

Guide to the expression of uncertainty in measurement
— Part 5: Examples

Guide pour l'expression de l'incertitude de mesure — Partie 5:
Exemples

JCGM GUM-5:2025-12-03

© JCGM 2025

Copyright of this JCGM guidance document is shared jointly by the JCGM member organizations (BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP and OIML).

Copyright

Even if electronic versions are available free of charge on the website of one or more of the JCGM member organizations, economic and moral copyrights related to all JCGM publications are internationally protected. The JCGM does not, without its written authorisation, permit third parties to rewrite or re-brand issues, to sell copies to the public, or to broadcast or use on-line its publications. Equally, the JCGM also objects to distortion, augmentation or mutilation of its publications, including its titles, slogans and logos, and those of its member organizations.

Official versions and translations

The only official versions of documents are those published by the JCGM, in their original languages.

The JCGM's publications may be translated into languages other than those in which the documents were originally published by the JCGM. Permission must be obtained from the JCGM before a translation can be made. All translations should respect the original and official format of the formulae and units (without any conversion to other formulae or units), and contain the following statement (to be translated into the chosen language):

All JCGM's products are internationally protected by copyright. This translation of the original JCGM document has been produced with the permission of the JCGM. The JCGM retains full internationally protected copyright on the design and content of this document and on the JCGM's titles, slogan and logos. The member organizations of the JCGM also retain full internationally protected right on their titles, slogans and logos included in the JCGM's publications. The only official version is the document published by the JCGM, in the original languages.

The JCGM does not accept any liability for the relevance, accuracy, completeness or quality of the information and materials offered in any translation. A copy of the translation shall be provided to the JCGM at the time of publication.

Reproduction

The JCGM's publications may be reproduced, provided written permission has been granted by the JCGM. A sample of any reproduced document shall be provided to the JCGM at the time of reproduction and contain the following statement:

This document is reproduced with the permission of the JCGM, which retains full internationally protected copyright on the design and content of this document and on the JCGM's titles, slogans and logos. The member organizations of the JCGM also retain full internationally protected right on their titles, slogans and logos included in the JCGM's publications. The only official versions are the original versions of the documents published by the JCGM.

Disclaimer

The JCGM and its member organizations have published this document to enhance access to information about metrology. They endeavor to update it on a regular basis, but cannot guarantee the accuracy at all times and shall not be responsible for any direct or indirect damage that may result from its use. Any reference to products of any kind (including but not restricted to any software, data or hardware) or links to websites, over which the JCGM and its member organizations have no control and for which they assume no responsibility, does not imply any approval, endorsement or recommendation by the JCGM and its member organizations.

Contents

	Page
Foreword	vii
Introduction	viii
0 Scope	1
1 Measurement of pH: linear interpolation	2
1.1 Preamble	2
1.2 pH of a test solution	2
1.3 Specification of the measurand	3
1.4 Measurement model	3
1.4.1 Generic approach to two-point calibration	3
1.4.2 Generic approach to multi-point calibration	5
1.4.3 Metrological extension	5
1.5 pH measured at a stipulated temperature	5
1.6 pH measurement accounting for temperature	6
1.7 Uncertainty propagation	7
1.8 pH estimation at a stipulated temperature and associated uncertainty evaluation . .	8
1.9 pH estimation accounting for temperature and associated uncertainty evaluation . .	9
1.10 Interpretation of results	10
1.10.1 Validation of results	10
1.10.2 Resolution	10
1.10.3 Correlation	10
2 Determination of benzo[a]pyrene	13
2.1 Preamble	13
2.2 Specification of the measurand	13
2.3 Measurement model	14
2.4 Uncertainty propagation	15
2.4.1 JCGM 100:2008 uncertainty framework (GUF)	15
2.4.2 Monte Carlo method (MCM)	16
2.5 Reporting the result	17
2.6 Interpretation of results	17
3 Relative molecular mass of glucose	20
3.1 Preamble	20
3.2 Measurement model	20
3.3 Calculation and uncertainty evaluation	21
3.4 Discussion	22
4 Gravimetric mixture preparation and calculation of composition	23
4.1 Preamble	23
4.2 Measurement model	23
4.3 Uncertainty evaluation	26
4.3.1 Correlations	26
4.3.2 General considerations	26
4.3.3 Weighing data	26
4.3.4 Molar masses	27
4.3.5 Composition of the parent gases	29
4.3.6 Composition of the gas mixture	29
4.4 Discussion	32
5 Greenhouse gas emission inventories	33
5.1 Preamble	33
5.1.1 General	33
5.1.2 Greenhouse gas inventories	33
5.1.3 Combination and correlation	35
5.1.4 IPCC and JCGM reporting guidelines	35
5.2 Specification of the measurand	36

5.3	Measurement model	36
5.4	Uncertainty propagation	36
5.5	Interpretation of results	40
5.6	Uncertainty guidance promoted by the IPCC and the JCGM	40
6	Simple linear measurement models	42
6.1	Preamble	42
6.2	Normally distributed input quantities	42
6.3	Rectangularly distributed input quantities with the same width	44
6.4	Rectangularly distributed input quantities with different widths	46
7	Second-order effects in a nonlinear measurement model: calibration of weights	48
7.1	Preamble	48
7.2	Measurement model	48
7.3	Measurand estimate and associated uncertainty	49
7.4	Uncertainty budget	51
	7.4.1 GUF	51
	7.4.2 MCM	52
8	Gauge block calibration	53
8.1	Preamble	53
8.2	Measurement model	53
8.3	GUF	55
8.4	MCM	55
	8.4.1 Assignment of PDFs	55
8.5	Results	57
9	GUM uncertainty evaluation for least-squares versus Bayesian inference - Calibration of a torque measuring system	59
9.1	Preamble	59
9.2	Experimental	59
9.3	Specification of the measurand	59
9.4	Measurement model	60
	9.4.1 General	60
	9.4.2 Ordinary and weighted least-squares	60
	9.4.3 Statistical model	61
9.5	Uncertainty evaluation	61
	9.5.1 Law of propagation of uncertainty	61
	9.5.2 Bayesian uncertainty analysis	62
9.6	Reporting the result	64
9.7	Discussion	64
10	Conformity assessment of mass concentration of total suspended particulate matter in air	67
10.1	Preamble	67
10.2	Objective and data	67
10.3	Specification of the measurand	67
10.4	Measured values and associated measurement uncertainty	67
10.5	Tolerance limits	68
10.6	Decision rule and conformity assessment	68
	10.6.1 General	68
	10.6.2 Bayesian framework	69
	10.6.3 Global risks	69
	10.6.4 Specific risks	71
10.7	Interpretation of results	71
11	Effect of considering a 2D image as a set of pixels on a computed quantity	73
11.1	Preamble	73
11.2	Estimation of organ or tumour mass	73
11.3	Measurement model	74
11.4	Model selection	76

11.5	Area estimation and associated uncertainty evaluation	78
11.6	Reporting the result	80
11.7	Interpretation of results	80
12	Between-bottle homogeneity of reference materials	82
12.1	Preamble	82
12.2	Specification of the measurand	82
12.3	Measurement model	83
12.4	Implementation	85
12.5	Reporting the result	88
13	Measurement of Celsius temperature using a resistance thermometer	90
13.1	Preamble	90
13.2	Measurement of a single Celsius temperature	90
13.3	Measurement of several Celsius temperatures	92
14	Activity of a radioactive source corrected for decay	96
14.1	Preamble	96
14.2	Measurement model	96
14.3	Uncertainty evaluation	96
14.4	Discussion	98
15	Breaking force of steel wire rope	99
15.1	Preamble	99
15.2	Measurement model	99
15.3	Assignment of PDFs	100
15.3.1	Breaking force observations	100
15.3.2	Correction effects	100
15.4	GUF	101
15.5	MCM	101
15.6	Discussion	102
16	Comparison loss in microwave power meter calibration	104
16.1	Preamble	104
16.2	Formulation	104
16.3	Propagation and summarizing: uncorrelated case	106
16.3.1	General	106
16.3.2	Input estimate $x_1 = 0$	106
16.3.3	Input estimate $x_1 = 0.010$	108
16.3.4	Input estimate $x_1 = 0.050$	109
16.3.5	Discussion	109
16.4	Propagation and summarizing: correlated case	110
16.4.1	General	110
16.4.2	Input estimates $x_1 = 0, 0.010$ and 0.050	111
16.4.3	Discussion	112
16.5	Analytic calculations	112
16.5.1	General	112
16.5.2	Expectation and standard deviation	112
16.5.3	Analytic solution for zero estimate of the voltage reflection coefficient having associated zero covariance	113
16.5.4	GUF — Uncorrelated input quantities	115
16.5.5	GUF — Correlated input quantities	115
16.6	Conclusion	116
Annexes		117
Annex A	Conventions and notation	117
Annex B	Glossary of principal symbols	119
B.1	Generic variables and functions	119
B.2	Probability distributions	119
References		120

Alphabetical index	131
Abbreviations	134

Final draft

Foreword

In 1997 a Joint Committee for Guides in Metrology (JCGM), chaired by the Director of the Bureau International des Poids et Mesures (BIPM), was created by the seven international organizations that had originally in 1993 prepared the ‘Guide to the expression of uncertainty in measurement’ and the ‘International vocabulary of basic and general terms in metrology’. The JCGM assumed responsibility for these two documents from the Technical Advisory Group 4 of the International Organization for Standardization (ISO/TAG4).

The Joint Committee is formed by the BIPM with the International Electrotechnical Commission (IEC), the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC), the International Laboratory Accreditation Cooperation (ILAC), the International Organization for Standardization (ISO), the International Union of Pure and Applied Chemistry (IUPAC), the International Union of Pure and Applied Physics (IUPAP), and the Organisation Internationale De Métrologie Légale (OIML).

JCGM has two Working Groups. Working Group 1, ‘Expression of uncertainty in measurement’, has the task to promote the use of the ‘Guide to the expression of uncertainty in measurement’ and to prepare documents for its broad application. Working Group 2, ‘Working Group on International vocabulary of basic and general terms in metrology’, has the task to revise and promote the use of the International vocabulary of basic and general terms in metrology (VIM).

In 2008 the JCGM made available a slightly revised version (mainly correcting minor errors) of the ‘Guide to the expression of uncertainty in measurement’, labelling the document ‘JCGM 100:2008’.

In 2017 the JCGM rebranded the documents in its portfolio that have been produced by Working Group 1 or are to be developed by that Group: the whole suite of documents is now known as the ‘Guide to the expression of uncertainty in measurement’ or ‘GUM’, and is concerned with the evaluation and expression of measurement uncertainty, as well as its application in science, trade, health, safety and other societal activities.

This part of the suite contains a collection of examples illustrating the methods for uncertainty evaluation described in the GUM.

This document has been prepared by Working Group 1 of the JCGM, and has benefited from detailed reviews undertaken by member organizations of the JCGM and National Metrology Institutes.

Introduction

The GUM suite of documents (parts overview in [14]) is concerned with the evaluation of uncertainty in measurement and how uncertainty is reported and used. Compact examples are given throughout the GUM suite that relate to specific areas of application and are intended to illustrate particular concepts. This document contains larger examples and case studies intended to illustrate the principles presented in the GUM suite for evaluating and reporting measurement uncertainty.

The examples are chosen to illustrate good practice in the evaluation of measurement uncertainty using the methods in the GUM, such as the JCGM 100:2008 uncertainty framework (GUF) [17, Definition 3.18] or the Monte Carlo method (MCM) [12, 17] or alternative methods not (yet) covered by the guidance in the suite. In some examples, methods for uncertainty evaluation are compared to highlight aspects of their suitability for the presented case.

The context and measurement in the examples are described to the extent relevant; simplifications can have been made in the interest of presenting a concise case. The data used in the examples is, where possible, taken from real measurements. These data are not necessarily the best, representing the state-of-the-art or in any other way to be interpreted as such. Measurement procedures and practices are presented in a concise format and not intended to enable reproducing the measurement. Where appropriate, references are provided to documents describing these procedures or practices.

All examples were edited prior to inclusion in this document.

Results are generally reported in the manner described in JCGM 101:2008 [17]. However, more than the recommended one or two significant digits are often given here to facilitate comparison of the results obtained from the various approaches.

Guide to the expression of uncertainty in measurement — Part 5: Examples

0 Scope

0.1 This document gives a number of examples, which are worked out in considerable detail to illustrate the principles presented in the Guide to the expression of uncertainty in measurement (GUM) for evaluating and reporting uncertainty in measurement. Together with the (shorter) examples included in the other parts of the GUM, they should enable users of the GUM to put these principles into practice in their own work.

0.2 Because the examples are for illustrative purposes, they have by necessity been simplified (but not over-simplified). Moreover, because they and the numerical data used in them have been chosen mainly to demonstrate the principles of the GUM, neither they nor the data should necessarily be interpreted as describing real measurements, let alone the state of the art for a particular measurement. Although the data were used as provided, all calculations retained the full computational precision of the systems employed. Final and intermediate results have been rounded appropriately to suit the intended applications.

0.3 The examples are used to demonstrate, among other:

- Using knowledge about a quantity to obtain an estimate of the quantity and its associated uncertainty;
- Propagation of uncertainty using the law of propagation of uncertainty (LPU) as in JCGM 100:2008 and JCGM 102:2011;
- Propagation of distributions using analytical and numerical methods, including the Monte Carlo method (MCM), as in JCGM 101:2008 and JCGM 102:2011;
- Applying law of propagation of uncertainty (LPU) and MCM for a univariate (scalar) and multivariate (vector) measurand;
- Applying LPU and MCM for independent and dependent input quantities;
- Using MCM to validate the results provided by LPU;
- Determining a coverage interval or a coverage region;
- Using measurement uncertainty in conformity assessment as in JCGM 106:2012;
- Reporting measurement results.

0.4 An overview of the parts of the GUM suite is given in [14]. The conventions and notation applied in this document are summarised in Annex A. The symbols used throughout the document are explained in the text and in Annex B.

1 Measurement of pH: linear interpolation

1.1 Preamble

1.1.1 This example is taken from a compendium of examples [1].

1.1.2 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 (LPU) 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 hydrogen ion activity (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.

1.1.3 The VIM concept of calibration [16, Definition 2.39] as constituting two stages is used. 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 y of the response given a value x of the stimulus. Uncertainties in both the stimulus and response variables are handled in both stages and propagated using the LPU in JCGM 100:2008 [10] and its generalization to vector measurands in JCGM 102:2011 [12].

1.1.4 The considerations of 1.1.3 are applied to pH measurement in which up to three two-point interpolations are required and uncertainties are tracked through the calculation.

1.2 pH of a test solution

1.2.1 pH, the negative logarithm to base 10 of the activity of hydrogen ion in a solution, is arguably the most measured quantity in chemistry [25]. The pH is a measure of the acidity of a solution. The pH of a solution is generally measured using a pH-sensitive glass electrode and a silver chloride (AgCl) reference electrode. The potential difference of such an electrode system is proportional to pH and forms the basis of pH measurement [28].

1.2.2 In 2002, the International Union of Pure and Applied Chemistry (IUPAC) issued a recommendation for revision of the pH scale based on the concept of a primary reference measurement procedure for pH [21]. 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 (SI), here the SI measurement unit 1.

1.2.3 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. Standard solutions are sold as certified reference materials (CRMs) to calibrate pH meters for routine use.

1.2.4 There are several approaches to pH measurement including the use of 1-point, 2-point and multi-point calibration, least-squares regression, and with or without temperature correction. Here the 2-point calibration approach, with and without temperature correction, is used. Two topics are considered: (a) when the temperature of the test solution matches that of the standard (reference) solutions and (b) when this is not the case.

1.3 Specification of the measurand

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.

1.4 Measurement model

There are two stages involved in calibration [16, definition 2.39]: (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 [15, Clause 8.4] and is necessary when the construction and use of the calibration model are carried out by different parties.

1.4.1 Generic approach to two-point calibration

1.4.1.1 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 1.1).

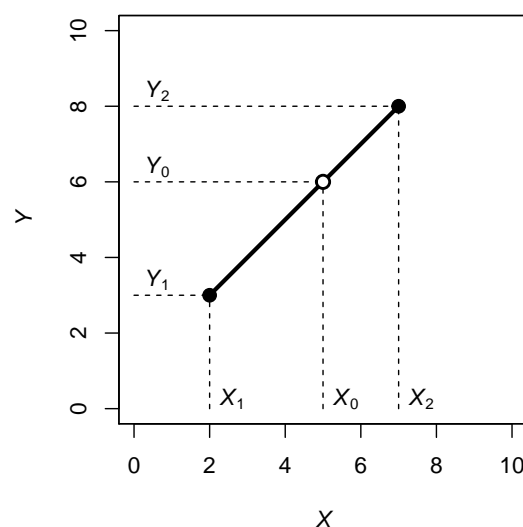


Figure 1.1: Two-point calibration using a linear interpolation function

1.4.1.2 By similar triangles, with $\delta X = X_2 - X_1$ and $\delta Y = Y_2 - Y_1$,

$$\frac{Y - Y_1}{X - X_1} = \frac{Y_2 - Y_1}{X_2 - X_1} = \frac{\delta Y}{\delta X}. \quad (1.1)$$

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

$$Y = a + bX, \quad (1.2)$$

where a is the intercept on the Y -axis and b is the slope [74].

NOTE The form (1.2) is used in the straight-line calibration standard ISO/TS 28037 [74] and will be familiar to many end users.

1.4.1.3 In the *single-stage model*, a single party has estimates of (X_1, Y_1) , (X_2, Y_2) and X_0 , and obtains an estimate of Y_0 , the Y -value on the line corresponding to X_0 (figure 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 .

1.4.1.4 In the *two-stage model*, however, one party has estimates of (X_1, Y_1) and (X_2, Y_2) , and provides estimates of a and b (intermediate measurands) to the second party using the straightforwardly verified

$$b = \frac{Y_2 - Y_1}{X_2 - X_1}, \quad (1.3)$$

$$a = Y_1 - bX_1. \quad (1.4)$$

The second party (possibly identical to the first party) has estimates of a , b and X_0 , and obtains the estimate of Y_0 using the expression

$$Y_0 = a + bX_0. \quad (1.5)$$

The measurement model is again described by the process to provide Y_0 . The resulting expression

$$Y_0 = Y_1 + b(X_0 - X_1), \quad (1.6)$$

and formula (1.3) constitute the measurement model with Y_0 as the measurand.

NOTE When X_1 and X_2 are far from the origin, that is, $|X_2 - X_1| \ll \max(|X_1|, |X_2|)$, an alternative form may be numerically more stable. One such form is given by working with a transformed X -variable

$$\tilde{X} = X - X_1.$$

Using expressions (1.2) and (1.4), the calibration function can then be expressed as

$$Y = Y_1 + b\tilde{X},$$

which is evaluated at the value X_0 of the independent variable.

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 [74, 75]) but is not required here. The roles of X and Y can be interchanged when permitted by the context.

1.4.2 Generic approach to multi-point calibration

1.4.2.1 Multi-point calibration is the treatment in clauses 1.4.1 and 1.4.1.3 extended to an arbitrary number of points. In these clauses, 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 [34], acting as the calibration function. For each interval between pairs of successive points, the treatment of clauses 1.4.1 and 1.4.1.3 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 and is naturally monotonic, a necessary condition. For $m > 2$, the ordered points may not form a monotonic sequence, a situation not considered here [75].

1.4.2.2 Alternatively, straight-line fitting by least squares can be used taking reported uncertainties associated with the calibration data into consideration [74]. Polynomial interpolation or polynomial regression can also be used [75]. See 1.10.1.

1.4.3 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 [10, 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 [10, clause 4.3]. Such covariances should be handled also to avoid producing invalid statements of uncertainty associated with predicted Y -values. The calibration data considered here are assumed to have independent errors.

1.5 pH measured at a stipulated temperature

1.5.1 When the cell potential is measured at one of the stipulated temperatures on a reference material certificate for standard solutions, the following process is applicable. A correction approach is used to provide the pH of a test solution [134] in which 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 :

$$\text{pH}_X = \text{pH}_S + \frac{E_X - E_S}{k}. \quad (1.7)$$

In expression (1.7), pH_S is the pH of S , and

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

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

1.5.2 Bracketing methods are used here since they are generally more accurate. Use is made of reference material certificates for the standard solutions, which give pH values and associated standard uncertainties at stipulated temperatures. It is assumed the specific temperature lies within the interval (here, 5 °C to 50 °C) covered by the certificates for

the Standard reference materials (SRMs) used here, namely, potassium dihydrogen phosphate (NIST SRM 186-I-g) and disodium dihydrogen phosphate (NIST SRM 186-II-g) [95]. Those certificates, issued by the National Institute of Standards and Technology (NIST), contain tables of pH values and associated uncertainties for temperatures from 5 °C to 50 °C at a spacing of 5 °C. (Temperatures are given here in °C as is common practice in pH measurement.) These temperatures are reference values with no assigned uncertainty.

1.5.3 The potential E_X of the test solution X is measured at temperature T_X , one of the stipulated temperatures in clause 1.5.2. 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 values of E_{S_1} and E_{S_2} bracket E_X and are as near as possible to it [31]. The pH of S_1 and S_2 at temperature T_X , namely, pH_{S_1, T_X} and pH_{S_2, T_X} , are taken from reference material certificates.

1.5.4 Assuming linearity between the points $(E_{S_1}, \text{pH}_{S_1, T_X})$ and $(E_{S_2}, \text{pH}_{S_2, T_X})$, linear interpolation is applied to estimate the pH value pH_X corresponding to the potential E_X . 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} , pH_{S_1, T_X} , pH_{S_2, T_X} and E_X , corresponding respectively to X_1 , X_2 , Y_1 , Y_2 and X_0 in the generic approach.

1.5.5 In an extended measurement model [15, clause 9], account is taken of further influences. In this case, main effects are pH instrument calibration, instrument resolution and interpolation model. 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,

$$\begin{aligned} E_{X, \text{corr}} &= E_X + \delta E_{\text{res}, X} + \delta E_{\text{cal}}, \\ E_{S_1, \text{corr}} &= E_{S_1} + \delta E_{\text{res}, S_1} + \delta E_{\text{cal}}, \\ E_{S_2, \text{corr}} &= E_{S_2} + \delta E_{\text{res}, S_2} + \delta E_{\text{cal}}. \end{aligned} \quad (1.8)$$

Interpolated pH and correction quantities in expressions (1.8) relating to cell potential are assumed independent as in common practice [21]. Clause 1.10.1 contains a discussion of possible correlation in these quantities.

1.6 pH measurement accounting for temperature

When the cell potential is not measured at one of the temperatures specified on a reference material certificate for standard solutions, the procedure in clause 1.5 can be extended as follows. A bracketing approach as in clause 1.5 is used in which three linear interpolations are carried out to estimate pH for solution X and can be described by the following procedure, which constitutes an algorithmic description of the measurand model.

Measure the temperature T_X of the test solution X.

Determine the closest bracketing temperatures T_1 and T_2 in the SRM certificates such that $T_X \in [T_1, T_2]$.

Measure the potential E_X in solution X and potentials E_{S_1} and E_{S_2} in standard solutions S_1 and S_2 .

Obtain the certified pH value pH_{S_1, T_1} at temperature T_1 for standard solution S_1 from the certificate for S_1 . Likewise, obtain the certified pH value pH_{S_2, T_1} at temperature T_1 for standard solution S_2 from the certificate for S_2 .

Repeat this process for temperature T_2 , that is, obtain the pH value pH_{S_1, T_2} for solution S_1 , and the pH value pH_{S_2, T_2} for solution S_2 .

Apply three steps of linear interpolation:

1. Interpolate between the points $(E_{S_1}, \text{pH}_{S_1, T_1})$ and $(E_{S_2}, \text{pH}_{S_2, T_1})$ to give the pH value pH_{X, T_1} at temperature T_1 corresponding to potential E_X .
2. Interpolate between the points $(E_{S_1}, \text{pH}_{S_1, T_2})$ and $(E_{S_2}, \text{pH}_{S_2, T_2})$ to give the pH value pH_{X, T_2} at temperature T_2 corresponding to potential E_X .
3. Interpolate between the points $(T_1, \text{pH}_{X, T_1})$ and $(T_2, \text{pH}_{X, T_2})$ so obtained to give the pH value pH_{X, T_X} at temperature T_X corresponding to potential E_X .

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

The measurement model is specified by the above algorithmic description where the input quantities in the model are $T_X, E_X, E_{S_1}, E_{S_2}, \text{pH}_{S_1, T_1}, \text{pH}_{S_2, T_1}, \text{pH}_{S_1, T_2}$ and pH_{S_2, T_2} .

1.7 Uncertainty propagation

1.7.1 Uncertainty propagation in this clause is based on the assumption that the input quantities $T_X, E_X, E_{S_1}, E_{S_2}, \text{pH}_{S_1, T_1}, \text{pH}_{S_2, T_1}, \text{pH}_{S_1, T_2}$ and pH_{S_2, T_2} , that is, the measured potentials and the pH values for the standard solutions, to the procedure in clause 1.6 are independent. This assumption is consistent with IUPAC recommendations for pH measurement [21]. There, procedures are given for accounting for input standard uncertainties based on the correlation-free variant of the law of propagation of uncertainty in [10, clause 5.1]. In practice, input quantities are likely to be correlated and account should be taken of that fact. See the discussion in clause 1.10.3.

1.7.2 Consider first the *two-stage model*. For the 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 \hat{a} and \hat{b} and their associated covariance matrix $U_{[\hat{a}, \hat{b}]}$. The model is bivariate (two output quantities):

$$b = \frac{Y_2 - Y_1}{X_2 - X_1} = \frac{\delta Y}{\delta X}, \quad a = Y_1 - bX_1. \quad (1.9)$$

For the uncertainty propagation, [12, clause 6.2.1.3, formula (3)] is applied to obtain the output covariance matrix for the model parameters a and b :

$$U_{[\hat{a}, \hat{b}]} = \begin{bmatrix} u^2(\hat{a}) & u(\hat{a}, \hat{b}) \\ u(\hat{a}, \hat{b}) & u^2(\hat{b}) \end{bmatrix} = \mathbf{C}_1 \mathbf{U}_{\text{in}} \mathbf{C}_1^\top. \quad (1.10)$$

In formula (1.10), $u(\hat{a}, \hat{b})$ denotes the covariance associated with \hat{a} and \hat{b} , \mathbf{U}_{in} denotes the covariance matrix of the input parameter values, namely, 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

$$\mathbf{C}_1 = \frac{1}{\delta x} \begin{bmatrix} -\hat{b}x_2 & x_2 & bx_1 & -x_1 \\ \hat{b} & -1 & -\hat{b} & 1 \end{bmatrix} \quad (1.11)$$

is the sensitivity matrix containing the first-order partial derivatives of a and b with respect to X_1, Y_1, X_2 and Y_2 , evaluated at the estimates x_1, y_1, x_2, y_2 and \hat{b} .

1.7.3 In the second stage, the inputs are the outputs from the first stage together with x_0 and $u(x_0)$. The measurement model is

$$Y_0 = a + bX_0. \quad (1.12)$$

Applying uncertainty propagation [12, Formula (3)] to the model (1.12), considering there are three input quantities, a , b and X_0 in that model, the output standard uncertainty $u(y_0)$ is given by

$$u^2(y_0) = \mathbf{C}_2 \mathbf{U}_{[\hat{a}, \hat{b}]} \mathbf{C}_2^\top + \hat{b}^2 u^2(x_0), \quad (1.13)$$

where \mathbf{C}_2 is the 1×2 sensitivity matrix

$$\mathbf{C}_2 = \begin{bmatrix} 1 & x_0 \end{bmatrix}. \quad (1.14)$$

1.7.4 The *single-stage model* can be obtained by combining the two stages above. The substitution of formula (1.10) into expression (1.13) yields

$$u^2(y_0) = \mathbf{C}_2 \mathbf{C}_1 \mathbf{U}_{\text{in}} \mathbf{C}_1^\top \mathbf{C}_2^\top + \hat{b}^2 u^2(x_0). \quad (1.15)$$

The use of formulae (1.11) and (1.14), together with expressions

$$q = \frac{x_0 - x_1}{\delta x} \quad (1.16)$$

and

$$\mathbf{U}_{\text{in}} = \text{diag}[u^2(x_1) \quad u^2(y_1) \quad u^2(x_2) \quad u^2(y_2)],$$

a diagonal matrix with the given diagonal entries, yields, after substituting into formula (1.15) and simplifying,

$$u^2(y_0) = \hat{b}^2 (1-q)^2 u^2(x_1) + (1-q)^2 u^2(y_1) + \hat{b}^2 q^2 u^2(x_2) + q^2 u^2(y_2) + \hat{b}^2 u^2(x_0). \quad (1.17)$$

The result (1.17) can also be confirmed from first principles.

1.8 pH estimation at a stipulated temperature and associated uncertainty evaluation

1.8.1 The process in clauses 1.5.3 and 1.5.4 is followed. Consider the estimation of pH at a specific temperature. Values of potential in the test and standard solutions S_1 and S_2 are

$$\hat{E}_X = -1.875 \text{ mV}, \quad \hat{E}_{S_1} = 6.15 \text{ mV}, \quad \hat{E}_{S_2} = -26.35 \text{ mV},$$

each of which was the average of 4 repeated observations. pH values at 25 °C for S_1 and S_2 from [95] are

$$\widehat{\text{pH}}_{S_1, 25^\circ\text{C}} = 6.8640, \quad \widehat{\text{pH}}_{S_2, 25^\circ\text{C}} = 7.4157.$$

Formulae (1.3) and (1.6) yield $\widehat{\text{pH}}_X = 7.0002$. The standard uncertainties associated with the input quantities are given in [95] and are as follows:

$$\begin{aligned} u(\hat{E}_X) &= 0.0250 \text{ mV}, & u(\hat{E}_{S_1}) &= 0.0289 \text{ mV}, & u(\hat{E}_{S_2}) &= 0.0289 \text{ mV}, \\ u(\widehat{\text{pH}}_{S_1, 25^\circ\text{C}}) &= 0.0051, & u(\widehat{\text{pH}}_{S_2, 25^\circ\text{C}}) &= 0.0051. \end{aligned}$$

The propagation of uncertainty carried out in accordance with expressions (1.16) and (1.17) yields $u(\widehat{\text{pH}}_X) = 0.0041$.

1.8.2 These results relate to the basic measurement model for pH. The extended model would work with the corrected quantities in (1.8) 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 (1.8), 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 $E_{X,\text{corr}}$, $E_{S_1,\text{corr}}$ and $E_{S_2,\text{corr}}$ as the ‘new’ X_0 , X_1 and X_2 , respectively, from expressions (1.8) into expressions (1.3) and (1.6).

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

1.9 pH estimation accounting for temperature and associated uncertainty evaluation

1.9.1 The process in clause 1.6 is followed. The measured temperature of the test solution is $\hat{T}_X = 23.7^\circ\text{C}$. The bracketing pair $T_1 = 25^\circ\text{C}$ and $T_2 = 20^\circ\text{C}$ is selected.

Potential measurement gives

$$\hat{E}_X = -1.875 \text{ mV}, \quad \hat{E}_{S_1} = 6.15 \text{ mV}, \quad \hat{E}_{S_2} = -26.35 \text{ mV},$$

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 [95] are

$$\begin{aligned} \widehat{\text{pH}}_{S_1,T_1} &= 6.8640, & \widehat{\text{pH}}_{S_2,T_1} &= 7.4157, \\ \widehat{\text{pH}}_{S_1,T_2} &= 6.8796, & \widehat{\text{pH}}_{S_2,T_2} &= 7.4323. \end{aligned}$$

Linear interpolation between the points

$$(\hat{E}_{S_1}, \widehat{\text{pH}}_{S_1,T_1}) \equiv (6.15 \text{ mV}, 6.8640) \quad \text{and} \quad (\hat{E}_{S_2}, \widehat{\text{pH}}_{S_2,T_1}) \equiv (-26.35 \text{ mV}, 7.4157)$$

gives the pH value $\widehat{\text{pH}}_{X,T_1} = 7.0002$ at temperature $\hat{T}_1 = 25^\circ\text{C}$ corresponding to potential $\hat{E}_X = -1.875 \text{ mV}$.

Likewise, linear interpolation between

$$(\hat{E}_{S_1}, \widehat{\text{pH}}_{S_1,T_2}) \equiv (6.15 \text{ mV}, 6.8796) \quad \text{and} \quad (\hat{E}_{S_2}, \widehat{\text{pH}}_{S_2,T_2}) \equiv (-26.35 \text{ mV}, 7.4323)$$

gives the pH value $\widehat{\text{pH}}_{X,T_2} = 7.0161$ at temperature $\hat{T}_2 = 20^\circ\text{C}$ corresponding to potential $\hat{E}_X = -1.875 \text{ mV}$.

Linear interpolation between the points $(T_1, \widehat{\text{pH}}_{X,T_1})$ and $(T_2, \widehat{\text{pH}}_{X,T_2})$ gives the pH value $\widehat{\text{pH}}_X = 7.0043$ at temperature $\hat{T}_X = 23.7^\circ\text{C}$ corresponding to potential $\hat{E}_X = -1.875 \text{ mV}$.

