By extracting the probability along this line, one has access to the probability distribution of r, given the measurement of the nanoparticle h/X: $P(r|h/X)$. A Bayesian inversion algorithm can thus be applied in order to estimate the true r distribution of the sample, where this model acts as a transfer function. This inversion can only be carried out numerically, by Monte Carlo methods. The final uncertainty will be derived from the obtained numerical distribution. The correction and uncertainty are directly determined from experimental data, without any need for a model parameter to be estimated posterior to measurement. This is a much better approach than the previously explained GUM approaches, where a priori knowledge of r was needed for the estimation of r (since it was based on $h/(2r)$ and $\widetilde{r} = r/X$). r was acting on an a priori knowledge parameter and being the estimated value at the same time an iterative estimation would be needed. This new approach is thus also more natural and rigorous.

It is also observed that a smaller pixel size X would tend to higher h/X and a reduced uncertainty: the associated hyperbola is more to the top-right corner in figure E5.1.11 (hyperbola not drawn on the figure).

E5.1.12 Conclusion

A model approach to uncertainty evaluation for a single source has been presented: it describes the effect of the choice of the pixel size for the measurement of the nanoparticle size distribution by AFM. The model shows an intrinsic probabilistic nature of the measurement. Several uses of the same model are here presented, targeting different applications and different refinement levels. A classical GUM approach to the pixel size uncertainty evaluation is first presented, with application for process optimization. Correction for the effect is also presented, for individual nanoparticles and the population. For the latter, some items under measurement can be missed but nevertheless corrected for by the model. Following a Bayesian approach, the pixel size effect can be corrected for and the uncertainty evaluated simultaneously by applying Bayesian inversion from purely experimental data (no a priori knowledge of the size of the sample needed). The latter approach will certainly appear more natural and rigorous to the reader than the frequentist approach.

E5.1.A Area enclosed by a Fourier curve

The area enclosed by the Fourier curve (E5.1.1) is

$$A = \int_{\alpha}^{\beta} x \mathrm{d}y = \int_{0}^{1} F(t)G'(t)\mathrm{d}t,$$

where α and β are the limits of the variable x.

Use of the basic trigonometric identities

$$\cos P \cos Q = \frac{1}{2}[\cos(P+Q) + \cos(P-Q)],$$
$$\sin P \cos Q = \frac{1}{2}[\sin(P+Q) + \sin(P-Q)],$$
$$\sin P \sin Q = -\frac{1}{2}[\cos(P+Q) - \cos(P-Q)]$$

together with the algebraic form for the derivative $G'(t)$ gives, after some lines of tedious algebra, expression (E5.1.3).

The correctness of expression (E5.1.3) was checked for degrees from 1 to 15 by in each case evaluating the Fourier series at 1000 uniformly spaced values of t and calculating the area of the polygon so formed. For all 15 degrees the areas agreed to better than 1 part in 10^5.

Example E5.2

Magnetic resonance-based electric properties tomography

A. Arduino, F. Pennecchi, L. Zilberti, U. Katscher, M.G. Cox

E5.2.1 Summary

The present example shows the uncertainty evaluation, under repeatability conditions, of a technique for Electric Properties Tomography (EPT), a quantitative imaging method based on Magnetic Resonance Imaging (MRI). Repeated acquisitions of MRI scans of a homogeneous cylindrical phantom are analysed with appropriate statistical techniques to evaluate the variance-covariance matrix of the EPT input. The latter is then propagated through the EPT technique according to the law of propagation of uncertainty (LPU) as described in the GUM [2, 4].

E5.2.2 Introduction of the application

EPT is a quantitative imaging method that, by elaborating maps acquired with an MRI scanner, provides an estimate of the spatial distribution of the electric properties within the human body [402, 403]. Amongst the different EPT techniques proposed in the literature, the phase-based Helmholtz-EPT [403, 404] is investigated here, because of its prevalence.

Biological tissues are dispersive media, which means that they exhibit electric properties varying with the frequency of the electromagnetic radiation which impinges on them. EPT estimates the spatial distribution of the electric properties only at a certain frequency, namely at the Larmor frequency of the used MRI scanner. Typical values of the Larmor frequency for clinical scanners are 64 MHz and 128 MHz, which correspond to the resonance frequency of hydrogen nuclei within 1.5 T and 3 T scanners, respectively.

A metrologically rigorous application of EPT would benefit the biomedical field. Indeed, it would allow the determination of the physiological values of the electric properties of *in vivo* tissues, which could differ from the values measured *ex vivo* on excised samples [405]. Moreover, it could allow to monitor the course of diseases (and therapies) that alter the electric properties, like for some kind of tumours [405].

The present example aims at evaluating the uncertainty in the EPT results under repeatability conditions. To this end, a two step approach is followed:

1. The **quantification of uncertainties and covariances** in the input data is achieved by analysing a dataset of experimental input maps acquired under repeatability conditions with a 3 T MRI scanner.
2. The **propagation of the input variance-covariance matrix through the EPT technique** is performed by LPU. Since the adopted EPT technique is linear, the use of the LPU is not an approximation.

E5.2.3 Specification of the measurand(s)

The phase-based Helmholtz-EPT provides an estimate of the spatial distribution of the electric conductivity σ by elaborating the phase distribution φ^+ of the complex positively rotating component B_1^+ of the radiofrequency magnetic field generated by the MRI scanner [403, 404]. If the magnitude $|B_1^+|$ is provided as an additional input, then the accuracy of the estimate of σ can be improved and an estimate of the permittivity ε can be obtained as well.

The present example is focused on the estimation of the electric conductivity σ of a homogeneous cylindrical phantom filled with a solution of $3.75\,\mathrm{g\,L^{-1}}$ NaCl in distilled water, which has a nominal (provided with no associated uncertainty) electric conductivity equal to $0.56\,\mathrm{S\,m^{-1}}$ at 128 MHz and for which no independent experimental characterization was available.

Since Helmholtz-EPT handles three-dimensional images, the elementary unit of its input and output is the voxel, namely the three-dimensional extension of the pixel. Each voxel is an elementary cuboid, whose dimensions correspond to the spatial resolution of the maps. The value of σ estimated by the EPT technique in each voxel is a distinct result of a measurement, possibly correlated to the estimates in the other voxels. Since the phantom is homogeneous, the spatial average $\overline{\sigma}$ of the estimated electric conductivity is computed, together with the associated uncertainty, and compared to the nominal electric conductivity. The choice of the spatial average as a representative value for the estimated electric conductivity of the phantom is corroborated by the fact that it is the least squares approximation of the collection of estimates. Moreover, the systematic error in the recovering of σ, typical of any phase-based EPT technique, is estimated by comparing the result with that obtained by using $|B_1^+|$ as an additional input.

Also the (generalized) weighted average $\overline{\sigma}_\mathrm{w}$ of the conductivity values in the several voxels is investigated as a further estimator of the phantom electric conductivity, since it accounts for the covariance matrix of the involved quantities and it is the minimum variance unbiased estimator of their expected value [366].

In order to avoid some well-known errors of Helmholtz-EPT at the boundary of the phantom, and to reduce the dimensionality of the problem, only a sphere of radius 30 mm centred in the middle of the phantom is considered in the present analysis.

E5.2.4 Measurement model

Helmholtz-EPT is based on the Maxwell equations, that describe the electromagnetic phenomenon. Precisely, given the frequency f of the electromagnetic radiation, the relevant time-harmonic Maxwell equations for a passive medium are

$$\begin{cases} \nabla \times \boldsymbol{B} = \mathrm{i}\omega\mu_0\left(\varepsilon - \mathrm{i}\dfrac{\sigma}{\omega}\right)\boldsymbol{E}\,, \\ \nabla \times \boldsymbol{E} = -\mathrm{i}\omega\boldsymbol{B}\,, \end{cases} \tag{E5.2.1}$$

where i is the imaginary unit, $\boldsymbol{B}$ is the magnetic flux density, $\boldsymbol{E}$ is the electric field, $\omega = 2\pi f$ is the angular frequency and μ_0 is the magnetic permeability of the vacuum. The two equations can be combined to obtain, under the assumption of spatially homogeneous electric properties σ and ε, the Helmholtz equation for the magnetic flux density. The algebraic inversion of the Helmholtz equation leads to the reference equation of Helmholtz-EPT

$$\varepsilon - \mathrm{i}\frac{\sigma}{\omega} = -\frac{\nabla^2 B_1^+}{\omega^2 \mu_0 B_1^+}\,, \tag{E5.2.2}$$

which relates explicitly the unknown electric properties to the measurable component B_1^+ of the magnetic flux density.

By looking only at the imaginary part of equation (E5.2.2), the explicit relation for the electric conductivity σ is found to be

$$\sigma = \frac{\nabla^2 \varphi^+}{\omega\mu_0} + 2\frac{\nabla|B_1^+| \cdot \nabla\varphi^+}{\omega\mu_0|B_1^+|}\,, \tag{E5.2.3}$$

being $B_1^+ = |B_1^+|\mathrm{e}^{\mathrm{i}\varphi^+}$. Finally, under the assumption of a spatially homogeneous $|B_1^+|$, as this is often the case for 1.5 T and 3 T scanners, the reference equation of the phase-based Helmholtz-EPT reduces to

$$\sigma = \frac{\nabla^2 \varphi^+}{\omega\mu_0}\,. \tag{E5.2.4}$$

The measurement model applied in this example is a discrete approximation of equation (E5.2.4). Indeed, it is a linear application, defined by the matrix $\boldsymbol{A}$, that elaborates a discrete three-dimensional image of φ^+, defined by the vector $\boldsymbol{x}$, and produces a discrete three-dimensional image of σ, defined by the vector $\boldsymbol{y}$, according to

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}\,. \tag{E5.2.5}$$

Each component of the vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ is associated through a one-by-one relation to a voxel of the three-dimensional images. Matrix $\boldsymbol{A}$ approximates the linear differential operator $(\omega\mu_0)^{-1}\nabla^2$, therefore it is in general a sparse matrix. In particular, each row of $\boldsymbol{A}$ is a local approximation of the differential operator and is completely defined by the set of voxels, called kernel, used to compute the approximation. In the following p denotes the dimension of vector $\boldsymbol{x}$, namely the number of voxels in which the measured phase is provided; whereas N denotes the dimension of vector $\boldsymbol{y}$, namely the number of voxels in which the electric conductivity is estimated.

The kernel is characterised by its shape and its size. Three shapes are analysed in this example: the cross, the sphere and the cube. They are represented in Figure E5.2.1. For each shape, a size ranging from $n = 1$ up to $n = 5$ has been considered, where n has the meaning depicted in Figure E5.2.1. The number of voxels in each kernel shape and size is depicted in Table E5.2.1. The values of the components of $\boldsymbol{A}$ are determined in each row from the Savitzky-Golay filter with the selected kernel [406]. In this example, the same selected kernel is used for each row of $\boldsymbol{A}$. The implementation of phase-based Helmholtz EPT collected in the open-source C++ library EPTlib 0.1.1 [407] has been adopted.

Finally, since the phantom is homogeneous, the spatial average of the estimated electric conductivity is computed as

$$\overline{\sigma} = \frac{1}{N}\sum_{i=1}^{N} y_i\,. \tag{E5.2.6}$$

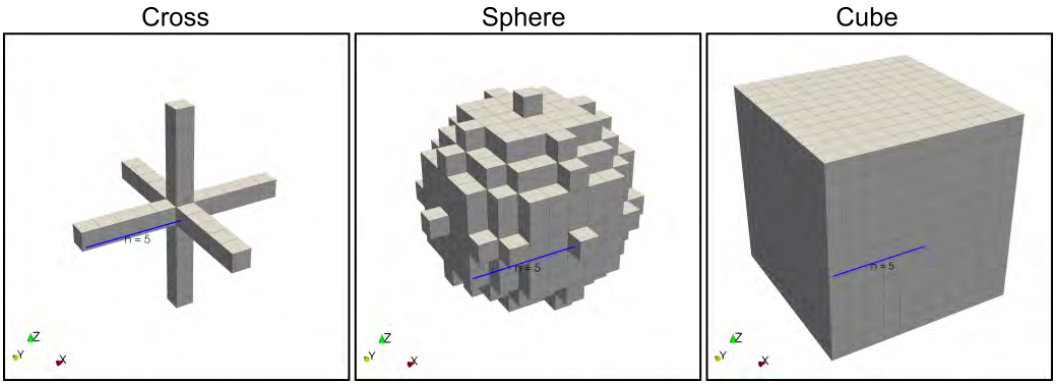


Figure E5.2.1: Considered shapes of the kernel of matrix A. For all the shapes, the size $n = 5$ is depicted.

Table E5.2.1: Number of voxels in each kernel shape and size.

n	Cross	Sphere	Cube
1	7	7	27
2	13	33	125
3	19	123	343
4	25	257	729
5	31	515	1331

As an alternative estimator of the phantom electrical conductivity, also the (generalized) weighted average [408] of the y_i values is investigated. Such estimator is a well-known kind of average that takes into account the variance-covariance matrix of the involved quantities, its analytical expression being

$$\overline{\sigma}_{\mathrm{w}} = \left(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{\Sigma}(\boldsymbol{y})^{-1}\boldsymbol{W}\right)^{-1}\boldsymbol{W}^{\mathrm{T}}\boldsymbol{\Sigma}(\boldsymbol{y})^{-1}\boldsymbol{y}\,, \tag{E5.2.7}$$

where $\boldsymbol{W}$ is a $(N \times 1)$ column vector of ones and $\boldsymbol{\Sigma}(\boldsymbol{y})$ is given in (E5.2.8).

E5.2.4.1 Uncertainty quantification in the input data

Experimental data

The input data to the EPT technique are the φ^+ values measured in the voxels of a three-dimensional image of a homogeneous cylindrical phantom. 25 scans of the phantom have been acquired, in groups of five, with a 3 T MRI scanner using a steady-state free precession (SSFP) sequence with isotropic resolution of 2 mm. The adopted MRI scanner is a 3 T Ingenia TX (Philips Healthcare, Best, The Netherlands) and a 32-channel RF receive head-coil has been used. The five scans of a given group have been acquired one after the other, without breaks, whereas the acquisitions were interrupted for a few minutes between different groups. During the entire procedure, the phantom was never moved. Therefore, the input data (25 φ^+ values for each voxel) were acquired under repeatabilty conditions.

The phase of the complex images acquired with the SSFP sequence provides a good estimate of $2\varphi^+$, so the measurement of the EPT input is practically direct [402]. Moreover, because of the cylindrical shape of the phantom, the magnitude of the complex images approximates well

$|B_1^+|^2$ [409]. The magnitude $|B_1^+|$ can be used to improve the accuracy of Helmholtz-EPT, so its measured value is used here to evaluate the systematic error introduced by performing EPT only on the phase data.

An example of the acquired phase and magnitude maps is reported in Figure E5.2.2. All the acquired data, restricted to the sphere of radius 30 mm centred in the middle of the phantom, are available in the repository [32].

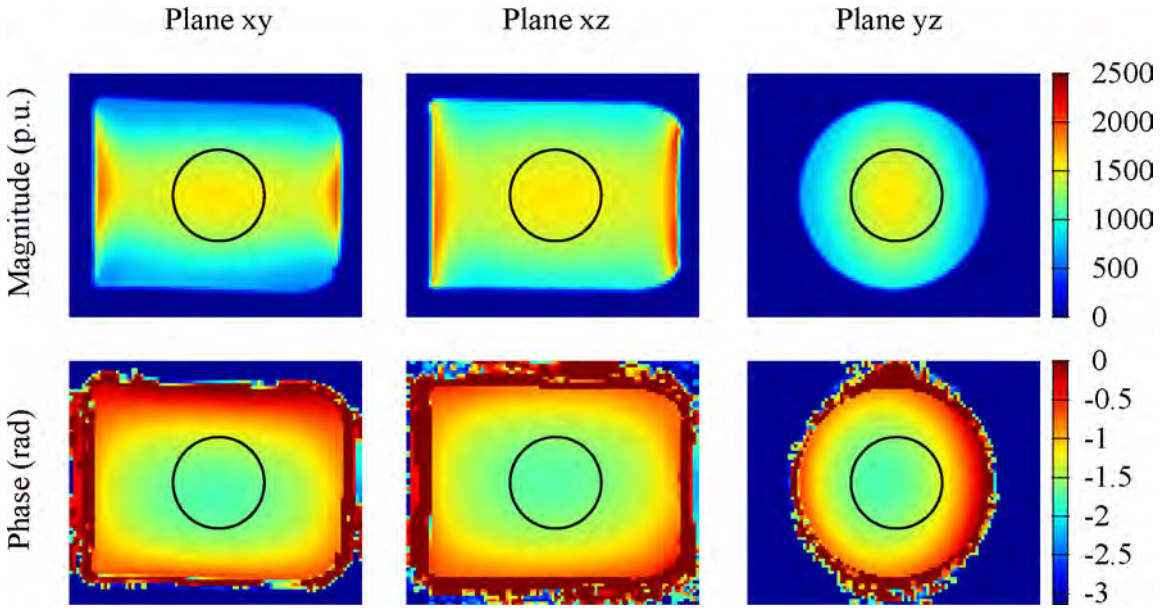


Figure E5.2.2: Three sections of a measured map proportional to the magnitude $|B_1^+|$ (top) and of a measured map of the phase φ^+ (bottom). The sections of the central sphere of voxels analysed in this example is pictured by the black circles.

It is worth saying that a constant difference in two acquired maps of φ^+ can appear. Indeed, for each voxel this phase corresponds to the direction of the local magnetic flux density vector, which rotates periodically around the MRI static field axis. Since the rotation takes place at a constant angular frequency (corresponding to the Larmor frequency), the absolute values of the measured phases depend on the time instant selected for the acquisition, while the phase-shift among voxels keeps the same. Hence, this constant variation of φ^+ does not affect the estimation of the electric conductivity by the Helmholtz-EPT method, since only the gradient of φ^+ is involved in equations (E5.2.2) and (E5.2.4). Nonetheless, in order to avoid an effect of sham correlation between phase values in different voxels due to this constant variation, each map of φ^+ has been translated in such a way to guarantee a null spatial average within the sphere under analysis. After this filtering of the average value, at a visual inspection the phase patterns of the 25 acquisitions appear almost identical, suggesting that the experiment has a high level of repeatability. In particular, throughout the 25 acquisitions, the phase of each single voxel exhibits fluctuations within ± 5 mrad (i.e., within ± 0.3°). Since the phase provides the direction of the rotating magnetic flux density vector (whose period, at the Larmor frequency of 128 MHz, is about 7.8 ns) at the acquisition instant, these data indicate that the timing of the 25 repetitions has a maximum error around 6.2 ps.

Finally, the averages of the 25 (translated) values of φ^+ were calculated, voxel-by-voxel, and provided as input to the EPT method.

Uncertainties and covariances in the input data

Since the number p of voxels in the considered sphere was larger than ten thousand but only 25 scans were available, that is, the number of repeated measurements was smaller than the number p of quantities to be estimated, the variance-covariance matrix associated with the mean value of the measured φ^+ could not be evaluated by the usual experimental variance-covariance matrix, which had at most rank 24.

In order to handle properly the dimension of the problem, a shrinkage estimator of the variance-covariance matrix has been applied, by means of the `cov.shrink` function of the R package corpcor [410]. It implements a James–Stein-type shrinkage estimator for the variance-covariance matrix, with separate shrinkage for the variances and the correlations according to the methods in [411] and [412], respectively.

The obtained input variance-covariance matrix was positive definite and well conditioned. All its terms were hence divided by 25, i.e. the number of repeated measurements, in order to get the variance-covariance matrix associated with the mean phases. The standard uncertainty in the mean phase measurement evaluated in each voxel varies from 0.3 mrad up to 1.1 mrad (the corresponding squared values, i.e. the variances, are on the matrix diagonal). The correlations are both positive and negative and are spatially arranged as shown in Figure E5.2.3 for a selection of reference voxels. Such patterns can be ascribed to a non-ideal response of the receiving coil that embraces the phantom and is used to detect the phase in each voxel. Being composed of a discrete number of annular loops, its response may be slightly anisotropic with respect to the instantaneous direction of the rotating magnetic flux density. As already discussed, the latter is highly stable throughout the repetitions, but inevitably suffers some minor fluctuation.

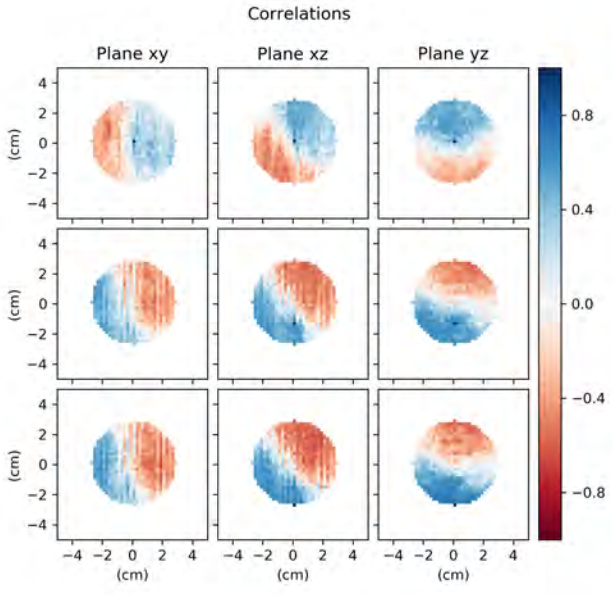


Figure E5.2.3: Spatial distributions of correlations in the input phases with respect to the central voxel (first row), a voxel 15 mm towards the negative z (second row) and a voxel on the sphere boundary (third row).

In the following, $\boldsymbol{x}$ and $\boldsymbol{\Sigma}(\boldsymbol{x})$ indicate the vector of dimension p of the voxel-by-voxel mean phases and the associated $p \times p$ variance-covariance matrix, respectively.

E5.2.5 Uncertainty propagation

The LPU as described in [4] has been applied to propagate the uncertainty through the model.

First, the variance-covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{y})$ of the estimated electric conductivity distribution is obtained by propagating the variance-covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{x})$ of the phase maps through equation (E5.2.5). This is accomplished according to [4, Eq. 3]:

$$\boldsymbol{\Sigma}(\boldsymbol{y}) = \boldsymbol{A}\boldsymbol{\Sigma}(\boldsymbol{x})\boldsymbol{A}^{\mathrm{T}} . \tag{E5.2.8}$$

Then, the output variance-covariance matrix is propagated through the averaging procedures (E5.2.6) and (E5.2.7) to evaluate the standard uncertainty $u(\overline{\sigma})$ and $u(\overline{\sigma}_{\mathrm{w}})$, associated with the corresponding electric conductivity estimates:

$$u(\overline{\sigma}) = \frac{1}{N}\sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N}\big(\boldsymbol{\Sigma}(\boldsymbol{y})\big)_{i,j}} , \tag{E5.2.9}$$

$$u(\overline{\sigma}_{\mathrm{w}}) = \sqrt{\big(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{\Sigma}(\boldsymbol{y})^{-1}\boldsymbol{W}\big)^{-1}} . \tag{E5.2.10}$$

The homogeneity of the phantom justifies the choice of a spatial average of y_i values as an estimator of the electrical conductivity in the phantom. It is worth noting that the covariances between y_i values in the different voxels prevent the uncertainty in equation (E5.2.9) from vanishing when N increases.

E5.2.6 Reporting the result

The propagation of the variance-covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{x})$ of the input phases through the phase-based Helmholtz-EPT method, computed according to equation (E5.2.8), leads to different results depending on the adopted kernel shape and size, as shown in Figure E5.2.4. In particular, larger variances can be seen when the kernel size is small or the shape involves few voxels. The variances are reduced by enlarging the kernel size. Significant covariances are present in the output. It is worth noting that the number N of voxels in which the electric conductivity is computed depends on the kernel adopted, because, in case a kernel partially falls outside the investigated sphere, the EPT reconstruction is not performed. For this reason, the number of rows and columns of the output variance-covariance matrices decreases by increasing the kernel size.

The variances collected in Figure E5.2.4 suggest that the standard uncertainty of the electric conductivity estimated in each voxel ranges from $1.8\,\mathrm{mS\,m^{-1}}$ to $0.4\,\mathrm{S\,m^{-1}}$, depending on the kernel adopted. However, despite the homogeneity of the phantom used, there is a significant spatial variability in the estimated electric conductivity that cannot be explained only by means of such small uncertainties. No regular geometrical pattern can be identified in the spatial distribution of the estimated electric conductivity, which varies irregularly in the domain. For any kernel shape and size, the collection of the electric conductivity values estimated within the sphere shows a symmetric distribution whose dispersion is reduced by increasing the kernel size, as summarised in the boxplots of Figure E5.2.5. The observed reduction does not scale linearly with the number of voxels in the kernel, which is reported in Table E5.2.1, suggesting that the degree to which

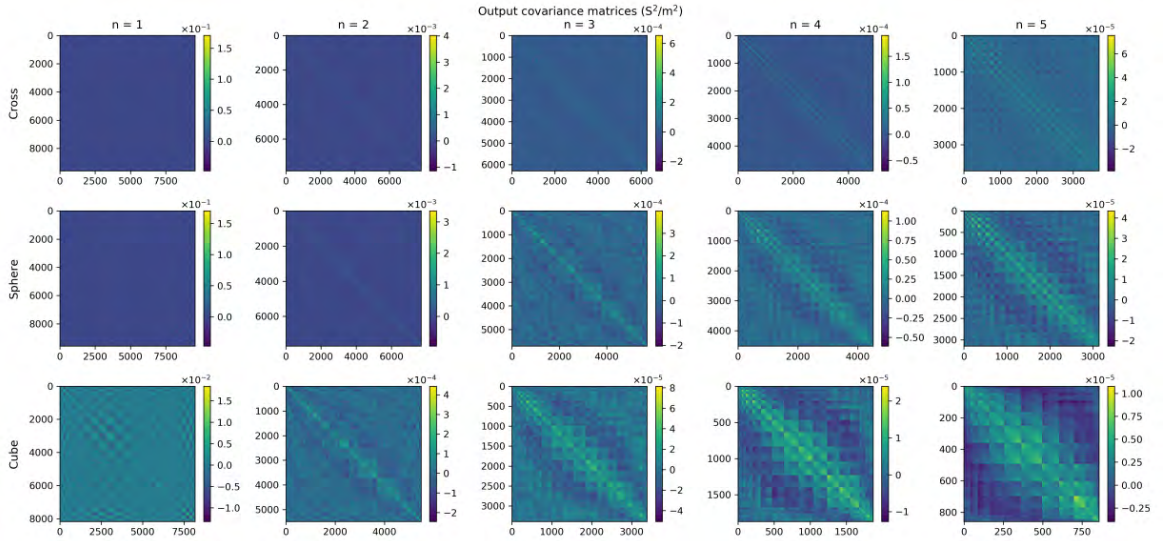


Figure E5.2.4: Visual representation of the variance-covariance matrices $\boldsymbol{\Sigma}(\boldsymbol{y})$ of the output $\boldsymbol{y}$ computed with kernels of different shapes and sizes.

each voxel in the kernel contributes to the overall homogeneity is not constant and depends on the kernel shape. The spatial variability of the estimated electric conductivity is a local error inherent to the Helmholtz-EPT method that can be reduced by changing the kernel shape and size. Such an error changes irregularly from point to point within the domain, so that it has the appearance of a random error when looking at the whole population of estimated conductivity values. Moreover, another systematic error is due to the phase-based approximation and can be corrected by providing the magnitude $|B_1^+|$ to the Helmholtz-EPT method as an additional input. Indeed, with the complete input, the average value (corresponding to the median for symmetric distributions) of the collection of the estimated electric conductivity values moves towards the nominal value, despite the spatial variability remains the same (cf. Figure E5.2.5).

By comparison with the nominal value, the boxplots of Figure E5.2.5 suggest that the spatial average $\overline{\sigma}$ of the electric conductivity estimated in the several voxels is a good estimate of the electric conductivity of the homogeneous phantom. The phase-based approximation introduces a systematic error in $\overline{\sigma}$, estimated between 3 % and 8 % of the nominal value, depending on the kernel adopted. The standard uncertainty $u(\overline{\sigma})$ under repeatability conditions, obtained through equation (E5.2.9), is very low, with a relative uncertainty of about 0.2 %. The computed values of $\overline{\sigma}$ and their uncertainties are collected in Table E5.2.2 for all the kernel shapes and sizes. It is worth noting that the uncertainty reported here does not take into account the spatial variability of the estimates; it expresses only the repeatability of the measurement. Thus, for a homogeneous phantom, the spatial average of the output of phase-based Helmholtz-EPT appears to be very precise, from the repeatability point of view, and to have an accuracy affected by an error smaller than 10 % due to the phase-based approximation.

Table E5.2.2 reports also the standard uncertainty of $\overline{\sigma}$ that would have been obtained by neglecting the correlations in the input phases. Its value, denoted by $\tilde{u}(\overline{\sigma})$, is always smaller than the actual standard uncertainty $u(\overline{\sigma})$, highlighting the importance of taking into account the correlations between the input quantities. Moreover, it is worth noting that the uncertainties propagated through the cross-shaped and the spherical kernels are quite similar to each other and different from the values obtained with the cubic kernel. This difference is only partially reflected by the kernel volumes collected in Table E5.2.1.

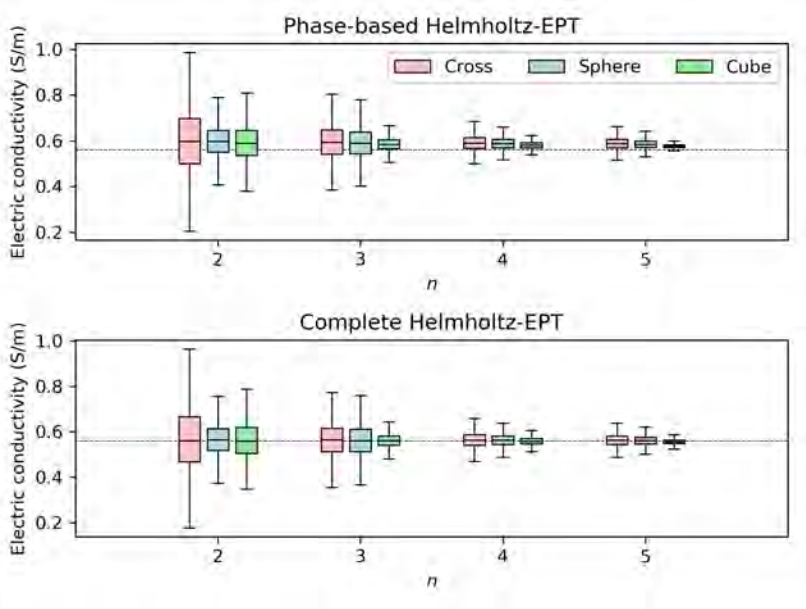


Figure E5.2.5: Boxplots of the electric conductivities $\boldsymbol{y}$ estimated in all the voxels of the analysed sphere, computed with kernels of different shapes and sizes. The dashed horizontal line denotes the nominal electric conductivity of the phantom. The result of the phase-based Helmholtz-EPT method (top) is compared with the result of the complete Helmholtz-EPT method (bottom).

In Table E5.2.2, the $\overline{\sigma}_{\mathrm{w}}$ estimates and the associated uncertainties are shown as well. The former are closer to the nominal value with respect to $\overline{\sigma}$ values (and more stable among shapes and sizes of the kernels); the latter are smaller than the uncertainties associated with $\overline{\sigma}$ values of about one order of magnitude, hence leading to a relative uncertainty of about 0.01 %.

E5.2.7 Interpretation of results

In this case study, the estimate of the electric conductivity obtained by the phase-based Helmholtz-EPT method in each voxel of the imaged phantom can be interpreted as a set of measurements of different measurands, which, by virtue of the *a priori* information on the homogeneity of the phantom, should ideally exhibit the same numerical value. Therefore, the values of the electric conductivity estimated in the voxels were compared to assess the spatial dispersion of the voxel-by-voxel estimates. Such variability proved to be reduced by increasing the kernel size (cf. Figure E5.2.5). Indeed, large kernels approximate the derivatives by averaging more input data, hence increasing the repeatability precision of the estimates and reducing their spatial variability. On the other hand, kernels too large could introduce significant numerical errors in the approximation of the derivatives and, especially in heterogeneous domains, could worse the estimation.

Moreover, in this example, it has been seen that the spatial average of the collection of the electric conductivity values estimated in the sphere was a stable estimate of the nominal value, weakly affected by the choice of the kernel and with a significantly small standard uncertainty under repeatability conditions (cf. Table E5.2.2). A systematic error in the average of less than 10 % with respect to the nominal value was due to the phase-based approximation and was almost completely corrected by the introduction of the magnitude $|B_1^+|$ as an additional input

Table E5.2.2: Average ($\overline{\sigma}$) and weighted average ($\overline{\sigma}_\mathrm{w}$) of the electric conductivity values obtained by the phase-based Helmholtz EPT and their standard uncertainties. The standard uncertainty ($\tilde{u}(\overline{\sigma})$) computed neglecting the input covariances is reported as well for the average. All data are provided for kernels with different shapes and sizes.

	n	1	2	3	4	5
Cross	$\overline{\sigma}$ / mS m^{-1}	605.102	600.247	595.509	592.737	590.278
	$u(\overline{\sigma})$ / mS m^{-1}	0.980	0.683	0.725	0.778	0.797
	$\tilde{u}(\overline{\sigma})$ / mS m^{-1}	0.963	0.287	0.171	0.126	0.105
	$\overline{\sigma}_\mathrm{w}$ / mS m^{-1}	586.238	586.609	586.788	586.666	586.765
	$u(\overline{\sigma}_\mathrm{w})$ / mS m^{-1}	0.051	0.052	0.056	0.064	0.071
Sphere	$\overline{\sigma}$ / mS m^{-1}	605.102	599.126	592.654	590.074	586.656
	$u(\overline{\sigma})$ / mS m^{-1}	0.980	0.669	0.712	0.801	0.862
	$\tilde{u}(\overline{\sigma})$ / mS m^{-1}	0.963	0.283	0.162	0.126	0.104
	$\overline{\sigma}_\mathrm{w}$ / mS m^{-1}	586.238	586.811	585.451	586.550	584.741
	$u(\overline{\sigma}_\mathrm{w})$ / mS m^{-1}	0.051	0.054	0.057	0.062	0.067
Cube	$\overline{\sigma}$ / mS m^{-1}	600.452	592.570	586.605	582.032	578.445
	$u(\overline{\sigma})$ / mS m^{-1}	0.645	0.709	0.819	0.949	1.101
	$\tilde{u}(\overline{\sigma})$ / mS m^{-1}	0.351	0.143	0.103	0.097	0.118
	$\overline{\sigma}_\mathrm{w}$ / mS m^{-1}	586.239	585.582	585.287	582.176	580.104
	$u(\overline{\sigma}_\mathrm{w})$ / mS m^{-1}	0.051	0.056	0.062	0.077	0.098

(cf. Figure E5.2.5). An even more accurate and precise estimate was provided by the weighted average $\overline{\sigma}_\mathrm{w}$ (cf. Table E5.2.2), which gave more importance to the voxel values with a smaller uncertainty by using the inverse of the variance-covariance matrix $\mathbf{\Sigma}(\boldsymbol{y})$ as the weight.

The results collected here suggest that the output maps of Helmholtz-EPT could be significantly improved by post-processing. In particular, for each voxel, a sphere of voxels around it can be selected and the average (or rather the weighted average, whenever possible) of the electric conductivity values estimated there can be assigned to that voxel. This operation would improve the local accuracy and the precision of the Helmholtz-EPT. Post-processing filters analogous to the one described here are already applied to Helmholtz-EPT results, for instance the median filter adopted in the clinical investigation conducted in [405]. Actually, when applied *in vivo*, being the domain heterogeneous, the sphere of voxels could be deformed into a volume locally adapted to the anatomy, in order to not cross the tissue boundaries.

Example E5.3

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

E5.3.1 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 method (WTLS) 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.

E5.3.2 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 E5.3.A. For external quality assurance of routine laboratories, inter-laboratory comparisons are performed in which the deviation from the reference value may not exceed 6 % [413]. 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 [414–416] 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 [417, 418] since a primary calibrator exists.

Previous comparisons of the HiCN and the AHD methods with high-accuracy procedures [417, 419] 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 [420–422] 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 $\mathrm{g\,L^{-1}}$ to 190 $\mathrm{g\,L^{-1}}$ 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 E5.3.1 and can be found online in repository [33]. Derivation of the total Hb concentration involves quantities common to both methods and all samples (cf. Appendix E5.3.B for background information). Some of these common quantities contribute significantly to the uncertainty of the Hb concentration [417, 419]. Therefore, it is reasonable to suspect significant correlation among the HiCN as well as among the AHD method (cf. clause 5.2.4 in [2]).

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 WTLS. The measurement model allows for uncertainty evaluation following the GUM. The WTLS accounts for uncertainties in both variables of a regression, as well as, for correlation among and between them. WTLS is recommended by multiple standards [77, 110] and applied in metrology (e.g. Refs. [423–425]).

[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.

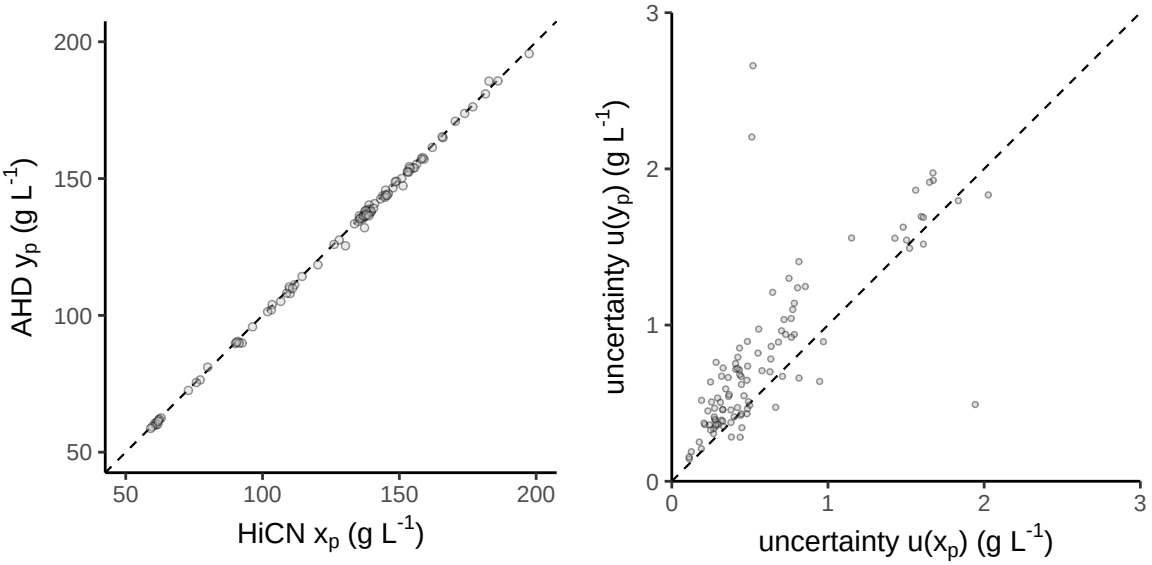


Figure E5.3.1: Left: Visualization of haemoglobin concentration measurements x_p, y_p performed at PTB on $P = 104$ blood samples by the two methods HiCN and AHD.
Right: Standard uncertainties $u(x_p)$ and $u(y_p)$ for both methods and all samples. These measurement results are available online in repository [33]. In both panels, the dashed line represents the identity $y = x$ and the markers are drawn as transparent; thus, overlayed markers appear darker.

E5.3.3 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

$$Y = \beta_0 + \beta_1 X \tag{E5.3.1}$$

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 methods (see [420–422] and references therein). The measurands are the intercept parameter β_0 and slope parameter β_1 of the straight-line model (E5.3.1). 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 E5.3.1). 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 E5.3.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 [3].)

E5.3.4 Measurement model

The measurement model for straight-line regression can be constructed from the appropriate least-squares method. The frequently applied ordinary least-squares method (OLS) and weighted least-squares method (WLS) 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 [310] and Passing-Bablok regression [426], 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 E5.3.1); 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.

WTLS is the method recommended by multiple standards [77, 110] when the uncertainty associated with the measured values x_p and y_p are both non-negligible. It also addresses correlation. The WTLS is based on minimizing the generalized sum of squares

$$Q = \left(\boldsymbol{x} - \widetilde{\boldsymbol{\xi}}\right)^{\top} \boldsymbol{U}_x^{-1} \left(\boldsymbol{x} - \widetilde{\boldsymbol{\xi}}\right) + \left(\boldsymbol{y} - (\widetilde{\beta}_0 + \widetilde{\beta}_1 \widetilde{\boldsymbol{\xi}})\right)^{\top} \boldsymbol{U}_y^{-1} \left(\boldsymbol{y} - (\widetilde{\beta}_0 + \widetilde{\beta}_1 \widetilde{\boldsymbol{\xi}})\right) \tag{E5.3.2}$$

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 (E5.3.2) defines the solution $(\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\boldsymbol{\xi}}^{\top})$ of the WTLS.

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,

$$(\beta_0, \beta_1, \boldsymbol{\xi}^{\top})^{\top} = \arg \min_{\widetilde{\beta}_0, \widetilde{\beta}_1, \widetilde{\boldsymbol{\xi}}} \left\{ \left(\boldsymbol{X} - \widetilde{\boldsymbol{\xi}}\right)^{\top} \boldsymbol{U}_x^{-1} \left(\boldsymbol{X} - \widetilde{\boldsymbol{\xi}}\right) + \left(\boldsymbol{Y} - (\widetilde{\beta}_0 + \widetilde{\beta}_1 \widetilde{\boldsymbol{\xi}})\right)^{\top} \boldsymbol{U}_y^{-1} \left(\boldsymbol{Y} - (\widetilde{\beta}_0 + \widetilde{\beta}_1 \widetilde{\boldsymbol{\xi}})\right) \right\}, \tag{E5.3.3}$$

where only (β_0, β_1) define the measurand.

E5.3.5 Estimation and uncertainty evaluation

Following the GUM [2, 4], estimates $\widehat{\beta}_0$ and $\widehat{\beta}_1$ of the measurands are obtained by evaluating measurement model (E5.3.3) 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 (E5.3.3) 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 [77]. 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

$$\left[\widehat{\beta}_i - 1.96\, u\left(\widehat{\beta}_i\right), \widehat{\beta}_i + 1.96\, u\left(\widehat{\beta}_i\right)\right].$$

[2] Cf. section 10.2.3 in [77] for the approximate validity of this Normality assumption.

A two-dimensional, joint 95 % coverage region can be calculated following clause 6.5.2 in [4].

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 [77] describes how these covariances 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 [4], where distributions for all input quantities are propagated through a joint measurement model to arrive at the $2P$-dimensional, joint distribution for the outputs $\boldsymbol{X}$ and $\boldsymbol{Y}$.

The uncertainty in HiCN and in AHD measurements is dominated by a common quantity, namely the molar extinction coefficient ϵ (see [417, 419]). We thus suspect that the covariance matrices $\boldsymbol{U}_x$ and $\boldsymbol{U}_y$ are governed by a common correlation coefficient ρ. That is, we set their elements $\boldsymbol{U}_{x,pq} = \rho u(x_p)u(x_q)$ and $\boldsymbol{U}_{y,pq} = \rho u(y_p)u(y_q)$ for all $p \neq q$. The diagonal elements contain the variances, i.e. $\boldsymbol{U}_{x,pp} = u^2(x_p)$ and $\boldsymbol{U}_{y,pp} = u^2(y_p)$. Further details are given in Appendix E5.3.C. First Monte Carlo evaluations of the joint uncertainty budget showed that correlation 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 [350]. The correlation between HiCN and AHD is dominated by two common quantities, viz., the cuvettes' absorption length d and the mean molar mass $M(\mathrm{Hb})$. According to [417], the amount of cross-correlation is much smaller compared to correlation 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 conditional 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 E5.3.B.

E5.3.6 Reporting the result

Let us now apply the measurement model (E5.3.3) to the estimates and uncertainties presented in figure E5.3.1 and to the above covariance structures $\boldsymbol{U}_x$ and $\boldsymbol{U}_y$. For selected correlation coefficients ρ, the results are listed in table E5.3.1. 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 [77] and application of the law of propagation of uncertainty [2]. R Markdown [356] code for this algorithm is available online in repository [33]. Figure E5.3.2 depicts the estimates $\widehat{\beta}_0$ and $\widehat{\beta}_1$ and the corresponding 95 % coverage interval.

Table E5.3.1: Results obtained by weighted total least-squares with uncertainty evaluation according 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 [4] to the measurement model (E5.3.3) and the algorithm in Ref. [77]. The non-linearity of (E5.3.3) could cause differences; however, this was not observed. Software is available that implements WTLS

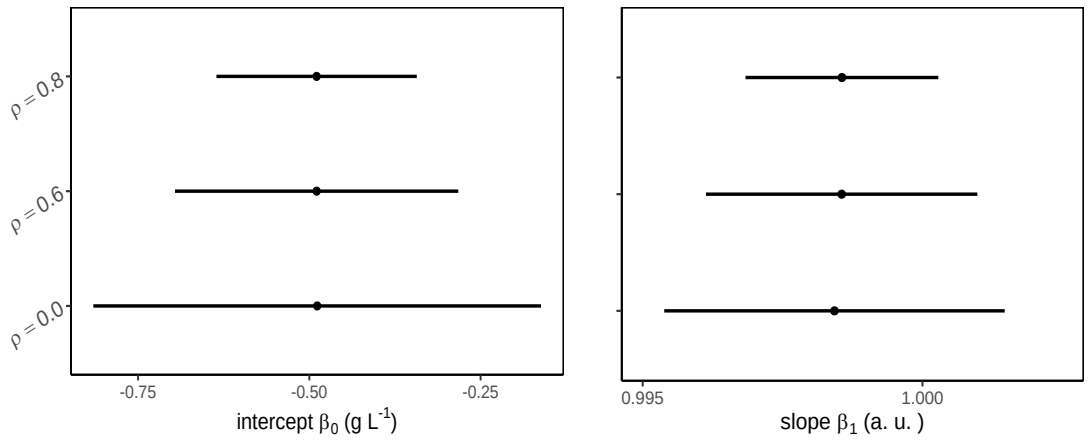


Figure E5.3.2: Displayed are the estimates $\widehat{\beta}_0$ and $\widehat{\beta}_1$ (dots) and their 95 % coverage intervals for the weighted total least-squares regression results listed in table E5.3.1.

and propagates uncertainties. For example, the CALIBRATION CURVE COMPUTING Software provided by INRIM [424] also produces the results in table E5.3.1, although a slightly different algorithm is implemented (which relies on an implicit set of normal equations).

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 (E5.3.1), since no systematic behaviour of these residuals were observed. A significant outcome of the χ^2 test, whose application is recommended in standard [77], 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 (E5.3.3). 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 (E5.3.1). The former does not contradict the assumptions of our analysis.

E5.3.7 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, OLS and WLS are inappropriate. Instead, WTLS 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 relationships are detailed in Appendix E5.3.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 ([427] 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.

E5.3.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 [428]. Deviations of the Hb concentration from the normal range (137 $\mathrm{g\,L^{-1}}$ – 162 $\mathrm{g\,L^{-1}}$ for men and 123 $\mathrm{g\,L^{-1}}$ – 145 $\mathrm{g\,L^{-1}}$ for women; c.f. [429, 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 [430], malaria [431] or thalassemia, the most common genetic disorder worldwide [432]. In addition, haemoglobin concentration is relevant to manage iron deficiency in pregnant women [433].

Total haemoglobin concentration is determined by a variety of methods [434], 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 [435]. 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 [436], has been considered as a gold standard for routine applications [434] and is also internationally accepted as higher-order method [414–416] to determine reference measurement values in external quality assurance of medical laboratories [413]. 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 [437, 438].

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 [413] 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 [416, p. 18] and was met in comparison experiments [417, 419]. 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 [416]. 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 [418] 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 [417, 418].

E5.3.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 $p = 1, \ldots, P$ and dilution i (cf. [419]). As recommended in DIN 58931:2010 [416], 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

$$w_{i,p}^k = \frac{\left(C_0^k + C_1^k a_{i,p}^k\right) M(\text{Hb})}{d\epsilon^k \phi_i}. \tag{E5.3.4}$$

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 (E5.3.4) can be found in [419]. 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. [419].

E5.3.C Influence of correlation for a common structure

If the covariance matrix for the HiCN method is given by

$$\boldsymbol{U}_x = (1-\rho)\text{diag}(\boldsymbol{u}_x^2) + \rho \boldsymbol{u}_x \boldsymbol{u}_x^\top$$

[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.

with $\boldsymbol{u}_x = (u(x_1), \ldots, u(x_P))^\top$, as described in section E5.3.5, the inverse of $\boldsymbol{U}_x$ is determined by

$$\boldsymbol{U}_x^{-1} = \frac{1}{1-\rho}\left[\operatorname{diag}\left(\frac{1}{u^2(x_1)}, \ldots, \frac{1}{u^2(x_P)}\right) - \frac{1}{P-1+1/\rho}\left(\frac{1}{\boldsymbol{u}_x}\right)\left(\frac{1}{\boldsymbol{u}_x}\right)^T\right].$$

The inverse $\boldsymbol{U}_y^{-1}$ can be determined by analogy. Then, the generalized sum of squares (E5.3.2) simplifies to

$$\begin{aligned} Q = \frac{1}{1-\rho}\Bigg(&\sum_{p=1}^{P}\frac{(x_p-\xi_p)^2}{u^2(x_p)} + \frac{(y_p-\beta_0-\beta_1\xi_p)^2}{u^2(y_p)} \\ &-\frac{1}{P-1+1/\rho}\left(\sum_{p,q=1}^{P}\frac{(x_p-\xi_p)(x_q-\xi_q)}{u(x_p)u(x_q)} + \sum_{p,q=1}^{P}\frac{(y_p-\beta_0-\beta_1\xi_p)(y_q-\beta_0-\beta_1\xi_q)}{u(y_p)u(y_q)}\right)\Bigg). \end{aligned}$$

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 $\operatorname{cov}(\widehat{\beta}_0,\widehat{\beta}_1)$ 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 E5.3.2 shows the influence of the correlation coefficient 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 E5.3.1 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. $\widehat{\beta}_1^{\text{corr}} - \widehat{\beta}_1 \approx u(\widehat{\beta}_1)/2$) and at the same time the uncertainty reduces considerably.

Table E5.3.2: Results obtained by weighted total least-squares with uncertainty evaluation according to the GUM for a subset of size $P = 20$ of the data in figure E5.3.1. 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 E5.4

Suitability of a Monte Carlo approach for uncertainty evaluation in rheology problems

J.A. Sousa, A. Furtado, J. Pereira, M.G. Cox, A.S. Ribeiro, M. Reader-Harris, A.M.H. van der Veen

E5.4.1 Summary

This example presents a method for the calibration of rheometers and evaluates the measurement uncertainty associated with the expression of viscosity derived for a rotational rheometer equipment using an alternative method from that in the GUM. This method is the Monte Carlo implementation of the propagation of distributions from GUM Supplement 1 (GUM-S1) 1, which can treat measurement models having any degree of non-linearity and possibly large input uncertainties. It is therefore particularly suited to handling the complexity of the applicable measurement model. The Monte Carlo method is used to validate the more commonly-used GUM approach, identifying conditions in which the application of the GUM might be acceptable and some difficulties associated with its implementation.

E5.4.2 Introduction of the application

This study is partly based on an accepted extended abstract [439], aiming at specifying general methods for the calibration of rheometers and, additionally, covers the uncertainty budget of viscosity measurements performed by rotational methods, since the complexity of the measurement model raises doubts on the validity of the GUM approach. The main quantity to be evaluated is the dynamic viscosity, expressed in Pa s.

In order to relate the measurement results obtained by viscosity sensors, specifically rotational rheometers, to SI base units, the following sequence of steps needs to be accomplished:

1. calibration of the standard rotational rheometer traceable to SI units;
2. use of Certified Reference Materials (CRMs) for viscosity (by tracing flow and viscosity curves);

3. calibration of a laboratory rheometer (each of the following quantities needs to be calibrated independently: viscosity by using a CRM; shear rate (indirectly through rotational velocity) and shear stress (indirectly through torque calibration)); temperature; dimensions of the measuring geometries, including gap geometry;
4. production of a reference material (RM) with specified viscosities by a standardised method and subsequent determination of its viscosities at specified shear rates by using the laboratory rheometer calibrated in step 3;
5. calibration of the viscosity sensors by using the RM from step 4.

Uncertainty necessarily increases with every step in this sequence. Steps 1 and 2 would be carried out by an national metrology institute (NMI) and steps 3-5 by the laboratory itself.

Input quantities, namely, those on which the required dynamic viscosity depends, are specific to a given measurement principle. The main quantities considered significant when determining the viscosity of non-Newtonian liquids by using a rotational rheometer are torque; rotational speed; temperature (due to the increase in uncertainty at high shear rates generated by frictional heating); dimensions (radii, lengths, angles, measuring gap); end effect correction [440–443] measurement model (choice of representative location, "narrow gap" approximation); measurement time (affects temperature via frictional heating and laminar flow field establishment within the gap) [444]; repeatability; reproducibility; sample shear history. These quantities influence the evaluation of the uncertainty of the viscosity of a CRM and represent parameters that must be controlled carefully when applying the CRM to the subsequent calibration of rheometers.

Alternative approaches are discussed, drawing attention to shortcomings of the GUM uncertainty framework (GUF) [2] and the advantages of the propagation of distributions approach, implemented using a Monte Carlo method [3].

E5.4.3 Specification of the measurand(s)

The measurand is the dynamic viscosity η at reference temperature T_{ref} , which, according to standards ISO 3219 (1993) [445] and DIN 53019-2 (2001) [446], is determined by

$$\eta(T_{\mathrm{ref}}) = \frac{M\left[\left(\frac{R_{\mathrm{a}}}{R_i}\right)^2 - 1\right]}{8\pi^2 1 n L R_i^2 c_{\mathrm{L}} \left(\frac{R_{\mathrm{a}}}{R_i}\right)^2} \left[1 - \beta_{\mathrm{T}}(T - T_{\mathrm{ref}})\right] \tag{E5.4.1}$$

based on the fundamental equation that defines the dynamic viscosity as the ratio between shear stress τ and shear rate $\dot{\gamma}$:

$$\eta = \frac{\tau}{\dot{\gamma}} \tag{E5.4.2}$$

where the (input) quantities on which η depends are:

M torque in N m,
R_{a} outer radius of the gap in m,
R_i inner radius of the gap in m,
n number of rotations per time unit in s^{-1}
L length of the cylinder mantle in m,
c_{L} end-effect correction factor,

β_T linear coefficient of change of viscosity with temperature in °C^{-1}
T measurement temperature in °C, and
T_{ref} reference temperature in °C

The measurement model (E5.4.1) is clearly non-linear and thus may depart from the conditions envisaged for the application of the GUM. Moreover, it has a degree of complexity that makes the analytical calculation of partial derivatives a cumbersome task that is error-prone.

E5.4.4 Measurement model

The calibration method considered here and expressed in section E5.4.1, corresponds to a direct method, where all the measured input quantities, i.e. torque, angular velocity, temperature and geometric dimensions are compared with reference values.

To be used as an example in this study, the direct method for calibration of a rotational viscometer or a rheometer requires a separate calibration of the torque and the angular velocity of the measuring geometry. Additionally, the dimensions (diameters and angles) of the measuring geometry and the temperature should also be calibrated by adequate means. The direct method is described in standard documents such as DIN 53019-2 (2001) [446]; ASTM E2510-07 (2013) [447] (for torque calibration) and ASTM E2509-14 (2014) [448] (for temperature calibration). Expression (E5.4.1) will be the measurement model used in this study for the evaluation of measurement uncertainty together with the data of table E5.4.1.

Table E5.4.1: Example of an uncertainty budget of the measurement of a Newtonian oil using a rheometer with a concentric cylinder system according to DIN 53019 standard [446] (and reference temperature 20 °C) using the GUM uncertainty framework (GUF).

Quantity and Symbol	Estimate x_i	PDF	$u(x_i)$	$\partial\eta/\partial x_i$	$u_{\mathrm{rel},i}(y)$
Torque M	5.7938×10^{-5} N m	R	2.89×10^{-7}	1.882 48	29.5 %
Number of rotations per time unit n	0.012 898 5 N^{-1}	G	5.00×10^{-7} N^{-1}	−0.008 46	0.2 %
Outer gap radius R_a	0.014 458 m	R	2.89×10^{-6} m	0.085 49	13.4 %
Inner gap radius R_i	0.013 329 5 m	R	2.89×10^{-7} m	−0.109 09	1.7 %
Cylinder mantle length L	0.040 009 m	R	2.89×10^{-7} m	−0.002 72	
End effect correction factor c_L	1.1	R	5.77×10^{-3}	−0.000 10	31.0 %
Measurement temperature T	20 °C	G	5×10^{-2} °C	−0.000 01	20.2 %
Reference temperature T_{ref}	19.99 °C	G	1×10^{-2} °C	0.000 01	4.0 %
Temperature coefficient of viscosity β_T	0.0683 K^{-1}	R	5.77×10^{-5} K^{-1}	−0.000 00	
Viscosity η	1.091		$u(y) = 0.009$	$U_{95\,\%} = 0.018$	

E5.4.5 Uncertainty propagation

Measurement uncertainty evaluation includes components arising from random and systematic effects, such as components associated with corrections and the assigned quantity values of measurement standards, as well as definitional uncertainty. In practice, systematic effects, which

Table E5.4.2: Example of an uncertainty budget of the measurement of a Newtonian oil using a rheometer with a concentric cylinder system according to DIN 53019 standard [446] (and reference temperature 20 °C) using the Monte Carlo method (MCM).

Quantity X_i	Unit	Estimate x_i	PDF	$u(x_i)$	$\partial\eta/\partial x_i$	$u_{\mathrm{rel},i}(y)/\%$
M	N m	5.7938×10^{-5}	R	2.89×10^{-7}	1.882 48	29.5
n	N^{-1}	0.012 898 5	G	5.00×10^{-7}	−0.008 46	0.2
R_a	m	0.014 458	R	2.89×10^{-6}	0.085 49	13.4
R_i	m	0.013 329 5	R	2.89×10^{-7}	−0.109 09	1.7
L	m	0.040 009	R	2.89×10^{-7}	−0.002 72	
c_L	1	1.1	R	5.77×10^{-3}	−0.000 10	31.0
T	°C	20.00	G	5×10^{-2}	−0.000 01	20.2
T_ref	°C	19.99	G	1×10^{-2}	0.000 01	4.0
β_T	K^{-1}	0.0683	R	5.77×10^{-5}	−0.000 00	
η	Pa s	1.091		$u(\eta) = 0.009$		
				$U_{95\,\%}(\eta) = 0.018$		

are treated here, are often overlooked with a consequent optimistic uncertainty for η. Definitional uncertainty is the uncertainty resulting from the specification of the measurand and is a very difficult uncertainty to treat in general because of its conceptual overtones. Here we can say that the measurement model and measurand are specified in an international and a national standard and so working correctly with those definitions ensures adherence to those standards. Two approaches for the evaluation of uncertainty are considered: the GUF, based on the law of propagation of uncertainty (LPU) [2], and the propagation of distributions of GUM-S1 based on the Monte Carlo method (MCM) [3].

Both approaches depend on knowledge of the PDFs for the input quantities, but whereas the GUF 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 purposes when the conditions for the use of the GUM approach, as given in [3] clause 5.8, are not fully met.

E5.4.5.1 GUM Uncertainty Framework

The GUF requires the calculation of sensitivity coefficients c_i, the first partial derivatives of the viscosity η measurement function (the right side of expression (E5.4.1)) with respect to the input quantities, evaluated at the estimates of those quantities. As for many complicated models, determining the required partial derivatives algebraically is not always practical and a numerical approach is recommended [449]. It is often a complicated step, which is completely avoided by the GUM-S1 approach. That approach, however, is computationally expensive for very complicated models.

The GUM method applied to a univariate model relates a single output quantity Y to input quantities $\boldsymbol{X} = [X_1, \ldots, X_N]^\top$ by a functional relationship f, here in the form of expression (E5.4.1). The estimate of the output quantity is taken as $y = f(\boldsymbol{X})$, evaluated at $\boldsymbol{X} = \boldsymbol{x}$, the estimate of $\boldsymbol{X}$.

The associated standard uncertainty $u(y)$ is evaluated from the LPU:

$$u^2(y) = \sum_{i=1}^{N}\sum_{j=1}^{N} c_i u(x_i, x_j) c_j, \tag{E5.4.3}$$

given in the GUM, where c_i is the partial derivative of f with respect to $\boldsymbol{X}$, evaluated at $\boldsymbol{X} = \boldsymbol{x}$, and is known as the ith sensitivity coefficient. The evaluation of the c_i require care since analytic expressions for the partial derivatives of the measurement function are often difficult to obtain. The usual approach uses a numerical method to approximate these derivatives. We recommend the complex-step method [449] based on the Taylor expansion of a function of a complex variable for this purpose: it is numerically extremely accurate, avoiding subtractive cancellation.

By defining the covariance matrix

$$\boldsymbol{U}_{\boldsymbol{x}} = \begin{bmatrix} u(x_1, x_2) & \cdots & u(x_1, x_N) \\ \vdots & \ddots & \vdots \\ u(x_N, x_1) & \cdots & u(x_N, x_N) \end{bmatrix}$$

of dimension $N \times N$, containing the covariances $u(x_i, x_j)$, and the (row) vector $\boldsymbol{c}^\top = [c_1, \ldots, c_N]$ of dimension $1\times$N, containing the sensitivity coefficients, a compact way of writing expression (E5.4.3) is

$$u^2(y) = \boldsymbol{c}^\top \boldsymbol{U}_{\boldsymbol{x}} \boldsymbol{c}. \tag{E5.4.4}$$

For independent input quantities, a simpler formulation is often used, where the variance $u^2(y)$ can be viewed as a sum of terms $u^2(x_i)$, so that each input quantity contributes to the final value of the squared standard uncertainty (variance). In this case, expression (E5.4.3) can written as

$$u^2(y) = \sum_{i=1}^{N} [c_i u(x_i)]^2 = \sum_{i=1}^{N} u_i^2(y)$$

with the various $u_i(y)$ in the uncertainty budget often being used to give a measure of the importance of each input quantity.

Given a coverage probability p, a $100p$% coverage interval for Y is defined as $y \pm U_p$ with expanded uncertainty $U_p = k_p u(y)$ and coverage factor k_p. In metrology it is common to take $p = 0.95$, yielding a value of $k_p = 2$ for a normal distribution (a scaled and shifted t distribution is used in some cases to determine k_p [3]).

E5.4.5.2 Propagation of distributions

A MCM is an implementation of the propagation of distributions and is generally considered a very reliable approach for uncertainty propagation. The PDFs for the input quantities are propagated through the measurement model to provide the PDF for the output quantity, with no restriction in relation to shape or symmetry of the input PDFs. The expectation (mean) of the resulting PDF is used as the best estimate of the quantity and the standard deviation of the PDF is used as the standard uncertainty associated with that estimate.

Appropriate use of MCMs will provide valid results when the applicability of the GUF is questionable. Aspects to be carefully considered include the number of samples taken of the model input quantities, and the related numerical accuracy of the results obtained. Guidance is provided in GUM-S1 [3].

From the PDF for the output quantity Y, a coverage interval for Y for any coverage probability p (usually taken as 0.95) can be obtained. Such an interval contains the value of Y with probability p. A coverage interval is not unique: a shortest 95 % coverage interval will provide values with the highest density, whereas for a symmetrical unimodal PDF the percentiles 0.025 and 0.975 should be used.

An advantage of this approach is that it provides a PDF for Y that is consistent with the input PDFs. The assumption in the GUM that the PDF for Y is Gaussian (or a scaled and shifted t-distribution) is not required. A further advantage is that the sensitivity coefficients (and thus the partial derivatives of the measurement function) required by the GUF are not required. It is only necessary to evaluate the measurement function itself.

E5.4.6 Reporting the result

Table 1 shows the input quantities for the viscosity model, the PDFs (with R = rectangular or G = Gaussian) associated with each of them, and their estimates and associated standard uncertainties. Most input quantities are experimental, except for the end-effect correction and the linear coefficient of change of viscosity with temperature, for which estimates and associated uncertainties were obtained from references [446–448]. The sensitivity coefficients were also calculated and listed in the table and were used together with the standard uncertainties associated with each quantity to estimate the relative contribution of each one of them. The application of LPU produced the results given in the last row of the same table for the measurand.

GUM Methodology. Consider the analysis of a measurement of a Newtonian oil, at 20 °C, using a rheometer with a concentric cylinder system, according to DIN 53019 [446], at a shear rate value of $1\,s^{-1}$, using the input quantities indicated in table E5.4.1 and the measurement model (E5.4.1). The result is a viscosity of 1.091 Pa s with an associated expanded uncertainty (for 95 % confidence) of 0.018 Pa s. The partial derivatives required as sensitivity coefficients were formed analytically and their values checked by the complex step method.

The results in table E5.4.1 show that 31 % of the standard uncertainty associated with the calculated viscosity originates from the end effect correction factor, 30 % from torque, 20 % from temperature stability and 13 % from R_a. These numbers can change drastically according to the measurement conditions: any conclusions drawn are only valid for these specific conditions. At lower shear rates, for example, the measured torque will be smaller, but its absolute uncertainty is constant, which increases its share in the viscosity uncertainty. Conversely, at higher shear rates, because the temperature stability becomes questionable due to frictional heating, the viscosity uncertainty will increase considerably and be dominated by temperature uncertainty. Also, as it will be shown in this study, there are input quantities which when having their uncertainties increased may change dramatically the output PDF: the Monte Carlo method reveals such features.

This example refers only to Newtonian fluids, and the measurement model is unlikely to cover all input quantities but is a first approach to give an impression of the complexity associated with the treatment of uncertainty in rheology. In the more complex cases of non-Newtonian fluids, cases exist where an appropriate measurement model does not yet exist, uncertainty evaluation relying mostly on interlaboratory comparisons or other experimental means. Additionally, it demonstrates that uncertainty evaluation needs to be supported by real measurements.

Monte Carlo propagation of distributions. Using the same input values of table E5.4.1, the results given by the MCM are displayed in table E5.4.2 and the output PDF in the form of a scaled histogram is shown in figure E5.4.1. Although there is considerable similarity to a Gaussian PDF,

the GUM overstates the uncertainty by some 4 %, which is hinted at by the graph and is not surprising because of the model non-linearity and the key input quantities with rectangular PDFs. The GUM assumes the approximate applicability of the central limit theorem, that is, the PDF for the output quantity will be approximately Gaussian when the input quantities are independent and the output standard uncertainty is much larger than that of any single input quantity with a non-Gaussian PDF. In a case such as the current example, validation should be mandatory, even though the difference may be acceptable for many applications. It should be pointed out, nevertheless, that in decision-making situations this difference may prove significant. Moreover, the difficulty in applying GUM because of the above assumption and entailed by the determination of the sensitivity coefficients can be avoided by using the MCM.

Table E5.4.3: Comparison between GUM and Monte Carlo method with data from table E5.4.1 (expanded uncertainties are displayed with an additional decimal digit to highlight differences)

Quantity	Estimate	$u(y)$/Pa s	$U_{95\,\%}(y)$/Pa s
GUM	1.091	0.0091	0.0182
Monte Carlo	1.091	0.0088	0.0175

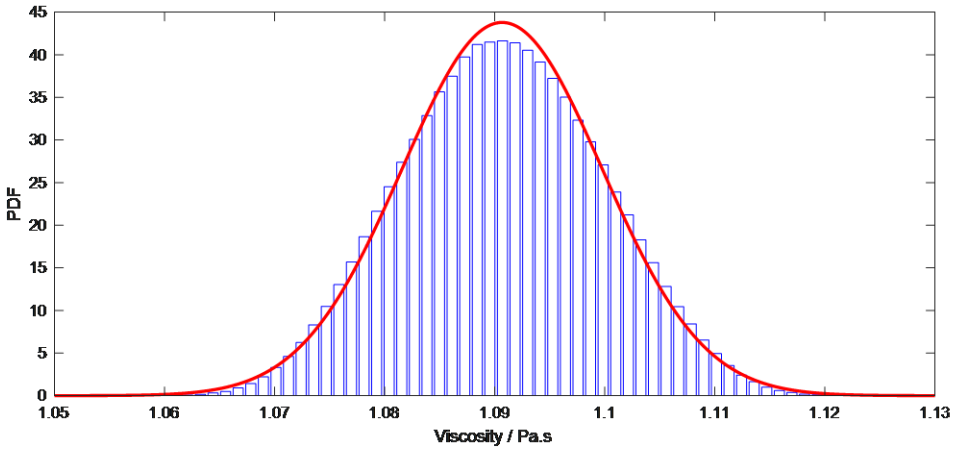


Figure E5.4.1: Output quantity PDF for the model (E5.4.2) using the GUM (continuous curve) and a Monte Carlo approach (scaled histogram).

For a very different set of conditions, a completely different result would have been obtained. Table E5.4.3 and figure E5.4.2 show the results for a hypothetical situation where the uncertainty associated with the input quantity R_i is increased by a factor of 10^3 (all other input quantities are unchanged). In this case, the use of the GUM approach would more visibly be totally inadequate for the evaluation of measurement uncertainty, in this instance overstating the expanded uncertainty by about 22 %. Similar results would be obtained if the standard deviations of the Gaussian distributions, particularly that for T_{ref} were reduced.

Table E5.4.4: Comparison between GUM and MC method with increased uncertainty associated with R_i

Quantity	Estimate	$u(y)$/Pa s	$U_{95\,\%}(y)$/Pa s
GUM	1.091	0.315	0.63
Monte Carlo	1.101	0.260	0.52

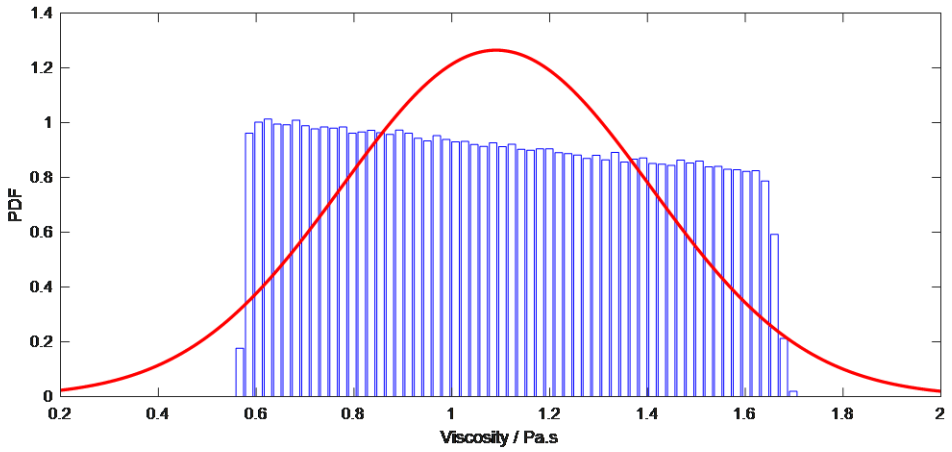


Figure E5.4.2: Output quantity PDF for the model (E5.4.2) using the GUM (continuous curve) and a Monte Carlo approach (scaled histogram), with a greatly increased uncertainty assigned to R_i.

E5.4.7 Interpretation of results

This study showed the complexity associated with measurement in the field of rheology. Even for the least complicated case, as is the case presented here for Newtonian liquids, there may be influencing quantities not reflected in the existing measurement models. However, for more complicated cases encountered for non-Newtonian liquids, sometimes there is no available "closed" measurement model. In such cases, proficiency testing data can be used to provide a realistic evaluation of measurement uncertainty [42] (sometimes called a "top-down" approach to measurement uncertainty evaluation as opposed to the "bottom-up" GUM approach). Proficiency tests and other forms of interlaboratory comparisons are a good external quality control measure of results in all cases.

The case studies presented have shown it is important to apply adequate statistical tools to evaluate measurement uncertainty. The advantages of MCM over GUF are apparent in this application, where the use of conventional recipes might bring erroneous results which, depending on the circumstances, can have serious implications, e.g., in the oil and mine industries or in health applications. Moreover, in the field of rheology there are decision-making situations, e.g., related to global trade in goods, especially involving liquids, where the accuracy and quality of results are of paramount importance [450].

Both the GUM and MCM approaches depend on knowledge of the PDFs for the input quantities, but whereas the GUF 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.

Example E5.5

Uncertainty calculation methodologies in microflow measurements: comparison of GUM, GUM-S1 and Bayesian approach

J.A. Sousa, E. Batista, S. Demeyer, N. Fischer, O. Pellegrino, A.S. Ribeiro, L.L. Martins, M. Reader-Harris, A.M.H. van der Veen, M.G. Cox

E5.5.1 Summary

The importance of measurement quality cannot be overemphasised in medical applications, as one is dealing with life issues and the well-being of society, from oncology to new-borns, and more recently to patients of the Covid-19 pandemic. In these dire situations the accuracy of fluid delivered according to a prescribed dose can be critical. Microflow applications are growing in importance for a wide variety of scientific fields, namely drug development and administration, Organ-on-a-Chip, or bioanalysis, but accurate and reliable measurements are a tough challenge for very small flow rates, from $1000\,\mu\mathrm{L\,h^{-1}}$ down to $1\,\mu\mathrm{L\,h^{-1}}$. Several sources of error have been established such as the mass measurement, the fluid evaporation dependent on the gravimetric methodology implemented and the repeatability, believed to be closely related to the operating mode of the stepper motor and drive screw pitch of a syringe pump. In addition, the difficulty in dealing with microflow applications extends to the evaluation of measurement uncertainty, which will quantify the quality of measurement. This difficulty is due to the conditions entailed when measuring very small values, close to zero, of a quantity such as flow rate, which is inherently positive. Alternative methods able to handle these features were developed and implemented, and their suitability will be discussed.

E5.5.2 Introduction of the application

The most used form of therapy in health care at hospitals is infusion therapy [451], which implies that drug delivery devices are very important instruments in this sector. Due to the widespread applications in critical health care, infusion errors are often made, with reported dramatic effects in different applications in the health sector. There are instances where adverse incidents,

morbidity and mortality, can be traced back to poor or inaccurate dosing [452] [453]. Important examples can be found in chemotherapy, oncology, anaesthesia, the operating theatre and nursery wards, especially for newborns. This situation is even more critical at very small flow rates, such as in neonatology, where small variations in the normal infusion procedure can lead to very large errors.

Furthermore, the unprecedented conditions that Public Health Institutions experience due to the COVID-19 pandemic crisis has forced hospital administrations to take measures beyond usual work practices, such as postponing maintenance and calibration deadlines in instruments with critical use, for example, to cope with shortage of equipment or the use of drug-delivery devices outside the patient room. These practices, if not performed under very controlled conditions, can lead to large dosing errors [454]. Therefore, any attempt to prevent adverse events by improving the knowledge of actual doses can make an enormous difference for the individual patient, especially new-born babies, and has a significant impact on the health sector. Measurement uncertainty plays a vital role in this situation since a reliable uncertainty evaluation will lead to an improved confidence in the actual prescribed and delivered dose.

In the EMPIR project "Metrology for Drug Delivery – MeDD II", metrological tools were developed and improved to assure the traceability of clinical data, allowing the metrological comparability of diagnostic and treatment information for flow rates from $30\,\mathrm{nL\,h^{-1}}$ to $60\,\mathrm{nL\,h^{-1}}$, targeting a relative expanded uncertainty of 1 % (95 % confidence level). The project delivered calibration methods to enable SI traceable flow rate measurements with high accuracy, which means that the uncertainty calculation should be properly implemented. This example interrelates the efforts in MeDD II and EMUE (Examples of measurement uncertainty evaluation), thus connecting the developed flow rate standard with best practices in the evaluation of measurement uncertainty.

The objective of this work is to compare uncertainty evaluation methodologies for very small flow rates, close to zero, during calibration. To exemplify the mathematical treatment of the uncertainty evaluation, experimental values arising from a syringe pump calibration will be used to assess the viability and advantage of using alternative methods to the traditional GUM approach (law of propagation of uncertainty), namely, Monte Carlo and Bayesian methods, to quantify measurement uncertainty when the conditions for the applicability of the GUM uncertainty framework [2] are not met. Such conditions relate to the extent of the non-linearity of the model, the number and magnitude of quantities with non-normal distributions and the ability to produce meaningful results for a specific application. Validation should always by attempted whenever the application of the GUM is questionable [3].

Here, due to the conditions entailed when measuring very small values, close to zero, of a quantity such as the flow rate, which is inherently positive, negative flow rate estimates may be made when inappropriate uncertainty evaluation methods are employed. This comparison work shows that, depending on the proportion of negative values obtained during calibration, alternative Bayesian uncertainty evaluation methods are preferred.

E5.5.3 Specification of the measurand

The measurand is volumetric flow rate Q in a syringe pump used under defined conditions.

E5.5.4 Measurement model

The measurement model associated with the syringe pump experiment can be described by [455]

$$Q = \frac{m_F - m_I}{t_F - t_I}\left[1 - \left(\frac{D_{tube}}{D_{tank}}\right)^2\right]\frac{1}{\rho_W - \rho_A}\left(1 - \frac{\rho_A}{\rho_B}\right)[1 - \gamma(T - 20)] + Q_{evap} + \delta Q_{rep}, \tag{E5.5.1}$$

relating the measurand (volumetric flow rate) Q to mass measurements (initial mass m_I and final mass m_F), densities of water ρ_W, air ρ_A and mass pieces ρ_B, evaporation rate Q_{evap}, water temperature T, time (initial time t_I and final time t_F), expansion coefficient γ, diameter D, and measurement repeatability δQ_{rep}.

The model (E5.5.1) is non-linear and thus may depart from the conditions prescribed for the application of the GUM. Moreover, relating to the remarks in section E5.5.2, this study considers the treatment of flow rates close to zero, as is typical of microflow applications, for which readings may result in negative values. This problem is one of measurement close to a physical limit, that can also be found in other areas, such as length (for example, roughness measurements) or chemistry (for example, measurement of the purity of a very pure substance), requiring to be properly handled to avoid erroneous results.

E5.5.5 Uncertainty propagation

E5.5.5.1 General

A typical uncertainty budget of a flow measurement is illustrated in table E5.5.1, resulting from a real experiment with a syringe pump calibrated by the gravimetric method at 2.77×10^{-7} mL s^{-1}. It relies on weighing the mass of the working fluid delivered by the instrument under test at a set time, which is then converted to volume at a reference temperature [456]. The volumetric flow rate Q is determined by the quotient of the volume of the reference liquid and the time interval, including the corrections indicated in equation (E5.5.1) [455]. The calibration procedure is conducted in a climate-controlled room insulated from vibration (reference temperature of 20 °C, above 50 % relative humidity). The syringe pump (that has a 1 mL glass removable syringe) is connected to stainless steel tubing of 0.16 cm (1/16″) diameter and the tube ending is immersed in the weighing vessel of the balance used, to prevent flow oscillations due to droplet detachment. A LabVIEW application is used for data acquisition, validation, online visualization of measured data and flow rate calculation [457]. Data acquisition starts approximately 10 min after steady flow is reached. Two tests were carried out, with 35 readings and 76 readings, for different sets of conditions. The test duration and number of acquisition points were increased to improve the knowledge of the experiment.

The experiments with the syringe pumps have highlighted two different patterns that may occur when dealing with measurements close to the physical limit of a quantity. Depending on the properties of the measuring instrument (for example, precision, accuracy), the conditions of the experiment (for example, temperature, operator, fluid), and the nominal flow rate targeted, the readings may show a proportion of negative values since in microflow experiments the range of values is very close to zero for the reasons given in section E5.5.4.

Thus, three different approaches will be applied using the data in table E5.5.1 for the measurement model (E5.5.1), one of which (the Bayesian approach) handles the situation where readings may be negative, and their suitability assessed. First a comparison will be made between the GUM

Table E5.5.1: Uncertainty budget for the syringe pump (case study 1); probability density function (PDF): N = normal, C = combined, R = rectangular. The unit of each element in the sensitivity coefficient column is that of the measurand, flow rate, divided by that of the corresponding input quantity

Quantity/unit	Best estimate x_i	PDF	Standard uncertainty $u(x_i)$	Sensitivity coefficient c_i	$u_i(Q) = \lvert c_i \rvert u(x_i)$
t_I/s	2.65×10^2	Normal	7.00×10^{-4}	1.15×10^{-10}	8.03×10^{-14}
m_I/g	3.799 544	Combined	2.86×10^{-5}	6.93×10^{-4}	1.97×10^{-8}
ρ_W/g mL^{-1}	0.997 615	Combined	6.26×10^{-4}	-1.67×10^{-7}	1.04×10^{-10}
ρ_A/g mL^{-1}	0.001 181	Rectangular	2.89×10^{-6}	1.46×10^{-7}	4.21×10^{-13}
ρ_B/g mL^{-1}	8.00	Normal	2.50×10^{-3}	3.07×10^{-12}	7.67×10^{-15}
t/°C	2.27×10^1	Combined	5.17×10^{-1}	-1.66×10^{-12}	8.60×10^{-13}
γ/°C^{-1}	1.00×10^{-5}	Rectangular	2.89×10^{-7}	-4.42×10^{-7}	1.28×10^{-13}
Q_{evap}/mL s^{-1}	1.04×10^{-7}	Rectangular	1.47×10^{-8}	1.00	1.47×10^{-8}
D_{tube}/cm	0.09	Normal	0.001	-1.62×10^{-8}	1.62×10^{-11}
D_{tank}/cm	1.36	Normal	0.001	1.07×10^{-9}	1.07×10^{-12}
δQ_{rep}/mL s^{-1}	2.13×10^{-7}	Normal	3.29×10^{-8}	1.00	3.29×10^{-8}
m_F/g	3.799 785	Combined	2.86×10^{-5}	-6.93×10^{-4}	1.97×10^{-8}
t_F/s	1.71×10^3	Normal	7.00×10^{-4}	-1.15×10^{-10}	8.03×10^{-14}
Flow rate Q/mL s^{-1}	2.70×10^{-7}			Combined standard uncertainty $u(Q)$	4.56×10^{-8}

(GUM uncertainty framework (GUF)) and GUM-S1 (propagation of distributions) when "normal" data are available, to assess whether the departure from ideal conditions for the application of the GUF will be a problem in this case. This is the situation when most data are positive (as a flow rate is expected to be). The second comparison will be drawn between these methods for uncertainty propagation and a Bayesian approach for the remaining case when the number of negative readings is significant, to assess which approach handles best this situation.

Both the GUF and MCM approaches depend on knowledge of the probability density functions (PDFs) for the input quantities, but whereas the GUF 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 MCM approach, making the latter generally more reliable and which should be used for validation purposes when the conditions for the use of the GUM approach, as given in [3] clause 5.8, are not fully met.

E5.5.5.2 GUM uncertainty framework

The GUF is based on the application of the law of propagation of uncertainty (LPU) [2, clause 5] and assuming a normal or Student t distribution for the PDF of the measurand [2,3]. Application of the LPU requires the calculation of sensitivity coefficients c_i, the first partial derivatives of the volumetric flow rate Q [the measurement function specified as the right-hand side of expression (E5.5.1)] with respect to the input quantities, evaluated at the estimates of those quantities. For full details of the GUF, see [2].

E5.5.5.3 Propagation of distributions

An MCM is an implementation of the propagation of distributions and is generally considered a very reliable approach for uncertainty propagation. Appropriate PDFs assigned to the input quantities are propagated through the measurement model to provide the PDF for the output quantity, with no restriction in relation to shape or symmetry of the input PDFs. The expectation (mean) of the resulting PDF is used as the best estimate of the quantity and the standard deviation of the PDF is used as the standard uncertainty associated with that estimate. When the applicability of the GUF is questionable, appropriate use of Monte Carlo methods will provide a PDF for Y that is consistent with the input PDFs. Full details of the method are given in [3].

E5.5.5.4 Bayesian evaluation

A Bayesian evaluation [65] determines a posterior distribution for the measurand (and any additional parameters) given observations and prior information on the measurand (and the additional parameters, if any). For linear measurement models and so-called non-informative prior distributions, the posterior distribution coincides with the GUM-S1 PDF. Here, the Bayesian approach is used to impose a positivity constraint on the measurand in the presence of a proportion of negative flow rates readings obtained during calibration (for reasons mentioned earlier).

According to table E5.5.1, we use the following "error-in-variables" formulation to represent the random variables $M_\mathrm{F} = m_\mathrm{F} + \zeta_{\mathrm{M_F}}$, $M_\mathrm{I} = m_\mathrm{I} + \zeta_{\mathrm{M_I}}$, $T_\mathrm{F} = t_\mathrm{F} + \zeta_{\mathrm{T_F}}$ and $T_\mathrm{I} = t_\mathrm{I} + \zeta_{\mathrm{T_I}}$, where ($m_\mathrm{F}$, m_I, t_F, t_I) are the measurements of the final mass, the initial mass, the final time and the initial time, respectively (given in the column "Best estimate" of table E5.5.1) and ($\zeta_{\mathrm{M_F}}$, $\zeta_{\mathrm{M_I}}$, $\zeta_{\mathrm{T_F}}$, $\zeta_{\mathrm{T_I}}$) are centered, normally distributed error terms.

Let q_obs denote the observed flowrate computed from measurements ($\zeta_{\mathrm{M_F}}$, $\zeta_{\mathrm{M_I}}$, $\zeta_{\mathrm{T_F}}$, $\zeta_{\mathrm{T_I}}$):

$$q_\mathrm{obs} = \frac{m_\mathrm{F} - m_\mathrm{I}}{t_\mathrm{F} - t_\mathrm{I}}$$

and Q_obs the associated random variable defined by

$$Q_\mathrm{obs} = \frac{M_\mathrm{F} - M_\mathrm{I}}{T_\mathrm{F} - T_\mathrm{I}} \sim N(q_\mathrm{obs}, u^2_{\mathrm{q_{obs}}}).$$

We show with Monte Carlo sampling from the distributions for (M_F, M_I, T_F, T_I) that Q_obs is normally distributed, with parameter values displayed in the Appendix for case study 1 and case study 2. The corresponding error-in-variables formulation writes

$$Q_\mathrm{obs} = q_\mathrm{obs} + \eta \text{ with } \eta \sim \mathrm{N}(0, u^2_{\mathrm{q_{obs}}}).$$

Replacing Q_obs in the measurement model gives

$$Q = Q_\mathrm{obs}\left[1 - \left(\frac{D_\mathrm{tube}}{D_\mathrm{tank}}\right)^2\right]\frac{1}{\rho_\mathrm{W} - \rho_\mathrm{A}}\left(1 - \frac{\rho_\mathrm{A}}{\rho_\mathrm{B}}\right)[1 - \gamma(T - 20)] + Q_\mathrm{evap}. \tag{E5.5.2}$$

The statistical model explaining the measurement Q_obs is obtained by introducing a repeatability term $\delta Q_\mathrm{rep} \sim \mathrm{N}(0, s^2/n)$ defined in section 3:

$$Q_\mathrm{obs} = \frac{Q - Q_\mathrm{evap}}{C(\theta)} + \delta Q_\mathrm{rep},$$

where

$$C(\theta) = \left[1 - \left(\frac{D_{\text{tube}}}{D_{\text{tank}}}\right)^2\right] \frac{1}{\rho_{\text{W}} - \rho_{\text{A}}} \left(1 - \frac{\rho_{\text{A}}}{\rho_{\text{B}}}\right) [1 - \gamma(T - 20)]$$

and

$$\theta = (D_{\text{tube}}, D_{\text{tank}}, \rho_{\text{W}}, \rho_{\text{A}}, \rho_{\text{B}}, \gamma, t).$$

By replacing the left-hand term, we obtain

$$q_{\text{obs}} + \eta = \frac{Q - Q_{\text{evap}}}{C(\theta)} + \delta Q_{\text{rep}}.$$

The statistical model associated with the measurement model is

$$q_{\text{obs}} = \frac{Q - Q_{\text{evap}}}{C(\theta)} + \zeta + \delta Q_{\text{rep}} \text{ (with } \zeta = -\eta), \qquad \zeta \sim N(0, u^2_{q_{\text{obs}}}), \qquad \delta Q_{\text{rep}} \sim N(0, s^2/n).$$

The quantity of interest is $\pi(Q|q_{\text{obs}})$ obtained from the joint posterior distribution

$$\pi(Q, \theta|q_{\text{obs}}) \propto l(q_{\text{obs}}|Q, \theta)\pi(Q)\pi(\theta),$$

where $l(q_{\text{obs}}|Q, \theta)$ is the likelihood of the data, $\pi(Q)$ is the prior distribution for the measurand and $\pi(\theta)$ is the joint distribution of the vector θ given by the PDFs marked Rectangular (and Combined) associated with the uncertainty sources in table E5.5.1.

To impose the non-negativity of the measurand, the prior distribution should have a support excluding negative values such as $\pi(Q) \sim \text{R}(0, \infty)$, where $\text{R}(a, b)$ denotes the rectangular PDF with limits a and b.

Since the posterior distribution has no closed form, Markov Chain Monte Carlo methods are employed to sample from the posterior distribution [458–460]. These methods construct a sequence of dependent values that form a Markov chain with stationary distribution equal to the sought distribution. The Metropolis-Hastings algorithm constitutes a popular class of MCMC methods as it only requires knowing the sensitivity coefficients and the uncertainties (and covariances, as appropriate) of the input quantities to sample from the posterior distribution. The sequence of values is usually considered only after a first period of "burn-in", and often the chains are thinned (that is, only each 10th value is used, say) to reduce the correlation between successive values. Various criteria have been suggested for assessing the convergence of MCMC methods.

We refer to [459] for a general introduction to these methods and to [460] for an introductory example in metrology. In this work, results are obtained with Python 3 using probabilistic programing with PyMC3 [461].

E5.5.6 Reporting the result

A typical set of uncertainty contributions, used for case study 1, is illustrated in Table E5.5.1 from which the best estimate and standard uncertainties can be used to evaluate the measurement uncertainty. A second set of data, for smaller values of Q was also used. In here, as for most situations, repeatability is treated as having a centred normal distribution (with mean 0), assuming that data can be well represented by the corresponding mean value and standard deviation [3]. The associated uncertainty is taken as the standard deviation of the mean. The model is mildly non-linear and there is no non-normal dominant source of uncertainty.

Experimental data from two case studies

The experiments with the syringe pumps have highlighted two different patterns that may occur when dealing with measurements close to the physical limit of a quantity. Depending on the measuring instrument (e.g., precision, accuracy), the conditions of the experiment (e.g., temperature, operator, fluid), and the nominal flow rate targeted, the readings may show a proportion of negative values since in microflow experiments the range of values is very close to zero.

Data from the first case study is displayed in figure E5.5.1, where most readings are of very small flow rates, as expected, but only a very small proportion of them are negative. The first 7 readings were ignored since the stabilization time mentioned above was only partially observed. The example of figure E5.5.1, though, represents a good response of a syringe pump with relatively small repeatability.

However, in the second case study illustrated in figure E5.5.2, there is a clear increase in the proportion of negative values, and a higher dispersion of values, which indicates a poorer response of the syringe pump used and thus should entail a different type of problem in the evaluation of measurement uncertainty, as it will be seen in section 6. It is important to point out that the negative values of flow rate are the result of working very close to the physical limit of a system and as a consequence the intrinsic noise and lack of precision will, in some instances, provide negative values for quantities that are inherently positive.

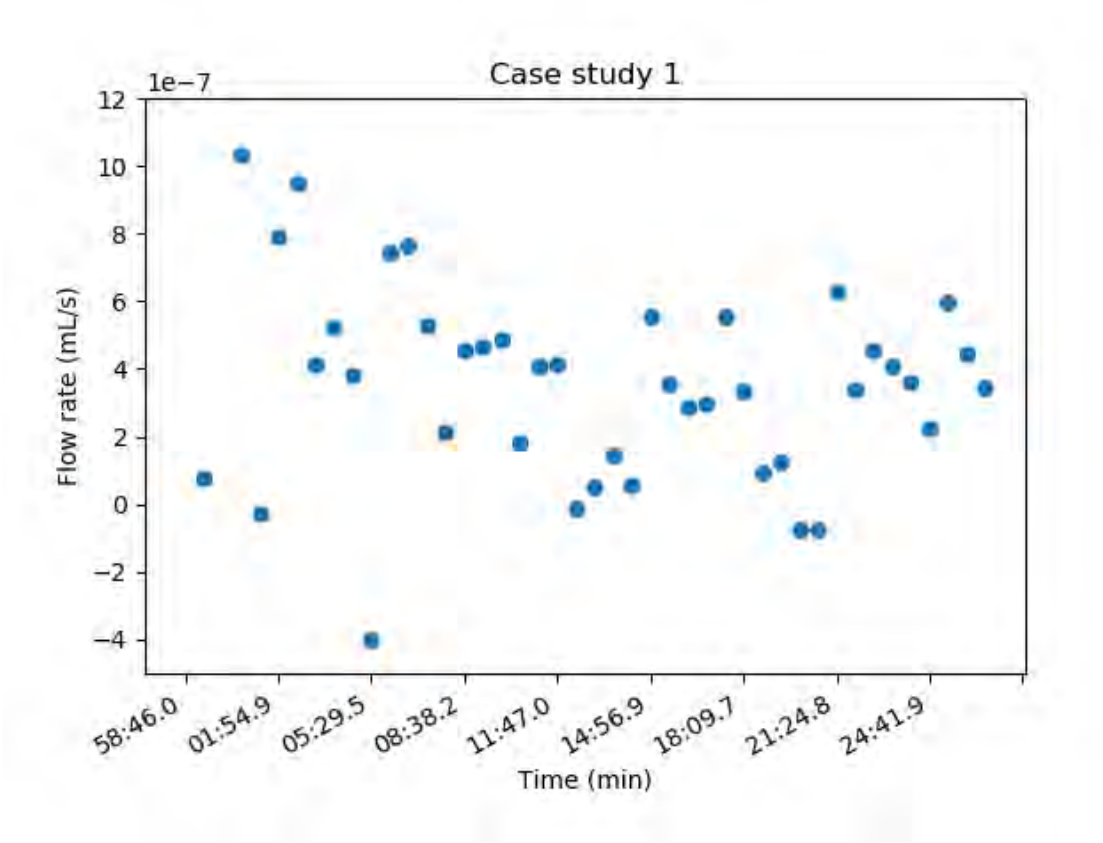


Figure E5.5.1: Flow rate for case study 1 when most readings are positive

For case study 1, the GUM is expected to perform well, whereas for case study 2, the conditions for the applicability of the GUM uncertainty framework are not met since part of the data is not meaningful. For each case study, the GUM, GUM-S1 (Monte Carlo) and Bayesian approach will

be compared to ascertain which of them can better handle the presence of a significant number of negative values in the readings and, as a consequence provide a more reliable estimate of the flow rate and associated standard uncertainty.

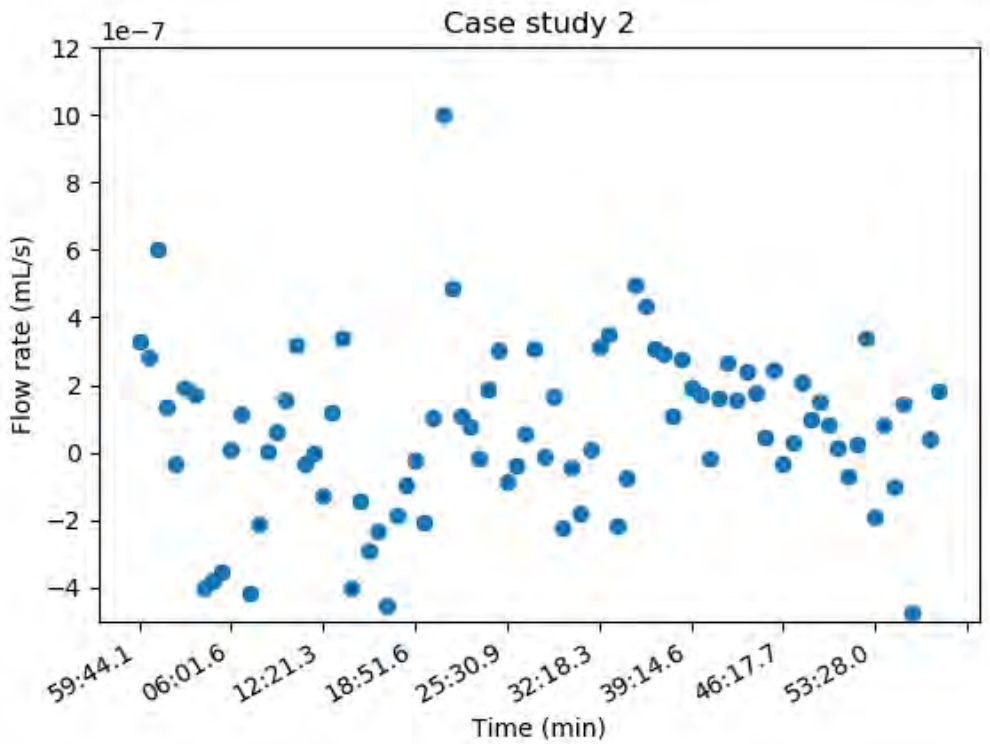


Figure E5.5.2: Flow rate readings for case study 2 with a significant number of negative readings

Case study 1

A comparison between the GUM and the Monte Carlo method from GUM-S1 is illustrated in figure E5.5.3 and results are provided in Table E5.5.2. For most purposes it would be reasonable to conclude that the GUM conventional approach is acceptable, which was expected since there is only a mild non linearity in the model (E5.5.1), and there is not a dominant normal source of uncertainty in Table E5.5.2 (in conformity assessment, these differences should be taken into account).

Table E5.5.2: GUM, Monte Carlo and Bayesian flow rate results (case study 1 – Gravimetric)

Approach	Best estimate $Q/\mathrm{mL\,s^{-1}}$	Standard uncertainty $u(Q)/\mathrm{mL\,s^{-1}}$	95 % coverage interval
GUM	2.6975×10^{-7}	5.0924×10^{-8}	1.6791×10^{-7} to 3.7160×10^{-7}
Monte Carlo	2.6972×10^{-7}	5.0906×10^{-8}	1.6898×10^{-7} to 3.6815×10^{-7}
Bayes ($Q > 0$)	2.6996×10^{-7}	5.1009×10^{-8}	1.7180×10^{-7} to 3.6970×10^{-7}

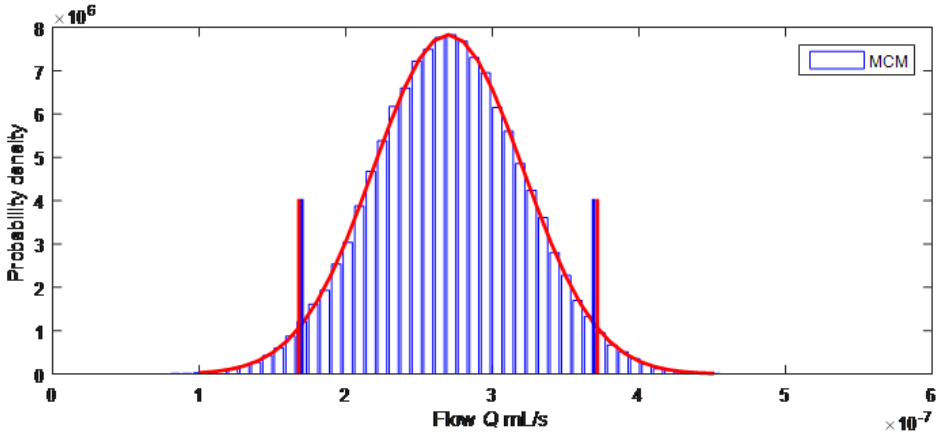


Figure E5.5.3: GUM and Monte Carlo for case study 1

It is important to underline that the conclusions drawn are only valid for the specific set of conditions defined in Table E5.5.2. Just for exemplifying purposes, in the case study 1, increasing by a factor of 5 the uncertainty associated with the evaporation quantity will change the output considerably and the suitability of the GUM may now be questionable, as illustrated in figure E5.5.4.

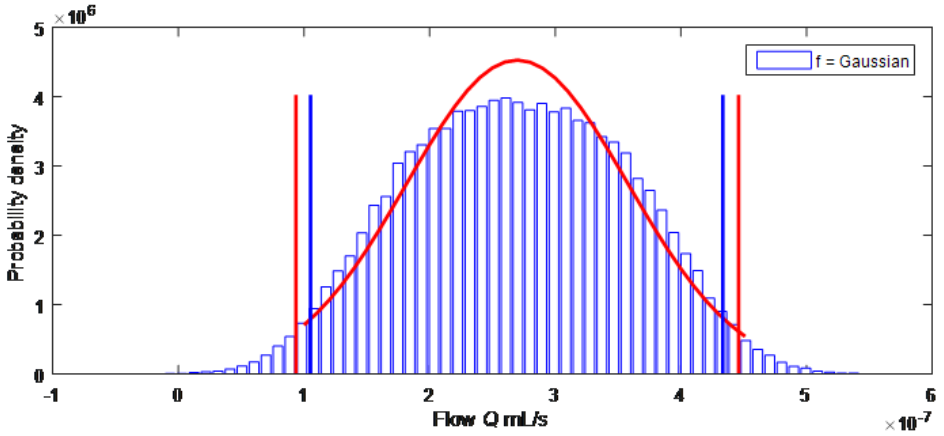


Figure E5.5.4: GUM and Monte Carlo comparison for an increased percentage of the rectangular distribution associated with Q_{evap}

Results obtained with the Bayesian method under the hypothesis that Q is positive are displayed in Table E5.5.3. The Bayesian approach provides results similar to those obtained by the other two methods.

Figure E5.5.5 groups together the GUM, the GUM-S1 and the Bayesian method, for the baseline conditions. It is apparent that for the conditions of this case study all different approaches provide similar results and therefore the GUF can be safely used for nominal flows of $2.7 \times 10^{-7}\,\text{mL}\,\text{s}^{-1}$ in these conditions.

Case study 2

A different situation arises when faced with data as depicted in figure E5.5.2. In this case, the number of negative values is significant, which means that negative values are expected to be found in the 95 % coverage interval as computed using the GUM uncertainty framework, which

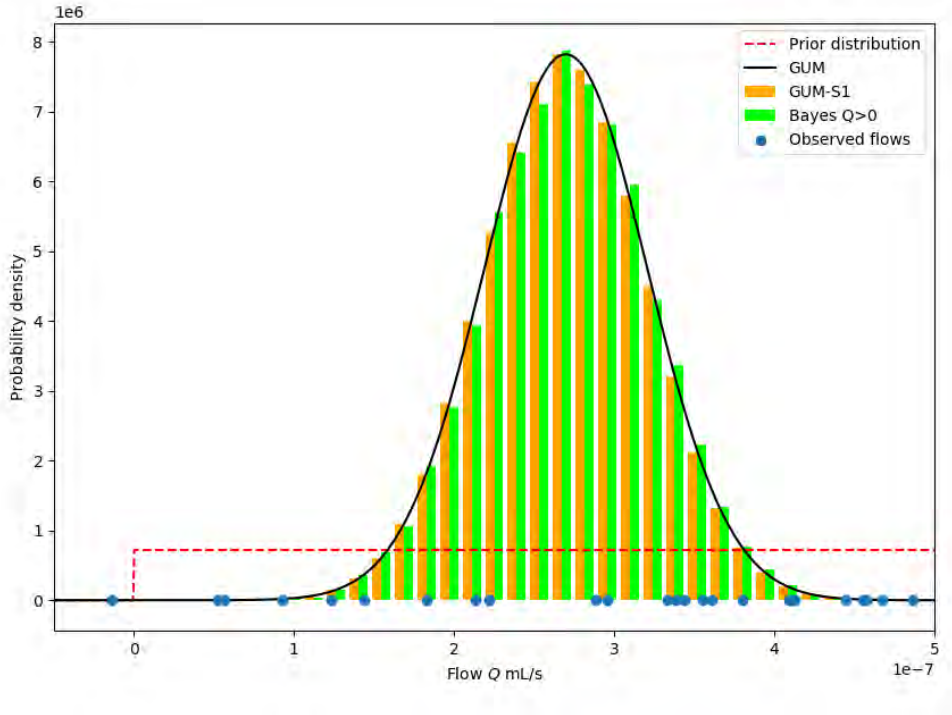


Figure E5.5.5: Measurand PDF for all approaches applied to case study 1

is a problem. The Monte Carlo method could deal with this situation by replacing negative values by zero, but this would distort the PDF for the measurand. Therefore, neither the GUM nor the MC method will be adequate to describe the latter situation, in terms of naturally representing the probability density function for the measurand. The Bayesian method, however, rather concerns the probability density of the measurand, considering the data produced by the measurement and possible prior information. Indeed, this method should handle well with this situation, where the distribution of values is truncated at zero (prior knowledge), and thus redistributes the values over the rest of the integral domain.

A comparison between the GUM, the Monte Carlo method from GUM-S1 and the Bayesian approach under the hypothesis that Q is positive is given in Table E5.5.3 and illustrated in figure E5.5.6.

Table E5.5.3: GUM, Monte Carlo and Bayesian flow rate results (case study 2)

Approach	Best estimate $Q/\mathrm{mL\,s^{-1}}$	Standard uncertainty $u(Q)/\mathrm{mL\,s^{-1}}$	95 % coverage interval
GUM	2.5325×10^{-8}	3.6951×10^{-8}	-4.8577×10^{-8} to 9.9227×10^{-8}
Monte Carlo	2.5392×10^{-8}	3.6937×10^{-8}	-4.6647×10^{-8} to 9.7517×10^{-8}
Bayes ($Q > 0$)	4.1893×10^{-8}	2.7728×10^{-8}	1.1994×10^{-12} to 9.3596×10^{-8}

The result of the simulation using the Bayesian method with the positivity constraint indicated above proves the advantage of the Bayesian approach in limit-of-detection problems. The Bayesian posterior, as expected, is very much like a truncated normal.

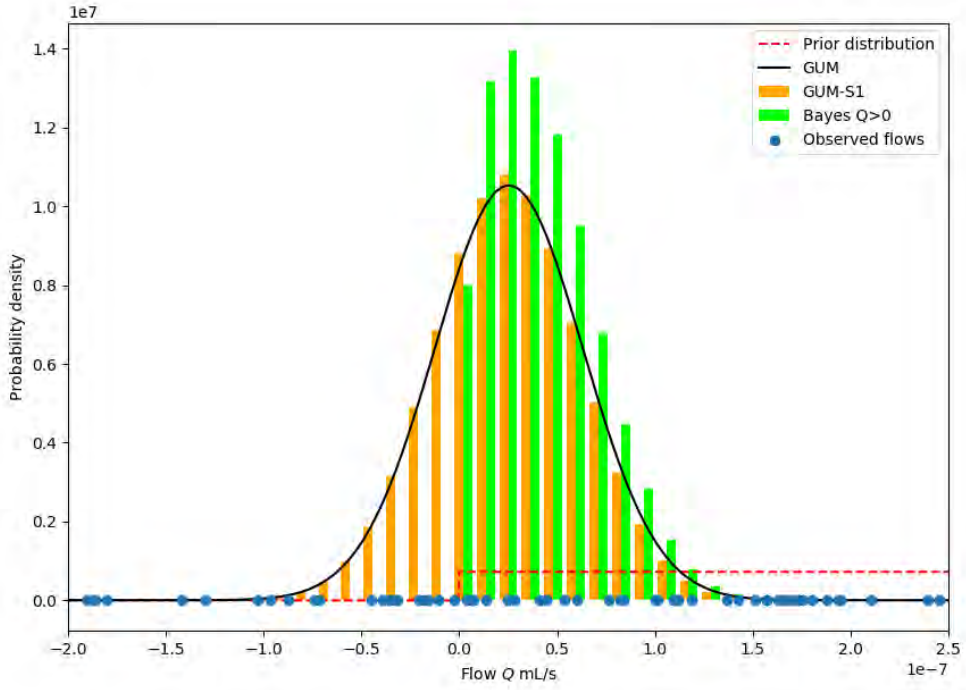


Figure E5.5.6: Measurand PDF for all approaches applied to case study 2

As hinted before, the MCM does not handle particularly well problems "close-to-the-physical-limit" as are micro flow rates or nano volumes which are widely used in health applications. It is therefore important to characterize errors and be able to qualify all measurements in this area through the correct evaluation of the measurement uncertainty.

E5.5.7 Interpretation of results

The aim of this work is to provide guidance on method selection with respect to the evaluation of measurement uncertainty. Clearly the choice of method depends on the problem under evaluation and no blind recipe should be used. Depending on the conditions of the problem, the GUM uncertainty evaluation may prove to be adequate, whereas in other circumstances alternative approaches should be applied instead, e.g., the MCM or the Bayesian method. What this study indicates is that the MCM approach is a convenient alternative to the GUM uncertainty evaluation approach for cases where the number of negative values is not significant, with the advantage of often having a simpler application, despite requiring a software implementation.

However, this robust alternative method may also prove inadequate in other circumstances, e.g., in cases too close to the physical limit of a system when out of bound values are significant. In the latter class of problems, the constraints imposed to the measurand by the use of a prior fits well to the Bayesian approach and is clearly a better method to evaluate measurement uncertainty in a number of similar problems where that evaluation is not a trivial matter.

For the budget in table E5.5.1, only four of the input quantities make meaningful contributions to the combined standard uncertainty. Those quantities are the initial time t_{I}, final time t_{F}, evaporation rate Q_{evap} and measurement repeatability δQ_{rep}. All other contributions are two or

more orders of magnitude smaller. A conclusion that can be drawn is that for data sets that are not too dissimilar from that considered here, the measurement model (E5.5.1) can be replaced by the much simpler model

$$Q = K(m_{\mathrm{F}} - m_{\mathrm{I}}) + Q_{\mathrm{evap}} + \delta Q_{\mathrm{rep}}, \tag{E5.5.3}$$

where, for purposes of uncertainty propagation,

$$K = \frac{1}{t_{\mathrm{F}} - t_{\mathrm{I}}}\left[1 - \left(\frac{D_{\mathrm{tube}}}{D_{\mathrm{tank}}}\right)^2\right]\frac{1}{\rho_{\mathrm{W}} - \rho_{\mathrm{A}}}\left(1 - \frac{\rho_{\mathrm{A}}}{\rho_{\mathrm{B}}}\right)[1 - \gamma(T - 20)],$$

regarded as a constant evaluated at the estimates of the input quantities. Since the measurement model (E5.5.3) is linear in its input quantities, LPU applied to it will give acceptable results in situations, as explained in section E5.5.6, when the bulk of the flow rate readings are positive.

Example E5.6

The role and use of measurement uncertainty in addressing specification requirements: medical temperature examples

J. Greenwood, M.G. Cox

E5.6.1 Summary

Measurement guides and standards are an essential source of information in most areas of testing. However the terminology used to describe technical requirements in terms of measurement results is often inadequate in situations where conformity decisions must be made that are consistent with the Guide to the expression of Uncertainty in Measurement (GUM), or with the requirements of standards used to support accreditation such as ISO/IEC 17025 and ISO 15189. This example illustrates the issue in terms of two highly regarded guidance documents from the healthcare sector.

E5.6.2 Introduction

Guidance documents, requirements and standards are found in all areas of testing and measurement. They are written by experts in the technical field with the aim of ensuring consistent and comparable results in the particular activity for which specifications are provided. Unfortunately, it is often not possible to establish whether these requirements are being met in a metrologically robust fashion as, for example, is required for accreditation to ISO/IEC 17025 [7] or ISO 15189 [39].

In the worst cases this difficulty occurs because there is no requirement for the evaluation (or even awareness) of measurement uncertainty and requirements are expressed in qualitative terms, such as 'accuracy'. This practice creates an immediate problem since the GUM [2] and standards such as ISO/IEC 17025 and ISO 15189 are generally concerned with uncertainty, not accuracy. These are fundamentally different concepts and, as is stated in the International Vocabulary of Metrology (VIM) [89] (in annotations to the html version [462]), there is no established

methodology for assigning a numerical value to measurement accuracy. Often the only approach that can be taken is to agree an ad hoc interpretation of the requirement, or to supplement the stated requirements with *additional* requirements such as limits in terms of measurement uncertainty [194].

Here, two examples are developed in order to illustrate how such guidance and standards requirements might be interpreted. These have not been selected as 'bad' examples; rather, they are chosen because they show that even in otherwise authoritative and carefully drafted guidance there can still be issues of interpretation left to resolve. The aim is to support drafting of future revisions and also to indicate options when working with the existing versions.

The first example concerns the decontamination of medical devices by steam sterilization. In this case, allowable limits on measurement uncertainty are defined. The only issue is to establish or formalise a clear and unambiguous rule to explain the role that measurement uncertainty must take in the decision process. The second example emerges during the mapping and monitoring of cold storage systems for blood products, where temperature is measured in order to decide whether an alarm condition is met. In this example, what at first may seem to be a clear 'accuracy' requirement is shown to be incomplete, requiring further information or assumptions before conformity can be decided.

The discussion and approach in both examples are equally applicable to other guidance and standards documents and to measurements of other quantities.

The document begins with a recapitulation of what 'measurement uncertainty' actually represents and how it relates to conformity decisions.

E5.6.3 Background: recapitulation on measurement uncertainty and its role in making conformity decisions

It is important to recognise that in most physical measurement scenarios the measured value y is only an estimate of the true value of the quantity of interest Y (the measurand). The measurement result comprises the estimated value and the associated standard measurement uncertainty u. In the LPU approach to uncertainty evaluation the standard measurement uncertainty corresponds to the standard deviation of a probability density function (PDF), which in many cases is approximately Gaussian. This PDF can be interpreted in terms of the relative likelihood of possible true values (of the measurand), given the measurement result.

Specifications generally relate to requirements concerning the measurand, for example 'the sample temperature Y must not exceed an upper limit S_U'; rather than to measured values for which separate 'acceptance criteria' need to be established. Therefore, any test of a specification based upon a measurement result is by its very nature a test in which probability must play a role. Consequently, when a conformity decision is made that is based upon the location of the measured value relative to a specification limit there is a finite risk of making a false decision. See figure E5.6.1.

In figure E5.6.1 the shaded region of the PDF corresponds to non-conforming values that, given the observed measured value, could reasonably be attributed to the measurand. Assuming that conformity has been accepted, it represents the probability of false acceptance [6], i.e., the risk that an incorrect decision has been made. Conversely, the unshaded region represents the conformance probability, i.e., the probability that given the measured value, the true value of the measurand lies within the specification interval.

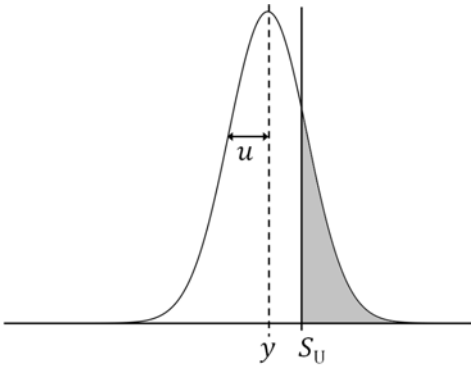


Figure E5.6.1: Measured value y and standard uncertainty u in relation to an upper specification limit S_U

The probabilities can be calculated from knowledge of the PDF. For those PDFs with a Gaussian or t distribution the probability represented by the various regions under the PDF can be readily calculated, e.g., using Excel cell functions [194]. For other PDFs, a numerical approach such as a Monte Carlo method (MCM) can be used [3].

The risk of making an incorrect decision can be minimised by restricting the range of measured values that will be accepted, for example by establishing an acceptance limit in terms of the specification limit and some guard band w [6]. For example, an acceptance limit A_U could be set to

$$A_U = S_U - w.$$

w is often chosen to be some multiple of the standard uncertainty u that defines the worst-case risk, which occurs when the value of y equals the acceptance limit (e.g., $y = A_U$). See Figure E5.6.2.

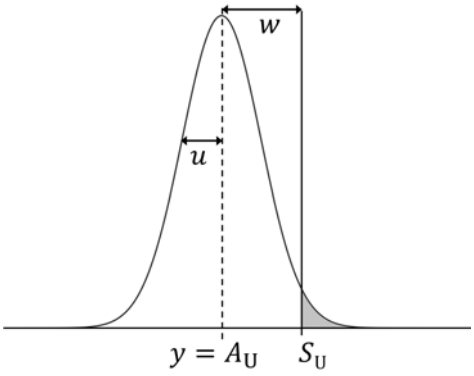


Figure E5.6.2: Measured value y and standard uncertainty u in relation to an upper specification limit S_U, a guard band w and upper acceptance limit A_U

A_U is the largest measured value to be accepted as indicating conformity.