Associated standard uncertainties are

$$\begin{aligned} u(\hat{T}_X) &= 0.0058^\circ\text{C}, & u(\hat{E}_X) &= 0.025 \text{ mV}, \\ u(\hat{E}_{S_1}) &= 0.029 \text{ mV}, & u(\hat{E}_{S_2}) &= 0.029 \text{ mV}, \\ u(\widehat{\text{pH}}_{S_1,T_1}) &= 0.0051, & u(\widehat{\text{pH}}_{S_2,T_1}) &= 0.0051, \\ u(\widehat{\text{pH}}_{S_1,T_2}) &= 0.0051, & u(\widehat{\text{pH}}_{S_2,T_2}) &= 0.0051. \end{aligned}$$

The above standard uncertainties associated with standard pH solutions are given in [95].

The application of the method of clauses 1.6 and 1.7 gives

$$\widehat{\text{pH}}_X = 7.0109, \quad u(\widehat{\text{pH}}_X) = 0.0041.$$

1.10 Interpretation of results

1.10.1 Validation of results

1.10.1.1 A check of the numerical accuracy of linear interpolation was made for the example in clause 1.6. Since NIST certificate [95] 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 on [75] was carried out using temperature values 15 °C, 20 °C, 25 °C and 30 °C (two values immediately 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, 0.0004, between these values is 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 considered adequate in this example.

1.10.1.2 In a study by Damasceno et al. [45], it was reported that a Monte Carlo method (MCM) applied to primary pH measurement gave similar results to the LPU [10]. The work of Wiora and Wiora [133] came to the same conclusion. To test these conclusions for the data used here, the MCM of JCGM 101:2008 [17] was applied to the example in clause 1.9. 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 clause to the number of decimal places stated.

1.10.2 Resolution

1.10.2.1 The instrument display gave results in volts with 3 significant decimal places. Assume a resolution error in the last digit, that is, in the interval $[-0.0005, 0.0005]$ V. Characterizing resolution by a rectangular distribution over this interval, the consequent resolution standard uncertainty E_{res} applying to all potential readings is $0.0005 \text{ V} / \sqrt{3} = 0.00029 \text{ V}$. This standard uncertainty is some one hundredth of the above potential standard uncertainties and so is negligible.

1.10.2.2 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 clause 1.9, $u(\widehat{\text{pH}}_x)$ becomes 0.0040 (originally 0.0041). Further repeated observations of the three potentials would do much to reduce $u(\widehat{\text{pH}}_x)$, assuming the repeated observations are genuinely independent.

1.10.3 Correlation

1.10.3.1 It is 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 [52, 59, 84, 127], for instance), but has adverse consequences in that evaluated pH uncertainties can be optimistically small. To obtain more-valid results, covariance effects should be quantified and incorporated.

1.10.3.2 Laboratories that follow IUPAC recommendations [21] will not take correlation into consideration and so might be reporting unreasonably small measurement uncertain-

ties. 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 JCGM 100:2008 [10, clause 5.2]. It is noted that correlation issues are discussed in [132], but they relate to correlations induced by the choice of parametrization rather than being associated with input quantities.

1.10.3.3 In terms of the 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 clauses 1.8 and 1.9) for an SI-traceable value and one, much smaller, that omits the uncertainty associated with the Bates-Guggenheim convention [8]. 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) is suggested as a generally conservative treatment. See clause 1.10.3.6 for a relevant discussion.

1.10.3.4 To indicate the effect of ignoring correlation, an exercise is carried out in which full correlation is present between the pH values for the standards. 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 [29, 37, 53], 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 clause 1.7.2, the input covariance matrix U_{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):

$$U_{\text{in}} = \begin{bmatrix} u^2(x_1) & & & \\ & u^2(y_1) & & u^2(y_1) \\ & & u^2(x_2) & \\ & u^2(y_1) & & u^2(y_1) \end{bmatrix}. \quad (1.18)$$

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 clause 1.7.4 but using the covariance matrix (1.18) gives

$$u^2(y_0) = \hat{b}^2(1-q)^2u^2(x_1) + (1-q)^2u^2(y_1) + 2q(1-q)u(y_1, y_2) + \hat{b}^2q^2u^2(x_2) + q^2u^2(y_2) + \hat{b}^2u^2(x_0). \quad (1.19)$$

The only difference is that the terms

$$(1-q)^2u^2(y_1) + q^2u^2(y_2) = [(1-q)^2 + q^2]u^2(y_1) \quad (1.20)$$

in the uncorrelated treatment [expression (1.17)] are replaced by

$$\begin{aligned} & (1-q)^2u^2(y_1) + 2q(1-q)u(y_1, y_2) + q^2u^2(y_2) \\ & = [(1-q)^2 + 2q(1-q) + q^2]u^2(y_1) \end{aligned} \quad (1.21)$$

in the fully correlated case [expression (1.19)]. Since expression (1.21) simplifies (exactly) to $u^2(y_1)$, expression (1.19) becomes

$$u^2(y_0) = \hat{b}^2(1-q)^2u^2(x_1) + u^2(y_1) + \hat{b}^2q^2u^2(x_2) + \hat{b}^2u^2(x_0) \quad (1.22)$$

in the fully correlated case. By applying a similar treatment to that in subclause 1.7 on the general single-stage model but using the covariance matrix (1.18) gives

$$u^2(y_0) = \hat{b}^2(1-q)^2u^2(x_1) + (1-q)^2u^2(y_1) + 2q(1-q)u(y_1, y_2) + \hat{b}^2q^2u^2(x_2) + q^2u^2(y_2) + \hat{b}^2u^2(x_0). \quad (1.23)$$

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 , expression (1.23) becomes

$$u^2(y_0) = \hat{b}^2(1-q)^2u^2(x_1) + u^2(y_1) + \hat{b}^2q^2u^2(x_2) + \hat{b}^2u^2(x_0). \quad (1.24)$$

1.10.3.5 In the application of expression (1.24) to the data in subclause 1.7, the standard uncertainty associated with $\widehat{\text{pH}}_X = 7.0109$ becomes $u(\widehat{\text{pH}}_X) = 0.0051$ (compared with 0.0041 when correlation between the pH standards is disregarded). Unsurprisingly, this value of $u(\widehat{\text{pH}}_X)$ is the same as the (identical) values of the ‘input’ standard uncertainties $u(\widehat{\text{pH}}_{S_1, 25^\circ\text{C}})$ and $u(\widehat{\text{pH}}_{S_2, 25^\circ\text{C}})$. Thus there is no reduction in uncertainty in the fully correlated case.

It is observed that the standard uncertainties $u(\widehat{\text{pH}}_{S_1, 25^\circ\text{C}})$ and $u(\widehat{\text{pH}}_{S_2, 25^\circ\text{C}})$ 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 (1.17)] in the case where correlation is ignored are

$$0.0004, \quad 0.0001, \quad 0.0038, \quad 0.0013, \quad 0.0004$$

to four decimal places. The values are to be compared with the corresponding four values from expression (1.24), namely,

$$0.0004, \quad 0.0001, \quad 0.0051, \quad 0.0004,$$

in the correlated case, demonstrating that the standard uncertainties of perfectly correlated input estimates are combined additively (see [10, clause 5.2.2, note 1]):

$$0.0038 + 0.0013 = 0.0051.$$

The situation is compounded in clause 1.9 where four (rather than two) pH values and two temperature values are involved.

1.10.3.6 Independence of the pH standards is not generally a valid assumption. A treatment such as given in [37, section 4.1] is suggested, that is, to work with a common correlation coefficient r associated with the input pH values. The basic change would be that the off-diagonal terms of U_{in} in formula (1.18) would become $ru^2(y_1)$. (The case $r = 0$ yields the uncorrelated case and $r = 1$ the case of perfect correlation.) A value for r 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.

1.10.3.7 A further, chemical, issue is that the NIST standard reference materials are not solutions; they are solids that must be weighed, mixed and fully dissolved in high-purity water. Buffer solutions are not particularly sensitive to minor dilution problems, but the preparation just mentioned will add further variation. Atmospheric carbon dioxide, CO_2 , for example, can shift the pH values, especially in neutral and alkaline pH test samples ($\text{pH} \geq 7$). Although a laboratory would exercise care in measuring secondary solutions, the analysis given here omits these important handling effects.

2 Determination of benzo[a]pyrene

2.1 Preamble

2.1.1 This example is taken from a compendium of examples [1].

2.1.2 The aim of this example is to show the uncertainty evaluation for the measurement of small masses of benzo[a]pyrene (BaP), an important polycyclic aromatic hydrocarbon (PAH) for ambient air monitoring. A comparison between the results obtained according to the JCGM 100:2008 uncertainty framework (GUF) [10] and the Monte Carlo method (MCM) for the propagation of distributions [12, 17] is made and discussed.

2.1.3 Determination of small masses of PAHs is an important issue since they are ubiquitous toxic contaminants that can be present in all the environmental compartments even at trace levels. Among the various PAHs, BaP is classified as a carcinogenic agent and is listed in the current European legislation [51] as a marker of the carcinogenic risk for the whole class of PAHs in ambient air.

2.1.4 This example aims at comparing the results obtained by application of the GUF [10] and the MCM for propagation of distributions [17] to real data sets derived from the quantification of a small mass of BaP spiked on filters commonly used for airborne particulate matter sampling.

2.1.5 The description of the analytical method to quantify BaP mass in ambient air can be found in [111], whereas details on the uncertainty evaluation, not explicitly reported in the present example, can be found in [115].

2.2 Specification of the measurand

2.2.1 A glass fibre 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 underwent Soxhlet extraction following the procedure described in [111]. The same filter was subsequently extracted a second time, thus obtaining a diluted sample. The mass of BaP present in the two extracts was determined using a gas chromatograph coupled with a mass spectrometer (GC-MS).

NOTE Spiking is an analytical procedure by which a known mass (or amount of substance) of one or more compounds is added to serve as a reference point (see [62, entry 3.72]).

2.2.2 The measurand in the present example is the mass m_E of BaP contained in a nominal volume of 1 μL of the second extract.

NOTE The uncertainty evaluation developed in the present example is relevant to the determination of BaP mass on a spiked, clean glass fibre filter. It may grossly underevaluate the uncertainty on real samples due to different extraction efficiencies or interferences. For the sake of simplicity, this example intentionally neglects important error sources in chemistry, such as matrix effects and recovery. The measurand, here, is the mass in an extract, rather than in the original sample. This allows neglecting the extraction efficiency in the uncertainty budget.

2.3 Measurement model

2.3.1 The 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 [49]. An aliquot of the NIST SRM 2270, containing deuterated benzo[a]pyrene (BaP- d_{12}), was added to the solution in order to obtain a concentration of BaP- d_{12} equal to $0.2455 \mu\text{g mL}^{-1}$, to be used as the internal standard. Then, 1 μL aliquots of the solution were repeatedly (three times) injected in the GC-MS. The quotient of the peak area corresponding to the analyte and that corresponding to the internal standard was used to determine the mass m_E of BaP present in the injected volume of the extracted sample, according to the model

$$m_E = f \frac{A_E}{A_{\text{ISE}}} m_{\text{ISE}}, \quad (2.1)$$

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

NOTE The considered measurement approach assumes that isotopic equilibrium has been reached in spiked samples. Further, it relies on the implicit assumption that substances of different isotopic composition display perfectly the same characteristics.

2.3.2 The calibration factor f was obtained as the arithmetic mean of three calibration factors corresponding to three reference solutions at different BaP concentrations. Details of the calibration procedure were reported in [115]. In the evaluation of the uncertainty associated with f (characterized by 9 degrees of freedom), covariance terms between the three factors were taken into account: they were due to the same mass of the internal standard used in the calibration model for each of the factors and to the same CRM used for preparing the three necessary reference solutions. For the same reason, f and m_{ISE} , as input quantities of measurement model (2.1), were correlated because of the use of the same internal standard both in the calibration and in the analysis.

2.3.3 The mass m_{ISE} and the associated uncertainty of the internal standard were obtained from its calibration certificate. The degrees of freedom was considered to be very large, so that it did not contribute to the effective degrees of freedom of the output uncertainty.

2.3.4 Mean chromatographic peak areas A_E and A_{ISE} of BaP and BaP- d_{12} were evaluated as the arithmetic means of the three repetitions of the area measurement of the relevant chromatographic peaks. Their standard uncertainties were calculated as the standard deviations of those means [10, clause 4.2.3] (hence, having two degrees of freedom). A strong linear relationship was observed between the areas of the BaP and those of the ISE in the same run; hence a corresponding covariance term for the two mean areas was calculated according to [10, clause 5.2.3]. The estimates of the means, standard uncertainties, and covariances associated with the aforementioned input quantities are reported in table 2.1.

Table 2.1: Estimate, standard uncertainty, degrees of freedom and covariance of the input quantities in model (2.1) (a.u.: arbitrary units)

Quantity	Estimate	Std. uncertainty	Deg. of freedom	Covariance
f	0.616	0.017	9	
m_{ISE}	0.2455 ng	0.0036 ng	∞	-3.3×10^{-5} ng
A_{E}	85 114 a.u.	9564 a.u.	2	
A_{ISE}	917 546 a.u.	44 492 a.u.	2	-2.03×10^8 (a.u.) ²

2.4 Uncertainty propagation

2.4.1 GUF

2.4.1.1 For calculating the uncertainty associated with the estimate $\widehat{m}_{\text{E}} = 0.014$ ng of the measurand, both the GUF [10] and the MCM for the propagation of probability distributions [12, 17] were applied and compared.

2.4.1.2 Applying the law of propagation of uncertainty (LPU) to model (2.1), taking into account the uncertainty and covariance contributions of the input quantities reported in table (2.1), the resulting standard uncertainty $u(\widehat{m}_{\text{E}})$ was 0.002 ng. Table 2.2 reports the uncertainty budget for the mass $\widehat{m}_{\text{E}} = 0.014$ ng of BaP of the sample obtained with the second extraction (the first two columns repeat part of the information already given in table 2.1). Specifically, in addition to the uncertainty contributions, the table gives the contributions arising from the covariances. Accounting for covariance contributions gives $u^2(\widehat{m}_{\text{E}}) = 4.1 \times 10^{-6}$ ng² whereas disregarding them gives 3.1×10^{-6} ng², some 25 % smaller.

Table 2.2: Uncertainty budget for the mass m_{E}

Quantity (x_i)	$u(\widehat{x}_i)$	$c_i = \partial \widehat{m}_{\text{E}} / \partial x_i$	$c_i^2 u^2(\widehat{x}_i) / \text{ng}^2$
f	1.7×10^{-2}	2.28×10^{-2} ng	1.50×10^{-7}
m_{ISE}	3.6×10^{-3} ng	5.71×10^{-2}	4.23×10^{-8}
A_{E}	9.6×10^3 a.u.	1.65×10^{-7} ng(a.u.) ⁻¹	2.50×10^{-6}
A_{ISE}	4.4×10^4 a.u.	-1.53×10^{-8} ng(a.u.) ⁻¹	4.53×10^{-7}
$u(\widehat{x}_i, \widehat{x}_j)$		$c_i c_j$	$2c_i c_j u(\widehat{x}_i, \widehat{x}_j) / \text{ng}^2$
f, m_{ISE}	-3.3×10^{-5} ng	1.30×10^{-3} ng	-8.59×10^{-8}
$A_{\text{E}}, A_{\text{ISE}}$	-2.03×10^8 (a.u.) ²	-2.52×10^{-15} ng ² (a.u.) ⁻²	1.02×10^{-6}
$u^2(\widehat{m}_{\text{E}})$			4.1×10^{-6}

2.4.1.3 The effective degrees of freedom ν_{eff} for $u(\widehat{m}_{\text{E}})$ was calculated using the Welch-Satterthwaite formula [10, eqn. (G2.b)], which was applied to the input uncertainties and their corresponding degrees of freedom:

$$\nu_{\text{eff}} = \frac{u^4(\widehat{m}_{\text{E}})}{\sum_i c_i^4 u^4(\widehat{x}_i) / \nu_i} \approx 5.2. \quad (2.2)$$

This evaluation, however, neglects accounting for the covariances between the input quantities. A generalization of the Welch–Satterthwaite formula for use with correlated quantities is available [131]:

$$\nu_{\text{eff},r} = \frac{u^4(\widehat{m}_E)}{\sum_i \sum_j r_{ij}^2 c_i^2 c_j^2 u^2(\widehat{x}_i) u^2(\widehat{x}_j) / \nu_i}. \quad (2.3)$$

The two non-trivial correlation coefficients between the estimates of the input quantities are

$$r(\widehat{f}, \widehat{m}_{\text{ISE}}) = \frac{u(\widehat{f}, \widehat{m}_{\text{ISE}})}{u(\widehat{f})u(\widehat{m}_{\text{ISE}})} = \frac{-3.3 \times 10^{-5}}{1.7 \times 10^{-2} \times 3.6 \times 10^{-3}} = -0.54,$$

$$r(\widehat{A}_E, \widehat{A}_{\text{ISE}}) = \frac{u(\widehat{A}_E, \widehat{A}_{\text{ISE}})}{u(\widehat{A}_E)u(\widehat{A}_{\text{ISE}})} = \frac{-2.03 \times 10^8}{9.6 \times 10^3 \times 4.4 \times 10^4} = -0.48,$$

whereas other correlation coefficients are either 0, as for $r(\widehat{f}, \widehat{A}_{\text{ISE}})$, or 1, as for $r(\widehat{f}, \widehat{f})$. The calculation of the effective degrees of freedom using equation (2.3) yields $\nu_{\text{eff},r} \approx 4.8$.

2.4.1.4 Coverage factors of a Student t -distribution with integer degrees of freedom are given in [10, table G.2]. Coverage factors for non-integer values can be recovered by means of common statistical software. $\nu_{\text{eff}} = 5.2$ (ignoring the correlations) and $\nu_{\text{eff},r} = 4.8$ (including the correlations) correspond to 95 % coverage factors $k = 2.5$ and $k = 2.6$, respectively. Using the latter coverage factor, the expanded uncertainty $U = k u(\widehat{m}_E)$ at a 95 % coverage probability is 0.0052 ng.

2.4.2 MCM

2.4.2.1 The MCM for propagation of probability distributions of the input quantities was applied in order to approximate the distribution for the measurand m_E , the mass of BaP in the extract. For this purpose, suitable probability distributions were assigned to the input quantities in model (2.1), according to the criteria prescribed in [12, 17].

2.4.2.2 Since the available information about f and m_{ISE} were estimates and their associated covariance matrix, a bivariate normal distribution was assigned to these quantities [12, Clause 6.5.2.2]. Hence, the mean of this bivariate normal distribution is a vector $[\widehat{f}, \widehat{m}_{\text{ISE}}]$ and its covariance matrix is

$$U = \begin{bmatrix} u^2(\widehat{f}) & u(\widehat{f}, \widehat{m}_{\text{ISE}}) \\ u(\widehat{f}, \widehat{m}_{\text{ISE}}) & u^2(\widehat{m}_{\text{ISE}}) \end{bmatrix},$$

whose components are available in table 2.1. Note that such a treatment is a simplification as it ignores the fact that $u(\widehat{f})$ is supported by a small degrees of freedom. One way to jointly model f using the t -distribution with 9 degrees of freedom and m_{ISE} using the normal distribution is with the use of copulas [103].

2.4.2.3 The $N = 2$ quantities A_E and A_{ISE} were estimated using $n = 3$ repeated indications (\widehat{A}_{E_i} and $\widehat{A}_{\text{ISE}_i}$, respectively). Therefore, according to [12, Clause 5.3.2], a scaled and shifted bivariate t -distribution with one degree of freedom ($\nu = n - N$) was assigned to

them. The expectation of the distribution is the vector $[\widehat{A}_E, \widehat{A}_{ISE}]^T$ and its scale matrix S/n is

$$S = \frac{1}{\nu} \begin{bmatrix} \sum_{i=1}^3 (\widehat{A}_{E_i} - \widehat{A}_E)^2 & \sum_{i=1}^3 (\widehat{A}_{E_i} - \widehat{A}_E)(\widehat{A}_{ISE_i} - \widehat{A}_{ISE}) \\ \sum_{i=1}^3 (\widehat{A}_{E_i} - \widehat{A}_E)(\widehat{A}_{ISE_i} - \widehat{A}_{ISE}) & \sum_{i=1}^3 (\widehat{A}_{ISE_i} - \widehat{A}_{ISE})^2 \end{bmatrix},$$

whose components are $S_{1,1} = 548\,861\,202$ (a.u.)², $S_{1,2} = S_{2,1} = -1\,220\,621\,757$ (a.u.)², and $S_{2,2} = 11\,877\,338\,582$ (a.u.)². Note that for $\nu = 1$, the mean and the covariance matrix of the t -distribution are not defined. In contrast, a coverage region for the distribution can always be determined [12, Clause 5.5.2, Note 1].

2.4.2.4 The propagation of the probability distributions for the input quantities through measurement model (2.1) was implemented in the R environment [107] by applying the R functions `rmvnorm` and `rmvt` available in the ‘mvtnorm’ package [58]. For each input quantity, $M = 10^7$ values were drawn. Since only positive values of the measurand are feasible, the joint input probability density functions were numerically truncated at zero by disregarding negative values drawn during the MCM simulation [17, Clause 9.4.2.1.1, Note], thus obtaining corresponding BaP mass values smaller in number than M . The number of values retained, however, was about 9×10^6 , sufficient to provide a reliable numerical approximation for the measurand distribution. From the MCM distribution, the shortest 95 % coverage interval was obtained and reported in table 2.3.

NOTE The rejection of unphysical MCM values for the input distributions is just an approximate and somewhat crude procedure to take into account the constraint $m_E > 0$. However, this procedure is aligned with the recommendation given in [48]. A rigorous treatment using Bayesian inference [126] would naturally include the constraint as prior knowledge about the measurand. Also see [17, Scope, NOTE 2].

2.5 Reporting the result

2.5.1 Figure 2.1 shows the results for m_E determined using GUF (bell-shaped curve) and MCM (frequency distribution shown as a histogram). The endpoints of the coverage intervals are also reported in table 2.3.

Table 2.3: Estimate, associated standard uncertainty and 95 % coverage interval for m_E as provided by GUF; median and shortest 95 % coverage interval from MCM

Method	\widehat{m}_E/ng	Median/ng	$u(\widehat{m}_E)/\text{ng}$	95 % C.I./ng
GUF	0.014	—	0.002	[0.009, 0.019]
MCM	—	0.014	—	[≈ 0 , 0.032]

2.5.2 In table 2.3, neither the mean nor the standard deviation of the MCM distribution are given, as they are unreliable. This is due to the presence in the input distributions of a bivariate for which these moments are not defined (see 2.4.2.3 and 2.6.1). By contrast, the median and the endpoints of coverage intervals, being quantiles (which always exist), are meaningful. It is worth noting that the MCM median (0.014 ng) is close to the GUF estimate.

2.6 Interpretation of results

2.6.1 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, ac-

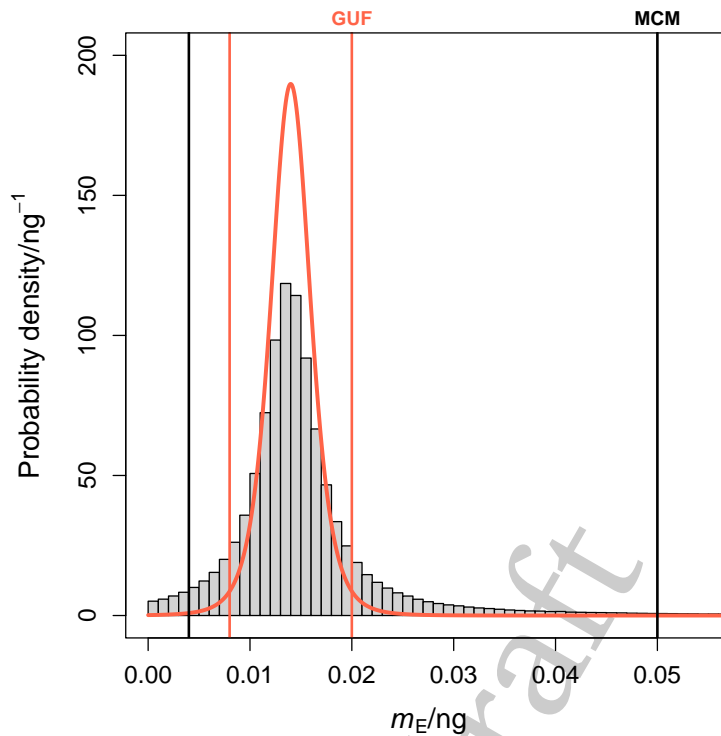


Figure 2.1: Results for m_E determined using GUF (bell-shaped curve) and MCM (frequency distribution shown as a histogram)

according to [17, eqns. (16) and (17)]. Nonetheless, [17, Clause 6, Note 2] states that in some special circumstances, such as when one of the input quantities has been assigned a Student 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 JCGM 101:2008 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 2.3, whereas feasible coverage intervals are those provided by MCM, as discussed shortly.

2.6.2 From figure 2.1, 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 3.2 times larger than that obtained in the GUF. Moreover, it appears asymmetric with respect to the measurand estimate because of the left-censoring of the values generated because of the physical constraint. This is a clear example of those situations in which the Central Limit Theorem does not hold, since the probability density function (PDF) for the output quantity is neither a normal distribution nor a scaled and shifted Student t -distribution.

2.6.3 Summarizing, this example is a clear case in which blind adherence either to the approach in JCGM 100:2008 [10] or to the MCM in [12, 17] would be dangerous. Care-

ful considerations of estimates, standard uncertainties and coverage intervals are always needed according to the specific problem under study. In this case, for example, the best solution is found to be a complementary use of both approaches, each one compensating the deficiencies of the other.

Final draft

3 Relative molecular mass of glucose

3.1 Preamble

3.1.1 This example is provided by the VSL, The Netherlands.

3.1.2 This example demonstrates how a relative molecular mass and the associated uncertainty are calculated using reference data, in this instance standard atomic weights as published by the Commission on Isotopic Abundances and Atomic Weights (CIAAW) of the International Union of Pure and Applied Chemistry (IUPAC). The case shows how to model ignorance about the isotopic composition of the elements in the material and how to propagate the measurement uncertainty using the law of propagation of uncertainty (LPU).

3.1.3 Relative molecular masses of substances, also known as ‘molecular weights’, are frequently used in science, for example, to study the thermodynamics of the conversion of glucose. Relative molecular masses can be calculated from the standard atomic weights of the elements present in the molecule, appreciating the coefficients describing the atomic composition. For glucose, whose chemical formula is $C_6H_{12}O_6$, these coefficients take the values 6 for carbon, 12 for hydrogen, and 6 for oxygen. Relative atomic and molecular masses are functions of the isotopic composition of the elements present in the material(s) at hand [20]. The standard atomic weights apply to ‘normal materials’, that is, materials that have not been artificially altered in isotopic composition and are a relevant source of the element. Such materials include laboratory reagents [130].

3.1.4 In this example, the isotopic composition of the elements in the material is unavailable. The material is a normal material, so that the standard atomic weights can be used [102, 123]. The calculation of the relative molecular mass of glucose will therefore be based on the assumption that the isotopic compositions of the elements in the molecule lie somewhere in the interval of values of isotopic compositions for normal materials. For elements with documented substantial variation in isotopic composition, the standard atomic weights are given in the form of an interval (see table 3.1).

Table 3.1: Lower (*a*) and upper (*b*) bounds for the standard atomic weight intervals of hydrogen, carbon and oxygen [105]

<i>i</i>	Element	<i>a</i>	<i>b</i>
1	hydrogen	1.007 84	1.008 11
2	carbon	12.009 6	12.011 6
3	oxygen	15.999 03	15.999 77

3.2 Measurement model

The relative molecular mass of glucose M_r is given by

$$M_r = \sum_{i=1}^3 \nu_i A_i, \quad (3.1)$$

where

ν_i denotes the atomic composition coefficient of element i in the molecule,
 A_i the standard atomic weight (relative atomic mass) of element i .

3.3 Calculation and uncertainty evaluation

3.3.1 For the calculation of the relative molecular mass, an estimate of the relative atomic mass of each element is required with an associated standard uncertainty. The measurement uncertainty should account for at least (a) the knowledge about the isotopic composition of the materials considered, and (b) the uncertainty associated with the measurement of the masses of the isotopes by mass spectroscopy.

3.3.2 For each element, both the estimate \hat{A}_i and the associated standard uncertainty $u(\hat{A}_i)$ can be obtained from a Type B evaluation of standard uncertainty using the atomic-weight intervals given in table 3.1. As the isotopic composition with respect to the elements H, C, and O can lie anywhere in the respective atomic-weight intervals, there is no specific information about the values within this interval. In this case, a rectangular (uniform) distribution is assigned to each input quantity A_i [10, clause 4.3.7]. For this probability distribution, the estimate of A_i and the standard uncertainty associated with this estimate are

$$\hat{A}_i = \frac{a_i + b_i}{2}, \quad u(\hat{A}_i) = \frac{b_i - a_i}{2\sqrt{3}}.$$

The resulting estimates and standard uncertainties associated with these input quantities are given in table 3.2.

Table 3.2: Estimates and standard uncertainties for the standard atomic weights of the elements found in glucose

i	Element	\hat{A}_i	ν_i	$u(\hat{A}_i)$	$\nu_i u(\hat{A}_i)$
1	hydrogen	1.007 98	12	0.000 08	0.000 94
2	carbon	12.010 6	6	0.000 58	0.003 46
3	oxygen	15.999 4	6	0.000 21	0.001 28

3.3.3 The estimate of the relative molecular mass of glucose is calculated using the measurement model (3.1) as

$$\hat{M}_r = 6 \times 12.0106 + 12 \times 1.00798 + 6 \times 15.9994 = 180.1557,$$

where the estimates of the standard atomic weights are taken from table 3.2.

3.3.4 The measurement model (3.1) is linear because the atomic composition coefficients are known exactly. In this case, the LPU can be applied to give an exact expression for the squared standard uncertainty associated with \hat{M}_r as follows:

$$u^2(\hat{M}_r) = \sum_{i=1}^3 \nu_i^2 u^2(\hat{A}_i). \quad (3.2)$$

The associated standard uncertainty is calculated from the contributions $\nu_i u(\hat{A}_i)$ given in the last column of table 3.2. The standard uncertainty is

$$u(\hat{M}_r) = (0.00094^2 + 0.00346^2 + 0.00128^2)^{1/2} = 0.0038.$$

3.4 Discussion

3.4.1 This example highlights how an estimate and the associated standard uncertainty can be evaluated from an atomic-weight interval. The probability density function chosen depends on the assumptions made, and these usually depend on the information at hand. Estimates and standard uncertainties thus derived for the three elements in question are independent of the isotopic composition and valid for any material that conforms with the description of a *normal material* of IUPAC [30, 130].

3.4.2 The use of the rectangular distribution underlines the lack of detailed information concerning the isotopic composition [102, 123] and is in agreement with earlier reports on the standard atomic weights [128, 129]. By no means can this choice of probability distribution be understood as depicting the actual isotopic composition. Rather, it depicts the state of knowledge about the composition, which is in this example very modest. The efforts of IUPAC in investigating the isotopic composition of elements in *normal materials* focus very much on establishing the interval end points, rather than modelling the isotopic composition of elements across materials [30]. The use of the rectangular distribution in connection with the standard atomic weights is not always sufficiently accurate to ignore completely the effects of the uncertainty associated with the relative molecular masses in a further application, for example, the calculation of the composition of a mixture expressed in amount-of-substance fractions [122].

3.4.3 Had the isotopic composition of the glucose been available, a more accurate value for the relative molecular mass M_r could have been obtained [102]. The estimate and standard uncertainty obtained in this example are for this application and many others in chemistry fully satisfactory and sufficiently accurate for further calculations [123]. With increasing accuracy of measurements, there will be a growing number of applications for which the approach to calculating relative molecular masses as outlined in this example is inadequate. In such cases, a batch-specific mass-spectrometric determination of the isotopic composition of the given sample used in the application would be required to obtain information about the isotopic composition of the elements in the material.

3.4.4 Calculations involving the standard atomic weights should be performed using the latest values published for these. At the time of writing, the standard atomic weights from 2021 [105] were the current.

4 Gravimetric mixture preparation and calculation of composition

4.1 Preamble

4.1.1 This example is provided by the VSL, The Netherlands.

4.1.2 This example demonstrates the uncertainty evaluation of the composition of a calibration gas mixture prepared by static gravimetry. It illustrates how data from weighing, purity analysis and molar masses are combined, and how correlations between quantities are taken into account. The propagation of uncertainty is carried out using the multivariate law of propagation of uncertainty (LPU) from JCGM 102:2011 [12].

4.1.3 Fundamental to the metrology of chemical composition is the preparation of mixtures having well-known compositions from pure substances. These mixtures are often used as (certified) reference materials for calibration [70]. There are two principal ways for preparing such mixtures: gravimetric and volumetric. In this example, the static gravimetric preparation [2, 76] of a gas mixture is considered. The mixture is prepared from two high-purity gases: carbon dioxide and nitrogen [72]. These parent gases have been checked for purity, so a correction can be applied for the contents of the impurities present in the gases. The objective is to calculate the amount fractions of the abundant components (the multivariate measurand) and their associated standard uncertainties and covariances.