E5.6.4 Example 1: Sterilisation temperature

As might be expected for such an important area of activity, there are numerous guidance documents relating to best or required practices in health services. For example, in the United Kingdom the requirements for decontamination of medical devices in a central decontamination unit

are set out in various National Health Service documents such as SHTM 01-01 Part C [463] (covering Scotland) and broadly equivalent documents covering England and Wales. SHTM 01-01 Part C2.12 defines the required time–temperature relationships for sterilization by steam. In this case, measurement uncertainty requirements are also stated within a related document SHTM 01-01 Part B [464] (clause 2.33 states that 'the uncertainty of measurement . . . should be no more than ±0.5 °C', which is interpreted as describing a 95 % coverage interval that in this case, for a Gaussian PDF, corresponds to a standard uncertainty of 0.25 °C).

In this example, a basic model for measurement uncertainty is presented. This is followed by a consideration of how a correctly evaluated measurement uncertainty might be used to test whether a measured value indicates conformity with a specification.

As has previously been stated, this example is not chosen as illustrating particularly 'poor' practice (it is actually one of the better guidance documents found); nevertheless it can be used to illustrate the difficulties around 'inadequately' defined requirements. The following points have general applicability: the intention here is to indicate how current requirement might be interpreted and how future requirements might be written in such a manner that 'interpretation' is not required.

E5.6.4.1 Measurand

The measurand, i.e., the quantity of interest, is the temperature achieved at the location of the item being sterilized.

E5.6.4.2 Model

The best estimate of the measurand is the measured temperature T, which is related to the indicated probe temperature T_o by a calibration correction and various metrological effects. A typical measurement model might involve an additive calibration correction. For example,

$$T = T_o + \Delta T_o + \delta T_D + \delta T_x + \delta T,$$

where

T_o is the indicated probe temperature,

ΔT_o is the calibration correction corresponding to indication T_o,

δT_D is the error due to drift of the probe behaviour since it was calibrated, i.e., drift in ΔT_o,

δT_x is the error due to installation and contact effects,

δT is the error due to errors of precision (such as measurement repeatability).

It is assumed that the terms in the measurement model are independent of each other, i.e., there is no correlation to include in the uncertainty evaluation. The δ 'error' terms represent unknown or unknowable quantities for which the best estimate of the value is zero; however, they have non-zero uncertainty.

E5.6.4.3 Uncertainty propagation

The LPU approach gives the standard uncertainty u_{T} associated with T:

$$u_{\mathrm{T}}^2 = u^2(T) = u^2(T_o) + u^2(\Delta T_o) + u^2(\delta T_D) + u^2(\delta T_x) + u^2(\delta T),$$

where, in this example,

$u(T_o)$	is associated with the finite resolution of the indication,
$u(\Delta T_o)$	is taken from the calibration certificate for the probe,
$u(\delta T_D)$	is evaluated from historical data,
$u(\delta T_x)$	is evaluated on the basis of physical data,
$u(\delta T)$	is the standard deviation for repeated measurements of T.

No covariance term is required as there are no correlated quantities in the model.

E5.6.4.4 Reporting measurement results

The measurement result can be reported in the conventional 'coverage interval' format

$$T \pm U,$$

where T here denotes the measured value. Assuming that the underlying PDF is Gaussian, $U = 2u_{\mathrm{T}}$ is the expanded uncertainty providing a coverage probability of approximately 95 %.

E5.6.4.5 Interpretation

The SHTM-C guidelines [464, clause 2.33] define an upper allowable limit of *measurement uncertainty* (although the equivalent earlier-dated English and Welsh versions do not yet do so). Although SHTM-C does not provide an explicit decision rule or mention a guard band, it can perhaps reasonably be assumed that the intended rule has acceptance limits set by so-called 'Simple Acceptance' criteria, i.e., $A_{\mathrm{L}} = S_{\mathrm{L}}$ and $A_{\mathrm{U}} = S_{\mathrm{U}}$ with a decision rule of the form (where t denotes the minimum duration of the test in minutes as stipulated in [464]):

> **PASS**: if $A_{\mathrm{L}} \leq T \leq A_{\mathrm{U}}$ for a duration exceeding t minutes AND uncertainty limits apply as detailed in SMTM 01-01 Part B, where T is the measured value, and A_{L} and A_{U} are lower and upper acceptance limits.
>
> **FAIL**: Otherwise.

For a PASS the maximum probability of false acceptance with Simple Acceptance criteria is (usually) 50 %, corresponding to $T = S_{\mathrm{L}}$ or $T = S_{\mathrm{U}}$. This risk reduces for values within the acceptance interval to a minimum at the central value. These 'risk-consequences' of Simple Acceptance criteria are *undefinable* if the measurement uncertainty (or at least some limit upon its value) is not incorporated in the decision process.

Note that in situations where a 'significant' portion of the PDF overlaps *both* limits of a double-sided specification, the risk will be higher than 50 %. The concept of equal 'shared risk' must therefore be used with caution and *cannot* be used without some consideration of measurement uncertainty).

E5.6.4.6 Alternative interpretation

Suppose instead that the intention of the guide writers had been that the worst case probability of false acceptance should be no more than 2.5 % (rather than 50 %), i.e., conformance probability should be at least 97.5 %. This intention could be achieved by defining a 'stringent' guard band w with corresponding acceptance limits defined by

$$A_{\mathrm{L}} = S_{\mathrm{L}} + w,$$
$$A_{\mathrm{U}} = S_{\mathrm{U}} - w.$$

For a Gaussian PDF, in those situations where a 'significant' portion of the PDF can only overlap one or other of the limits, i.e., when the uncertainty is small compared with the acceptance interval, the limit of 2.5 % is achieved with $w = 2u_{\mathrm{T}}$.

(Note: the figure of 2.5 % corresponds (approximately) to the area in *each* tail of a Gaussian PDF, where a 'tail' is located beyond a limit two standard deviations from the central value, corresponding to $2u_{\mathrm{T}}$ in this case.)

For example, suppose that $S_{\mathrm{L}} = 121\,^{\circ}\mathrm{C}$, $S_{\mathrm{U}} = 124\,^{\circ}\mathrm{C}$ and the standard uncertainty associated with T is $u_{\mathrm{T}} = 0.25\,^{\circ}\mathrm{C}$. These values give acceptance limits

$$A_{\mathrm{L}} = S_{\mathrm{L}} + 0.5\,^{\circ}\mathrm{C} = 121.5\,^{\circ}\mathrm{C},$$
$$A_{\mathrm{U}} = S_{\mathrm{U}} - 0.5\,^{\circ}\mathrm{C} = 123.5\,^{\circ}\mathrm{C}.$$

The associated decision rule could be written as (where t denotes the minimum duration of the test in minutes):

> **PASS**: when the measured temperature is within the acceptance interval (at or) between the acceptance limits A_{L} and A_{U} for a duration exceeding t minutes. The worst-case risk of a false pass is less than about 2.5 %, which occurs for a measured value at an acceptance limit, and decreases rapidly for conforming values away from the acceptance limits.
>
> **FAIL**: Otherwise.

Note that SHTM-C requires that expanded (95 %) uncertainty is not allowed to exceed $U_{\mathrm{max}} = 0.5\,^{\circ}\mathrm{C}$. This uncertainty defines the maximum guard band (i.e., narrowest acceptance interval) consistent with the worst case risk limit. If the prevailing uncertainty is smaller than U_{max} then, for the same maximum specific risk, a narrower guard band and consequently wider acceptance interval can be established.

E5.6.5 Example 2: Storage of blood products

This second example is presented in terms of another highly respected guidance document, namely, the 'PHSS Guidance Document for Cold Storage Temperature Monitoring and Mapping for Blood Products' [465].

As was the case for the previous example, this example is representative of many, otherwise authoritative, guides and standards in which conformity requirements are defined in such a way that it is not possible to test them in terms of the GUM infrastructure.

In this measurement scenario, temperature measurement is part of a wider process in which decisions such as whether to commission a storage device or raise an alarm for an operating storage unit are based upon independent temperature measurements.

PHSS provides extensive and clear guidance on where to measure and on what process to apply when an alarm condition is met. It embodies a conformity test ('Is an alarm condition met?') that is based upon specified limits for measured values of the measurand. It does not, however, provide a decision rule, i.e., it does not say how *measurement uncertainty* is to be taken into account when making such a decision. This omission means that the risk associated with such a decision and therefore the risk associated with subsequent actions is not immediately definable.

In general terms the PHSS process can be depicted schematically as shown in Figure E5.6.3. Two decisions need to be made, the first of these corresponding to a decision on whether, based upon the measured temperature, an alarm condition has been met. This particular decision is the subject of example 2. The second decision, concerning delay time, can be treated in a similar fashion.

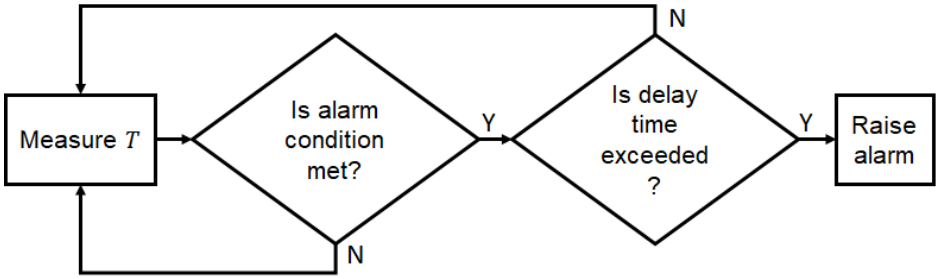


Figure E5.6.3: Alarm process for blood storage devices

E5.6.5.1 Measurand and model

As in example 1, the measurand, i.e., the quantity of interest, is the temperature prevailing at a location of interest either in air or within a sample such as a bag of fluid). Temperature is measured (monitored) by an alarm system when the store is in use.

A measurement model similar to that in example 1 could easily apply for the temperature measurements that are made. Consequently, the propagation and reporting of uncertainty would also be similar.

Where the examples differ significantly is that the PHSS guidance provides no information concerning measurement uncertainty or its role in the decision process. Instead it only refers to an 'accuracy' requirement.

E5.6.5.2 The problem with 'accuracy'

The term 'accuracy' is probably intended by authors to describe the expected quality of a measurement value. Unfortunately, however, its use creates immediate problems. Most obviously, difficulties arise due to the various meanings attached in common practice to the word 'accuracy'. Sometimes the intended use can be inferred from the context in which it is used, but often the meaning is ambiguous. For example if a requirement states that a weighing instrument 'should

be capable of measuring loads up to 4 kg with an accuracy of 0.1 kg' this requirement might commonly be interpreted as referring to the size of the measurement error, to the resolution of the display, or to the measurement uncertainty.

More fundamentally, from a metrological (GUM) standpoint, uncertainty and accuracy are entirely different concepts. This distinction is clear from the definitions, notes and annotations of the International Vocabulary of Metrology (VIM) [89] in which accuracy is described as a qualitative concept. For example, in the HTML version of the VIM [462] it is stated that:

> 'Historically, the term 'measurement accuracy' has been used in related but slightly different ways. Sometimes a single measured value is considered to be accurate (as in the VIM definition), when the measurement error is assumed to be small (in magnitude). In other cases, a set of measured values is considered to be accurate when both the measurement trueness and the measurement precision are assumed to be good. Sometimes a measuring instrument or measuring system is considered to be accurate, in the sense that it provides accurate indications. Care must therefore be taken in explaining in which sense the term 'measurement accuracy' is being used. In no case is there an established methodology for assigning a numerical value to measurement accuracy.'

The above matter is not simply a hypothetical problem or an issue of semantics — the GUM framework [2, 6] and standards such as ISO/IEC 17025 [7], ISO 15189 [39] are concerned with measurement uncertainty, not accuracy.

Faced with this situation in practice, the ideal approach is to obtain further information, such as formal clarification from the publishers of the guidance or standard on how 'accuracy' is to be interpreted and, if not also stated, how *uncertainty* is to be taken into account when specifications are tested using measured values. Unfortunately, this approach is often not feasible.

In such cases all that remains is for assumptions to be made and stated transparently, and for interpretations and additional requirements to be agreed between the testing body and its customer.

E5.6.5.3 Possible interpretations of the 'accuracy' requirement: metrologist

A metrologist might perhaps take a view that the term 'accuracy' should be interpreted as a misnaming of 'expanded uncertainty' — perhaps intended to provide a more 'familiar' language to less experienced users of a guide or standard. That interpretation might seem reasonable on the grounds that best practice involves making corrections for all know systematic effects (such as calibration errors), so it might be reasoned (by a metrologist) that the 'accuracy' requirement cannot be referring to calibration errors if best practice is to be followed. This assumption (that corrections have been made for all known systematic effects) is not unreasonable — it is made in all authoritative guidance documents on the subject of evaluating measurement uncertainty.

In this case an 'accuracy' requirement of say '±0.5 °C' [465, Clause 7.5.1] might then be interpreted as a requirement that (corrected) measured values of temperature should have an expanded uncertainty of no larger than $U_{\max} = 0.5\,°\mathrm{C}$, where we might suppose that the standard uncertainty has been expanded by a factor $k = 2$ (for a 95 % coverage interval).

The PHSS guidance does not explain how to take account of measurement uncertainty in deciding whether an alarm condition is met whilst the system is in use. Various Decision Rules are possible, as was seen in example 1. For example a rule based upon Simple Acceptance criteria might be adopted, in which the acceptance limit (for measured values) equates to a limit specified in the guidance, as at [465, clause 8.1.4]

Alarm condition not met: if $A_{\mathrm{L}} \leq T \leq A_{\mathrm{U}}$ AND $U \leq U_{\max}$

Alarm condition met: Otherwise

Alternatively a guard band might be applied for the given limit, defined in terms of the measurement uncertainty

E5.6.5.4 Possible interpretations of the 'accuracy' requirement: practitioner

In the context of this example the metrologist's interpretation of the intended meaning of 'accuracy' is perhaps less likely to be correct. Practitioners might take a view that making a correction for the calibration error ΔT_o was not expected by the authors of the guidance, nor it is feasible to do so on a routine basis. They might further choose to interpret the term 'accuracy' as describing a limit, say $\Delta T_{\max}$, to the maximum allowable measurement error [89, Clause 2.16].

This relatively common interpretation of 'accuracy' leaves open the question of measurement uncertainty, and the role it must have in subsequent decisions.

As has already been stated, best practice dictates that correction should be made for all known biases to establish a measurement result — albeit common practice often ignores this. Often instead, an 'uncorrected error' (such as an unused calibration correction) is incorporated into the uncertainty evaluation. Various approaches and rationales for this practice have been proposed [154], although strictly speaking these approaches do not usually establish a measurement uncertainty that is consistent with the principles within the GUM uncertainty framework [157] and can lead to significant unforeseen consequences [167].

However, in situations where the only (recognised) knowledge of calibration error ΔT_o is that its value is within some limiting value, $\Delta T_{\max}$ and that its standard uncertainty is $u(\Delta T_o)$, this information can be incorporated into a measurement model as two independent terms [168] . Thus

$$T = T_o + \delta T_{\Delta T} + \delta T_D + \delta T_x + \delta T,$$

where the squared standard uncertainty can be evaluated from

$$u^2(T) = u^2(T_o) + [u^2(\Delta T_{\max}) + u^2(\Delta T_o)] + u^2(\delta T_D) + u^2(\delta T_x) + u^2(\delta T).$$

Here $\delta T_{\Delta T}$ represents the 'poorly known' calibration correction whose value is known to be within an interval $[-\Delta T_{\max}, \Delta T_{\max}]$. The 'best estimate' of $\delta T_{\Delta T}$ is therefore taken as zero.

Given the measurement uncertainty and the specification, Decision Rules can now be established, such as the rules based upon Simple Acceptance criteria or guard bands described above. These rules assume that the PDF associated with T remains Gaussian. If this assumption does not hold then other approaches, such as the use of a Monte Carlo method [3], can be applied to evaluate conformance probability, risk or acceptance limits.

E5.6.6 Conclusions

The examples here have demonstrated some of the difficulties that are routinely faced when attempting to test for conformity against requirements described in guidance documents and standards. The documents cited here were not selected because they are particularly poor examples — they are in fact representative of the vast majority which, in other respects, provide

invaluable information to users. These examples show that, in order to make conformity decisions that are consistent with the GUM and with requirements for accreditation, some additional information, clarification or interpretation is often required.

Such a situation can be avoided in future guidance and standards by (i) suitably defining the role of measurement uncertainty, (ii) avoiding the use of the term 'accuracy', and (iii) ensuring correction is made for any bias in the measurement. Point (i) can be achieved, for example, 'indirectly', as a prerequisite for applying Simple Acceptance criteria, or 'directly' through its role in defining a guard band. Point (ii) can be addressed by using a quantitative concept such as measurement error or expanded uncertainty (for a given coverage probability such as 95 %) in place of the term 'accuracy'. Point (iii) can be handled by making a correction for any systematic effect and incorporating an uncertainty associated with the correction.

E5.6.7 Acknowledgement

Health Services Scotland provided valuable input to the development of this example.

Part E6

Industry and society

Example E6.1

Measurement uncertainty evaluation for turbofan nozzle thrust derived from non-intrusive flow measurements

M.G. Cox, J. Dawkins, J. Gillespie, T. Lowe, W. Ng, J. Roberts

E6.1.1 Summary

The estimation of nozzle thrust and the evaluation of the associated uncertainty in a turbofan jet engine is considered given information derived from acoustic sensors. There are two scenarios: the first is uniform or homogeneous flow, and the second non-uniform flow. The first scenario is entertained here, whereas the second lies beyond the scope of the current study.

Although nozzle thrust is the primary measurand, several other measurands — jet velocity, speed of sound, nozzle exit Mach number and exhaust temperature — are considered. For one experimental configuration, formulæ are given for these measurands in terms of (input) quantities defining the geometry of the set-up, acoustic times of flight and properties of the medium. Expressions are developed for the uncertainties associated with estimates of these quantities in accordance with the internationally acknowledged Guide to the expression of Uncertainty in Measurement (GUM) published by the Joint Committee on Guides in Metrology (JCGM).

For data representative of an actual engine test, estimates of nozzle thrust and the other four measurands, along with evaluations of their associated uncertainties, are provided.

Possibilities are considered for adapting the experimental configuration to yield reduced uncertainties associated with estimates of the measurands. These possibilities are illustrated by redoing the calculations for a simple change in the geometry of the set-up. Even though this change is not optimized, a reduction in the relative standard uncertainty in nozzle thrust of some 60 % is achieved, with corresponding reductions for the other measurands.

E6.1.2 Introduction of the application

The thrust of a turbofan engine is a key parameter to understanding both the performance of a new or improved propulsion system design, and to ensuring that an engine returning to service after repair or overhaul is still performing as intended.

Off the wing, that is, on the ground, an individual engine's net thrust can be determined by direct measurement of the forces imparted by the engine on load cells in a bespoke structure in a test cell. Figure E6.1.1 (left) shows a jet engine under test in an indoor test cell with thrust cradle, used for data acquisition, and (right) an outdoor test facility for future target applications.

Figure E6.1.1: Test facilities for indoor (left) and outdoor (right) data acquisition

However, direct measurement of force exerted by the power plant on the mountings is difficult to carry out with adequate accuracy for a power plant installed in an air-frame. Firstly, the load path(s) between the power plant and air-frame are more complex than on the test stand, making direct measurement of load challenging. Secondly, any changes in aircraft attitude modifies the forces in the engine/air-frame interface, which will in turn influence the errors in the measured thrust.

Nozzle thrust can also be derived from measurements of exhaust flow, and determined from multiple measurements of total pressure and total temperature. These measurements are traditionally made using multiple 'rakes' of individual pressure and temperature sensing 'Kiel' heads [466], in combination with static wall pressure tappings. The methods of determining thrust from aerodynamic performance measurements are well established: a typical traditional approach is described in [467]. To achieve an acceptable level of uncertainty, this approach requires a large number of individual temperature and pressure measurement points to ensure good capture of the complex flow, but this in turn becomes increasingly intrusive to the flow being measured. The rakes required introduce blockage to the flow, especially if they need to be designed to be suitable for flight. Additionally, a large volume of signal-conditioning equipment needs to be accommodated on board the aircraft.

Both these traditional approaches are therefore difficult to realize practically on an engine installed in an air-frame.

A seminal document [467] concerns the indirect measurement of thrust of turbojet and turbofan engines at steady conditions. It covers the calculation or determination of thrust as opposed to the *measurement* of thrust and despite its age remains the most complete summary of previous methods for thrust determination. There are many more recent papers on the better understanding or optimization of total pressure or of total temperature probes or both, but contain little specific information relating to thrust determination using them.

An alternative, non-intrusive, approach to determining nozzle thrust, is therefore highly desirable. However, the uncertainties in the results so obtained of the approach need to be understood. Here we describe a non-intrusive acoustic techniques known as sonic anemometry that uses distance and acoustic propagation time-of-flight (TOF) measurements. The TOF is a function of propagation distance and direction, flow velocity and speed of sound.

Sonic anemometry is a well known and often used measurement technique in low-speed flow applications such as atmospheric meteorology and hydrology [468–474]. It is often coupled with sonic thermometry so as also to estimate the gas temperature [472, 474]. Sonic thermometry is also used to measure temperatures in some industrial applications, such as power plant boilers, and is usually called acoustic pyrometry on these fields [475, 476]. Most of these applications involve flows at low speeds (typically Mach numbers below 0.1). There are several significant challenges to extending the technique to the compressible regime. These challenges include typically high background noise, as well as significant attenuation of the transmitted signal by turbulence in the flow. Additionally, under some circumstances, it may be necessary to account for the refraction of the acoustic rays by the moving fluid. The application of sonic anemometry and sonic thermometry to higher speed flows, like those present in turbofan engines, was pioneered by Otero et al. [477–480], targeted for future in-flight measurement of engine exhaust thrust.

A sonic anemometry approach is used together with an uncertainty assessment for a current experimental configuration. Two cases are possible: uniform flow, namely, constant flow and constant (thermodynamic) speed of sound propagation throughout the medium, and non-uniform flow. Only uniform flow is considered in detail although some comments are made on the non-uniform case.

Potential advantages of changing the experimental configuration are considered. It is shown that the uncertainty associated with estimated nozzle thrust can usefully be reduced.

E6.1.3 Specification of the measurands

E6.1.3.1 Assumptions

A Cartesian xy co-ordinate system is used in which x denotes the direction of the flow with y orthogonal to the flow. Assumptions made are:

1. The exhaust gas is calorically perfect (the specific heat capacity of the gas is constant),
2. The speed of sound in the medium is constant and independent of direction,
3. Flow velocity in the x-direction is constant,
4. Flow velocity in the y-direction is negligible when compared to that in the x-direction.

E6.1.3.2 Primary and secondary measurands

The measurands are shown as rectangular boxes in figure E6.1.2, which gives a workflow for the required calculations. The primary measurands are the nozzle thrust F and exhaust temperature T shown at the bottom and bottom-right of the figure. Intermediate measurands are jet velocity v_x, speed of sound C in the medium and nozzle exit Mach number M on which v_x and T depend. The input quantities to the calculations, on which the measurands depend, are shown as trapezoidal boxes in the figure.

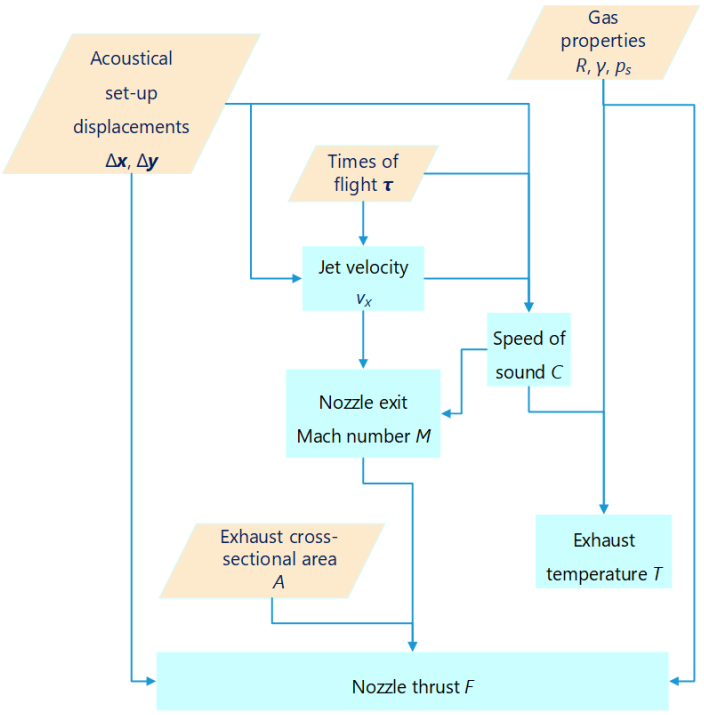


Figure E6.1.2: Computational workflow. Input quantities are shown in trapezoidal boxes. Output quantities (measurands) appear in rectangular boxes: nozzle thrust and exhaust temperature are primary measurands; the other three quantities are intermediate measurands

E6.1.3.3 Coordinate system

The coordinate system used for the analysis is that in [477] and displayed in figure E6.1.3:

- x points axially, pointing aft (the direction in which the exhaust points),
- y points horizontally, from the side of the exhaust with the source to the side with the acoustic receivers,
- z points vertically, away from the ground,
- The origin is centred on the turbofan exhaust, and in the same axial plane as the sound source, that is, the centre of exhaust is located at $(x, 0, 0)$ and the source is at $(0, y, 0)$,
- The source and receivers all lie in the xy plane ($z = 0$).

E6.1.3.4 Acoustic components

Figure E6.1.4 depicts an arrangement of the acoustic components. That particular configuration contains a (single) source located on one side of the flow field and two receivers ('downstream' and 'upstream') on the other, displaced in the stream-wise direction.

E6.1.4 Measurement model

The measurement model for the primary measurand F and T and the intermediate measurands v_x, C and M (see figure E6.1.2) comprises the following formulæ

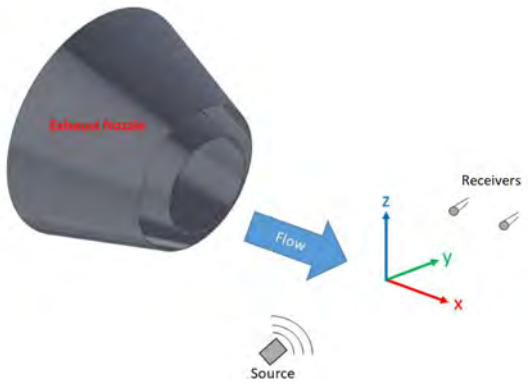


Figure E6.1.3: Coordinate system, with exhaust, sound source and receivers shown

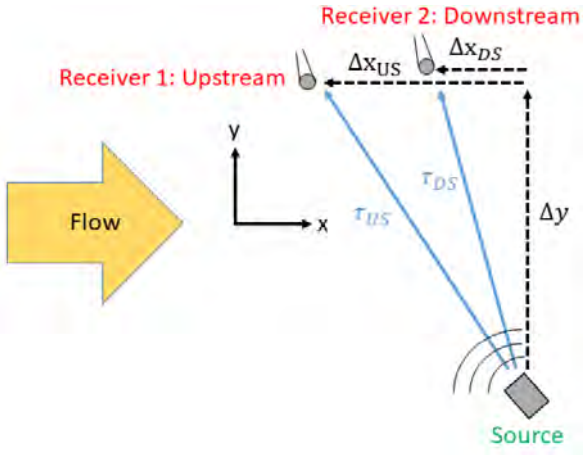


Figure E6.1.4: Upstream and downstream acoustic component displacements and times of flight

1. Jet velocity v_x in the direction of flow given displacements Δx_{US}, Δx_{DS}, Δy_{US} and Δy_{DS}, and times of flight τ_{US} and τ_{DS}:

$$v_x = \frac{1}{2}\frac{\dfrac{\Delta x_{\mathrm{DS}}^2 + \Delta y_{\mathrm{DS}}^2}{\tau_{\mathrm{DS}}^2} - \dfrac{\Delta x_{\mathrm{US}}^2 + \Delta y_{\mathrm{US}}^2}{\tau_{\mathrm{US}}^2}}{\dfrac{\Delta x_{\mathrm{DS}}}{\tau_{\mathrm{DS}}} - \dfrac{\Delta x_{\mathrm{US}}}{\tau_{\mathrm{US}}}}. \tag{E6.1.1}$$

2. Speed of sound C in the medium given v_x from formula (E6.1.1) and (once more) Δx_{US}, Δy_{US} and τ_{US}:

$$C = \left[\left(\frac{\Delta x_{\mathrm{US}}}{\tau_{\mathrm{US}}} - v_x\right)^2 + \left(\frac{\Delta y_{\mathrm{US}}}{\tau_{\mathrm{US}}}\right)^2\right]^{1/2}. \tag{E6.1.2}$$

3. Exhaust temperature T given C from formula (E6.1.2) and the gas properties γ and R:

$$T = \frac{C^2}{\gamma R}. \tag{E6.1.3}$$

4. Nozzle exit Mach number M given v_x from formula (E6.1.1) and C from formula (E6.1.2):

$$M = \frac{v_x}{C}, \tag{E6.1.4}$$

5. Nozzle thrust F given M from formula (E6.1.4), and the gas properties p_s and R, and the cross-sectional area A of the exhaust:

$$F = p_s \gamma M^2 A. \tag{E6.1.5}$$

In the above expressions, p_s is atmospheric static pressure, γ is the specific heat ratio, R is the specific gas constant for the exhaust gas and A is the cross-sectional area of the exhaust.

formulæ (E6.1.1) and (E6.1.2) are derived in appendix E6.1.A by an analysis involving the relative displacement of components and times of flight along two acoustic propagation paths.

E6.1.4.1 Data

In terms of dimensional quantities and times of flight, fundamental quantities relating to figure E6.1.4 are Δx_{US}, Δx_{DS}, Δy_{US}, Δy_{DS}, τ_{US} and τ_{DS}. Δx_{US} and Δx_{DS} are the upstream and downstream displacements from the source of the receivers in the x-direction, Δy_{US} and Δy_{DS} are the counterparts in the y-direction (they are both equal to Δy in figure E6.1.4), and τ_{US} and τ_{DS} denote the corresponding acoustic times of flight.

Δx_{US}, Δx_{DS}, Δy_{US} and Δy_{DS} are measured before the experiment, while τ_{US} and τ_{DS} are the times of flight measured during the experiment. Associated standard uncertainties are available. Values and associated uncertainties are also available for the quantities A, R, γ and p_s.

The ten quantities of which the data items Δx_{US}, Δx_{DS}, Δy_{US}, Δy_{DS}, τ_{US}, τ_{DS}, A, R, γ and p_s constitute measured values are regarded as statistically independent input quantities in the measurement model specified by formulæ (E6.1.1) to (E6.1.5).

E6.1.5 Measurement uncertainty evaluations

E6.1.5.1 Generalized law of propagation of uncertainty

In applying the law of propagation of uncertainty (LPU) in the Guide to the expression of Uncertainty in Measurement (GUM) [2] to the measurement model specified by the workflow in figure E6.1.2 in section E6.1.4, the standard uncertainties $u(F)$ and $u(T)$ associated with the primary measurands F and T can be obtained together with those, $u(v_x)$, $u(C)$ and $u(M)$, associated with the secondary measurands v_x, C and M.

To apply LPU to evaluate these standard uncertainties, partial derivatives of the relevant measurands with respect to the input quantities are required. These derivatives constitute sensitivity coefficients representing measures of dependence of the measurands upon the input quantities. Since there are $N = 10$ input quantities and $m = 5$ measurands, there will be $10 \times 5 = 50$ such partial derivatives. These derivatives are arranged to form a sensitivity matrix $\boldsymbol{C}_{\boldsymbol{x}}$ of dimension $N \times m$. Some of these derivatives will be zero since some measurands do not depend on certain input quantities. For instance, jet velocity v_x is not influenced by the gas property γ.

Appendix E6.1.B gives an outline of the generalization of the LPU to multivariate measurands, m in number, and N input quantities.

For this application, the $N = 10$ input quantities and $m = 5$ output quantities are respectively

$$\boldsymbol{X} \equiv [\Delta x_{US} \;\; \Delta x_{DS} \;\; \Delta y_{US} \;\; \Delta y_{DS} \;\; \tau_{US} \;\; \tau_{DS} \;\; A \;\; R \;\; \gamma \;\; p_s]^\top, \qquad \boldsymbol{Y} \equiv [v_x \;\; C \;\; T \;\; M \;\; F]^\top \tag{E6.1.6}$$

and, since the input quantities have been assumed to be independent, the covariance matrix $\boldsymbol{V}_{\boldsymbol{x}}$ of dimension $N \times N$ (10×10) is a diagonal matrix with diagonal elements

$$[u^2(\Delta x_{\mathrm{US}})\ u^2(\Delta x_{\mathrm{DS}})\ u^2(\Delta y_{\mathrm{US}})\ u^2(\Delta y_{\mathrm{DS}})\ u^2(\tau_{\mathrm{US}})\ u^2(\tau_{\mathrm{DS}})\ u^2(A)\ u^2(R)\ u^2(\gamma)\ u^2(p_{\mathrm{s}})]^{\top}.$$

Equivalently, indicating zero elements by blanks,

$$\boldsymbol{V}_{\boldsymbol{x}} = \begin{bmatrix} u^2(x_1) & & \\ & \ddots & \\ & & u^2(x_N) \end{bmatrix} = \begin{bmatrix} u^2(\Delta x_{\mathrm{US}}) & & \\ & \ddots & \\ & & u^2(p_{\mathrm{s}}) \end{bmatrix}. \tag{E6.1.7}$$

The other component in LPU is the above sensitivity matrix $\boldsymbol{C}_{\boldsymbol{X}}$ evaluated at $\boldsymbol{X} = \boldsymbol{x}$. Considerable care is required when obtaining the elements of $\boldsymbol{C}_{\boldsymbol{X}}$. Although section E6.1.4 gives formulæ for the measurands in terms of the input quantities, these formulæ do not individually indicate that dependence *explicitly*. Apart from in formula (E6.1.1) there are 'hidden' variables, which must be taken into account in obtaining valid uncertainty evaluations. Formula (E6.1.2) expresses C is terms of three input quantities Δx_{US}, Δy_{US} and τ_{US}, and jet velocity v_x, but C also depends indirectly on Δx_{DS}, Δy_{DS} and τ_{DS} since C depends on v_x, which in turn depends on these three further quantities. Thus, C depends on all displacement and time of flight data. Similarly, T, M and F depend on eight, six and nine input quantities rather than the three, two and four, respectively, that appear explicitly in formulæ (E6.1.3) to (E6.1.5).

These observations imply a high degree of complexity in the analytical evaluation of the partial derivatives of the measurands with respect to the quantities that influence them. Analytical differentiation will require repeated careful application of the chain rule of differential calculus to ensure that all dependencies are fully considered. This degree of complexity is to be contrasted with the simplicity of the formulæ themselves for v_x, C, T, M and F in section E6.1.4, which can be evaluated, depending on the precise way the formulæ are arranged, with only some 24 basic arithmetic operations (plus, minus, times, divide) and one square root.

The safest way to capture all essential dependencies is to automate the process of producing first-order partial derivatives or sensitivity coefficients. Accordingly, we considered the advice given in [40, 481], which covers automatic differentiation techniques and their application in metrology. We examine three approaches, symbolic differentiation, the complex step method and finite differences, in appendix E6.1.C. These methods avoids this complexity and the associated high risk of accompanying algebraic error. Two of these methods are capable of providing a degree of numerical accuracy comparable to that provided by analytical methods. The third method provides less numerical accuracy but is more than sufficient for practical requirements.

A computationally intensive approach (Monte Carlo method (MCM)) to confirming the correctness of the evaluation is also considered. This approach requires assignment of probability distributions to the input quantities. Normal distributions were chosen for this purpose with means equal to the input estimates $x_1, \ldots, x_N$ and standard deviations equal to the associated standard uncertainties $u(x_1), \ldots, u(x_N)$. All that is then required is the evaluation (a very large number of times) of the model equations (E6.1.1) to (E6.1.5), from which measurand estimates, standard uncertainties and covariances can readily be extracted. Appendix E6.1.D gives a brief description of the approach.

E6.1.5.2 Data

Table E6.1.1 gives typical data. For proprietary reasons, the data are not actual engine test data but representative of what can be obtained from such tests. The uncertainties quoted for the gas constants in table E6.1.1 take account of the gas composition at the exit of the engine not being

constant. A varying proportion of combustion products, mostly CO_2 and H_2O, will be present depending on engine operating conditions. For a modern high by-pass ratio turbofan the majority of the flow will be entrained air.

Table E6.1.1: Typical data: values and associated expanded uncertainties

Quantity	Unit	Value	Expanded uncertainty (≈ 95 % confidence)
Δx_{US}	m	−0.200	0.002
Δx_{DS}	m	−0.050	0.002
Δy_{US}	m	1.000	0.002
Δy_{DS}	m	1.000	0.002
τ_{US}	ms	3.03	0.010
τ_{DS}	ms	2.79	0.010
A	m^2	0.7854	0.0031
R	$J\,kg^{-1}\,K^{-1}$	287	2
γ	—	1.38	0.03
p_s	Pa	101 325	100

Standard uncertainties are taken as half the reported expanded uncertainties, that is, by assuming the corresponding quantities are characterized by normal distributions and that the expanded uncertainties correspond to a confidence level of approximately 95 %.

It is worth mentioning more about the nature of the uncertainties for τ_{US} and τ_{DS}, given their central importance in the measurement. Their expanded uncertainty is listed as constant. This behaviour stems from the fundamental measurement being one of acoustic signal phase. Changes to the measuring system will leave phase uncertainty of the signal processing scheme unaltered, given that the signal-to-noise ratio and the acoustic carrier waveform remain the same.

E6.1.5.3 Results

Using the data in column 3 of table E6.1.1, formulæ (E6.1.1) to (E6.1.5) in section E6.1.4 were used to form estimates of the five measurands. These estimates are given in column 2 of table E6.1.2.

Table E6.1.2: Measurand estimates and their associated standard uncertainties

Quantity	Estimate	Std. unc.		Rel. std. unc./%	
		Current	New	Current	New
Jet velocity $v_x/m\,s^{-1}$	161	8	3	4.7	1.9
Speed of sound $C/m\,s^{-1}$	401	4	2	1.1	0.6
Exhaust temperature T/K	405	10	7	2.5	1.6
Nozzle exit Mach number M	0.402	0.015	0.005	3.7	1.3
Nozzle thrust F/kN	17.8	1.3	0.5	7.4	2.9

As said, the expanded uncertainties in column 4 of table E6.1.1 were halved, assuming underlying normality, to give the standard uncertainties associated with the values in column 3 of that table. These standard uncertainties were squared to form the corresponding variances, which were used as the diagonal elements of the diagonal covariance matrix $\boldsymbol{V}_{\boldsymbol{x}}$ in expression (E6.1.7).

The sensitivity matrix $\boldsymbol{C}_x$, given in table E6.1.3, was formed using the complex step method (appendix E6.1.C.3). The ith numbered row corresponds to the ith element of the vector input quantity in expression (E6.1.6) and the jth numbered column to the jth vector output quantity. The unit of the element in row i and column j is the unit of the jth measurand divided by that of the ith input quantity. A check was made on the calculation using the finite-difference method outlined in section E6.1.C.2 to form an approximate sensitivity matrix. The relative difference between the 2-norms of the two matrices was $\mathcal{O}(1 \times 10^{-8})$, so confirming the adequacy of these calculations.

Table E6.1.3: Sensitivity matrix provided by the complex step method (blank = zero element). The unit of the element in row i and column j is the unit of the jth measurand divided by that of the ith input quantity

	1	2	3	4	5
1	1.560×10^3	6.975×10^2	1.412×10^3	3.192	2.822×10^5
2	-2.265×10^3	-1.013×10^3	-2.050×10^3	−4.635	-4.098×10^5
3	-1.336×10^3	-7.576×10^2	-1.533×10^3	−2.572	-2.274×10^5
4	2.672×10^3	1.515×10^3	3.066×10^3	5.145	4.548×10^5
5	8.505×10^5	3.803×10^5	7.696×10^5	1.741×10^3	1.539×10^8
6	-9.815×10^5	-5.567×10^5	-1.126×10^6	-1.890×10^3	-1.671×10^8
7					2.265×10^4
8			−1.413		
9			-2.938×10^2		1.289×10^4
10					1.755×10^{-1}

Formula (E6.1.11) in appendix E6.1.B gives the generalization of the LPU to a multivariate measurand. Insertion of $\boldsymbol{V}_x$ and $\boldsymbol{C}_x$ into that formula gives the covariance matrix $\boldsymbol{V}_y$ of dimension 5×5 associated with the estimates of the five measurands. The square roots of the diagonal elements of $\boldsymbol{V}_y$ are the standard uncertainties associated with those estimates. They are given in column 3, marked 'Current', of table E6.1.2. Relative standard uncertainties, that is, these standard uncertainties divided by the corresponding estimates in column 2 are given in column 5.

$\boldsymbol{V}_y$ is a full matrix (units not given here but they are the appropriate products of the units of the elements of $\boldsymbol{y}$):

$$\boldsymbol{V}_y = \begin{bmatrix} 58.6559 & 30.1900 & 61.0905 & 0.1161 & 10258.7839 \\ 30.1900 & 15.7465 & 31.8636 & 0.0595 & 5261.7129 \\ 61.0905 & 31.8636 & 85.8935 & 0.1205 & 9795.2098 \\ 0.1161 & 0.0595 & 0.1205 & 0.0002 & 20.3164 \\ 10258.7839 & 5261.7129 & 9795.2098 & 20.3164 & 1834599.2814 \end{bmatrix}.$$

The diagonal elements of $\boldsymbol{V}_y$ are the squares of the standard uncertainties associated with $\boldsymbol{y}$, that is $[u^2(v_x)\ u^2(C)\ u^2(T)\ u^2(M)\ u^2(F)]$, and the covariances between pairs of estimate are the off-diagonal elements. For instance, the element in row 1 and column 2 is the covariance between the first and second elements of the measurand vector $\boldsymbol{y}$, namely, that between v_x and C.

It is virtually impossible to say anything concrete about the magnitudes of the correlations between the various quantities following a cursory examination of the elements of $\boldsymbol{V}_y$. However, the correlation matrix associated with $\boldsymbol{y}$ obtained for the original data is

$$\boldsymbol{R}_y = \boldsymbol{D}^{-1}\boldsymbol{V}_y\boldsymbol{D}^{-1} = \begin{bmatrix} 1.000 & 0.993 & 0.861 & 1.000 & 0.989 \\ 0.993 & 1.000 & 0.866 & 0.989 & 0.979 \\ 0.861 & 0.866 & 1.000 & 0.857 & 0.780 \\ 1.000 & 0.989 & 0.857 & 1.000 & 0.989 \\ 0.989 & 0.979 & 0.780 & 0.989 & 1.000 \end{bmatrix},$$

where $\boldsymbol{D}$ is a diagonal matrix with diagonal $[u(v_x)\ \ u(C)\ \ u(T)\ \ u(M)\ \ u(F)]$. For example, the correlation between v_x and C is element (1, 2) of $\boldsymbol{R}_y$, namely, 0.993. This and several other correlations are very close to unity, indicating the strong mutual dependence on the input data.

Monte Carlo calculations for 1×10^7 trials, requiring about 60 s on a modest laptop computer, reproduced the above results to within two to three significant decimal digits.

Table E6.1.4 gives the uncertainty budget for the estimate of each measurand in terms of the contributing quantities. Rows in the table correspond to the ten input quantities and columns to the measurands of interest. Note that the zero-nonzero structure is identical to that of the sensitivity matrix $\boldsymbol{C}_x$ in table E6.1.3 for the following reason. The entry in row i and column j of the uncertainty budget is the product of the standard uncertainty $u(x_i)$ associated with the ith input quantity x_i and the magnitude of $c_{i,j}$, the first-order partial derivative (sensitivity coefficient) of the jth measurand with respect to the ith input quantity evaluated at the input estimates. By comparison, the corresponding element in the sensitivity matrix in table E6.1.3 is simply $c_{i,j}$. So, since the input standard uncertainties are all non-zero, $u(x_i)|c_{i,j}|$ will be zero or non-zero according to whether $c_{i,j}$ has this property. The standard uncertainties associated with the measurand estimates take the values given in the antepenultimate row of the table, marked 'Combined'.

Inspection of table E6.1.4 reveals that the standard uncertainties associated with measured times of flight, τ_{US} and τ_{DS}, make the largest contributions to the estimates of all measurand.

To improve measurand uncertainties, it is well worth reducing the uncertainties associated with times of flight. If hypothetically they could be reduced to zero, the standard uncertainties associated with the measurand estimates would take the values given in the penultimate row of the table, marked 'Combined*'. Consequently, all measurand uncertainties would be reduced by over 45 % with the exception of that for gas temperature T, which would be reduced by over 30 %.

Of course, such reductions are totally unrealistic but give an indication of the limit of achievement consequent on improving time of flight uncertainties.

E6.1.6 Interpretation of results

E6.1.6.1 Validation of results

Attention has been paid to validating the results of the uncertainty evaluation. The use of the LPU requires first-order partial derivatives of the five measurands with respect to the ten input quantities. The measurands are given by relatively simple formulæ (E6.1.1) to (E6.1.5) but they form a cascade of calculations as indicated in figure E6.1.2. Calculating partial derivatives analytically thus involves repeated application of the chain rule of differential calculus, with considerable scope for algebraic error. As a consequence, several approaches for *automatic* calculation of

Table E6.1.4: Uncertainty budget for measurand estimates in terms of the contributing quantities

	Measurand (output quantity)				
Input quantity	$v_x/(\mathrm{m\,s^{-1}})$	$C/(\mathrm{m\,s^{-1}})$	T/K	M	F/kN
$\Delta x_{\mathrm{US}}/\mathrm{m}$	1.6	0.7	1.4	0.0032	282
$\Delta x_{\mathrm{DS}}/\mathrm{m}$	2.3	1.0	2.0	0.0046	410
$\Delta y_{\mathrm{US}}/\mathrm{m}$	1.3	0.8	1.5	0.0026	227
$\Delta y_{\mathrm{DS}}/\mathrm{m}$	2.7	1.5	3.1	0.0051	455
$\tau_{\mathrm{US}}/\mathrm{ms}$	4.3	1.9	3.8	0.0087	769
$\tau_{\mathrm{DS}}/\mathrm{ms}$	4.9	2.8	5.6	0.0095	835
$A/\mathrm{m^2}$					36
$R/(\mathrm{J\,kg^{-1}\,K^{-1}})$			1.4		
γ			4.4		193
$p_{\mathrm{s}}/\mathrm{Pa}$					9
Combined std. unc.	7.7	4.0	9.3	1.52	1354
Combined*	4.1	2.1	6.3	8.0	738
Reduction/%	47	47	32	47	46

these derivatives were reviewed: symbolic algebra, the complex step method and finite differences. The latter two approaches were applied and found to be in agreement to a numerical precision far surpassing that required in the application. As a further test, the MCM was applied with a large number of trials (1×10^7) after assigning normal distributions to input quantities with parameters defined in terms of the provided estimates and associated standard uncertainties. Agreement of results with those for the complex step method was found to the numerical precision expected of the MCM with that number of trials.

We also checked some of the derivative calculations by obtaining manually analytical expressions for some of the partial derivatives, employing the chain rule of differential calculus. We also used the symbolic calculation engine [482] to check these derivatives.

Concerning the formulæ themselves that specify the measurands, it is not possible to make any statement about their faithful implementation on the computer apart from the authors themselves verifying that the results are reasonable in the context of the application. This check was undertaken.

E6.1.6.2 Experimental design considerations

Some progress can be made on implementing experimental design principles. Times of flight in terms of the considered configuration can be determined from equation (E6.1.2). If the determined values of C and v_x are used in this equation, the computed time of flight τ_{US} (and τ_{DS} similarly) would be recovered:

$$\tau_{\mathrm{US}} = \frac{-\Delta x_{\mathrm{US}} v_x + [(\Delta x_{\mathrm{US}} v_x)^2 + (C^2 - v_x^2)(\Delta x_{\mathrm{US}}^2 + \Delta y_{\mathrm{US}}^2)]^{1/2}}{C^2 - v_x^2},$$

which we express in terms of nozzle exit Mach number $M = v_x/C$ [expression (E6.1.4)] as

$$\tau_{\mathrm{US}} = \frac{-M\Delta x_{\mathrm{US}} + [\Delta x_{\mathrm{US}}^2 + (1-M^2)\Delta y_{\mathrm{US}}^2]^{1/2}}{(1-M^2)C}.$$

To avoid the potential loss of numerical precision for M close to unity, we rationalize the expression by multiplying both numerator and denominator by the conjugate of the former, giving

$$\tau_{\mathrm{US}} = \frac{\Delta x_{\mathrm{US}}^2 + \Delta y_{\mathrm{US}}^2}{C\{M\Delta x_{\mathrm{US}} + [\Delta x_{\mathrm{US}}^2 + (1-M^2)\Delta y_{\mathrm{US}}^2]^{1/2}\}}, \tag{E6.1.8}$$

with a comparable expression for τ_{DS}.

In terms of the assumption in section E6.1.3.1 that C and v_x are constant over the medium, other configurations can be considered by changing the displacements Δx_{US}, Δy_{US}, Δx_{DS} and Δy_{DS}. Suppose a single change is made to the configuration: move receiver 1 further upstream, that is, change Δx_{US} from −0.2 m to a different value. Expression (E6.1.8) and and its counterpart for τ_{DS} provide the corresponding times of flight for the new configuration: Δx_{US} will change in value but, as expected, τ_{US} will be unaltered. The estimates of the primary measurands v_x, C, T, M and F will be unaltered but their associated relative standard uncertainties will change. Figure E6.1.5 shows the manner in which the relative standard uncertainty in two of the measurands, exhaust thrust F and jet velocity v_x, varies as a function of Δx_{US}.

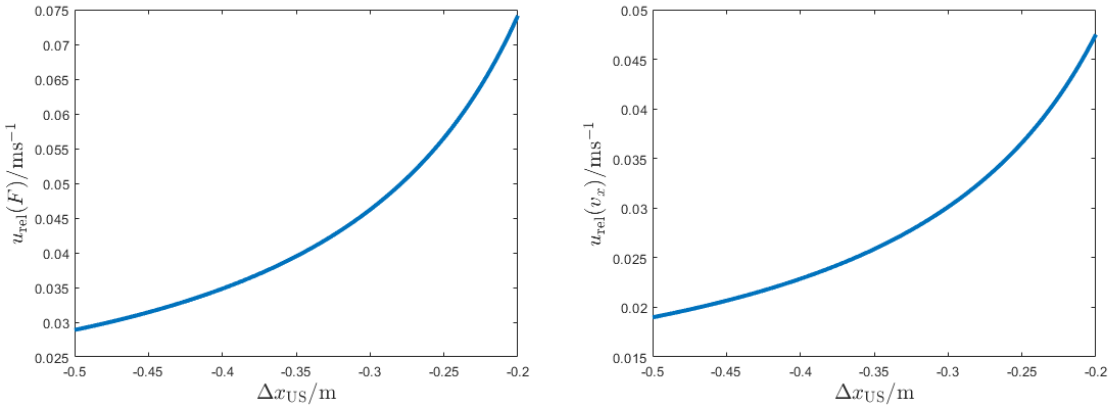


Figure E6.1.5: Relative standard uncertainty $u_{\mathrm{rel}}(v_x)$ in exhaust thrust F and (right) jet velocity v_x as functions of displacement Δx_{US} of receiver 1 from source in x-direction

Changing Δx_{US} from the original value of −0.2 m to −0.3 m has a relatively large benefit: $u_{\mathrm{rel}}(F)$ changes from 7.4 % to 4.6 % and $u_{\mathrm{rel}}(v_x)$ from 4.7 % to 3.0 %.

The results of the calculation with the data revised such that only Δx_{US} is changed, from −0.2 m to −0.5 m, are given in columns 4 and 6 marked 'New' in table E6.1.2. In that instance,

$$\tau_{\mathrm{US}} = 3.71\,\mathrm{ms}, \qquad \tau_{\mathrm{DS}} = 2.79\,\mathrm{ms}$$

and, as expected, τ_{US} has changed and τ_{DS} is unaltered. The estimates of the primary measurands v_x, C, T, M and F are unchanged, as anticipated, but their associated relative standard uncertainties have improved, mostly lying between approximately one third and one half of their previous values. The only exception is T, for which the standard uncertainty is about two thirds of its original value.

E6.1.6.3 Suggestions for further work

1. The two receivers are generally arranged such that they are shifted axially from each other, that is, the horizontal spacing is the same for both receivers ($\Delta y_{\mathrm{DS}} = \Delta y_{\mathrm{US}}$). As stated, the approach considered is general and would operate with alternative spacings in the x- and y-directions. Optimization of the measurand standard uncertainties over a prescribed region for the location of the sources and receivers is possible, but the practicalities of re-positioning sources and receivers must be respected.

2. If further calculations are to be carried out with some or all of the measurands, namely, v_x, C, T, M and F, account must be taken of the especially high correlations associated with their values. Otherwise, any resulting uncertainties will be invalid.

3. Further study of the methods used for measuring displacements and times of flight might identify some common effects across individual observations, which would induce correlation in the input data. Using the principles expounded in [5], the input covariance matrix (section E6.1.5) would be modified to take account of the necessary covariance terms, and the process would proceed in the same manner.

4. Only the homogeneous case is covered in the treatment here. Subsequent work would usefully consider the case when the flow and temperature become non-uniform throughout the exhaust. The use of a number of sources and receivers and determining TOFs between them could be beneficial. With a criss-cross pattern of paths and the use of least squares, the non-uniform effects would be averaged out to some extent.

E6.1.A Derivation of formulæ for the speed of sound and flow velocity

Formulæ are derived for the speed of sound and flow velocity.

Let β_1 denote the angle between the y-axis and path 1 (from the source to receiver 1) in figure E6.1.4. Using 'distance = velocity × time' in the x- and y-directions,

$$\Delta x_{\mathrm{US}} = (v_x + C_1 \cos\beta_1)\tau_{\mathrm{US}},$$
$$\Delta y_{\mathrm{US}} = C_1 \tau_{\mathrm{US}} \sin\beta_1,$$

where C_1 is the (average) speed of sound along the path (also see [477, 478, 483]). There are similar expressions for path 2 (from source 1 to receiver 2) involving angle β_2 between the y-axis and path 2. Upon applying $\cos^2\theta + \sin^2\theta = 1$ for $\theta = \beta_1$ and $\theta = \beta_2$,

$$C_1^2 = \left(\frac{\Delta x_{\mathrm{US}}}{\tau_{\mathrm{US}}} - v_x\right)^2 + \left(\frac{\Delta y_{\mathrm{US}}}{\tau_{\mathrm{US}}}\right)^2, \qquad C_2^2 = \left(\frac{\Delta x_{\mathrm{DS}}}{\tau_{\mathrm{DS}}} - v_x\right)^2 + \left(\frac{\Delta y_{\mathrm{DS}}}{\tau_{\mathrm{DS}}}\right)^2.$$

Then, using assumption 2 in section E6.1.3.1, the speed of sound in the medium $C = C_1 = C_2$, implying formula (E6.1.2) in section E6.1.4.

$$C^2 = \left(\frac{\Delta x_{\mathrm{US}}}{\tau_{\mathrm{US}}} - v_x\right)^2 + \left(\frac{\Delta y_{\mathrm{US}}}{\tau_{\mathrm{US}}}\right)^2 = \left(\frac{\Delta x_{\mathrm{DS}}}{\tau_{\mathrm{DS}}} - v_x\right)^2 + \left(\frac{\Delta y_{\mathrm{DS}}}{\tau_{\mathrm{DS}}}\right)^2. \tag{E6.1.9}$$

Solving equation (E6.1.9) for v_x gives

$$v_x = \frac{1}{2} \frac{\dfrac{\Delta x_{\mathrm{DS}}^2 + \Delta y_{\mathrm{DS}}^2}{\tau_{\mathrm{DS}}^2} - \dfrac{\Delta x_{\mathrm{US}}^2 + \Delta y_{\mathrm{US}}^2}{\tau_{\mathrm{US}}^2}}{\dfrac{\Delta x_{\mathrm{DS}}}{\tau_{\mathrm{DS}}} - \dfrac{\Delta x_{\mathrm{US}}}{\tau_{\mathrm{US}}}} \tag{E6.1.10}$$

as the uniform flow velocity v_x in the direction of flow (x-direction).

Thus, the speed of sound depends on flow velocity and the downstream (or upstream) displacements and times of flight. In turn, v_x is calculated from both upstream and downstream displacements and times of flight.

E6.1.B The generalization of LPU to multivariate measurands

GUM Supplement 2 (GUM-S2) [4] gives an extension of the LPU to multivariate measurands, m in number and N input quantities, for which a brief outline is given here.

The relationship between a multivariate output quantity $\boldsymbol{Y} = (Y_1, \ldots, Y_m)^\top$ and a multivariate input quantity $\boldsymbol{X} = (X_1, \ldots, X_N)^\top$ is specified by a multivariate measurement model

$$\boldsymbol{Y} = \boldsymbol{f}(\boldsymbol{X}), \qquad \boldsymbol{f} = (f_1, \ldots, f_m)^\top,$$

where $\boldsymbol{f}$ denotes the multivariate measurement function.

Given an estimate $\boldsymbol{x}$ of $\boldsymbol{X}$, the 'plug-in' estimate of $\boldsymbol{Y}$ is

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x}).$$

The covariance matrix of dimension $m \times m$ associated with $\boldsymbol{y}$ is

$$\boldsymbol{V}_y = \begin{bmatrix} u(y_1, y_1) & \cdots & u(y_1, y_m) \\ \vdots & \ddots & \vdots \\ u(y_m, y_1) & \cdots & u(y_m, y_m) \end{bmatrix},$$

where $u(y_i, y_j)$ is the covariance associated with y_i and y_j and $u(y_j, y_j) = u^2(y_j)$, given by [81, page 29]

$$\boldsymbol{V}_y = \boldsymbol{C}_x \boldsymbol{V}_x \boldsymbol{C}_x^\top \tag{E6.1.11}$$

and $\boldsymbol{C}_x$ is the sensitivity matrix given by evaluating

$$\boldsymbol{C}_X = \begin{bmatrix} \dfrac{\partial f_1}{\partial X_1} & \cdots & \dfrac{\partial f_1}{\partial X_N} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial f_m}{\partial X_1} & \cdots & \dfrac{\partial f_m}{\partial X_N} \end{bmatrix}$$

at $\boldsymbol{X} = \boldsymbol{x}$.