4.2 Measurement model

4.2.1 The composition of a gravimetrically prepared gas mixture is commonly expressed in amount fractions, which are calculated as the ratio of the amount n_k of component k ($k = 1, \dots, q$) in the gas mixture to the sum n_{tot} of the amounts of all components in the mixture. In a slightly amended notation, the measurement model of ISO 6142-1 [2, 76] for the amount fraction y_k of a component k in a gravimetric mixture of p parent gases is

$$y_k = \frac{n_k}{n_{\text{tot}}}, \quad \text{with} \quad n_k = \sum_{j=1}^p \frac{m_j x_{kj}}{M_{p,j}}, \quad n_{\text{tot}} = \sum_{j=1}^p \frac{m_j}{M_{p,j}}, \quad (4.1)$$

where

y_k denotes the amount fraction of component k in the mixture,

n_k the amount of component k ,

n_{tot} the (total) amount of substance in the mixture,

m_j the mass of parent gas j added to the mixture,

x_{kj} the amount fraction of component k in parent gas j ,

$M_{p,j}$ the molar mass of parent gas j

p the number of parent gases,

q the number of components in the mixture.

Equation (4.1) is an example of a multivariate measurement model. The molar mass of parent gas j , $M_{p,j}$ is determined by the chemical composition and the molar masses of its constituent components, M_i ($i = 1, \dots, q$):

$$M_{p,j} = \sum_{i=1}^q x_{ij} M_i, \quad (4.2)$$

where

x_{ij} denotes the amount fraction of component i in parent gas j ,

M_i the molar mass of component i .

In the current example, there are two parent gases ($p = 2$) comprising eight components ($q = 8$) comprised of $e = 5$ different chemical elements.

4.2.2 The weighing is performed using the substitution method [5, 91, 122], where the sample and reference cylinders are weighed alternately a number of times. The sample cylinder is the cylinder in which the mixture is prepared, whereas the reference cylinder is a nominally identical cylinder. From these weighing data, the differences between the indications of the balance are calculated for the sample and reference cylinder (see table 4.1).

In doing so, a possible linear drift in the indications is minimised. The use of a reference cylinder of the same type with the same valve compensates for most of the buoyancy effects [5, 122]. The differences in the indications thus obtained are shown in table 4.1.

Table 4.1: Differences (\widehat{d}_{ij}) between the indications from weighing the sample and reference cylinders

i	$\widehat{d}_{i0}/\text{g}$ evacuated	$\widehat{d}_{i1}/\text{g}$ +parent 1	$\widehat{d}_{i2}/\text{g}$ +parent 2
1	-152.025	93.497	717.273
2	-152.025	93.495	717.274
3	-152.024	93.494	717.275
4	-152.024	93.494	717.276
5	-152.025	93.496	717.275
6	-152.025	93.495	717.274
7	-152.025	93.495	717.273

4.2.3 From the differences in table 4.1, the masses r_0 , r_1 and r_2 are obtained as the average of the differences, viz.,

$$r_j = \frac{1}{7} \sum_{i=1}^7 d_{ij}.$$

The standard uncertainty associated with \widehat{r}_j is evaluated as the combination of the standard deviation of the average differences r_j and the resolution of the balance (1 mg). The uncertainty contribution due to the finite resolution of the balance is modelled using the rectangular distribution, with standard uncertainty 0.29 mg. Table 4.2 summarises the mass differences and associated standard uncertainties.

Table 4.2: Mass differences \hat{r}_j and associated standard uncertainties of the sample cylinder during the filling process. The uncertainty contributions due to the repeatability of weighing ($u_{\text{rpt},j}$) and the finite resolution of the balance ($u_{\text{res},j}$) are given as well

j	0	1	2
\hat{r}_j/g	-152.0243	93.4951	717.2740
$u_{\text{rpt},j}/\text{g}$	0.0004	0.0007	0.0006
$u_{\text{res},j}/\text{g}$	0.0003	0.0003	0.0003
$u(\hat{r}_j)/\text{g}$	0.0005	0.0008	0.0007

4.2.4 This example deals with the preparation of a calibration gas mixture from two nominally pure gases. The amount fraction of the most abundant component is calculated by subtracting the amount fractions of the impurities from one [72]

$$x_k = 1 - \sum_{i=1, i \neq k}^q x_i. \quad (4.3)$$

Using the LPU and assuming that the amount fractions of the impurities are mutually independent [11, equation (10)],

$$u^2(x_k) = \sum_{i=1, i \neq k}^q u^2(x_i). \quad (4.4)$$

Using the LPU [11, equation (F.2)], the covariance between x_k and x_i is

$$u(x_k, x_i) = -u^2(x_i). \quad (4.5)$$

4.2.5 The chemical composition of the two parent gases is given in table 4.3. It is noted that the standard uncertainty associated with the amount fraction nitrogen in the nitrogen gas is only marginally larger than the standard uncertainty associated with the amount fraction argon (the dominant uncertainty component). The difference in these standard uncertainties vanishes when rounding these standard uncertainties in accordance with applicable guidance [11, clause 7].

Table 4.3: Chemical composition of the two parent gases expressed as amount fractions of their components in $\mu\text{mol mol}^{-1}$

Component	Carbon dioxide gas		Nitrogen gas	
	\hat{x}	$u(\hat{x})$	\hat{x}	$u(\hat{x})$
Ar			5.0	3.0
CH ₄	0.10	0.06	0.008	0.005
CO	0.5	0.3	0.015	0.009
CO ₂	999 988.0	3.4	0.010	0.006
H ₂			0.025	0.015
H ₂ O	1.9	0.1	0.010	0.006
N ₂	4.3	2.1	999 994.9	3.0
O ₂	5.3	2.6	0.005	0.003

4.3 Uncertainty evaluation

4.3.1 Correlations

The data in Tables 4.2 and 4.3 are used as estimates and associated standard uncertainties of the corresponding quantities. There are correlations between the input quantities in the measurement model (4.1). These correlations should be appreciated in the uncertainty evaluation. First, there are correlations between the molar masses of the components. For example, identical isotopic composition of oxygen in the components CO and CO₂ is assumed; mutual independence between the standard atomic weights is also assumed. Second, the masses of the transferred parent gases are correlated because the weighings are done sequentially and the estimates of both parent gas masses,

$$m(\text{CO}_2) = r_1 - r_0 \quad \text{and} \quad m(\text{N}_2) = r_2 - r_1,$$

are dependent on r_1 . Last, the estimates of the impurity amount fractions in the parent gases are correlated since the amount fraction of the most abundant component is estimated by mass balance.

4.3.2 General considerations

4.3.2.1 The multivariate LPU in JCGM 102:2011 uses vectors and matrices. For the measurement model (4.1) of ISO 6142-1, the vector of input quantities can be considered to be partitioned into several input vectors, namely one containing the masses of the transferred parent gases $\mathbf{m} = [m_1, \dots, m_p]^\top$, another containing the molar masses $\mathbf{M} = [M_1, \dots, M_q]^\top$, and finally $j = 1, \dots, p$ vectors \mathbf{x}_j . As for this case, there are no dependencies between the vectors \mathbf{m} , \mathbf{M} and the \mathbf{x}_j , the LPU from JCGM 102:2011 can be used on each matrix separately to obtain the covariance matrix $U_{\hat{\mathbf{n}}}$ associated with the amounts estimate $\hat{\mathbf{n}}$.

4.3.2.2 Correlations between the compositions of the parent gas mixtures can arise in multi-stage gas mixture preparation, where for instance one gas mixture is diluted with one of the parent gases used in a previous stage [122]. In that case, it can be necessary to consider further dependencies among the input quantities of equation (4.1).

4.3.3 Weighing data

4.3.3.1 The mass of a parent gas m_j is given by the difference between the weighings of a cylinder before and after adding the parent gas to it. Since the relationship between these two observed quantities is linear, it can be expressed in matrix notation as follows:

$$\mathbf{m} = \mathbf{C}_r \mathbf{r}, \tag{4.6}$$

where

$\mathbf{r} = [r_0, \dots, r_p]^\top$ denotes a vector of mass readings of sample cylinder with respect to the reference cylinder shown in table 4.2,

\mathbf{m} a vector of the masses of the parent gases,

\mathbf{C}_r the sensitivity matrix relating \mathbf{m} to \mathbf{r} .

In this example, \mathbf{m} is a vector of length $p = 2$ and \mathbf{C}_r the $p \times (p + 1)$ matrix

$$\mathbf{C}_r = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}.$$

Using the LPU from JCGM 102:2011, the covariance matrix associated with $\widehat{\mathbf{m}}$ is

$$\mathbf{U}_{\widehat{\mathbf{m}}} = \mathbf{C}_r \mathbf{U}_{\widehat{\mathbf{r}}} \mathbf{C}_r^\top,$$

where

$\mathbf{U}_{\widehat{\mathbf{r}}}$ denotes the covariance matrix associated with $\widehat{\mathbf{r}}$,

$\mathbf{U}_{\widehat{\mathbf{m}}}$ the covariance matrix associated with $\widehat{\mathbf{m}}$.

4.3.3.2 As the mass measurements \widehat{r}_j ($j = 1 \dots p+1$), shown in table 4.2, can be assumed to be uncorrelated [122], the covariance matrix $\mathbf{U}_{\widehat{\mathbf{r}}}$ is the diagonal matrix

$$\mathbf{U}_{\widehat{\mathbf{r}}} = \text{diag} \{u^2(\widehat{r}_1), u^2(\widehat{r}_2), u^2(\widehat{r}_3)\},$$

where $u^2(\widehat{r}_j)$ denotes the squared standard uncertainty associated with \widehat{r}_j .

4.3.4 Molar masses

4.3.4.1 The molar masses are calculated from the standard atomic weights \mathbf{A} and the subscripts in the formulæ of the molecules involved [76, 122], using

$$\mathbf{M} = \boldsymbol{\nu} \mathbf{A} M_{\text{u}}, \quad (4.7)$$

where

\mathbf{M} denotes the vector holding the molar masses of the q components in the final mixture,

$\boldsymbol{\nu}$ the matrix holding the coefficients in the q molecular formulæ of the components (see table 4.6),

\mathbf{A} denotes the vector holding the e standard atomic weights of the elements occurring in the components, as shown in table 4.5,

M_{u} denotes the molar mass constant, taken for practical purposes as $M_{\text{u}} = 1 \text{ g mol}^{-1}$ and ignoring its uncertainty, as it is negligible for this application.

The subscripts in the molecule formulæ, presented in a $q \times e$ matrix $\boldsymbol{\nu}$, are constants, where each row corresponds to a component of the prepared gas mixture and columns correspond to different chemical elements, as shown in table 4.5.

4.3.4.2 The estimate $\widehat{\mathbf{M}}$ of \mathbf{M} is given by $\widehat{\mathbf{M}} = \boldsymbol{\nu} \widehat{\mathbf{A}} M_{\text{u}}$, where $\widehat{\mathbf{A}}$ is the estimate of \mathbf{A} . The covariance matrix associated with $\widehat{\mathbf{M}}$ is obtained by applying the LPU to equation (4.7):

$$\mathbf{U}_{\widehat{\mathbf{M}}} = \boldsymbol{\nu} \mathbf{U}_{\widehat{\mathbf{A}}} \boldsymbol{\nu}^\top, \quad (4.8)$$

where $\mathbf{U}_{\widehat{\mathbf{A}}} = \text{diag} \{u^2(\widehat{A}_1), \dots, u^2(\widehat{A}_e)\}$. The standard atomic weights of the elements are assumed to be independent. This is an appropriate assumption with the standard atomic weights, although not applicable generally to atomic weights in normal materials (see clause 3.4.2) [102, 123].

4.3.4.3 The standard atomic weights are modelled using the rectangular distributions, whose bounds are provided by International Union of Pure and Applied Chemistry (IUPAC) (see table 4.4). Recall that a rectangular distribution $R(a, b)$ with endpoints a and b ($a < b$) has expectation $(a + b)/2$ and standard deviation $(b - a)/\sqrt{12}$ [10, clause 4.3.7]. Guidance on the use of atomic weights and assigning probability density functions (PDFs) is given in [102, 123].

Table 4.4: Standard atomic weights of the makeup elements for components in the two parent gases

Element	Distribution	a	b	\hat{A}	$u(\hat{A})$
H	$R(a, b)$	1.00784	1.00811	1.00798	0.00008
C	$R(a, b)$	12.0096	12.0116	12.0106	0.0006
N	$R(a, b)$	14.00643	14.00728	14.00686	0.00025
O	$R(a, b)$	15.99903	15.99977	15.99940	0.00021
Ar	$R(a, b)$	39.792	39.963	39.878	0.049

4.3.4.4 Combining the subscripts in the molecule formulæ of the components and the atomic weights of the elements (Eqs. 4.7 and 4.8) provides molecular weight estimates which are summarised in table 4.5.

Table 4.5: Atomic composition coefficients and molar masses of the components in the two parent gases

Component	$\nu(\text{H})$	$\nu(\text{C})$	$\nu(\text{N})$	$\nu(\text{O})$	$\nu(\text{Ar})$	$\hat{M}/$ (g mol^{-1})	$u(\hat{M})/$ (g mol^{-1})
Ar	0	0	0	0	1	39.878	0.049
CH ₄	4	1	0	0	0	16.0425	0.0007
CO	0	1	0	1	0	28.0100	0.0006
CO ₂	0	1	0	2	0	44.0094	0.0007
H ₂	2	0	0	0	0	2.01595	0.00016
H ₂ O	2	0	0	1	0	18.0154	0.0003
N ₂	0	0	2	0	0	28.0137	0.0005
O ₂	0	0	0	2	0	31.9988	0.0004

4.3.4.5 The covariance matrix $U_{\hat{M}}$ is related with the correlation matrix $R_{\hat{M}}$ in the following way [12, Definition 3.20, Note 3]:

$$U_{\hat{M}} = D R_{\hat{M}} D, \quad (4.9)$$

where the matrix D is defined as

$$D = \text{diag}\{u(\hat{M}_1) \dots u(\hat{M}_q)\}.$$

Conversely, the correlation matrix is obtained from the covariance matrix as follows:

$$R_{\hat{M}} = D^{-1} U_{\hat{M}} D^{-1}. \quad (4.10)$$

The covariance matrix $U_{\widehat{M}}$ is usually not full rank, as the formulæ used for computing the molar masses can be linearly dependent (e.g., as in the case of the molar masses of alkanes, which have as general molecular formula C_nH_{2n+2} ($n \geq 1$)).

4.3.4.6 The upper triangular part of the correlation matrix $R_{\widehat{M}}$ associated with the molar masses is

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & 1 & 0.825 & 0.707 & 0.475 & 0.280 & 0 & 0 \\ & & 1 & 0.960 & 0 & 0.280 & 0 & 0.347 \\ & & & 1 & 0 & 0.481 & 0 & 0.595 \\ & & & & 1 & 0.589 & 0 & 0 \\ & & & & & 1 & 0 & 0.808 \\ & & & & & & 1 & 0 \\ & & & & & & & 1 \end{bmatrix}.$$

There is a large correlation (0.960) between the third and fourth components, CO and CO₂, whereas the second-largest correlation (0.825) is between the molar masses of the second and third components, CH₄ and CO.

4.3.5 Composition of the parent gases

4.3.5.1 When preparing a gas mixture, the composition of the parent gases plays an important role and three distinct scenarios can be considered [122]:

1. The parent gas is a near-pure gas whose impurities have been characterized;
2. The parent gas is a gravimetrically prepared gas mixture;
3. All components in the parent gas have been analysed.

4.3.5.2 In all cases, the covariance matrix $U_{\widehat{x}_j}$ of the amount fractions in each parent gas is not diagonal [122]. Due to the fact that the amount fractions of all components in the parent gas mixtures add up to unity, these covariance matrices are always singular [4].

4.3.5.3 The covariance matrices associated with the compositions of the two parent gases can be readily formed once the amount fractions of the impurities are known. The formulæ (4.3), (4.4) and (4.5) provide the expressions for obtaining the estimates and the associated covariance matrices for both parent gases.

4.3.6 Composition of the gas mixture

4.3.6.1 Consider a vector \mathbf{n} , containing the amounts of the components and their total amount in the mixture. That is, $\mathbf{n} = (n_1, \dots, n_q, n_{\text{tot}})^T$. From this vector, one can compute the vector \mathbf{y} , that is, the amount fractions of the individual components in the gas mixture. The sensitivity matrices are obtained with respect to the vector \mathbf{n} .

4.3.6.2 All input quantities can be represented by a vector \mathbf{a} formed by ‘stacking’ the vectors \mathbf{m} , \mathbf{M} , $\mathbf{x}_1, \dots, \mathbf{x}_p$:

$$\mathbf{a} = \begin{bmatrix} \mathbf{m} \\ \mathbf{M} \\ \mathbf{x} \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_p \end{bmatrix}. \quad (4.11)$$

The covariance matrix $U(\hat{\mathbf{a}})$ associated with the estimate $\hat{\mathbf{a}}$ of \mathbf{a} holds the variances associated with the elements of $\hat{\mathbf{a}}$ and the covariances between pairs of these elements.

4.3.6.3 If there are no dependencies between the vectors \mathbf{m} , \mathbf{M} , and \mathbf{x} , then $U_{\hat{\mathbf{a}}}$ can be written in block matrix notation as follows:

$$U_{\hat{\mathbf{a}}} = \text{diag}\{U_{\hat{\mathbf{m}}}, U_{\hat{\mathbf{M}}}, U_{\hat{\mathbf{x}}}\}. \quad (4.12)$$

The construction of the covariance matrices $U_{\hat{\mathbf{m}}}$, $U_{\hat{\mathbf{M}}}$, and $U_{\hat{\mathbf{x}}}$ is discussed briefly in Clauses 4.3.3, 4.3.4, and 4.3.5, respectively. The method of construction is based on the LPU from JCGM 102:2011 [12] considering multivariate sub-models. Equations (4.11) and (4.12) are not the most efficient for this data set, but can be readily adapted when, e.g., calculating the composition of a calibration gas mixture prepared by gravimetrically diluting a calibration gas mixture.

4.3.6.4 The expression for $U_{\hat{\mathbf{n}}}$ is

$$U_{\hat{\mathbf{n}}} = \mathbf{C}_a U_{\hat{\mathbf{a}}} \mathbf{C}_a^\top, \quad (4.13)$$

where \mathbf{C}_a can be written in block matrix notation as

$$\mathbf{C}_a = \begin{bmatrix} \mathbf{C}_m & \mathbf{C}_M & \mathbf{C}_{x_1} & \dots & \mathbf{C}_{x_p} \end{bmatrix}.$$

4.3.6.5 The matrices \mathbf{C}_m , \mathbf{C}_M , \mathbf{C}_{x_j} hold the sensitivity coefficients of all input parameters (see also [2, 76]). The expressions for the elements of these matrices have been derived elsewhere [122]. The sensitivity matrix \mathbf{C}_m of dimension $(q+1) \times p$ is obtained as follows:

$$\mathbf{C}_m = \begin{bmatrix} \frac{x_{11}}{M_{p,1}} & \frac{x_{12}}{M_{p,2}} & \dots & \frac{x_{1p}}{M_{p,p}} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{x_{q1}}{M_{p,1}} & \frac{x_{q2}}{M_{p,2}} & \dots & \frac{x_{qp}}{M_{p,p}} \\ \frac{1}{M_{p,1}} & \frac{1}{M_{p,2}} & \dots & \frac{1}{M_{p,p}} \end{bmatrix},$$

where $M_{p,j}$ is the average molar mass of a parent gas j , as given by equation (4.2). Similarly, the sensitivity matrix \mathbf{C}_M of dimension $(q+1) \times q$ is obtained as follows:

$$\mathbf{C}_M = - \begin{bmatrix} \sum_{j=1}^p \frac{x_{1j}m_j}{M_{p,j}^2} x_{1j} & \dots & \sum_{j=1}^p \frac{x_{1j}m_j}{M_{p,j}^2} x_{qj} \\ \vdots & \vdots & \vdots \\ \sum_{j=1}^p \frac{x_{qj}m_j}{M_{p,j}^2} x_{1j} & \dots & \sum_{j=1}^p \frac{x_{qj}m_j}{M_{p,j}^2} x_{qj} \\ \sum_{j=1}^p \frac{m_j}{M_{p,j}^2} x_{1j} & \dots & \sum_{j=1}^p \frac{m_j}{M_{p,j}^2} x_{qj} \end{bmatrix},$$

and the sensitivity matrix \mathbf{C}_{x_j} of dimension $(q+1) \times q$ is given as follows:

$$\mathbf{C}_{x_j} = - \frac{m_j}{M_{p,j}^2} \begin{bmatrix} x_{1j}M_1 - M_{p,j} & x_{1j}M_2 & \dots & x_{1j}M_q \\ x_{2j}M_1 & x_{2j}M_2 - M_{p,j} & \dots & x_{2j}M_q \\ \vdots & \vdots & \vdots & \vdots \\ x_{qj}M_1 & x_{qj}M_2 & \dots & x_{qj}M_q - M_{p,j} \\ M_1 & M_2 & \dots & M_q \end{bmatrix}.$$

4.3.6.6 The propagation of the uncertainty associated with the amount of substance of the components to the amount fractions can be carried out as follows. The covariance matrix $U_{\hat{y}}$ associated with the vector \hat{y} of the estimates of the amount fractions of the q components in the mixture is calculated from the covariance matrix $U_{\hat{n}}$ associated with \hat{n} using

$$U_{\hat{y}} = C_n U_{\hat{n}} C_n^T. \quad (4.14)$$

The sensitivity matrix C_n of dimension $q \times (q + 1)$ is given in block matrix notation by

$$C_n = \begin{bmatrix} n_{\text{tot}}^{-1} I & -n_{\text{tot}}^{-2} \boldsymbol{\eta} \end{bmatrix}, \quad (4.15)$$

where I denotes the identity matrix of dimension $q \times q$ and $\boldsymbol{\eta} = (n_1, \dots, n_q)^T$.

4.3.6.7 The covariance matrix $U_{\hat{y}}$ associated with a complete composition \hat{y} is singular, as the amount fractions of the components sum to unity [4]. Due to round-off effects, the covariance matrix corresponding to the correlation matrix $R_{\hat{y}}$ can appear to be non-singular (for the relationship between the correlation matrix and covariance matrix, see JCGM 102:2011 [12, definition 3.21]).

4.3.6.8 The results for the amount fractions for the mixture are given in table 4.6.

Table 4.6: Chemical composition of CO₂ and N₂ mixture expressed in amount fractions and their associated standard uncertainties

Component	$\hat{y}/$ ($\mu\text{mol mol}^{-1}$)	$u(\hat{y})/$ ($\mu\text{mol mol}^{-1}$)	$u_{\text{rel}}(\hat{y})/$ %
Ar	4.0	2.4	60
CH ₄	0.026	0.013	48
CO	0.112	0.061	54
CO ₂	200345	5	0.003
H ₂	0.020	0.012	60
H ₂ O	0.379	0.021	5
N ₂	799649	5	0.0006
O ₂	1.07	0.52	49

4.3.6.9 The upper triangular part of the correlation matrix for the chemical composition of the gas mixture $R_{\hat{y}}$ is

$$\begin{bmatrix} 1 & 0.00 & 0.00 & 0.04 & 0.00 & 0.00 & -0.50 & 0.00 \\ & 1 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ & & 1 & -0.01 & 0.00 & 0.00 & 0.00 & 0.00 \\ & & & 1 & 0.00 & 0.00 & -0.86 & -0.08 \\ & & & & 1 & 0.00 & 0.00 & 0.00 \\ & & & & & 1 & 0.00 & 0.00 \\ & & & & & & 1 & -0.02 \\ & & & & & & & 1 \end{bmatrix}.$$

The correlation matrix associated with the composition of the gas mixture shows that most correlations between pairs of amount fraction are rather weak; only the correlations between the amount fractions of argon and nitrogen (-0.50), and between carbon dioxide

and nitrogen (-0.86) have appreciable influence. The joint PDF of the latter is shown in figure 4.1.

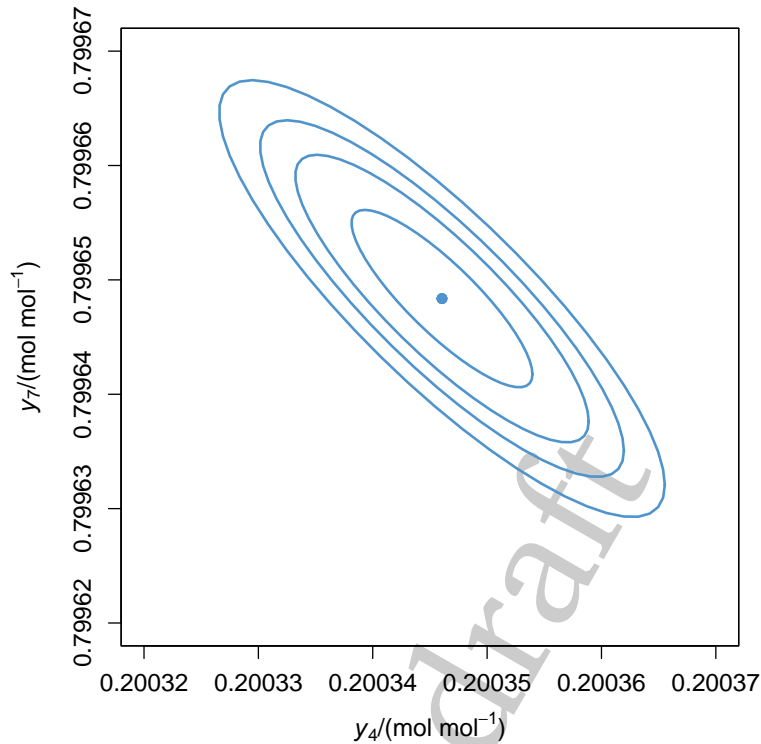


Figure 4.1: Contour plot of the bivariate normal probability density function for $\text{N}_2(y_7)$ and $\text{CO}_2(y_4)$ amount fractions in the gas mixture with contours corresponding to 68 %, 95 %, 99 % and 99.9 % coverage probability

4.4 Discussion

This example illustrates some of the advantages of using the multivariate form of the LPU (JCGM 102:2011). First, the LPU does not only provide the variances associated with the estimates of the various output quantities but also their covariances. Second, the multivariate modelling aids in structuring the model in a fashion that can be readily implemented. Important submodels are the formation of the covariance matrices of the transferred masses, molar masses, and the composition of the parent gases. The use of multivariate modelling is particularly useful in describing multi-stage gas mixture preparation.

5 Greenhouse gas emission inventories

5.1 Preamble

This example is taken from a compendium of examples [1].

5.1.1 General

5.1.1.1 The compilation of country-scale greenhouse gas emissions' inventories involves combining standard emissions' factors and activity data to provide sector emissions' estimates. This example 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 for Guides in Metrology (JCGM).

5.1.1.2 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.

5.1.1.3 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.

5.1.1.4 Current practice provided by the IPCC regarding uncertainty propagation in the area is considered in the context of the IPCC Guidelines for National Greenhouse Gas Inventories [66]. The IPCC practice is contrasted with that in the Guide to the expression of uncertainty in measurement (GUM) [10] and other related JCGM guides. In particular, the possible effects of correlation, often ignored in practice, on the estimate are considered.

5.1.2 Greenhouse gas inventories

5.1.2.1 GHG inventories of the United Kingdom (UK) and other countries are compiled according to IPCC 2006 Guidelines [66]. Each year the inventory is updated to include the latest data available. Together with many other countries, the UK submits a report to the United Nations Framework Convention on Climate Change (UNFCCC) annually via a consolidated report.

5.1.2.2 The GHG inventory includes seven gases under the Kyoto Protocol, namely, carbon dioxide (CO₂), methane (CH₄), nitrous oxide (N₂O), hydrofluorocarbons (HFCs), sulfur hexafluoride (SF₆), perfluorocarbons (PFCs) and nitrogen trifluoride (NF₃).

5.1.2.3 A bottom-up calculation of emissions is based on contributions of the form

$$\text{Emissions (mass)} = \text{Emissions' factor (mass per volume)} \times \text{Activity data (volume)}. \quad (5.1)$$

In all, there are 743 emissions' factors and more than 1700 sources of emissions, and 17 key data sources together with involvement from industry, government and academics [19].

5.1.2.4 In terms of emissions from the three largest GHG contributors — CO₂, CH₄ and N₂O — the majority of the uncertainty in the GHG inventory comes from the agriculture, land use and waste sectors [23], with these three sectors contributing nearly 90 % of the uncertainty in the total inventory emissions.

5.1.2.5 European countries, especially the UK, are moving towards metrological assessment of current uncertainty quantification, including accounting for the effect of correlated quantities. Also, they are extending the current top-down validation activity [7] to cover the complete GHG inventory.

5.1.2.6 Figure 5.1 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 proportional contribution.

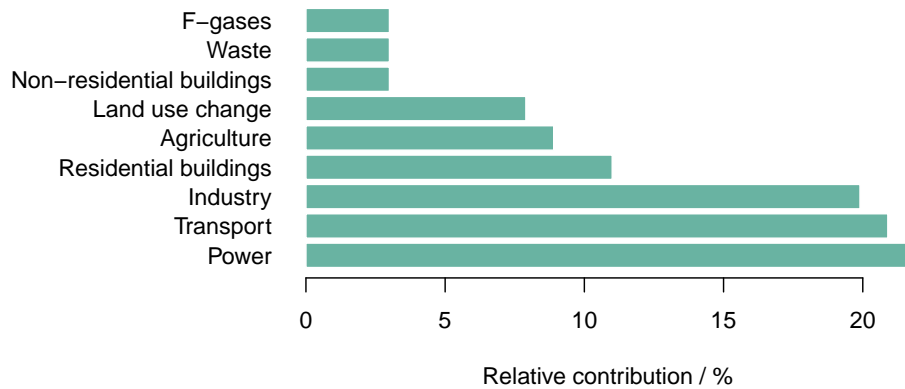


Figure 5.1: Inventory contributions across sectors

5.1.2.7 An emissions' inventory can be expressed as a linear combination of terms of the form (5.1):

$$E = \sum_i F_i A_i, \quad (5.2)$$

with

- E Total GHG emission for a given sector, geographic area, and time period,
- F_i Emission factor for the emissions of a given pollutant from source category i ,
- A_i Activity data for source category i .

5.1.2.8 In practice, the model (5.2) is often enlarged to incorporate scaling factors s_i to allow total GHG emission to be given in terms of 'CO₂ equivalent' (CO₂e). That is, each GHG in the table has a conversion or scaling factor associated with it, giving

$$E(\text{CO}_2\text{e}) = \sum_i s_i F_i A_i, \quad (5.3)$$

An example incorporating such scaling factors is given in clause 5.4.

5.1.3 Combination and correlation

5.1.3.1 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 [10, clause 5.1.2, formula (10)].

5.1.3.2 Figure 5.2 shows the expanded uncertainty contributions across sectors, with agriculture being the biggest contributor.

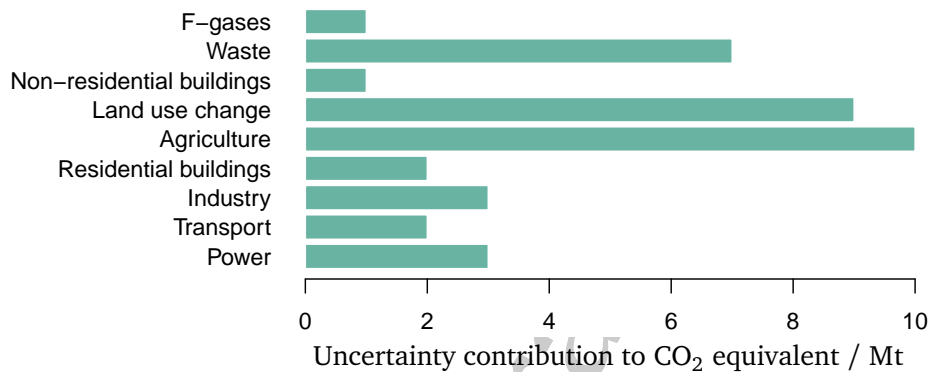


Figure 5.2: Expanded uncertainty contributions (at 95% confidence level) across sectors, expressed in CO₂-equivalent per Mt

5.1.3.3 Emissions' inventories are critical to many environmental decision-making processes [18]. 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 measurement, modelling or expert judgment, the main sources of 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 (5.3) that comprise an emission inventory is needed.

5.1.4 IPCC and JCGM reporting guidelines

5.1.4.1 The IPCC reporting guidelines refer to the 'error propagation approach' and the 'Monte Carlo approach' for uncertainty propagation. An extract from [67], 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.'

5.1.4.2 In this example, the IPCC approach to uncertainty evaluation in the context of greenhouse gas emissions is considered. Statements are made concerning the extent of

the alignment with guidance provided by the JCGM. Wherever measurement uncertainty is used, it is important that it is evaluated in the same manner by different practitioners so that users' interpretation is consistent and reliable. Universally agreed methods for uncertainty evaluation are hence required.

5.2 Specification of the measurand

A meaningful interpretation of the measurand, used here, is 'total emissions for a given pollutant, geographic area and time period' (see clause 5.1).