E6.1.C Methods for obtaining sensitivity coefficients

E6.1.C.1 Symbolic differentiation

Symbolic differentiation uses algebraic expressions or formulæ (such as in section E6.1.4) for the functions involved. It applies the rules for differentiation together with a database of derivatives of special function, to provide formulæ for the derivatives. Software for symbolic differentiation essentially mimics the steps taken in deriving the formulæ analytically. Further, we checked some of the partial differential calculations using the symbolic algebra package [482].

An advantage is that formulæ are provided that can be directly incorporated in software. Disadvantages are that the formulæ produced can contain unwieldy expressions and there is no guarantee that they are in a numerically stable form for evaluating the derivatives.

E6.1.C.2 Partial derivatives by finite differences

The forward difference approximation to the first derivative of a function f is [484]

$$f'(x) \approx \frac{f(x+h)-f(x)}{h} \tag{E6.1.12}$$

with discretization error $-hf''(\eta)/2$, where h is the step size, for some η in the interval $[x,\ x+h]$. Generally, a better, very popular central difference approximation is [484]

$$f'(x) \approx \frac{f(x+h)-f(x-h)}{2h}$$

with discretization error $-h^2 f'''(\eta)/6$ for some η in the interval $[x,\ x+h]$. h is usually selected such that the sum of the magnitudes of the round-off error and the discretization error is minimized. Its optimal choice in that regard is

$$h = \left(\frac{3\epsilon}{|M|}\right)^{1/3}, \tag{E6.1.13}$$

where ϵ is the machine round-off ($\epsilon \approx 2 \times 10^{-16}$ for many computers) and $M \approx f'''(\eta)$. However, equation (E6.1.13) is primarily of theoretical value and can hardly be used in practice to determine h because to obtain the third derivative or a bound for it is a more demanding task than approximating the first derivative itself. We have found empirically that $h = \epsilon^{1/2}|x|$ seems to be a reasonable choice for the current application.

E6.1.C.3 Partial derivatives by the complex step method

The use of the complex step method in metrology is considered in [40,481] (see section E6.1.C.3), where the original references to the method are also to be found. There also is interest in the aeronautics industry with the method [485,486], where its advantages have been expounded.

The method utilizes a minute imaginary step size in expression (E6.1.12), which means that the rounding error associated with the conventional finite difference formulation is completely avoided and thus the derivative is approximated to very close to full machine precision. A brief description of the method is given. It is capable of delivering first derivatives to a relative numerical accuracy of $\mathcal{O}(\epsilon)$.

Consider the output quantity Y in a univariate measurement model $Y = f(X_1, \ldots, X_N)$ or one component of a multivariate measurement model having input quantities $X_1, \ldots, X_N$, where the measurement function f is given by a mathematical expression or algorithm.

Rather than approximate $f'(x)$ using expression (E6.1.12), it is defined [487] by

$$f'(x) \approx \frac{1}{h}\Im f(\boldsymbol{x} + \mathrm{i}h),$$

where $\Im$ denotes 'imaginary part' and h is a positive value very much smaller than unity.

This method is applicable when the real types in the software that implements the model can be replaced by complex types. Such types are supported by several popular computer languages.

The complex-step method is similar to finite differences but uses complex arithmetic to obtain first derivatives. The Taylor expansion of a function f of a complex variable is

$$f(z + w) = \sum_{r=0}^{\infty} \frac{w^r}{r!} f^{(r)}(z),$$

where z and w are complex. Setting $z = x$ and $w = \mathrm{i}h$, where x is real and h is real, positive and small, and taking real and imaginary parts,

$$\Re f(x + \mathrm{i}h) = f(x) - \frac{h^2}{2} f'''(x) + \mathcal{O}(h^4), \qquad \Im f(x + \mathrm{i}h) = h f'(x) - \frac{h^3}{6} f'''(x) + \mathcal{O}(h^5),$$

from which

$$f(x) = \Re f(x + \mathrm{i}h), \qquad f'(x) = \frac{1}{h}\Im f(x + \mathrm{i}h),$$

with truncation errors of $\mathcal{O}(h^2)$. Unlike the use of a finite-difference formula for $f'(x)$, h is chosen to be *very* small with no concern about the loss of significant digits through subtraction cancellation. In the work of the first author, the method is routinely applied with $h = 10^{-100}$ [488], suitable for all but pathologically-scaled problems. Also see references [40, 489].

A simple example of the use of the ih method is as follows. Consider the first derivative $f'(x)$ of the mathematical function

$$f(x) = \frac{(1 + x)(1 + 2x)}{(1 + 3x)(1 + 4x)}$$

evaluated at $x = 1$. A MATLAB implementation is

```
h = 1e-100;
x = 1 + sqrt(-1)*h;
f = (1 + x)*(1 + 2*x)/((1 + 3*x)*(1 + 4*x));
dfdx = imag(f)/h;
```

The exact value of $f'(1)$ is $-23/200$, which the above code delivers correct to within one binary digit of full machine precision.

E6.1.D The Monte Carlo method for uncertainty evaluation

From the Bayesian perspective, the MCM is the 'gold standard' for uncertainty evaluation. The input quantities in a measurement model are assigned appropriate probability distributions as befits knowledge of those quantities. The method involves evaluating the measurement model very many times for values of the input quantities chosen randomly from their probability distributions, from which the required results are derived. Considerable advantages are that no derivative calculation is required and, with a sufficiently large number of evaluations, any measurement model can be treated no matter how complicated and 'how non-linear' it is. Conversely, the GUM uncertainty framework (GUF) applies only for linear models or models that can safely be linearized.

The Monte Carlo method for uncertainty evaluation can be summarized as follows [3]. Consider a univariate measurement model, or one component of a multivariate measurement model, of the form

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

Assign a probability distribution to each of the input quantities $X_1, \ldots, X_N$. Consider a large number M of Monte Carlo trials. In the rth trial, draw randomly from the probability distribution for each input quantity, and use the measurement model to compute the corresponding value Y_r of Y. The values Y_r, $r = 1, \ldots, M$, constitute a sample from the probability distribution for Y. M should be chosen to be sufficiently large so that a representative sample of the probability distribution of the output quantity Y is obtained. The approach here applies to independent input quantities. For details of its extension to dependent input quantities, see GUM-S1 [3], and for a multivariate output quantity, see GUM-S2 [4].

Example E6.2

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

S. Martens, K. Klauenberg, C. Elster

E6.2.1 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.

E6.2.2 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 [490]. 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 E6.2.1 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 [490].

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 E6.2.1: Summary statistics for part[1] of the measurement data given in the guideline [490][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 [35].

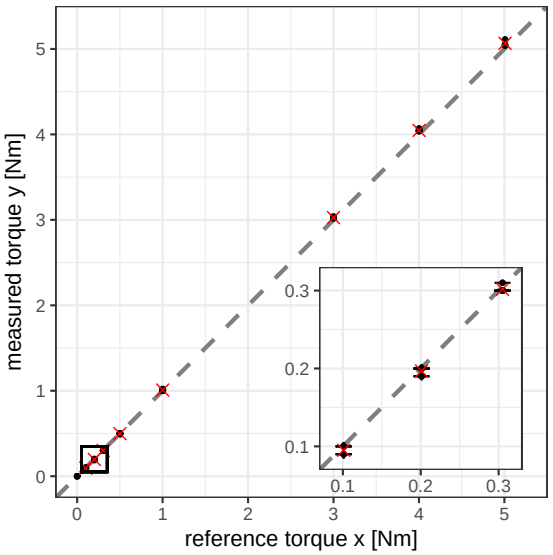


Figure E6.2.1: Visualization of the data. Repeated measurements (dots) are indistinguishable from the mean values (crosses) in this presentation. The dashed line represents the identity $y = x$.

This example provides guidance for the evaluation of uncertainty in the estimation of a calibration curve from data like those in table E6.2.1. Two approaches are provided: (i) ordinary and weighted least-squares estimation (see, e.g., ISO 28037 [77]) accompanied with an uncertainty evaluation based on the GUM [2], and (ii) a statistical approach applying Bayesian inference (cf., for example, [65]). Explicit expressions are given for the Bayesian uncertainty analysis which simplifies its application.

E6.2.3 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

$$Y = \beta X \tag{E6.2.1}$$

is assumed to model the relationship between the measured responses of the considered sensor and the applied stimulus. Model (E6.2.1) represents a straight line with zero offset. The latter has been chosen for physical reasons which is supported by the observed data (cf. figure E6.2.1). The measurand is the slope parameter β of the particular straight line model (E6.2.1). 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 (E6.2.1). To prevent double counting information, we omit the data point $(0,0)$ in the our consideration.

E6.2.4 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.

E6.2.4.1 Ordinary and weighted least-squares

Application of weighted least-squares estimation determines an estimate $\widehat{\beta}$ for the measurand by minimizing

$$Q = \sum_{i=1}^{p} \sum_{j=1}^{n_i} W_i \left(y_{ij} - \beta x_i\right)^2 = \sum_{i=1}^{p} W_i \{n_i (y_i - \beta x_i)^2 + (n_i - 1) S_i^2\} \tag{E6.2.2}$$

with respect to β. In (E6.2.2), 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

$$\widehat{\beta} = \left(\sum_{i=1}^{p} n_i W_i x_i^2 \right)^{-1} \sum_{i=1}^{p} n_i W_i x_i y_i . \tag{E6.2.3}$$

The measurement model will now be *defined* as

$$\beta = \left(\sum_{i=1}^{p} n_i W_i x_i^2 \right)^{-1} \sum_{i=1}^{p} n_i W_i x_i Y_i , \tag{E6.2.4}$$

i.e., by replacing the estimates y_i in (E6.2.3) 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 (E6.2.4) has already been replaced with the known values x_i in this example. Ordinary least-squares estimation is obtained by choosing weights $W_i = 1, i = 1, \ldots, p$.

E6.2.4.2 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.

$$Y_{ij}|\beta,\sigma_i^2 \sim N\left(\beta x_i,\sigma_i^2\right). \tag{E6.2.5}$$

In (E6.2.5), 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 (E6.2.5) the likelihood function is given by

$$l(\beta,\boldsymbol{\sigma}^2;\text{data}) \propto \prod_{i=1}^{p}(\sigma_i^2)^{-n_i/2}\exp\left(-\frac{1}{2\sigma_i^2}\left((n_i-1)S_i^2+n_i(y_i-\beta x_i)^2\right)\right), \tag{E6.2.6}$$

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 E6.2.1.

E6.2.5 Estimation and uncertainty evaluation

E6.2.5.1 GUM uncertainty propagation

The measurement model in (E6.2.4) 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 E6.2.2.

According to the GUM [2], the estimate $\widehat{\beta}$ for β is taken as value of the measurement model (E6.2.4) when inserting the estimates from table E6.2.2 for the input quantities, i.e.

$$\widehat{\beta}=\left(\sum_{i=1}^{p}n_iW_ix_i^2\right)^{-1}\sum_{i=1}^{p}n_iW_ix_iy_i\,. \tag{E6.2.7}$$

The (squared) standard uncertainty is obtained by

$$u^2\left(\widehat{\beta}\right)=\sum_{i=1}^{p}\left(\left.\frac{\partial\beta}{\partial Y_i}\right|_{Y_i=y_i}\right)^2u^2(y_i)=\frac{\sum_{i=1}^{p}(n_iW_ix_i)^2u^2(y_i)}{\left(\sum_{i=1}^{p}n_iW_ix_i^2\right)^2}.$$

If the weights are chosen according to $n_iW_i=1/u^2(y_i)$, one obtains

$$u^2\left(\widehat{\beta}\right)=\left(\sum_{i=1}^{p}\frac{x_i^2}{u(y_i)^2}\right)^{-1}, \tag{E6.2.8}$$

and for ordinary least-squares estimation with $W_i=1$

$$u^2(\widehat{\beta})=\frac{\sum_{i=1}^{p}(n_ix_i)^2u^2(y_i)}{\left(\sum_{i=1}^{p}n_ix_i^2\right)^2}. \tag{E6.2.9}$$

Assuming a Gaussian distribution for β, a 95% coverage interval is given by

$$\left[\widehat{\beta}-1.96u\left(\widehat{\beta}\right),\widehat{\beta}+1.96u\left(\widehat{\beta}\right)\right]. \tag{E6.2.10}$$

Table E6.2.2: For the data in table E6.2.1 and measurement model (E6.2.4), 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

E6.2.5.2 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 (E6.2.5)) conditional on the observed data. In our case, the posterior is given through the following probability density function (PDF)

$$\pi(\beta, \boldsymbol{\sigma}^2|\text{data}) \propto \pi(\beta, \boldsymbol{\sigma}^2) l(\beta, \boldsymbol{\sigma}^2; \text{data}), \tag{E6.2.11}$$

where $l(\beta, \boldsymbol{\sigma}^2; \text{data})$ denotes the likelihood function (E6.2.6) for the assumed statistical model (E6.2.5), and $\pi(\beta, \boldsymbol{\sigma}^2)$ the employed prior for β and $\boldsymbol{\sigma}^2 = (\sigma_1^2, \ldots, \sigma_p^2)^\top$. From the joint posterior (E6.2.11), the marginal posterior $\pi(\beta|\text{data})$ for the measurand is obtained through marginalization according to

$$\pi(\beta|\text{data}) = \int_0^\infty \ldots \int_0^\infty \pi(\beta, \boldsymbol{\sigma}^2|\text{data}) \mathrm{d}\sigma_1^2 \ldots \mathrm{d}\sigma_p^2 . \tag{E6.2.12}$$

The marginal posterior (E6.2.12) 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,

$$\widehat{\beta} = \int_{-\infty}^\infty \pi(\beta|\text{data}) \beta \mathrm{d}\beta \tag{E6.2.13}$$

as the Bayesian estimate, and the posterior standard deviation as the associated standard uncertainty $u(\beta)$, where

$$u^2(\beta) = \int_{-\infty}^\infty \pi(\beta|\text{data})(\beta - \widehat{\beta})^2 \mathrm{d}\beta . \tag{E6.2.14}$$

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 equation (E6.2.14) rather than $u(\widehat{\beta})$. Finally, a 95 % credible interval $[\underline{\beta}, \overline{\beta}]$ can be calculated from the posterior (E6.2.12) which satisfies

$$\int_{\underline{\beta}}^{\overline{\beta}} \pi(\beta|\text{data}) \mathrm{d}\beta = 0.95 . \tag{E6.2.15}$$

Equation (E6.2.15) 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 [3].

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

$$a_i = 2 + \frac{1}{c_i^2}\,, \quad b_i = (a_i - 1)\widehat{\sigma}_i^2\,, \tag{E6.2.16}$$

i.e., the prior knowledge about each σ_i^2 is modelled 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

$$\pi(\beta|\text{data}) \propto \pi(\beta)\prod_{i=1}^{p} t_{n_i-1+2a_i}\left(\beta; y_i/x_i, ((n_i-1)S_i^2 + 2b_i)/(n_i x_i^2[n_i - 1 + 2a_i])\right)\,, \tag{E6.2.17}$$

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.

$$t_\nu(x; m, s^2) \propto \left(1 + \frac{1}{\nu}\frac{(x-m)^2}{s^2}\right)^{-\frac{\nu+1}{2}}\,, \tag{E6.2.18}$$

cf. also [3]. The univariate PDF (E6.2.17) is easily evaluated, and the summary statistics (E6.2.13)–(E6.2.15) can immediately be obtained through standard procedures of numerical quadrature. Note that for evaluating (E6.2.17) 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 (E6.2.17) then approaches

$$\pi(\beta|\text{data}) \propto \prod_{i=1}^{p} t_{n_i-1}(\beta; y_i/x_i, S_i^2/(n_i x_i^2))\,, \tag{E6.2.19}$$

which is also formally obtained when using the following non-informative prior [65]

$$\pi(\beta, \boldsymbol{\sigma}^2) \propto \prod_{i=1}^{p} \frac{1}{\sigma_i^2} \tag{E6.2.20}$$

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]

Table E6.2.3: 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.

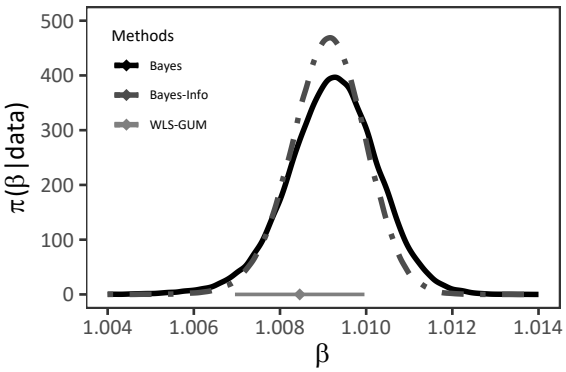


Figure E6.2.2: Marginal posterior distribution $\pi(\beta|\text{data})$ for the measurand according to (E6.2.19) (solid line) and (E6.2.17) (dashed line) using a non-informative or a vague informative prior, respectively. The symbols in light gray display the estimate $\widehat{\beta}$ and 95% coverage interval of the weighted least-squares approach for comparison.

E6.2.6 Reporting the result

Table E6.2.3 contains the estimate, its associated standard uncertainty, and the 95% coverage interval obtained by application of the GUM to ordinary least-squares method (OLS) and weighted least-squares method (WLS) 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 E6.2.2 shows the PDFs for the measurand obtained by Bayesian uncertainty analyses in comparison with the results achieved by WLS estimation with uncertainty evaluated according to the GUM.

E6.2.7 Discussion and recommendation

The results obtained by application of the GUM to OLS and WLS estimation are different. This difference is due to the difference of the corresponding measurement models (E6.2.4) used. Specifically, WLS 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 OLS estimation. In fact, these weights are "optimal" in the sense that they lead to a minimum uncertainty under all measurement models (E6.2.4).

On the other hand, OLS 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 OLS estimator under hypothetical repeated sampling from the statistical model (E6.2.5). From the perspective of the type A evaluation of the GUM, the corresponding uncertainty evaluation can thus be recommended. However, OLS 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 WLS estimate can be also be viewed as a solution to the statistical model (E6.2.5) 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 (E6.2.5) 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 (E6.2.5) 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 (E6.2.5) 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 [490] reports non-vanishing uncertainties for them. The statistical model (E6.2.5) 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

The authors would like to thank Nicolas Fischer (LNE) for reviewing, as well as Paola Pedone, Fabrizio Manta (ACCREDIA) and Alessandro Germak (INRIM) for commenting on the example.

Example E6.3

Calibration and measurement uncertainty in hardness verification

M. Griepentrog, A. Germak, A. Bošnjaković, V. Vedran Karahodžić, F. Manta, P. Pedone, M.G. Cox

Summary

The example refers to the evaluation of the uncertainty in hardness verification considering that repetition of tests at the same location is in general not possible. In order to improve the methods provided by standards for the Vickers Hardness Test, Knoop Hardness Test, Rockwell Hardness Test, Brinell Hardness Test and the Instrumented Indentation Test, a harmonised approach for uncertainty evaluation has been developed for application to all hardness tests. Two methods, using the GUM uncertainty framework (GUF) and Monte Carlo method (MCM), are used for the uncertainty evaluation. Further, proposals are made on how to use measurement uncertainty in hardness testing in conformity assessment.

E6.3.1 Introduction to the application

A statement concerning the evaluation of uncertainty is given in the normative part of all related standards; for the Vickers Hardness Test [491], Knoop Hardness Test [492], Rockwell Hardness Test [493], Brinell Hardness Test [494] and the Instrumented Indentation Test [495] it is stated that a complete evaluation of the uncertainty should be carried out according to the GUM [2].

In general, for hardness tests, there are two possibilities for the evaluation of uncertainty described in the standards:

- indirect calibration based on hardness reference blocks for the overall checking of the machine. In all related standards guidance for the evaluation of uncertainty is given in informative annexes;
- direct calibration based on the evaluation of single parameters of the machine (that is, preliminary test force, total test force, indentation depth, indentor geometry, frame stiffness, etc.). In such cases general guidance is given only in the Instrumented Indentation Test standard.

In hardness measurement, traceability has a four-level structure of the metrological chain necessary to define and disseminate hardness scales. The chain starts at the international level using international definitions of the various hardness scales to carry out international intercomparisons. A number of primary hardness standard machines at the national level "produce" primary hardness reference blocks for the calibration laboratory level. This third level provides calibrated hardness reference blocks that are used at the fourth level to calibrate the hardness testing machines.

For all hardness tests specific hardness reference blocks for indirect calibration of hardness testing machines are available and guidelines for uncertainty evaluation related to indirect calibration traceability chain are given as informative annexes in all hardness standards.

Analysing the content of the Annexes in the standards it was found that all guidance is comparable in general, but there are significant differences in detail making understanding, application and comparison for users of the standards difficult. The main differences were found in the wording and the symbols used and for calculation and considering the uncertainty of the testing machine, as well as the treatment of the bias.

A proposal for a standard procedure for the evaluation of uncertainty applicable to all five hardness tests is presented in this example.

The approach for evaluating uncertainty presented in the Annexes to the standards considers only those uncertainties associated with the overall measurement performance of the hardness testing machine with respect to the hardness reference blocks (designated as certified reference material (CRM)). These performance uncertainties reflect the combined effect of all the separate uncertainties (indirect verification).

The proposed approach, applicable to indirect hardness verification methods, has the aim to harmonize and improve the already existing procedures described in the standards.

The harmonized approach begins with the calibration of the reference block and the testing machine in order to evaluate the uncertainty of a hardness measurement carried out on a test piece. In addition, the proposed approach can be used for uncertainty evaluation when a statement of conformity to specification is required.

In general in metrology, measurement uncertainty is understand as the dispersion of the values that could reasonably be attributed to the measurand [2,89]. For hardness testing, the measurand is the hardness at a point of the test piece surface obtained by applying a standardized method with the dispersion relating to measurements repeated in the same way and in the same circumstances. The uncertainty is given as a standard deviation assuming that the measured hardness values follow an assumed probability distribution. Usually a normal distribution is used for this purpose.

In hardness testing it is not possible to repeat the test at the same place. This inability could lead to a, mostly unknown, possibly significant contribution to the statistical dispersion. The main sources of this additional contribution are (i) local changes of material properties and (ii) change of test machine performance because of sample movement.

In daily laboratory practice, the quantitative assessment of the contributions of these sources on a single test is not possible, at least because local changes in hardness are often the object of the investigation.

The Student's t-test is a statistical test to determine, to a reasonable degree of confidence, whether two sets of observations come from different populations. Using the t-test for analysis of the measured hardness values it is possible to decide whether the difference in measured values from sample test sets carried out at different locations describes a real change in hardness.

The main assumption is, if the analysis using the t-test gives no evidence, that sample sets carried out at different locations describing a real hardness change, the influence of the non-repeatability of hardness tests at the same place on the evaluation of uncertainty can be neglected.

Example: two sets of Instrumented Indentation Hardness measurements

Two sets of Instrumented Indentation Hardness measurements A and B with $N_\mathrm{A} = 6$ and $N_\mathrm{B} = 7$ tests in the sets were carried out at different locations in a test piece. Calculated values for the mean hardness values $\overline{H}_{\mathrm{IT_A}}$, $\overline{H}_{\mathrm{IT_B}}$ and standard deviations $s_\mathrm{A}, s_\mathrm{B}$ were

$$\overline{H}_{\mathrm{IT_A}} = 313.0\,\mathrm{N\,mm^{-2}}, \qquad \overline{H}_{\mathrm{IT_B}} = 338.6\,\mathrm{N\,mm^{-2}}, \qquad s_\mathrm{A} = 28.2\,\mathrm{N\,mm^{-2}}, \qquad s_\mathrm{B} = 29.1\,\mathrm{N\,mm^{-2}}.$$

The according t statistic is

$$t = \frac{\overline{H}_{IT_A} - \overline{H}_{IT_B}}{\sqrt{\frac{s_A^2}{N_A} - \frac{s_B^2}{N_B}}} = 1.58. \tag{E6.3.1}$$

The appropriate reference distribution for the t statistic is the t distribution. For a degrees of freedom $N_\mathrm{A} + N_\mathrm{B} - 2 = 11$ and an expected confidence level of 95 %, the critical t-value is $t_\mathrm{crit} = 1.80$. Since $t \leq t_\mathrm{crit}$ we can say with 95 % confidence that the analysis using the t test gives no evidence that the sample sets carried out at different locations describe a real change in hardness. Hence, the influence of the non-repeatability of hardness tests at the same place on the evaluation of uncertainty can be neglected, certainly for the data set under consideration.

For the Rockwell Hardness Test, guidance for the evaluation of the uncertainty, based on the evaluation of all relevant sources appearing during a direct calibration, is available [496]. In the case of this test, a mathematical relationship connecting measured quantities with hardness is not known. The relationship between measured values and the estimated hardness value is given by the scale definitions. Using the appropriate sensitivity coefficients, namely, the partial derivatives of the dependent variable hardness in terms of the independent variables, the GUM [2] can be applied to propagate standard uncertainties in the case of uncorrelated input quantities. In the absence of a mathematical formula, the sensitivity factors are estimated experimentally.

As far as possible, the software packages delivered with the testing machines offered by the main suppliers have been checked. Only one supplier offers a procedure based on the evaluation of relevant sources appearing during direct calibration. Evaluating this procedure, a harmonized and general procedure was developed, This procedure combines uncertainty from the following contributions:

- displacement and force;
- maximum sensed displacement;
- maximum displacement;
- frame compliance;
- thermal drift;
- support spring stiffness;
- maximum force;
- stiffness;
- contact depth;
- contact area.

The advantages of the evaluation of uncertainty based on all relevant sources appearing during a direct calibration are:

– the evaluation of uncertainty can be made without any reference material (hardness block);

– knowing the sensitivity factors for individual sources of uncertainty it is possible to evaluate their contribution to the combined uncertainty;

– the uncertainty contributions may be grouped into those corresponding to random and systematic effects.

E6.3.2 Specification of the measurand

The measurand for all indentation hardness tests is hardness defined as the ratio of applied force to contact area for a given point on the surface of the test piece under prescribed conditions. Depending on the test, the contact area is a defined function of one or two characteristic dimensions.

There are several standardized methods including the Rockwell, Vickers, Brinell, Knoop and Instrumented Indentation Test. For the first four of these methods, the size of the indentation is obtained after the force application process. For the Instrumented Indentation Test, indentation size is obtained during the force application process.

The size of the indentation is determined by measuring the average length of the diagonals of the indentation (Vickers), the longest diagonal (Knoop), the average diameter (Brinell) or the depth (Rockwell, Instrumented Indentation Test).

E6.3.3 Measurement model

The characteristic dimensions mentioned in section E6.3.2 together with the applied force are the measured quantities.

E6.3.3.1 Vickers hardness test

In the Vickers hardness test, the indentor is made from diamond in the form of a pyramid. The measurand is HV. The quantities measured are the applied test force F and the length d of the diagonal of the indentation. The measurement model is

$$\mathrm{HV} = 0.1891\frac{F}{d^2}. \tag{E6.3.2}$$

E6.3.3.2 Knoop hardness test

The Knoop hardness test also uses a diamond indentor in the form of a pyramid. The measurand is HK. The quantities measured are the same as for the Vickers hardness test. The measurement model is

$$\mathrm{HK} = 1.451\frac{F}{d^2}. \tag{E6.3.3}$$

E6.3.3.3 Rockwell hardness test

In the Rockwell hardness test, the indentor has a specific size and shape. The measurand is HR. The quantities measured are the applied test force F and the indentation depth h. The measurement model, where N and S are given constants, is

$$\mathrm{HR} = N - \frac{h}{S}. \tag{E6.3.4}$$

E6.3.3.4 Brinell hardness test

In a Brinell hardness test, the indentor takes the form of a ball with a radius D. The measurand is HBW. The quantities measured are the applied test force F and the diameter d of the indentation. The measurement model is

$$\mathrm{HBW} = \frac{0.102}{\pi} \frac{2F}{D^2\left(1 - \sqrt{1 - \frac{d^2}{D^2}}\right)}. \tag{E6.3.5}$$

E6.3.3.5 Instrumented Indentation hardness test

In an Instrumented Indentation hardness test, a diamond indentor in the form of a pyramid (as for the Vickers hardness test) is used. The measurand is $\mathrm{H_{IT}}$. The quantities continuously measured during force increasing (loading) and force decreasing (unloading) are the applied test force F and the indentation depth h. The measurement model is

$$\mathrm{H_{IT}} = \frac{F_{\max}}{A_{\mathrm{p}}(h_{\mathrm{c}})}. \tag{E6.3.6}$$

In equation (E6.3.6),

$$A_{\mathrm{p}}(h_{\mathrm{c}}) = 24.50 h_{\mathrm{c}}^2$$

is the projected (cross-sectional) area of contact between the indentor and the test piece calculated for the depth of contact h_{c} at maximum applied test force $F_{\max}$, and

$$h_{\mathrm{c}} = h_{\max} - \frac{F_{\max}}{S}. \tag{E6.3.7}$$

where $h_{\max}$ is the maximum indentation depth and S is the slope (stiffness) of the unloading curve.

E6.3.4 Uncertainty evaluation

This document discusses a general procedure of uncertainty evaluation based on indirect calibration using a hardness reference block for the overall checking of the machine. The procedure is applicable to all related standards. The metrological traceability chain of hardness measurement, as described in clause 1 of the standards, starts from a reference material (RM) or CRM, usually termed a hardness block, as described in clause 3 of the standards, and passed through a calibrated Hardness Testing Machine (HTM), as described in clause 2 of the standards.

This procedure calculates an expanded uncertainty U (at the 95 % level of confidence) associated with the measured hardness value X of the test piece.

The general expression for the expanded uncertainty is

$$U = k\left(u_{\mathrm{HTM}}^2 + u_{\mathrm{x}}^2 + u_{\mathrm{msx}}^2\right)^{1/2}, \tag{E6.3.8}$$

where k is the coverage factor for a confidence level of 95 %. The contributions to the standard measurement uncertainty u associated with X are

- u_{HTM}, the contribution generated by the hardness testing machine;
- u_{x}, the contribution due to the combination of the lack of measurement repeatability of the hardness testing machine and the hardness non-uniformity of the sample under investigation;
- u_{msx}, the contribution due to the resolution of the hardness testing machine when measuring a characteristic dimension of the indentation for the sample under investigation.

The effective degrees of freedom ν associated with u is calculated with the Welch-Satterthwaite formula [2]. The coverage factor k is taken as 2 when ν exceeds 30. Otherwise, k is calculated using the t distribution [2, annex G].

The treatment of uncertainty made in the annexes of the standards as well as in EURAMET Cg-16 [496] implicitly assumes all contributions are uncorrelated. The same assumption is adopted in this proposed procedure for indirect traceability.

E6.3.4.1 Standard uncertainty due to the hardness testing machine

The standard uncertainty u_{HTM} due to the hardness testing machine (HTM) is given by

$$u_{\mathrm{HTM}}^2 = u_{\mathrm{CRM}}^2 + u_{\mathrm{HCRM}}^2 + u_{\mathrm{CRM-D}}^2 + u_{\mathrm{ms}}^2. \tag{E6.3.9}$$

$u_{\mathrm{CRM}} = U_{\mathrm{CRM}}/2$ is the standard uncertainty corresponding to the calibration expanded uncertainty U_{CRM} associated with the certified hardness values H_{CRM} of the CRM, according to the calibration certificate.

u_{HCRM} is the standard uncertainty due to the combination of the lack of measurement repeatability of the HTM and the hardness non-uniformity of the CRM when measuring the CRM. Note that the non-uniformity of the CRM is already included in u_{CRM}; consequently, the contribution of non-uniformity of the CRM is doubly counted. It would be desirable to have separate information about this contribution in the CRM calibration standard uncertainty u_{CRM} in order to avoid this double counting.

$u_{\mathrm{CRM-D}}$ is the standard uncertainty due to the drift of values certified for the CRM. Usually, the estimate of the contribution is assumed to be 0.

u_{ms} is the standard uncertainty due to the resolution of the hardness testing machine when measuring the characteristic dimension of the residual indentation in the CRM.

Standard uncertainty due to calibration

From the results of n hardness measurements H_i, $(i = 1, \ldots, n)$ on the CRM, the mean value is

$$\overline{H} = \frac{1}{n}\sum_{i=1}^{n} H_i. \tag{E6.3.10}$$

Then, the standard deviation s_{HCRM} and the standard uncertainty u_{HCRM} are calculated from

$$s^2_{\mathrm{HCRM}} = \frac{1}{n-1}\sum_{i=1}^{n}(H_i - \overline{H})^2, \tag{E6.3.11}$$

$$u_{\mathrm{HCRM}} = \frac{s_{\mathrm{HCRM}}}{n^{1/2}}. \tag{E6.3.12}$$

In the informative annexes "Uncertainty of measured hardness values" of all mentioned standards, a Student t factor of 1.14 is incorporated in the right-hand side of formula (E6.3.12). This factor corresponds to a confidence level of 68.27 % and 4 degrees of freedom. This approach is used in the standards in order to be combined with other contributions to the uncertainty (i) avoiding the calculation of the effective degrees of freedom with the Welch-Satterthwaite formula and (ii) enabling a coverage factor $k = 2$ to be used for evaluating the expanded uncertainty.

This procedure is *not* GUM-compliant and is not adopted here. It leads to the same expanded uncertainty as delivered by the GUM uncertainty framework (GUF) only in the case where the contribution u_{HCRM} is dominated by the remaining contributions. Otherwise, its use leads to an overstatement of the uncertainty.

Standard uncertainty due to resolution of measurement of characteristic distance

All hardness tests are based on the measurement of some characteristic distance d (length of diagonal, diameter, indentation depth), to be able to calculate some characteristic area, where d in this section is defined differently from elsewhere. For all indentation tests the hardness value H is a function of d:

$$H = H(d). \tag{E6.3.13}$$

The resolution for the measurement of d is denoted by δ_{ms} for which a rectangular distribution is assumed. So,

$$u_{\mathrm{ms}} = C\frac{\delta_{\mathrm{ms}}}{2\sqrt{3}} \tag{E6.3.14}$$

with sensitivity coefficient

$$C = \frac{\partial H}{\partial d}. \tag{E6.3.15}$$

When considering the measurement models for the various hardness measurands, sensitivity coefficients must be calculated. Table E6.3.1 shows the sensitivity coefficients for the various hardness tests calculated in terms of measured values of the input quantities only and also in terms of the measurand value.

NOTE: If the measurement of the distance is obtained from a difference (or sum) of two single measurements, like for Vickers measurements, the effect of resolution shall be considered twice.

E6.3.4.2 Standard uncertainty due to repeated measurements on the sample

From the result of n measurements X_{i}, $i = 1, \ldots, n$, on the sample under investigation the mean value is

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i. \tag{E6.3.16}$$

Table E6.3.1: Sensitivity coefficients for the various hardness tests

	Sensitivity coefficient C	
Hardness test	Measured values known only	Measurand value also known
Vickers	$-\frac{2\times 0.1891F}{d^3}$	$-\frac{2}{d}\mathrm{HV}$
Knoop	$\frac{2\times 1.451F}{d^3}$	$-\frac{2}{d}\mathrm{HK}$
Rockwell	$-\frac{1}{S}$	
Brinell	$-\frac{0{,}102}{\pi}\frac{2Fd}{D^4\sqrt{1-\frac{d^2}{D^2}}\left(1-\sqrt{1-\frac{d^2}{D^2}}\right)^2}$	$-\mathrm{HBW}\frac{d}{D^2\sqrt{1-\frac{d^2}{D^2}}}$
Instrumented indentation	$-\frac{2F_{\max}}{24.50h_c^3}$	$-\frac{2}{h_c}\mathrm{H_{IT}}$

Then, the standard deviation s_X and the standard uncertainty u_X are given by

$$s_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2. \tag{E6.3.17}$$

$$u_X = \frac{s_X}{n^{1/2}}. \tag{E6.3.18}$$

E6.3.4.3 Standard uncertainty due to resolution of the hardness testing machine

The standard uncertainty due to resolution of the hardness testing machine is similar to that, u_{ms}, of the measurement of characteristic distance in section E6.3.4.1:

$$u_{msx} = C\frac{\delta_{ms}}{2\sqrt{3}}. \tag{E6.3.19}$$

$$C = \frac{\partial H}{\partial d}. \tag{E6.3.20}$$

E6.3.4.4 Bias of the testing machine

The bias of the HTM under the particular verification conditions is

$$b = \overline{H} - H_{CRM}. \tag{E6.3.21}$$

Because the usual calibration of the HTM is performed at several hardness values, q in number, using different CRMs, there are different values for the bias referred to the specific hardness values for the same HTM. The bias can be computed as the actual bias for each specific hardness value. Alternatively, the bias $\overline{b}$ can be calculated as the arithmetic mean of the q biases $b_1, \ldots, b_q$:

$$\overline{b} = \frac{1}{q}\sum_{j=1}^{q} b_j. \tag{E6.3.22}$$

Further, the standard deviation s_b and the standard uncertainty u_b are formed using

$$s_b^2 = \frac{1}{q-1}\sum_{j=1}^{q}(b_j - \overline{b})^2, \tag{E6.3.23}$$

$$u_b = \frac{s_b}{q^{1/2}}. \tag{E6.3.24}$$

Other methods can be used to take into consideration the systematic effect of the HTM, like the correction function obtained by a linear regression of the biases at the specific hardness values. In such a case, the bias is given by the correction function and its associated standard uncertainty can be calculated from the regression results.

E6.3.4.5 Application of the Monte Carlo method

The above analysis method requires the evaluation of the standard uncertainty for each input quantity and its influence on the combined standard uncertainty through a sensitivity analysis. To provide the sensitivity coefficients, either first-order partial derivatives of the output quantity with respect to each input quantity need to be calculated analytically or estimated experimentally.

The glsmcm is an alternative approach for obtaining the required results that avoid the sometimes-laborious sensitivity calculations. In a specific hardness test, a limited amount of data is gathered that is used in conjunction with the appropriate measurement model. Using the MCM the mean and the uncertainty of the estimated hardness value can be calculated using the measured mean and standard deviations of the input quantities. A general treatment is given in [3]. MCM is considered below for the Vickers' Hardness Test. Other tests are using MCM are conducted analogously.

Example: GUM and Monte Carlo method for the evaluation of uncertainty for the Vickers' Hardness Test

The GUM's LPU and the MCM for the evaluation of uncertainty were both applied to data from the Vickers' Hardness Test using the same measurement model in each case. It is assumed that force and diagonal measurement are the most significant sources of uncertainty. Further, it is assumed that the input quantities force and displacement are not correlated and, for MCM, the mean values of the input quantities follow normal distributions.

For both approaches, means and standard uncertainties are based on 20 repeated measurements during direct calibration. 2×10^5 Monte Carlo trials were carried out and for each trial a sample of the input quantities from their respective PDFs was drawn. The according value of the measurand is calculated from the measurement model. Table E6.3.2 summarizes the results obtained.

Table E6.3.2: GUM and MCM estimates and standard uncertainties for Vickers hardness test

	Force/N	Displacement/mm	Vickers hardness/HV
GUM estimate	294.18	0.3730	401.40
GUM standard uncertainty	1.18	0.0005	1.10
MC estimate	294.22	0.3729	399.82
MC standard uncertainty	1.21	0.0005	1.06

Taking the respective standard uncertainties in table E6.3.2 into consideration, there is no significant difference between the GUM and MC results are. The implication is that the GUM approach is satisfactory for this data set and other data sets that are sufficiently similar.

E6.3.4.6 Uncertainty evaluation based on all relevant sources

The evaluation of uncertainty based on all relevant sources appearing during a direct calibration assumes that the uncertainty contributions from all source are known.

If the quantities measured in indentation hardness tests for estimation of a hardness value H are force F and some characteristic geometric dimension d, a general measurement model is

$$H = f(F, d). \tag{E6.3.25}$$

In applying the law of propagation of uncertainty [2], the standard uncertainty u_H associated with an estimate of H can be calculated knowing the standard uncertainties u_F and u_D and covariance $u_{F,d}$:

$$u_H^2 = \left(\frac{\partial H}{\partial F}\right)^2 u_F^2 + \left(\frac{\partial H}{\partial d}\right)^2 u_d^2 + 2\frac{\partial H}{\partial F}\frac{\partial H}{\partial d}u_{F,d}. \tag{E6.3.26}$$

The covariance $u_{F,d}$ is the uncertainty contribution due to possible correlation between the measured values of F and d, with $\overline{F}$ and $\overline{d}$ denoting the mean values of n simultaneous measurements:

$$u_{F,d} = \frac{1}{n(n-1)}\sum_{i=1}^{n}(F_i - \overline{F})(d_i - \overline{d}). \tag{E6.3.27}$$

The treatment of uncertainty in the annexes of the standards as well as in EURAMET cg-16 [496] implicitly assumes all contributions to be uncorrelated. The same assumption is adopted in all proposed procedures. This assumption can be verified with the following considerations.

Usually, the correlation coefficient r is calculated as

$$r(F, d) = \frac{u_{F,d}}{u_F\, u_d}. \tag{E6.3.28}$$

Using the correlation coefficient, the law of propagation (E6.3.26) can be written as

$$u_H^2 = \left(\frac{\partial H}{\partial F}\right)^2 u_F^2 + \left(\frac{\partial H}{\partial d}\right)^2 u_d^2 + 2\frac{\partial H}{\partial F}\frac{\partial H}{\partial d}u_F u_h r(F, d). \tag{E6.3.29}$$

If $|r| \ll 1$, the input quantities F and d can be assumed to be uncorrelated.

Example: Test of correlation for data from an Instrumented Indentation Hardness tester

With an Instrumented Indentation Hardness tester, $n = 20$ tests on a sample were conducted. The quantities maximum force $F_{\max}$ and maximum displacement $h_{\max}$ were measured simultaneously for every test. From these quantities the mean values $\overline{F}_{\max} = 20.130\,\mathrm{mN}$ and $\overline{h}_{\max} = 1610.1\,\mathrm{nm}$ and the corresponding standard uncertainties $u_{\mathrm{F_{max}}} = 1.89\,\mu\mathrm{N}$ and $u_{\mathrm{h_{max}}} = 51.5\,\mathrm{nm}$ were calculated. Applying equation (E6.3.28), $r = 0,0021$. For this example, the input quantities can be assumed to be uncorrelated.

E6.3.5 Reporting the result

In section E6.3.4, the method for evaluating the uncertainty of hardness measurement was presented. The measurement is made using a calibrated hardness testing machine whose uncertainty and error, called bias, have been evaluated. Depending on how the bias is considered, the uncertainty on the hardness measurement must be treated accordingly.

If the bias is used to correct for systematic effects in case the actual bias b was calculated, the confidence interval is expressed as

$$(\overline{X} - b) \pm U. \tag{E6.3.30}$$

When the average bias $\overline{b}$ is calculated, the confidence interval is expressed as

$$(\overline{X} - \overline{b}) \pm U_{\text{corr}}. \tag{E6.3.31}$$

where U_{corr} is calculated considering the contribution due to the calculation of the mean bias $\overline{b}$:

$$U_{\text{corr}} = k\left(u_{\text{HTM}}^2 + u_{\text{X}}^2 + u_b^2 + u_{\text{msx}}^2\right)^{1/2} \tag{E6.3.32}$$

with k the coverage factor for a confidence level of 95 %.

If the actual bias b is used to compute the uncertainty, the confidence interval is expressed as

$$\overline{X} \pm (U + |b|), \tag{E6.3.33}$$

When the average bias $\overline{b}$ is calculated, the confidence interval is expressed as

$$\overline{X} \pm (U_{\text{corr}} + |\overline{b}|), \tag{E6.3.34}$$

E6.3.6 Examples of calculation

E6.3.6.1 General

Examples of hardness estimation and the evaluation of the associated uncertainty are carried out in sections E6.3.6.2 to E6.3.6.7 using the harmonized procedure summarized in section E6.3.4.6. For the determination of the coverage factor k (usually for a confidence level of 95 %), the effective degrees of freedom ν_{eff} must be calculated considering the actual number of direct measurements of the measured quantities. Note that in the Vickers Hardness test the diagonal length used in the measurement model is the mean of the measurement of two diagonals. The force is measured directly only in the Instrumented Indentation Test. Because in all other tests force is only controlled to be in the required limits of the direct force calibration, the uncertainty and the degrees of freedom for the direct calibration must be considered. Further uncertainties and degrees of freedom for the calibration of the CRM and for the resolution of the dimension measuring device must be considered.

In each example, the data relates to the construction and properties of the testing machine used.

Table E6.3.3: Rockwell Hardness values calculated from five tests made on a CRM

Test number	Hardness value H/HRC
1	45.2
2	45.4
3	45.6
4	45.4
5	45.8
Mean value $\overline{H}$	45.48
Standard deviation s_H	0.23

E6.3.6.2 Rockwell calculation

Using a Rockwell Hardness Testing machine ($\delta_{\text{ms}} = 0.2$ HRC) calibrated according to ISO 6508 [493], five tests on a CRM ($H_{\text{CRM}} = 45.1$ HRC, $U_{\text{CRM}} = 0.3$ HRC) were conducted with a maximum applied force of 1.47 kN. The quantities measured are the applied test force F and indentation depth h under the final test force. Applying the measurement model (E6.3.4), the Rockwell hardness values for all tests were calculated (table E6.3.3).

Applying equations (E6.3.14) and (E6.3.15), $u_{\text{ms}} = 0.06$ HRC. Using the mean value $\overline{H} = 45.48$ HRC and the standard deviation $s_{\text{H}} = 0.23$ HRC, the contribution $u_{\text{HCRM}} = 0{,}10$, HRC was calculated using equation (E6.3.12). The contribution $u_{\text{HTM}} = 0.27$ HRC was finally calculated using equation (E6.3.9).

Using the same Rockwell Hardness Testing machine, five tests on the sample under investigation were conducted. Applying the measurement model (E6.3.4), the Rockwell hardness values for all tests were calculated (table E6.3.4).

Table E6.3.4: Rockwell Hardness values calculated from five tests made on the sample under investigation

Test number	Hardness value H/ HRC
1	42.4
2	42.6
3	42.0
4	42.8
5	42.2
Mean value $\overline{X}$	42.40
Standard deviation s_X	0.32

Applying equations (E6.3.19) and (E6.3.20), $u_{\text{msx}} = 0.06$ HRC. Using the mean value $\overline{X} = 42.4$ HRC and the standard deviation $s_X = 0.32$ HRC, the contribution $u_X = 0.29$ HRC was calculated using equation (E6.3.18). Then, applying equation (E6.3.8), $u = 0.31$ HRC with expanded uncertainty $U = ku = 0.64$ HRC ($k = 2.06$, $\nu_{\text{eff}} = 25$, confidence level 95 %).

E6.3.6.3 Instrumented Indentation Test calculation

Using an Instrumented Indentation Hardness tester ($\delta_{ms} = 1$ nm) calibrated according ISO 14577-2 [495], five tests on a CRM ($H_{CRM} = 9.4$ GPa, $U_{CRM} = 0.25$ GPa) were conducted with a maximum applied force of 300 mN. The quantities measured are the maximum test force F_{max} and the maximum indentation depth h_{max} under maximum force. Applying the measurement model (E6.3.6) the Indentation Hardness values H_{IT} for all tests were calculated (table E6.3.5).

Table E6.3.5: Indentation Hardness values calculated from five tests made on a CRM

Test number	Hardness value H GPa
1	9.385
2	9.619
3	9.446
4	9.204
5	9.447
Mean value $\overline{H}$	9.240
Standard deviation s_H	0.149

Applying equation (E6.3.14), $u_{ms} = -0.011$ GPa. Using the mean value $\overline{H} = 9.420$ and the standard deviation $s_H = 0.149$ GPa from table E6.3.4, $u_{HCRM} = 0.067$ GPa was calculated. The contribution $u_{HTM} = 0.147$ GPa was finally calculated applying equation (E6.3.9).

Using the same Instrumented Indentation Testing machine, five tests on the sample under investigation were conducted. Applying the measurement model (E6.3.6), the Indentation Hardness values H_{IT} for all tests were calculated.

Table E6.3.6: Indentation Hardness values calculated from five tests made on the sample under investigation

Test number	Hardness value H/GPa
1	1.533
2	1.409
3	1.419
4	1.328
5	1.272
Mean value $\overline{X}$	1.392
Standard deviation s_x	0.099

Applying equation (E6.3.14), $u_{ms} = -0.05$ GPa. Using the mean value $\overline{X} = 1.392$ GPa and the standard deviation $s_x = 0.099$ GPa from table E6.3.5, $u_x = 0.044$ GPa. Finally, applying equation (E6.3.8), $u = 0.156$ GPa and the appropriate expanded uncertainty $U = ku = 0.32$ GPa ($k = 2.06$, $\nu_{eff} = 25$, confidence level 95 %).

E6.3.6.4 Vickers calculation

Using a Vickers Hardness Testing machine ($\delta_{ms} = 0.1\,\mu m$, calibrated according to ISO 6507 [491], five tests on a CRM ($H_{CRM} = 400\,HV$, $U_{CRM} = 5\,HV$) were conducted. The quantities measured are the applied test force $F = 292.4\,N$ and the mean length d of two measured diagonals after removal of the test force: see table E6.3.7. Applying the measurement model (E6.3.2), the Vickers hardness values for all tests were calculated.

Table E6.3.7: Vickers hardness values calculated from five tests made on a CRM

Test number	Mean indentation diagonal d/mm	Hardness value H/HV
1	0.3716	402.9
2	0.3724	401.1
3	0.3728	400.3
4	0.3719	402.2
5	0.3722	401.5
Mean value $\overline{H}$	0.3722	401.6
Standard deviation s_H	0.000 46	1.00

Applying equations (E6.3.14) and (E6.3.15), $u_{ms} = 0.06\,HV$. Using the mean value $\overline{H} = 401.6\,HV$ and the standard deviation $s_H = 1.00\,HV$, the contribution $u_{HCRM} = 0.45\,HV$ was determined using equation (E6.3.12). The contribution $u_{HTM} = 2.54\,HV$ was finally calculated using equation (E6.3.9).

Using the same Vickers Hardness Testing machine, five tests on the sample under investigation were conducted. Applying the measurement model (E6.3.2), the Vickers hardness values for all tests were calculated.

Table E6.3.8: Vickers hardness values calculated from five tests made on the sample under investigation

Test number	Mean indentation diagonal d/mm	Hardness value H/HV
1	0.3746	398.6
2	0.3731	399.6
3	0.3713	401.4
4	0.3725	400.9
5	0.3731	399.6
Mean value $\overline{X}$	0.3729	400.0
Standard deviation s_x	0.0119	1.12

Applying equations (E6.3.19) and (E6.3.20), $u_{msx} = 0.06\,HV$ was calculated. Using the mean value $\overline{X} = 400.0\,HV$ and the standard deviation $s_X = 1.12\,HV$, the contribution $u_{\overline{X}} = 0.5\,HV$ was calculated using equation (E6.3.18). Finally applying equation (E6.3.8), $u = 2.6\,HV$ and the associated expanded combined uncertainty $U = ku = 5.3\,HV$ ($k = 2.04$, $\nu_{eff} = 33$, confidence level 95 %).

E6.3.6.5 Knopp calculation

Using a Knopp Hardness Testing machine ($\delta_{ms} = 0.5\,\mu m$), calibrated according to ISO 4545 part 2 [492], five tests on a CRM ($H_{CRM} = 802.7\,HK1$, $U_{CRM} = 12\,HK1$) were conducted. The quantities measured are the applied test force $F = 9.807\,N$ and the length d of the larger diagonal of the indentation after removal of the test force: table E6.3.9. Applying the measurement model (E6.3.4), the Knopp hardness values for all tests were calculated.

Table E6.3.9: Knopp hardness values calculated from five tests made on a CRM

Test number	Measured indentation diagonal d/mm	Hardness value H/HK
1	0.1332	802.0
2	0.1333	800.8
3	0.1335	798.4
4	0.1330	804.4
5	0.1331	803.2
Mean value $\overline{H}$	0.1332	801.76
Standard deviation s_H	0.000 19	2.41

Applying equations (E6.3.14) and (E6.3.15), $u_{ms} = 1.77\,HK$. Using the mean value $\overline{H} = 801.76\,HK$ and the standard deviation $s_H = 2.41\,HK$ from table E6.3.8, the contribution $u_{HCRM} = 1.08\,HK$ was determined. The contribution $u_{HTM} = 6.25\,HK$ was finally calculated.

Using the same Knopp Hardness Testing machine, five tests on the sample under investigation were conducted. Applying the measurement model (E6.3.3), the Knopp hardness values for all tests were calculated: table E6.3.10.

Table E6.3.10: Knopp Hardness values calculated from five tests made on the sample under investigation

Test number	Measured indentation diagonal d/mm	Hardness value H/HK
1	0.1881	402.2
2	0.1876	404.3
3	0.1882	401.7
4	0.1885	400.5
5	0.1876	404.3
Mean value $\overline{X}$	0.1880	402.6
Standard deviation s_x	0.000 39	1.67

Applying equation (E6.3.19), $u_{msx} = 1.88\,HK$. Using the mean value $\overline{X} = 402.6\,HK$ and the standard deviation $s_x = 1.67\,HK$ from table E6.3.9, the contribution $u_X = 0.75\,HK$ was calculated using equation (E6.3.18). Finally applying equation (E6.3.8), $u = 6.32\,HK$ and the associated expanded combined uncertainty $U = ku = 13.01\,HK$ ($k = 2.06$, $\nu_{eff} = 25$, confidence level 95 %) were calculated.

E6.3.6.6 Brinell calculation

Using a Brinell Hardness Testing machine ($\delta_{ms} = 0.5\,\mu m$), calibrated according to ISO 6507 [491], five tests on a CRM ($H_{CRM} = 597.1$ HBW 2.5/187.5, $U_{CRM} = 3.6$ HBW 2.5/187.5) were conducted. The quantities measured are the applied test force F and the diameter d of the indentation after removal of the test force $F = 1839$ N. Applying the measurement model (E6.3.5) with $d = 2.5$ mm, the Brinell hardness values for all tests were calculated: table E6.3.11.

Table E6.3.11: Brinell Hardness values calculated from five tests made on a CRM

Test number	Measured indentation diameter d/mm	Hardness value H/HBW
1	0.6305	591.4
2	0.63	592.3
3	0.6295	593.3
4	0.6297	592.9
5	0.6295	593.3
Mean value $\overline{H}$	0.6298	592.6
Standard deviation s_H	0.000 42	0.80

Applying equations (E6.3.14) and (E6.3.15), $u_{ms} = -0.28$ HBW. Using the mean value $\overline{H} = 592.6$ HBW and the standard deviation $s_H = 0.8$ HBW from table E6.3.10, the contribution $u_{HCRM} = 0.36$ HBW was determined. The contribution $u_{HTM} = 1.86$ HBW was finally calculated using equation (E6.3.9).

Using the same Brinell Hardness Testing machine, five tests on the sample under investigation were conducted. Applying the measurement model (E6.3.5) the Brinell hardness values for every test were calculated: table E6.3.12.

Table E6.3.12: Brinell Hardness values calculated from five test made on the sample under investigation

Test number	Measured indentation diameter d/mm	Hardness value H/HBW
1	0.6304	591.0
2	0.6301	591.6
3	0.6294	592.9
4	0.6296	592.5
5	0.6297	592.3
Mean value $\overline{X}$	0.6298	592.1
Standard deviation s_x	0.000 40	0.76

Applying equation (E6.3.19), $u_{msx} = -0.28$ HBW. Using the mean value $\overline{X} = 592.1$ HBW and the standard deviation $s_X = 0.76$ HBW, the contribution $u_X = 0.34$ HBW was calculated using equation (E6.3.18). Finally applying equation (E6.3.8), $u = 1.91$ HBW and the associated expanded combined uncertainty $U = ku = 3.93$ HBW ($k = 2.06$, $\nu_{eff} = 25$, confidence level 95 %) were calculated.

E6.3.6.7 Instrumented Indentation Test calculation

Most of the measuring heads of Instrumented Indention Testing machines are working-force controlled. The indentor displacement is sensed using various types of sensor. The sensed displacement z is given by the measured signal and the calibration function.

An evaluation of the standard uncertainty $u_{z_{\max}}$ of the maximum sensed displacement $z_{\max}$ is made.

Assume the estimation of the zero point z_0, the position of first physical contact between indentor and sample, yields a displacement range Δz in which it is very likely that contact occurred. Assuming further that within this range the probability density function is rectangular, the standard uncertainty due to surface detection is

$$u_{z_0} = \frac{\Delta z}{2\sqrt{3}}. \tag{E6.3.35}$$

The standard uncertainty for the position just before unloading can be calculated by considering the scatter about a straight-line fit $z(t)$ to the time-displacement data over a period prior to unloading. The maximum displacement is

$$h_{\max} = z_{\max} - h_{\text{frame}} - h_{\text{drift}}, \qquad h_{\text{frame}} = (F_1 - F_2)c_{\text{f}}. \tag{E6.3.36}$$

In equation (E6.3.36), F_1 and F_2 are the forces applied at contact and at maximum displacement, c_{f} is the frame stiffness and h_{drift} is the drift of the measuring system. The standard uncertainty $u_{h_{\max}}$ is given by

$$u^2_{h_{\max}} = \left(\frac{\partial h_{\max}}{\partial z_{\max}}\right)^2 u^2_{z_{\max}} + \left(\frac{\partial h_{\max}}{\partial h_{\text{frame}}}\right)^2 u^2_{h_{\text{frame}}} + \left(\frac{\partial h_{\max}}{\partial h_{\text{drift}}}\right)^2 u^2_{h_{\text{drift}}} \tag{E6.3.37}$$

$$= u^2_{z_{\max}} + c_{\text{f}}^2 u^2_{F_1} + c_{\text{f}}^2 u^2_{F_2} + (F_2 - F_1)^2 u^2_{c_{\text{f}}} + u^2_{h_{\text{drift}}}. \tag{E6.3.38}$$

The frame compliance is determined according to the procedure described in ISO 14577-1 [495]. Together with the mean value, the standard uncertainty $u_{c_{\text{f}}}$ is calculated.

An estimate of the drift h_{drift} and evaluation of the associated standard uncertainty $u_{h_{\text{drift}}}$ are obtained.

If the force is applied by an electromagnetic system, the measured force F is estimated using the applied current and the calibration function. An evaluation of the standard uncertainty u_{F} associated with the estimate is made.

Analyzing the force balance at contact and at maximum displacement, the force at maximum displacement $F_{\max}$ is calculated using

$$F_{\max} = F_2 - F_1 - K_S Z_{\max}. \tag{E6.3.39}$$

An estimate of the support stiffness K_{S} is made and the associated standard uncertainty u_{K_S} evaluated. The standard uncertainty $u_{F_{\max}}$ is then given by

$$u^2_{F_{\max}} = \left(\frac{\partial F_{\max}}{\partial F_2}\right)^2 u^2_{F_2} + \left(\frac{\partial F_{\max}}{\partial F_1}\right)^2 u^2_{F_1} + \left(\frac{\partial F_{\max}}{\partial Z_{\max}}\right)^2 u^2_{Z_{\max}} + \left(\frac{\partial F_{\max}}{\partial K_{\text{S}}}\right)^2 u^2_{K_{\text{S}}} \tag{E6.3.40}$$

$$= u^2_{F_1} + u^2_{F_2} + K_{\text{S}}^2 u^2_{z_{\max}} + Z^2_{\max} u^2_{K_S}. \tag{E6.3.41}$$

Assume the estimation of the force applied at the zero point yields a force range ΔF for which it is very likely that contact occurred. Assuming further that within this range the PDF is rectangular, the standard uncertainty due to surface detection is

$$u_{F_1} = \frac{\Delta F}{2\sqrt{3}}. \tag{E6.3.42}$$

Since stiffness S is the slope of the force-removal curve at $h_{\max}$, S is usually estimated by fitting a power-law to the force-displacement data and then analytically differentiating that expression. In general, the stiffness standard uncertainty u_ϵ depends on the accuracy of the fitting procedure used.

Assuming u_c to be negligible, the contact depth h_c and the associated standard uncertainty u_{h_c} are given by

$$h_c = h_{\max} - \frac{\epsilon F_{\max}}{S}, \tag{E6.3.43}$$

$$u_{h_c}^2 = \left(\frac{\partial h_c}{\partial h_{\max}}\right)^2 u_{h_{\max}}^2 + \left(\frac{\partial h_c}{\partial F_{\max}}\right)^2 u_{F_{\max}}^2 + \left(\frac{\partial h_c}{\partial S}\right)^2 u_S^2 + \left(\frac{\partial h_c}{\partial \epsilon}\right)^2 u_\epsilon^2 \tag{E6.3.44}$$

$$= u_{h_{\max}}^2 + \left(\frac{\epsilon}{S}\right)^2 u_{F_{\max}}^2 + \left(\frac{\epsilon F_{\max}}{S^2}\right)^2 u_S^2 + \left(\frac{F_{\max}}{S}\right)^2 u_\epsilon^2. \tag{E6.3.45}$$

The contact area is normally expressed as a mathematical function relating the projected surface area A_p to the contact depth h_c. A procedure for the verification of the area function is given in [5]. The area function is given as a polynomial-like function with m adjustable coefficients $a_0, \ldots, a_{m-1}$:

$$A_p(h_c) = \sum_{j=1}^{m} a_{j-1} h_c^{2^{2-j}} = a_0 h_c^2 + a_1 h_c + a_2 h_c^{1/2} + a_3 h_c^{1/4} + \cdots + a_{m-1} h_c^{2^{2-m}}. \tag{E6.3.46}$$

The associated standard uncertainty u_{A_p} is given by

$$u_{A_p}^2 = \sum_{j=1}^{m} \left(h_c^{2^{2-j}} u_{a_{j-1}}\right)^2 + \sum_{j=1}^{m} \left(2^{2-j} a_{j-1} h_c^{2^{2-j}-1} u_{h_c}\right)^2. \tag{E6.3.47}$$

The measurement model for Instrumented Indentation hardness is

$$H_{IT} = \frac{F_{\max}}{A_{p(h_c)}}. \tag{E6.3.48}$$

The combined standard uncertainty $u_{H_{IT}}$ is given by

$$u_{H_{IT}}^2 = \left(\frac{\partial H_{IT}}{\partial F_{\max}}\right)^2 u_{F_{\max}}^2 + \left(\frac{\partial H_{IT}}{\partial A_p(h_c)}\right)^2 u_{A_{p(h_c)}}^2 \tag{E6.3.49}$$

$$= \left(\frac{1}{A_{p(h_c)}}\right)^2 u_{F_{\max}}^2 + \left(\frac{F_{\max}}{A_{p(h_c)}^2}\right)^2 u_{A_{p(h_c)}}^2 \tag{E6.3.50}$$

E6.3.7 Interpretation of results

The most significant impact of this work relates to uncertainty evaluation in the framework of normative development and conformity assessment.

E6.3.7.1 Uncertainty evaluation in hardness testing standardization

Given the close connection between this work and the guideline 'Metallic Materials – Strategy for a Common Framework to determine Measurement Uncertainty in Mechanical Testing' drafted in the Advisory Group 'Uncertainty' of ISO/TC 164, the proposed harmonized procedure will be used as the starting point for discussion in TC 164/SC 3 Hardness testing. Following the resolutions approved by TC 164/SC 3 in 2020, the systematic review of all mentioned standards starts in 2021. One of the main purposes of these systematic reviews is the harmonization of the evaluation of uncertainty for all hardness testing standards. This report has been submitted to TC 164/SC 3 for consideration in that activity.

E6.3.7.2 Uncertainty evaluation in conformity assessment

Most product specifications have tolerances that have been developed over the past years based mainly on the requirements of the product but also, in part, on the performance of the testing machines used to make the hardness measurement. These specifications, therefore, incorporate a contribution due to the uncertainty of the hardness measurement and, generally, it would be inappropriate to make any further allowance for this uncertainty by, for example, modifying the limits in the specifications by the evaluated uncertainty of the hardness measurement.

In special circumstances, adjusting specification limits by the measurement uncertainty is appropriate. In any case, a proper and harmonized evaluation of uncertainty is clearly necessary for understanding the actual specifications.

In all evaluated hardness standards informative guidance on how to evaluate an uncertainty is given. As long as there is no harmonized procedure in hardness standards, evaluated uncertainties can create ambiguity in conformity assessment.

At the ISO/TC 164 level, the Advisory Group 'Uncertainty' has drafted a guidance document [497] to assist user standards in the interpretation of terms and definitions applied to uncertainty evaluation, application in decision rules and determination of risk in applying a decision rule. The main recommendations are the following:

- the method standards themselves are not recommended for specifying requirements of assessment criteria for use in product standards;
- the method standards are not to be considered normative in the application of measurement uncertainty when testing or assessing products;
- if measurement uncertainty in the method standards is determined uniformly, they can be considered in a conformity assessment if required;
- a statement of conformity can be made based on the standards dealing with general product specifications (e.g. ISO 14253-1:2018 [199] and JCGM 106:2012 [6]).

Example E6.4

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

E6.4.1 Summary

This example illustrates the evaluation of measurement uncertainty related to the calibration of a mobile optical measurement system, based on the use of an SI-traceable reference standard bar measured in specific spatial positions. The measurement system studied (Krypton, model K610) contains three linear CCD (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.

E6.4.2 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 [498, 499] – 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 [499] 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 [498, 499].

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 E6.4.1, which displays seven spatial distances $d_1 \ldots d_7$ that are measured.

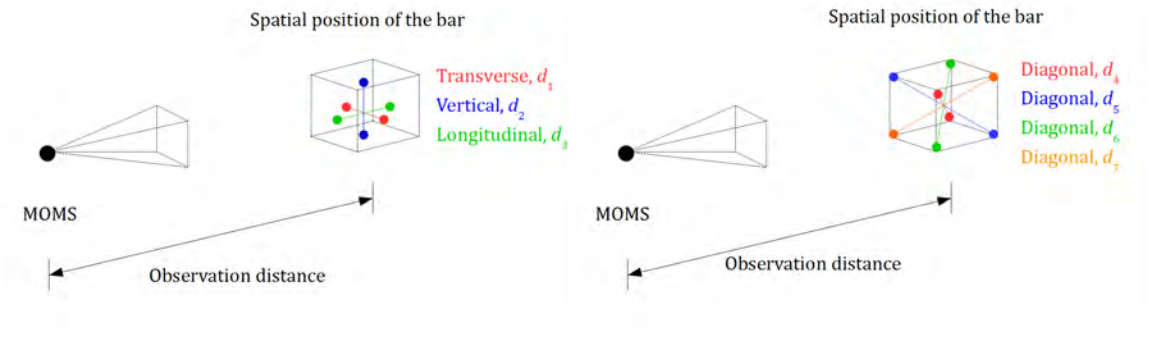


Figure E6.4.1: Schematic representation of the MOMS calibration

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 [499].

E6.4.3 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:

$$d = l - l_s(1 + \alpha_s \cdot \theta_s) \tag{E6.4.1}$$

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.

E6.4.4 Measurement model

The measurement model can be obtained from expression (E6.4.1) and is given by

$$l = d + l_s(1 + \alpha_s \cdot \theta_s) \tag{E6.4.2}$$

E6.4.5 Uncertainty propagation

The application of the law of propagation of uncertainty [3] to expression (E6.4.2) yields

$$u^2(l) = c_d^2 \cdot u^2(d) + c_{l_s}^2 \cdot u^2(l_s) + c_{\alpha_s}^2 \cdot u^2(\alpha_s) + c_{\theta_s}^2 \cdot u^2(\theta_s) \tag{E6.4.3}$$

with

$$c_d = \frac{\partial l}{\partial d} = 1 \tag{E6.4.4}$$

$$c_{l_s} = \frac{\partial l}{\partial l_s} = 1 + \alpha_s \cdot \theta_s \tag{E6.4.5}$$

$$c_{\alpha_s} = \frac{\partial l}{\partial \alpha_s} = l_s \cdot \theta_s \tag{E6.4.6}$$

$$c_{\theta_s} = \frac{\partial l}{\partial \theta_s} = l_s \cdot \alpha_s \tag{E6.4.7}$$

and thus

$$u^2(l) = u^2(d) + (1 + \alpha_s \cdot \theta_s)^2 \cdot u^2(l_s) + l_s^2 \cdot \theta_s^2 \cdot u^2(\alpha_s) + l_s^2 \cdot \alpha_s^2 \cdot u^2(\theta_s) \tag{E6.4.8}$$

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 (E6.4.8) can be simplified:

$$u^2(l) = u^2(d) + u^2(l_s) + l_s^2 \cdot \theta_s^2 \cdot u^2(\alpha_s) \tag{E6.4.9}$$

Table E6.4.1 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 E6.4.1) in the measurement volume, at a nominal observation distance of 3.5 m. Table E6.4.1 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 E6.4.1: 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 E6.4.1, correlation coefficients, $r(d_i, d_j)$, were determined between pairs of spatial positions of the reference standard bar,

$$r\left(d_i, d_j\right) = \frac{u\left(d_i, d_j\right)}{s\left(d_i\right) \cdot s\left(d_j\right)}, \tag{E6.4.10}$$

where $u\left(d_i, d_i\right)$ is the covariance, which can be calculated by

$$u\left(d_i, d_j\right) = \frac{1}{n(n-1)} \sum_{k=1}^{n} \left(d_{i_k} - \bar{d}_i\right)\left(d_{j_k} - \bar{d}_j\right) \tag{E6.4.11}$$

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 E6.4.2. 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 E6.4.2: 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 E6.4.1. In a global perspective, an estimate of the difference between reading and reference values can be obtained by averaging. The corresponding standard uncertainty [2] is given by

$$u^2(d) = \sum_{i=1}^{7} c_i^2 \cdot u^2\left(d_i\right) + 2 \sum_{i=1}^{6} \sum_{j=i+1}^{7} c_i \cdot c_j \cdot u\left(d_i\right) \cdot u\left(d_j\right) \cdot r\left(d_i, d_j\right) \tag{E6.4.12}$$

where $c_i = c_j = \frac{1}{7}$, $r\left(d_i, d_j\right)$ is the correlation coefficient and $u^2\left(d_i\right) = s^2\left(d_i\right)$ (see table E6.4.2), allowing to simplify expression (12) to

$$u^2(d) = \frac{1}{49}\left[\sum_{i=1}^{7} s^2\left(d_i\right) + 2 \sum_{i=1}^{6} \sum_{j=i+1}^{7} s\left(d_i\right) \cdot s\left(d_j\right) \cdot r\left(d_i, d_j\right)\right] \tag{E6.4.13}$$

The use of the values in tables E6.4.1 and E6.4.2 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: $l_s = 1550.313\,\text{mm}$. 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 $\left(\alpha_s = 0.0 \times 10^{-6}\,^{\circ}\text{C}^{-1}\right)$. However, it has an associated standard uncertainty represented by a uniform distribution with a semi-width of $2 \times 10^{-6}\,^{\circ}\text{C}^{-1}$, 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 [2] corresponds to

$$u(\alpha_s) = \frac{1}{\sqrt{3}} 2 \times 10^{-6}\,^{\circ}\text{C}^{-1} = 1.2 \times 10^{-6}\,^{\circ}\text{C}^{-1} \tag{E6.4.14}$$

The calibration of the MOMS was performed in a controlled laboratory environment, with a nominal temperature of 20 °C. Room temperature time records show an average temperature deviation of $\theta_s = 0.1\,^{\circ}\text{C}$, 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

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

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

Table E6.4.3: Summary of the standard uncertainty components

Standard uncertainty component $u(x_i)$	Source of uncertainty	Standard uncertainty $u(x_i)$	$c_i \equiv \frac{\partial l}{\partial x_i}$	$u_i(l) \equiv \lvert c_i \rvert \cdot u(x_i)$	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	$1.2 \times 10^{-6}\,^{\circ}\text{C}^{-1}$	155 mm °C	0.0002 mm	50
$u(\theta_s)$	Temperature deviation from reference value	0.41 °C	0	0	∞

E6.4.6 Reporting the result

Based on the results shown in table E6.4.3, the combined standard uncertainty, $u_c(l)$, of the length reading is determined from expression (E6.4.9), 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

$$U_{95\%}(l) = k \cdot u_c(l) = 2.04 \cdot 0.010\,\text{mm} = 0.021\,\text{mm}$$

E6.4.7 Interpretation of results

Table E6.4.3 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 $11.5 \times 10^{-6}\,°\mathrm{C}^{-1}$, 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 E6.5

Evaluation of measurement uncertainty associated with the quantification of ephedrine in anti-doping testing

O. Barroso, A. Danion, B. Garrido, S. Westwood, M.G. Cox, A.M.H. van der Veen

E6.5.1 Summary

In this example, an anti-doping laboratory's evaluation of the combined standard uncertainty associated with the measured concentration of the prohibited (in sports) stimulant ephedrine at levels close to the threshold in urine samples is described. For this purpose, laboratory validation data for intermediate precision and bias are used. The evaluation is verified through the laboratory's participation in three rounds of WADA's proficiency testing programme, EQAS. The study concludes with the acceptance of the evaluated standard uncertainty in accordance with WADA requirements.

E6.5.2 Introduction of the application

E6.5.2.1 Regulatory framework

The WADA Prohibited List defines which substances and methods are prohibited in sports. In addition, some of these substances are classified as a threshold substance. That is, for an anti-doping laboratory to decide an adverse analytical finding (AAF), the substance (or its characteristic metabolite) shall be measured in a sample at a concentration higher than a pre-defined threshold T (with at least a 95 % level of confidence). This 95 % level of confidence in exceeding the stated T is achieved by comparing the measured concentration x with a decision limit (DL) [500, 501]. The comparison takes into consideration the analytical standard uncertainty $u_c(x)$ of a result obtained with the validated measurement procedure (see figure E6.5.1).

The narrower distribution (red) in figure E6.5.1 illustrates a normal PDF with mean equal to T and standard deviation equal to a testing laboratory's reported standard uncertainty $u_c(x)$. The broader distribution (blue) shows a normal PDF with the same mean and standard deviation equal to the maximum allowed measurement uncertainty $u_{c,max}$ (see section E6.5.2.4). The

decision limit DL, equal to $T + 1.645u_{c,max}$, where $u_{c,max}$ is the target standard uncertainty, is also shown in the figure. The value 1.645, taken from the normal distribution, ensures that DL constitutes the upper endpoint of the upper one-sided 95 % limit of the probability distribution.

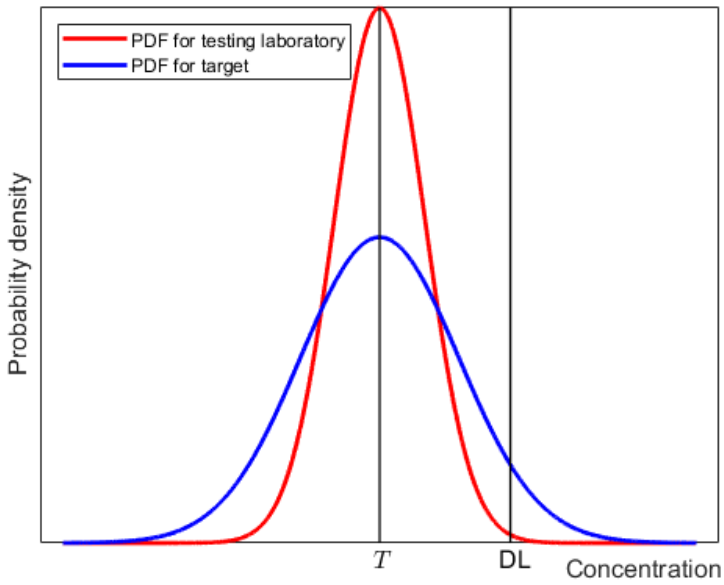


Figure E6.5.1: Probability distribution for (a) mass concentration ephedrine in urine for the test laboratory and (b) corresponding to the target uncertainty showing the threshold T and decision limit DL

The applicable decision rule [7, clause 7.8.6] is based on a result at the threshold T, for which a target measurement uncertainty [502] is established. For the substances concerned, this target measurement uncertainty is given in [501]. Based on this target measurement uncertainty, a guard band [6] is defined, which is used to establish a DL that is applied to assess compliance of the analytical result with the applicable rules. This guard band is established using a coverage factor k of 1.645 to comprise the extreme left 95 % coverage range of a normal distribution. This example is mainly concerned with the task of the laboratory to demonstrate that its analytical standard uncertainty $u_c(x)$ at the threshold is no greater than the target standard uncertainty, so that the DL can be used in the reporting of an AAF [501]:

> 'Where a threshold T has been established for a prohibited substance, the DL represents the value for that substance above which it can be decided that the result in a given sample, obtained using a validated measurement procedure, has exceeded T with a statistical confidence of at least 95 %, and hence that an AAF is justified.'

To evaluate $u_c(x)$, WADA-accredited laboratories may use any approach consistent with the Guide to the expression of uncertainty in measurement (GUM) [2]. Such approaches include 'top-down' or 'empirical' approaches using data derived, for example, from intra- or inter-laboratory method validation studies or from the WADA External Quality Assessment Scheme (EQAS) programme, outlined in section E6.5.2.2. In this regard, the use of laboratory validation data for method performance characteristics (namely intermediate precision and bias) is the recommended procedure. The evaluated $u_c(x)$ is re-assessed periodically using quality control (QC) data to account for expected improvements in method performance over time as well as any other factors that may influence its application.

Evaluations of $u_c(x)$ are verified through the laboratory's regular participation in the WADA EQAS, which may include the analysis of blind, double-blind or educational (open) samples. In this example, a laboratory's evaluations associated with the measurement of the prohibited stimulant ephedrine at levels close to the threshold in urine samples is described. Laboratory validation data are used for this purpose, and the evaluation verified by the laboratory's participation in three rounds of WADA EQAS (two blind, one double-blind). The study concludes with the acceptance of the evaluated $u_c(x)$ in accordance with the requirements of the relevant WADA Technical Document (TD DL) [501].

E6.5.2.2 WADA External Quality Assessment Scheme (EQAS)

WADA regularly distributes EQAS samples to WADA-accredited laboratories and, when applicable, to probationary laboratories. The WADA EQAS is designed to monitor continually the capabilities of the accredited laboratories and probationary laboratories to evaluate their proficiency and to improve test result uniformity between laboratories.

The WADA EQAS includes three different types of EQAS:

1. Blind EQAS (bEQAS)

   Laboratories will be aware that the sample is a WADA EQAS sample since it is delivered by WADA's EQAS Sample Provider; however, the content of the EQAS samples is not known to the participants.

2. Double-blind EQAS (dbEQAS)

   The dbEQAS samples are packaged and distributed to laboratories by testing authorities (that is, laboratory clients) in order to be indistinguishable from routine anti-doping samples and, therefore, the participants are not aware that they are processing an EQAS sample.

3. Educational EQAS

   Educational EQAS samples may be provided as open (in which case the content of the EQAS sample is known), blind or double-blind samples and therefore the participants may or may not be made aware of the educational EQAS sample or its contents. This programme aims at providing educational opportunities to the participants or data-gathering opportunities to WADA.

E6.5.2.3 Evaluation of measurement uncertainty in anti-doping quantification procedures

Since the methods employed for doping control analyses must be validated and determined to be fit-for-purpose (WADA International Standard for Laboratories [500]), the use of laboratory validation data for method performance characteristics (namely intermediate precision and bias) is the recommended procedure.

Where the laboratory result x is assigned as the mean of n replicate measurements, the combined standard uncertainty $u_c(x)$ associated with x is evaluated as the root square sum of a contribution s_w derived from the within-laboratory (im)precision estimate of an individual result using the measurement procedure under intermediate precision conditions and a contribution u_{bias} accounting for measurement bias:

$$u_c(x) = \sqrt{\frac{s_w^2}{n} + u_{bias}^2}.$$

Relevant points are as follows:

- s_w may be determined by repeat analysis of a typical sample under intermediate precision conditions (that is, over time by different operators and using different equipment, if applicable) and represents the intra-laboratory precision of an individual analysis.
- To determine u_{bias}, repeated analysis of a sample having an independently assigned reference value x_{ref} (preferably a Certified Reference Material[1]) using the complete measurement procedure under repeatability conditions is performed. Under these conditions, the bias contribution u_{bias} to $u_c(x)$ is evaluated according to

$$u_{bias}^2 = \frac{s_{ref}^2}{n} + \Delta^2 + u^2(x_{ref}),$$

where

s_{ref}	standard deviation (SD) under repeatability conditions of the measured values obtained for the CRM
Δ	deviation $\Delta = x - x_{ref}$ of the mean value x of n replicate measurement results from the corresponding reference value x_{ref}
$u(x_{ref})$	standard uncertainty associated with x_{ref}.

- To evaluate method performance and check for the absence of unaccounted bias, a secondary material (CRM or usually EQAS material), also having independently assigned concentration x_{test} of the compound concerned, is analyzed using the laboratory's method as it is applied in practice. In this case the absolute value of $\Delta x_{test} = x - x_{test}$ is compared with the expanded uncertainty $U(\Delta x_{test})$ associated with Δx_{test}. $U(\Delta x_{test})$ is taken as $2u(\Delta x_{test})$, where $u(\Delta x_{test})$ is the standard uncertainty associated with Δx_{test}, assuming a normal distribution. $u(\Delta x_{test})$ combines the standard uncertainty $u(\Delta x_{test})$ associated with the assigned value x_{test} and the standard uncertainty $u(x)$ associated with x:

$$u^2(\Delta x_{test}) = u^2(x_{test}) + u^2(x).$$

- If $\Delta_{test} \leq U_\Delta$ there is no significant difference between the measurement and the certified/assigned value and there is no evidence of unaccounted bias.
- In addition, the evaluated $u_c(x)$ should be re-assessed periodically using QC data. This periodic evaluation would account for expected improvements in method performance over time as well as any other factors that may influence its application (for example, replacement of technicians or analysts).

E6.5.2.4 Example of laboratory estimation and verification of measurement uncertainty for quantification of ephedrine in urine samples

Ephedrine is a stimulant, which is included under class S6.b – Specified Stimulants of the WADA Prohibited List [503] and is prohibited in-competition in all sports. In addition, ephedrine is classified as a threshold substance, that is, for an anti-doping laboratory to conclude an adverse analytical finding (AAF), ephedrine shall be measured in a sample at a concentration higher than (with at least 95 % level of confidence) a threshold T set at $10\,\mu g\,mL^{-1}$. This 95 % level of

[1] If a suitable CRM is not available, a QC sample prepared to be traceable to a CRM or a value-assigned EQAS sample can be used.

confidence in exceeding the stated T is achieved by comparing the measured concentration with a decision limit DL set at 11 $\mu g\,mL^{-1}$. The DL takes into consideration the analytical uncertainty of a result obtained with the validated measurement procedure, using a coverage factor k of 1.645 to comprise the 95 % coverage range of a single-tailed normal distribution.

The measured concentration of ephedrine shall be determined using a method having an associated combined standard uncertainty $u_c(x)$ for a result at levels close to T, which does not exceed the maximum allowed measurement uncertainty $u_{c,max}$ of 0.5 $\mu g\,mL^{-1}$ (5 %). This value of $u_{c,max}$ is a conservative estimate derived from inter-laboratory reproducibility (robust) data obtained from EQAS measurement performance data.

For evaluation of the $u_c(x)$ of the measurement procedure of ephedrine, a WADA-accredited laboratory applies a top-down intra-laboratory data approach based on the estimation of intermediate precision s_w and method bias. The uncertainty contribution due to the intermediate precision of a determination was estimated using a urine QC sample, containing ephedrine at a level close to T, which was analyzed over an extended period under intermediate precision conditions. The relative intermediate precision $s_{w,rel}$ for the result of an individual analysis, obtained from the combined results for this sample is 4.2 %.

The uncertainty contribution due to the precision of a determination equals $0.042x/\sqrt{n}$, where x is the mean of n replicate determinations.[2]

The uncertainty contribution due to bias was determined from the results of replicate determinations of a CRM solution for ephedrine in urine under repeatability conditions as described below.

E6.5.3 Specification of the measurand(s)

The measurand is the mass concentration ephedrine in urine. The threshold T is set at 10 $\mu g\,mL^{-1}$. The regulation requires the laboratories to demonstrate that their standard measurement uncertainty does not exceed 0.5 $\mu g\,mL^{-1}$ at levels close to T. The DL is established based on a one-sided normal distribution and is set at 11 $\mu g\,mL^{-1}$ [501].

E6.5.4 Measurement model

The laboratory establishes its precision using a QC material and its bias using a certified reference material. The resulting uncertainty budget is deemed valid for the estimation of the concentration of ephedrine μ in urine samples collected from athletes during doping control.

The statistical model of a laboratory result is given by [232]

$$x = \mu + B + \varepsilon,$$

where x denotes the measured value, B a bias term and ε a random error term. ε is evaluated as precision under within-laboratory conditions.

The estimation of B can be performed in different ways, and hence it can be viewed also in different ways:

– as a bias term, evaluated using a certified reference material (CRM);

[2] According to the WADA International Standard for Laboratories (ISL) [500], the result of a quantitative confirmation procedure is reported as the mean of a triplicate determination.

– as a laboratory effect in the EQAS.

The evaluation summarised in table E6.5.1 uses a CRM to establish the bias Δ in the measured value using the certified value as the reference. This bias is included in the uncertainty budget for B. A target uncertainty is set to ensure that the associated standard uncertainty $u(B)$, in combination with a precision component, is fit for purpose to implement the WADA rules. The target uncertainty is used as standard uncertainty in the subsequent steps, including the conformity assessment.

Table E6.5.1: Calculation of bias component of $u_c(x)$ using a CRM

Step	Action	Formulae	Result
1	Use the certified value of ephedrine in the CRM		$x_{ref} = 10.00\,\mu g\,mL^{-1}$ $U(x_{ref}) = 0.06\,\mu g\,mL^{-1}$ $u(x_{ref}) = 0.03\,\mu g\,mL^{-1}$ $u_{rel}(x_{ref}) = 0.3\,\%$
2	Derive the relative uncertainty component due to precision for a mean value x of triplicate results using this procedure	$x = \frac{1}{3}(x_1 + x_2 + x_3)$ $u_{prec} = \frac{s_w}{\sqrt{n}} = \frac{0.042x}{\sqrt{3}}$	$u_{rel}(x) = \frac{4.2\,\%}{\sqrt{3}} = 2.4\,\%$
3	Determine the bias component of the method by analyzing the ephedrine urine CRM 30 times under repeatability conditions	$x = \text{mean}(x_1, \ldots, x_{30})$ $s_{ref} = \text{SD}(x_1, \ldots, x_{30})$ $u_{\text{bias precision}} = \frac{s_{ref}}{\sqrt{30}}$	$x = 10.3\,\mu g\,mL^{-1}$ $s_{ref} = 0.203\,\mu g\,mL^{-1}$ RSD= 1.97 % $u_{\text{bias precision}} = 0.36\,\%$
4	Form absolute deviation	$\Delta = \lvert x - x_{ref}\rvert$	$\Delta = 0.3\,\mu g\,mL^{-1}$
5	Form relative deviation	$\Delta_{rel} = \frac{\lvert x - x_{ref}\rvert}{x_{ref}}$	$\Delta_{rel} = 3.0\,\%$
6	Form u_{bias} contribution to $u_c(x)$	$u_{bias} = \left[\frac{s_{ref}^2}{\sqrt{n}} + \Delta^2 + u^2(x_{ref})\right]^{1/2}$	$u_{bias} = 3.0\,\%$ Note: $u_{bias} = \Delta$ (other contributions negligible)
7	Evaluate combined relative standard uncertainty $u_c(x)$ for mean of triplicate results	$u_c(x) = (u_{prec}^2 + u_{bias}^2)^{1/2}$	$u_c(x) = 3.8\,\%$ Note: $u_c(x) < u_{c,max}$

In table E6.5.2, the verification of the laboratory's performance in an EQAS, the magnitude of B, the laboratory effect, is assessed using the target standard uncertainty as the standard deviation for performance rating in the EQAS [69]. The z-score shall be satisfactory, that is, $|z| \leq 2$. This criterion ensures that the laboratory result in the EQAS is fit for purpose.

From

$$u_c^2 = \frac{s_w^2}{n} + u_{bias}^2,$$

the statistical model in section E6.5.4 follows. s_w, the intermediate precision, is estimated by the repeated analysis of urine samples under within-laboratory reproducibility conditions. The estimate of $\widehat{B}$ and its evaluated standard uncertainty $u(\widehat{B})$ are given by

$$\widehat{B} = 0,$$
$$u^2(\widehat{B}) = \frac{s_{ref}^2}{n'} + \Delta^2 + u_{ref}^2.$$

The bias estimate $\widehat{B}$ is set at zero by including Δ in the expression of its associated standard uncertainty and to confirm that this standard uncertainty, combined with the intermediate precision, does not exceed the target standard uncertainty. (See also ISO 21748 [232] for the application of this model.)

Table E6.5.2: Verification of u_c assignment from Laboratory EQAS data

Step	Action	Formulae	Result of 3 trials
1	Make laboratory triplicate determinations of ephedrine in EQAS Sample	x_1 x_2 x_3	(11.8, 12.1, 11.6) µg mL^{-1} (11.5, 11.9, 11.6) µg mL^{-1} (11.5, 11.4, 11.5) µg mL^{-1}
2	Form laboratory mean value and standard uncertainty for mean of triplicates	$x = \frac{1}{3}(x_1 + x_2 + x_3)$ $u_c(x) = 0.038 \times x$	(11.8, 11.6, 11.5) µg mL^{-1} (0.45, 0.44, 0.44) µg mL^{-1}
3	Form standard deviation s_r of laboratory repeatability determinations of EQAS sample. Check consistency with validated intermediate precision s_w	$s_r = \mathrm{SD}(x_1, x_2, x_3)$ $s_{rel,r} = s_r/x$ s_w	(0.252, 0.208, 0.45) µg mL^{-1} (2.1, 1.8, 0.4) % (4.2, 4.2, 4.2) %
4	Number of participant laboratories in the EQAS		31, 27, 30
5	EQAS Assigned value X_{EQAS}	Robust average of results reported by all participants in EQAS study	(12.3, 12.3, 11.1) µg mL^{-1}
6	Report inter-laboratory reproducibility standard deviation s_R	Robust SD of results reported by all participants in EQAS study	(0.53, 0.53, 0.35) µg mL^{-1} (4.3, 4.3, 3.2) %
7	Report and check z-score	$z = \dfrac{x - x_{EQAS}}{s_{EQAS}}$	0.27, 0.71, 0.47 All $\lvert z\rvert -$ scores < 2: satisfactory
8	Form absolute deviation ΔX_{EQAS} of laboratory value x from assigned value X_{EQAS}	$\Delta X_{EQAS} = \lvert x - X_{EQAS}\rvert$	(0.5, 0.6, 0.4) µg mL^{-1}
9	Form expanded uncertainty for $U(\Delta X_{EQAS})$	$2\left(u^2(X_{EQAS}) + u^2(x)\right)^{1/2}$ $\Delta \le U(\Delta X_{EQAS})$?	(0.93, 0.92, 0.89) µg/mL Yes, Yes, Yes
10	Calculate E_n-score for verification of $u_c(x)$	$\dfrac{x - X_{EQAS}}{[u^2(x) + u^2(X_{EQAS})]^{1/2}}$ $\Delta \le U(\Delta X_{EQAS})$?	 Yes, Yes, Yes

E6.5.5 Uncertainty propagation

The uncertainty propagation for the calculation of the bias component of $u_c(x)$ using a CRM proceeds by following the series of steps in tables E6.5.1 and E6.5.2.

Stage 1 Calculation of bias component of $u_c(x)$ using a CRM: table E6.5.1;

Stage 2 Verification of $u_c(x)$ assignment from Laboratory EQAS data: table E6.5.2.

In the calculations, the law of propagation of uncertainty for independent input quantities from the GUM [2, eqn. (10)] is used.

For verification of $u_c(x)$ using EQAS data, the Laboratory considers data from each EQAS round that includes samples containing ephedrine. In this case, the Laboratory has participated in three recent EQAS rounds:

- Blind EQAS-01/2019 (April 2019): Sample prepared by spiking a urine blank pool with ephedrine standard at 10.6 $\mu g\,mL^{-1}$ (concentration determined by EQAS sample provider) (and norephedrine at 0.85 $\mu g\,mL^{-1}$);
- Double-blind EQAS-02/2019 (September 2019): same sample distributed in bEQAS-01/2019, but this time distributed in double-blind fashion;
- Blind EQAS 01/2020 (January 2020): New sample prepared by spiking a urine blank pool with ephedrine standard at 11.7 $\mu g\,mL^{-1}$ (again concentration determined by EQAS sample provider) and norephedrine at 0.8 $\mu g\,mL^{-1}$.

In each EQAS round, the inter-laboratory EQAS consensus value and reproducibility (SD) are obtained by robust statistics from the results reported by all participants in the EQAS study.

E6.5.6 Reporting the result

The laboratory reports the measured value and the target uncertainty. Based on the applicable decision rule, the laboratory states that the result constitutes an adverse analytical finding when the decision limit is exceeded.

E6.5.7 Interpretation of results

This example demonstrates how a laboratory can seek support for meeting a given performance requirement. In the WADA rules [500,501], such a requirement is given as a maximum value for the standard uncertainty, a target standard uncertainty.

The inclusion of the bias observed when using a CRM in an uncertainty budget can be consequential (see also example E1.8). Other approaches when using a CRM are permitted, which implies that the guidance of ISO Guide 33 [10] can also be followed. This approach requires an evaluation of the uncertainty associated with the measured value of the CRM. Unbiasedness is then demonstrated if

$$|\Delta| \leq 2\sqrt{u_{\text{meas}}^2 + u_{\text{CRM}}^2},$$

and, given the similarity of the procedure for analysing the CRM, many components of uncertainty in this assessment can be used in a GUM uncertainty budget, following for example the guidance in the EuraChem/CITAC guide [42].

Alternatively, the guidance of ISO 21748 [232] can be used to establish the uncertainty of a laboratory result from the participation in the EQAS of WADA. The only component that needs in such an approach further attention is a possible bias of all laboratories. With the current practices in place, this possibility is ruled out by using a CRM to establish that the laboratory results are not significantly biased.

To conclude, there are several options to establish an uncertainty budget for the measurand, which can range from practically completely 'top-down', to mostly 'bottom-up'. In the end, the rigorous quality control and quality assurance measures ensure that laboratories meet the set performance requirements, thereby guaranteeing that the decision limits set by WADA can be maintained.

Example E6.6

Measurement uncertainty in a multiplexed data-acquisition system

A. Carullo, S. Corbellini, A. Vallan

E6.6.1 Summary

This example refers to the measurement of the magnitude of an electrical impedance by comparison to a standard resistor. The voltage signals across the impedance and the resistor are acquired by means of a multiplexed data-acquisition board. The measurement uncertainty is estimated according to the GUM uncertainty framework (GUF).

E6.6.2 Introduction of the application

The magnitude of the impedance of an electrical component is one of the quantities that is often required to be measured in Alternate Current (AC) in order to characterize the behaviour of the component at different operating frequencies. The device under test could be used as a current shunt that converts an unknown AC current into an AC voltage signal and then its impedance magnitude has to be known at the frequencies of the unknown current. In another scenario, the device under test could be the load of an AC power supply and then the knowledge of its frequency behaviour gives important information related to the driving capabilities of the power supply.

Among the different techniques that can be used to measure the magnitude of an electrical impedance, the comparison method in AC current is here considered, which is based on the measurement circuit that is shown in the figure E6.6.1. The unknown impedance $\mathbf{Z_X}$[1] is connected in series to a known standard resistor R_S and the voltage source $\mathbf{V_G}$ supplies the circuit with a sinusoidal signal at the frequency f_0. Two true root-mean square (true rms) voltmeters are connected in parallel to $\mathbf{Z_X}$ and R_S, respectively, thus measuring the voltage amplitudes V_X and V_S.

[1]Hereafter bold face character is used for complex quantities, while normal character denotes the magnitude of the same quantities.

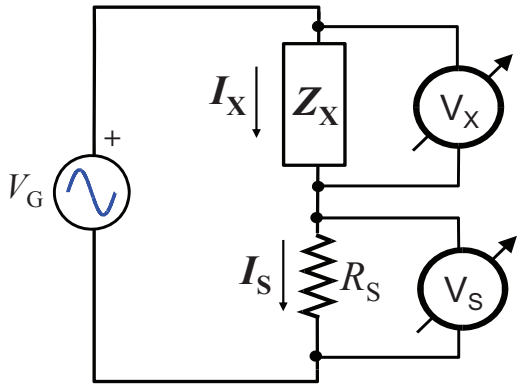


Figure E6.6.1: The basic circuit for the measurement of the impedance magnitude.

Provided that the load effect of the two voltmeters is negligible, the current $\mathbf{I_X}$ that flows through $\mathbf{Z_X}$ is the same as the current $\mathbf{I_S}$ that flows through R_S:

$$I_X = \frac{V_X}{Z_X} = I_S = \frac{V_S}{R_S}, \tag{E6.6.1}$$

and then the magnitude of the unknown impedance can be obtained as:

$$Z_X = R_S \frac{V_X}{V_S}. \tag{E6.6.2}$$

According to the measurement model (E6.6.2), the main uncertainty contributions are related to the standard resistor R_S and to the two voltmeters that measure the voltages V_S and V_X. Other possible uncertainty contributions could be related to the load effect of the voltmeters and to the effect of the measurement frequency on the standard resistor R_S.

In this example, the described method is implemented using a two-channel multiplexed Data-AcQuisition (DAQ) board and taking into account the uncertainty contributions related to the different components of the measuring chain.

E6.6.3 Specification of the measurand

The impedance of an electrical component is usually represented by means of a complex number, where the real part R represents the resistive component of the impedance, while the imaginary part X represents the reactive component or reactance. The reactance increases as the frequency f increases for an inductance L according to this rule:

$$X_L = 2\pi f L, \tag{E6.6.3}$$

while the reactance decreases as the frequency increases for a capacitance according to this expression:

$$X_C = \frac{1}{2\pi f C}. \tag{E6.6.4}$$

In general, an impedance $\mathbf{Z}$ is represented as a complex number in this way:

$$\mathbf{Z} = R \pm jX, \tag{E6.6.5}$$

where the imaginary number j satisfies the equation $j^2 = -1$ and the sign is plus for an inductive reactance and minus for a capacitive reactance.

The magnitude of an impedance, which is the measurand of this example, can be expressed as:

$$Z = \sqrt{R^2 + X^2}, \tag{E6.6.6}$$

that for an ohmic-inductive impedance becomes:

$$Z = \sqrt{R^2 + (2\pi f L)^2}, \tag{E6.6.7}$$

while for an ohmic-capacitive impedance becomes:

$$Z = \sqrt{R^2 + \left(\frac{1}{2\pi f C}\right)^2}. \tag{E6.6.8}$$

The last two expressions highlight that the magnitude of an impedance is strongly dependent on the frequency, thus requiring a series of measurement in the operating frequency range of the impedance to be performed in order to obtain its full characterization.

In this example, the measurand is the magnitude of an impedance in the frequency range from 100 Hz to 20 kHz.

E6.6.4 Measurement model

The basic circuit of figure E6.6.1 is arranged using a dual-channel multiplexed Data AcQuisition (DAQ) board according to the scheme of figure E6.6.2. The input channels CH0 and CH1 of the DAQ board are configured as differential inputs, in order to measure the voltage signals v_X and v_S. The samples v_{Xi} and v_{Si} collected at the two input channels have to be processed in order to estimate the rms value of the two voltage signals.

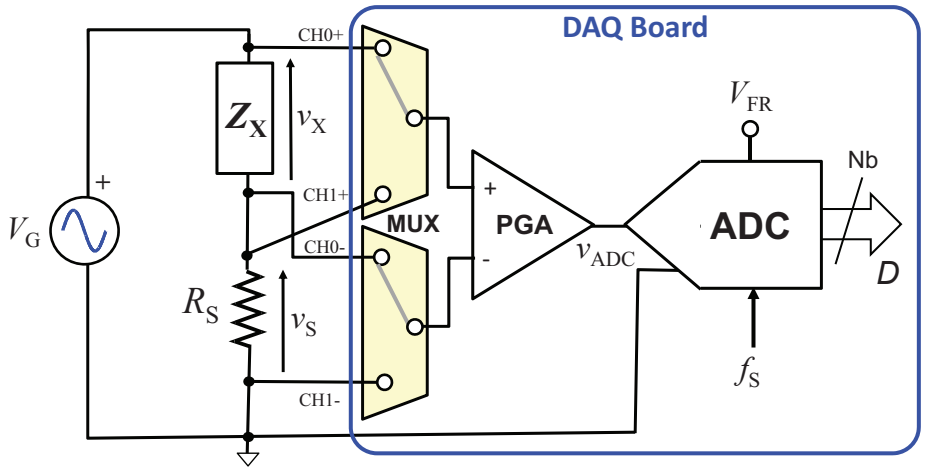


Figure E6.6.2: The measurement circuit arranged using a dual-channel multiplexed DAQ board.

One should note that because of the presence of the multiplexer (MUX), the actual sampling frequency f_{CH} on each channel is lower than the sampling rate f_S of the Analogue-to-Digital Converter (ADC). In the situation of settling time of the system that is negligible with respect to the sample period T_S, the sampling frequency can be obtained as $f_{CH} \approx f_S/2$, i.e. the ADC sampling rate divided by the number of active channels[2].

[2]In the opposite situation (settling time S_T higher than T_S), the sampling frequency is $f_{CH} < f_S/2 \approx 1/(2S_T)$.

The codes D_{Xi} and D_{Si} at the output of the ADC can be expressed as:

$$D_{\mathrm{Xi}} = \frac{v_{\mathrm{Xi}} G_0}{V_{\mathrm{q}}} \; ; \; D_{\mathrm{Si}} = \frac{v_{\mathrm{Si}} G_1}{V_{\mathrm{q}}}, \tag{E6.6.9}$$

where G_0 and G_1 are the nominal gains of the Programmable Gain Amplifier (PGA) for the channels CH0 and CH1, respectively, while V_{q} is the nominal quantization voltage of the ADC, i.e. $V_{\mathrm{q}} = V_{\mathrm{FR}}/2^{Nb}$.

Starting from the nominal expressions (E6.6.9), the samples of the two voltage signals are obtained as:

$$v_{\mathrm{Xi}} = \frac{D_{\mathrm{Xi}} V_{\mathrm{q}}}{G_0} \; ; \; v_{\mathrm{Si}} = \frac{D_{\mathrm{Si}} V_{\mathrm{q}}}{G_1}, \tag{E6.6.10}$$

and the estimation of their rms values v_{X} and v_{S} are obtained according to the following processing algorithm:

$$v_{\mathrm{X}} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(\frac{D_{\mathrm{Xi}} V_{\mathrm{q}}}{G_0} \right)^2} \; ; \; v_{\mathrm{S}} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(\frac{D_{\mathrm{Si}} V_{\mathrm{q}}}{G_1} \right)^2}, \tag{E6.6.11}$$

where $N = kT_0/T_{\mathrm{S}}$ is the number of acquired samples that correspond to an integer number k of periods T_0 of the two voltage signals[3].

Replacing the expressions (E6.6.11) in the measurement model (E6.6.2), the magnitude of the unknown impedance is eventually obtained through the following expression:

$$Z_{\mathrm{X}} = R_{\mathrm{S}} \frac{G_1}{G_0} \sqrt{\frac{\sum_{i=1}^{N} D_{\mathrm{Xi}}^2}{\sum_{i=1}^{N} D_{\mathrm{Si}}^2}}. \tag{E6.6.12}$$

E6.6.5 Uncertainty propagation

The codes D_{Xi} at the output of the measuring chain of figure E6.6.2 when the channel CH0 is selected can be expressed according to the following expression:

$$D_{\mathrm{Xi}} = \frac{G_0}{V_{\mathrm{q}}} (v_{\mathrm{Xi}} + V_{\mathrm{off0,RTI}} + n_{\mathrm{i}}) + D_{\mathrm{off}} + \delta_{\mathrm{qi}}, \tag{E6.6.13}$$

where

$V_{\mathrm{off0,RTI}}$ is the Referred-To-Input voltage offset of the PGA on the channel CH0 (gain G_0);

D_{off} is the offset of the ADC;

δ_{qi} is the *i-th* realization of the quantization noise of the ADC;

n_{i} is the *i-th* realization of the electronic noise superimposed to the input signal.

[3] This condition, which is referred as coherent sampling [504], is considered valid in this example. In non-coherent sampling conditions, the acquired samples can be weighted by means of suitable windowing functions in order to minimize the error due to the processing algorithm [505].

Inverting the expression (E6.6.13), each voltage sample can be expressed as:

$$v_{\mathrm{Xi}} = \frac{D_{\mathrm{Xi}} V_{\mathrm{q}}}{G_0} - V_{\mathrm{off0,RTI}} - \frac{D_{\mathrm{off}} V_{\mathrm{q}}}{G_0} - \frac{\delta_{\mathrm{qi}} V_{\mathrm{q}}}{G_0} - n_{\mathrm{i}}. \tag{E6.6.14}$$

For commercial DAQ boards, manufacturer usually provide specifications in terms of gain error ϵG, offset error e_{OFF} and noise $(e_{\mathrm{q}} + n)$ of the whole measuring chain:

$$e_{\mathrm{OFF},0} = V_{\mathrm{off0,RTI}} + \frac{D_{\mathrm{off}} V_{\mathrm{q}}}{G_0};\ \ (e_{\mathrm{q}} + n)_{\mathrm{i}} = e_{\mathrm{qn,i}} = \frac{\delta_{\mathrm{qi}} V_{\mathrm{q}}}{G_0} + n_{\mathrm{i}}, \tag{E6.6.15}$$

and then the expression (E6.6.14) can be rewritten as:

$$v_{\mathrm{Xi}} = \frac{D_{\mathrm{Xi}} V_{\mathrm{q}}}{G_0}(1 + \epsilon G_0) + e_{\mathrm{OFF},0} + e_{\mathrm{qn,i}}. \tag{E6.6.16}$$

According to the processing algorithm (E6.6.11), the sum of the squared values of the samples v_{Xi} has to be estimated and then each squared value can be estimated as:

$$\begin{aligned} v_{\mathrm{Xi}}^2 = &\left[\frac{D_{\mathrm{Xi}} V_{\mathrm{q}}}{G_0}(1 + \epsilon G_0)\right]^2 + e_{\mathrm{OFF},0}^2 + e_{\mathrm{qn,i}}^2 + \\ &+ 2\frac{D_{\mathrm{Xi}} V_{\mathrm{q}}}{G_0}(1 + \epsilon G_0) e_{\mathrm{OFF},0} + 2\frac{D_{\mathrm{Xi}} V_{\mathrm{q}}}{G_0}(1 + \epsilon G_0) e_{\mathrm{qn,i}} + 2 e_{\mathrm{OFF},0} e_{\mathrm{qn,i}}. \end{aligned} \tag{E6.6.17}$$

The square of the rms value of v_X can be obtained as:

$$v_{\mathrm{X}}^2 = \frac{1}{N}\sum_{i=1}^{N} v_{\mathrm{Xi}}^2. \tag{E6.6.18}$$

Replacing the expression (E6.6.17) of v_{Xi}^2 in the previous relationship, it can be obtained:

$$\begin{aligned} v_{\mathrm{X}}^2 = &\frac{(1 + \epsilon G_0)^2}{N}\sum_{i=1}^{N}\left(\frac{D_{\mathrm{Xi}} V_{\mathrm{q}}}{G_0}\right)^2 + e_{\mathrm{OFF},0}^2 + \frac{1}{N}\sum_{i=1}^{N} e_{\mathrm{qn,i}}^2 + \\ &+ 2\frac{(1 + \epsilon G_0) e_{\mathrm{OFF},0}}{N}\sum_{i=1}^{N}\left(\frac{D_{\mathrm{Xi}} V_{\mathrm{q}}}{G_0}\right) + 2\frac{(1 + \epsilon G_0)}{N}\sum_{i=1}^{N}\left(\frac{D_{\mathrm{Xi}} V_{\mathrm{q}}}{G_0}\right) e_{\mathrm{qn,i}} + \\ &+ \frac{2 e_{\mathrm{OFF},0}}{N}\sum_{i=1}^{N} e_{\mathrm{qn,i}}. \end{aligned} \tag{E6.6.19}$$

In the first term of the second row of the expression (E6.6.19), the mean value of the voltage signal v_{X} is present, which is zero. Furthermore, in the third row the mean value of the noise estimated on N samples can be considered negligible. The second term of the second row, hereafter indicated with $n_{\mathrm{mod,X}}$, is a random noise with zero mean modulated by a sine wave with zero mean, then its expected value is zero and its standard deviation is obtained as [506]:

$$u(n_{\mathrm{mod,X}}) = \frac{v_{\mathrm{X}} u(e_{\mathrm{qn}})}{\sqrt{N}}. \tag{E6.6.20}$$

For the random variable ϵG_0, the expected value is considered equal to zero ($E[\epsilon G_0] = 0$), while its standard uncertainty is indicated as $u(\epsilon G_0)$. For the random variable $e_{\mathrm{OFF},0}$, a normal distribution is considered that is characterized by a zero expected value ($E[e_{\mathrm{OFF},0}] = 0$) and a standard

uncertainty $u(e_{\mathrm{OFF},0})$. Under this assumption, the random variable $e^2_{\mathrm{OFF},0}$ is characterized by a scaled chi-square distribution with one degree of freedom, whose expected value and standard deviation are equal to [507]:

$$E[e^2_{\mathrm{OFF},0}] = u^2(e_{\mathrm{OFF},0});\ u(e^2_{\mathrm{OFF},0}) = \sqrt{2}u^2(e_{\mathrm{OFF},0}). \tag{E6.6.21}$$

The third term in the first row of the expression (E6.6.19) can be considered a new random variable $e^2_{\mathrm{qn0,rms}}$ that is characterized by a scaled chi-square distribution with N degree of freedom, whose expected value and standard deviation are equal to [507]:

$$E[e^2_{\mathrm{qn0,rms}}] = Nu^2(e_{\mathrm{qn,rms}});\ \ u(e^2_{\mathrm{qn0,rms}}) = \sqrt{2N}u^2(e_{\mathrm{qn,rms}}). \tag{E6.6.22}$$

The expression (E6.6.19) can be rewritten as:

$$\begin{aligned} v_{\mathrm{X}}^2 &= (1+2\epsilon G_0+\epsilon G_0^2)v_{\mathrm{X0}}^2 + e^2_{\mathrm{OFF},0} + e^2_{\mathrm{qn0,rms}} + 2(1+\epsilon G_0)n_{\mathrm{mod,X}} \approx \\ &\approx (1+2\epsilon G_0)v_{\mathrm{X0}}^2 + e^2_{\mathrm{OFF},0} + e^2_{\mathrm{qn0,rms}} + 2(1+\epsilon G_0)n_{\mathrm{mod,X}}. \end{aligned} \tag{E6.6.23}$$

where the approximation is due to the assumption $\epsilon G_0^2 \ll \epsilon G_0$.

Repeating the same considerations for the voltage signal v_{S}, the measurement model of the impedance magnitude Z_{X} is eventually expressed as:

$$Z_{\mathrm{X}} = R_{\mathrm{S}}\sqrt{\frac{(1+2\epsilon G_0)v_{\mathrm{X0}}^2 + e^2_{\mathrm{OFF},0} + e^2_{\mathrm{qn0,rms}} + 2(1+\epsilon G_0)n_{\mathrm{mod,X}}}{(1+2\epsilon G_1)v_{\mathrm{S0}}^2 + e^2_{\mathrm{OFF},1} + e^2_{\mathrm{qn1,rms}} + 2(1+\epsilon G_1)n_{\mathrm{mod,S}}}}. \tag{E6.6.24}$$

When the proposed measurement technique is implemented, a very convenient choice consists in using a standard resistor R_{S} with a value that is very close to the impedance magnitude Z_{X}. In this condition it is possible to use the same gain configuration for the two active channels, thus obtaining a partial compensation of the systematic effects due to the gain error and the offset error, since:

$$\epsilon G_0 = \epsilon G_1 = \epsilon G;\ \ e_{\mathrm{OFF},0} = e_{\mathrm{OFF},1} = e_{\mathrm{OFF}}, \tag{E6.6.25}$$

and then the expression (E6.6.24) becomes:

$$Z_{\mathrm{X}} = R_{\mathrm{S}}\sqrt{\frac{(1+2\epsilon G)v_{\mathrm{X0}}^2 + e^2_{\mathrm{OFF}} + e^2_{\mathrm{qn0,rms}} + 2(1+\epsilon G)n_{\mathrm{mod,X}}}{(1+2\epsilon G)v_{\mathrm{S0}}^2 + e^2_{\mathrm{OFF}} + e^2_{\mathrm{qn1,rms}} + 2(1+\epsilon G)n_{\mathrm{mod,S}}}}. \tag{E6.6.26}$$

The expression (E6.6.26) represents the measurement model that is used to propagate the uncertainty contributions according to the GUM framework. The standard uncertainty of the impedance magnitude Z_{X} is estimated under the assumption of negligible correlation among the random variables R_{S}, ϵG, e^2_{OFF}, $e^2_{\mathrm{qn0,rms}}$, $e^2_{\mathrm{qn1,rms}}$, $n_{\mathrm{mod,X}}$ and $n_{\mathrm{mod,S}}$, thus obtaining:

$$\begin{aligned} u^2(Z_{\mathrm{X}}) = &\left(\frac{\partial Z_{\mathrm{X}}}{\partial R_{\mathrm{S}}}\right)^2 u^2(R_{\mathrm{S}}) + \left(\frac{\partial Z_{\mathrm{X}}}{\partial \epsilon G}\right)^2 u^2(\epsilon G) + \left(\frac{\partial Z_{\mathrm{X}}}{\partial e^2_{\mathrm{OFF}}}\right)^2 u^2(e^2_{\mathrm{OFF}}) + \\ &+ \left(\frac{\partial Z_{\mathrm{X}}}{\partial e^2_{\mathrm{qn0,rms}}}\right)^2 u^2(e^2_{\mathrm{qn0,rms}}) + \left(\frac{\partial Z_{\mathrm{X}}}{\partial e^2_{\mathrm{qn1,rms}}}\right)^2 u^2(e^2_{\mathrm{qn1,rms}}) + \\ &+ \left(\frac{\partial Z_{\mathrm{X}}}{\partial n_{\mathrm{mod,X}}}\right)^2 u^2(n_{\mathrm{mod,X}}) + \left(\frac{\partial Z_{\mathrm{X}}}{\partial n_{\mathrm{mod,S}}}\right)^2 u^2(n_{\mathrm{mod,S}}), \end{aligned} \tag{E6.6.27}$$

where[4]:

$$\frac{\partial Z_X}{\partial R_S} = \frac{Z_X}{R_S}. \tag{E6.6.28}$$

$$\frac{\partial Z_X}{\partial \epsilon G} = \frac{R_S^2}{Z_X} \frac{(v_X^2 - v_S^2)(e_{OFF}^2 + e_{qn,rms}^2)}{(v_S^2 + e_{OFF}^2 + e_{qn,rms}^2)^2}. \tag{E6.6.29}$$

$$\frac{\partial Z_X}{\partial e_{OFF}^2} = \frac{R_S^2}{2Z_X} \frac{(v_S^2 - v_X^2)}{(v_S^2 + e_{OFF}^2 + e_{qn,rms}^2)^2}. \tag{E6.6.30}$$

$$\frac{\partial Z_X}{\partial e_{qn0,rms}^2} = \frac{R_S^2}{2Z_X} \frac{1}{v_S^2 + e_{OFF}^2 + e_{qn,rms}^2}. \tag{E6.6.31}$$

$$\frac{\partial Z_X}{\partial e_{qn1,rms}^2} = -\frac{R_S^2}{2Z_X} \frac{v_X^2 + e_{OFF}^2 + e_{qn,rms}^2}{(v_S^2 + e_{OFF}^2 + e_{qn,rms}^2)^2}. \tag{E6.6.32}$$

$$\frac{\partial Z_X}{\partial n_{mod,X}} = \frac{R_S^2}{Z_X} \frac{1}{v_S^2 + e_{OFF}^2 + e_{qn,rms}^2}. \tag{E6.6.33}$$

$$\frac{\partial Z_X}{\partial n_{mod,S}} = -\frac{R_S^2}{Z_X} \frac{v_X^2 + e_{OFF}^2 + e_{qn,rms}^2}{(v_S^2 + e_{OFF}^2 + e_{qn,rms}^2)^2}. \tag{E6.6.34}$$

In the previous expressions of the sensitivity coefficients it has been taken into account that the expected values of ϵG, $n_{mod,X}$ and $n_{mod,S}$ are zero.

One should note that the expressions (E6.6.29) and (E6.6.30) highlight a partial compensation of the uncertainty contributions due to gain and offset errors of the DAQ board. Such a compensation is complete, i.e. null uncertainty contributions related to gain and offset errors, if $v_X = v_S$.

E6.6.6 Reporting the result

The results here reported refer to the data that have been assigned in the MATLAB script unc_zx.m. Other input data can be used accessing to the variables in the script according to the instructions explained in the help section of the script.