5.3 Measurement model

The IPCC does not seem explicitly to use the concept of 'measurement model' in their documents. However, expression (5.3) can be regarded as a basic measurement model central to IPCC considerations.

5.4 Uncertainty propagation

5.4.1 The use of the model specified as expression (5.3), in which the sector is taken as mobile machinery in agriculture (IPCC category A4ci), the region as the UK and the time period as 2018, is considered. Uncertainties are propagated using the LPU in JCGM 100:2008 [10], comparable to the IPCC's 'error propagation' method. The quantities concerned are taken as independent. Then, perceived correlation is considered.

5.4.2 Data for this sector from the UK National Atmospheric Emissions Inventory site (2018) are given in table 5.1 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 (the A_i) for those fuels. The table also specifies the Emissions' Factors (the F_i) for each of the key GHGs (CO_2 , CH_4 and N_2O) for those fuels. The reported relative expanded uncertainties $U_{\text{rel}}(\cdot)$ for the individual activity data and emission factors are given for 95% coverage probability.

Table 5.1: Reported activity data and emissions' factors with associated relative expanded uncertainties for 95% coverage probability

Fuel	Gas	A_i/TJ	$U_{\text{rel}}(A_i)/\%$	$F_i/(\text{kt TJ}^{-1})$	$U_{\text{rel}}(F_i)/\%$
Gas oil	CO_2	56 317.0920	38.6	2.0438×10^{-2}	2.7
Gas oil	CH_4	56 317.0920	1.6	3.5368×10^{-6}	80.0
Gas oil	N_2O	56 317.0920	1.6	3.0984×10^{-6}	216.3
Petrol	CO_2	67.1856	50.7	1.9127×10^{-2}	2.0
Petrol	CH_4	67.1856	1.6	4.8654×10^{-5}	80.0
Petrol	N_2O	67.1856	1.6	3.3578×10^{-7}	216.3

5.4.3 Table 5.1 incorporates the scaling mentioned in clause 5.1.2. The scaling factors to give 'CO₂ equivalent' emissions are $s = 11/3$ for CO₂ as C (carbon), $s = 25$ for methane (CH₄) and $s = 298$ for nitrous oxide (N₂O) [44]. These values, for which no uncertainty is reported, have recently been superseded [6]. The value 11/3 is identical to the molecular weight ratio 44/12 used to convert a mass of elemental carbon (C) emissions into the equivalent mass of carbon dioxide (CO₂) emissions. The updated values are recommended for

future climate reporting and inventory purposes, as they reflect the latest scientific understanding. The values used in this example are retained for use with previous inventories, as here. For the provenance of the data used in this example, see [23].

5.4.4 Since the A_i and the F_i denote the activity data values and the corresponding emissions' factors, respectively, in table 5.1,

$$E(\text{CO}_2\text{e}) = \sum_{i=1}^6 E_i(\text{CO}_2\text{e}) = \sum_{i=1}^6 s_i A_i F_i, \quad (5.4)$$

where the s_i are the scaling factors, assumed to have no associated uncertainty.

5.4.5 Total CO_2 equivalent emissions for this sector were reported to be 4282 kt. The reported expanded uncertainty (at the 95 % level of confidence) associated with this total was 1635 kt (relative expanded uncertainty of 38 %).

5.4.6 Table 5.2 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).

Table 5.2: Emissions as CO_2 equivalent for each contribution and the total CO_2 equivalent

GHG	Fuel	E_i/kt	s_i	$E_i(\text{CO}_2\text{e})/\text{kt}$
CO_2 as C	Gas oil	1150.9806	11/3	4220.2623
CH_4	Gas oil	0.1992	25	4.9795
N_2O	Gas oil	0.1745	298	51.9987
CO_2 as C	Petrol	1.2851	11/3	4.7120
CH_4	Petrol	0.0033	25	0.0817
N_2O	Petrol	0.0000	298	0.0067
				$E(\text{CO}_2\text{e}) = 4282.0409 \text{ kt}$

The application of the LPU [10, equation (10)] for independent quantities gives the associated standard uncertainty $u_{\text{ind}}(E(\text{CO}_2\text{e}))$:

$$u_{\text{ind}}^2(E(\text{CO}_2\text{e})) = \sum_{i=1}^6 s_i^2 [A_i^2 u^2(F_i) + F_i^2 u^2(A_i)]. \quad (5.5)$$

The use of the data in tables 5.1 and 5.2 gives $E(\text{CO}_2\text{e})$, as above, and by applying expression (5.2),

$$E(\text{CO}_2\text{e}) = 4282.04 \text{ kt}, \quad u_{\text{ind}}(E(\text{CO}_2\text{e})) = 835.16 \text{ kt}. \quad (5.6)$$

The uncertainty budget is given in table 5.3.

5.4.7 Perceived correlations associated with the various quantities, which seem not to be strongly considered in the IPCC documents, are now considered. It appears that correlation is not entertained with the 'error propagation' approach, although it is stated, without giving detail, that the Monte Carlo method may be used for this purpose [65, clause 6.3].

Table 5.3: Uncertainty budget

GHG	Fuel	$u(E_i(\text{CO}_2\text{e}))/\text{kt}$
CO ₂ as C	Gas oil	833.1844
CH ₄	Gas oil	2.0329
N ₂ O	Gas oil	57.3860
CO ₂ as C	Petrol	1.2198
CH ₄	Petrol	0.0334
N ₂ O	Petrol	0.0074
$u(E(\text{CO}_2\text{e}))/\text{kt}$		835.16

5.4.8 Examining the activity contributions in table 5.1, A_2 and A_3 are identical numerically as are their associated standard uncertainties $u(A_2)$ and $u(A_3)$. A similar statement applies to A_5 and A_6 and their associated standard uncertainties. This observation raises the possibility that A_2 and A_3 have a common origin and might even be the same quantity, and similarly for A_5 and A_6 . In discussion with GHG experts, this perception is considered to be more than reasonable.

5.4.9 Assuming this perception to be correct, an 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, whether the (numerically identical) activities A_2 and A_3 , and A_5 and 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. So, taking A_2 and A_3 as the same quantity, and similarly for A_5 and A_6 , instead of the model (5.3), we consider the model

$$E = s_1 A_1 F_1 + A_2 (s_2 F_2 + s_3 F_3) + s_4 A_4 F_4 + A_5 (s_5 F_5 + s_6 F_6).$$

For the current data, $E(\text{CO}_2\text{e})$ is identical to the previously calculated value, as expected and, applying LPU, since the s_i are constants,

$$\begin{aligned} u_{\text{cor}}^2(E(\text{CO}_2\text{e})) &= \sum_{i=1,4} s_i^2 [A_i^2 u^2(F_i) + F_i^2 u^2(A_i)] \\ &+ \sum_{i=2,5} \{A_i^2 [s_i^2 u^2(F_i) + s_{i+1}^2 u^2(F_{i+1})] + (s_i F_i + s_{i+1} F_{i+1})^2 u^2(A_i)\}, \end{aligned} \quad (5.7)$$

that is,

$$u_{\text{cor}}(E(\text{CO}_2\text{e})) = 835.16 \text{ kt}, \quad (5.8)$$

the same value to the number of decimal digits displayed as obtained when not taking correlation into consideration. Algebraically, however, using expressions (5.5) and (5.7),

$$u_{\text{cor}}^2(E(\text{CO}_2\text{e})) - u^2(E(\text{CO}_2\text{e})) = 2s_2 F_2 s_3 F_3 u^2(A_2) + 2s_5 F_5 s_6 F_6 u^2(A_5), \quad (5.9)$$

which is positive and demonstrated by displaying $u^2(E(\text{CO}_2\text{e}))$ and $u_{\text{cor}}^2(E(\text{CO}_2\text{e}))$ to a greater number of decimal places. This result suggests that considering correlation is not meaningful for such calculations. However, rationale is presented why accounting for covariance in emissions' calculation is generally appropriate.

5.4.10 The reason for the increase in the standard uncertainty associated with total emissions being so small is that one particular contribution to the budget, CO₂ as C for Gas oil, the first in table 5.2, is by far dominant. This effect has the consequence that the consideration of correlation is *unnecessary* in this example.

5.4.11 The principle holds that such correlations should generally be taken into account, as they may make a significant difference where one uncertainty contribution is less dominant. Such an instance was given in a presentation to the 2022 WMO-BIPM Workshop on Metrology for Climate Action. It was reported that a similar analysis was conducted for Waste Incineration (IPCC category 5C). The study, as yet unpublished, found that the estimated equivalent standard uncertainty contribution from CO₂ was 24 % higher when covariance was accounted for, compared to when it was ignored.

5.4.12 It would be possible to handle the uncertainty propagation using [10, formula (13)] taking correlation into account for the quantities concerned. The use of variable substitution as above reparametrizes the problem in such a way that the need to take correlation directly into account is avoided. The assumption is made that the remaining quantities are independent, which should be verified in practice to the extent possible.

5.4.13 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.

5.4.14 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 (5.5) becomes

$$u_{\text{ind}}^2(E(\text{CO}_2\text{e})) = 6s^2A^2u_F^2 + 6F^2s^2u_A^2 = s^2A^2F^2 [6u_{\text{rel}}^2(F) + 6u_{\text{rel}}^2(A)] \quad (5.10)$$

with relative standard uncertainties

$$u_{\text{rel}}(F) = \frac{u_F}{F}, \quad u_{\text{rel}}(A) = \frac{u_A}{A}.$$

On the other hand, from expression (5.7),

$$u_{\text{cor}}^2(E(\text{CO}_2\text{e})) = 6s^2A^2u_F^2 + 18F^2s^2u_A^2 = s^2A^2F^2 [6u_{\text{rel}}^2(F) + 18u_{\text{rel}}^2(A)]. \quad (5.11)$$

The quotient of the terms in square brackets in expressions (5.10) and (5.11), namely,

$$\lambda = \frac{u_{\text{cor}}^2(E(\text{CO}_2\text{e}))}{u_{\text{ind}}^2(E(\text{CO}_2\text{e}))} = \frac{6u_{\text{rel}}^2(F) + 18u_{\text{rel}}^2(A)}{6u_{\text{rel}}^2(F) + 6u_{\text{rel}}^2(A)} = \frac{u_{\text{rel}}^2(F) + 3u_{\text{rel}}^2(A)}{u_{\text{rel}}^2(F) + u_{\text{rel}}^2(A)}, \quad (5.12)$$

is informative regarding the respective contributions from the emissions' factor and the activity data under the above correlation and independence assumptions. Examining expression (5.12), for an activity data relative uncertainty that is small compared with that for the emissions' factor, 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, $u_{\text{cor}}(E(\text{CO}_2\text{e}))$ is $\sqrt{3} \approx 1.7$ times $u_{\text{ind}}(E(\text{CO}_2\text{e}))$, a 70 % increase. It is emphasized that such an increase is optimistic. For such cases, where better balanced data in terms of the values of the quantities involved is available, the expected increase is bounded by 70 %.

5.5 Interpretation of results

5.5.1 Relative expanded uncertainties of 100 % or more are not uncommon in emissions' inventories, which is why it is important that these large uncertainty sources are carefully taken into consideration. The UK introduction to GHG inventories [108], shows a highly asymmetric interval for inventory N_2O figures, in that emissions' estimates for N_2O are far from the centre of their 95 % coverage intervals, in its figure 5.2. The implication is that the probability distributions for N_2O are strongly asymmetric, but in the previous clause (5.1) of that document uncertainties are expressed as 95 % coverage intervals (for example, table 5.1), with no indication that they may be asymmetric.

5.5.2 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 explain the figure. After some investigation, these aspects seem to be based on methodology described in Milne et al. [90]. Figure 2 in that paper shows the distribution for the emissions' factors given by Monte Carlo modelling, with asymmetry visually evident. Log-normal probability distributions are used to represent knowledge of such quantities.

5.5.3 Reference [90] is focused on agricultural emissions (that is, manure, enteric fermentation and field burning), table 5.1 in this document being based on non-road vehicle emissions (that is, from tractors, combine harvesters, etc.), which burn gas-oil and diesel. Meagre information has been found by the authors of this chapter on the probability distributions for the quantities under consideration here. It is emphasized that if the PDF is changed for one (or more) of the input quantities, provided it has the same standard uncertainty, the standard uncertainty associated with the estimate of the measurand will be unchanged using JCGM 100 [10]. The same is not true for PDF propagation using JCGM 101 [17].

5.5.4 The concentration here is on an agricultural subsector for which sufficient publicly available data allowed correlation evaluation to be carried out. While correlation effects appear to be more significant in other agricultural subsectors, the necessary raw data to quantify correlation were not accessible. In contrast, in the area of Waste Incineration (IPCC Category 5C), it was reported in 5.4.11 that accounting for correlation led to a substantial increase — approximately 24 % — in the standard uncertainty for CO_2 emissions obtained under an assumption of independence. It is understood that supporting data for this finding will be made available in due course.

5.6 Uncertainty guidance promoted by the IPCC and the JCGM

5.6.1 In terms of regarding expression (5.3) as the measurement model, JCGM 100:2008 [10] refers to the F_i and A_i as input quantities and $E(CO_2e)$ as the output quantity or measurand. In the model (5.3), the expression on the right-hand side constitutes the *measurement function*, with sensitivity coefficients given by the first-order partial derivatives of that function. IPCC documents seem not to emphasize that propagation of uncertainty should be carried out in the context of a *measurement model*. It is only by using measurement models as a basis, that measurement uncertainty evaluation can properly be conducted.

5.6.2 The IPCC 2000 Good Practice Guidance [112] and the 2006 Guidelines [66] set out the ‘error propagation approach’ referred to as the LPU in the legacy GUM [10]. For example, the LPU occurs specifically in [65, clause 6.3, formula (6.3)].

5.6.3 The attitude of the JCGM is that the Monte Carlo method (MCM), as described in JCGM 101:2008 and JCGM 102:2011 [12, 17], is better suited to handle situations where the uncertainties are large and the measurement model is nonlinear. In contrast, the error propagation approach (the LPU in JCGM 100:2008 [10]) can produce invalid uncertainty statements for such models. Thus, MCM can be used as a reference against which other methods can be assessed, such as evaluation using LPU, which requires linearization of the model. This way of thinking is very different from that of the IPCC, which regards the error approach as a reference. Indeed, MCM is seen, at least by some Bayesian statisticians, as a ‘gold standard’ for uncertainty propagation: Huggins et al. [63] 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 [41, 61, 101, 113]) to the Monte Carlo method as a ‘gold standard’ for uncertainty propagation. In addition, non-pathological examples are well documented to show disagreement of the results produced by the two approaches. One example in mass calibration, from JCGM 101:2008 [17], shows that the LPU can understate the standard uncertainty by some 40% compared to the MCM.

5.6.4 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’.

JCGM 101:2008 provides a different view on the matter [17]:

‘Whereas there are some limitations to the JCGM 100:2008 uncertainty framework (GUF), 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.’

A comparison of MCM and the GUF 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 (5.3), which is nonlinear.

6 Simple linear measurement models

6.1 Preamble

6.1.1 This example is provided by the JCGM-WG1.

6.1.2 The primary aim of this example is to quantify how a change of probability density function (PDF) for an input quantity — whilst retaining its mean and standard deviation — affects the standard uncertainty associated with an estimate of the measurand.

6.1.3 The evaluation of uncertainty based on the simple linear measurement model

$$Y = X_1 + \cdots + X_N. \quad (6.1)$$

when $N = 4$ is considered. The propagation of measurement uncertainty and the determination of coverage intervals using the provisions of the JCGM 100:2008 uncertainty framework (GUF) and the Monte Carlo method (MCM) are shown. MCM is used to validate the application of JCGM 100:2008. Standard uncertainties are reported to three significant decimal digits to facilitate their comparison.

6.1.4 A specific case of the generic linear model is considered, in which the input quantities X_i are independent and dimensionless. Three different sets of PDFs $g_{x_i}(\xi_i)$ for them are discussed.

In the first set, each $g_{x_i}(\xi_i)$ is a standard normal PDF having expectation zero and standard deviation unity. In the second set, each $g_{x_i}(\xi_i)$ is a rectangular PDF, also having expectation zero and standard deviation unity. The third set is identical to the second except that the PDF for $g_{x_4}(\xi_4)$ has a standard deviation of 10.

NOTE Further information concerning linear models, such as the model (6.1), where the PDFs are normal or rectangular or a combination of both, is available [47].

6.2 Normally distributed input quantities

6.2.1 A standard normal PDF is assigned to each X_i . Thus, the estimates of the X_i are $x_i = 0$, $i = 1, 2, 3, 4$, with associated standard uncertainties $u(x_i) = 1$.

6.2.2 The law of propagation of uncertainty (LPU) [10, clause 5.1.2] gives the standard uncertainty $u(y) = 2.00$ associated with the estimate $y = 0.00$ of Y , using a numerical tolerance of two significant decimal digits for $u(y)$ [17, clause 7.9.2]. A probabilistically symmetric 95% coverage interval for Y is $[-3.92, 3.92]$ based on a coverage factor of $k = 1.96$. These results are given in the first (non-header) row of table 6.1.

6.2.3 The application of MCM [17, clause 7] with $M = 10^5$ trials and a numerical tolerance of $\delta = 0.05$ gives the estimate $y = 0.00$, its standard uncertainty $u(y) = 2.00$ and the probabilistically symmetric 95% coverage interval $[-3.92, 3.92]$. See table 6.1, row 2. Two further applications of the method, with $M = 10^6$ trials, agree with these results to within the numerical tolerance used. See table 6.1, rows 3 and 4. These two further applications (different random samplings being taken from the PDFs) were made to demonstrate variation in the results obtained.

6.2.4 Two applications of the adaptive Monte Carlo procedure [17, clause 7.9] with the use of a numerical tolerance of $\delta/5$ as recommended in [17, clause 8.2] were then made, taking 1.23×10^6 and 1.02×10^6 trials. See table 6.1, rows 5 and 6.

6.2.5 The PDF for Y obtained analytically is the normal PDF with expectation zero and standard deviation two. See table 6.1, bottom row.

Table 6.1: Uncertainty evaluation for the linear model (6.1) with a standard normal PDF assigned to each input quantity X_i

Method	M	y	$u(y)$	Prob. sym. 95 % cov. int.	d_{lo}	d_{hi}	GUF validated ($\delta = 0.05$)?
GUF		0.00	2.00	[-3.92, 3.92]			
MCM	10^5	0.00	2.00	[-3.92, 3.92]			
MCM	10^6	0.00	2.00	[-3.92, 3.92]			
MCM	10^6	0.00	2.00	[-3.92, 3.92]			
Adaptive MCM	1.23×10^6	0.00	2.00	[-3.92, 3.93]	0.00	0.01	Yes
Adaptive MCM	1.02×10^6	0.00	2.00	[-3.92, 3.92]	0.00	0.00	Yes
Analytical		0.00	2.00	[-3.92, 3.92]			

6.2.6 Figure 6.1 shows the (normal) PDF for Y resulting from the GUF. This PDF coincides with the analytical solution, since the Central Limit Theorem (CLT) holds exactly. The figure also shows one of the approximations (density histogram of $M = 10^6$ values of Y) constituting a discrete representation of this PDF provided by MCM [17, clause 7.5]. The endpoints of the probabilistically symmetric 95 % coverage interval provided by both methods are shown as vertical lines. The PDF and the approximation are visually indistinguishable, as are the respective coverage intervals.

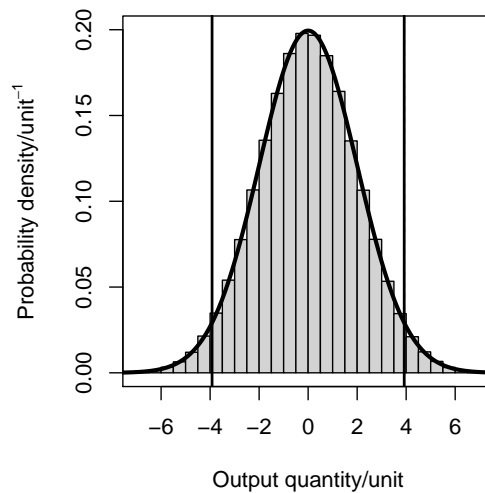


Figure 6.1: Approximations for the model (6.1), with each X_i assigned a standard normal PDF, to the PDF for Y provided by (a) the GUF (black line), (b) MCM (density histogram, clause 6.2.6). The analytical solution here coincides with the GUF solution. “Unit” denotes any unit

The probabilistically symmetric 95 % coverage interval is determined. This step is taken

since the general linear measurement model (6.1) with independent input quantities has the property that the PDF for Y is symmetric when the PDFs for the X_i are symmetric. This property is proved as follows. For random variables X_1 and X_2 having PDFs that are symmetric about zero, their sum $Y = X_1 + X_2$ is also symmetric about zero. To establish this result (a more complicated proof appears in [106]), g_Y is given by the convolution integral

$$g_Y(\eta) = \int_{-\infty}^{\infty} g_{X_1}(\eta - \xi_2) g_{X_2}(\xi_2) d\xi_2. \quad (6.2)$$

To prove that $g_Y(\eta) = g_Y(-\eta)$ for expression (6.2), using symmetry, negate the arguments of g_{X_1} and g_{X_2} , and set $\tau = -\xi_2$. Comparison with the right side of (6.2) proves the result in this case. Observing that the result also applies for symmetry about nonzero expectations, by induction the result holds for the more general model (6.1).

6.2.7 The validation procedures of [17, Clauses 8.1, 8.2] were applied. Using the terminology of [17, clause 7.9.2], $n_{\text{dig}} = 2$, since two significant decimal digits in $u(y)$ are sought. Hence, $u(y) = 2.0 = 20 \times 10^{-1}$, and so $c = 20$ and $\ell = -1$ in [17, clause 7.9.2, item a)]. Thus, according to that clause, the numerical tolerance is $\delta = \frac{1}{2} \times 10^{-1} = 0.05$. The magnitudes d_{lo} and d_{hi} of the endpoint differences [17, equations. (19), (20)] are shown in table 6.1, columns 6 and 7, for the two applications of adaptive MCM. Whether the GUF has been validated for $\delta = 0.05$ is shown in column 8.

6.3 Rectangularly distributed input quantities with the same width

6.3.1 Assign a rectangular PDF to each X_i , so that X_i has an expectation of zero and a standard deviation of unity. The estimates of the X_i are $x_i = 0$ ($i = 1, 2, 3, 4$), with associated standard uncertainties $u(x_i) = 1$. By following the analogous steps to those in clause 6.2, the results in table 6.2 were obtained. It is noted that adaptive MCM validates the GUM uncertainty framework in one instance but not in another. This divergence reflects the inherent variability of stochastic simulation and underscores that complete assurance in any single Monte Carlo estimate cannot be guaranteed. See 6.2.3.

Table 6.2: Uncertainty evaluation for the linear model (6.1), with each input quantity X_i assigned identical rectangular PDFs

Method	M	y	$u(y)$	Prob. sym. 95% cov. int.	d_{lo}	d_{hi}	GUF validated ($\delta = 0.05$)?
GUF		0.00	2.00	[-3.92, 3.92]			
MCM	10^5	0.00	2.01	[-3.90, 3.89]			
MCM	10^6	0.00	2.00	[-3.89, 3.88]			
MCM	10^6	0.00	2.00	[-3.88, 3.88]			
Adaptive MCM	1.02×10^6	0.00	2.00	[-3.88, 3.89]	0.04	0.03	Yes
Adaptive MCM	0.86×10^6	0.00	2.00	[-3.87, 3.87]	0.05	0.05	No
Analytical		0.00	2.00	[-3.88, 3.88]			

The analytic solution for the endpoints of the probabilistically symmetric 95% coverage interval, namely, $\pm 2\sqrt{3}[2 - (3/5)^{1/4}] \approx \pm 3.88$, is obtained as described below.

The PDF of a rectangular distribution $R(a, b)$ takes the constant value $(b - a)^{-1}$ for $a \leq \xi \leq b$ and is zero otherwise.

The n -fold convolution of $R(0, 1)$ is the B-spline $B_n(x)$ of order n (degree $n - 1$) with knots $0, \dots, n$ [114] whose explicit expression is [33]

$$B_n(x) = \frac{1}{(n-1)!} \sum_{r=0}^n {}^n C_r (-1)^r (x-r)_+^{n-1}, \quad (6.3)$$

where ${}^n C_r = n!/[r!(n-r)!]$ and $z_+ = \max(z, 0)$. A piecewise cubic-polynomial representation of $B_4(x)$ obtained from expression (6.3) or by applying the B-spline recurrence relation in [33] is

$$B_4(x) = \frac{1}{6} \begin{cases} x^3, & 0 \leq x < 1, \\ 4 - 12x + 12x^2 - 3x^3, & 1 \leq x < 2, \\ -44 + 60x - 24x^2 + 3x^3, & 2 \leq x < 3, \\ (4-x)^3, & 3 \leq x \leq 4, \\ 0, & \text{otherwise.} \end{cases} \quad (6.4)$$

The left-hand endpoint y_{10} of the probabilistically symmetric 95 % coverage interval lies between zero and one, because

$$\int_0^1 B_4(x) dx = \left[\frac{1}{24} x^4 \right]_0^1 = \frac{1}{24},$$

and $0.025 = \frac{1}{40} < \frac{1}{24}$. The area under the PDF that lies to the left of y_{10} is given by

$$\int_0^{d_L} B_4(x) dx = \frac{1}{24} d_L^4 = \frac{1}{40}.$$

That is, $d_L = (3/5)^{1/4}$. By symmetry, the right-hand endpoint is $d_R = 4 - (3/5)^{1/4}$. Thus, the probabilistically symmetric 95 % coverage interval is

$$\left[(3/5)^{1/4}, 4 - (3/5)^{1/4} \right] \quad \text{or} \quad 2 \pm (2 - (3/5)^{1/4}).$$

The corresponding coverage interval for the four-fold convolution of the rectangular PDF $R(-\sqrt{3}, \sqrt{3})$ (which has zero expectation and unit standard deviation) is given by shifting this result by two units and scaling it by $2\sqrt{3}$ units.

6.3.2 Figure 6.2 shows the counterpart of figure 6.1 in this case, again based on $M = 10^6$ trials and a numerical tolerance of $\delta = 0.05$. By comparison with figure 6.1, some modest differences between the approximations to the PDFs can be seen. Since normality of the measurand is assumed, the GUF provides exactly the same PDF for Y whether the PDFs for the X_i are normal or rectangular, because the expectations of these quantities are identical, as are the standard deviations, in the two cases. The probabilistically symmetric 95 % coverage interval determined on the basis of the GUF is in this case slightly more conservative than that obtained analytically. The reason for this behaviour is that, although the four-fold convolution of a uniform PDF closely approximates a normal PDF, the coverage factor to be used to construct a 95 % coverage interval is still slightly smaller than that for a normal (see also [10, annex G.2.2]).

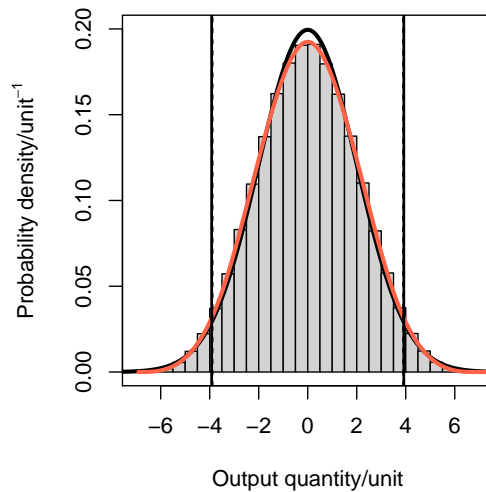


Figure 6.2: Counterpart of figure 6.1 for quantities having the same expectations and standard deviations (clause 6.3.2), but distributed as rectangular PDFs. The black curve corresponds to GUF, the scaled histogram represents the MCM result and the red curve (close to the histogram) is the analytic solution

6.3.3 As for normally distributed quantities, the validation procedure was applied (columns 6–8 of table 6.2). As before, $n_{\text{dig}} = 2$, $u(y) = 20 \times 10^{-1}$, $c = 20$, $\ell = -1$ and $\delta = 0.05$. The endpoint differences d_{lo} and d_{hi} are larger than for the case of normally distributed quantities (table 6.1). For the first of the two applications of the adaptive Monte Carlo method, the GUF is validated. For the second application, it is not validated, although d_{lo} and d_{hi} for this application are close to the numerical tolerance $\delta = 0.05$ (seen if more decimal digits than in table 6.2 are considered). Different validation results such as these are an occasional consequence of the stochastic nature of the Monte Carlo method, especially in a case such as that here.

6.4 Rectangularly distributed input quantities with different widths

6.4.1 Consider the example of clause 6.3, except that X_4 has a standard deviation of ten rather than unity. Table 6.3 contains the results obtained.

6.4.2 The numbers M of Monte Carlo trials taken by the adaptive procedure (0.03×10^6 and 0.08×10^6) are much smaller than they were for the two previous cases in this example. The main reason is that, in this case, $\delta = 0.5$, the numerical tolerance resulting from requesting, as before, two significant decimal digits in $u(y)$, is ten times the previous value. Were the previous value to be used, M would be of the order of 100 times greater.

6.4.3 Figure 6.3 shows the (normal) PDF for Y (black curve) resulting from the GUF. It also shows a (scaled) histogram obtained from the MCM results. They differ appreciably owing to the dominance of the PDF for X_4 . The PDF for Y resembles that for X_4 , but there is an effect in the flanks resulting from the PDFs for the other X_i . The PDF for Y , closely resembling the histogram, is also shown as the red curve. It is based on an analytical solution for the distribution of the sum of n non-identically distributed uniform random variables given in [92]. The probabilistically symmetric 95 % coverage interval

Table 6.3: As table 6.2, except that the fourth input quantity has a standard deviation of ten rather than unity, and no analytic solution is provided (clause 6.4.1)

Method	M	y	$u(y)$	Prob. sym. 95 % cov. int.	d_{lo}	d_{hi}	GUF validated ($\delta = 0.5$)?
GUF		0.0	10.1	[-19.9, 19.9]			
MCM	10^5	0.0	10.2	[-17.0, 17.0]			
MCM	10^6	0.0	10.2	[-17.0, 17.0]			
MCM	10^6	0.0	10.1	[-17.0, 17.0]			
Adaptive MCM	0.03×10^6	0.1	10.2	[-17.1, 17.1]	2.8	2.8	No
Adaptive MCM	0.08×10^6	0.0	10.1	[-17.0, 17.0]	2.9	2.9	No

determined on the basis of the GUF in this case is more conservative than that obtained using MCM. Again, the reason is the dominance of the uniform PDF for X_4 , so that the PDF for Y is quasi-uniform. As it is well known, a uniform is ‘narrower’ than a normal (see also [10, G.1.3, Note]).

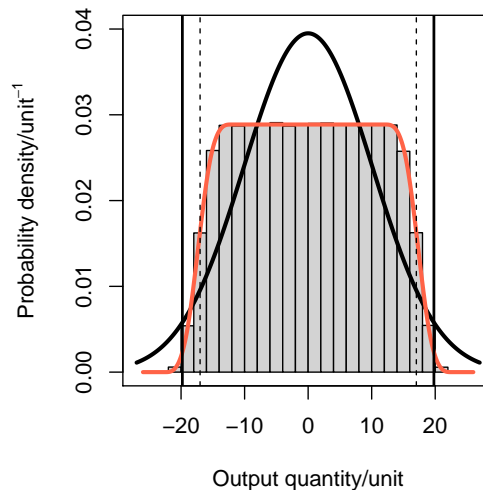


Figure 6.3: Counterpart of figure 6.2, except that the fourth input quantity (X_4) has a standard deviation of ten rather than unity (clause 6.4.3). The black, bell-shaped curve corresponds to GUF, the (scaled) histogram represents the MCM result, and the red line is the analytical solution. Full and broken vertical lines denote the limits of the 95 % coverage intervals for the GUF and analytical solution, respectively

6.4.4 The validation procedure was applied (columns 6–8 of table 6.3). Now, $n_{\text{dig}} = 2$, $u(y) = 10 \times 10^0$, $c = 10$, $\ell = 0$ and $\delta = 1/2 \times 10^0 = 0.5$. For the two applications of the adaptive Monte Carlo procedure, the GUF is not validated. For a numerical tolerance of one significant decimal digit in $u(y)$, i.e. $n_{\text{dig}} = 1$, for which $\delta = 5$, the validation status would be positive in both cases, the 95 % coverage intervals all being $[-2 \times 10^1, 2 \times 10^1]$. The conditions for the CLT to apply are not met in this circumstance [10, clause G.6.5], because of the dominating effect of the rectangular PDF for X_4 . Therefore, application of the GUF provides reliable estimate and associated standard uncertainty whereas confidence intervals (including expanded uncertainty) are inaccurate (conservative in this case). Consideration should thus be given to using MCM (see [17, clause 8.1.1]).

7 Second-order effects in a nonlinear measurement model: calibration of weights

7.1 Preamble

7.1.1 This example is provided by the JCGM-WG1.

7.1.2 This example considers a nonlinear measurement model — of a mass calibration — and compares three ways of obtaining the measurement result: the JCGM 100:2008 uncertainty framework (GUF) considering only first-order terms (GUF₁), the GUF considering both first- and second-order terms (GUF₂), and the Monte Carlo method (MCM).