The results reported in this example refers to an impedance $\mathbf{Z_X}$ that is characterized by a resistance R connected in series to an inductance L, which have the nominal values $R = 100\ \Omega$ and $L = 1$ mH. According to the expression (E6.6.7), the impedance magnitude Z_X has the nominal values

[4]The sensitivity coefficients have been analytically obtained using the basic rules of derivatives [508].

reported in the table E6.6.1 at the measurement frequencies f. The same table also shows the nominal values of the measured voltage signals v_X and v_S (rms value and peak value), which have been obtained considering a supply voltage with a sinusoidal waveform and an rms value $v_G = 5$ V and a standard resistor with the nominal value $R_S = 100\ \Omega$ and standard uncertainty $u(R_S) = 0.2$ mΩ.

A commercial multiplexed DAQ board is selected for this application that is characterized by the following specifications:

- bipolar full-range from -5 V to $+5$ V; $N_b = 16$;
- maximum ADC sample rate: 250 kSa/s (negligible settling time);
- $\epsilon G = 250 \times 10^{-6}$ (maximum admitted error);
- $U(e_{OFF}) = 5$ mV (95% confidence level specification);
- $u(e_{qn,rms}) = 118\ \mu$V
- $R_{IN} > 10$ GΩ; $C_{IN} = 100$ pF
- Common-Mode Rejection Ratio (CMRR) = 100 dB;
- Cross Talk (CT) $= -80$ dB @ 100 kHz (adjacent channels);
- CT $= -95$ dB @ 100 kHz (non-adjacent channels).

Table E6.6.1: Nominal values of the impedance magnitude Z_X and of the measured voltage signals v_X and v_S at the measurement frequencies f. The supply voltage is a sinusoidal signal with an rms value $v_G = 5$ V and the standard resistor has the nominal value $R_S = 100\ \Omega$.

f (kHz)	$Z_X(\Omega)$	$v_X(V)$	$v_S(V)$	$v_{X,pk}(V)$	$v_{S,pk}(V)$
0.1	100.002	2.500	2.500	3.536	3.535
0.2	100.008	2.500	2.500	3.526	3.535
0.5	100.049	2.501	2.499	3.536	3.535
1	100.197	2.502	2.498	3.539	3.532
2	100.787	2.510	2.490	3.549	3.522
5	104.819	2.559	2.441	3.619	3.452
10	118.101	2.707	2.293	3.829	3.242
20	160.597	3.081	1.919	4.358	2.713

One should note that the available input range is suitable for the peak-to-peak values of the two voltage signals at any measurement frequency. Furthermore, the sampling rate on each channel (maximum value equals to 125 kSa/s) allows the Shannon theorem to be met up to the maximum measurement frequency of 20 kHz. The acquisition process is managed in order to collect $N = 1000$ samples of the two voltage signals v_X and v_S at each frequency of interest.

Among the DAQ board specifications, the parameters CMRR (Common-Mode Rejection Ratio) of the PGA and CT (Cross Talk) are provided, which can be responsible for further uncertainty contributions.

The parameter CMRR is defined in the following way:

$$\text{CMRR} = 20\log_{10}\left(\frac{G_d}{G_{CM}}\right)\ \text{(dB)}. \tag{E6.6.35}$$

where G_d and G_{CM} are the differential and the common-mode gains of the PGA, respectively.
The voltage output of the PGA when the channel CH0 is selected can be represented as:

$$v_{ADC,0} = G_d v_X + G_{CM} v_{CM} \tag{E6.6.36}$$

where v_{CM} is the common-mode voltage at the PGA input:

$$v_{CM} = \frac{v^+ + v^-}{2} = \frac{v_G + v_S}{2} \tag{E6.6.37}$$

The relative measurement error (worst case condition) of the voltage v_X can be obtained as:

$$\epsilon v_X = \frac{G_{CM} v_{CM}}{G_d v_X} = 10^{-\frac{CMRR}{20}} \frac{v_{CM}}{v_X} = 10^{-5} \frac{v_{CM}}{v_X} \tag{E6.6.38}$$

With reference to the voltage values summarized in the table E6.6.1, the relative error ϵv_X due to the CMRR is 15×10^{-6} for frequencies lower than 5 kHz and becomes 11×10^{-6} @ 20 kHz.

The same considerations for the voltage signals v_S bring to an estimation of the relative error ϵv_S due to the CMRR that is equal to 5×10^{-6}, since in this case $v_{CM} = v_S/2$.

The parameter CT is related to the inter-channel phenomena and is defined as:

$$CT = 20\log_{10}\left(\frac{v_{CH,adj}}{v_{adj}}\right) \text{ (dB)} \tag{E6.6.39}$$

where $v_{CH,adj}$ is the voltage error that is superimposed on the voltage v_{CH} of the active channel due to the non perfect insulation with respect to the adjacent channel, where the signal v_{adj} is present. The relative measurement error (worst case condition) of the voltage v_{CH} can be obtained as:

$$\epsilon v_{CH} = \frac{v_{CH,adj}}{v_{CH}} = 10^{\frac{CT}{20}} \frac{v_{adj}}{v_{CH}} \tag{E6.6.40}$$

According to the manufacturer specifications, the parameter CT is equal to −80 dB for adjacent channels and since v_X and v_S have the same order of magnitude, a relative error of almost 100×10^{-6} is expected, which seems to be not negligible with respect to the other contributions. However, the manufacturer provides this parameter at the frequency of 100 kHz and no information is provided about the behaviour of the parameter CT at lower frequencies, which is expected to be better. For this reason, the cross-talk of the DAQ board has been estimated at different frequencies through the experimental set-up described in [509], obtaining the results that are shown in the figure E6.6.3. In the figure, the blue line and the dashed black line represent the measured cross-talk between adjacent channels (CH0 and CH1) and non-adjacent channels (CH0 and CH7). The manufacturer specifications @ 100 kHz are also reported for adjacent (continuous red line) and non-adjacent (dashed red line) channels.

For adjacent channels (the configuration that is shown in the figure E6.6.2), CT values of about −110 dB have been obtained up to 1 kHz, which corresponds to a relative error ϵv_{CH} lower than 3×10^{-6}. In the frequency range from 1 kHz to 6 kHz the parameter CT is lower than −100 dB and the relative error ϵv_{CH} is lower than 10×10^{-6}. Eventually, at the measurement frequencies of 10 kHz and 20 kHz the measured CT was about −96 dB and −90 dB, respectively, and the corresponding relative errors ϵv_{CH} are 19×10^{-6} and 51×10^{-6}, respectively.

The standard uncertainty of the magnitude impedance Z_X is initially estimated without considering the effects of CMRR and CT. For this evaluation, the propagation uncertainty law expressed in the equation (E6.6.27) requires the standard uncertainty of each random variable to be estimated. According to the available specifications, the following values are considered:

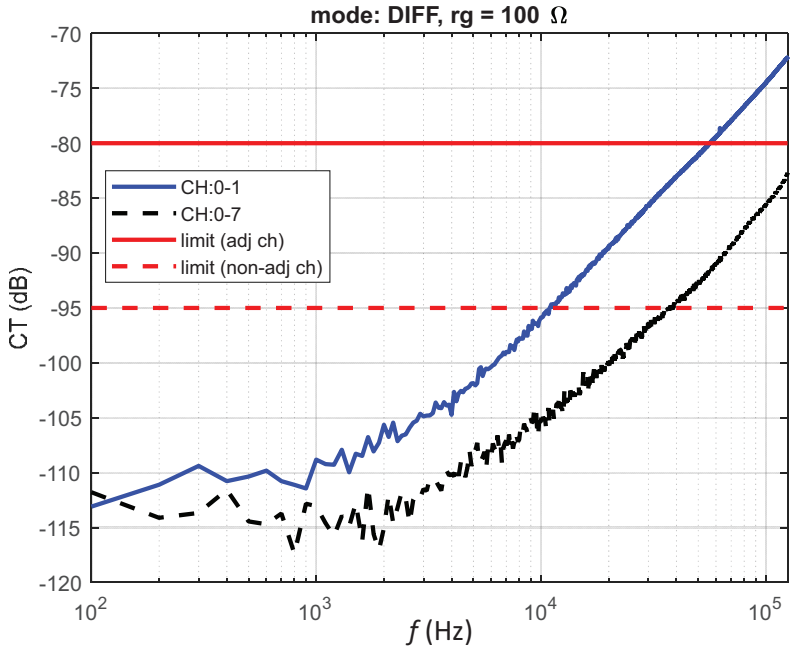


Figure E6.6.3: Experimental results related to the parameter CT of the multiplexed DAQ board. The blue line and the dashed black line represent the measured cross-talk for adjacent and non-adjacent channels; the continuous red line (adjacent channels) and the dashed red line (non-adjacent channels) are the manufacturer specification @ 100 kHz .

- $u(R_S) = 0.2\ \text{m}\Omega$
- $u(\epsilon G) = \frac{\epsilon G_{\max}}{\sqrt{3}} \approx 144 \times 10^{-6}$
- $u(e_{0FF}^2) = \sqrt{2}u^2(e_{\text{OFF}}) = \sqrt{2}\frac{U^2(e_{\text{OFF}})}{4} \approx 8.8 \times 10^{-6}\ \text{V}^2$
- $u(e_{\text{qn,rms}}^2) = \sqrt{2N}u^2(e_{\text{qn,rms}}) = \sqrt{2000}u^2(e_{\text{qn,rms}}) \approx 6.3 \times 10^{-7}\ \text{V}^2$
- $u(n_{\text{mod,X}}) = v_X\frac{118\times10^{-6}}{\sqrt{1000}} \approx [9.3\ 9.3\ 9.3\ 9.3\ 9.4\ 9.5\ 10.1\ 11.5] \times 10^{-6}\ \text{V}^2$ at the different measurement frequencies
- $u(n_{\text{mod,S}}) = v_S\frac{118\times10^{-6}}{\sqrt{1000}} \approx [9.3\ 9.3\ 9.3\ 9.3\ 9.3\ 9.1\ 8.6\ 7.2] \times 10^{-6}\ \text{V}^2$ at the different measurement frequencies

About the random variables e_{OFF}^2 and $e_{\text{qn,rms}}^2$, their expected values are obtained as:

- $E[(e_{0FF}^2] = u^2(e_{\text{OFF}}) \approx 6.25 \times 10^{-6}\ \text{V}^2$
- $E[e_{\text{qn,rms}}^2] = Nu^2(e_{\text{qn,rms}}) \approx 1.4 \times 10^{-5}\ \text{V}^2$

The numerical values of the different terms (squared standard uncertainty contributions) of the expression (E6.6.27) are summarized in the figure E6.6.4 in the frequency range of interest, while the figure E6.6.5 reports the standard uncertainty $u(Z_X)$ in the same frequency range.

The effects of CMRR and CT are now considered interpreting them as uncertainty contributions on the voltage signals v_X and v_S. According to the previous considerations, the relative errors ϵv_X and ϵv_S due to the parameter CMRR have the following values at the different measurement frequencies:

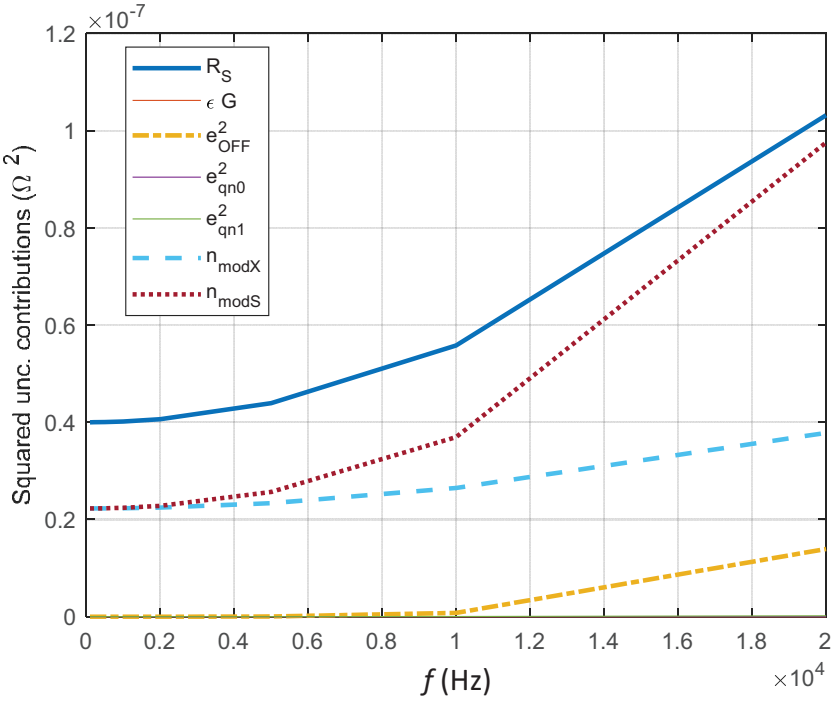


Figure E6.6.4: Numerical values of the squared standard uncertainty contributions of the expression (E6.6.27) in the frequency range of interest.

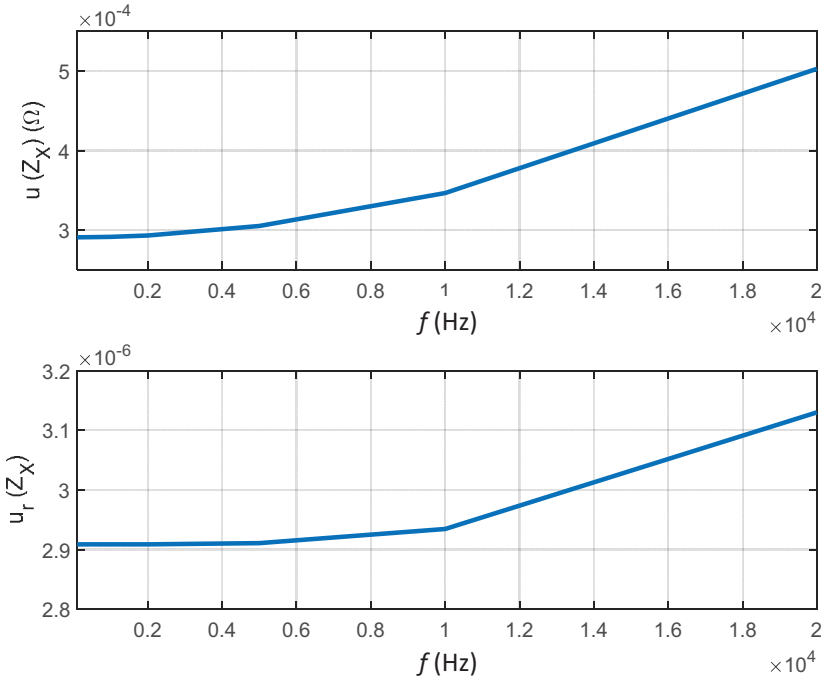


Figure E6.6.5: The standard uncertainty $u(Z_X)$ in the frequency range of interest without considering the effects of CMRR and CT.

– $\epsilon v_X = [15\ 15\ 15\ 15\ 15\ 15\ 14\ 11] \times 10^{-6}$

– $\epsilon v_S = [5.0\ 5.0\ 5.0\ 5.0\ 5.0\ 5.0\ 5.0\ 5.0] \times 10^{-6}$

If the cross-talk effects are taken into account for adjacent channels, the relative error ϵv_{CH} on the two voltage signals have similar values, since the signals have the same order of magnitude:

– $\epsilon v_{CH} = [3.2\ 3.2\ 3.2\ 3.2\ 10\ 10\ 19\ 51] \times 10^{-6}$

The absolute errors due to CMRR and CT are then estimated as:

– $\delta v_{X,CMRR} = \epsilon v_X v_X = [38\ 38\ 38\ 38\ 38\ 38\ 38\ 34]$ µV

– $\delta v_{S,CMRR} = \epsilon v_S v_S = [12\ 12\ 12\ 12\ 12\ 12\ 11\ 9]$ µV

– $\delta v_{X,CT} = \epsilon v_{CH} v_S = [8\ 8\ 8\ 8\ 25\ 24\ 44\ 98]$ µV

– $\delta v_{S,CT} = \epsilon v_{CH} v_X = [8\ 8\ 8\ 8\ 25\ 26\ 51\ 157]$ µV

The assumption of uniform distribution in the range $[-\delta v, +\delta v]$ V for each of the previous error contribution is considered, thus estimating their expected value and standard deviation as:

$$\mu_v = 0;\ \sigma(v) = \frac{\delta v}{\sqrt{3}}. \tag{E6.6.41}$$

Eventually, the absolute standard uncertainties $u(v_X)$ and $u(v_S)$ due to CMRR and CT are obtained as:

$$u(v_X) = \sqrt{\frac{\delta^2 v_{X,CMRR}}{3} + \frac{\delta^2 v_{X,CT}}{3}} = [22\ 22\ 22\ 22\ 26\ 26\ 33\ 60]\ \mu V. \tag{E6.6.42}$$

$$u(v_S) = \sqrt{\frac{\delta^2 v_{S,CMRR}}{3} + \frac{\delta^2 v_{S,CT}}{3}} = [9\ 9\ 9\ 9\ 16\ 16\ 30\ 91]\ \mu V. \tag{E6.6.43}$$

The standard uncertainties $u(v_X)$ and $u(v_S)$ are then multiplied by the corresponding sensitivity coefficients:

$$\frac{\partial Z_X}{\partial v_X} = \frac{R_S^2}{Z_X} \frac{v_X}{v_S^2 + e_{OFF}^2 + e_{qn,rms}^2}. \tag{E6.6.44}$$

$$\frac{\partial Z_X}{\partial v_S} = -\frac{R_S^2}{Z_X} \frac{v_S(v_X^2 + e_{OFF}^2 + e_{qn,rms}^2)}{(v_S^2 + e_{OFF}^2 + e_{qn,rms}^2)^2}. \tag{E6.6.45}$$

obtaining the squared standard uncertainty contributions reported in the figure E6.6.6 and the standard uncertainty $u(Z_X)$ that is shown in the figure E6.6.7.

The same considerations are now repeated taking into account the cross-talk effects for non-adjacent channels, whose approximated values are (see figure E6.6.3):

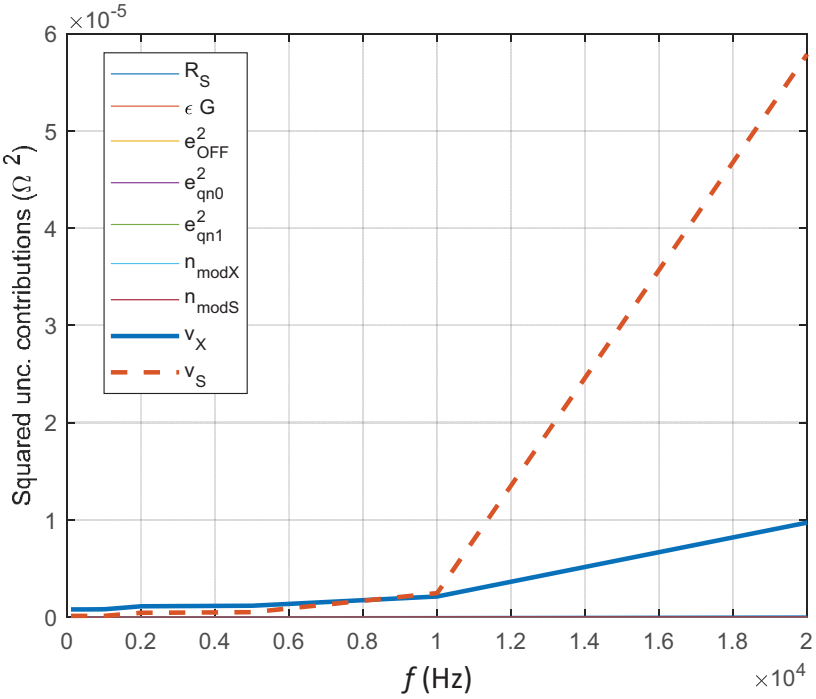


Figure E6.6.6: Numerical values of the squared standard uncertainty contributions of the expression (E6.6.27) and the ones related to CMRR and CT effects for adjacent channels.

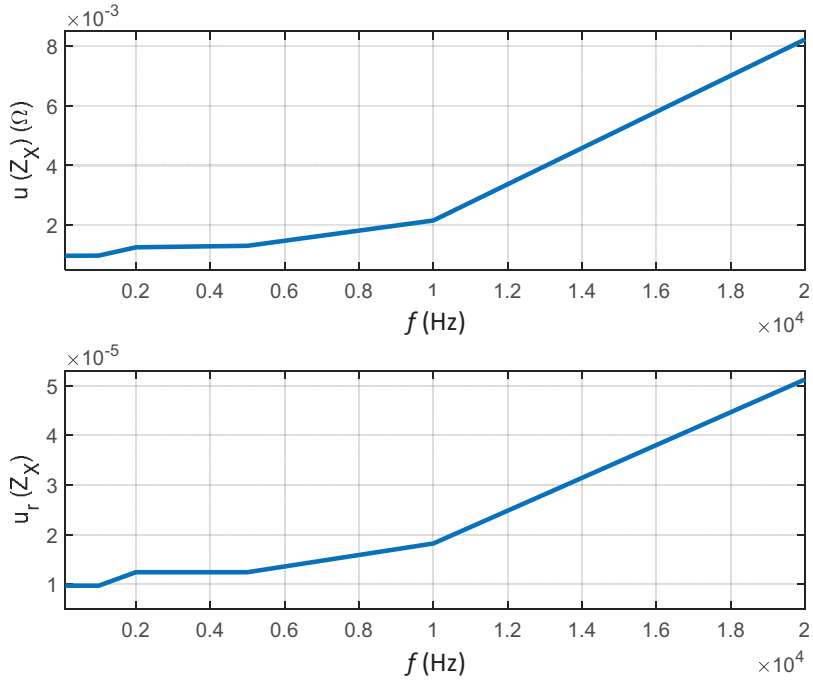


Figure E6.6.7: The standard uncertainty $u(Z_X)$ in the frequency range of interest including the effects of CMRR and CT for adjacent channels.

– $\mathrm{CT}_{\mathrm{non-adj}} = [-112\ -114\ -114\ -114\ -114\ -108\ -105\ -100]\,\mathrm{dB}$

The corresponding relative error $\epsilon v_{\mathrm{CH,non-adj}}$ is estimated as:

$$\epsilon v_{\mathrm{CH,non-adj}} = [2.5\ 2.0\ 2.0\ 2.0\ 2.0\ 4.2\ 6.6\ 16] \times 10^{-6},$$

and the absolute standard uncertainties $u(v_{\mathrm{X,non-adj}})$ and $u(v_{\mathrm{S,non-adj}})$ becomes:

$$u(v_{\mathrm{X,non-adj}}) = [22\ 22\ 22\ 22\ 22\ 23\ 24\ 26]\ \mu\mathrm{V} \tag{E6.6.46}$$

$$u(v_{\mathrm{S,non-adj}}) = [8\ 8\ 8\ 8\ 8\ 9\ 12\ 29]\ \mu\mathrm{V}. \tag{E6.6.47}$$

Figure E6.6.8 shows the squared standard uncertainty contributions for the multiplexed DAQ board that uses non-adjacent channels. Eventually, the standard uncertainty $u(Z_{\mathrm{X}})$ in the same configuration is shown in the figure E6.6.9.

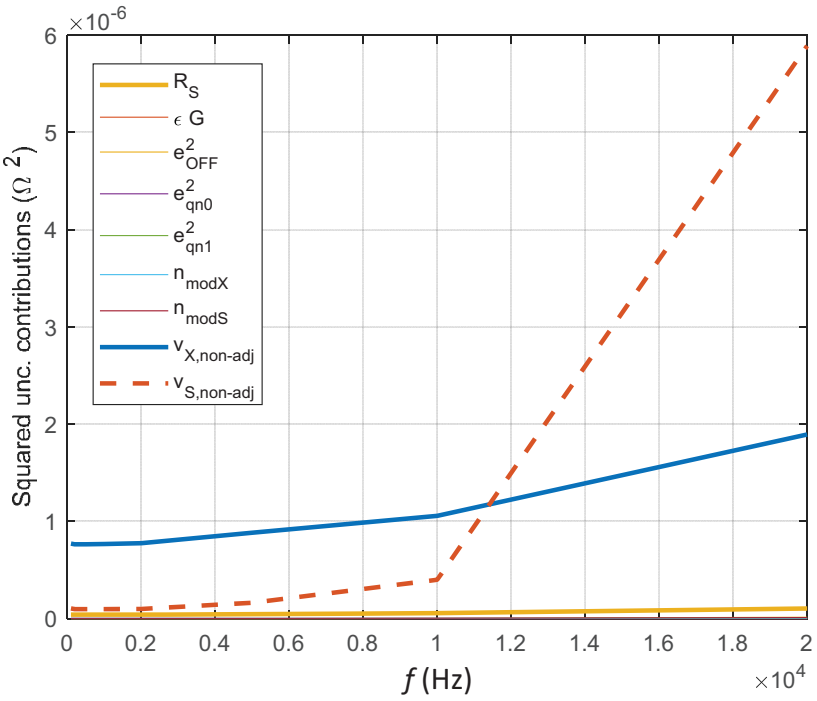


Figure E6.6.8: Numerical values of the squared standard uncertainty contributions of the expression (E6.6.27) and the ones related to CMRR and CT effects for non-adjacent channels.

E6.6.7 Interpretation of results

If the effects related to the parameters CMRR and CT of the multiplexed DAQ board are not taken into account, the uncertainty contributions that are summarized in the figure E6.6.4 have to be analyzed. In this case, the main contributions are the ones related to the standard resistor R_{S} and to the two random variables $n_{\mathrm{mod,X}}$ and $n_{\mathrm{mod,S}}$, which depend on the noise superimposed to the acquired signals. A small contribution is also given by the offset error of the measuring chain at frequencies higher than 10 kHz, where the signals v_{X} and v_{S} becomes significantly different and then the offset-error compensation is less effective. The compensation remains instead always effective for the gain error, which gives a negligible contribution. The contributions of the random

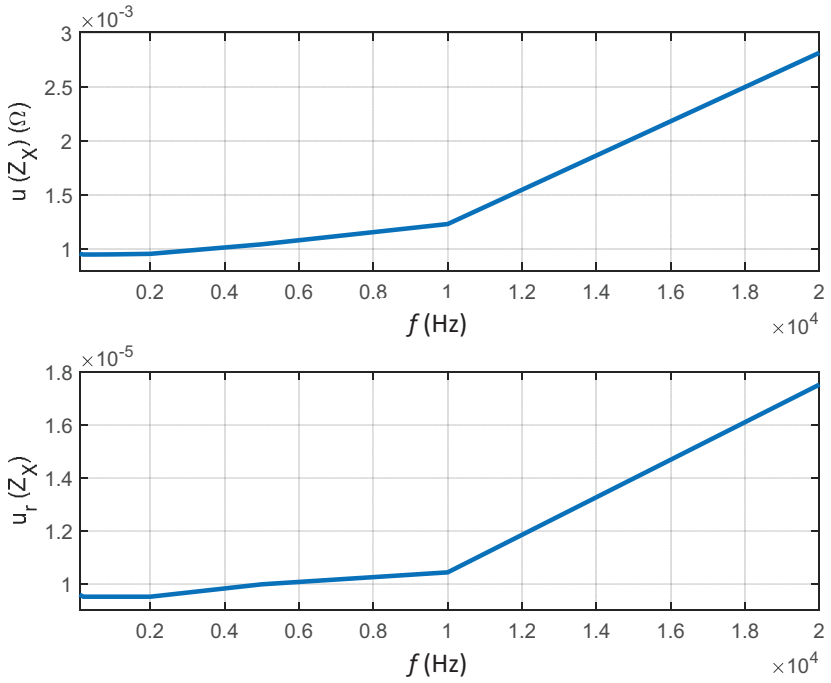


Figure E6.6.9: The standard uncertainty $u(Z_X)$ in the frequency range of interest including the effects of CMRR and CT for non-adjacent channels.

variables e^2_{qn0} and e^2_{qn1} are negligible in the whole frequency range. The figure E6.6.5, where absolute and relative standard uncertainties of the measurand Z_X are reported, shows that the implemented technique allows relative uncertainties of the order of few part per million (ppm) to be obtained even though the gain error of the DAQ board is two order of magnitude higher. The most effective intervention to reduce further the uncertainty consists in using a more accurate standard resistor, but its relative uncertainty (2×10^{-6}) is close to the uncertainty of primary-level standards. Another possibility is the acquisition of a higher number N of samples in order to reduce the contributions $u(n_{mod,X})$ and $u(n_{mod,S})$, even though it is not very convenient if the contribution related to R_S is not reduced.

When the contributions of Common Mode Rejection Ration and Cross Talk are estimated, the main outcome is that these effects are the main factors that limit the overall uncertainty. The obtained results suggest to avoid the use of adjacent channels, since in this configuration (see figure E6.6.6) the uncertainty of the estimated rms values v_X and v_S becomes the most important contributions and the overall relative standard uncertainty $u_r(Z_X)$ (see figure E6.6.7) becomes almost an order of magnitude higher than the previous obtained value.

If non-adjacent channels are instead used, CMRR and CT effects remain the main contributions (see figure E6.6.8), but they are of the same order of magnitude of the other contributions up to 10 kHz. About the relative standard uncertainty $u_r(Z_X)$ (see figure E6.6.9), it is lower than 11×10^{-6} up to 10 kHz and becomes about 18×10^{-6}at 20 kHz.

In conclusion, it can be stated that the proposed measurement technique is very effective, since using a low-cost commercial multiplexed DAQ board it is possible to obtain a relative uncertainty of the measurand Z_X that is almost an order of magnitude higher than the uncertainty of the used standard resistor R_S, which is the most expensive component of the proposed circuit. In order to obtain a relative uncertainty that is of the same order of magnitude of $u(R_S)$ it is necessary to

use a DAQ board with independent channels (non multiplexed), thus minimizing the cross-talk effects, and with a PGA characterized by a higher CMRR. This also mean that the cost of the system becomes about an order of magnitude higher than the previous one.

Example E6.7

Temperature measurement with a micro-controller based board

A. Carullo, M.G. Cox

E6.7.1 Summary

This example refers to the measurement of the temperature inside a climatic chamber by means of a measuring chain that includes a Resistive Temperature Detector (RTD), a simple conditioning circuitry and a commercial micro-controller (µ-C) based board. The advantages that are related to the characterization of the Analogue-to-Digital Converter internal to the µ-C board and the implementation of a dithering technique are highlighted. The measurement uncertainty is evaluated according to the GUM uncertainty framework (GUF).

E6.7.2 Introduction to the application

The application of thermal cycles inside a climatic chamber are very common tests that are performed to verify the correct behaviour of a device in the operating temperature range stated by the manufacturer. During these tests, the air temperature inside the climatic chamber as well as the temperature of different parts of the device under test are measured using different type of sensors, such as thermistors, resistive temperature detectors, thermocouples or integrated electronic devices.

In this example, a platinum resistive thermal detector Pt100 is considered as the device that converts the measured temperature θ into an electrical resistance R according to the Callendar-Van Dusen equation[1] [510]:

$$R = R_0(1 + A\theta + B\theta^2), \tag{E6.7.1}$$

where

$$R_0 = 100\,\Omega,$$
$$A = 3.9083 \times 10^{-3}\,^{\circ}\mathrm{C}^{-1},$$
$$B = -5.775 \times 10^{-7}\,^{\circ}\mathrm{C}^{-2}.$$

[1] The measurement of temperatures higher than 0 °C is here considered.

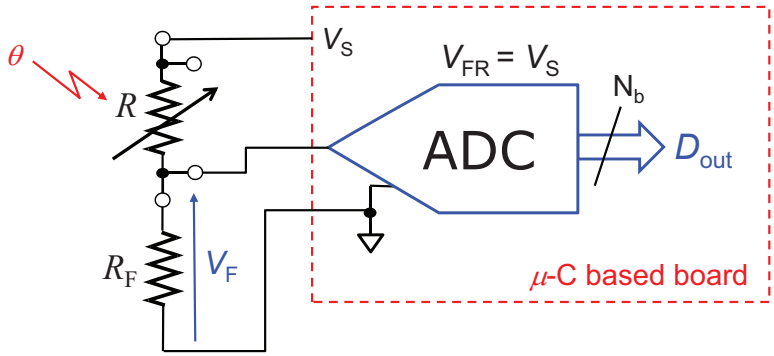


Figure E6.7.1: The measuring chain of the temperature sensed through the sensor Pt100.

A 4-wire class A [510] Pt100 based on the film technological process is available, which is characterized by a time constant $\tau = 0.1\,\text{s}$, a thermal resistance of $20\,^{\circ}\text{C}\,\text{W}^{-1}$, a temperature range $-30\,^{\circ}\text{C}$ to $300\,^{\circ}\text{C}$ and a maximum admitted error expressed as:

$$\delta\theta_{\text{sensor}} = \pm(0.15 + 0.002|\theta|)\,^{\circ}\text{C}. \tag{E6.7.2}$$

The resistance change at the output of the sensor Pt100 is converted into a voltage signal by means of a simple conditioning circuitry, which is based on a voltage divider. The voltage output of this circuit is converted into a digital code through the Analogue-to-Digital Converter (ADC) internal to a μ-C based board. The scheme of the whole measuring chain is shown in the figure E6.7.1, where R is the sensor resistance, R_{F} is a fixed known resistance, the ADC has a unipolar range and its reference voltage V_{FR} is set to the same value of the voltage V_{S} (nominal value 5 V) that supplies the voltage divider, thus obtaining a ratiometric measurement. The main specifications of the ADC are:

- maximum sampling rate: $f_{\text{s,max}} = 76.9\,\text{kSa/s}$
- resolution: $N_{\text{b}} = 10\,\text{bit}$ (8 bit if $f_{\text{s}} > 15\,\text{kSa/s}$)
- Integral Non Linearity: $\text{INL} = \pm 0.5\,\text{LSB}$
- Total Unadjusted Error: $\text{TUE} = \pm 2\,\text{LSB}$
- input resistance: $R_{\text{IN}} = 100\,\text{M}\Omega$

E6.7.3 Specification of the measurand

The thermal cycle that is considered in this example is shown in figure E6.7.2, where the red line represents the temperature profile the device under test is subjected to.

The quantity under measurement is the air temperature inside the climatic chamber, which is included in the range 10 °C to 70 °C. Information about the temperature rate during the different parts of the test is also provided in figure E6.7.2, where the highest rate is equal to $1\,^{\circ}\text{C}\,\text{s}^{-1}$ in the time interval from 12.0 min to 12.5 min. Other measurement requirements include the standard uncertainty, which is 2.0 °C, and the temperature resolution, which has to be not higher than 4.0 °C.

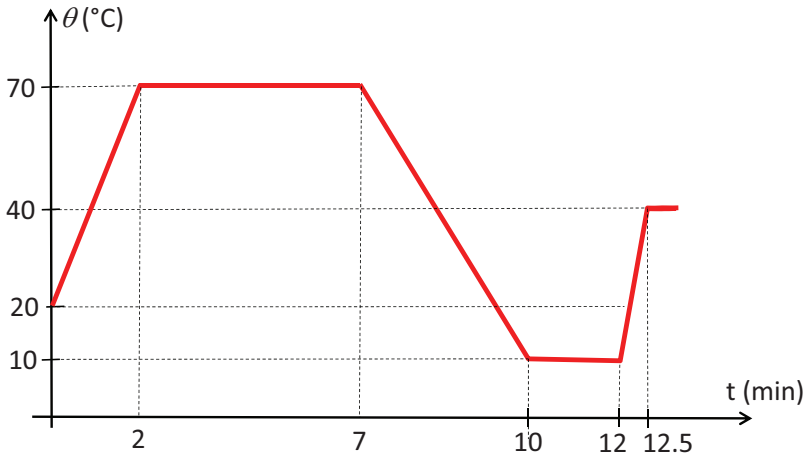


Figure E6.7.2: The temperature profile to be measured inside a climatic chamber.

E6.7.4 Measurement model

The measurement circuit of figure E6.7.1 has been arranged using a resistor R_F with a nominal value of 1 kΩ and a relative standard uncertainty of $u_r(R_F) = 1 \times 10^{-3}$. The value of R_F has been set taking into account the sensitivity $S^{\theta}_{V_F}$ of the voltage output V_F with respect to the input quantity θ and the self-overheating of the sensor Pt100.

The sensitivity $S^{\theta}_{V_F}$ is defined as:

$$S^{\theta}_{V_F} = \frac{\partial V_F}{\partial \theta} \quad (\mathrm{V/^\circ C}), \tag{E6.7.3}$$

where the voltage V_F is obtained as:

$$V_F = V_S \frac{R_F}{R_F + R} = V_S \frac{R_F}{R_F + R_0(1 + A\theta + B\theta^2)}. \tag{E6.7.4}$$

Calculating the partial derivative of the previous expression with respect to the temperature, the sensitivity (E6.7.3) can be expressed as:

$$S^{\theta}_{V_F} = -V_S \frac{R_F R_0 (A + 2B\theta)}{[R_F + R_0(1 + A\theta + B\theta^2)]^2}. \tag{E6.7.5}$$

Expression (E6.7.5) shows that a high sensitivity can be obtained using a high voltage V_S, as expected. However, the supply voltage is limited by the self-overheating of the sensor. In addition, it can be shown that the second term of the expression (E6.7.5) has its extreme value when $R_F = R_0$, thus suggesting a nominal value of 100 Ω for the fixed resistance. In this condition, the maximum voltage V_S that allows the self-overheating to be neglected is 1 V and the corresponding sensitivity is of about $-0.8\,\mathrm{V\,^\circ C^{-1}}$.

Another possible choice consists in selecting a value of the fixed resistor R_F higher than R_0, thus having the possibility to set values of the voltage V_S higher than 1 V. This is the design criterion used in this example, since the supply voltage provided by the µ-C based board is $V_S = 5$ V. The value of the fixed resistor R_F has been set in order to obtain a sensor self-overheating not higher than $(\delta\theta_{sensor}/4)$, where $\delta\theta_{sensor}$ is obtained by means of the expression (E6.7.2). Figure E6.7.3 shows the comparison between the sensor uncertainty $\delta\theta_{sensor}$ (blue line) and the sensor self-overheating $\delta\theta_{self}$ (black line) for a value of the resistor R_F that is equal to 1 kΩ, where $\delta\theta_{self}$ has

been evaluated by multiplying the power on the sensor R by its thermal resistance ($20\,°C\,W^{-1}$). One should note that the sensor self-overheating is always lower than the fixed threshold (red line), which is equal to $\delta\theta_{sensor}/4$.

Once the conditioning circuitry has been designed, the expected voltage V_F can be evaluated according to the equation (E6.7.4), thus obtaining the result that is shown in the top chart of figure E6.7.4. The bottom chart of the same figure shows the sensitivity of V_F with respect to the measured temperature, which is evaluated by the expression (E6.7.5) in the range from about $-1.6\,mV\,°C^{-1}$ to $-1.5\,mV\,°C^{-1}$.

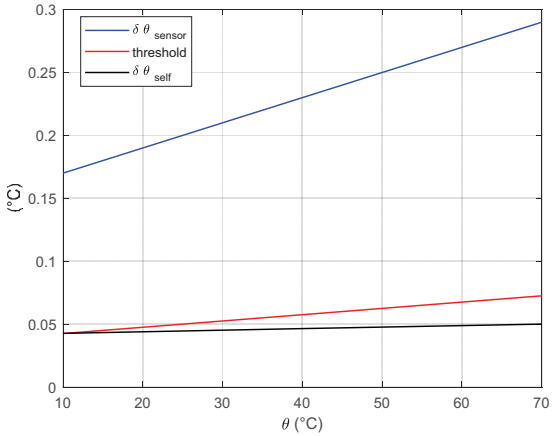


Figure E6.7.3: The effect of the sensor self-overheating (black line) compared to the sensor uncertainty (blue line) when $R_F = 1\,k\Omega$ and $V_S = 5\,V$.

The Input/Output (I/O) relationship of the whole measuring chain that has the temperature θ as the input quantity and the code D_{out} provided by the ADC as the output quantity (see figure E6.7.1) is:

$$D_{out} = \frac{V_F}{V_q} = \frac{V_S}{V_q}\frac{R_F}{R_F + R} = \frac{V_S}{V_q}\frac{R_F}{R_F + R_0(1 + A\theta + B\theta^2)}. \tag{E6.7.6}$$

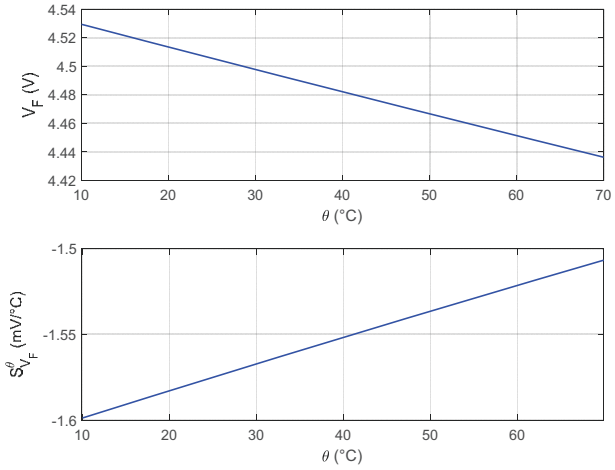


Figure E6.7.4: The output voltage of the conditioning circuitry (top chart) and its sensitivity with respect to the measured temperature (bottom chart) when $R_F = 1\,k\Omega$ and $V_S = 5\,V$.

Since the reference voltage of the ADC is the same voltage V_S that supplies the conditioning circuitry, the quantization voltage V_q is obtained as:

$$V_q = \frac{V_S}{2^{N_b}}. \tag{E6.7.7}$$

Replacing the expression (E6.7.7) in the equation (E6.7.6), the output code of the ADC is obtained as:

$$D_{out} = \frac{\cancel{V_S}}{\cancel{V_S}} \frac{R_F 2^{N_b}}{R_F + R} = \frac{R_F 2^{N_b}}{R_F + R_0(1 + A\theta + B\theta^2)}, \tag{E6.7.8}$$

that is then independent of the voltage V_S (radiometric measurement), which has no effect in the uncertainty budget of the measuring chain.

Eventually, inverting the I/O relationship (E6.7.8), the calibration function (O/I relationship) of the whole measuring chain is obtained:

$$\theta = \frac{-A}{2B} - \sqrt{\frac{A^2}{4B^2} - \frac{1}{R_0 B}\left(R_0 + R_F - R_F \frac{2^{N_b}}{D_{out}}\right)}. \tag{E6.7.9}$$

The calibration function (E6.7.9) has to be implemented in the firmware of the micro-controller in order to provide the temperature measurements starting from the nominal values of the sensor parameters A, B and R_0, the actual value of the resistance R_F and the output code D_{out} of the ADC; then it represents the measurement model.

E6.7.5 Uncertainty propagation

The measurement model (E6.7.9) allows the main uncertainty contributions to be identified, which are:

- the sensor uncertainty, which is evaluated through expression (E6.7.2);
- the tolerance of the fixed resistor R_F;
- the uncertainty of the code D_{out} at the output of the ADC.

The last term depends on offset and gain errors E_{off} and E_G of the ADC, on its Integral Non Linearity (INL) and on the quantization error E_q. If none of the first three error terms of the ADC is compensated, the total unadjusted error provided by the manufacturer (maximum admitted error ±2 LSB) is considered as a reliable evaluation of the effects of the errors E_{off}, E_G and INL. For the quantization error E_q, if single readings of the ADC are processed according to the model (E6.7.9), its value can be considered in the range −0.5 LSB to 0.5 LSB.

The sensitivity coefficient of θ with respect to D_{out} can be provided using expression (E6.7.9) as:

$$S_\theta^{D_{out}} = \frac{\partial \theta}{\partial D_{out}} = \frac{\dfrac{-2^{N_b} R_F}{R_0 B D_{out}^2}}{2\sqrt{\dfrac{A^2}{4B^2} - \dfrac{1}{R_0 B}\left(R_0 + R_F - R_F \dfrac{2^{N_b}}{D_{out}}\right)}}, \tag{E6.7.10}$$

that has the behaviour shown in figure E6.7.5 for the two possible configurations of the ADC, i.e. $N_b = 10$ for sampling rate $f_s \leq 15\,\text{kSa/s}$ (blue line) or $N_b = 8$ for $f_s > 15\,\text{kSa/s}$ (red line). The important difference between the two configurations requires the sampling rate of the ADC to be set before the uncertainty evaluation is performed.

As observed from figure E6.7.2, the highest rate during the thermal test is equal to $1\,°\text{C}\,\text{s}^{-1}$ and since the resolution requirement is 0.4 °C, the minimum sampling rate is obtained as:

$$f_{S,\min} = \frac{\text{Rate}_{\max}}{\text{Resolution}} = \frac{1\,°\text{C}\,\text{s}^{-1}}{0.4\,°\text{C}} = 2.5\,\text{Sa/s}. \tag{E6.7.11}$$

The ADC can be then configured to work with a sampling rate lower than 15 kSa/s, thus taking advantage of its maximum resolution, i.e. $N_b = 10$. However, in single-reading mode the ADC resolution is $\Delta D_{out} = 1\,\text{LSB}$, which corresponds to a temperature resolution that is equal to $\Delta\theta = S_\theta^{D_{out}} \Delta D_{out} \approx 3\,°\text{C}$, which does not allow the resolution requirement to be met. Furthermore, due to the total unadjusted error of the ADC (±2 LSB) it is not possible to meet the temperature uncertainty requirement, which is equal to 2 °C (absolute standard uncertainty).

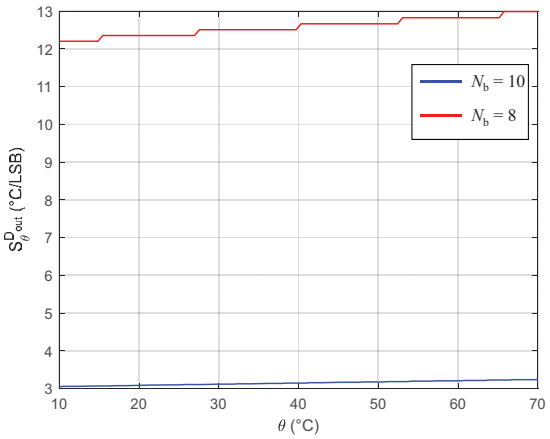


Figure E6.7.5: The sensitivity coefficient of θ with respect to D_{out} in the temperature range of interest for ADC resolution equal to 10 bit (blue line) and 8 bit (red line).

The proposed system cannot be used *as is*, but a suitable characterization is required in order to compensate for offset and gain errors of the ADC, thus reducing its uncertainty contribution to the parameter INL, which is stated to be not higher than 0.5 LSB. In addition, also the ADC resolution has to be decreased since it is one of the main uncertainty contribution.

About the ADC characterization, a two-step procedure is proposed that is based on the evaluation of the offset error D_{off} and the actual quantization step $V_{q,act}$. In the first step (right side of figure E6.7.6), a short circuit is connected to the input of the ADC and the output code D_{off} is stored, which is the evaluation of the ADC offset error. In the second step (left side of figure E6.7.6), a known voltage V_1 is applied to the ADC input and the corresponding output code D_1 is acquired, thus evaluating the actual quantization step as:

$$V_{q,act} = \frac{V_1}{D_1}. \tag{E6.7.12}$$

In order to maintain the advantage of a ratiometric measurement, the voltage V_1 is obtained by means of a voltage divider made up of two resistors R_1 and R_2 that are supplied by the voltage V_S. In addition, the nominal values of the two resistors are set in order to obtain a voltage input of

the ADC that is very similar to the voltage V_F (see top chart of figure E6.7.4), that are $R_1 = 10\ \text{k}\Omega$ and $R_2 = 1\ \text{k}\Omega$. In this condition, the expression (E6.7.12) can be rewritten as:

$$V_{q,act} = \frac{V_S}{D_1}\frac{R_1}{R_1 + R_2}. \tag{E6.7.13}$$

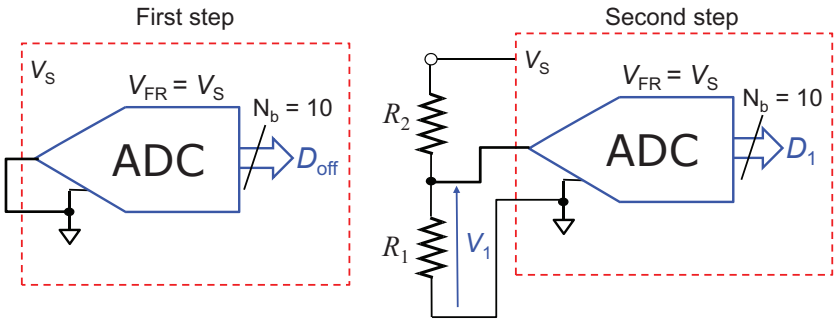


Figure E6.7.6: The two steps of the ADC characterization: in the first step the ADC offset error is evaluated, while in the second step the actual quantization voltage of the ADC is obtained.

The new I/O relationship of the characterized measuring chain can be expressed as

$$D_{out} = \frac{\cancel{V_S}}{\cancel{V_S}}\left(1 + \frac{R_2}{R_1}\right) D_1 \frac{R_F}{R_F + R} + D_{off}, \tag{E6.7.14}$$

and the calibration function becomes:

$$\theta = \frac{-A}{2B} - \sqrt{\frac{A^2}{4B^2} - \frac{1}{R_0 B}\left(R_0 + R_F - \frac{K D_1 R_F}{D_{out} - D_{off}}\right)}, \tag{E6.7.15}$$

where $K = 1 + R_2/R_1$.

In the new measurement model (E6.7.15), the uncertainty contributions related to the sensor and the fixed resistor R_F remain the same, the new terms K, D_{off} and D_1 are present, while for the term D_{out} the parameter INL and the quantization error of the ADC have to be taken into account.

About the term K (nominal value 1.1), if the two resistors R_1 and R_2 are known with a relative standard uncertainty $u_r(R_1) = u_r(R_2) = 1 \times 10^{-3}$, it can be easily shown[2] that the absolute standard uncertainty of the term K is $u(K) \approx 1.4 \times 10^{-4}$.

The uncertainty of the terms D_{off} and D_1 obtained during the characterization of the ADC mainly depends on the reading resolution and repeatability. The ADC resolution is also responsible for the uncertainty of the term D_{out} and, as previously shown, the value of 1 LSB that corresponds to a single-reading mode does not allow to meet either the temperature resolution requirement or the uncertainty requirement.

With the aim of reducing the ADC resolution, a dithering technique is implemented that is based on the evaluation of the mean value of multiple readings. The effectiveness of this technique in reducing the ADC resolution is related to the number M of multiple readings that are acquired at the output of the ADC and mainly to the characteristics of the noise that is superimposed on the sampled signal [511]. Such a noise must have a zero mean value, in order not to bias

[2] The correlation between the measurements of R_1 and R_2 is considered negligible.

the sampled signal, and a standard deviation that is similar to the quantization step V_q of the ADC. The dithering technique is often used in data acquisition boards, where an internal noise generator can be turned on when multiple readings are acquired and averaged, and in digital storage oscilloscope (high-resolution mode), taking advantage of the internal noise of the input analogue conditioning circuitry.

One should note that the implementation of the dithering technique requires to over-sample the signal at the input of the ADC, i.e. a sampling rate that is much higher than the Nyquist limit has to be set; otherwise the average processing of multiple readings acts as a low-pass filter and then modifies the behaviour of the sampled signal. In this example, a minimum sampling rate has been evaluated by means of the expression (E6.7.11), which is 2.5 Sa/s, that sets the over-sampling conditions. If the sampling rate of the ADC internal to the µ-C board is set to 10 kSa/s, a maximum number of samples $M = 4000$ can be acquired and averaged over a time interval of 0.4 s, thus ensuring the compliance to the minimum sampling rate.

If a zero mean-value noise with a normal distribution and a standard deviation σ_n is superimposed on the sampled signal V_F, the number of bit of resolution of the ADC is improved of the quantity N_b^+ that can be expressed as[3]:

$$N_b^+ = \frac{1}{2}\log_2\left[\frac{M}{12(\sigma_n/V_q)^2+1}\right]. \tag{E6.7.16}$$

With the assumption of a unitary ratio (σ_n/V_q), according to the previous expression the quantity N_b^+ is in the range from about 2 bit to about 3 bit for a number of samples M equal to 200 and 800, respectively.

In the proposed system, the dithering technique is implemented taking advantage of the internal noise of the µ-C analogue circuitry. The standard deviation of such a noise has been evaluated during the characterization of offset and gain errors of the ADC. The figure E6.7.7 shows the code D_1 acquired during 1000 single conversions of the ADC. This result refers to the second step of the ADC characterization (see right side of figure E6.7.6) and to the nominal value of the resistor R_1 and R_2 that are equal to 10 kΩ and 1 kΩ, respectively. The experimental standard deviation σ_{n+e_q} of the acquired samples, which includes the effects of the internal noise n of the analogue circuitry and the quantization noise e_q, has a value of about 0.87 LSB. Since the quantization noise is characterized by a standard deviation $\sigma_{e_q} = 1/\sqrt{12}$ LSB, the standard deviation σ_n of the noise can be obtained as

$$\sigma_n = \sqrt{\sigma_{n+e_q}^2 - \sigma_{e_q}^2} = \sqrt{0.87^2 - \frac{1}{12}} \approx 0.82\,\text{LSB}. \tag{E6.7.17}$$

The ratio σ_n/V_q is then equal to 0.82, which is suitable for an effective implementation of the dithering technique.

The sampling rate of the ADC has been set to about[4] 9615 Sa/s and 1000 sets of 200 samples have been acquired and averaged, obtaining the result that is shown in figure E6.7.8 during the characterization of the ADC.

The result of figure E6.7.8 shows that averaging 200 multiple readings of the ADC allows the resolution ΔD_{out} to be decreased to 0.1 LSB, that corresponds to a temperature resolution of about 0.3 °C.

[3] This expression can be considered valid for values of the ratio (σ_n/V_q) greater than 0.5.

[4] This value derives from the main clock of the µ-C board divided by a integer division factor.

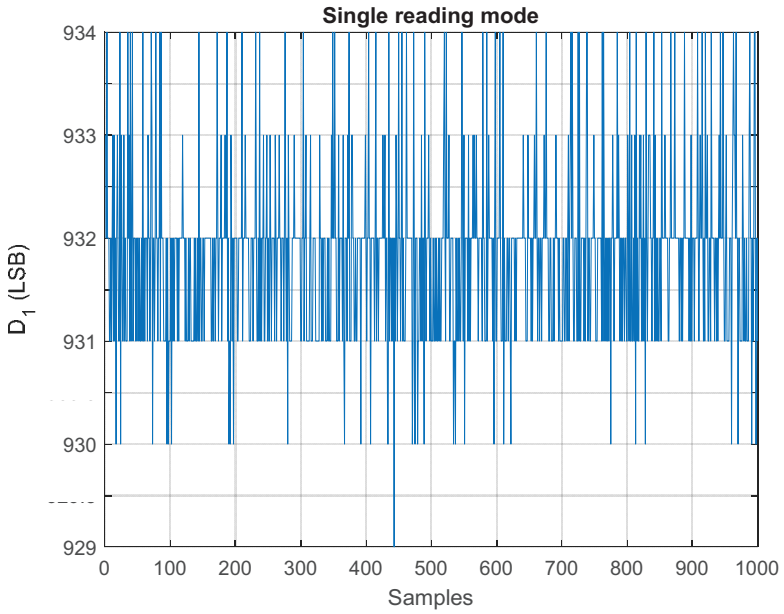


Figure E6.7.7: Characterization of the gain error of the ADC in single conversion mode.

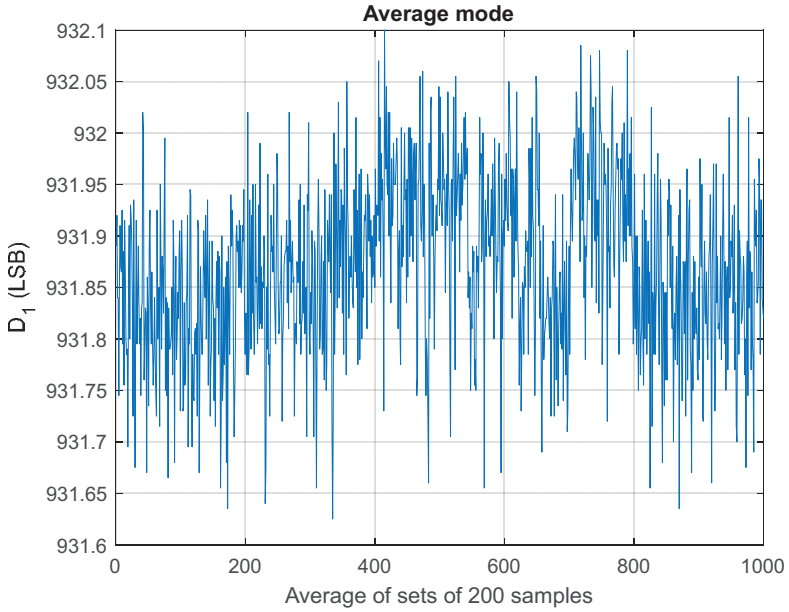


Figure E6.7.8: Characterization of the gain error of the ADC in free-running mode (average of sets of 200 samples).

The mean value and the experimental standard deviation of the parameter D_1 have been evaluated as $\bar{D}_1 = 931.87$ and $\sigma_{D_1} = u(D_1) \approx 0.09$, respectively.

During the first step of the ADC characterization, the parameter D_{off} has been evaluated as 1.00 LSB with a standard uncertainty $u(D_{\text{off}}) \approx 0.09$ LSB.