7.2 Measurement model

7.2.1 Consider the calibration of a weight W, whose density is ρ_W , against a reference weight R, whose density is ρ_R . The calibration is performed using a balance operating in air whose density is ρ_a . Both weights have the same nominal mass $m_{\text{nom}} = 100$ g. Since ρ_W and ρ_R are generally different, it is necessary to account for buoyancy effects [97]. This example appeared originally in [17, clause 9.3].

7.2.2 Applying Archimedes' principle, the measurement model takes the form

$$m_W(1 - \rho_a/\rho_W) = (m_R + \delta m_R)(1 - \rho_a/\rho_R), \quad (7.1)$$

where δm_R is the mass of a small weight of density ρ_R added to R to balance it with W.

7.2.3 The conventional mass $m_{W,c}$ of W is the mass of a (hypothetical) weight of density $\rho_0 = 8000$ kg/m³ that balances W in air at density $\rho_{a_0} = 1.2$ kg/m³ (these assigned value have no associated uncertainty). Thus,

$$m_W(1 - \rho_{a_0}/\rho_W) = m_{W,c}(1 - \rho_{a_0}/\rho_0).$$

7.2.4 In terms of conventional masses $m_{W,c}$, $m_{R,c}$ and $\delta m_{R,c}$, model (7.1) becomes

$$m_{W,c} \frac{1 - \rho_a/\rho_W}{1 - \rho_{a_0}/\rho_W} = (m_{R,c} + \delta m_{R,c}) \frac{1 - \rho_a/\rho_R}{1 - \rho_{a_0}/\rho_R} \quad (7.2)$$

or, to an approximation adequate for most practical purposes ($\rho_W, \rho_R \gg \rho_a$) [97],

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

from which, introducing $\delta m = m_{W,c} - m_{\text{nom}}$, the measurement model used in this example is eventually obtained:

$$\delta m = (m_{R,c} + \delta m_{R,c}) \left[1 + (\rho_a - \rho_{a_0}) \left(\frac{1}{\rho_W} - \frac{1}{\rho_R} \right) \right] - m_{\text{nom}}. \quad (7.3)$$

7.2.5 The only information concerning $m_{R,c}$ and $\delta m_{R,c}$ is an estimate and an associated standard uncertainty for each of these quantities. Accordingly, normal distributions are assigned to them, with estimates and standard uncertainties as expectations and standard deviations, respectively [17, clause 6.4.7.1]. As for quantities ρ_a , ρ_W and ρ_R , only lower and upper limits are known for each of these quantities. Accordingly, rectangular distributions are assigned to these quantities, with endpoints equal to the known limits [17, clause 6.4.2.1]. Table 7.1 summarizes the input quantities and their probability density functions (PDFs). In the table, a normal distribution $N(\mu, \sigma^2)$ is described in terms of expectation μ and standard deviation σ , and a rectangular distribution $R(a, b)$ with endpoints a and b ($a < b$) in terms of expectation $(a + b)/2$ and half-width $(b - a)/2$ or standard deviation $(b - a)/\sqrt{12}$.

Table 7.1: Input quantities and PDFs assigned to them for the mass calibration measurement model (7.3)

Quantity	Unit	Distribution	Distribution parameters		
			Expectation	Std. dev.	Half-width
$m_{R,c}$	mg	$N(\mu, \sigma^2)$	100 000.000	0.050	
$\delta m_{R,c}$	mg	$N(\mu, \sigma^2)$	1.234	0.020	
ρ_a	kg m^{-3}	$R(a, b)$	1.20	$0.10/\sqrt{3}$	0.10
ρ_W	kg m^{-3}	$R(a, b)$	8000	$1000/\sqrt{3}$	1000
ρ_R	kg m^{-3}	$R(a, b)$	8000.00	$0.05/\sqrt{3}$	0.05

7.3 Measurand estimate and associated uncertainty

7.3.1 The GUF that considers only first-order terms (GUF_1), applicable to uncorrelated input quantities, corresponds to the following evaluations:

$$y_{\text{GUF}_1} = f(\mathbf{x}), \quad (7.4)$$

$$u^2(y_{\text{GUF}_1}) = \sum_{i=1}^N \left(\frac{\partial f}{\partial X_i} \right)^2 u^2(x_i), \quad (7.5)$$

where $\mathbf{x} = [x_1, \dots, x_N]^T$ and $y = f(\mathbf{x})$ denote the vector of estimates of the input quantities X_i , $i = 1, \dots, N$, and the corresponding estimate of the measurand Y , respectively. Partial derivatives are evaluated at the estimates of the input quantities.

7.3.2 The GUF that considers first- and second-order terms (GUF_2), applicable to uncorrelated and normally distributed input quantities, corresponds to the following evaluations:

$$y_{\text{GUF}_2} = y_{\text{GUF}_1} + \frac{1}{2} \sum_{i=1}^N \frac{\partial^2 f}{\partial X_i^2} u^2(x_i), \quad (7.6)$$

$$u^2(y_{\text{GUF}_2}) = u^2(y_{\text{GUF}_1}) + \sum_{i=1}^N \sum_{j=1}^N \left[\frac{1}{2} \left(\frac{\partial^2 f}{\partial X_i \partial X_j} \right)^2 + \frac{\partial f}{\partial X_i} \frac{\partial^3 f}{\partial X_i \partial X_j^2} \right] u^2(x_i) u^2(x_j). \quad (7.7)$$

Formulae (7.6) and (7.7) form a self-consistent pair since, as stated in [85], they are both based on the third-order Taylor expansion of the measurement function $f(\mathbf{X})$. The use of the estimate (7.4) in conjunction with formula (7.7), as in [10, clause 5.1.2], represents inconsistent working, since it is biased, not taking higher-order terms into consideration.

7.3.3 The GUF and the adaptive MCM [17, clause 7.9] were each used to obtain an estimate $\widehat{\delta m}$ of δm , the associated standard uncertainty $u(\widehat{\delta m})$, and the shortest 95 % coverage interval for δm . The results obtained are shown in table 7.2. There, the GUF₁ and GUF₂ estimates of δm are equal since the partial derivative for the second-order effect is zero in this instance.

Table 7.2: Results of the calculation stage for the mass calibration model (7.3), δm

Method	$\widehat{\delta m}$ /mg	$u(\widehat{\delta m})$ /mg	Shortest 95 % coverage interval/mg	GUF validated ($\delta = 0.005$)?
GUF ₁	1.2340	0.0539	[1.1285, 1.3395]	No
MCM	1.2340	0.0755	[1.0843, 1.3838]	
GUF ₂	1.2340	0.0750	[1.0870, 1.3810]	Yes

7.3.4 A total of 0.72×10^6 trials were taken by the adaptive MCM [17, subclause 7.9] with the use of a numerical tolerance of $\delta/5$ [17, subclause 8.2] with δ set for the case where one significant decimal digit in $u(\widehat{\delta m})$ is regarded as meaningful.

7.3.5 Figure 7.1 shows the PDFs for δm obtained from GUF₁ (sharply peaked curve), GUF₂ (more rounded curve) and MCM (scaled histogram). The inner pair of (broken) vertical lines indicates the endpoints of the shortest 95 % coverage interval for δm based on GUF₁. The outer pair of (continuous) vertical lines indicates the endpoints of the shortest 95 % coverage interval for δm based on the discrete representation of the distribution function determined as in [17, subclause 7.5]. The endpoints of the shortest 95 % coverage interval for δm based on GUF₂ essentially coincide with those provided by MCM.

7.3.6 Although GUF₁ and MCM both give estimates of δm in good agreement, the associated standard uncertainties are noticeably different. The value of $u(\widehat{\delta m})$ returned by either MCM or GUF₂ is 40 % larger than that returned by GUF₁. The latter method is therefore understating the uncertainty associated with $\widehat{\delta m}$.

7.3.7 Table 7.2 also shows in the right-most two columns the results of applying the JCGM 101:2008 validation procedure [17, clause 8] in the case where one significant decimal digit in $u(\widehat{\delta m})$ is regarded as meaningful. Since a numerical tolerance of one significant decimal digit in $u(\widehat{\delta m})$ is required, $n_{\text{dig}} = 1$. Hence, $u(\widehat{\delta m}) = 0.08 \text{ mg} = 8 \times 10^{-2} \text{ mg}$, and so the c in [17, subclause 7.9 item a)] equals 8 and $\ell = -2$. Thus $\delta = 1/2 \times 10^{-2} = 0.005$. Whether the results were validated to one significant decimal digit in $u(\widehat{\delta m})$ is indicated in the final column of the table. Thus, if only first-order terms are considered, the application of GUF provides inadequate results. If, however, higher-order terms are considered [10, clause 5.1.2 note], the GUF becomes valid. Thus, the non-linearity of the model is such that accounting only for first-order terms is inadequate and higher-order terms cannot be neglected.

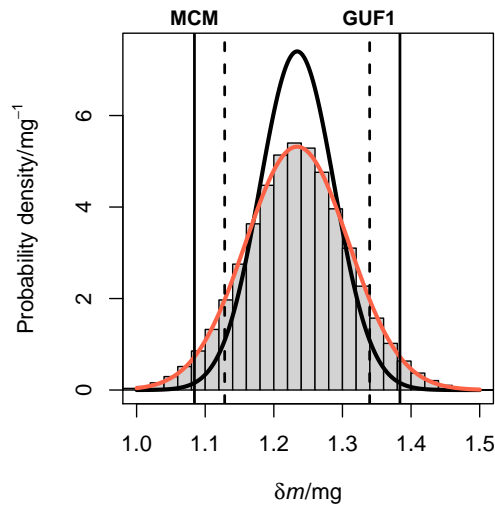


Figure 7.1: Approximations to the PDF for the output quantity δm obtained using GUF₁ (sharply peaked curve), GUF₂ (more rounded curve) and MCM scaled histogram, sub-clause 7.3.5)

7.4 Uncertainty budget

7.4.1 GUF

In many measurement contexts it is common practice to list the uncertainty components $u_i(y) = |c_i|u(x_i)$, $i = 1, \dots, N$, where c_i is the i th sensitivity coefficient and $u(x_i)$ the standard uncertainty associated with the i th input estimate x_i , contributing to the standard uncertainty $u(y)$. Usually these are presented in a table (such as table 7.3), the ‘uncertainty budget’. This practice may be useful to identify the dominant terms contributing to $u(y)$ associated with the estimate of the output quantity.

Table 7.3 lists the partial derivatives of first order for the model (7.3) with respect to the input quantities together with the sensitivity coefficients, namely, these derivatives evaluated at the estimates of the input quantities.

Table 7.3: Uncertainty budget – GUF₁

Input quantity X_i	Std. unc. $u(x_i)$	Partial derivative $c_i = \partial Y / \partial X_i$	Rel. contrib./% $c_i^2 u^2(x_i) / u^2(y)$
$m_{R,c}$	0.050 mg	$1 + (\rho_a - \rho_{a_0})(1/\rho_W - 1/\rho_R) = 1$	86
$\delta m_{R,c}$	0.020 mg	$1 + (\rho_a - \rho_{a_0})(1/\rho_W - 1/\rho_R) = 1$	14
ρ_a	0.058 kg m^{-3}	$(m_{R,c} + \delta m_{R,c})(1/\rho_W - 1/\rho_R) = 0 \text{ m}^3$	0
ρ_W	580 kg m^{-3}	$-(m_{R,c} + \delta m_{R,c})(\rho_a - \rho_{a_0})/\rho_W^2 = 0 \text{ m}^3$	0
ρ_R	0.029 kg m^{-3}	$(m_{R,c} + \delta m_{R,c})(\rho_a - \rho_{a_0})/\rho_R^2 = 0 \text{ m}^3$	0
$u(y) = 0.0539 \text{ mg}$			100

Since the uncertainty associated with the output quantity is the sum of squared individual uncertainty contributions, $u_i(y)$, the squared uncertainties of the two input quantities $m_{R,c}$ and $\delta m_{R,c}$ contribute to the squared uncertainty $u^2(y)$ in the respective proportions $0.050^2 / (0.050^2 + 0.020^2) = 86\%$ and $0.020^2 / (0.050^2 + 0.020^2) = 14\%$.

The uncertainty budget can also include higher-order uncertainty contributions in accordance with expression (7.7). Table 7.4 shows that the interaction of variables ρ_a and ρ_W has significant impact on the overall uncertainty estimate $u(y)$.

Table 7.4: Uncertainty budget – GUF₂ (only terms having non-zero partial derivatives are shown)

Input quantity X_i	Partial derivative $c_i = \partial Y / \partial X_i$	Rel. contrib./% $c_i^2 u^2(x_i) / u^2(y)$
$m_{R,c}$	$1 + (\rho_a - \rho_{a_0})(1/\rho_W - 1/\rho_R) = 1$	44
$\delta m_{R,c}$	$1 + (\rho_a - \rho_{a_0})(1/\rho_W - 1/\rho_R) = 1$	7
Input quantities X_i, X_j	Partial derivative $c_{ij} = \partial^2 Y / (\partial X_i \partial X_j)$	Rel. contrib./% $c_{ij}^2 u^2(x_i) u^2(x_j) / u^2(y)$
ρ_a, ρ_W	$-(m_{R,c} + \delta m_{R,c}) / \rho_W^2 = -1.56 \times 10^{-9} \text{ m}^6 \text{ kg}^{-1}$	49
ρ_a, ρ_R	$+(m_{R,c} + \delta m_{R,c}) / \rho_R^2 = +1.56 \times 10^{-9} \text{ m}^6 \text{ kg}^{-1}$	0
$u(y) = 0.0750 \text{ mg}$		100

In cases for which a valid implementation of the propagation of distributions is more appropriate, an uncertainty budget should be regarded as a qualitative tool.

7.4.2 MCM

MCM, contrary to the GUF, does not require sensitivity coefficients to obtain the measurement result. Therefore, their calculation represents a separate task in this approach. A procedure is given in [17, Annex B].

8 Gauge block calibration

8.1 Preamble

8.1.1 This example is provided by the JCGM-WG1.

8.1.2 The example shows the uncertainty evaluation for the calibration of a gauge block using a reference standard by applying the JCGM 100:2008 uncertainty framework (GUF) [10, clause 5] and the Monte Carlo method (MCM) [17, clause 7]. The example starts with the transformation of a model for the difference between the length of the blocks, the quantity actually measured, into an explicit measurement model. Since some of the input quantities in that model are correlated, a re-parametrization [15, clause 8.3.2] is used to remove that correlation. Probability density functions (PDFs) are assigned to the transformed quantities based on the metrologist's knowledge. Using those PDFs, GUF and MCM are applied and the results discussed.

8.2 Measurement model

8.2.1 The length of a gauge block is determined by comparing it with a known reference standard having the same nominal length $L_{\text{nom}} = 50$ mm. The direct output of the comparison of two gauge blocks is the difference d in their lengths given by

$$d = L(1 + \alpha\theta) - L_s(1 + \alpha_s\theta_s), \quad (8.1)$$

where

L denotes the length (at 20 °C temperature [68]) of the gauge block being calibrated,

L_s denotes the length of the reference standard, also at 20 °C, as given in its calibration certificate,

α and α_s are the coefficients of thermal expansion for the two blocks,

θ and θ_s are the deviations in temperature of the two gauge blocks from the 20 °C reference temperature.

(JCGM 100:2008 refers to a gauge block as an end gauge and the symbol L for the length of a gauge block is used in this document in place of the symbol l used in JCGM 100:2008 and in ISO 3650:1998 for that quantity.)

8.2.2 From expression (8.1), the output quantity L is given by

$$L = \frac{L_s(1 + \alpha_s\theta_s) + d}{1 + \alpha\theta}. \quad (8.2)$$

By using $1/(1 + x) \approx 1 - x$ when $|x|$ is small compared with unity, model (8.2) becomes

$$L = L_s + d + L_s(\alpha_s\theta_s - \alpha\theta) - d\alpha\theta - L_s\alpha_s\theta_s\alpha\theta. \quad (8.3)$$

The term $L_s\alpha_s\theta_s\alpha\theta$ is a second-order term and can be safely neglected. The term $d\alpha\theta$ is the thermal-expansion correction for the difference d between the two gauge blocks. Even considering an extreme situation (say, $d = 1$ μm and $\theta = 3$ °C), both the value and

the associated standard uncertainty lie well below a negligible 0.1 nm. Model (8.3) thus reduces to

$$L = L_s + d + L_s(\alpha_s \theta_s - \alpha \theta), \quad (8.4)$$

an approximation adequate for most practical purposes.

If the difference in temperature between the gauge block being calibrated and the reference standard is written as $\delta\theta = \theta - \theta_s$ and the difference in their thermal expansion coefficients as $\delta\alpha = \alpha - \alpha_s$, model (8.4) becomes

$$L = L_s + d - L_s(\delta\alpha\theta + \delta\theta\alpha_s). \quad (8.5)$$

This re-parametrization [15, clause 8.3.2] avoids the introduction of covariances between θ and θ_s and between α and α_s .

8.2.3 The difference d in the lengths of the gauge block being calibrated and the reference standard is determined as the average of a series of five indications, obtained independently, of the difference using a calibrated comparator. The difference d can be expressed as

$$d = D + d_1 + d_2, \quad (8.6)$$

where D denotes the observed length difference, and d_1 and d_2 are additive corrections modelling, respectively, the random and systematic effects associated with using the comparator [15, clause 10.4.3].

8.2.4 The quantity θ , representing deviation of the temperature from 20 °C of the gauge block being calibrated, can be expressed as

$$\theta = \theta_0 + \Delta, \quad (8.7)$$

where θ_0 represents the average temperature deviation of the gauge block from 20 °C and Δ describes a cyclic variation of the temperature deviation from θ_0 .

8.2.5 Substituting expressions (8.6) and (8.7) into model (8.5) and introducing $\delta L = L - L_{\text{nom}}$ gives

$$\delta L = L_s + D + d_1 + d_2 - L_s[\delta\alpha(\theta_0 + \Delta) + \delta\theta\alpha_s] - L_{\text{nom}} \quad (8.8)$$

as the measurement model employed here.

8.2.6 This measurement problem is treated in terms of model (8.8) with output quantity δL and input quantities L_s , D , d_1 , d_2 , α_s , θ_0 , Δ , $\delta\alpha$ and $\delta\theta$. This treatment differs from JCGM 100:2008 example H.1 where the models (8.6) and (8.7) are regarded as sub-models to model (8.5). The approach here avoids having to use the results obtained from MCM applied to the sub-models (8.6) and (8.7) to provide information about the distributions for the input quantities d and θ in model (8.5).

8.3 GUF

8.3.1 The application of the GUF is based on the first-order Taylor series approximation to model (8.8), the use of the Welch-Satterthwaite formula to evaluate the effective degrees of freedom (rounded towards zero) associated with the uncertainty obtained from the law of propagation of uncertainty, and assigning a scaled and shifted t -distribution with the above degrees of freedom to the output quantity. For a detailed account of the GUF approach, including the determination of the sensitivity coefficients, the assignment of estimates and associated uncertainties, and the consideration of second-order terms, see [10, Annex H.1].

8.4 MCM

The application of MCM requires sampling from probability distributions describing available knowledge about the input quantities 8.4.1.

8.4.1 Assignment of PDFs

8.4.1 In the following clauses the available information about each input quantity in model (8.8) is provided. For each item of information, the JCGM 100:2008 [10] subclause from which the item is extracted is identified. Also provided is an interpretation of the information in terms of an assignment of a distribution to the quantity. Table 8.1 summarizes the assignments made.

Table 8.1: PDFs assigned to input quantities for the gauge block model (8.8), based on available information

Quantity	PDF	μ	σ	ν	a	b	d
L_s/nm	$t_\nu(\mu, \sigma^2)$	50 000 623	25	18			
D/nm	$t_\nu(\mu, \sigma^2)$	215	5.8	24			
d_1/nm	$t_\nu(\mu, \sigma^2)$	0	3.9	5			
d_2/nm	$t_\nu(\mu, \sigma^2)$	0	6.7	8			
$\alpha_s/(10^{-6} \text{ }^\circ\text{C}^{-1})$	$R(a, b)$				9.5	13.5	
$\theta_0/^\circ\text{C}$	$N(\mu, \sigma^2)$	-0.1	0.2				
$\Delta/^\circ\text{C}$	$U(a, b)$				-0.5	0.5	
$\delta\alpha/(10^{-6} \text{ }^\circ\text{C}^{-1})$	$\text{CTrap}(a, b, d)$				-1.0	1.0	0.1
$\delta\theta/^\circ\text{C}$	$\text{CTrap}(a, b, d)$				-0.050	0.050	0.025

8.4.2 The calibration certificate for the reference standard gives $\widehat{L}_s = 50.000\,623$ mm as its length at 20°C [10, Annex H.1.5]. It gives $U(\widehat{L}_s) = 75$ nm as the expanded uncertainty of the reference standard length and states that it was obtained using a coverage factor of $k = 3$ [10, Annex H.1.3.1]. The certificate states that the effective degrees of freedom associated with the combined standard uncertainty, from which the quoted expanded uncertainty was obtained, is $\nu_{\text{eff}}(u(\widehat{L}_s)) = 18$ [10, Annex H.1.6]. Thus, a scaled and shifted t -distribution $t_\nu(\mu, \sigma^2)$ is assigned to \widehat{L}_s [17, clause 6.4.9.7]. Here μ , σ and ν are the location, scale and degrees of freedom, respectively, with values

$$\mu = 50\,000\,623 \text{ nm}, \quad \sigma = \frac{U}{k} = \frac{75}{3} \text{ nm} = 25 \text{ nm}, \quad \nu = 18.$$

8.4.3 The average of the five indications of the length difference between the gauge block being calibrated and the reference standard is 215 nm [10, Annex H.1.5]. The pooled experimental standard deviation characterizing the comparison of L and L_s was determined from 25 indications, obtained independently, of the length difference of two standard gauge blocks, and equalled 13 nm [10, Annex H.1.3.2]. Thus, a scaled and shifted t -distribution $t_\nu(\mu, \sigma^2)$ is assigned to D [17, clause 6.4.9.2], with

$$\mu = 215 \text{ nm}, \quad \sigma = \frac{13}{\sqrt{5}} \text{ nm} = 5.8 \text{ nm}, \quad \nu = 24.$$

8.4.4 According to the calibration certificate of the comparator used to compare L with L_s , the associated uncertainty due to random effects is 10 nm for a coverage probability of 95 % and is obtained from six replicate measurements [10][Annex H.1.3.2]. Thus, a scaled and shifted t -distribution $t_\nu(\mu, \sigma^2)$ is assigned to d_1 [17, clause 6.4.9.2], with

$$\mu = 0 \text{ nm}, \quad \sigma = \frac{U}{k} = \frac{10}{2.57} \text{ nm} = 3.9 \text{ nm}, \quad \nu = 5.$$

Here, k is obtained from table G.2 of JCGM 100:2008 with $\nu = 5$ degrees of freedom and $p = 0.975$.

8.4.5 The uncertainty of the comparator due to systematic effects is given in the certificate as 20 nm at the ‘three sigma level’ [10, Annex H.1.3.2]. This uncertainty may be assumed to be reliable to 25 %, and thus the degrees of freedom is $\nu_{\text{eff}}(u(d_2)) = 8$ [10, Annex H.1.6]. Thus, a scaled and shifted t -distribution $t_\nu(\mu, \sigma^2)$ is assigned to d_2 [17, clause 6.4.9.7], with

$$\mu = 0 \text{ nm}, \quad \sigma = \frac{U}{k} = \frac{20}{3} \text{ nm} = 6.7 \text{ nm}, \quad \nu = 8.$$

8.4.6 The coefficient of thermal expansion of the length reference standard is given as $\hat{\alpha}_s = 11.5 \times 10^{-6} \text{ }^\circ\text{C}^{-1}$ with possible values of this quantity represented by a rectangular distribution with limits $\pm 2 \times 10^{-6} \text{ }^\circ\text{C}^{-1}$ [10, Annex H.1.3.3]. Thus, a rectangular distribution $R(a, b)$ is assigned to α_s [17, clause 6.4.2] with limits $a = 9.5 \times 10^{-6} \text{ }^\circ\text{C}^{-1}$ and $b = 13.5 \times 10^{-6} \text{ }^\circ\text{C}^{-1}$.

8.4.7 The temperature of the test bed is reported as $(19.9 \pm 0.5) \text{ }^\circ\text{C}$. The average temperature deviation $\hat{\theta}_0 = -0.1 \text{ }^\circ\text{C}$ is reported as having an associated standard uncertainty due to the uncertainty associated with the average temperature of the test bed of $u(\hat{\theta}_0) = 0.2 \text{ }^\circ\text{C}$ [10, Annex H.1.3.4]. Thus, a normal distribution $N(\mu, \sigma^2)$ is assigned to θ_0 [17, clause 6.4.7], with $\mu = -0.1 \text{ }^\circ\text{C}$ and $\sigma = 0.2 \text{ }^\circ\text{C}$. Since there is no information about the source of the evaluation of the uncertainty, a normal distribution is assigned.

8.4.8 The temperature of the test bed is reported as $(19.9 \pm 0.5) \text{ }^\circ\text{C}$. The stated maximum offset of $-0.5 \text{ }^\circ\text{C}$ for Δ is said to represent the amplitude of an approximately cyclical variation of temperature under a thermostatic system. The cyclic variation of temperature results in a U-shaped (arc sine) distribution [10, Annex H.1.3.4]. Thus, an arc sine distribution $U(a, b)$ is assigned to Δ [17, clause 6.4.6], with limits $a = -0.5 \text{ }^\circ\text{C}$ and $b = 0.5 \text{ }^\circ\text{C}$.

8.4.9 The estimated bounds on $\delta\alpha$ are $\pm 1.0 \times 10^{-6} \text{ }^\circ\text{C}^{-1}$ with an equal probability of $\delta\alpha$ having any value within those bounds [10, Annex H.1.3.5]. These bounds are deemed to be reliable to 10 %, giving $\nu(u(\widehat{\delta\alpha})) = 50$ [10, Annex H.1.6]. Thus, a curvilinear trapezoid, *i.e.*, a rectangular distribution with inexactly prescribed limits, is assigned to $\delta\alpha$ [17, clause 6.4.3], with parameters $a = -1.0 \times 10^{-6} \text{ }^\circ\text{C}^{-1}$, $b = 1.0 \times 10^{-6} \text{ }^\circ\text{C}^{-1}$ and $d = 0.1 \times 10^{-6} \text{ }^\circ\text{C}^{-1}$. The stated reliability of 10 % on the estimated bounds provides the basis for this value of the parameter d .

8.4.10 The reference standard and the gauge block being calibrated are expected to be at the same temperature, but the temperature difference $\delta\theta$ could lie with equal probability anywhere in the estimated interval $-0.05 \text{ }^\circ\text{C}$ to $0.05 \text{ }^\circ\text{C}$ [10, Annex H.1.3.6]. This difference is believed to be reliable only to 50 %, giving $\nu(u(\widehat{\delta\theta})) = 2$ [10, Annex H.1.6]. Thus, a curvilinear trapezoid, *i.e.*, a rectangular distribution with inexactly prescribed limits is assigned to $\delta\theta$ [17, clause 6.4.3], with parameters $a = -0.050 \text{ }^\circ\text{C}$, $b = 0.050 \text{ }^\circ\text{C}$, and $d = 0.025 \text{ }^\circ\text{C}$. The stated reliability of 50 % provides the basis for this value of the parameter d .

8.5 Results

8.5.1 Table 8.2 gives the results obtained using the information summarized in table 8.1. Figure 8.1 shows the PDFs for δL obtained from the application of the GUF (red solid curve) and MCM (blue scaled histogram). For the latter, 1.3×10^6 trials were taken by the adaptive MCM [17, clause 7.9] with a numerical tolerance ($\delta = 0.5$) set to deliver $n_{\text{dig}} = 2$ significant decimal digits in the standard uncertainty.

Table 8.2: Results for model (8.8) using the information summarized in table 8.1

Method	$\widehat{\delta L}$ /nm	$u(\widehat{\delta L})$ /nm	Shortest 99 % coverage interval for δL /nm
GUF	838	32	[745, 931]
MCM	838	36	[745, 932]

There are some modest differences in the results obtained between the GUF and the MCM. Taking into account higher-order terms in the GUF does not affect the estimate and yields a smaller difference between the standard uncertainties provided by the GUF and MCM (2 nm as compared to 4 nm considering first-order terms only; see [10, Annex H.1.7]). Whether such differences are important is to be judged in terms of the manner in which the results are to be used.

8.5.2 The estimate $\widehat{\delta L}$ is the sum of three terms of model (8.8), L_s , D and L_{nom} . The remaining terms have estimates equal to zero, thus not contributing to the measurand estimate. Yet, the associated uncertainties are different from zero and do contribute to the associated uncertainty. See also [15, clauses 10.3.4 and 10.4.3].

8.5.3 The distribution obtained from the GUF is a t -distribution with $\nu = 16$ degrees of freedom. The endpoints of the shortest 99 % coverage intervals for δL obtained from the PDFs are indicated as (broken, red) vertical lines (obtained from the GUF) and (continuous, blue) vertical lines (obtained from MCM).

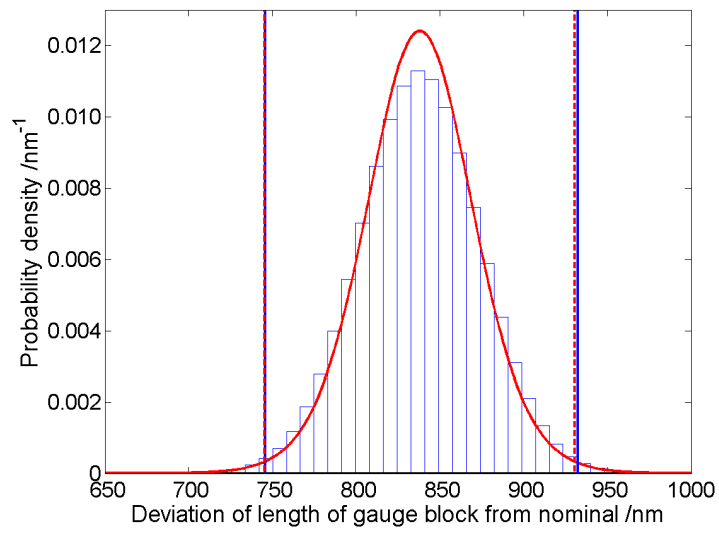


Figure 8.1: PDF for δL obtained using the GUF (red curve) and scaled histogram from MCM for the approximate model (8.8) using the information summarized in table 8.1

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

9.1 Preamble

9.1.1 This example is taken from a compendium of examples [1].

9.1.2 This example addresses the straight-line calibration of a torque-measuring sensor against a reference system using measurements taken at different torque values. The goal is to determine a linear relationship that relates the results of the torque measuring sensor with those of the reference system. Two approaches are provided: (i) ordinary and weighted least-squares estimation (see, for example, ISO/TS 28037 [73]), accompanied by an uncertainty evaluation based on the law of propagation of uncertainty (LPU) [10], and (ii) a statistical approach applying Bayesian inference (compare, for example, [57]).

9.2 Experimental

Measurements have been carried out at different values of torque by a sensor and a reference system. For the i th torque level, let x_i denote the indication provided by the reference system and y_{ij} the n_i repeated measurement results of the sensor, where the latter are summarized through their means y_i and standard deviations s_i , respectively. These summary statistics are displayed in Table 9.1 and are available online in repository [86]. For further details about the measurements the reader is referred to [125]. We observe that the standard deviation differs by up to an order of magnitude between different torque levels.

Table 9.1: Summary statistics for measurement data given in the guideline [125][table B6, page 43].

i	x_i /(Nm)	Mean y_i /(Nm)	s_i /(Nm)	n_i
1	0.101	0.0950	0.0055	6
2	0.201	0.1966	0.0052	6
3	0.305	0.3016	0.0041	6
4	0.501	0.4983	0.0041	6
5	1.001	1.0083	0.0098	6
6	3.000	3.0266	0.0082	6
7	4.001	4.0466	0.0121	6
8	5.007	5.0666	0.0379	3

9.3 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, denoted below as response. The linear relation

$$Y = \beta X \tag{9.1}$$

is assumed to model the relationship between the measured responses of the sensor and the applied stimulus. Model (9.1) represents a straight line with zero offset. Such a straight

line may be chosen for physical reasons, or may be supported by external data. The measurand is the slope parameter β in the particular straight line model (9.1). The input quantities are Y_1, \dots, Y_p (with $p = 8$ in this example), which correspond to the measured values at the considered torques X_1, \dots, X_p . While the variability associated with the estimates y_1, \dots, y_p differs by an order of magnitude, the variability associated with the measured values x_1, \dots, x_p from the reference device are considered sufficiently small to be neglected.

9.4 Measurement model

9.4.1 General

9.4.1.1 The uncertainty evaluations presented in this example are based on two different models. The uncertainty evaluation following the JCGM 100:2008 [10] 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.

9.4.1.2 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. Bayesian inference can account for prior knowledge about the measurand, resulting 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 Bayesian inference, rather than by a process of propagating input uncertainties through a measurement model in the sense of JCGM 100:2008 [10].