Using the expression (E6.7.15) as the measurement model, where D_{out} is obtained as the average of 200 multiple readings, the standard uncertainty of the temperature θ is evaluated under the assumption of negligible correlation among the random variables R_{F}, K, D_{off}, D_1, D_{out} and the parameters of the sensor Pt100, thus obtaining:

$$u^2(\theta) = \left(\frac{\partial\theta}{\partial R_{\text{F}}}\right)^2 u^2(R_{\text{F}}) + \left(\frac{\partial\theta}{\partial K}\right)^2 u^2(K) + \left(\frac{\partial\theta}{\partial D_{\text{off}}}\right)^2 u^2(D_{\text{off}}) + \tag{E6.7.18}$$

$$+ \left(\frac{\partial\theta}{\partial D_1}\right)^2 u^2(D_1) + \left(\frac{\partial\theta}{\partial D_{\text{out}}}\right)^2 u^2(D_{\text{out}}) + u^2(\theta_{\text{sensor}}), \tag{E6.7.19}$$

where[5]

$$\frac{\partial\theta}{\partial R_{\text{F}}} = \frac{\dfrac{1}{R_0 B}\left(1 - \dfrac{KD_1}{D_{\text{out}} - D_{\text{off}}}\right)}{2\sqrt{\dfrac{A^2}{4B^2} - \dfrac{1}{R_0 B}\left(R_0 + R_{\text{F}} - \dfrac{KD_1 R_{\text{F}}}{D_{\text{out}} - D_{\text{off}}}\right)}}, \tag{E6.7.20}$$

$$\frac{\partial\theta}{\partial K} = \frac{\dfrac{-D_1 R_{\text{F}}}{R_0 B(D_{\text{out}} - D_{\text{off}})}}{2\sqrt{\dfrac{A^2}{4B^2} - \dfrac{1}{R_0 B}\left(R_0 + R_{\text{F}} - \dfrac{KD_1 R_{\text{F}}}{D_{\text{out}} - D_{\text{off}}}\right)}}, \tag{E6.7.21}$$

$$\frac{\partial\theta}{\partial D_{\text{off}}} = \frac{\dfrac{-KD_1 R_{\text{F}}}{R_0 B(D_{\text{out}} - D_{\text{off}})^2}}{2\sqrt{\dfrac{A^2}{4B^2} - \dfrac{1}{R_0 B}\left(R_0 + R_{\text{F}} - \dfrac{KD_1 R_{\text{F}}}{D_{\text{out}} - D_{\text{off}}}\right)}}, \tag{E6.7.22}$$

$$\frac{\partial\theta}{\partial D_1} = \frac{\dfrac{-KR_{\text{F}}}{R_0 B(D_{\text{out}} - D_{\text{off}})}}{2\sqrt{\dfrac{A^2}{4B^2} - \dfrac{1}{R_0 B}\left(R_0 + R_{\text{F}} - \dfrac{KD_1 R_{\text{F}}}{D_{\text{out}} - D_{\text{off}}}\right)}}, \tag{E6.7.23}$$

$$\frac{\partial\theta}{\partial D_{\text{out}}} = \frac{\dfrac{KD_1 R_{\text{F}}}{R_0 B(D_{\text{out}} - D_{\text{off}})^2}}{2\sqrt{\dfrac{A^2}{4B^2} - \dfrac{1}{R_0 B}\left(R_0 + R_{\text{F}} - \dfrac{KD_1 R_{\text{F}}}{D_{\text{out}} - D_{\text{off}}}\right)}}. \tag{E6.7.24}$$

E6.7.6 Reporting the result

The results here reported, which refer to the temperature range from 10 °C to 70 °C, have been obtained using the MATLAB script pt100_emue.m. Other temperature intervals and other input data can be used accessing to the variables in the script according to the instructions explained in the help section.

[5] The sensitivity coefficients have been analytically obtained using the basic rules of derivatives [508].

The absolute standard uncertainties of the input quantities R_F, K, D_{off} and D_1 are here summarized:

$$\begin{aligned} u(R_F) &= 1\,\Omega \\ u(K) &= 1.4 \times 10^{-4} \\ u(D_{off}) &= 0.09\,\text{LSB} \\ u(D_1) &= 0.09\,\text{LSB} \end{aligned}$$

About the code D_{out}, which is obtained as the average of 200 multiple readings of the ADC, the main uncertainty contributions are related to the Integral Non Linearity (INL) of the ADC and to the standard deviation $\sigma_{D_{out}}$ of the repeated readings. The first contribution, which is stated by the manufacturer as 0.5 LSB, is associated to a uniform probability density function, thus obtaining a corresponding standard uncertainty $u(\text{INL})$ that is equal to:

$$u(\text{INL}) = \frac{1}{\sqrt{12}} \approx 0.29\,\text{LSB}. \tag{E6.7.25}$$

The standard deviation $\sigma_{D_{out}}$ is evaluated as the same value obtained during the ADC characterization (0.09 LSB), thus obtaining the standard uncertainty $u(D_{out})$ as:

$$u(D_{out}) = \sqrt{u^2(\text{INL}) + \sigma^2_{D_{out}}} = \sqrt{0.29^2 + 0.09^2} \approx 0.30\,\text{LSB}. \tag{E6.7.26}$$

Eventually, the contribution due to the sensor is evaluated starting from its maximum admitted error $\delta\theta_{sensor}$ (see the expression (E6.7.2)) and making also for this contribution the assumption of uniform probability density function, thus obtaining:

$$u(\theta_{sensor}) = \frac{\delta\theta_{sensor}}{\sqrt{3}}\ (^\circ\text{C}). \tag{E6.7.27}$$

The numerical values of the different terms (squared standard uncertainty contributions) of the expression (E6.7.18) are summarized in figure E6.7.9 in the temperature range of interest, while figure E6.7.10 reports the standard uncertainty $u(\theta)$ in the same temperature range.

E6.7.7 Interpretation of results

From an analysis of figure E6.7.9, the first outcome is that the main uncertainty contributions is that related to the code D_{out}, which is due to the Integral Non Linearity of the ADC. The other contributions are relatively negligible, thus suggesting the possibility to relax the uncertainty requirements of the fixed resistor R_F and of the fixed resistors R_1 and R_2 used during the ADC characterization. The same figure shows that the lowest contribution is that related to the sensor Pt100; then it is possible to use a cheap sensor that belongs to a worse uncertainty class than class A, thus reducing the cost of the whole system.

Figure E6.7.10 eventually highlights that the evaluated standard uncertainty meets the measurement requirement (2 °C), thanks to

- the minimization of the sensor self-overheating;
- the use of a ratiometric conditioning circuitry;
- the compensation of offset and gain errors of the ADC internal to the micro-controller[6];

[6]Note that in this example the contributions related to possible drifts of the components of the measuring chain after the characterization have not been taken into account.

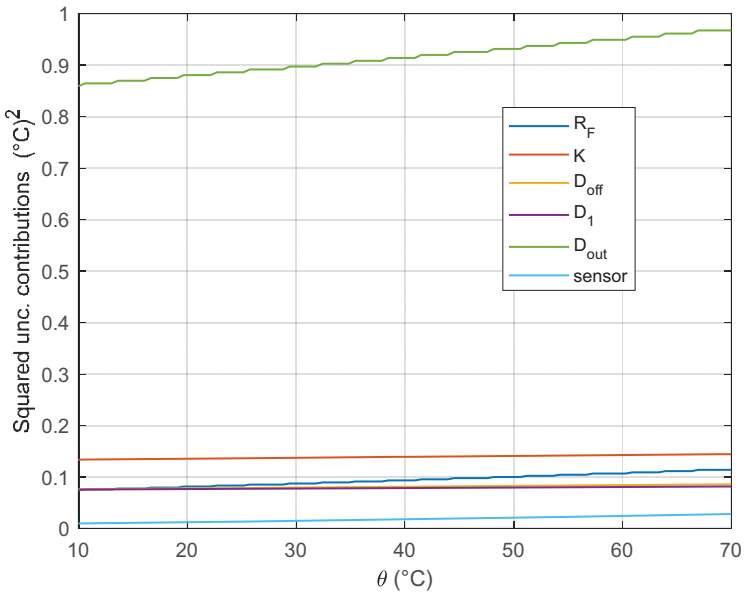


Figure E6.7.9: Numerical values of the squared standard uncertainty contributions of expression (E6.7.18) in the temperature range of interest.

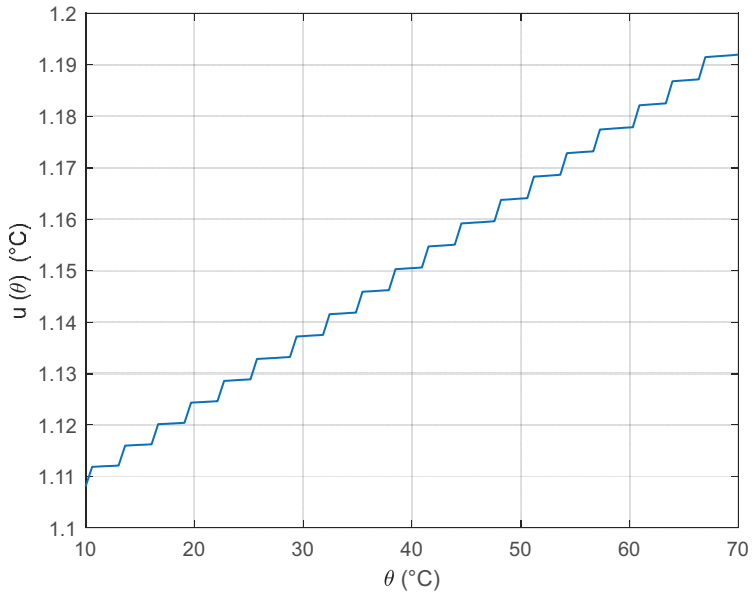


Figure E6.7.10: Absolute standard uncertainty $u(\theta)$ in the temperature range of interest.

– the implementation of the dithering technique.

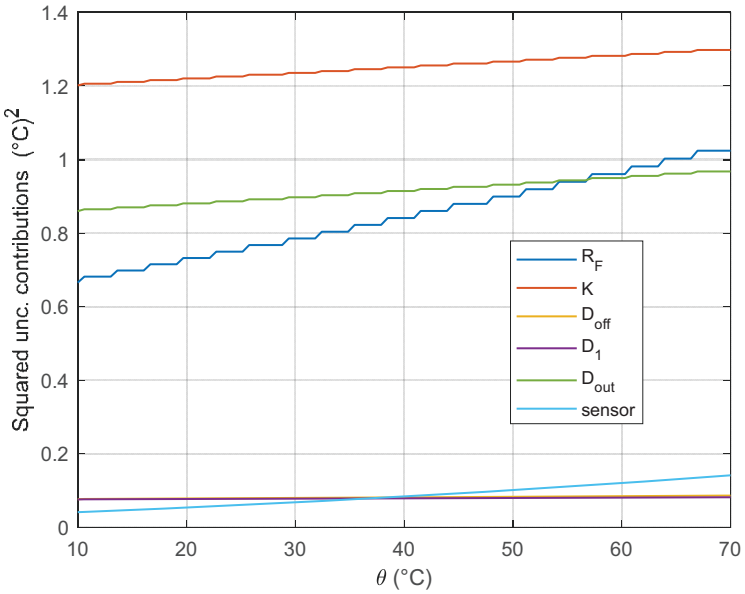


Figure E6.7.11: Numerical values of the squared standard uncertainty contributions of the expression (E6.7.18) in the temperature range of interest obtained with a class B Pt100 and fixed resistors known with a relative standard uncertainty equal to 3×10^{-3}.

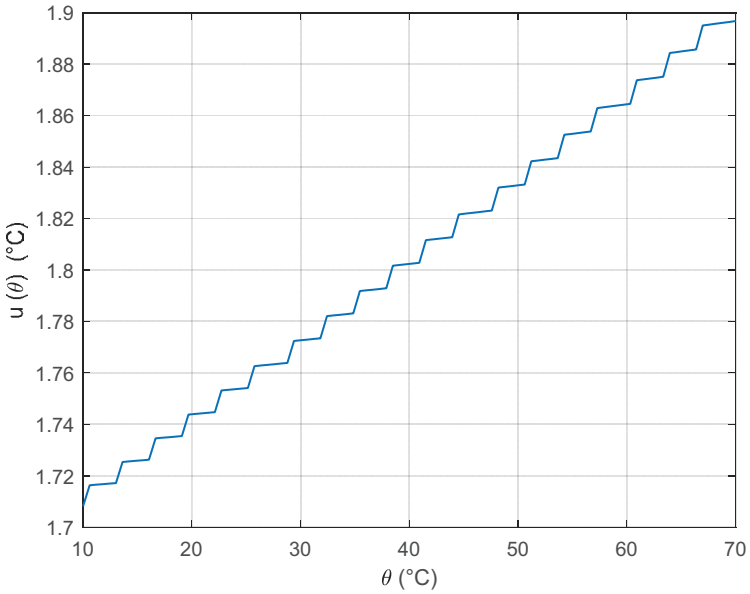


Figure E6.7.12: Absolute standard uncertainty $u(\theta)$ in the temperature range of interest obtained with a class B Pt100 and fixed resistors known with a relative standard uncertainty equal to 3×10^{-3}.

As an example of results that can be obtained relaxing the requirements of the components of the measuring chain, figure E6.7.11 shows the squared standard uncertainty contributions in expression (E6.7.18) when the relative standard uncertainty of the resistors R_F, R_1 and R_2 is

3×10^{-3} (instead of 1×10^{-3}) and the sensor Pt100 belongs to class B [510], which corresponds to a maximum admitted error expressed as:

$$\delta\theta_{\text{sensor}} = \pm(0.30 + 0.005|\theta|)\ ^{\circ}\text{C}. \tag{E6.7.28}$$

The corresponding absolute standard uncertainty is instead shown in figure E6.7.12, which shows that the uncertainty requirement is met in the whole temperature range.

The last consideration is related to the uncertainty contribution that is due to the wire resistance of the sensor Pt100. Even though the sensor is provided with four wires, the measurement circuit of figure E6.7.1 allows a three-wire connection to be arranged. The main consequence of this connection is related to the resistance R_{W} of the wire that connects the sensor to the voltage supply V_{S}. The resistance R_{W} is connected in series to the sensor resistance R and then it acts as a systemic effect that biases the temperature measurement. A raw evaluation of this effect can be obtained evaluating the absolute resistance change of the sensor Pt100 as about $0.4\,\Omega\,^{\circ}\text{C}^{-1}$, which corresponds to a temperature error of $0.25\,^{\circ}\text{C}$ for a value of $R_{\text{W}} = 0.1\,\Omega$. When this systematic effect becomes non-negligible, it is possible to measure the resistance of the two wires of the sensor Pt100 and modify the measurement model to take into account the voltage drop across the resistance R_{W}.

Bibliography

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. *Guide to the Expression of Uncertainty in Measurement – Part 6: Developing and using measurement models, JCGM GUM-6:2020*. BIPM, 2020.

[6] 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.

[7] ISO/IEC 17025 General requirements for the competence of testing and calibration laboratories. ISO, International Organization for Standardization, Geneva, Switzerland, 2017. Third edition.

[8] ISO 17034 General requirements for the competence of reference material producers. ISO, International Organization for Standardization, Geneva, Switzerland, First edition 2016.

[9] ISO/IEC 17043 Conformity assessment – General requirements for proficiency testing. ISO, International Organization for Standardization, Geneva, Switzerland, 2010. First edition.

[10] ISO Guide 33 Reference materials – Good practice in using reference materials. ISO, International Organization for Standardization, Geneva, Switzerland, 2015. Third edition.

[11] 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/`.

[12] 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`.

[13] L. L. Martins, A. S. Ribeiro, M. G. Cox, J. A. Sousa, D. Loureiro, M. C. Almeida, M. A. Silva, R. Brito, and A. C. Soares. EMUE-D1-3-SingleBurningItem, April 2020. doi:10.5281/zenodo.3736602.

[14] K. Shirono and M. Cox. EMUE-D1-4-KeyComparisonGaugeBlocks, July 2020. The example relies on the publication: Katsuhiro Shirono and Maurice Cox. Statistical reassessment of calibration and measurement capabilities based on key comparison results. Metrologia, 56(4):045001, 2019 and relates to the provisions of the MRA: BIPM 1999 Mutual recognition of national measurement standards and of measurement certificates issued by national metrology institutes (MRA) Technical Report Bureau International des Poids et Mesures, Sèvres, France. doi:10.5281/zenodo.3948342.

[15] F. Pennecchi, M. G. Cox, P. Harris, A. M. H. van der Veen, and S. L. R. Ellison. EMUE-D2-1-MulticomponentMaterials, March 2020. The example relies on the following publication and IUPAC Project: - 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 - IUPAC Project n: 2016-007-1-500, Risks of conformity assessment of a multicomponent material or object in relation to measurement uncertainty of its test results, https://iupac.org/projects/project- details/?project_nr=2016-007-1-500. doi:10.5281/zenodo.3723507.

[16] A. Bošnjaković, V. Karahodžić, J. Greenwood, and M. G. Cox. EMUE-D2-2-CorrectionFactors, May 2021. Calculations in the majority of this example are performed by substitution of values into equations provided in the text or through use of standard tabulated values. No software was written for the purpose of performing these calculations which can be made, for example, using standard Excel function cell functions. Section 7 is presented in terms of matrices. The calculation has been performed using Mathcad 7. In view of the lack of backward compatibility and the existence of multiple versions of Mathcad, the calculation is provided as a pdf 'script' which should be straightforward to reproduce. doi:10.5281/zenodo.4742171.

[17] F. Pennecchi, F. Rolle, A. Allard, and S. L. R. Ellison. EMUE-D2-3-TSPConcentration, November 2020. The example relies on the following publications: - Ilya Kuselman, Shamai Shpitzer, Francesca Pennecchi and Cathy 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 (https://doi.org/10.1007/s11869-010-0103-6) - Ilya Kuselman, Francesca Pennecchi, Cathy Burns, Aleš Fajgelj and Paolo de Zorzi. IUPAC/CITAC Guide: Investigating out-of-specification test results of chemical composition based on metrological concepts (IUPAC Technical Report). Pure and Applied Chemistry, 84(9), 2012 (https://doi.org/10.1351/PAC-REP-11-10-04). doi:10.5281/zenodo.4242988.

[18] T. Caebergs, B De Boeck, J. Pétry, N. Sebaïhi, M. Cox, N. Fischer, and J. Greenwood. EMUE-D2-4-NanoparticleHeightMeasurements, June 2021. The example closely relates to the following publication: J. Pétry, B. De Boeck, N. Sebaïhi, M. Coenegrachts, T. Caebergs, and M. Dobre. "Uncertainty evaluation in atomic force microscopy measurement of nanoparticles based on statistical mixed model in a Bayesian framework". Measurement Science and Technology, 32(8):085008, 2021. doi:10.1088/1361-6501/abe47f. doi:10.5281/zenodo.5027540.

[19] J. Greenwood, A. Bošnjaković, V. Karahodžić, P. Pedone, and F. Manta. Emue-d2-5-traceabilityfromconformity, May 2021. doi:10.5281/zenodo.4793471.

[20] F. Pennecchi, F. Rolle, M. Sega, S.L.R. Ellison, and A.M.H. van der Veen. EMUE-D3-2-LowMassBaPEvaluation, July 2020. The example relies on the following publication: Michela Sega, Francesca Pennecchi, Sarah Rinaldi, 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, 2016 (http://dx.doi.org/10.1016/j.aca.2016.03.032). doi:10.5281/zenodo.3968940.

[21] F. Pennecchi, F. Rolle, M. Sega, P. G. Spazzini, I. de Krom, and A. M. H. van der Veen. EMUE-D3-3-CalibrationGasMixtures, June 2020. doi:10.5281/zenodo.3875469.

[22] S. L. R. Ellison, M. Singh, and M. G. Cox. EMUE-D3-4-SoilContaminantsMeasurement, April 2021. doi:10.5281/zenodo.4665685.

[23] M. Cauševic, H. Meuzelaar, A. M. H. van der Veen, and M. G. Cox. EMUE-RMG-3-CalibrationGasMixtures, October 2020. doi:10.5281/zenodo.4084956.

[24] A.S. Ribeiro, M. G. Cox, J. A. Sousa, L. L. Martins, D. Loureiro, M. C. Almeida, M. A. Silva, R. Brito, and A. C. Soares. EMUE-D4-1-WaterVolumeMeasurement, April 2021. doi:10.5281/zenodo.4700500.

[25] M.J. Reader-Harris, C. Forsyth, and T. Boussouara. EMUE-D4-2-OrificePlate, July 2021. The discharge-coefficient data are described in Chapter 5 of Reader-Harris, M. J., Orifice plates and Venturi tubes, Springer, 2015, which gives references to the original data. doi:10.5281/zenodo.5139372.

[26] S. Martens, K. Klauenberg, B. Mickan, C. Yardin, N. Fischer, and C. Elster. EMUE-D4-3-QuantifyUncertaintiesInCalibration, September 2020. doi:10.5281/zenodo.4016916.

[27] A. Bošnjaković, V. Karahodžić, M. Čaušević, and A.M.H. van der Veen. EMUE-D4-4-TransformerPowerLoadLoss, June 2021. doi:10.5281/zenodo.4896452.

[28] J. A. Sousa, M. G. Cox, A. S. Ribeiro, and L. L. Martins. EMUE-D4-5-ThermalComfort, March 2020. The example relies on the following publications: - Ribeiro, A.S.; Alves e Sousa, J.; Cox, M.G.; Forbes, A.B.; Matias, L.C.; Martins, L.L. Uncertainty Analysis of Thermal Comfort Parameters. Int. J. Thermophys. 2015, 36, 2124–2149. - Manuel Gameiro da Silva, Maria Marreno Santana, João Alves e Sousa. Uncertainty Analysis of the mean Radiant Temperature Measurement based on Globe Temperature Probes, Journal of Physics: Conf. Series, 1065 (2018), 072036 (doi: 10.1088/1742-6596/1065/7/072036). doi:10.5281/zenodo.3733403.

[29] M. Čaušević, M. G. Cox, and A. M. H. van der Veen. EMUE-RMG-1-GasFlowCalibration, October 2020. In the "Flow_meter_calibration_using_the_master_me ter_method.pdf" Table 1 and Table 2 the data on flow, pressure, temperature and similar can be found. These data can serve as additional input data when running the program MainScript.mat. Since all the .mat files except MainScript.mat are MATLAB functions necessary for running the MainScript, all of them including MainScript should be located in one folder when running the program. doi:10.5281/zenodo.4114438.

[30] M. Čaušević, M. G. Cox, and J. Greenwood. EMUE-RMG-2-Pressure drop measurement, November 2020. It should be noticed that there are two types of .mat files: "WithoutCorrelations" and "WithCorrelations". When running the program, the program UKASPressureDropExample_WithoutCorrelations will use MATLAB functions that have "WithoutCorrelations" in their name. The other will use only functions that have WithCorrelations in their name. However, when running the program it is necessary to ensure that all the functions that are used and the main script (UKASPressureDropExample.mat) are located in one folder. doi:10.5281/zenodo.4243049.

[31] T. Caebergs and M. Cox. EMUE-D5-1-PixelVoxelUncertainty, June 2021. doi:10.5281/zenodo.5027868.

[32] A. Arduino, F. Pennecchi, L. Zilberti, U. Katscher, and M. G. Cox. EMUE-D5-3-EPTTissueCharacterization, November 2020. The example relies on the implementation of phase-based Helmholtz EPT collected in the open-source C++ library EPTlib 0.1.1 (https://eptlib.github.io/), developed in the framework of the EMPIR Project 18HLT05 QUIERO (https://quiero-project.eu/). doi:10.5281/zenodo.4248879.

[33] S. Martens, K. Klauenberg, J. Neukammerand S. Cowen, S.L.R. Ellison, and C.Elster. EMUE-D5-4-MethodComparisonWithCorrelation, 2020. URL: https://zenodo.org/record/3911584, doi:10.5281/zenodo.3911584.

[34] J. Greenwood and M. Cox. EMUE-D5-8-MedicalTemperatureExample, April 2021. No software was developed in support of this example. doi:10.5281/zenodo.4664687.

[35] S. Martens, K. Klauenberg, and C. Elster. EMUE-D6-2-CalibrationUncertaintyGUMvsBayesian, May 2020. URL: https://doi.org/10.5281/zenodo.3858120, doi:10.5281/zenodo.3858121.

[36] L. L. Martins, A. S. Ribeiro, M. G. Cox, J. A. Sousa, D. Loureiro, M. C. Almeida, M. A. Silva, R. Brito, and A. C. Soares. EMUE-D6-4-MobileOpticalMeasurement, April 2020. doi:10.5281/zenodo.3756658.

[37] A. Carullo, S. Corbellini, and A. Vallan. EMUE-D6-6-DAQBoardElectricalQuantities, September 2020. The results reported in the example refer to the data that have been assigned in the MATLAB script unc_zx.m. Other input data can be used accessing to the variables in the script according to the instructions explained in the help section of the script. The experimental results that refer to the characterization of the cross talk between adjacent and non-adjacent channels of a commercial DAQ board have been obtained with the set-up described in the paper: A. Carullo, S. Corbellini, A. Vallan and A. Atzori, Uncertainty Issues in Multi-Channel Data Acquisition Systems, in Proceedings of IEEE 2020 I2MTC Conference, May 25 - 28, 2020, Dubrovnik, Croatia. doi:10.5281/zenodo.4028283.

[38] A. Carullo and M. Cox. EMUE-D6-6-DAQBoardTemperature, February 2021. The results reported in the example refer to the data that have been assigned in the MATLAB script pt100_emue.m. Other input data can be used accessing to the variables in the script. doi:10.5281/zenodo.4556763.

[39] ISO 15189 Medical laboratories – Requirements for quality and competence. ISO, International Organization for Standardization, Geneva, Switzerland, 2012. Third edition.

[40] 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.

[41] A. 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.

[42] S. L. R. Ellison and A. Williams, editors. *Quantifying Uncertainty in Analytical Measurement*. Number QUAM:2012.P1. 2012. Third edition.

[43] 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.

[44] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical Recipes in C: The Art of Scientific Computing, Second Edition*. Cambridge University Press, 1992.

[45] G. Wübbeler, P. M. Harris, M. G. Cox, and C. 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.

[46] M. Matsumoto and T. 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.

[47] G. 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.

[48] R Core Team. The R Project for Statistical Computing, 2019. Accessed 2019-07-05. URL: https://www.r-project.org/.

[49] RStudio Team. *RStudio: Integrated Development Environment for R*. RStudio, Inc., Boston, MA, 2015. URL: http://www.rstudio.com/.

[50] J. T. Hetzel. *trapezoid: The Trapezoidal Distribution*, 2012. R package version 2.0-0. URL: https://CRAN.R-project.org/package=trapezoid.

[51] A. Genz and F. Bretz. *Computation of Multivariate Normal and t Probabilities*. Lecture Notes in Statistics. Springer-Verlag, Heidelberg, 2009.

[52] R. 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.

[53] J. Meija, T. B. Coplen, M. Berglund, W. A. Brand, P. De Bièvre, M. Gröning, N. E. Holden, J. Irrgeher, R. D. Loss, T. Walczyk, and T. Prohaska. Atomic weights of the elements 2013 (IUPAC technical report). *Pure and Applied Chemistry*, 88(3), jan 2016. doi:10.1515/pac-2015-0305.

[54] 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.

[55] A. Possolo, A. M. H. van der Veen, J. Meija, and D. B. 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.

[56] V. Bloomfield. *Using R for numerical analysis in science and engineering*. CRC Press, Taylor & Francis Group, Boca Raton, 2014.

[57] EA Laboratory Committee. EA 4/02 Evaluation of the uncertainty of measurement in calibration. European Cooperation for Accreditation, September 2013.

[58] A. Swishchuk. Table of basic derivatives. http://people.ucalgary.ca/ aswish/A-MAT219TABLES_W11.pdf. Accessed 2019-06-05.

[59] R. C. Weast. *CRC Handbook of Chemistry and Physics: A Ready-Reference Book of Chemical and Physical Data*. CRC Press Inc, 64th edition, 1984.

[60] 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.

[61] Stan Development Team. RStan: the R interface to Stan, 2018. R package version 2.17.3. URL: http://mc-stan.org/.

[62] B. Carpenter, A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell. Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1), 2017. doi:10.18637/jss.v076.i01.

[63] A. 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.

[64] A. 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.

[65] A. Gelman, J. Carlin, H. Stern, D. Dunson, A. Vehtari, and D. Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, Boca Raton, Florida, USA, 3rd edition, 2013.

[66] W. Bich, M. G. Cox, R. Dybkaer, C. Elster, W. T. Estler, B. Hibbert, H. Imai, W. Kool, C. Michotte, L. Nielsen, L. Pendrill, S. Sidney, A. M. H. van der Veen, and W. 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.

[67] 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.

[68] G. Mana and M. Pizzocaro. The least informative distribution and correlation coefficient of measurement results. *Metrologia*, 58(1):015012, 2021. doi:10.1088/1681-7575/abcbe9.

[69] ISO 13528 Statistical methods for use in proficiency testing by interlaboratory comparison. ISO, International Organization for Standardization, Geneva, Switzerland, 2015. Second edition.

[70] M G Cox. The evaluation of key comparison data. *Metrologia*, 39(6):589–595, dec 2002. doi:10.1088/0026-1394/39/6/10.

[71] 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.

[72] A. Hornikova and N. F. Zhang. The relation between theEnvalues including covariance and the 'exclusive' e_n statistic. *Metrologia*, 43(1):L1–L2, nov 2005. doi:10.1088/0026-1394/43/1/n01.

[73] A. Hornikova and N. F. Zhang. The relation between the en values including covariance and the 'exclusive' e_n statistic. *Metrologia*, 43(2):S157–S157, mar 2006. doi:10.1088/0026-1394/43/2/c01.

[74] R. P. Buck, S. Rondini, A. K. Covington, F. G. K. Bauke, C. M. A. Brett, M. F. Camoes, M. J. T. Milton, T. Mussini, R. Naumann, K. W. Brett, P. Spitzer, and G. S. Wilson. The measurement of pH. Definition standards and procedures. *Pure Appl. Chem.*, 74:2169–2200, 2002.

[75] F. Baucke. New IUPAC recommendations on the measurement of pH - background and essentials. *Analytical and Bioanalytical Chemistry*, 374(5):772–777, 2002. doi:10.1007/s00216-002-1523-4.

[76] N. R. Draper and H. Smith. *Applied Regression Analysis*. Wiley-Blackwell, apr 1998. URL: http://dx.doi.org/10.1002/9781118625590, doi:10.1002/9781118625590.

[77] ISO/TS 28037 Determination and use of straight-line calibration functions. ISO, International Organization for Standardization, Geneva, Switzerland, 2010.

[78] CIPM. Mutual recognition of national measurement standards and of calibration and measurement certificates issued by national metrology institutes, October 1999.

[79] MATLAB. *Release R2015a*. The MathWorks Inc., Natick, Massachusetts, 2015.

[80] Guido Van Rossum and Fred L Drake Jr. *Python tutorial*. Centrum voor Wiskunde en Informatica Amsterdam, The Netherlands, 1995.

[81] K. V. Mardia, J. T. Kent, and J. M. Bibby. *Multivariate Analysis*. Academic Press, London, 1979.

[82] W. Bich, M. G. Cox, and P. M. Harris. Uncertainty modelling in mass comparisons. *Metrologia*, 30:1–12, 1993.

[83] G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore, MD, USA, 1996. Third edition.

[84] P. E. Gill, W. Murray, and M. H. Wright. *Practical Optimization*. Academic Press, London, 1981.

[85] S. Lewis and G. Peggs. *The Pressure Balance: A Practical Guide to its Use. Second edition*. HMSO, London, 1991.

[86] EA Laboratory Committee. EA 4/17 Calibration of pressure balancens EAL-G26. European Cooperation for Accreditation, March 1997.

[87] B. E. Poling and J. M. Prausnitz. *The Properties of Gases and Liquids*. McGraw-Hill Education, fifth edition edition, 2000.

[88] A. M. H. van der Veen. Evaluating measurement uncertainty in fluid phase equilibrium calculations. *Metrologia*, 55(2):S60–S69, feb 2018. doi:10.1088/1681-7575/aaa6dd.

[89] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. *International Vocabulary of Metrology – Basic and General Concepts and Associated Terms, JCGM 200:2012 (JCGM 200:2008 with minor corrections)*. BIPM, 2012.

[90] ISO/TS 28037:2010 Determination and use of straight-line calibration functions. ISO, International Organization for Standardization, Geneva, Switzerland, 2010. First edition.

[91] ILAC. ILAC Policy for Uncertainty in Calibration, ILAC-P14:01/2013. Technical report, ILAC-P14:01/2013, 2013.

[92] UKAS. The expression of uncertainty and confidence in measurement, Edition 4. Technical Report M 3003, United Kingdom Accreditation Service, 2019.

[93] A. Possolo. Five examples of assessment and expression of measurement uncertainty. *Applied Stochastic Models in Business and Industry*, 29(1):1–18, 2013. URL: http://dx.doi.org/10.1002/asmb.1947, doi:10.1002/asmb.1947.

[94] A. B. Forbes. Measurement uncertainty and optimized conformance assessment. *Measurement*, 39(9):808–814. doi:10.1016/j.measurement.2006.04.007.

[95] T. Lafarge and A. Possolo. The NIST uncertainty machine. *NCSLI Measure*, 10(3):20–27, sep 2015. doi:10.1080/19315775.2015.11721732.

[96] M. G. Cox and A. O'Hagan. Meaningful expressions of uncertainty in measurement. Technical Report MS 27, National Physical Laboratory, 2021. doi:10.47120/npl.ms27.

[97] ISO 19229 Gas analysis – Purity analysis and the treatment of purity data. ISO, International Organization for Standardization, Geneva, Switzerland, 2018. Second edition.

[98] Measurement uncertainty arising from sampling: a guide to methods and approaches. Eurachem/EUROLAB/ CITAC/Nordtest/AMC Guide, 2019. second edition.

[99] B. D. Hall. On the propagation of uncertainty in complex-valued quantities. *Metrologia*, 41:173–177, 2004.

[100] M. Sené, I. Gilmore, and J. Janssen. Metrology is key to reproducing results. *Nature*, 547(7664):397–399, 2017. doi:10.1038/547397a.

[101] A. M. H. van der Veen and M. G. Cox. Using the Monte Carlo method. In A. M. H. van der Veen and M. G. Cox, editors, *Good Practice in Evaluating Measurement Uncertainty – Compendium of examples*. Teddington, United Kingdom, 1st edition, 2021.

[102] I. Lira and D. Grientschnig. Bayesian assessment of uncertainty in metrology: a tutorial. *Metrologia*, 47(3):R1–R14, 2010. URL: http://stacks.iop.org/0026-1394/47/i=3/a=R01.

[103] A. M. H. van der Veen. Bayesian inference in R and RStan. In A. M. H. van der Veen and M. G. Cox, editors, *Good Practice in Evaluating Measurement Uncertainty – Compendium of examples*. Teddington, United Kingdom, 1st edition, 2021.

[104] ISO 80000-1:2009. Quantities and units — Part 1: General. ISO, International Organization for Standardization, Geneva, Switzerland, 2009. First edition.

[105] R. J. C. Brown and P. J. Brewer. Reporting uncertainty for gas certified reference materials: balancing customer requirements with calibration and measurement capabilities. *Accreditation and Quality Assurance*, 26(1):41–45, February 2021. doi:10.1007/s00769-021-01456-w.

[106] ISO/TS 28038:2018 Determination and use of polynomial calibration functions. ISO, International Organization for Standardization, Geneva, Switzerland, 2018. First edition.

[107] C. W. Clenshaw. *Mathematical Tables Volume 5. Chebyshev Series for Mathematical Functions*. Her Majesty's Stationery Office, 1962.

[108] BIPM. The International System of Units (SI), 2006. (SI Brochure).

[109] A century of pH measurements. *Chemistry International – News magazine for IUPAC*, 32(2), 2010. doi:10.1515/ci.2010.32.2.3.

[110] ISO/TS 28038 Determination and use of polynomial calibration functions. ISO, International Organization for Standardization, Geneva, Switzerland, 2018. First edition.

[111] M. G. Cox. A survey of numerical methods for data and function approximation. In D. A. H. Jacobs, editor, *The State of the Art in Numerical Analysis*, pages 627–668, London, 1977. Academic Press.

[112] Y. C. Wu and W. F. Koch and R. A. Durst. Standard Reference Materials: Standardization of pH Measurements. Spec. Publ. 260-53, NBS, 1988.

[113] A. K. Covington, R. G. Bates, and R. A. Durst. Definition of pH scales, standard reference values, measurement of pH and related terminology (Recommendations 1984). *Pure and Applied Chemistry*, 57(3):531–542, 1985. doi:10.1351/pac198557030531.

[114] NIST. Certificate of Analysis, Standard Reference Material186g, pH Standards, Potassium Dihydrogen Phosphate (186-I-g), Disodium Hydrogen Phosphate (186-II-g). URL: https://www-s.nist.gov/srmors/certificates/186g.pdf.

[115] NIST. Certificate of Analysis, Standard Reference Material189c, Potassium Tetroxalate Dihydrate pH Standard. URL: https://www-s.nist.gov/srmors/certificates/189c.pdf.

[116] J. C. Damasceno, R. M. H. Borges, P. R. G. Couto, A. P. Ordine, M. A. Getrouw, P. P. Borges, and I. C. S. Fraga. Estimation of primary pH measurement uncertainty using Monte Carlo simulation. *Metrologia*, 43(3):306–310, 2006. doi:10.1088/0026-1394/43/3/014.

[117] J. Wiora and A. Wiora. Measurement uncertainty calculations for pH value obtained by an ion-selective electrode. *Sensors*, 18(6):1915, 2018. doi:10.3390/s18061915.

[118] L. Fan, J. Smethurst, P. Atkinson, and W. Powrie. Propagation of vertical and horizontal source data errors into a TIN with linear interpolation. *International Journal of Geographical Information Science*, 28(7):1378–1400, 2014. doi:10.1080/13658816.2014.889299.

[119] A. Gustavo González, M. Ángeles Herrador, A. García Asuero, and J. Martín. A practical way to ISO/GUM measurement uncertainty for analytical assays including in-house validation data. In *Quality Control in Laboratory*. InTech, 2018. doi:10.5772/intechopen.72048.

[120] R. Lundblad. *Handbook of biochemistry and molecular biology*. CRC Press, Boca Raton, FL, 2010.

[121] D. R. White. Propagation of uncertainty and comparison of interpolation schemes. *International Journal of Thermophysics*, 38(3), 2017. doi:10.1007/s10765-016-2174-6.

[122] J. Wiora. Problems and risks occurred during uncertainty evaluation of a quantity calculated from correlated parameters: a case study of pH measurement. *Accreditation and Quality Assurance*, 21(1):33–39, 2015. doi:10.1007/s00769-015-1183-7.

[123] M. G. Cox, C. Eiø, G. Mana, and F. Pennecchi. The generalized weighted mean of correlated quantities. *Metrologia*, 43:S268–S275, 2006.

[124] M. A. Collett, M. G. Cox, T. J. Esward, P. M. Harris, and J. A. Sousa. Aggregating measurement data influenced by common effects. *Metrologia*, 44:308–318, 2007.

[125] I. Farrance and R. Frenkel. Uncertainty of Measurement: A Review of the Rules for Calculating Uncertainty Components through Functional Relationships. *Clin. Biochem. Rev.*, 33:49–75.

[126] G.E.P. Box and G.C. Tiao. *Bayesian Inference in Statistical Analysis*. Wiley, 1992.

[127] 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.

[128] 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.

[129] 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.

[130] A. Possolo and B. Toman. Assessment of measurement uncertainty via observation equations. *Metrologia*, 44(6):464–475, 2007. doi:10.1088/0026-1394/44/6/005.

[131] 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.

[132] EN/TR 16988:2016 Estimation of uncertainty in the single burning item test. EN, European Organization for Standardization, Brussels, Belgium, 2016. First edition.

[133] 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.

[134] M. Kochsiek and M. Gläser. *Comprehensive mass metrology*. Wiley-VCH, Weinheim New York, 2000.

[135] 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.

[136] K. Shirono and M. 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.

[137] M. Cox, P. Harris, and M. 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.

[138] S. P. Robinson, P. M. Harris, J. Ablitt, G. Hayman, A. Thompson, A. Lee van Buren, J. F. Zalesak, R. M. Drake, A. E. Isaev, A. M. Enyakov, C. Purcell, Z. Houqing, W. Yuebing, Z. Yue, P. Botha, and D. 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.

[139] M. G. Cox. The evaluation of key comparison data. *Metrologia*, 39(6):589–595, 2002. doi:10.1088/0026-1394/39/6/10.

[140] 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.

[141] A. 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.

[142] BIPM. Mutual recognition of national measurement standards and of measurement certificates issued by national metrology institutes (MRA). Technical report, Bureau International des Poids et Mesures, Sèvres, France, 1999.

[143] D. R. White. On the analysis of measurement comparisons. *Metrologia*, 41(3):122, 2004. URL: http://stacks.iop.org/0026-1394/41/i=3/a=003.

[144] C. M. Sutton. Analysis and linking of international measurement comparisons. *Metrologia*, 41(4):272–277, 2004. URL: http://stacks.iop.org/0026-1394/41/i=4/a=008.

[145] C. Elster, A. G. Chunovkina, and W. Wöger. Linking of a RMO key comparison to a related CIPM key comparison using the degrees of equivalence of the linking laboratories. *Metrologia*, 47(1):96, 2010. URL: http://stacks.iop.org/0026-1394/47/i=1/a=011.

[146] I. A. Kharitonov and A. G. Chunovkina. Evaluation of regional key comparison data: two approaches for data processing. *Metrologia*, 43(5):470, 2006. URL: http://stacks.iop.org/0026-1394/43/i=5/a=019.

[147] J. E. Decker, A. G. Steele, and R. J. Douglas. Measurement science and the linking of CIPM and regional key comparisons. *Metrologia*, 45(2):223, 2008. URL: http://stacks.iop.org/0026-1394/45/i=2/a=012.

[148] Th Bruns. Final report on the key comparison EUROMET-AUV.V-K1.1. *Metrologia*, 47(1A):09001, 2010. URL: http://stacks.iop.org/0026-1394/47/i=1A/a=09001.

[149] H.-J. von Martens, C. Elster, A. Link, A. Täubner, and W. Wabinski. CCAUV.V-K1 Final report. *Metrologia*, 40(1A):09001, 2003. URL: http://stacks.iop.org/0026-1394/40/i=1A/a=09001.

[150] ISO 20486:2017 Non-destructive testing — Leak testing — Calibration of reference leaks for gases. ISO, International Organization for Standardization, Geneva, Switzerland, 2017. First edition.

[151] EMRP IND 12 consortium. Metrology of the leak detection - Practical guide. Technical report, 2015.

[152] I. H. Lira and W. Wöger. The evaluation of the uncertainty in knowing a directly measured quantity. *Meas. Sci. Technol.*, 9:1167–1173, 1998.

[153] W. Haesselbarth. Accounting for bias in measurement uncertainty estimation. *Accreditation and Quality Assurance*, 9(8), 2004. doi:10.1007/s00769-004-0782-5.

[154] B. Magnusson and S. 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.

[155] 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.

[156] 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.

[157] K. Klauenberg, G. Wübbeler, and C. Elster. About not Correcting for Systematic Effects. *Measurement Science Review*, 19(5):204–208, 2019. doi:10.2478/msr-2019-0026.

[158] G. Molinar, M. Beciet, and F. 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.

[159] S. Westgard and D. 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.

[160] W. F. Guthrie. NIST/SEMATECH e-Handbook of Statistical Methods (NIST Handbook 151), 2020. doi:10.18434/M32189.

[161] G. E. O'Donnell and D. B. Hibbert. Treatment of bias in estimating measurement uncertainty. *The Analyst*, 130(5):721, 2005. doi:10.1039/b414843f.

[162] I. Kuselman, D. B. Hibbert, F. R. Pennecchi, and R. 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.

[163] I. Kuselman, F. R. Pennecchi, R. J. N. B. da Silva, and D. B. 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.

[164] Drugs.com Overview: What is NyQuil Severe Cold & Flu? https://www.drugs.com/mtm/nyquil-severe-cold-flu.html. Accessed: 2019-06-26.

[165] ISO/IEC 27893 Vacuum Technology – Vacuum Gauges – Evaluation of the uncertainties of results of calibrations by direct comparison with a reference gauge. ISO, International Organization for Standardization, Geneva, Switzerland, 2011.

[166] ISO 3567:2011 Vacuum gauges — Calibration by direct comparison with a reference gauge. ISO, International Organization for Standardization, Geneva, Switzerland, 2011.

[167] J. Greenwood, M. G. Cox, and N. Fischer. E1.8 Factoring effects such as calibration corrections and drift into uncertainty evaluations. In A. M. H. van der Veen and M. G. Cox, editors, *Good Practice in Evaluating Measurement Uncertainty – Compendium of examples*, pages 109–118. Teddington, United Kingdom, 1st edition, 2021. URL: http://empir.npl.co.uk/emue/wp-content/uploads/sites/49/2020/06/Compendium_M36.pdf.

[168] J. Greenwood, A. Bosnjakovic, V. Karahodzic, P. Pedone, and F. Manta. E2.5 GUM-LPU uncertainty evaluation - importing measurement traceability from a conformity statement. In A. M. H. van der Veen and M. G. Cox, editors, *Good Practice in Evaluating Measurement Uncertainty – Compendium of examples*, pages XX–YY. Teddington, United Kingdom, 1st edition, 2021. URL: http://empir.npl.co.uk/emue/wp-content/uploads/sites/49/2020/06/Compendium_M36.pdf.

[169] M. G. Cox and A. M. H. van der Veen. Reporting measurement results. In A. M. H. van der Veen and M. G. Cox, editors, *Good Practice in Evaluating Measurement Uncertainty – Compendium of examples*. Teddington, United Kingdom, 1st edition, 2021.

[170] 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.

[171] 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.

[172] I. Kuselman, F. 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.

[173] F. R. Pennecchi, I. Kuselman, R. J. N. B. da Silva, and D. B. 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.

[174] A. Allard, N. Fischer, I. Smith, P. Harris, and L. Pendrill. CASoft software. https://www.lne.fr/en/software/CASoft, 2020. version 1.1.

[175] J. Pétry, B. De Boeck, N. Sebaïhi, M. Coenegrachts, T. Caebergs, and M. Dobre. Uncertainty evaluation in atomic force microscopy measurement of nanoparticles based on statistical mixed model in a Bayesian framework. *Measurement Science and Technology*, 32(8):085008, 2021. doi:10.1088/1361-6501/abe47f.

[176] Regulation (EC) No 1223/2009 of the European Parliament and of the Council of 30 November 2009 on cosmetic products. *OJ*, L342:59–209. Text with EEA relevance. URL: http://data.europa.eu/eli/reg/2009/1223/oj.

[177] Regulation (EU) No 1169/2011 of the European Parliament and of the Council of 25 October 2011 on the provision of food information to consumers. *OJ*, L 304:18–63. amending Regulations (EC) No 1924/2006 and (EC) No 1925/2006 of the European Parliament and of the Council, and repealing Commission Directive 87/250/EEC, Council Directive 90/496/EEC, Commission Directive 1999/10/EC, Directive 2000/13/EC of the European Parliament and of the Council, Commission Directives 2002/67/EC and 2008/5/EC and Commission Regulation (EC) No 608/2004 Text with EEA relevance. URL: http://data.europa.eu/eli/reg/2011/1169/oj.

[178] Regulation (EU) No 528/2012 of the European Parliament and of the Council of 22 May 2012 concerning the making available on the market and use of biocidal products. *OJ*, L 167:1–123, 2012. Text with EEA relevance. URL: http://data.europa.eu/eli/reg/2012/528/oj.

[179] Regulation (EU) No 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices. *OJ*, L 117:1–175, 2017. amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC (Text with EEA relevance.). URL: http://data.europa.eu/eli/reg/2017/745/2017-05-05.

[180] R-Nano. *Code de l'environnement*, 2012. created by décret n°2012-232 of 17 February 2012. URL: https://www.legifrance.gouv.fr/eli/decret/2012/2/17/DEVP1123456D/jo/texte.

[181] Sécurité de la Chaîne alimentaire et Environnement Service publique fédéral Santé publique. Arrêté royal relatif à la mise sur le marché des substances manufacturées à l'état nanoparticulaire. *Moniteur belge*, 2014. URL: http://www.ejustice.just.fgov.be/eli/arrete/2014/05/27/2014024329/moniteur.

[182] Bekendtgørelse om register over blandinger og varer, der indeholder nanomaterialer samt producenter og importørers indberetningspligt til registeret. 2014. URL: https://www.retsinformation.dk/eli/lta/2014/644.

[183] Commission Recommendation of 18 October 2011 on the definition of nanomaterial. *OJ*, L275:38–40. Text with EEA relevance. URL: http://data.europa.eu/eli/reco/2011/696/oj.

[184] T. Caebergs and M.G. Cox. E5.1 Set of pixels or voxels. In A. M. H. van der Veen and M. G. Cox, editors, *Good Practice in Evaluating Measurement Uncertainty – Compendium of examples*. Teddington, United Kingdom, 1st edition, 2021.

[185] A. Delvallée, N. Feltin, S. Ducourtieux, M. Trabelsi, and J.-F. Hochepied. Toward an uncertainty budget for measuring nanoparticles by AFM. *Metrologia*, 53(1):41–50, December 2016. doi:10.1088/0026-1394/53/1/41.

[186] Stan Development Team. RStan: the R interface to Stan, 2018. R package version 2.18.2. URL: http://mc-stan.org/.

[187] J. Small and R. Watters. Reference Material® 8012. Report of Investigation, National Institude of Standards and Technology (NIST), 2015. URL: https://www-s.nist.gov/srmors/certificates/8012.pdf.

[188] J. Grobelny, F. DelRio, N. Pradeep, D.-I. Kim, V. Hackley, and R. Cook. Size Measurement of Nanoparticles Using Atomic Force Microscopy. In *Methods in Molecular Biology*, pages 71–82. Humana Press, October 2010. doi:10.1007/978-1-60327-198-1_7.

[189] *ISO 5436-1:2000 Geometrical Product Specifications (GPS) - Surface texture: Profile method; Measurement standards - Part 1: Material measures*. Geneva, Switzerland.

[190] SAS Institute Inc. *JMP®, v.14*.

[191] D. Hensher, J. Rose, and W. Greene. *Applied Choice Analysis*. Cambridge University Press, 2nd edition, 2015. URL: https://www.cambridge.org/be/academic/subjects/economics/econometrics-statistics-and-mathematical-economics/applied-choice-analysis-2nd-edition?format=PB&isbn=9781107092648, doi:10.1017/CBO9781316136232.

[192] D. Montgomery. *Design and Analysis of Experiments*. John Wiley & sons, 6th edition, 2005.

[193] SAS Institude Inc. *SAS/STAT® 14.1 User's Guide*, chapter 77: The MIXED Procedure, pages 6047–10032. July 2015. https://support.sas.com/documentation/onlinedoc/stat/141/mixed.pdf, visited 2021-01-21. URL: https://support.sas.com/documentation/onlinedoc/stat/141/mixed.pdf.

[194] UKAS. Decision rules and statements of conformity. UKAS LAB 48, Ed 3, 2020.

[195] OIML. Weights of classes E_1, E_2, F_1, F_2, M_1, M_{1-2}, M_2, M_{2-3} and M_3 part 1: Metrological and technical requirements. 2004.

[196] ISO13385-1 Geometrical product specifications (GPS) — Dimensional measuring equipment — Part 1: Design and metrological characteristics of callipers. 2019.

[197] ISO14253-5 Geometrical product specifications (GPS) — Inspection by measurement of workpieces and measuring equipment — Part 5: Uncertainty in verification testing of indicating measuring instruments. 2015.

[198] ISO14978 Geometrical product specifications (GPS) — General concepts and requirements for GPS measuring equipment. 2018.

[199] ISO14253-1 Geometrical product specifications (GPS) — Inspection by measurement of workpieces and measuring equipment — Part 1: Decision rules for verifying conformity or nonconformity with specifications. 2017.

[200] DESA and UN et al. Transforming our world: The 2030 agenda for sustainable development. 2016.

[201] Guide to Hydrological Practices. Volume I. WMO-No. 168, World Meteorological Organization, Geneva, Switzerland, 2008.

[202] L.G. Lanza, M. Leroy, C. Alexandropoulos, L. Stagi, and W. Wauben. WMO Laboratory Intercomparison of Rainfall Intensity Gauges. WMO/TD-No. 1304, World Meteorological Organization, Geneva, Switzerland, 2006. Instruments and Observing Methods Report No. 84.

[203] Catalogue of National Standard Precipitation Gauges. WMO/TD-No. 313, World Meteorological Organization, Geneva, Switzerland, 1989a. Instruments and Observing Methods Report No. 39.

[204] WMO Guidelines on the Calculation of Climate Normals. WMO–No. 1203, World Meteorological Organization, Geneva, Switzerland, 2017.

[205] Guides to Instruments and Methods of Observation. Vol. 1 Measurement of Meteorological Variables. Technical report, World Meteorological Organization, Geneva, Switzerland, 2018. WMO-No. 8.

[206] A. Molini, P. La Barbera, L.G. Lanza, and L. Stagi. Rainfall intermittency and the sampling error of tipping-bucket rain gauges. *Physics and Chemistry of the Earth, Part C: Solar, Terrestrial & Planetary Science*, 26(10-12):737–742, January 2001. doi:10.1016/s1464-1917(01)95018-4.

[207] B. Sevruk. Methods of correction for systematic error in point precipitation measurement for operational use. WMO-No. 589, World Meteorological Organization, Geneva, Switzerland, 1982. Operational Hydrology Report No. 21.

[208] E. Vuerich, C. Monesi, L.G. Lanza, L. Stagi, and E. Lanzinger. WMO Field Intercomparison of Rainfall Intensity Gauges. WMO/TD-No. 1504, World Meteorological Organization, Geneva, Switzerland, 2009. Instruments and Observing Methods Report No. 99.

[209] D. R. Legates and C. J. Willmott. Mean seasonal and spatial variability in gauge-corrected, global precipitation. *International Journal of Climatology*, 10(2):111–127, March 1990. doi:10.1002/joc.3370100202.

[210] V. Nešpor and B. Sevruk. Estimation of wind-induced error of rainfall gauge measurements using a numerical simulation. *Journal of Atmospheric and Oceanic Technology*, 16(4):450–464, April 1999. doi:10.1175/1520-0426(1999)016<0450:eowieo>2.0.co;2.

[211] Z. Sen. *Spatial Modeling Principles in Earth Sciences*. Springer Netherlands, 2009. doi:10.1007/978-1-4020-9672-3.

[212] A. González-Álvarez, O.o Viloria-Marimón, O. Coronado-Hernández, A. Vélez-Pereira, K. Tesfagiorgis, and J. Coronado-Hernández. Isohyetal maps of daily maximum rainfall for different return periods for the colombian caribbean region. *Water*, 11(2):358, February 2019. doi:10.3390/w11020358.

[213] B. Sevruk. Evaporation losses from storage gauges. In *Distribution of Precipitation in Mountainous Areas, Geilo Symposium (Norway, 31 July–5 August 1972), Volume II – technical papers*. World Meteorological Organization, Geneva, Switzerland, 1972. WMO-No. 326.

[214] M. Colli, L.G. Lanza, and P.W. Chan. Co-located tipping-bucket and optical drop counter RI measurements and a simulated correction algorithm. *Atmospheric Research*, 119:3–12, January 2013. doi:10.1016/j.atmosres.2011.07.018.

[215] M. Colli, L.G. Lanza, and P. La Barbera. Performance of a weighing rain gauge under laboratory simulated time-varying reference rainfall rates. *Atmospheric Research*, 131:3–12, September 2013. doi:10.1016/j.atmosres.2013.04.006.

[216] E. Habib, W. F. Krajewski, and A. Kruger. Sampling errors of tipping-bucket rain gauge measurements. *Journal of Hydrologic Engineering*, 6(2):159–166, April 2001. doi:10.1061/(asce)1084-0699(2001)6:2(159).

[217] R. E. Rinehart. Out-of-level instruments: Errors in hydrometeor spectra and precipitation measurements. *Journal of Climate and Applied Meteorology*, 22(8):1404–1410, August 1983. doi:10.1175/1520-0450(1983)022<1404:ooliei>2.0.co;2.

[218] B. Sevruk and V. Nešpor. The effect of dimensions and shape of precipitation gauges on the wind-induced error. In M. Desbois and F. Désalmand, editors, *Global Precipitations and Climate Change*, pages 231–246, Berlin, Heidelberg, 1994. Springer Berlin Heidelberg.

[219] J. Niemczynowicz. The dynamic calibration of tipping-bucket raingauges. *Hydrology Research*, 17(3):203–214, June 1986. doi:10.2166/nh.1986.0013.

[220] L. C. Sieck, S. J. Burges, and M. Steiner. Challenges in obtaining reliable measurements of point rainfall. *Water Resources Research*, 43(1), January 2007. doi:10.1029/2005wr004519.

[221] M. Stagnaro, M. Colli, L. Giovanni Lanza, and P. Wai Chan. Performance of post-processing algorithms for rainfall intensity using measurements from tipping-bucket rain gauges. *Atmospheric Measurement Techniques*, 9(12):5699–5706, November 2016. doi:10.5194/amt-9-5699-2016.

[222] 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.

[223] 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.

[224] M. Sega, F. Pennecchi, S. Rinaldi, and F. 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.

[225] 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.

[226] 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.

[227] 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.

[228] 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.

[229] EN 14211:2012, Ambient air - Standard method for the measurement of the concentration of nitrogen dioxide and nitrogen monoxide by chemiluminescence. Technical report, 2012.

[230] 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.

[231] ISO 11466:1995 Soil quality – Extraction of trace elements soluble in aqua regia. ISO, International Organization for Standardization, Geneva, Switzerland, 1995.

[232] ISO 21748:2017 – Guidance for the use of repeatability, reproducibility and trueness estimates in measurement uncertainty evaluation. ISO, International Organization for Standardization, Geneva, Switzerland, 2017.

[233] ISO 11464:2006: Soil quality – Soil quality — Pretreatment of samples for physico-chemical analysis. ISO, International Organization for Standardization, Geneva, Switzerland, 2006.

[234] S. L. R. Ellison, V. J. Barwick, P. Norris, and M. Griffiths. Complete curve fitting of extraction profiles for estimating uncertainties in recovery estimates. *Analyst*, 128:493–498, 2003. URL: http://dx.doi.org/10.1039/B212784A, doi:10.1039/B212784A.

[235] Eurachem/CITAC guide: Use of uncertainty information in compliance assessment, 207. URL: https://www.eurachem.org.

[236] UK Environment Agency. Performance standard for laboratories undertaking chemical testing of soil, version 5. Mcerts standard, UK Environment Agency, Apr 2018. URL: http://www.mcerts.net.

[237] ISO 5725-2:2019 – Accuracy (trueness and precision) of measurement methods and results – Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method. ISO, International Organization for Standardization, Geneva, Switzerland, 2019.

[238] ISO 5725-4:2020 – Accuracy (trueness and precision) of measurement methods and results – Part 4: Basic methods for the determination of the trueness of a standard measurement method. ISO, International Organization for Standardization, Geneva, Switzerland, 2020.