9.4.2 Ordinary and weighted least-squares

9.4.2.1 The application of weighted least squares (WLS) estimation determines an estimate $\hat{\beta}$ for the measurand by minimizing

$$Q = \sum_{i=1}^p \sum_{j=1}^{n_i} W_i (y_{ij} - \beta x_i)^2 = \sum_{i=1}^p W_i [n_i (y_i - \beta x_i)^2 + (n_i - 1) s_i^2] \quad (9.2)$$

with respect to β . In expression (9.2), W_i denote some weights, $i = 1, \dots, p$. The solution $\hat{\beta}$ to this minimization problem is

$$\hat{\beta} = \left(\sum_{i=1}^p n_i W_i x_i^2 \right)^{-1} \sum_{i=1}^p n_i W_i x_i y_i. \quad (9.3)$$

The measurement model will be defined as

$$\beta = \left(\sum_{i=1}^p n_i W_i x_i^2 \right)^{-1} \sum_{i=1}^p n_i W_i x_i Y_i, \quad (9.4)$$

that is, by replacing the estimates y_i in expression (9.3) with corresponding quantities Y_i . Note that a measurement model in the sense of JCGM 100:2008 [10] 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 (9.4) has already been replaced with the known value x_i in this example.

Ordinary least squares (OLS) estimation is obtained by choosing weights $W_i = 1$ for all observations $i = 1, \dots, p$ in (9.4).

9.4.3 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 individual observations y_{ij} , for $j = 1, \dots, n_i$, $i = 1, \dots, p$, are realizations of independently and normally distributed random variables Y_{ij} with means equal to βx_i and variances σ_i^2 , that is,

$$y_{ij} | \beta, \sigma_i^2 \sim N(\beta x_i, \sigma_i^2). \quad (9.5)$$

In expression (9.5), the x_i denote the known stimuli and σ_i^2 are the variances to be inferred. The likelihood function is the probability for the observed data viewed as a function of the model parameters. For the statistical measurement model (9.5), the likelihood function is

$$\ell(\beta, \sigma^2 | \text{data}) \propto \prod_{i=1}^p (\sigma_i^2)^{-n_i/2} \exp \left\{ -\frac{1}{2\sigma_i^2} [(n_i - 1)s_i^2 + n_i(y_i - \beta x_i)^2] \right\}, \quad (9.6)$$

where $\sigma^2 = (\sigma_1^2, \dots, \sigma_p^2)^\top$ and 'data' summarizes the sufficient statistics $y_1, \dots, y_p, s_1, \dots, s_p$ of the data.

9.5 Uncertainty evaluation

9.5.1 Law of propagation of uncertainty

9.5.1.1 The measurement model (9.4) contains input quantities Y_i , $i = 1, \dots, p$. For each of these input quantities, a series of repeated measurement results y_{ij} , $j = 1, \dots, n_i$, is available. In following JCGM 100:2008, the mean is taken as an estimate y_i of Y_i and the standard deviation of the mean is taken as standard uncertainty $u(y_i) = s_i / \sqrt{n_i}$. The estimates and standard uncertainties for the input quantities are listed in table 9.2.

Table 9.2: Estimates y_i and associated standard uncertainties $u(y_i)$ for the data in table 9.1 and used in measurement model (9.4)

Input quantity	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8
$y_i / (\text{Nm})$	0.0950	0.1966	0.3016	0.4983	1.0083	3.0266	4.0466	5.0666
$u(y_i) / (\text{Nm})$	0.0022	0.0021	0.0017	0.0017	0.0040	0.0033	0.0049	0.0219

9.5.1.2 According to JCGM 100:2008 [10], the estimate $\hat{\beta}$ of β is taken as the value given by the measurement model (9.4) when inserting the estimates from table 9.2 for the input quantities, that is,

$$\hat{\beta} = \left(\sum_{i=1}^p n_i W_i x_i^2 \right)^{-1} \sum_{i=1}^p n_i W_i x_i y_i. \quad (9.7)$$

The associated variance (squared standard uncertainty) is

$$u^2(\hat{\beta}) = \sum_{i=1}^p \left(\frac{\partial \beta}{\partial Y_i} \Big|_{Y_i=y_i} \right)^2 u^2(y_i) = \frac{\sum_{i=1}^p (n_i W_i x_i)^2 u^2(y_i)}{\left(\sum_{i=1}^p n_i W_i x_i^2 \right)^2}. \quad (9.8)$$

9.5.1.3 The WLS estimate assigns weights as $n_i W_i = 1/u^2(y_i)$, whereas OLS estimate corresponds to setting $W_i = 1$ for all i .

9.5.1.4 According to JCGM 100 Annex G.3, $\hat{\beta}$ follows a t -distribution with ν_{eff} degrees of freedom, such that a 95% coverage interval is given by $\hat{\beta} \pm t_{0.975, \nu_{\text{eff}}} u(\hat{\beta})$. For OLS $\nu_{\text{eff}} = 2.7$ and for WLS $\nu_{\text{eff}} = 14$ is obtained by applying JCGM 100 (G.2.b) with

$$\nu_{\text{eff}} = \frac{u^4(\hat{\beta})}{\sum_{i=1}^p \frac{u_i^4(\hat{\beta})}{\nu_i}} \quad (9.9)$$

where $\nu_i = n_i - 1$ and $u_i(\hat{\beta}) = \frac{x_i/u(y_i)}{\sum_{j=1}^p (x_j/u(y_j))^2}$ or $u_i(\hat{\beta}) = \frac{n_i x_i u(y_i)}{\sum_{j=1}^p n_j x_j^2}$, respectively, using (9.8).

9.5.2 Bayesian uncertainty analysis

9.5.2.1 In a Bayesian inference, prior knowledge about the measurand (and other unknowns) is combined with the information contained in the data through application of Bayes' theorem. The result is the posterior distribution that summarizes the knowledge about the measurand (and other unknowns in the statistical model (9.5)), conditional on the observed data. In our case, the posterior is given through the following probability density function (PDF):

$$p(\beta, \sigma^2 | \text{data}) \propto \pi(\beta, \sigma^2) \ell(\beta, \sigma^2 | \text{data}), \quad (9.10)$$

where $\ell(\beta, \sigma^2 | \text{data})$ is the likelihood function (9.6) for the assumed statistical model (9.5), and $\pi(\beta, \sigma^2)$ is the adopted prior for the model parameters β and σ^2 where σ^2 denotes a vector $(\sigma_1^2, \dots, \sigma_p^2)$.

9.5.2.2 Once the posterior distribution of all model parameters is obtained, expression (9.10), the posterior $p(\beta | \text{data})$ for the parameter of interest (the measurand) is obtained through marginalization:

$$p(\beta | \text{data}) = \int_0^\infty \cdots \int_0^\infty p(\beta, \sigma^2 | \text{data}) d\sigma_1^2 \dots d\sigma_p^2. \quad (9.11)$$

The marginal posterior (9.11) is a PDF that constitutes the complete Bayesian uncertainty analysis for the measurand. Summary statistics of this PDF may be sufficient in many cases, and the posterior mean

$$\hat{\beta} = \int_{-\infty}^\infty p(\beta | \text{data}) \beta d\beta \quad (9.12)$$

can be considered as the Bayesian estimate of the slope β , and the posterior standard deviation as its associated standard uncertainty, $u(\hat{\beta})$:

$$u^2(\hat{\beta}) = \int_{-\infty}^\infty p(\beta | \text{data}) (\beta - \hat{\beta})^2 d\beta. \quad (9.13)$$

Finally, a 95 % credible interval $[\beta_L, \beta_H]$ can be calculated from the posterior (9.11), which satisfies

$$\int_{\beta_L}^{\beta_H} p(\beta|\text{data}) d\beta = 0.95. \quad (9.14)$$

Since Eq. (9.14) does not uniquely determine a credible interval, further conditions need to be imposed, for example, that the credible interval is probabilistically symmetric, that is, symmetric around the Bayes estimate, or is the 95 % credible interval of shortest length.

9.5.2.3 Informative prior In the following, assume that $\pi(\beta, \sigma^2) = \pi(\beta)\pi(\sigma^2)$, and that prior knowledge about the variances σ_i^2 is available in a form of prior estimates $\hat{\sigma}_i^2$ ($i = 1, \dots, p$). Assume also that the reliability of these prior estimates can be expressed in terms of coefficients of variation c_i . This set of prior knowledge can be modelled using independent inverse Gamma distributions

$$\pi(\sigma_i^2) = \frac{b_i^{a_i}}{\Gamma(a_i)} \sigma_i^{-2(a_i+1)} \exp\left(-\frac{b_i}{\sigma_i^2}\right), \quad \text{for } \sigma_i^2 > 0, \quad (9.15)$$

with parameters a_i and b_i given as follows:

$$a_i = 2 + 1/c_i^2, \quad b_i = (a_i - 1)\hat{\sigma}_i^2. \quad (9.16)$$

That is, the prior knowledge about each σ_i^2 is modelled by an inverse Gamma distribution with mean $\hat{\sigma}_i^2$ and variance $c_i^2(\hat{\sigma}_i^2)^2$.

In this example no true prior knowledge has been available. For the purpose of illustration, hypothetical prior knowledge about β in the form of a normal distribution with mean 1 and standard deviation 0.1 is used, accompanied with guesses $\hat{\sigma}_i^2$ 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 $c_i = 1$, which results in $a_i = 3$ and $b_i = 2s_i^2$. Prior knowledge is information available before measurements are performed. Assigning the prior from the observed data is not a recommended practice.

9.5.2.4 The resulting marginal posterior for the slope is then obtained as

$$p(\beta|\text{data}) \propto \pi(\beta) \prod_{i=1}^p t_{n_i-1+2a_i}\left(\beta; y_i/x_i, \frac{(n_i-1)s_i^2 + 2b_i}{n_i x_i^2 (n_i - 1 + 2a_i)}\right), \quad (9.17)$$

where $\pi(\beta)$ denotes the prior PDF for β , and $t_\nu(x; m, s)$ stands for the PDF of a scaled and shifted t -distribution with ν degrees of freedom, a mean m and a scale parameter s , that is,

$$t_\nu(x; m, s^2) \propto \left(1 + \frac{1}{\nu} \frac{(x-m)^2}{s^2}\right)^{-(\nu+1)/2}. \quad (9.18)$$

9.5.2.5 The univariate PDF (9.17) can be evaluated and the summary statistics (9.12) to (9.14) obtained through standard procedures of numerical quadrature. For evaluating (9.17), it is advantageous to calculate the logarithm of $p(\beta|\text{data})$ first, applying the exponential function afterwards.

9.5.2.6 Markov Chain Monte Carlo method (MCMC) sampling can also be employed to draw samples from the posterior distribution arising from the above considerations.

9.5.2.7 Non-informative prior The case of vague prior knowledge can be modelled by choosing a vague prior with large variance for the measurand β , and parameters a_i and b_i of the inverse Gamma distributions for the variances σ_i^2 which approach zero. For example, for β a normal distribution with mean 1 and standard deviation 100 can be used.

Then numerical quadrature or Markov chain Monte Carlo sampling can be performed as before to obtain the posterior (9.17). The limiting posterior is formally obtained by using the non-informative prior $\pi(\beta, \sigma^2) \propto \prod_{i=1}^p 1/\sigma_i^2$.

9.6 Reporting the result

Table 9.3 contains the estimate, its associated standard uncertainty, and the 95 % coverage interval obtained by application of the JCGM 100:2008 uncertainty framework (GUF) to OLS and 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 9.1 shows the PDF for the measurand obtained by Bayesian uncertainty analyses in comparison with the results achieved by WLS estimation with uncertainty evaluated according to JCGM 100:2008 [10].

Table 9.3: Results obtained by ordinary least-squares (OLS) and weighted least-squares (WLS) with uncertainty evaluation according to JCGM 100:2008, as well as results from a Bayesian uncertainty evaluation with (Inf-Bayes) and without (NI-Bayes) accounting for prior knowledge

Method	β	$u(\beta)$	95 % coverage (credible) interval
OLS	1.0107	0.0015	[1.0055, 1.0158]
WLS	1.0085	0.0008	[1.0069, 1.0101]
NI-Bayes	1.0092	0.0011	[1.0070, 1.0112]
Inf-Bayes	1.0091	0.0009	[1.0073, 1.0108]

9.7 Discussion

9.7.1 The results obtained by application of the GUF to OLS and WLS estimation are quite different, which is due to the difference of the corresponding models (9.4) used. Specifically, the WLS estimation with weights $n_i W_i = 1/u^2(y_i)$ leads to a different estimate for the slope and a smaller uncertainty than OLS estimation. These weights are ‘optimal’ in the sense that they take into account the different variabilities of the data and thus lead to a minimum uncertainty under all measurement models (9.4).

9.7.2 On the other hand, OLS does not apply ‘optimal’ weights and results in a larger uncertainty associated with its different estimate of 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 (9.5). However, OLS estimation utilizes a measurement model that does not account for the fact that different observations have different

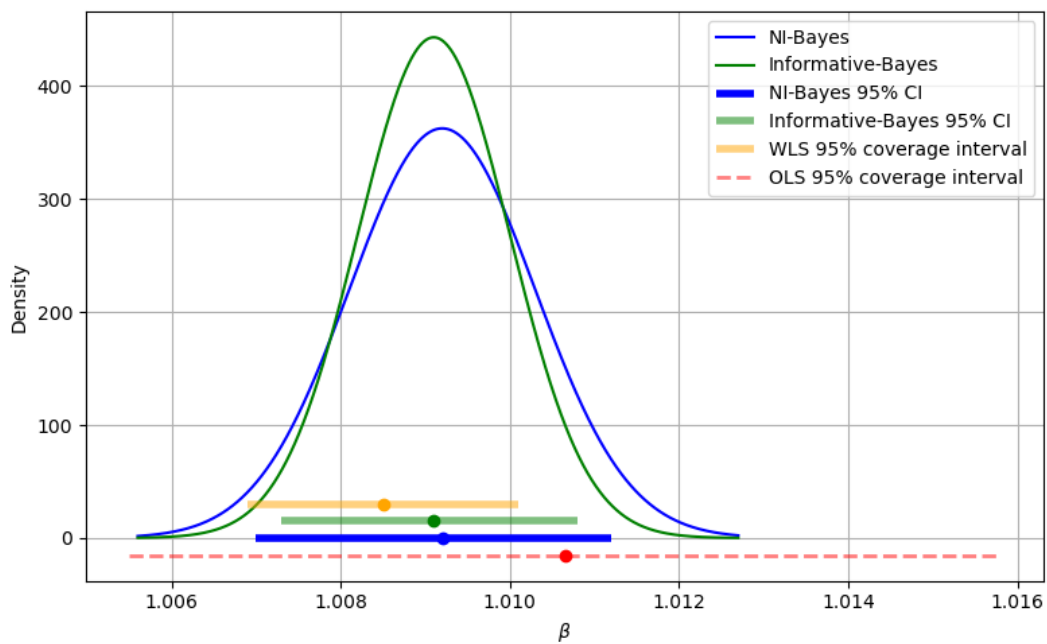


Figure 9.1: Marginal posterior distributions $p(\beta|\text{data})$ for the regression slope from Bayesian inference using an informative and non-informative prior, with the associated 95% credible intervals. For OLS and WLS the estimates and 95% coverage intervals are displayed

variability. That is, measurements are assigned the same weight although their variability differs by an order of magnitude.

9.7.3 The WLS estimate can also be viewed as a solution to the statistical model (9.5) if the variances σ_i^2 were known. For unknown variances, however, the optimal weights $n_i W_i = 1/(\sigma_i^2/n_i)$ are also unknown. While the standard uncertainties $u(y_i)$ provide approximate estimates of these optimal weights in the form of $\sigma_i/\sqrt{n_i}$, they can become unreliable when derived from a small number of repeated measurements.

9.7.4 The Bayesian uncertainty analysis is based on the statistical model (9.5) 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, the Bayesian uncertainty analysis is recommended for this example. It should be noted that methods from classical statistics can also be used to analyse the data on the basis of the statistical model (9.5), which have not been considered in this example.

9.7.5 Bayesian inference using the informative prior described in clause 9.5.2.3 yields very similar results to those using the non-informative prior (9.5.2.7). The reason is that the data dominate the prior information taken for the parameters and that prior knowledge

about the variances was obtained in accordance with the observed variances. If either of these two latter conditions for the prior for the variances is removed, the results of an informative Bayesian inference might look appreciably 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 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 procedure 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.

9.7.6 The statistical model (9.5) used here does not directly account for possible errors in the measured values provided by the reference system. In fact, the example B2 of the guideline VDI/VDE 2600 part 2 [125] reports non-vanishing uncertainties for them. The statistical model (9.5) accounts 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.

Final draft

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

10.1 Preamble

10.1.1 This example is taken from a compendium of examples [1].

10.1.2 The main goal of this example is to calculate, according to JCGM 106:2012 [13], risks of false decisions in conformity assessment when a normal distribution is not adequate to model prior information about the measurand. As a case study, mass concentration of total suspended particulate matter (TSPM) in ambient air is considered.

10.2 Objective and data

10.2.1 A total of 220 measurement results of the mass concentration of TSPM in ambient air, collected in 2009 in the proximity of a stone quarry located in Israel, was obtained according to Environmental Protection Agency (EPA) method IO-2.1 [50]. Such results were compared with the national (Israeli) regulation limit for air quality to study the occurrence of out-of-specification (OOS) results, as detailed in [80, 81].

10.2.2 In this example, the focus is on the calculation of global and specific risks of false decision in the conformity assessment of the air quality, based on such kind of measurement results. The risk of underestimating the pollutant concentration is the consumer's (inhabitants') risk and that of overestimating it is the producer's (owner's) risk. Calculation of such risks is as important for the regulator protecting the inhabitants' quality of life in the area surrounding the quarry as for the manufacturers' association acting in the interests of the stone producers in the country.

10.3 Specification of the measurand

For the characterization of TSPM, the EPA method IO-2.1 [50] indicates the use of a high-volume sampler for collection of particles with aerodynamic diameters of 100 μm or less. For each measurement, a large volume V of air between 1600 m^3 and 2400 m^3 was sampled at a constant flow 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 mass concentration of TSPM over the sampling period: $\rho = m/V$ (mg m^{-3}).

10.4 Measured values and associated measurement uncertainty

10.4.1 The measured values of the TSPM mass concentration, $\hat{\rho}_i$, ($i = 1, \dots, 220$), are reported in mg m^{-3} within the Q1data.txt file (available at [100]) and depicted in figure 10.1. (The running index i will be dropped hereafter for simplicity.) Note further that in JCGM 106:2012 a measured value would be denoted as ρ_m , whereas in this document it is denoted by $\hat{\rho}$.

10.4.2 A full uncertainty budget for the considered results is available in [81], where it is shown that the major contribution to the uncertainty is that from the measurement of the sampled air volume. The relative standard uncertainty associated with any $\hat{\rho}_i$ is 7.0%.

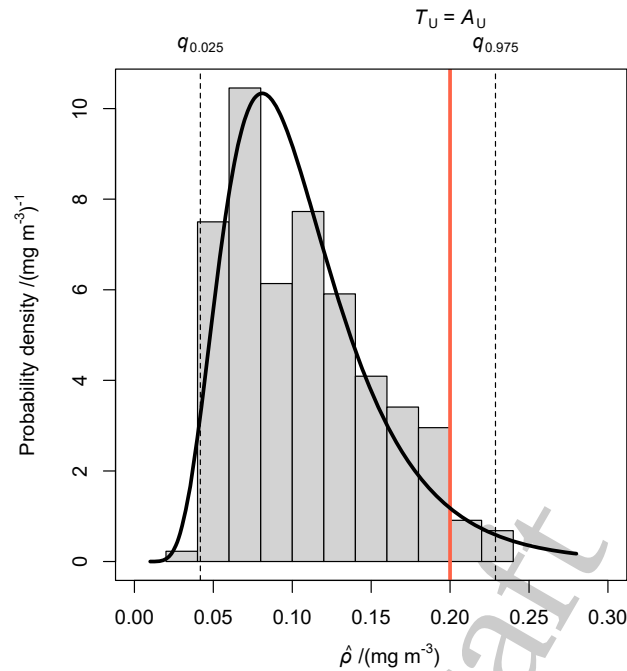


Figure 10.1: Histogram of the measured values of TSPM mass concentration and the corresponding lognormal probability density function (expression 10.1). The quantiles of this distribution are $q_{0.025} = 0.042 \text{ mg m}^{-3}$ and $q_{0.975} = 0.229 \text{ mg m}^{-3}$

10.5 Tolerance limits

The Israeli national regulations of ambient air quality prescribe an upper tolerance (regulation) limit $T_U = 0.2 \text{ mg m}^{-3}$ for TSPM mass concentration averaged over 24 h. This limit applies for any location, including those close to the quarry.

10.6 Decision rule and conformity assessment

10.6.1 General

10.6.1.1 Regulations require direct comparison of measured values $\hat{\rho}_i$ with T_U , as shown in figure 10.1. In the present example, acceptance limits A_U 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, $A_U = T_U$), a ‘shared risk’ rule is considered as the decision rule for conformity assessment [13, clause 8.2.1].

10.6.1.2 In this instance, the inhabitants living in the area are considered the ‘consumers’, whereas the stone quarry owners are the ‘producers’. So, the ‘consumer’s risk’ relates to the inhabitants, whereas the ‘producer’s risk’ relates to the quarry owners.

10.6.1.3 The global and specific risks of false decisions in conformity assessment are defined in [13, clause 3.3] for both the consumer and the producer, and have different interpretations. While a specific risk is the risk of an incorrect decision based on a measurement result of a particular air sample, global risk refers to the probability of incorrect

decisions based on future measurement results. 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 [57]) is used for specific risks while the joint distribution is used for global risks. The posterior distribution is for the true value ρ given the measured value $\hat{\rho}$. The joint distribution is for the true value *and* the measured value. See clauses 10.6.3 and 10.6.4.

10.6.2 Bayesian framework

10.6.2.1 The evaluation of specific and global risks of false decisions on a characteristic of an item is described in JCGM 106:2012 [13].

The underlying approach requires defining a prior probability density function (PDF) $g_0(\rho)$ for TSPM mass concentration. Here, the prior PDF is based on experimental data (see JCGM 106:2012 [13, clause 6.2.2]) and describes the distribution of the population of the TSPM mass concentration. An adequate fit for the experimental results, among possible candidate distributions, is the lognormal distribution:

$$g_0(\rho) = \frac{1}{\rho\sigma\sqrt{2\pi}} \exp\left[-\frac{(\ln\rho - \mu)^2}{2\sigma^2}\right]. \quad (10.1)$$

The lognormal prior PDF (10.1) is depicted as the curve approximating the histogram in figure 10.1. The location and scale parameter estimates of this distribution are $\hat{\mu} = -2.325$ and $\hat{\sigma} = 0.434$, respectively. They are the sample mean and standard deviation, respectively, of the log-transformed data.

10.6.2.2 Each measurement result $\hat{\rho}$ was modelled by a normal distribution with expectation equal to ρ and standard deviation equal to the standard measurement uncertainty $u = 0.07\hat{\rho}$ [81]. The corresponding likelihood is hence

$$h(\hat{\rho}|\rho) = \frac{1}{u\sqrt{2\pi}} \exp\left[-\frac{(\hat{\rho} - \rho)^2}{2u^2}\right]. \quad (10.2)$$

10.6.2.3 When both the prior PDF and the likelihood are normal distributions, the posterior PDF [13, expression (1)] is normal [13, clause 7.2.1]. In this example, the prior PDF is log-normal. However, the method described in JCGM 106:2012 [13] to evaluate risks applies for any PDF.

10.6.3 Global risks

10.6.3.1 For any considered upper acceptance limit A_U , global risks for the consumer and the producer were calculated by combining the lognormal prior PDF (10.1) and the normal likelihood (10.2), according to [13, expressions (19) and (20)]. The joint PDF can be obtained in a numerically stable way by taking the logarithms of both the prior and the likelihood, summing the two transformed distributions and taking the exponential of the result, as follows:

```
P = function(rho_rho_hat){
  p1 = dlnorm(rho_rho_hat[1], meanlog = -2.325, sdlog = 0.434,
             log = TRUE)
  p2 = dnorm(rho_rho_hat[2], mean = rho_rho_hat[1],
```

```

    sd = 0.07 * rho_rho_hat[2], log = TRUE)
  exp(p1 + p2)
}

```

10.6.3.2 The integration of this joint PDF can be carried out using the R function `cubintegrate` defined in the package `cubature` [94]. In the considered case, since all the involved PDFs were defined on the positive axis only, the lower integration limits were both taken as zero:

```

T_U = 0.2
A_U = 0.2

```

```

require(cubature)
Int = cubintegrate # rename the function cubintegrate
C = 1/Int(P, lower = c(0,0), upper = c(Inf, Inf))$int
risk_c = C * Int(P, lower = c(T_U, 0), upper = c(Inf, A_U))$int
risk_p = C * Int(P, lower = c(0, A_U), upper = c(T_U, Inf))$int

```

10.6.3.3 The obtained consumer's (red upward sloping line) and producer's (blue downward sloping line) global risks are displayed in figure 10.2, for A_U values varying in the interval $[T_U - 0.05 \text{ mg m}^{-3}, T_U + 0.05 \text{ mg m}^{-3}]$.

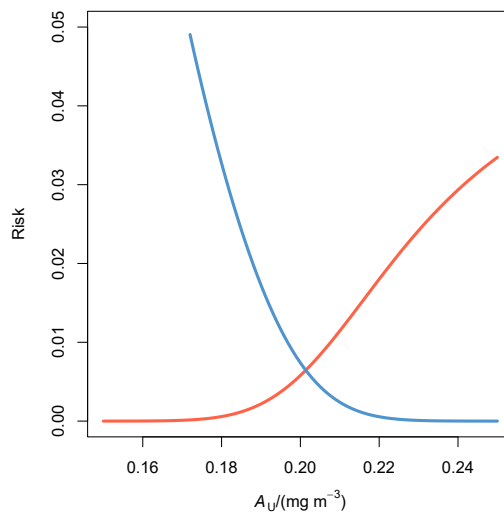


Figure 10.2: Global consumer's risk (red upward sloping line) and global producer's risk (blue downward sloping line) as functions of the acceptance limit

Considering, for example, the special case in which $A_U = T_U$, consumer's and producer's global risks were 0.5% and 0.9%, respectively.

10.6.3.4 Consider determining the maximum acceptable value of A_U corresponding to a small consumer's risk, such as 0.002%. The result is that such an acceptance limit should not exceed 0.17 mg m^{-3} . However, in this instance, the global producer's risk would increase from 0.9% to 5.8%. The other way round, A_U should be at least equal to 0.22 mg m^{-3} in order to assure a producer's risk smaller than 0.1%. In this case, the global consumer's risk would increase from 0.5% to 1.7%.

10.6.4 Specific risks

10.6.4.1 Whereas global risks are concerned with future outcomes from this quarry, specific risks are concerned with a particular measured value at hand, as discussed in [13, clause 9.3.2].

10.6.4.2 For a specific value $\hat{\rho} < A_U$ (that is, the measured TSPM mass concentration is assessed as conforming to the regulatory limit), the consumer's specific risk is the integral of the posterior PDF $g(\rho|\hat{\rho})$ over the region $[T_U, \infty]$ — the interval of true values that would be deemed as nonconforming. For a specific value $\hat{\rho} > A_U$ (that is, a value not conforming to the regulatory limit), the producer's specific risk is the integral of the posterior PDF over the interval $[0, T_U]$ — the region of actually conforming true values.

10.6.4.3 In both cases, the posterior PDF $g(\rho|\hat{\rho})$ [13, expression (A.11)] is needed, which is not available in closed form. The integral of the posterior PDF is calculated by means of the R function `integrate`:

```
P = function(rho, rho_hat){
  p1 = dlnorm(rho, meanlog = -2.325, sdlog = 0.434, log=TRUE)
  p2 = dnorm(rho_hat, mean = rho, sd = 0.07 * rho_hat, log=TRUE)
  exp(p1 + p2)
}

T_U = 0.2
# specific consumer's risk for rho_hat = 0.190
rc1 = integrate(P, T_U, Inf, rho_hat=0.190)$value
rc2 = integrate(P, 0, Inf, rho_hat=0.190)$value
risk_c = rc1/rc2

# specific producer's risk for rho_hat = 0.225
rp1 = integrate(P, 0, T_U, rho_hat=0.225)$value
rp2 = integrate(P, 0, Inf, rho_hat=0.225)$value
risk_p = rp1/rp2
```

The obtained consumer's and producer's specific risks for various values of $\hat{\rho}$ are displayed in figure 10.3 (for $A_U = T_U$).

Owing to the comparatively large width of the prior distribution, and thus to the meagre prior information, the posterior is essentially determined by the likelihood. The implication is that the specific risks as calculated in a purely frequentist context would be very close to those given here.

10.7 Interpretation of results

10.7.1 Studies on global risks, such as that conducted in clause 10.6.3, allow the involved parties (consumers and producers) to agree on an acceptance limit, thus balancing the safeguarding of the inhabitants' health and the economical interests of the quarry owners, in the considered example.

10.7.2 The approach in clause 10.6.4 provides risks of false decision for a specific result and for a particular acceptance limit ($A_U = T_U$, in the considered case). From a practical

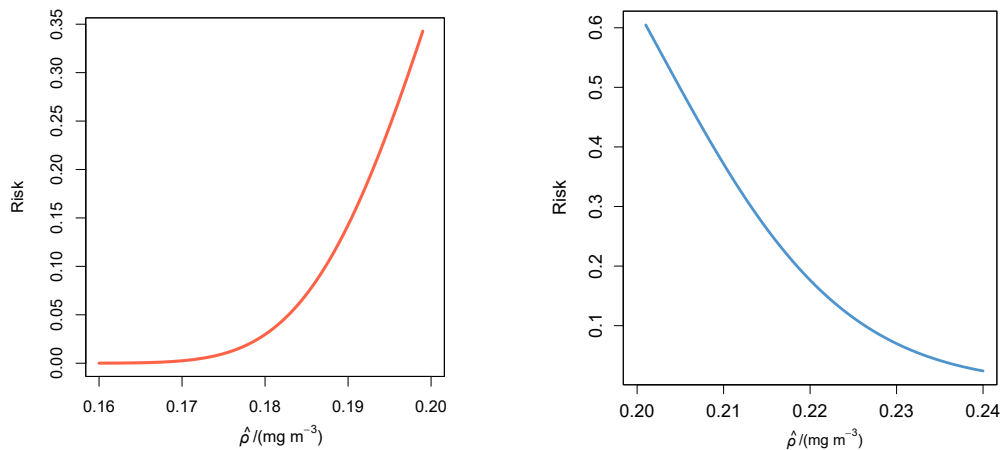


Figure 10.3: Consumer's (red line, left) and producer's (blue line, right) specific risks at $A_U = T_U$ depending on the measurement results for the TSPM concentration in the quarry

point of view, no action will be undertaken when a measurement result falls below the acceptance limit, that is, when it is conforming with requirements. However, when a result exceeds the limit, it will be declared as non-conforming and some corrective action will be required. In this case, producers have at hand a tool for assessing the extent of their responsibility for such failure and possibly elaborate an appropriate reaction. As an example, for a non-conforming measured value $\hat{\rho} = 0.225 \text{ mg m}^{-3}$, the specific producer's risk is about 11 %, meaning that there is a non-negligible 11 % probability that such a value corresponds to an actually conforming true value of the TSPM mass concentration.

10.7.3 Reducing uncertainties associated with measurement results undergoing conformity assessment decreases specific consumer's and producer's risks as well as global risks.

11 Effect of considering a 2D image as a set of pixels on a computed quantity

11.1 Preamble

11.1.1 This example is taken from a compendium of examples [1].

11.1.2 This example is concerned with properties of a physical object extracted from a pixel-based image. With a greater number of pixels 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 and 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.

11.1.3 In radiation therapy anatomical segmentation, image pixel size impacts on the area of a 2D section of a tumour (or an organ) 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.

11.2 Estimation of organ or tumour mass

11.2.1 The estimation of tumour mass necessitates some form of outlining of the tumour on an image. Under assumptions of homogeneity, tumour 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 [60].

11.2.2 The area of a tumour section is generally obtained from an ROI (region of interest) outlined on anatomical imaging data in the form of an array of pixels [55]. Knowing that outline, the corresponding area 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 area. 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. When determining areas manually, an outline is typically drawn by an operator across all images that comprise the data set of concern.

11.2.3 In practice, a set of planar sections of the ROI, usually at a constant spacing, is considered, from each of which the area is deduced and hence, by considering all such sections, the volume is estimated. Here, attention is paid to a single section as one step in the overall process.

11.2.4 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 [64] and the

references therein. Here, just boundary-based methods are considered, such as gradient-based edge detection, which require an initial region of interest (ROI) to be defined inside which an operator or algorithm estimates the object boundary [55].

11.3 Measurement model

11.3.1 The measurand is the area of a plane section of the tumour.

11.3.2 Figure 11.1 shows an outline of a tumour obtained from a pixelated image, at two image resolutions (left: 224 pixels; right: 1840 pixels), produced by scanning equipment, which will be used to provide numerical results and examine the effect of pixel size on the reconstructed shape. It is assumed that the outline of the tumour is contained within the darker-shaded pixels chosen by an operator or algorithmically. Each pixel is taken as of unit size (dimension 1×1).

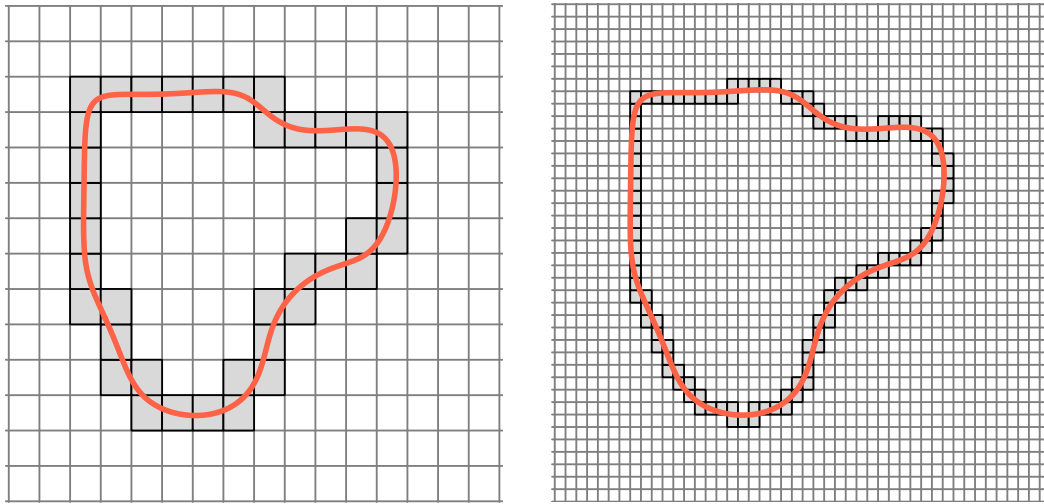


Figure 11.1: Tumour profiles for fine and coarse pixel sizes: 85 and (right) 585 pixels defining the outline and interior

11.3.3 The measurement model is specified by the algorithm used to compute the area of the tumour section. The area depends on knowledge of the boundary of the section, specifically a mathematical description of a smooth curve representing the boundary, for which a parametric representation is used:

$$x = F(t), \quad y = G(t), \quad (11.1)$$

where the parameter t is (ideally) taken as arc length from some starting point.

11.3.4 The tumour section area could also be computed, more approximately, as some average (such as the arithmetic mean) of two extreme areas. Referring to either panel of figure 11.1, one area would be that of the union of the grey pixels and the white pixels they surround. The other area would be that of the union of those white pixels. Such a method may not be effective for a small number of pixels. In clause 11.4, an argument is made in favour of representing the boundary by a smooth curve (11.1). Clause 11.7.4 makes a simple comparison of those approaches in terms of the outlines in figure 11.1.

11.3.5 Various empirical functions such as periodic splines and parametric Fourier series can be used for F and G to form the profile. Parametric Fourier series are selected based on considerable experience [34, 36, 43, 79] with the use of Fourier series for data analysis. For this purpose, some guidance in [75] concerned with the determination and use of polynomial calibration functions is adapted. That standard gives advice on determining the degree of polynomial most appropriate for the data by first constructing polynomials of all degrees n from one up to some value n_{\max} and then selecting a particular degree from that set. Here, parametric Fourier series involving an increasing number of harmonics are determined, choosing the number regarded as suitable according to criteria to be discussed. By analogy with polynomials, the number of harmonics is called *the degree* [98].

11.3.6 A parametric Fourier series (11.1) of degree n has $2n+1$ parameters defining $F(t)$ and an identical number defining $G(t)$. The x - and y -components can be expressed as

$$F_n(t) = b_1 + \sum_{r=1}^n [b_{2r} \cos(2\pi r t) + b_{2r+1} \sin(2\pi r t)], \quad (11.2)$$

$$G_n(t) = c_1 + \sum_{r=1}^n [c_{2r} \cos(2\pi r t) + c_{2r+1} \sin(2\pi r t)]. \quad (11.3)$$

11.3.7 Here, the outline is approximated by such a parametric Fourier series. The area A enclosed by the curve can be expressed in terms of the coefficients in the parametric Fourier series (11.2) and (11.3) by

$$A = \pi \sum_{r=1}^n r (b_{2r+1} c_{2r} - b_{2r} c_{2r+1}), \quad (11.4)$$

an expression derived in [124, annex E5.1.A].

11.3.8 The data comprises a set of $m = 38$ ordered values x_i and y_i , $i = 1, \dots, m$, taken as the midpoints of the grey pixels in figure 11.1 (right panel) through which the tumour boundary passes. Since the true values of each coordinate pair lie somewhere in the containing pixel, it is assumed that these coordinates can be modelled by rectangular distributions $R(-1/2, 1/2)$ relative to that midpoint. Accordingly, the standard uncertainties $u(x_i)$ and $u(y_i)$, denoted by u_0 , associated with the x_i and y_i are taken as the standard deviations of these distributions, namely, $1/(2\sqrt{3}) \approx 0.29$.

The points (x_i, y_i) are regarded as values of the dependent variables x and y corresponding to an increasing set of values t_i , $i = 1, \dots, m$ of the independent variable t . The values of t_i used are the cumulative chordal distance from (x_1, y_1) to (x_i, y_i) :

$$t_1 = 0, \quad t_i = t_{i-1} + [(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2]^{1/2}, \quad i = 2, \dots, m.$$

These t_i -values are normalized so that the total cumulative chord length becomes unity:

$$t_i := \frac{t_i}{t_m}, \quad i = 1, \dots, m.$$

Since, for sufficiently closely-spaced data, cumulative chord length approximates well arc length and is often used for that purpose, normalized cumulative chord length is employed as realizations of the parameter t [54].

Parametric Fourier series of degrees $n = 1, \dots, n_{\max}$, are fitted to the data and a degree that represents a compromise between fidelity and smoothness is established. The x -data and y -data can be treated separately since they are modelled by the functions (11.2) and (11.3), which have no parameter in common.

A goodness-of-fit measure is formed for each degree. A common measure, used here, is the chi-squared statistic χ_{obs}^2 , the sum of squares of the deviations of the fitted series of degree n from the data (x_i, y_i) , weighted inversely by the squared standard uncertainties associated with those values (ISO/TS 28037 [74] provides an analogous description for polynomial regression). When covariances associated with these data are present, a modified measure is used, but that is not the case here.

In terms of the computed Fourier series for degree n , the observed chi-squared value is

$$\chi_{\text{obs}}^2(n) = \sum_{i=1}^m \frac{[x_i - F_n(t_i)]^2 + [y_i - G_n(t_i)]^2}{u_0^2} = \frac{1}{12} \sum_{i=1}^m \{[x_i - F_n(t_i)]^2 + [y_i - G_n(t_i)]^2\}$$

with degrees of freedom $\nu = 2m - 4n - 2$ (the number of data less the number of adjustable parameters), from which is obtained the root-mean-square residual RMSR for each n given by

$$\left[\frac{\chi_{\text{obs}}^2(n)}{2m - 4n - 2} \right]^{1/2}. \quad (11.5)$$

11.3.9 In addition to using these RMSR-values to choose a value of n , various model-selection criteria such as Akaike's Information Criterion (AIC), corrected Akaike's Information Criterion (AICc) and the Bayesian Information Criterion (BIC) [22] are used. Generically, for m data points and a model with $n + 1$ parameters, these criteria are given by

$$\begin{aligned} \text{AIC: } & \chi_{\text{obs}}^2(n) + 2(n + 1), \\ \text{AICc: } & \chi_{\text{obs}}^2(n) + 2(n + 1) + 2(n + 1)(n + 2)/(m - n - 2), \\ \text{BIC: } & \chi_{\text{obs}}^2(n) + (n + 1) \ln(m). \end{aligned}$$

In this application, there are $2m$ data, the x - and the y -values, and $4n + 2$ parameters in the functions (11.2) and (11.3). So, the generic m and n in the above information criteria and expression (11.5) for RMSR and S are replaced by $2m$ and $4n + 2$, respectively.

11.3.10 All three information criteria are designed to balance reproducing the data and simplicity of model [93]. Given a number of candidate models, here parametric Fourier series of degrees $n = 1, \dots, n_{\max}$, the model having the smallest value of AIC (or AICc or BIC) would usually be selected. According to [75], experience with polynomial calibration problems indicates that the same degree of polynomial is often selected by all three criteria, although there may be some exceptions such as when the data set is small.

11.4 Model selection

11.4.1 Consider the profile of the tumour section shown in the left panel (85 pixels) of figure 11.1. The midpoints of the boundary pixels, each of unit size, are given by the $m = 38$ coordinate pairs with the bottom-left corner of the grid corresponding to a location (1, 1) and, consequently, the top-left corner pixel of the tumour outline corresponding to location (3, 12).

11.4.2 Consider approximations to this set of discrete data points by continuous profiles in the form of curves represented by the parametric Fourier functions (11.2) and (11.3) with the values of t taken as cumulative chord length as in clause 11.3.

11.4.3 Parametric Fourier series (11.4) with $n = 1, 2, \dots, 15$ harmonics were fitted to the data by least squares. Ordinary least squares was used because of the common uncertainty. The corresponding values of RMSR, AIC, AICc and BIC, and the estimate \hat{A} of area A , were calculated and given in table 11.1. Figure 11.2 displays the parametric Fourier series for these degrees.

Table 11.1: Root-mean-square residuals and information criteria for various numbers, n , of harmonics and tumour section area estimates \hat{A} in the example of the coarse-pixelated tumour image in figure 11.1 (left)

n	RMSR	AIC	AICc	BIC	\hat{A}
1	0.96	38	39	46	65.1
2	0.63	<u>23</u>	<u>25</u>	<u>34</u>	63.9
3	0.60	25	30	40	64.2
4	0.43	23	31	41	65.7
5	0.43	27	39	48	65.7
6	0.39	30	48	55	65.2
7	0.40	33	59	62	65.2
8	0.39	37	73	68	65.4
9	0.40	41	90	76	65.4
10	0.41	45	112	83	65.4
11	0.43	49	141	90	65.3
12	0.44	52	180	97	65.3
13	0.44	56	236	104	65.3
14	0.40	59	325	110	65.2
15	0.37	63	485	117	65.4

11.4.4 The intention is to select an appropriate number of harmonics that offers a balance between accounting for the data uncertainties and yielding an acceptably smooth representation of the profile. For the coarsely pixelated image, the sequence of RMSR values (column 2 of table 11.1) initially shows a tendency to decrease to a minimum value of approximately 0.4 when $n = 6$ (six harmonics). Then, the sequence saturates at about that level for a few degrees, subsequently tending to behave erratically for larger n due in part to the errors in the data being followed too closely. The Fourier series of degree 6 was selected, for which the RMSR value is 0.39. This value is framed in table 11.1 as is the corresponding area estimate of 65.2. Curiously, the three information criteria (AIC, AICc and BIC) all exhibit minimum values for $n = 2$, shown underlined in table 11.1, for which the RMSR value of 0.63 is some 50 % higher than that, 0.39, for $n = 6$.

11.4.5 This observation about the information criteria is not consistent with general experience of selecting an appropriate degree in polynomial calibration studies. Further, this choice of degree ($n = 2$) does not seem particularly consistent with the data on which it was based as seen from the poor fit when this number of harmonics are used to fit the

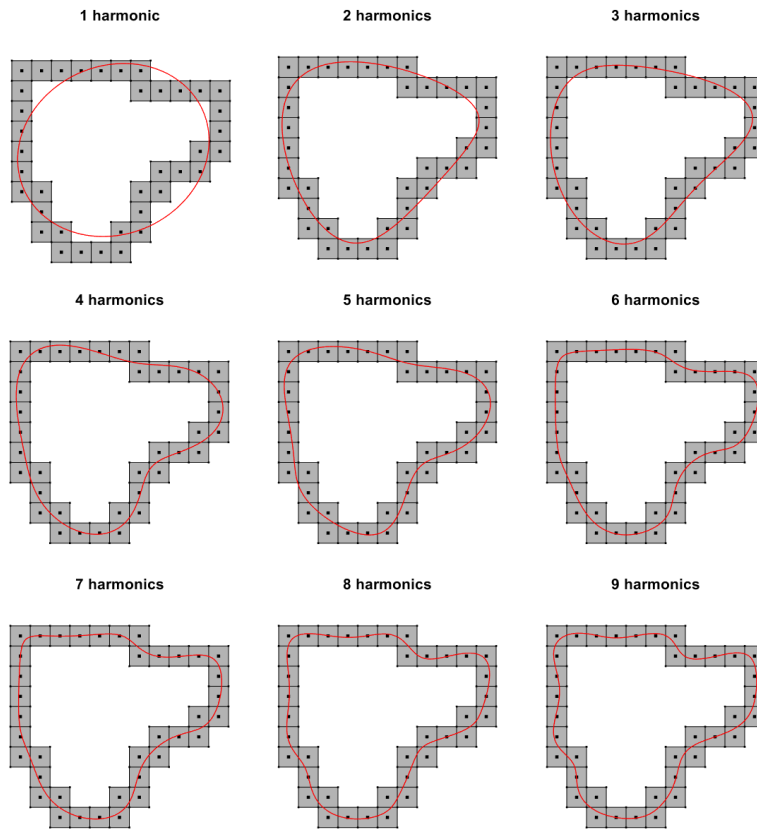


Figure 11.2: Fourier representations of a section of a tumour using 1 to 9 Fourier harmonics

tumour outline (figure 11.2). For instance, the re-entrant feature in the ‘north-east’ region of the outline is not reproduced. This and other features, however, are better reproduced by Fourier series of degree $n \geq 5$. By the time $n = 15$ is reached, spurious effects (not shown), due to following the ‘noise’ in the data, appear in the curves that are not apparent in the perceived outline of the tumour.

11.4.6 It would be unwise to select a low degree such as $n \leq 3$ since the corresponding Fourier series over-smooth the tumour outline. At the other extreme, the choice of a large number of harmonics such as 13 or greater means that the Fourier representations follow 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 between 5 and 9 might be suitable. Degree six was selected for the reasons above.

11.4.7 Recently [93], modified AIC and BIC criteria for model selection have come available based on incorporating the Shannon entropy in the criteria. Although their use for synthetic data sets containing 50 to 550 points has been demonstrated to provide better discrimination in the choice of model, to the knowledge of the JCGM they have not yet been applied in metrology.

11.5 Area estimation and associated uncertainty evaluation

11.5.1 The area enclosed by the curve modelled using a parametric Fourier series containing $n = 6$ harmonics is estimated to be 65.2 (see clause 11.4.4). For comparison, a

simplistic approach is to use the inner and outer profiles of the boundary pixels to yield minimum and maximum areas of 47 and 85, respectively, with a mean of 66, in close agreement, coincidentally, with the above value.

11.5.2 The measurement model for the area A of the tumour section based on the use of Fourier series is described in 11.3.3, 11.3.6 and 11.3.7 and used for the selected value, 6, of n . The input quantities in the model are realized by the x - and y -coordinates of the midpoints of each boundary pixel (unit square). The estimate of A in 11.5.1 is given by evaluating the area at those x - and y -values.

11.5.3 Consistent with 11.3.8, the x - and y -coordinates of the midpoints of each boundary pixel (unit square) are each modelled by a rectangular distribution, $R(-1/2, 1/2)$, relative to the midpoint. It is complicated to apply the GUF to propagate uncertainties in pixel midpoint locations through the measurement model. Accordingly, a Monte Carlo Method (MCM) [17] is adopted, which provides a relatively straightforward computational approach.

11.5.4 By sampling from these distributions, a new set of equally plausible (x, y) coordinates is generated. A Fourier series with $n = 6$ harmonics is computed, and the corresponding enclosed area is evaluated. Repeating this process a large number of times (10^5) yields a distribution of area values, whose mean serves as the Monte Carlo estimate and whose standard deviation represents the associated standard uncertainty.

11.5.5 The calculation in (11.5.4) gave the estimated PDF for tumour section area shown in figure 11.3. From the Monte Carlo results, the area estimate was 65.4 and the associated standard uncertainty was 1.6. This estimate is practically indistinguishable from the aforementioned point-estimate of 65.2 square pixels (table 11.1), indicating little bias in the process.

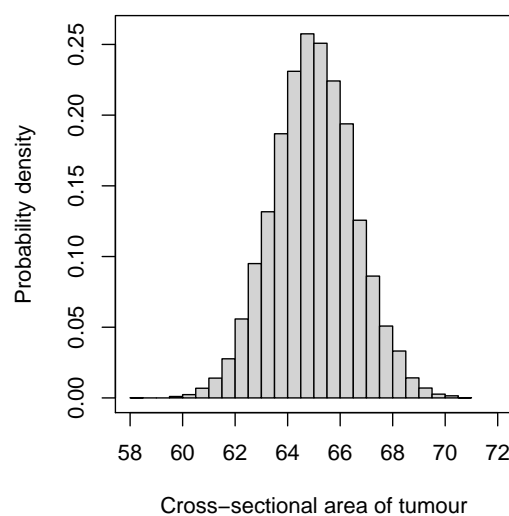


Figure 11.3: Scaled histogram approximation to the PDF for the cross-sectional area of the tumour given by Monte Carlo sampling using parametric Fourier series with six harmonics

11.6 Reporting the result

11.6.1 For the example presented in clause 11.4, two estimates of the area a of the considered section of the tumour and the associated standard uncertainties were provided. One result (clause 11.5.1) was

$$\hat{a} = 66, \quad u(\hat{a}) = 11$$

using the crude approach of taking the mean of the areas defined by the inner and outer perimeters of the shaded pixels in figure 11.1 (centre) and taking these two extremes as endpoints of a rectangular distribution of possible values. The second result was

$$\hat{a} = 65.0, \quad u(\hat{a}) = 1.6,$$

which was obtained by modelling the midpoints of the boundary pixels by a parametric Fourier series of degree six and employing Monte Carlo as in section 11.4.

11.6.2 To demonstrate stability of the results with respect to the model used, Monte Carlo method (MCM) results were also obtained for neighbouring degrees. For $n = 5$, the estimate 65.3 and standard uncertainty 1.5 were obtained. For $n = 7$, the corresponding values were 65.0 and 1.5. These estimates are consistent.

11.7 Interpretation of results

11.7.1 The volume or mass of an organ or tumour is generally obtained from a region of interest (ROI) outlined on anatomical imaging data [55]. 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 adopted here uses regression techniques to fit parametric Fourier functions to the midpoints of the pixels approximating the outline while taking into account that the midpoints are a crude guide to points on the boundary. A Fourier series of degree six appears to capture the given outline quite satisfactorily bearing in mind that only data points representing the midpoint of the boundary pixels are used.

11.7.2 In 11.6.1, the reported standard uncertainty of 1.6 for the area, derived via Fourier regression, is considerably smaller than the approximate value of 11 obtained by averaging the areas enclosed by the inner and outer perimeters of the shaded pixels in Figure 11.1 (left), and interpreting these bounds as the limits of a rectangular distribution. While this discrepancy may initially appear unexpected, it is important to note that the Fourier series provides a close approximation of the tumour boundary. The reduced uncertainty reflects the fidelity of this approximation and the associated reduction in variability.

11.7.3 The importance of the principles used to select a measurement model of an appropriate complexity cannot be overemphasized. Identification of the point where the RMSR values start to saturate with respect to model complexity is generally reliable [32, 34, 35, 39, 40, 42, 89]. That approach is better than the use of traditional information criteria such as AIC for the example considered here. The reason is believed to be the relative scarcity of profile boundary data, whereas the information criteria generally function better for large quantities of data. However, it is accepted that identification of where the RMSR values start to saturate is difficult to automate and there is much value in having the analysts involved in such decision making.

11.7.4 To illustrate the numerical accuracy of the parametric Fourier representation, the cross-sectional area of the fine pixelated image shown in figure 11.1 (right), taking each pixel as unit size, is also analyzed. This contour consists of $m = 113$ darkly shaded pixels, each of dimension 0.35×0.35 . Table 11.2 shows the results of fitting parametric Fourier series to these data along with RMSR and the other model selection criteria.

Table 11.2: Root-mean-square residuals, information criteria for various numbers, n , of harmonics and tumour section area estimates \hat{A} , in the case of a finer-pixelated image

n	RMSR	AIC	AICc	BIC	\hat{A}
1	1.28	91	92	101	65.6
2	0.76	39	40	53	64.8
3	0.67	35	39	54	65.2
4	0.42	<u>26</u>	<u>31</u>	<u>49</u>	66.1
5	0.43	30	37	56	66.1
6	<u>0.40</u>	33	43	63	<u>66.1</u>
7	0.39	36	50	71	66.1
8	0.39	40	58	78	66.0
9	0.40	44	68	86	66.0
10	0.41	48	78	94	66.0
11	0.41	51	90	102	66.1
12	0.41	55	103	110	66.0
13	0.42	59	119	118	66.0
14	0.40	62	137	125	65.9
15	0.40	66	158	133	66.0

The RMSR residuals again saturated essentially at degree 6. The information criteria were consistent in selecting degree 4, again different from that according to RMSR saturation, but better than for the coarser-pixel data, presumably because of the larger quantity of data involved. The corresponding area estimates in table 11.2 are corrected by a factor that accounts for the relative pixel size between the two images. For the finer image, the estimated area was 65.6 with an associated standard uncertainty of 0.8 provided by Monte Carlo. The area estimates are almost identical, but the standard uncertainty has halved.

11.7.5 The sequence of boundary coordinates might have associated correlation because there is pattern in the arrangement of the corresponding pixels as a consequence of underlying smoothness in the tumour boundary. This correlation has no tangible effect on the results presented here and is discussed in greater detail in [124, example E5.1], where information is also given on the extension of the approach to tumour volumes.

11.7.6 The Fourier-series approach works well for smooth tumour outlines, but struggles with sharp corners, which require very fine pixelation. Fourier series are inefficient for such features; piecewise methods like periodic splines [46] are more suitable, though less straightforward to implement.

12 Between-bottle homogeneity of reference materials

12.1 Preamble

12.1.1 This example is taken from a compendium of examples [1].

12.1.2 An essential element in the production of reference materials (RMs), certified reference materials (CRMs) and proficiency test materials in batch form is the evaluation of the between-bottle homogeneity [120]. In ISO 33405 [3], this kind of inhomogeneity is known as between-unit inhomogeneity and in ISO/IEC 17043 [71] it would be between-item inhomogeneity. This form of (in)homogeneity accounts for the (small) differences in the property of interest between the bottles (or more generally, items [3]) 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 [120,121].

12.1.3 This example shows how a Bayesian hierarchical model can be used to determine the between-bottle standard deviation of the amount fraction nitrogen in a set of synthetic natural gas mixtures. The model is demonstrated for a data set involving nitrogen where the classical analysis does not provide a solution. Any within-bottle homogeneity can contribute to the dispersion of the measured values for the items subject to the homogeneity study. In the case of gas mixtures and solutions, such effects are usually small. Evaluation of within-bottle homogeneity requires a different experimental approach and is not addressed in this example.

12.2 Specification of the measurand

12.2.1 The measurand is the amount fraction nitrogen in a set of 10 synthetic natural gas mixtures. This example focuses on determining an uncertainty component that characterizes the between-bottle homogeneity in a batch of a reference material [3]. The data from the homogeneity study are shown in table 12.1 [118]. For each of the 10 mixtures, 5 values for the amount fraction nitrogen were obtained (called ‘replicates’).

Table 12.1: Amount fraction of nitrogen in 10 gas mixtures expressed in cmol mol^{-1}

Mixture	Replicate 1	Replicate 2	Replicate 3	Replicate 4	Replicate 5
D520472	0.424 577	0.425 167	0.425 379	0.424 522	0.424 805
H95396	0.425 572	0.425 411	0.423 638	0.425 301	0.424 527
VSL190663	0.424 152	0.425 517	0.425 638	0.424 207	0.425 135
D520467	0.426 320	0.424 672	0.425 211	0.425 533	0.425 864
D520834	0.424 855	0.425 079	0.425 413	0.424 729	0.424 725
D520361	0.425 104	0.424 773	0.426 424	0.424 266	0.424 632
D520270	0.425 750	0.424 917	0.424 779	0.425 086	0.425 318
D520446	0.425 547	0.426 483	0.424 631	0.425 968	0.424 620
VSL190485	0.426 326	0.424 646	0.425 205	0.426 302	0.425 020
VSL190977	0.425 968	0.424 069	0.425 988	0.425 489	0.423 936

12.2.2 Traditionally, classical analysis of variance (ANOVA) is used [3, 69] for this purpose, which is more fully described in [120]. The parameter of interest is the between-group standard deviation, which in this specific case is called the between-bottle standard deviation [118, 120]. 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 [118]. Such situations should be avoided [3, 120] but that is not always possible [118].

12.2.3 Using the traditional ANOVA, the following values of sum-of-squares (SS), degrees of freedom (ν), and mean squares (MS) are obtained (see table 12.2).

Table 12.2: Analysis of variance results for the homogeneity data for the amount fraction nitrogen

	SS/(cmol mol ⁻¹) ²	ν	MS/(cmol mol ⁻¹) ²
Between groups	2.92×10^{-6}	9	3.25×10^{-7}
Within groups	1.96×10^{-5}	40	4.90×10^{-7}

12.2.4 The standard deviation corresponding to the between-group effect is, in this case,

$$s_{\text{between}} = \sqrt{\frac{MS_{\text{between}} - MS_{\text{within}}}{n}},$$

where n denotes the number of members in one group. s_{between} cannot be computed since MS_{between} is smaller than MS_{within} .

12.2.5 The data set in table 12.1 suffers from poor repeatability, but that does not rule out that the between-bottle standard deviation is not negligible when compared to the uncertainty associated with the batch average for the amount fraction nitrogen [3]. The use of a Bayesian hierarchical model which, apart from the use of prior probability density distributions for the parameters, is similar to the traditional one-way ANOVA model widely used in the evaluation of homogeneity studies [120], but does not suffer from the shortcomings of ANOVA [57, 118].

12.3 Measurement model

12.3.1 The statistical model relating the observed amount fractions y_{ij} for mixture i ($i = 1 \dots 10$) and replicate j ($j = 1 \dots 5$) 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 [120]

$$y_{ij} = \mu + B_i + \varepsilon_{ij}. \quad (12.1)$$

12.3.2 The objective of the evaluation is to determine the between-group variance $\tau^2 = \text{Var}(B_i)$ and the within-group variance $\sigma^2 = \text{Var}(\varepsilon_{ij})$. τ is the counterpart of s_{between} introduced previously. If no pooling is used, then $\sigma_i^2 = \text{Var}(\varepsilon_{ij})$, that is, a variance is computed for each mixture. In a between-bottle homogeneity study, the assumption is

typically made that the within-group standard deviations are independent and identically distributed [3, 120]. In that case, σ is the pooled within-group standard deviation. The Bayesian approach presented here also makes such an assumption [118].

12.3.3 The third parameter, μ denotes the mean of the group means. It is comparable to the grand mean in classical ANOVA [116, 120, 121]. It generally plays a subordinate role in between-bottle homogeneity studies [3], although in the case of gas analysis, with a proper calibration of the gas chromatograph (GC) used for making the measurements, it could be used to assign batch values to the amount fractions.

12.3.4 In this example, a Bayesian model is used, which implies that a joint prior probability density function (PDF) should be chosen for the model parameters. In the case that the parameters are assumed to be mutually independent a priori, then this joint prior PDF can be factored as three PDFs, one for each of the parameters. These probability density functions are specified as follows:

$$\mu \sim N(\mu_0, \sigma_{\text{target}}^2), \quad (12.2)$$

$$\tau \sim \text{Cauchy}(0, \tau_0), \quad (12.3)$$

$$\sigma \sim \text{Cauchy}(0, \sigma_0), \quad (12.4)$$

where N denotes the normal distribution and Cauchy the Cauchy distribution. The Cauchy distribution is the Student t distribution with $\nu = 1$ degree of freedom [56, 57].

12.3.5 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 ($\mu_0 = 0.4250 \text{ cmol mol}^{-1}$). 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 [118, 119].

12.3.6 The prior for the between-bottle standard deviation τ is chosen to be the half-Cauchy distribution with location parameter 0 and scale parameter τ_0 . τ_0 is obtained from the specification for the production of the batch gas mixtures, which is usually larger than the value expected for τ . The half-Cauchy distribution concentrates half of the density between 0 and the scale parameter, taking only non-negative values [56].

12.3.7 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.

12.3.8 The likelihood is, conditionally on the parameters, a normal distribution [119]:

$$y_{ij} | \mu, \tau, \sigma \sim N(\mu, \tau^2 + \sigma^2). \quad (12.5)$$

12.4 Implementation

12.4.1 The Bayesian model as described in the previous sections is used with Bayes' rule. From the weakly informative priors for μ , τ , and σ (expressions (12.2)-(12.4)), using the data and the likelihood (equation (12.5)), a joint posterior for the model parameters is obtained. From this posterior, the value for τ , the between-bottle standard deviation, is calculated. The can be implemented in various computational environments supporting Markov Chain Monte Carlo method (MCMC). An implementation in R [107] using the package RStan [117] was used to obtain the posterior for μ , τ and σ [118].

12.4.2 In Stan [117] code, the statistical model of the between-bottle homogeneity study with pooling of the within-group standard deviations reads as (after [117])

```

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;
}

```

12.4.3 The model consists of the following blocks

1. `data`, declaring the data used by the model
2. `transformed data`, used here to rescale the data by dividing the observed amount fractions by their mean
3. `parameters`, declaring the model parameters and any auxiliary parameters needed for running the calculations
4. `transformed parameters`, declaring the scaled group means
5. `model`, specifying the Bayesian model in terms of the priors and the likelihood
6. `generated quantities`, declaring and computing the unscaled model parameters

In the `data` block, the variables are declared needed for transferring the data. In this block, the number of gas mixtures ('bottles') N (n) and the (maximum) number of replicates K (k) are declared, followed by the table with amount fractions y . `mu0`, `tau0` and `sigma0` are the (hyper)parameters of the priors assigned to μ , τ , and σ respectively.

The next block, `transformed data`, performs a rescaling on the data in y . The transformation consists of calculating the mean of all observed amount fractions and to use this to rescale the data (variable `scale`). The rescaled variables is $y_$. This transformation could also have been performed in R before transferring the data to the Bayesian model. Including it in the model enables for the user of the model to transfer the original data and, as it will be seen, the model also returns the unscaled model parameters (see the `generated quantities` block). The last line in the `transformed data` performs the rescaling of the data and is written in vectorised form, as this is the fastest way to perform the rescaling [26, 117].

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 [57]. These group means are declared in the block `transformed parameters`. The transformation of the parameters back and forth is done to enhance the performance of the MCMC [57, 117].

12.4.4 Experience from previous between-bottle homogeneity studies for nitrogen in natural gas has indicated that the relative repeatability standard deviation for the nitrogen amount fraction is 0.20 % and the specification for the between-bottle homogeneity is 0.30 %. The latter is interpreted as an expanded uncertainty with coverage factor $k = 2$, thus $\tau/\mu = 0.15\%$ and $\sigma/\mu = 0.20\%$.

12.4.5 Running the MCMC sampling with 25 000 iterations and a warm-up of 5000 iterations, using 4 chains [118, 119] yields the posterior PDF for μ , τ , and σ shown in figure 12.1 and 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 uncertainty due to the MCMC is given in the column `se_mean` and is smaller than the last digit stated in the columns `mean` and `sd`. Repeated runs of the MCMC are expected to reproduce these values with the number of decimals displayed.

12.4.6 A sensitivity analysis, aimed at assessing how influential the choices of the priors are, revealed that the posterior of τ is quite sensitive with respect to the choice of the scale parameter of the Cauchy distribution [118]. Yet, in this example, the scale parameter was chosen based on the specification of the performance of the preparation process. As this process had been used several times before, these specifications were deemed to be adequate for eliciting this prior.

12.4.7 Where traditional ANOVA fails at quantifying the between-bottle homogeneity effect, the Bayesian approach provides a probability density function for τ , from which the between-bottle standard deviation can be derived. Using Bayesian inference does not substitute for using measurement methods with good repeatability for homogeneity studies [3, 82, 120], yet it provides more information given the information at hand. Underrating of the between-bottle homogeneity standard deviation can still occur [83].