[239] Y. Sakamoto, M. Ishiguro, and G. Kitagawa. *Akaike information criterion statistics*. Mathematics and its applications (Japanese series). KTK Scientific Publishers ; D. Reidel ; Sold and distributed in the U.S.A. and Canada by Kluwer Academic Publishers, Tokyo : Dordrecht ; Boston : Hingham, MA, 1986.

[240] C. M. Hurvich and C.-L. Tsai. Regression and time series model selection in small samples. *Biometrika*, 76(2):297–307, 06 1989. arXiv:https://academic.oup.com/biomet/article-pdf/76/2/297/737009/76-2-297.pdf, doi:10.1093/biomet/76.2.297.

[241] S. L. R. Ellison. Including correlation effects in an improved spreadsheet calculation of combined standard uncertainties. *Accreditation and Quality Assurance*, 10(7):338–343, 2005. doi:10.1007/s00769-005-0008-5.

[242] F. Frenzel, H. Grothey, C. Habersetzer, M. Hiatt, W. Hogrefe, M. Kirchner, G. Lütkepohl, W. Marchewka, U. Mecke, M. Ohm, et al. Industrial flow measurement basics and practice. Technical report, ABB automation products Gmbh, 2011.

[243] J. G. Webster and H. Eren, editors. *Measurement, Instrumentation, and Sensors Handbook: Electromagnetic, Optical, Radiation, Chemical, and Biomedical Measurement*. CRC Press, 2nd edition edition, 2014.

[244] H. Zhang, C. Guo, and J. Lin. Effects of velocity profiles on measuring accuracy of transit-time ultrasonic flowmeter. *Applied Sciences*, 9(8):1648, 2019. doi:10.3390/app9081648.

[245] S.W. Rienstra and A. Hirschberg. An introduction to acoustics. Lecture notes, Eindhoven University of Technology, Eindhoven, the Netherlands, 2004.

[246] UKAS. The expression of uncertainty and confidence in measurement. UKAS M3003, Ed 4, 2019.

[247] ISO 9104 Measurement of fluid flow closed conduit methods of evaluating the performance of electromagnetic flow meter for liquids. ISO, International Organization for Standardization, Geneva, Switzerland, 1991. First edition.

[248] ISO 5168 Measurement of fluid flow – Procedures for the evaluation of uncertainties. ISO, International Organization for Standardization, Geneva, Switzerland, 2005. Second edition.

[249] ISO 20456 Measurement of fluid flow in closed conduits — guidance for the use of electromagnetic flowmeters for conductive liquids. ISO, International Organization for Standardization, Geneva, Switzerland, 2017. First edition.

[250] A. L. Sotero Salustiano Martim, J. G. Dalfré Filho, Y. de Faria Lemos De Lucca, and A. I. Borri Genovez. Electromagnetic flowmeter evaluation in real facilities: Velocity profiles and error analysis. *Flow Measurement and Instrumentation*, 66:44–49, 2019. doi:10.1016/j.flowmeasinst.2019.01.001.

[251] A. Weissenbrunner, A. Fiebach, S. Schmelter, M. Bär, P.U. Thamsen, and T. Lederer. Simulation-based determination of systematic errors of flow meters due to uncertain inflow conditions. *Flow Measurement and Instrumentation*, 52:25–39, 2016. doi:10.1016/j.flowmeasinst.2016.07.011.

[252] L. Ma, J. Liu, and J. Wang. Study of the accuracy of ultrasonic flowmeters for liquid. *AASRI Procedia*, 3:14–20, 2012. doi:10.1016/j.aasri.2012.11.004.

[253] ISO 20456 Measurement of fluid flow in closed conduits — ultrasonic transit-time meters for liquid. ISO, International Organization for Standardization, Geneva, Switzerland, 2012. First edition.

[254] B. Iooss, C. Lhuillier, and H. Jeanneau. Numerical simulation of transit-time ultrasonic flowmeters: uncertainties due to flow profile and fluid turbulence. *Ultrasonics*, 40(9):1009–1015, 2002. doi:10.1016/s0041-624x(02)00387-6.

[255] IPCC. IPCC Guidelines for National Greenhouse Gas Inventories Volume 1 General Guidance and Reporting, 2006.

[256] D. Butterfield. Understanding the UK Greenhouse Gas Inventory: An assessment of how the UK inventory is calculated and the implications of uncertainty. Technical Report CCM 2, National Physical Laboratory, Teddington, UK, 2017.

[257] D.L. Bloom, D.M. Byrne, and J.M. Andreson. Communicating Risk to Senior EPA Policy-Makers: A Focus Group Study. Technical report, Bloom Research and the Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, 1993.

[258] IPCC. UK Greenhouse Gas Inventory 1990 to 2016: Annual Report for submission under the Framework Convention on Climate Change, 2018.

[259] IPCC. Good Practice Guidance and Uncertainty Management in National Greenhouse Gas Inventories, 2000.

[260] Ricardo-AEA. An Introduction to the UK's Greenhouse Gas Inventory, 2014.

[261] A. E. Milne, M. J. Glendining, R. Murray Lark, S. A. M. Perryman, T. Gordon, and A. P. Whitmore. Communicating the uncertainty in estimated greenhouse gas emissions from agriculture. *Journal of Environmental Management*, 160:139–153, 2015. doi:10.1016/j.jenvman.2015.05.034.

[262] S. Eggleston and L. Buendia and Kyoko Miwa and Todd Ngara and Kiyoto Tanabe. Guidelines for National Greenhouse Gas Inventories, 2000.

[263] J. H. Huggins, T. Campbell, M. Kasprzak, and T. Broderick. Practical bounds on the error of Bayesian posterior approximations: A nonasymptotic approach. *arXiv e-prints*, page arXiv:1809.09505, Sep 2018. arXiv:1809.09505.

[264] D. Hangleiter, I. Roth, D. Nagaj, and J. Eisert. Easing the Monte Carlo sign problem. *Science Advances*, 6(33):eabb8341, 2020. doi:10.1126/sciadv.abb8341.

[265] V. Periyasamy and M. Pramanik. Advances in Monte Carlo Simulation for Light Propagation in Tissue. *IEEE Reviews in Biomedical Engineering*, 10:122–135, 2017. doi:10.1109/rbme.2017.2739801.

[266] J. Saini, E. Traneus, D. Maes, R. Regmi, S. R. Bowen, C. Bloch, and T. Wong. Advanced Proton Beam Dosimetry Part I: review and performance evaluation of dose calculation algorithms. *Translational Lung Cancer Research*, 7(2):171–179, 2018. doi:10.21037/tlcr.2018.04.05.

[267] D. E. Farrar and R. R. Glauber. Multicollinearity in regression analysis: The problem revisited. *The Review of Economics and Statistics*, 49(1):92, 1967. doi:10.2307/1937887.

[268] Intergovernmental Panel on Climate Change, editor. *Climate Change 2013 - The Physical Science Basis*. Cambridge University Press, 2009. doi:10.1017/cbo9781107415324.

[269] T. Arnold, A. J. Manning, J. Kim, S. Li, H. Webster, D. Thomson, J. Mühle, R. F. Weiss, S. Park, and S. O'Doherty. Inverse modelling of CF_4 and NF_3 emissions in East Asia. *Atmospheric Chemistry and Physics*, 18(18):13305–13320, 2018. doi:10.5194/acp-18-13305-2018.

[270] M. Rigby, S. Park, T. Saito, L. M. Western, A. L. Redington, X. Fang, S. Henne, A. J. Manning, R. G. Prinn, G. S. Dutton, P. J. Fraser, A. L. Ganesan, B. D. Hall, C. M. Harth, J. Kim, K.-R. Kim, P. B. Krummel, T. Lee, S. Li, Q. Liang, M. F. Lunt, S. A. Montzka, J. Mühle, S. O'Doherty, M.-K. Park, S. Reimann, P. K. Salameh, P. Simmonds, R. L. Tunnicliffe, R. F. Weiss, Y. Yokouchi, and D. Young. Increase in CFC-11 emissions from eastern China based on atmospheric observations. *Nature*, 569(7757):546–550, 2019. doi:10.1038/s41586-019-1193-4.

[271] T. Arnold, C. M. Harth, J. Mühle, A. J. Manning, P. K. Salameh, J. Kim, D. J. Ivy, L. P. Steele, V. V. Petrenko, J. P. Severinghaus, D. Baggenstos, and R. F. Weiss. Nitrogen trifluoride global emissions estimated from updated atmospheric measurements. *Proceedings of the National Academy of Sciences*, 110(6):2029–2034, 2013. doi:10.1073/pnas.1212346110.

[272] J. Mühle, A.L. Ganesan, B.R. Miller, P.K. Salameh, C.M. Harth, B.R. Greally, M. Rigby, L.W. Porter, L.P. Steele, C.M. Trudinger, P.B. Krummel, S. O'Doherty, P.J. Fraser, P.G. Simmonds, R.G. Prinn, and R.F. Weiss. Perfluorocarbons in the global atmosphere: tetrafluoromethane, hexafluoroethane, and octafluoropropane, atmos. *Atmos. Chem. Phys*, 10:5145–5164, 2013. doi:10.5194/acp-10-5145-2010.

[273] D. R. Worton, W. T. Sturges, L. K. Gohar, K. P. Shine, P. Martinerie, D. E. Oram, S. P. Humphrey, P. Begley, L. Gunn, J.-M. Barnola, J. Schwander, and R. Mulvaney. Atmospheric Trends and Radiative Forcings of CF_4 and C_2F_6 Inferred from Firn Air. *Environmental Science & Technology*, 41(7):2184–2189, 2007. doi:10.1021/es061710t.

[274] J. Kim, P. J. Fraser, S. Li, J. Mühle, A. L. Ganesan, P. B. Krummel, L. P. Steele, S. Park, S.-K. Kim, M.-K. Park, T. Arnold, C. M. Harth, P. K. Salameh, R. G. Prinn, R. F. Weiss, and K.-R. Kim. Quantifying aluminum and semiconductor industry perfluorocarbon emissions from atmospheric measurements. *Geophysical Research Letters*, 41(13):4787–4794, 2014. doi:10.1002/2014gl059783.

[275] R. G. Prinn, R. F. Weiss, J. Arduini, T. Arnold, H. L. DeWitt, P. J. Fraser, A. L. Ganesan, J. Gasore, C. M. Harth, O. Hermansen, J. Kim, P. B. Krummel, S. Li, Z. M. Loh, C. R. Lunder, M. Maione, A. J. Manning, B. R. Miller, B. Mitrevski, J. Mühle, S. O'Doherty, S. Park, S. Reimann, M. Rigby, T. Saito, P. K. Salameh, R. Schmidt, P. G. Simmonds, L. P. Steele, M. K. Vollmer, R. H. Wang, B. Yao, Y. Yokouchi, D. Young, and L. Zhou. History of chemically and radiatively important atmospheric gases from the advanced global atmospheric gases experiment (AGAGE). *Earth System Science Data*, 10(2):985–1018, 2018. doi:10.5194/essd-10-985-2018.

[276] T. Arnold, J. Mühle, P. K. Salameh, C. M. Harth, D. J. Ivy, and R. F. Weiss. Automated Measurement of Nitrogen Trifluoride in Ambient Air. *Analytical Chemistry*, 84(11):4798–4804, 2012. doi:10.1021/ac300373e.

[277] B. R. Miller, R. F. Weiss, P. K. Salameh, T. Tanhua, B. R. Greally, J. Mühle, and P. G. Simmonds. Medusa: A Sample Preconcentration and GC/MS Detector System for in Situ Measurements of Atmospheric Trace Halocarbons, Hydrocarbons, and Sulfur Compounds. *Analytical Chemistry*, 80(5):1536–1545, 2008. doi:10.1021/ac702084k.

[278] X. Fang, A. Stohl, Y. Yokouchi, J. Kim, S. Li, T. Saito, S. Park, and J. Hu. Multiannual Top-Down Estimate of HFC-23 Emissions in East Asia. *Environmental Science & Technology*, 49(7):4345–4353, 2015. doi:10.1021/es505669j.

[279] Y. S. Kim, D. I. Kim, S.Y. Cho, M.H. Kim, K.M. Yang, H.Y. Lee, and S.H. Han. Statistical Analysis for Organ Weights in Korean Adult Autopsies. *Korean J Anat*, 42:219–224, 2009.

[280] S. Li, J. Kim, K.-R. Kim, J. Muühle, S.-K. Kim, M.-K. Park, A. Stohl, D.-J. Kang, T. Arnold, C. M. Harth, P. K. Salameh, and R. F. Weiss. Emissions of Halogenated Compounds in East Asia Determined from Measurements at Jeju Island, Korea. *Environmental Science & Technology*, 45(13):5668–5675, 2011. doi:10.1021/es104124k.

[281] D. J. Thomson. Criteria for the selection of stochastic models of particle trajectories in turbulent flows. *Journal of Fluid Mechanics*, 180(-1):529, 1987. doi:10.1017/s0022112087001940.

[282] B. R. Miller, M. Rigby, L. J. M. Kuijpers, P. B. Krummel, L. P. Steele, M. Leist, P. J. Fraser, A. McCulloch, C. Harth, P. Salameh, J. Mühle, R. F. Weiss, R. G. Prinn, R. H. J. Wang, S. O'Doherty, B. R. Greally, and P. G. Simmonds. HFC-23 (CHF_3 emission trend response to HCFC-22 ($CHClF_2$ production and recent HFC-23 emission abatement measures. *Atmospheric Chemistry and Physics*, 10(16):7875–7890, 2010. doi:10.5194/acp-10-7875-2010.

[283] J. Lin, D. Brunner, C. Gerbig, A. Stohl, A. Luhar, and P. Webley, editors. *Lagrangian Modeling of the Atmosphere*. American Geophysical Union, Washington, DC, 2012. doi:10.1029/gm200.

[284] D.B. Ryall and R.H. Maryon. Validation of the UK Met. Office's NAME model against the ETEX dataset. *Atmospheric Environment*, 32(24):4265–4276, 1998. doi:10.1016/s1352-2310(98)00177-0.

[285] A. Jones, D. Thomson, M. Hort, and B. Devenish. The U.K. Met Office's Next-Generation Atmospheric Dispersion Model, NAME III. In *Air Pollution Modeling and Its Application XVII*, pages 580–589. Springer US, 2007. doi:10.1007/978-0-387-68854-1_62.

[286] M. J. P. Collen. The unified forecast/climate model. *Meteorol. Mag*, 122:81–94, 1993.

[287] C. L. Lawson, R. J. Hanson, F. T. Krogh, and O. R. Kincaid. Basic linear algebra subprograms for FORTRAN usage. *ACM Trans. Math. Software*, 5:308–323, 1979.

[288] Scientific Data Curation Team. Metadata record for: High resolution temporal profiles in the Emissions Database for Global Atmospheric Research, 2020. doi:10.6084/M9.FIGSHARE.12052887.

[289] K. Klauenberg, G. Wübbeler, B. Mickan, P. Harris, and C. Elster. A tutorial on Bayesian Normal linear regression. *Metrologia*, 52(6):878–892, 2015. doi:10.1088/0026-1394/52/6/878.

[290] A. Stohl, J. Kim, S. Li, S. O'Doherty, J. Mühle, P. K. Salameh, T. Saito, M. K. Vollmer, D. Wan, R. F. Weiss, B. Yao, Y. Yokouchi, and L. X. Zhou. Hydrochlorofluorocarbon and hydrofluorocarbon emissions in East Asia determined by inverse modeling. *Atmospheric Chemistry and Physics*, 10(8):3545–3560, 2010. doi:10.5194/acp-10-3545-2010.

[291] B. Yao, M. K. Vollmer, L. X. Zhou, S. Henne, S. Reimann, P. C. Li, A. Wenger, and M. Hill. In-situ measurements of atmospheric hydrofluorocarbons (HFCs) and perfluorocarbons (PFCs) at the Shangdianzi regional background station, China. *Atmospheric Chemistry and Physics*, 12(21):10181–10193, 2012. doi:10.5194/acp-12-10181-2012.

[292] J. G. J. Olivier, K. M. Schure, and J. A. H. W. Peters. Trends in global CO_2 emissions, PBL Netherlands EnvironmentalAssessment Agency, The Hague/Bilthoven500114022, 2017.

[293] M. Rigby, R. G. Prinn, S. O'Doherty, B. R. Miller, D. Ivy, J. Mühle, C. M. Harth, P. K. Salameh, T. Arnold, R. F. Weiss, P. B. Krummel, L. P. Steele, P. J. Fraser, D. Young, and P. G. Simmonds. Recent and future trends in synthetic greenhouse gas radiative forcing. *Geophysical Research Letters*, 41(7):2623–2630, 2014. doi:10.1002/2013gl059099.

[294] SEMI. World Fab Forecast. http://www.semi.org/en/Store/MarketInformation/fabdatabase/ctr_027238.

[295] C. Michotte, A. K. Pearce, M. G. Cox, and J.-J. Gostely. An approach based on the SIR measurement model for determining ionization chamber efficiency curves, and a study of ^{65}Zn and ^{201}Tl photon emission intensities. *Applied Radiation and Isotopes*, 64:1147–1155, 2006.

[296] 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.

[297] ISO 6145 Gas analysis – Preparation of calibration gas mixtures using dynamic methods. ISO 6145, International Organization for Standardization, Geneva, Switzerland. 11 parts.

[298] J. Viallon, E. Flores, F. Idrees, P. Moussay, R. I. 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.

[299] A. M. H. van der Veen, G. Nieuwenkamp, R. M. Wessel, M. Maruyama, G. S. Heo, Y. Kim, D. M. Moon, B. Niederhauser, M. Quintilii, M. J. T. Milton, M. G. Cox, P. M. Harris, F. R. Guenther, G. 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.

[300] 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.

[301] 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.

[302] ISO 14167 Gas analysis – General quality aspects and metrological traceability of calibration gas mixtures. ISO, International Organization for Standardization, Geneva, Switzerland, 2018. First edition.

[303] S. 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.

[304] S. 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.

[305] A. D. McNaught and A. Wilkinson. *Compendium of Chemical Terminology – IUPAC Recommendations (IUPAC Chemical Data) – Gold Book*. Wiley, 2014.

[306] 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,.

[307] P. 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.

[308] 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.

[309] 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.

[310] 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.

[311] A. M H van der Veen and K. Hafner. Atomic weights in gas analysis. *Metrologia*, 51(1):80, 2014. URL: http://stacks.iop.org/0026-1394/51/i=1/a=80.

[312] K. O. 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.

[313] A. M. H. van der Veen and G. 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.

[314] M. L. Boas. *Mathematical Methods in the Physical Sciences*. John Wiley & Sons Inc, 2005.

[315] P. Reig, T. Shiao, and F. Gassert. Aqueduct water risk framework. Technical report, World Resources Institute, Washington DC, USA, 2013.

[316] E. Borgomeo. *Climate change and water resources: risk-based approaches for decision-making*. PhD thesis, University of Oxford, 2015.

[317] H. Alegre, J. M. Baptista, E. Cabrera Jr, F. C. Patricia Duarte, W. Hirner, W. Merkel, and R. Parena. *Performance Indicators for Water Supply Services*. IWA Publishing, third edition edition, 2016. URL: https://www.ebook.de/de/product/26834154/helena_alegre_jaime_m_baptista_enrique_cabrera_jr_performance_indicators_for_water_supply_services_third_edition.html.

[318] M. Almeida Silva, C. Amado, and D. Loureiro. Propagation of uncertainty in the water balance calculation in urban water supply systems – a new approach based on high-density regions. *Measurement*, 126:356–368, oct 2018. doi:10.1016/j.measurement.2018.05.061.

[319] J. Thornton. *Water Loss Control Manual*. McGraw-Hill Education - Europe, 2002. URL: https://www.ebook.de/de/product/3668496/julian_thornton_water_loss_control_manual.html.

[320] F. Arregui, E. Cabrera, and R. Cobacho. *Integrated water meter management*. IWA Publishing, London, 2006.

[321] ILAC G8:2019 – Guidelines on decision rules and statements of conformity. ILAC, Silverwater, Australia, 2019.

[322] A. S. Ribeiro, D. Loureiro, M. C. Almeida, M. G. Cox, J A. Sousa, M. A. Silva, L. Martins, R. Brito, and A. C. Soares. Uncertainty evaluation of totalization of flow and volume measurements in drinking water supply networks. In *FLOMEKO*, Lisbon, Portugal, 2019.

[323] ISO 5167-1 Measurement of fluid flow by means of pressure differential devices inserted in circular cross-section conduits running full — Part 1: General. ISO, International Organization for Standardization, Geneva, Switzerland, 2003. Second edition.

[324] ISO 5167-2 measurement of fluid flow by means of pressure differential devices inserted in circular cross-section conduits running full — Part 2: Orifice plates. ISO, International Organization for Standardization, Geneva, Switzerland, 2003. First edition.

[325] M. Reader-Harris. ISO flow measurement standards — report on the ISO/TC 30 meeting in November 2006. *Flow Measurement and Instrumentation*, 18(3-4):114–120, June 2007. doi:10.1016/j.flowmeasinst.2007.05.003.

[326] M. J. Reader-Harris and J. A. Sattary. The orifice plate discharge coefficient equation - the equation for ISO 5167-1. In *Proc. 14th North Sea Flow Measurement Workshop*, National Engineering Laboratory, East Kilbride, Glasgow, UK, 1996.

[327] M. Reader-Harris. *Orifice Plates and Venturi Tubes*. Springer-Verlag GmbH, 2015. URL: https://www.ebook.de/de/product/25034945/michael_reader_harris_orifice_plates_and_venturi_tubes.html.

[328] American Petroleum Institute. Manual of Petroleum Measurement Standards, Chapter 14 - Natural Gas Fluids Measurement, Section 3 - Concentric, Square-edged Orifice Meters, Part 1 - General equations and uncertainty guidelines. MPMS API 14.3.1:1990, American Petroleum Institute, Washington DC, USA, 1990. Third edition.

[329] ISO/TR 12767 Measurement of fluid flow by means of pressure differential devices – Guidelines on the effect of departure from the specifications and operating conditions given in ISO 5167. ISO, International Organization for Standardization, Geneva, Switzerland, 2007. Second edition.

[330] American Petroleum Institute. Orifice Metering of Natural Gas and Other Related Hydrocarbon Fluids – Concentric, Square-edged Orifice Meters, Part 1 - General Equations and Uncertainty Guidelines. Manual of Petroleum Measurement Standards API 14.3.1:2012, American Petroleum Institute, Washington DC, USA, 2012. Fourth edition.

[331] J. M. Hobbs and J. S. Humphreys. The effect of orifice plate geometry upon discharge coefficient. *Flow Measurement and Instrumentation*, 1(3):133–140, April 1990. doi:10.1016/0955-5986(90)90002-o.

[332] P. Jepson and R. Chipchase. The effect of plate buckling on orifice meter accuracy. Technical Report No ERS.R.467, British Gas Engineering Research Station, 1973.

[333] P. Jepson and R. Chipchase. Effect of plate buckling on orifice meter accuracy. *Journal of Mechanical Engineering Science*, 17(6):330–337, December 1975. doi:10.1243/jmes_jour_1975_017_047_02.

[334] ISO/TR 9464 Guidelines for the use of ISO 5167:2003. ISO, International Organization for Standardization, Geneva, Switzerland, 2008. Second edition.

[335] M. J. Reader-Harris and W. C. Brunton. The effect of diameter steps in upstream pipework on orifice plate discharge coefficients. In *Proceedings of 5th International Symposium Fluid Flow Measurement*, Washington, DC, 2002.

[336] D. L. George and T. B. Morrow. Orifice meter calibration for backwards-facing orifice plates. GRI Report No 01/0074 on SwRI Project No 18-8890, Gas Research Institute, Chicago, USA, 2001.

[337] T. B. Morrow. Orifice meter expansion factor tests in 4-inch and 6-inch meter tubes. GRI Report GRI-04/0042, Gas Research Institute, Chicago, USA, 2004.

[338] T. B. Morrow. Additional studies of orifice meter installation effects and expansion factor. GRI Report GRI-04/0246 on SwRI Project No 18.10315, Gas Research Institute, Chicago, USA, 2005.

[339] T. B. Morrow and J. T. Park. Baseline conditions for orifice meter calibration (as amended by errata, 1993). GRI Report GRI-92/0097, Gas Research Institute, Chicago, USA, 1992.

[340] M. J. Reader-Harris. The effect on orifice plates of diameter steps in upstream pipework and of swirl. Report No 2002/75 on Project No FDDP02, NEL, East Kilbride, Glasgow UK, 2002.

[341] M. J. Reader-Harris, W. C. Brunton, I. G. Nicholson, and R. Rushworth. Ageing effects on orifice metering. In *Proc. 21st North Sea Flow Measurement Workshop*, Norway, 2003.

[342] M. J. Reader-Harris, D. Hodges, and R. Rushworth. The effect of drain holes in orifice plates on the discharge coefficient. In *Proc. 26th Int. North Sea Flow Measurement Workshop*, St Andrews. NEL, East Kilbride, Glasgow, 2008.

[343] M. J. Reader-Harris N. Barton and D. Hodges. The effect of contaminated orifice plates on the discharge coefficient. In *Proc. 15th FLOMEKO*, Taipei, 2010.

[344] W. Studzinski, U. Karnik, P. La Nasa, T. Morrow, D. Goodson, Z. Husain, and J. Gallagher. White paper on orifice meter installation configurations with and without flow conditioners. White paper prepared for API 14.3 Part 2. GRI Report GRI 99/0262, Gas Research Institute, Chicago, USA, 2001.

[345] W. Studzinski, M. Weiss, J. Geerlings, and J. Attia. Effect of reducers, expanders, a gate valve and two bends in perpendicular planes on orifice meter performance. In *Proc. Flow Measurement 2001 Int. Conf.*, National Engineering Laboratory, East Kilbride, Glasgow, UK, 2001. Peebles: paper 3.1.

[346] 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.

[347] ISO 9300 Measurement of gas flow by means of critical flow Venturi nozzles. ISO, International Organization for Standardization, Geneva, Switzerland, 2005. Second edition.

[348] D. Geropp. Laminare Grenzschichten in Ebenen und Rotationssymmetrischen Lavaldüsen. *Deutsche Luft- und Raumfahrt, Forschungsbericht*, pages 71–90, 1971.

[349] D. Geropp. Laminare Grenzschichten in Überschalldüsen. *Deutsche Luft- und Raumfahrt, Forschungsbericht*, 01 TM 8603-AK/PA 1, 1987.

[350] S. Martens, K. Klauenberg, B. Mickan, C. Yardin, N. Fischer, and C. Elster. EMUE-D4-3-QuantifyUncertaintiesInCalibration, 2020. URL: `https://zenodo.org/record/4016915`, `doi:10.5281/ZENODO.4016915`.

[351] 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.

[352] 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`.

[353] Calibration and Measurement Capabilities of PTB for gas flow rate are documented at www.bipm.org/kcdb/. Last visit on 06/08/2020.

[354] 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.

[355] 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.

[356] 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`.

[357] R. B. Schnabel and E. Eskow. A new modified Cholesky factorization. *SIAM J. Sci. and Stat. Comput.*, 11(6):1136–1158, 1990.

[358] 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`.

[359] 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.

[360] Directive 2009/125/EC of the European Parliament and of the Council on establishing a framework for the setting of ecodesign requirements for energy-related products, 2009.

[361] Comission Regulation (EU) 548/2014 on implementing Directive 2009/125/EC of the European Parliament and of the Council with regard to small, medium and large power transformers, 2014.

[362] IEC 60076-19 Power transformers – Part 19: Rules for the determination of uncertainties in the measurement of the losses on power transformers and reactors. IEC, International Electrotechnical Commission, Geneva, Switzerland, 2013.

[363] A. S. Ribeiro, J. Alves e Sousa, M. G. Cox, A. B. Forbes, L. C. Matias, and L. L. Martins. Uncertainty analysis of thermal comfort parameters. *International Journal of Thermophysics*, 36(8):2124–2149, jul 2015. doi:10.1007/s10765-015-1888-1.

[364] 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.

[365] 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.

[366] J. Rice. *Mathematical statistics and data analysis*. Duxbury Press, Belmont, California, USA, 2nd edition, 1995.

[367] ISO 7726 Ergonomics of the thermal environment – Istruments for measuring physical quantities. ISO, International Organization for Standardization, Geneva, Switzerland, 1998. Second edition.

[368] D. Andrich and P. Pedler. On a law of ordinal error. *Journal of Physics: Conference Series*, 1044:012055, Jun 2018. doi:10.1088/1742-6596/1044/1/012055.

[369] M. Gameiro da Silva, M. M. Santana, and J. 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.

[370] 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.

[371] ISO Guide 35 Reference materials – Guidance for characterization and assessment of homogeneity and stability. ISO, International Organization for Standardization, Geneva, Switzerland, 2017. Fourth edition.

[372] A. 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.

[373] A. Koepke, T. Lafarge, A. Possolo, and B. 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.

[374] 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.

[375] A. 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.

[376] M. Plummer, N. Best, K. Cowles, and K. Vines. Coda: Convergence diagnosis and output analysis for mcmc. *R News*, 6(1):7–11, 2006. URL: https://journal.r-project.org/archive/.

[377] M. Meredith and J. Kruschke. *HDInterval: Highest (Posterior) Density Intervals*, 2020. R package version 0.2.2. URL: https://CRAN.R-project.org/package=HDInterval.

[378] S. 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.

[379] H. Dietrich, H.J. Hotze, B. Jarosch, F.J. Juenger, M. Kaempf, R. Kramer, B. Mickan, B. Nath, H. Polzin, and G. Wendt. 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.

[380] K. A. Masavetas. The mere concept of an ideal gas. *Mathematical and Computer Modelling*, 12(6):651–657, 1989.

[381] EN 12480 Gas meters. Rotary displacement gas meters. CEN, European Committee for Standardization, 2015.

[382] EN 12261 Gas meters. Turbine gas meters. CEN, European Committee for Standardization, 2015.

[383] 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.

[384] J. V. Hajnal and D. L. G. Hill. *Medical Image Registration*. Taylor & Francis Ltd, 2019. URL: https://www.ebook.de/de/product/37917038/medical_image_registration.html.

[385] Jonathan I. Gear, Maurice G. Cox, Johan Gustafsson, Katarina Sjögreen Gleisner, Iain Murray, Gerhard Glatting, Mark Konijnenberg, and Glenn D. Flux. EANM practical guidance on uncertainty analysis for molecular radiotherapy absorbed dose calculations. *European Journal of Nuclear Medicine and Molecular Imaging*, 2018. doi:10.1007/s00259-018-4136-7.

[386] IAEA. *Dosimetry for Radiopharmaceutical Therapy*. International Atomic Energy Agency, Vienna, 2020. To appear.

[387] M. J. Korczynski, M. G. Cox, and P. M. Harris. Convolution and uncertainty evaluation. In P. Ciarlini, E. Felipe, A. B. Forbes, and F. Pavese, editors, *Advanced Mathematical Tools in Metrology VII*, pages 188–195, Singapore, 2006. World Scientific.

[388] Maurice G. Cox and Annarita Lazzari. Modelling and uncertainty of high-accuracy roundness measurement. In *10th Symposium on Advances of Measurement Science, St. Petersburg, Russia*, 2004. www.imeko.org/publications/tc7-2004/IMEKO-TC7-2004-028.pdf. URL: `www.imeko.org/publications/tc7-2004/IMEKO-TC7-2004-028.pdf`.

[389] M. G. Cox, M. P. Dainton, A. B. Forbes, and P. M. Harris. Validation of CMM form and tolerance assessment software. In G. N. Peggs, editor, *Laser Metrology and Machine Performance V*, pages 367–376, Southampton, 2001. WIT Press.

[390] B. Osgood. *Lectures on the Fourier transform and its applications*. American Mathematical Society, Providence, Rhode Island, 2019.

[391] K. P. Burnham and D. R. Anderson. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach 2nd edn*. New York: Springer, 2002.

[392] M. Cox and P. Harris. Polynomial calibration functions revisited: numerical and statistical issues. volume Advanced Mathematical and Computational Tools in Metrology X of *Series on Advances in Mathematics for Applied Sciences*, pages 9–16. World Scientific, 2015.

[393] M. G. Cox, C. Michotte, and A. K. Pearce. Measurement modelling of the international reference system (SIR) for gamma-emitting radionuclides. Technical report, Bureau International des Poids et Mesures, Sèvres, France, 2007. BIPM Monographie BIPM-7.

[394] M. G. Cox, P. M. Harris, P. D. Kenward, and Emma Woolliams. Spectral characteristic modelling. Technical Report CMSC 27/03, National Physical Laboratory, Teddington, UK, 2003.

[395] Maurice G. Cox. Modelling clinical decay data using exponential functions. In A. Iske E. H. Georgoulis and J. Levesley, editors, *Approximation algorithms for complex systems*, pages 184–203, Berlin, 2011. Springer-Verlag.

[396] M. G. Cox. The area under a curve specified by measured values. *Metrologia*, 44:365–378, 2007.

[397] G. Dai, L. Xu, and K. Hahm. Accurate tip characterization in critical dimension atomic force microscopy. *Measurement Science and Technology*, 31(7):074011, May 2020. `doi:10.1088/1361-6501/ab7fd2`.

[398] C. Odin, J.-P. Aimé, Z. El Kaakour, and T. Bouhacina. Tip's finite size effects on atomic force microscopy in the contact mode: simple geometrical considerations for rapid estimation of apex radius and tip angle based on the study of polystyrene latex balls. *Surface Science*, 317(3):321–340, October 1994. `doi:10.1016/0039-6028(94)90288-7`.

[399] T. Caebergs, N. Sebaïhi, and J. Pétry. Modelling pixel size uncertainty for nanoparticle sizing by AFM. *to be published*.

[400] A. Delvallée, N. Feltin, S. Ducourtieux, M. Trabelsi, and J.-F. Hochepied. Toward an uncertainty budget for measuring nanoparticles by AFM. *Metrologia*, 53(1):41–50, December 2015. URL: `https://doi.org/10.1088%2F0026-1394%2F53%2F1%2F41`, `doi:10.1088/0026-1394/53/1/41`.

[401] L. Crouzier, A. Delvallée, S. Ducourtieux, L. Devoille, G. Noircler, C. Ulysse, O. Taché, E. Barruet, C. Tromas, and N. Feltin. Development of a new hybrid approach combining AFM and SEM for the nanoparticle dimensional metrology. *Beilstein Journal of Nanotechnology*, 10:1523–1536, July 2019. doi:10.3762/bjnano.10.150.

[402] U. Katscher, D.-H. Kim, and J. K. Seo. Recent progress and future challenges in MR electric properties tomography. *Computational and mathematical methods in medicine*, 2013, 2013. doi:10.1155/2013/546562.

[403] A. Arduino. *Mathematical methods for magnetic resonance based electric properties tomography*. PhD thesis, Politecnico di Torino, 2018. doi:10.6092/polito/porto/2698325.

[404] T. Voigt, U. Katscher, and O. Doessel. Quantitative conductivity and permittivity imaging of the human brain using electric properties tomography. *Magnetic Resonance in Medicine*, 66(2):456–466, 2011. doi:10.1002/mrm.22832.

[405] K. K. Tha, U. Katscher, S. Yamaguchi, C. Stehning, S. Terasaka, N. Fujima, K. Kudo, K. Kazumata, T. Yamamoto, M. Van Cauteren, et al. Noninvasive electrical conductivity measurement by MRI: a test of its validity and the electrical conductivity characteristics of glioma. *European radiology*, 28(1):348–355, 2018. doi:10.1007/s00330-017-4942-5.

[406] A. Savitzky and M. J. E. Golay. Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry*, 36(8):1627–1639, 1964. URL: http://doi.org/10.1021/ac60214a047, doi:10.1021/ac60214a047.

[407] Alessandro Arduino. EPTlib 0.1.1. https://eptlib.github.io/, 2020. Accessed: 2020-09-25.

[408] M. G. Cox, C. Eiø, G. Mana, and F. Pennecchi. The generalized weighted mean of correlated quantities. *Metrologia*, 43(4):S268–S275, 2006. URL: https://iopscience.iop.org/article/10.1088/0026-1394/43/4/S14, doi:10.1088/0026-1394/43/4/S14.

[409] S.-K. Lee, S. Bulumulla, F. Wiesinger, L. Sacolick, W. Sun, and I. Hancu. Tissue electrical property mapping from zero echo-time magnetic resonance imaging. *IEEE transactions on medical imaging*, 34(2):541–550, 2015. doi:10.1109/TMI.2014.2361810.

[410] J. Schafer, R. Opgen-Rhein, V. Zuber, M. Ahdesmaki, A. P. Duarte Silva, and K. Strimmer. corpcor 1.6.9: Efficient Estimation of Covariance and (Partial) Correlation. http://www.strimmerlab.org/software/corpcor/, 2017. Accessed: 2020-09-29.

[411] R. Opgen-Rhein and K. Strimmer. Accurate ranking of differentially expressed genes by a distribution-free shrinkage approach. *Statistical Applications in Genetics and Molecular Biology*, 6(1), 2007. doi:10.2202/1544-6115.1252.

[412] J. Schafer and K. Strimmer. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical Applications in Genetics and Molecular Biology*, 4(1), 2005. doi:10.2202/1544-6115.1175.

[413] *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.

[414] S. M. Lewis and S. Kumari. Chapter 7: Haemoglobinometry. In *Guidelines on standard operating procedures for haematology*. WHO, 1999.

[415] BS 3985:2003 Haemoglobincyanide (cyanmethhaemoglobin) preparation as a standard for spectrometric haemoglobin. BS, British Standards Institution, BSI Group, UK, 2003.

[416] DIN 58931:2010 Haematology – determination of haemoglobin concentration in blood – reference method. DIN, German Standards Institution, Beuth-Verlag, Berlin, Germany, 2010.

[417] 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.

[418] 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.

[419] 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`.

[420] 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.

[421] 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.

[422] 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.

[423] 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.

[424] 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`.

[425] R Feistel, J. W Lovell-Smith, P Saunders, and S Seitz. Uncertainty of empirical correlation equations. *Metrologia*, 53(4):1079, 2016.

[426] 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`.

[427] 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.

[428] 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.

[429] 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.

[430] R. Enns. Capsule endoscopy in patients with iron deficiency. *J. Gastroenterol. Hepatol.*, 8:847–849, 2012.

[431] N. J. White. Anaemia and malaria. *Malaria Journal*, 17:371–387, 2018. doi:10.1186/s12936-018-2509-9.

[432] 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.

[433] 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.

[434] 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.

[435] 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.

[436] 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.

[437] 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.

[438] 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.

[439] A. Furtado, J. Pereira, J.A. Sousa, M.G. Cox, and A.S. Ribeiro. Uncertainty evaluation in rheology measurements. In *Proc. AMCTM XII*. World Scientific, 2000.

[440] D. J. Highgate and R. W. Whorlow. End effects and particle migration effects in concentric cylinder rheometry. *Rheologica Acta*, 8(2):142–151, July 1969. doi:10.1007/bf01984650.

[441] Harro Bauer and Norbert Boese. Rheological properties of a micelle system in solution to be used as reference liquid with viscoplastic behaviour. In *Third European Rheology Conference and Golden Jubilee Meeting of the British Society of Rheology*, pages 37–40. Springer Netherlands, 1990. doi:10.1007/978-94-009-0781-2_20.

[442] V.C. Kelessidis, R. Maglione, and G. Bandelis. On the end-effect correction for couette type oil-field direct-indicating viscometers for newtonian and non-newtonian fluids. *Journal of Petroleum Science and Engineering*, 71(1-2):37–46, March 2010. doi:10.1016/j.petrol.2010.01.001.

[443] Ondřej Wein, Věra Pěnkavová, and Jaromír Havlica. End effects in rotational viscometry II. pseudoplastic fluids at elevated reynolds number. *Rheologica Acta*, 54(11-12):903–914, October 2015. doi:10.1007/s00397-015-0878-3.

[444] H. Giesekus and G. Langer. Die bestimmung der wahren fließkurven nicht-newtonscher flüsigkeiten und plastischer stoffe mit der methode der repräsentativen viskosität. *Rheologica Acta*, 16(1):1–22, January 1977. doi:10.1007/bf01516925.

[445] ISO 3219 Plastics - Polymers/resins in the liquid state or as emulsions or dispersions – Determination of viscosity using a rotational viscometer with defined shear rate. ISO, International Organization for Standardization, Geneva, Switzerland, 1993. Second edition.

[446] DIN 53019–2 Viscosimetry - Measurement of viscosities and flow curves by means of rotation viscosimeters – Part 2: Viscosimeter calibration and determination of the uncertainty of measurement. DIN, German Institute for Standardisation, Berlin, Germany, 2001. First edition.

[447] ASTM e2510–07 Standard Test Method for Torque Calibration or Conformance of Rheometers. ASTM, American Society for Testing and Materials, West Conshohocken, PA, 2013.

[448] ASTM e2509–14 Standard Test Method for Temperature Calibration of Rheometers in Isothermal Mode. ASTM, American Society for Testing and Materials, West Conshohocken, PA, 2014.

[449] R. Boudjemaa, M.G. Cox, A.B. Forbes, and P.M. Harris. Automatic differentiation and its applications to metrology. In *Advanced Mathematical and Computational Tools in Metrology VI*, pages 170–179, Singapore, 2004. World Scientific.

[450] Patrick Ballereau, Nikola Pelevic, Jan Jette Blange, Inge Van Andel, Ernad Borovac, Salvatore Lorefice, Andreia Furtado, Marc de Huu, Hugo Bissig, Henning Wolf, Ronald Pagel, Svetlana Hageraats-Ponomareva, Fionn Iversen, and Maria Teresa Cidade. Sensor development and calibration method for inline detection of viscosity and solids content of non-newtonian fluids. In Bernard Larquier, editor, *17th International Congress of Metrology*. EDP Sciences, 2015. doi:10.1051/metrology/20150003004.

[451] P. T. Lee, F. Thompson, and H. Thimbleby. Analysis of infusion pump error logs and their significance for health care. *British Journal of Nursing*, 21(Sup8):S12–S20, April 2012. doi:10.12968/bjon.2012.21.sup8.s12.

[452] L. Heinemann, G. A. Fleming, J. R. Petrie, R. W. Holl, R. M. Bergenstal, and A. L. Peters. Insulin pump risks and benefits: A clinical appraisal of pump safety standards, adverse event reporting, and research needs. *Diabetes Care*, page dc150168, March 2015. doi:10.2337/dc15-0168.

[453] H. Ma, M. A. Lovich, and R. A. Peterfreund. Quantitative analysis of continuous intravenous infusions in pediatric anesthesia: safety implications of dead volume, flow rates, and fluid delivery. *Pediatric Anesthesia*, 21(1):78–86, December 2010. doi:10.1111/j.1460-9592.2010.03475.x.

[454] M. Husch. Insights from the sharp end of intravenous medication errors: implications for infusion pump technology. *Quality and Safety in Health Care*, 14(2):80–86, April 2005. doi:10.1136/qshc.2004.011957.

[455] I. Godinho E. Batista, N. Almeida and E. Filipe. Uncertainty calculation in gravimetric microflow measurements. *Advanced Mathematical and Computational Tools in Metrology and Testing X*, pages 98–104, 2015. URL: https://www.worldscientific.com/doi/abs/10.1142/9789814678629_0011, doi:10.1142/9789814678629_0011.

[456] ISO 4787 – Laboratory Glassware — Volumetric Glassware — Methods for use and Testing of Capacity. ISO, International Organization for Standardization, Geneva, Switzerland, 2010.

[457] H. Bissig, H. T. Petter, P. Lucas, E. Batista, E. Filipe, N. Almeida, L. F. Ribeiro, J. Gala, R. Martins, B. Savanier, F. Ogheard, A. Koustrup Niemann, J. Lötters, and W. Sparreboom. Primary standards for measuring flow rates from 100 nl/min to 1 ml/min – gravimetric principle. *Biomedical Engineering / Biomedizinische Technik*, 60(4), January 2015. doi:10.1515/bmt-2014-0145.

[458] S. Demeyer, N. Fischer, and C. Elster. Guidance on Bayesian uncertainty evaluation for a class of GUM measurement models. *Metrologia*, 58, 08 2020. doi:10.1088/1681-7575/abb065.

[459] C. Robert and G. Casella. *Monte Carlo statistical methods*. Springer Science & Business Media, 2013.

[460] K. Klauenberg and C. Elster. Markov chain Monte Carlo methods: an introductory example. *Metrologia*, 53(1):S32–S39, jan 2016. doi:10.1088/0026-1394/53/1/s32.

[461] C. Fonnesbeck J. Salvatier, T.V. Wiecki. Probabilistic programming in python using pymc3. *PeerJ Computer Science*, 2:e55, 2016. doi:10.7717/peerj-cs.55.

[462] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. JCGM 200: International vocabulary of metrology – Basic and general concepts and associated terms (VIM), ANNOTATION (informative) 9 June 2016, https://jcgm.bipm.org/vim/en/2.13.html , 2012.

[463] SHTM-01-01-C. Scottish Health Technical Memorandum 01-01: Decontamination of medical devices in a Central Decontamination Unit. Part C: Sterilization by steam. Health Facilities Scotland, 2018.

[464] SHTM-01-01-B. Scottish Health Technical Memorandum 01-01: Decontamination of medical devices in a Central Decontamination Unit. Part B: Test equipment/methods. Health Facilities Scotland, 2018.

[465] PHSS. Guidance Document for Cold Storage Temperature Monitoring and Mapping for Blood Products. The Pharmaceutical and Healthcare Sciences Society, 2013.

[466] Sabine Bauinger, Andreas Marn, Emil Göttlich, and Franz Heitmeir. Influence of Pressure Fluctuations on the Mean Value of Different Pneumatic Probes. *International Journal of Turbomachinery, Propulsion and Power*, 2(3):13, 2017. doi:10.3390/ijtpp2030013.

[467] NATO. Guide to in-flight thrust measurement of turbojets and fan engines, NATO Advisory Group for Aerospace Research and Development, report AGARD-AG-237. resreport AGARD-AG-237, NATO Advisory Group for Aerospace Research and Development, 1979.

[468] Alvaro Cuerva and Angel Sanz-Andrés. On sonic anemometer measurement theory. *Journal of Wind Engineering and Industrial Aerodynamics*, 88(1):25–55, 2000. doi:10.1016/s0167-6105(00)00023-4.

[469] Alvaro Cuerva and Angel Sanz-Andrés. Sonic anemometry of planetary atmospheres. *Journal of Geophysical Research*, 108(E4), 2003. doi:10.1029/2002je001944.

[470] D. Banfield and R. Dissly. A Martian sonic anemometer. In *2005 IEEE Aerospace Conference*. IEEE, 2005. doi:10.1109/aero.2005.1559354.

[471] J C Wyngaard. Cup, propeller, vane, and sonic anemometers in turbulence research. *Annual Review of Fluid Mechanics*, 13(1):399–423, 1981. doi:10.1146/annurev.fl.13.010181.002151.

[472] R. Sozzi and M. Favaron. Sonic anemometry and thermometry: theoretical basis and data-processing software. *Environmental Software*, 11(4):259–270, 1996. doi:10.1016/s0266-9838(96)00046-9.

[473] Beverley McKeon, Geneviève Comte-Bellot, John Foss, Jerry Westerweel, Fulvio Scarano, Cameron Tropea, James Meyers, Joseph Lee, Angelo Cavone, Richard Schodl, Manoochehr Koochesfahani, Yiannis Andreopoulos, Werner Dahm, John Mullin, James Wallace, Petar Vukoslavčević, Scott Morris, Eric Pardyjak, and Alvaro Cuerva. Velocity, vorticity, and mach number. In *Springer Handbook of Experimental Fluid Mechanics*, pages 215–471. Springer Berlin Heidelberg, 2007. doi:10.1007/978-3-540-30299-5_5.

[474] J. C. Kaimal and J. E. Gaynor. Another look at sonic thermometry. *Boundary-Layer Meteorology*, 56(4):401–410, 1991. doi:10.1007/bf00119215.

[475] M. Bramanti, E. A. Salerno, A. Tonazzini, S. Pasini, and A. Gray. An acoustic pyrometer system for tomographic thermal imaging in power plant boilers. *IEEE Transactions on instrumentation and measurement*, 45(1):159–167, Feb 1996. doi:10.1109/19.481329.

[476] S.P. Zhang, L.S. An, G.Q. Shen, and Y.G. Niu. Acoustic pyrometry system for environmental protection in power plant boilers. *Boundary-Layer Meteorology*, 23(2):24–35, 2014. doi:10.3808/jei.201400265.

[477] R Otero, K T Lowe, and W F Ng. Extension of sonic anemometry to high subsonic mach number flows. *Measurement Science and Technology*, 28(3):035306, 2017. doi:10.1088/1361-6501/aa54ed.

[478] R Otero, K T Lowe, and W F Ng. Non-intrusive acoustic measurement of flow velocity and temperature in a high subsonic mach number jet. *Measurement Science and Technology*, 29(1):015106, 2017. doi:10.1088/1361-6501/aa92a9.

[479] R Otero, K T Lowe, W F Ng, L Ma, and C-Y Kim. Nonintrusive gas-turbine engine-exhaust characterization using acoustic measurements. *Journal of Propulsion and Power*, 34(3):730–738, 2018. doi:10.2514/1.b36579.

[480] R Otero, K T Lowe, W F Ng, and K A Silas. Coupled velocity and temperature acoustic tomography in heated high subsonic mach number flows. *Measurement Science and Technology*, 30(10):105901, 2019. doi:10.1088/1361-6501/ab24a3.

[481] R. Boudjemaa, M. G. Cox, A. B. Forbes, and P. M. Harris. Automatic differentiation techniques and their application in metrology. Technical Report CMSC 26/03, National Physical Laboratory, Teddington, UK, 2003.

[482] The MathWorks Inc. *Symbolic Math Toolbox*. Natick, Massachusetts, United States, 2019. URL: `https://www.mathworks.com/help/symbolic/`.

[483] R. D. Otero. *Compressible Flow Characterization Using Non-Intrusive Acoustic Measurements*. PhD thesis, Virginia Tech., 2017.

[484] S. D. Conte and C. de Boor. *Elementary Numerical Analysis: An Algorithmic Approach*. McGraw-Hill, 1972.

[485] J. R. R. A. Martins, I. Kroo, and J. Alonso. An automated method for sensitivity analysis using complex variables. In *38th Aerospace Sciences Meeting and Exhibit*. American Institute of Aeronautics and Astronautics, 2000. `doi:10.2514/6.2000-689`.

[486] J. R. R. A. Martins, P. Sturdza, and J. J. Alonso. The connection between the complex-step derivative approximation and algorithmic differentiation. In *39th Aerospace Sciences Meeting and Exhibit*, number AIAA-2000-0921. American Institute of Aeronautics and Astronautics, 2001.

[487] J. N. Lyness and C. B. Moler. Numerical differentiation of analytic functions. *SIAM J. Numer. Anal.*, 4:202–210, 1967.

[488] A. H. Al-Mohy and N. J. Higham. The complex step approximation to the Fréchet derivative of a matrix function. *Numer. Algor.*, 53:133–148, 2010.

[489] M. Cox, A. Forbes, P. Harris, and C. Matthews. Numerical aspects in the evaluation of measurement uncertainty. In *Uncertainty Quantification in Scientific Computing*, pages 180–194. Springer, 2012.

[490] 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.

[491] ISO 6507-1:2018 Metallic materials - Vickers hardness test - Part 1: Test method. ISO, International Organization of Standardization, Geneva, Switzerland, 2018.

[492] ISO 4545-1:2017 Metallic materials - Knopp hardness test - Part 1: Test method. ISO, International Organization of Standardization, Geneva, Switzerland, 2017.

[493] ISO 6508-1:2016 Metallic materials - Rockwell hardness test - Part 1: Test method. ISO, International Organization of Standardization, Geneva, Switzerland, 2016.

[494] ISO 6506-1:2014 Metallic materials - Brinell hardness test - Part 1: Test method. ISO, International Organization of Standardization, Geneva, Switzerland, 2014.

[495] ISO 14577:2015 Metallic materials - Instrumented indentation test for hardness and material parameters - Part 1: Test method and Part 2: Verification and calibration of testing machines. ISO, International Organization of Standardization, Geneva, Switzerland, 2015.

[496] EURAMET cg-16 Version 2.0 (03/2011) Guidelines on the Estimation of Uncertainty in Hardness Measurements, 2011.

[497] ISO/AWI TR 8463:2021 Metallic Materials – Strategy for a Common Frame Work to determine Measurement Uncertainty in Mechanical Testing. ISO, International Organization of Standardization, Geneva, Switzerland, 2021.

[498] B. Peeters, K. Peeters, H. Van der Auweraer, T. Olbrechts, F. Demeester, and L. 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.

[499] Krypton Help Pages on K400/K600 Hardware & Software Guide, April 2003.

[500] International standard for laboratories (ISL). WADA, Montreal, Quebec, Canada, 2021.

[501] Decision limits for the confirmatory quantification of exogenous threshold substances by chromatography-based analytical methods. WADA, Montreal, Quebec, Canada, 2021. Version 1.

[502] R. Bettencourt da Silva and A. Williams. Eurachem/CITAC Guide: Setting and using target uncertainty in chemical measurement. Technical report, 2015. 1st edition.

[503] WADA Prohibited List. WADA, Montreal, Quebec, Canada, 2021.

[504] A theoretical view of coherent sampling. Application Note AN9705, RENEAS, 1997. Rev. 0. URL: https://www.renesas.com/sg/en/www/doc/application-note/an9705.pdf.

[505] Coherent sampling vs. window sampling. Application Note 1040, MAXIM, 2002. URL: https://www.maximintegrated.com/en/design/technical-documents/tutorials/1/1040.html.

[506] M. D. Springer. *The algebra of random variables*. Wiley, New York, 1979.

[507] A. Papoulis and S. Unnikrishna Pillai. *Probability, Random Variables and Stochastic Processes*. McGraw-Hill, Europe, 2002.

[508] H. Dwight. *Tables of integrals and other mathematical data*. Macmillan, New York, 1961.

[509] Uncertainty issues in multi-channel data acquisition systems. In *IEEE 2020 I^2MTC Conference*, Dubrovnik, Croatia, May 2020.

[510] IEC 60751 Industrial platinum resistance thermometers and platinum temperature sensors. IEC, International Electrotechnical Commission, Geneva, Switzerland, 2008. Second edition.

[511] S. Lipshitz, R. Wannamaker, and J. Vanderkooy. Quantization and dither: A theoretical survey. *Journal of the Audio Engineering Society*, 40:355–374, 05 1992.

Alphabetical index

Acronyms

AFM Atomic Force Microscopy. 171–175, 177, 180, 187, 409, 418–420, 423, 426, 427

ANOVA analysis of variance. 166, 174, 247, 379, 380, 383, 384

BIPM Bureau International des Poids et Mesures. 99

CIPM International Committee of Weights and Measures. 99–102, 104, 106–108

CMC calibration and measurement capability. 91–95, 97, 98, 106

CMM coordinate measuring machine. 345

CRM certified reference material. 220, 510, 513, 514, 521, 524, 538, 541, 542

DoE Design of Experiment. 171, 174, 176–178, 181, 185–187

EA European co-operation for Accreditation. 21

GLS generalized least squares. 102

GUF GUM uncertainty framework. 77, 90, 163, 217, 219, 223, 224, 371, 374, 375, 377, 378, 395, 450–454, 456, 462, 463, 467, 499, 509, 515, 545, 561

GUM Guide to the expression of Uncertainty in Measurement. xv, 1, 2, 29, 57, 109–111, 115–118, 122, 123, 125, 126, 129, 131, 133, 137, 151, 152, 161, 162, 173, 185, 189, 196, 200, 205, 213, 216, 225, 244, 263, 269, 273–275, 278, 312, 315, 318, 341, 342, 345, 348, 350, 353, 369–371, 374, 378, 384, 403, 419, 425, 427, 429, 439, 440, 442, 449, 451–457, 461, 462, 467, 469, 471, 476, 483, 488, 501–505, 507, 508, 511, 515, 517, 542

GUM-S1 GUM Supplement 1. 21, 205, 210, 213, 216, 217, 224, 273, 278, 344, 353, 370, 371, 374, 375, 395, 403, 425, 449, 452, 453, 457, 462, 463, 467, 499

GUM-S2 GUM Supplement 2. 57, 109, 111, 115, 122, 161, 344, 347, 496, 499

IEC International Electrotechnical Commission. 45

ILAC International Laboratory Accreditation Cooperation. 45

ISO International Organization for Standardization. 45

JCGM Joint Committee on Guides in Metrology. xv, 1, 99, 133, 269, 273, 483

KC key comparison. 91–95, 97–104, 106–108

KCRV key comparison reference value. 91, 93, 96, 98, 100–102, 108

LIMS laboratory information management system. 2

LPU law of propagation of uncertainty. 29, 57, 71, 73, 111, 114, 117, 152, 161, 173, 185, 189, 200, 201, 210, 219, 221, 222, 241, 262, 263, 266, 271, 274, 275, 278, 279, 315, 318, 345, 353, 369, 371, 429, 430, 452, 462, 472, 475, 488, 489, 491, 492, 496, 517

MAP maximum a posteriori. 95, 97

MCM Monte Carlo method. 29, 71, 73, 76, 77, 82, 89, 90, 210, 216, 217, 219, 221, 223–225, 266, 273, 278, 279, 343, 345, 353, 374, 375, 378, 395, 403, 416, 417, 452–456, 462, 463, 469, 473, 489, 493, 499, 509, 517

MCMC Markov Chain Monte Carlo. 21, 22, 174, 178, 179, 181–184, 187, 384

MPE maximum permissible error. 196, 199

MRA CIPM Mutual Recognition Arrangement. 99, 100

NMI national metrology institute. 92, 99, 110, 450

OLS ordinary least-squares method. 178, 348, 442, 444, 507

PDF probability density function. 53, 145, 146, 150, 153–155, 157, 167–169, 178, 179, 181–185, 187, 189, 191–195, 197, 198, 200, 201, 212, 213, 216, 224, 263, 266, 371, 374, 375, 377, 378, 380, 384, 385, 422, 423, 425, 426, 452–456, 462, 463, 472–474, 505–507, 517, 526

PID Proportional Integral Derivative. 172, 419

RM reference material. 513

RMO regional metrology organisation. 99–104, 106, 108

SBI single burning item. 77–79, 83

SI Système International d'unités. 171, 172, 174, 196, 449, 529

SPM Scanning Probe Microscopy. 419, 423

VIM International Vocabulary of Metrology. 471

WLS weighted least-squares method. 345, 348, 442, 444, 507, 508

WM weighted mean. 101, 104

WMO World Meteoroligical Organization. 205, 209, 210

WTLS weighted total least-squares method. 341, 343, 345, 347, 439, 440, 442–445