12.4.8 As the uncertainty evaluation in accordance with the GUM:1995 [10] and ISO 34405 [3] requires only a point estimate for the uncertainty component τ , there are several ways to obtain such an estimate from the posterior probability density function – the mean, the mode, or the median. Given the skewness of the posterior of τ/μ (figure 12.1),

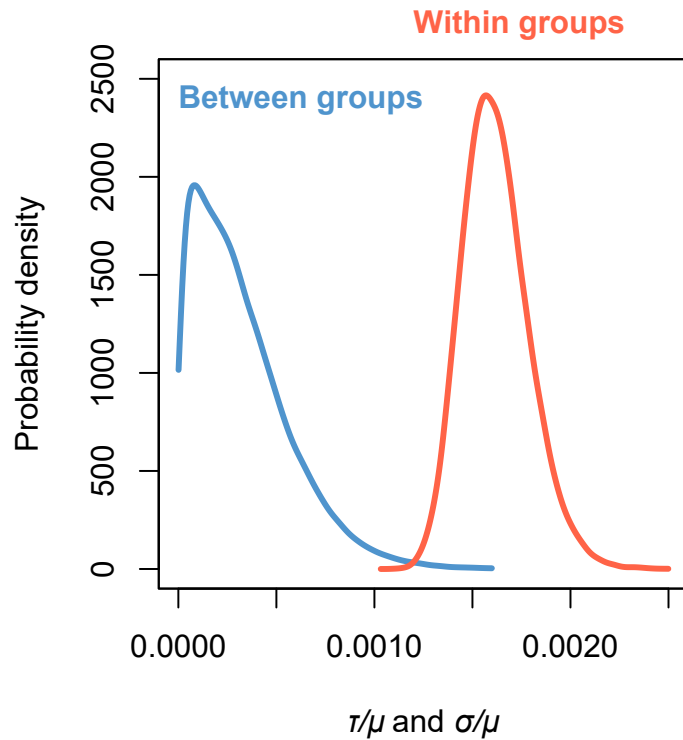


Figure 12.1: Posterior densities for the relative between-group and within-group standard deviations for the amount fraction of nitrogen

these three estimates are not equivalent: the mode of the posterior of τ is $0.3 \mu\text{mol mol}^{-1}$, the median is $1.1 \mu\text{mol mol}^{-1}$, and the mean is $1.4 \mu\text{mol mol}^{-1}$.

12.4.9 From the MCMC draws, the shortest coverage intervals can also be obtained. The following code, requiring the R packages `HDInterval` and `rstan` performs the calculation:

```
require(HDInterval)
require(rstan)
fit.mcmc = extract(fit)
tau.hdi = hdi(fit.mcmc$tau/fit.mcmc$mu, credMass = 0.95)
sigma.hdi = hdi(fit.mcmc$sigma/fit.mcmc$mu, credMass = 0.95)
```

12.4.10 The third line converts the output from Stan into the appropriate form and then the function `hdi` is used to compute the highest posterior density intervals [88]. The lower (L) and upper (H) limits of the 95 % highest posterior density intervals are shown in table 12.3.

12.5 Reporting the result

12.5.1 The primary result is the value for the between-bottle standard deviation τ . In a previous paper [118], the mean of the posterior for τ was chosen as estimate for τ , which is the most cautious option (it leads to the largest value for this uncertainty contribution). For data sets where the between-bottle variability is farther away from zero, the differences

Table 12.3: Highest posterior density intervals for the uncertainty components τ/μ and σ/μ

Parameter	$L_{95\%}$	$H_{95\%}$	Unit
τ/μ	0.0000	0.0008	1
σ/μ	0.0013	0.0020	1

between the three options become smaller.

12.5.2 Alternatives to using the mean include the use of the median or mode of the posterior probability density function for τ . The mean and median can be readily obtained from the default output from Stan, whereas the mode can be obtained using the following R code (v denotes the MCMC samples for τ):

```
# Create the function for calculating the mode
getmode <- function(v) {
  uniqv <- unique(v)
  uniqv[which.max(tabulate(match(v, uniqv)))]
}
```

Even using $\tau = 0$ could be justified, as the lower end of the 95 % highest posterior density interval is practically zero.

12.5.3 There are different choices for the credible interval as well. The example provides two intervals, the shortest in table 12.3 and the probabilistically symmetric in 12.4.5. Shortest intervals are generally more appropriate for skewed posteriors, as in the case of the posterior for τ .

12.5.4 To take full advantage of the Bayesian inference, the marginal posterior distribution of τ can be used in a subsequent uncertainty evaluation. It is difficult to find a point estimate that comprehensively represents the marginal posterior [57].

12.5.5 Whereas this example illustrates how a probability distribution for the between-bottle homogeneity standard deviation can be obtained if the between-bottle homogeneity effect is small in comparison to the repeatability of the measurement method, it does not substitute for adhering to the requirement in ISO 33405 [3] to select methods with good repeatability [82, 120].

13 Measurement of Celsius temperature using a resistance thermometer

13.1 Preamble

13.1.1 This example is provided by the JCGM-WG1.

13.1.2 This example is concerned with the measurement of Celsius temperature by comparing the resistance of an industrial platinum resistance thermometer with that of a standard resistor using a resistance bridge. The measurement of a single Celsius temperature (see 13.2) is described by an implicit univariate measurement model [17, Clause 4.1, note 2] and that of several temperatures (see 13.3) by an implicit multivariate measurement model [12, clause 6.3.1]. The example illustrates the treatment of these (univariate and multivariate) measurement models using the law of propagation of uncertainty (LPU) generalized to implicit measurement models and models with more than one measurand [10, 12].

13.1.3 The application of LPU in this example depends on the linearization of a non-linear (univariate or multivariate) measurement model. Ideally, either a statement should be made that any replacement of a non-linear measurement model by a linear model has been checked to be acceptable, or that such a check has not been made.

13.2 Measurement of a single Celsius temperature

13.2.1 In this example, the Celsius temperature t is measured by comparing the resistance $R(t)$ of a resistance thermometer with the resistance R_S of a standard resistor using a resistance bridge. The quantity measured with the resistance bridge is the resistance ratio

$$W(t) = \frac{R(t)}{R_S}. \quad (13.1)$$

In the temperature interval from 0 °C to 30 °C, the thermometer resistance $R(t)$ is approximated by a parabolic function of its Celsius temperature t :

$$R(t) = (1 + At + Bt^2)R_0. \quad (13.2)$$

Table 13.1: Estimates of the input quantities $\mathbf{X} \equiv (R_0, A, B, R_S, W(t))^T$ for the measurement of a single Celsius temperature and associated standard uncertainties (13.2.1, 13.2.2)

	\widehat{R}_0/Ω	$\widehat{A}/^\circ\text{C}^{-1}$	$\widehat{B}/^\circ\text{C}^{-2}$	\widehat{R}_S/Ω	$\widehat{W}(t)$
Estimate	99.996 10	0.003 909 6	-6.0×10^{-7}	99.999 47	1.078 005 7
Std unc	0.000 50	0.000 002 7	1.1×10^{-7}	0.000 10	0.000 005 0

13.2.2 Table 13.1 gives estimates of R_0 , A , B , R_S and $W(t)$ with the associated standard uncertainties. \widehat{R}_0 , \widehat{A} and \widehat{B} and associated uncertainties are determined from a previous calibration of the thermometer. \widehat{R}_S and the associated standard uncertainty are determined by a previous calibration of the reference resistor.

Table 13.2: Non-zero correlation coefficients associated with pairs of estimates of the input quantities for the measurement of a single Celsius temperature (13.2.1, 13.2.2 and 13.2.3)

	\widehat{R}_0	\widehat{A}	\widehat{B}
\widehat{R}_0	1	-0.155	0.092
\widehat{A}		1	-0.959
\widehat{B}			1

13.2.3 The non-zero correlation coefficients associated with pairs of estimates are given in table 13.2. R_S is independent of the parameters R_0 , A and B . $W(t)$ is independent of the parameters R_0 , A and B of the resistance thermometer and the resistance R_S of the standard resistor; there are thus no further non-zero correlation coefficients beyond those given in table 13.2.

13.2.4 By combining expressions (13.1) and (13.2), the following measurement model for the Celsius temperature t is obtained:

$$h(Y, X) = (1 + At + Bt^2)R_0 - W(t)R_S = 0. \quad (13.3)$$

The measurement model is an implicit univariate model with $N = 5$ input quantities $X \equiv (R_0, A, B, R_S, W(t))^T$ and a single output quantity (measurand) $Y = t$.

13.2.5 Although the above measurement model can be solved for t by considering the roots of a quadratic equation, the numerical evaluation of the transformed model can suffer from subtractive cancellation, and the evaluation of sensitivity coefficients is made more difficult by the transformation. Thus, measurement model 13.3 will be treated as implicit.

13.2.6 The estimate $y \equiv \widehat{t}$ of the Celsius temperature generating the measured resistance ratio $\widehat{W}(t)$ is found by inserting the estimates given in table 13.1 into equation 13.3 and solving this equation for t . The solution obtained is $\widehat{t} = 20.0232^\circ\text{C}$.

13.2.7 The standard uncertainty $u_y \equiv u(\widehat{t})$ associated with the estimate y is evaluated using [12, expression (8)], namely

$$C_y U_y C_y^T = C_x U_x C_x^T \quad (13.4)$$

with the matrices of first-order partial derivatives given as follows:

$$C_y = \frac{\partial h}{\partial Y} = \frac{\partial h}{\partial t} = (A + 2Bt)R_0,$$

and

$$C_x = \frac{\partial h}{\partial X} = \left[\frac{\partial h}{\partial R_0}, \frac{\partial h}{\partial A}, \frac{\partial h}{\partial B}, \frac{\partial h}{\partial R_S}, \frac{\partial h}{\partial W(t)} \right] = [1 + At + Bt^2, R_0 t, R_0 t^2, -W(t), -R_S].$$

These expressions are evaluated at the estimates of the input quantities given in table 13.1 and the corresponding estimate of the output quantity, which gives the following sensitivity matrices:

$$C_y = 0.389 \Omega^\circ\text{C}^{-1}$$

and

$$C_x = [1.078, 2.002 \times 10^3 \Omega^\circ\text{C}, 4.009 \times 10^4 \Omega^2\text{C}^2, -1.078, -99.999 \Omega].$$

The elements of the covariance matrix U_x associated with the estimates of the input quantities are calculated from the standard uncertainties in table 13.1 and the correlation coefficients in table 13.2. The result obtained after matrix algebra of expression (13.4) is $u(\hat{t}) = 0.0045^\circ\text{C}$.

13.3 Measurement of several Celsius temperatures

13.3.1 The resistance thermometer, standard resistor and resistance bridge described in 13.2 are used to measure the resistance ratios $W(t_1), \dots, W(t_{10})$ generated by ten Celsius temperatures t_1, \dots, t_{10} .

13.3.2 The estimates $\hat{R}_0, \hat{A}, \hat{B}$ and \hat{R}_S , and the associated standard uncertainties remain as before (table 13.1) and the estimates $\hat{W}(t_1), \dots, \hat{W}(t_{10})$ and their associated standard uncertainties are shown in table 13.3. The only non-zero correlation coefficients associated with pairs of the estimates remain those given in table 13.2. The resistance ratios are assumed to be independent, an assumption that is valid if the magnitudes of the random errors in the measured resistance ratios dominate.

Table 13.3: Estimates of the resistance ratios for the measurement of several Celsius temperatures and associated standard uncertainties (13.3.2)

	$\hat{W}(t_1)$	$\hat{W}(t_2)$	$\hat{W}(t_3)$	$\hat{W}(t_4)$	$\hat{W}(t_5)$
Estimate	1.000 005 3	1.015 005 4	1.030 005 5	1.045 005 6	1.060 005 6
Std unc	0.000 005 0	0.000 005 0	0.000 005 0	0.000 005 0	0.000 005 0
	$\hat{W}(t_6)$	$\hat{W}(t_7)$	$\hat{W}(t_8)$	$\hat{W}(t_9)$	$\hat{W}(t_{10})$
Estimate	1.078 005 7	1.090 005 8	1.105 005 9	1.120 006 0	1.078 005 7
Std unc	0.000 005 0	0.000 005 0	0.000 005 0	0.000 005 0	0.000 005 0

13.3.3 Each resistance ratio $W(t_j)$ is related to the corresponding Celsius temperature t_j by an expression of the form (13.3):

$$(1 + At_j + Bt_j^2)R_0 - W(t_j)R_S = 0, \quad j = 1, \dots, 10. \quad (13.5)$$

There are $N = 14$ input quantities $X \equiv (R_0, A, B, R_S, W(t_1), \dots, W(t_{10}))^\top$ and $m = 10$ output quantities $Y \equiv (t_1, \dots, t_{10})^\top$, all of which are related by the following implicit multivariate measurement model:

$$\begin{bmatrix} h_1(Y, X) \\ \vdots \\ h_{10}(Y, X) \end{bmatrix} \equiv \begin{bmatrix} R_0(1 + At_1 + Bt_1^2) - W(t_1)R_S \\ \vdots \\ R_0(1 + At_{10} + Bt_{10}^2) - W(t_{10})R_S \end{bmatrix} = \mathbf{0}.$$

As noted in clause 13.2.5, transformation of the implicit measurement model (13.5) into an explicit form is possible but inadvisable.

Table 13.4: Estimates of the output quantities Y and associated standard uncertainties for the measurement of several Celsius temperatures (13.3.4 and 13.3.5)

	\hat{t}_1	\hat{t}_2	\hat{t}_3	\hat{t}_4	\hat{t}_5	\hat{t}_6	\hat{t}_7	\hat{t}_8	\hat{t}_9	\hat{t}_{10}
Estimate/ $^{\circ}\text{C}$	0.0100	3.8491	7.6928	11.5410	15.3938	20.0232	23.1131	26.9797	30.8509	20.0232
Std. unc./ $^{\circ}\text{C}$	0.0018	0.0027	0.0040	0.0046	0.0047	0.0045	0.0046	0.0060	0.0089	0.0045

13.3.4 The estimates $y = (\hat{t}_1, \dots, \hat{t}_{10})$ of the Celsius temperatures Y are given by inserting the estimates given in columns 1 to 4 of table 13.1 and in table 13.3 into equations (13.5) and solving these equations. These estimates are given in table 13.4.

13.3.5 The covariance matrix C_y associated with y is evaluated using expression (13.4) where C_y and C_x are sensitivity matrices given by evaluating C_Y and C_X , respectively, at the estimates of the input and output quantities. C_Y is a diagonal matrix of dimension 10×10 with diagonal entries $R_0(A + 2Bt_1), \dots, R_0(A + 2Bt_{10})$. C_X is a matrix of dimension 10×14 given by

$$C_X = \begin{bmatrix} C_X^{(1)} & C_X^{(2)} \end{bmatrix},$$

where

$$C_X^{(1)} = \begin{bmatrix} 1 + At_1 + Bt_1^2 & R_0t_1 & R_0t_1^2 & -W(t_1) \\ \vdots & \vdots & \vdots & \vdots \\ 1 + At_{10} + Bt_{10}^2 & R_0t_{10} & R_0t_{10}^2 & -W(t_{10}) \end{bmatrix}$$

is a matrix of dimension 10×4 , and $C_X^{(2)}$ is a diagonal matrix of dimension 10×10 with diagonal elements all equal to $-R_S$. The covariance matrix U_x is calculated from the standard uncertainties given in columns 1 to 4 of table 13.1 and in table 13.2. The standard uncertainties associated with the estimates of the Celsius temperatures and the correlation coefficients associated with pairs of estimates, derived from the matrix C_y , are given in table 13.4 and table 13.5, respectively. A numerically stable method to form the covariance matrix C_y is given in [12, annex B].

Table 13.5: Correlation coefficients associated with pairs of estimates of the output quantities Y for the measurement of several Celsius temperatures (13.3.5)

	\hat{t}_1	\hat{t}_2	\hat{t}_3	\hat{t}_4	\hat{t}_5	\hat{t}_6	\hat{t}_7	\hat{t}_8	\hat{t}_9	\hat{t}_{10}
\hat{t}_1	1	0.252	0.127	0.079	0.059	0.054	0.056	0.054	0.050	0.054
\hat{t}_2		1	0.815	0.800	0.755	0.580	0.312	-0.092	-0.358	0.580
\hat{t}_3			1	0.902	0.868	0.691	0.400	-0.057	-0.365	0.691
\hat{t}_4				1	0.909	0.766	0.495	0.040	-0.281	0.766
\hat{t}_5					1	0.847	0.629	0.208	-0.115	0.847
\hat{t}_6						1	0.841	0.549	0.264	0.918
\hat{t}_7							1	0.812	0.613	0.841
\hat{t}_8								1	0.909	0.549
\hat{t}_9									1	0.264
\hat{t}_{10}										1

13.3.6 The results given in table 13.4 and figure 13.1 show how the standard uncertainty $u(\hat{t}_j)$ varies with the estimate \hat{t}_j of Celsius temperature t_j . The uncertainty is smallest around 0 °C and increases rapidly for temperatures greater than 25 °C. This effect is due to the fact that the resistance thermometer was calibrated at the temperatures 0 °C, 15 °C, 20 °C and 25 °C, and that the Celsius temperature 0 °C during calibration was generated using an ice bath with a standard uncertainty three times smaller than those associated with the other three temperatures, which were generated using an oil bath.

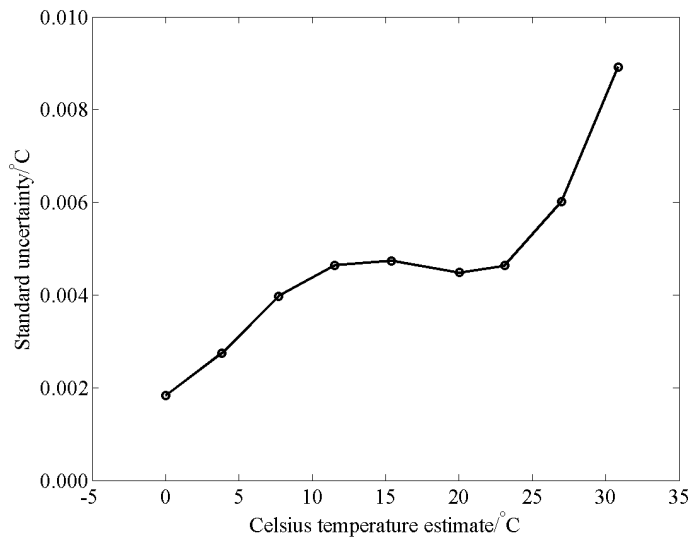


Figure 13.1: Standard uncertainty $u(\hat{t}_j)$ associated with the estimate \hat{t}_j of Celsius temperature t_j (13.3.6)

NOTE In figures 13.1 and 13.2 the straight-line segments joining the plotted points are included for purpose of visualization.

13.3.7 Using the results given in the final column of table 13.5, figure 13.2 shows how the correlation coefficient associated with the pair of estimates \hat{t}_j and \hat{t}_{10} of Celsius temperatures t_j and $t_{10} = 20$ °C varies with \hat{t}_j , $j = 1, \dots, 9$. The correlation coefficient has a maximum for $\hat{t}_j = \hat{t}_6$ and approaches zero as the absolute difference $|\hat{t}_j - \hat{t}_{10}|$ becomes large. The example demonstrates that quantities measured with the same instrument can be highly correlated.

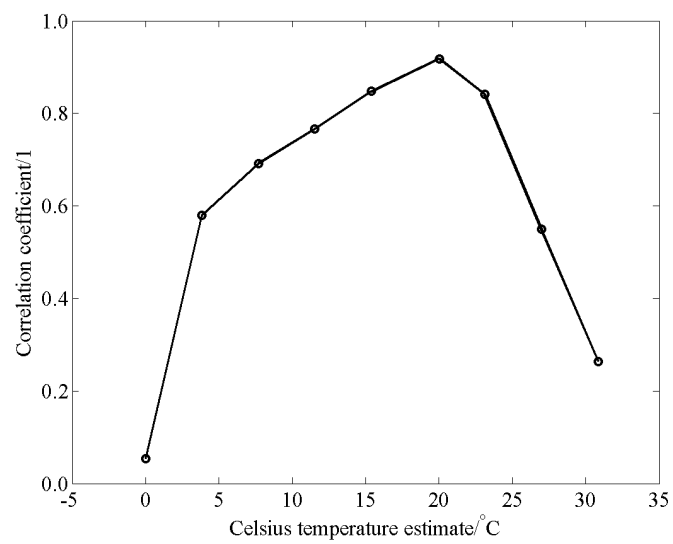


Figure 13.2: Correlation coefficient associated with the pair of estimates \hat{t}_j and \hat{t}_{10} of Celsius temperatures t_j and $t_{10} = 20^\circ\text{C}$ (13.3.7)

14 Activity of a radioactive source corrected for decay

14.1 Preamble

14.1.1 This example is provided by the BIPM.

14.1.2 A ^{207}Bi (bismuth-207) standard is used at a time t_1 to calibrate a detector. The calibration certificate of this radioactive source reports an activity \hat{A}_0 at the reference date t_0 , with associated standard uncertainty $u(\hat{A}_0)$.

14.1.3 This example provides an estimate \hat{A}_1 of the activity A_1 at time t_1 and evaluates the associated standard uncertainty $u(\hat{A}_1)$, illustrating the application of the law of propagation of uncertainty (LPU). Because the model is non-linear in time and half-life $T_{1/2}$, the example also verifies whether the LPU is adequate for this task.

14.2 Measurement model

The measurement model for the activity A_1 at time t_1 is

$$A_1 = f(A_0, T_{1/2}, \Delta t) = A_0 \exp\left(-\frac{\Delta t \ln 2}{T_{1/2}}\right), \quad (14.1)$$

with $\Delta t = t_1 - t_0$. Both quantities t_0 and t_1 , and thus Δt , are considered to have negligible uncertainty.

Introducing the *decay correction* $E = \exp\left(-\frac{\Delta t \ln 2}{T_{1/2}}\right)$,

$$A_1 = A_0 E. \quad (14.2)$$

14.3 Uncertainty evaluation

14.3.1 Estimates \hat{A}_0 and $\hat{T}_{1/2}$ of A_0 and $T_{1/2}$ and their associated standard uncertainties are taken from a calibration certificate and the literature [24], respectively:

$$\begin{aligned} \hat{A}_0 &= 11.450 \text{ kBq}, & u(\hat{A}_0) &= 0.050 \text{ kBq}, \\ \hat{T}_{1/2} &= 32.9 \text{ a}, & u(\hat{T}_{1/2}) &= 1.4 \text{ a} \end{aligned}$$

and Δt is taken as $\Delta t = 2.73 \text{ a}$. The symbol a is used for the year in accordance with ISO 80000-3:2020 [77]. No further information is taken into consideration.

14.3.2 Substitution of the above values into measurement model (14.1) gives

$$\hat{A}_1 = 11.450 \text{ kBq} \times \exp\left(\frac{-2.73 \text{ a} \times \ln 2}{32.9 \text{ a}}\right) = 0.944 11 \hat{A}_0 = 10.810 \text{ kBq}.$$

Applying the LPU [10, Clauses 5.1.2, 5.1.3] to measurement model (14.1) gives

$$u^2(\hat{A}_1) = u_{\hat{A}_0}^2(\hat{A}_1) + u_{\hat{T}_{1/2}}^2(\hat{A}_1) \quad \text{with} \quad u_{\hat{A}_0}(\hat{A}_1) = |c_{\hat{A}_0}| u(\hat{A}_0), \quad u_{\hat{T}_{1/2}}(\hat{A}_1) = |c_{\hat{T}_{1/2}}| u(\hat{T}_{1/2}).$$

$c_{\hat{A}_0}$ and $c_{\hat{T}_{1/2}}$ are sensitivity coefficients given by differentiating the measurement model (14.1), namely,

$$c_{\hat{A}_0} = \frac{\partial f}{\partial A_0} = E = 0.944\,11,$$

$$c_{\hat{T}_{1/2}} = \frac{\partial f}{\partial T_{1/2}} = \hat{A}_1 \frac{\Delta t \ln 2}{\hat{T}_{1/2}^2} = 10.810\,\text{kBq} \times \frac{2.73\,\text{a} \times \ln 2}{(32.9\,\text{a})^2} = 0.018\,898\,\text{kBq a}^{-1},$$

where the partial derivatives are evaluated at the estimates \hat{A}_0 and $\hat{T}_{1/2}$. Thus,

$$u_{\hat{A}_0}(\hat{A}_1) = 0.944\,11 \times 0.050\,\text{kBq} = 0.0472\,\text{kBq},$$

$$u_{\hat{T}_{1/2}}(\hat{A}_1) = 0.018\,898\,\text{kBq a}^{-1} \times 1.4\,\text{a} = 0.0265\,\text{kBq}$$

and hence

$$u(\hat{A}_1) = [(0.0472)^2 + (0.0265)^2]^{1/2}\,\text{kBq} = 0.054\,\text{kBq}.$$

14.3.3 To investigate whether the LPU is adequate in this case, the behaviour of the non-linear model (14.1) is compared with the linearized model in the neighbourhood of $\hat{T}_{1/2} = 32.9\,\text{a}$. The linearized model is

$$A'_1 = f_{\text{lin}}(A_0, T_{1/2}) = \hat{A}_1 + c_{\hat{A}_0}(A_0 - \hat{A}_0) + c_{\hat{T}_{1/2}}(T_{1/2} - \hat{T}_{1/2}). \quad (14.3)$$

The criterion used here for adequacy of LPU is that the values of the non-linear model (14.1) and the linearized model (14.3) differ negligibly in the neighbourhood of $\hat{T}_{1/2}$. This neighbourhood is taken as comprising values of the half-life $T_{1/2}$ up to several multiples of $u(\hat{T}_{1/2})$ away from its estimate. Accordingly, the linearized and the exact model values are tabulated at several positive and negative multiples of $u(\hat{T}_{1/2}) = 1.4\,\text{a}$ away from $\hat{T}_{1/2} = 32.9\,\text{a}$. See table 14.1. The approach used is comparable to that in [15, annex F.2]; see also [10, clause 5.1.3, note 2].

Table 14.1: Estimates \hat{A}'_1 and \hat{A}_1 from linearized and exact models for ^{207}Bi decay, respectively

$\hat{T}_{1/2}/\text{a}$	\hat{A}'_1/kBq	\hat{A}_1/kBq	Difference/kBq
28.7	10.731	10.719	0.011
30.1	10.757	10.752	0.005
31.5	10.784	10.782	0.001
32.9	10.810	10.810	0.000
34.3	10.836	10.835	0.001
35.7	10.863	10.859	0.004
37.1	10.889	10.881	0.009

14.3.4 As seen from table 14.1, the values provided by the linearized and non-linear models differ at most by 0.011 kBq and 0.009 kBq at $T_{1/2} + 3u(T_{1/2})$ and $T_{1/2} - 3u(T_{1/2})$, respectively. These differences are much smaller than $u(\hat{A}_0) = 0.050\,\text{kBq}$ and $u(\hat{A}_1) = 0.054\,\text{kBq}$ and therefore it can be concluded that the validity of LPU is confirmed in this case.

14.4 Discussion

14.4.1 The time interval $\Delta t = 2.73$ a is small compared with the radionuclide half life $\hat{T}_{1/2} = 32.9$ a. As a consequence, the value 0.944 11 of E in expression (14.1) is reasonably close to unity, and so $u(\hat{A}_1)$ differs little from $u(\hat{A}_0)$.

14.4.2 When the time difference Δt becomes closer to $\hat{T}_{1/2}$, LPU may not provide acceptable results for ^{207}Bi as judged by such a check. This happens because the relative uncertainty of its half-life of 4% is large and the range of values from $\hat{T}_{1/2} - 3u(\hat{T}_{1/2})$ to $\hat{T}_{1/2} + 3u(\hat{T}_{1/2})$ is more likely to highlight non-linearities. However, in most cases encountered in radionuclide metrology and its applications, half-lives are known more precisely (relative uncertainty of 0.5% or better) and LPU can be used safely unless $\Delta t/\hat{T}_{1/2}$ is very large like in ^{14}C dating of very old samples.

When $\Delta t/\hat{T}_{1/2}$ and the relative uncertainty of the half-life $u(\hat{T}_{1/2})/\hat{T}_{1/2}$ are both sufficiently small, LPU can be used safely. In practice, the relative uncertainty of the decay correction,

$$u(E)/E = \ln 2 [u(\hat{T}_{1/2})/\hat{T}_{1/2}] \Delta t/\hat{T}_{1/2},$$

is a good measure of the non-linearity of the model. It can be shown that in general, when the relative uncertainty of the decay correction $u(E)/E$ reaches about 1%, it is important to check the validity of LPU as shown in the present example and compare the difference between the linearized and exact models with the uncertainty of the activity in order to decide whether the non-linear effect on LPU is significant or not. In the affirmative, JCGM 101:2008 [17], concerned with the propagation of distributions using a Monte Carlo method, should be applied.

15 Breaking force of steel wire rope

15.1 Preamble

15.1.1 This example is provided by the NLA, South Africa.

15.1.2 With thousands of workers transported up and down mine shafts each day (figure 15.1), measuring the safety of hoisting ropes forms an integral part of mine management. Strict laws cover the maintenance and testing of steel wire rope, the latter focussing mainly on the destructive testing of these ropes. The source of this information and the example is the Rope Testing Laboratory of the Council for Scientific and Industrial Research, South Africa (CSIR), from whom permission was obtained to use the example.



Figure 15.1: Steel wire rope used for hoisting in vertical mine shafts

15.2 Measurement model

15.2.1 The measurand is the breaking force F_0 of a sample constituting the length of a piece of rope of a particular type. An observation of the breaking force cannot be repeated since the test is destructive on the sample. Because it is impossible to make a series of observations of the measurand under repeatability conditions, one observation is instead made of each of a number, n , of samples given by cutting pieces from the same rope. The pragmatic approach is therefore taken of testing six pieces from the same coil of rope and accepting that the variability across the six specimens will be embedded in the variability of the test results.

15.2.2 The breaking forces that are recorded are influenced by two effects: the variability of the samples due to the inhomogeneity of the rope, and intrinsic measurement effects, essentially measurement corrections, such as the resolution of the force measuring system. Following the provisions of JCGM GUM-6:2020 [15], $n = 6$ breaking force observation equations are

$$F_i = F_0 + E_i + B + B_{\text{ref}} + e_{\text{res}} + e_T, \quad i = 1, \dots, n, \quad (15.1)$$

where

the F_i are observed breaking forces,

F_0 is the true value of the measurand, the ‘true’ breaking force,

E_i are the (unobservable) measurement errors associated with the F_i ,

B is the error of the test machine load cell,

B_{ref} is the error of the reference load cell,

e_{res} is the resolution error of the force measuring system,

e_T is the error associated with the temperature.

15.3 Assignment of PDFs

15.3.1 Breaking force observations

The available data comprise $n = 6$ observations F_i ,

10 006 kN, 10 007 kN, 10 005 kN, 10 008 kN, 10 003 kN, 10 005 kN,

whose variability is ascribed to the associated random errors E_i , which are modelled as (independent) realisations from a random variable E having mean 0 kN and unknown standard deviation σ . The sample mean is $\hat{F} = 10\,005.7$ kN and the associated uncertainty, taken as the sample standard deviation of the mean, is $u(\hat{F}) \equiv s(\hat{F}) = 1.8 \text{ kN} / \sqrt{6} = 0.7$ kN. Figure 15.2 shows a normal distribution with mean 10 005.7 kN and standard deviation 0.7 kN as the full blue curve. An alternative choice of distribution, the Weibull, displayed as the broken black curve, is considered in clause 15.6. Using the provisions of [17, clause

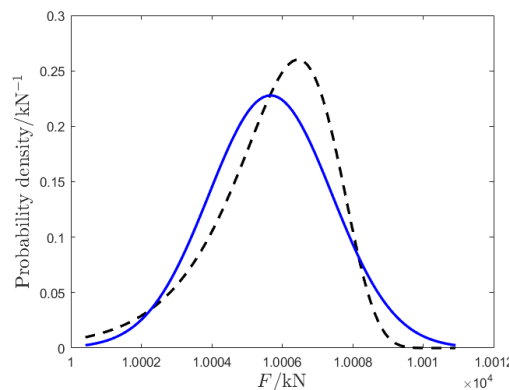


Figure 15.2: Normal distribution (full blue curve) and Weibull distribution (broken black curve) fitted to the breaking force data

6.4.7]), the mean 10 005.7 kN and standard deviation 0.7 kN with $\nu = n - 1 = 5$ degrees of freedom are used as the parameters of the resulting t -distribution for breaking force F .

15.3.2 Correction effects

All other input quantities, representing measurement corrections, are modelled on the basis of available information with rectangular distributions centred at zero and having specified semi-widths: see table 15.1 (data provided by NLA, South Africa).

Table 15.1: Rectangular PDFs assigned to correction effects for the rope breaking-force model on the basis of available information

Quantity	Semi-width/kN	Standard uncertainty/kN
B	6.8	3.9
B_{ref}	8.2	4.7
e_{res}	8.7	5.0
e_T	10.2	5.9

15.4 JCGM 100:2008 uncertainty framework (GUF)

15.4.1 A trivial rearrangement of expression (15.1) gives

$$F_0 = F_i - E_i - B - B_{\text{ref}} - e_{\text{res}} - e_T, \quad (15.2)$$

an overdetermined system of equations. A classical solution, under the assumption that the terms on the right-hand side are independent, yields the measurand estimate as

$$\hat{F}_0 = \frac{1}{6} \sum_{i=1}^6 F_i - B - B_{\text{ref}} - e_{\text{res}} - e_T = 10\,005.7 \text{ kN}. \quad (15.3)$$

Applying the law of propagation of uncertainty (LPU) to expression 15.3 yields

$$u(\hat{F}_0) = [u^2(\hat{F}) + u^2(\hat{B}) + u^2(\hat{B}_{\text{ref}}) + u^2(\hat{e}_{\text{res}}) + u^2(\hat{e}_T)]^{1/2} = 9.9 \text{ kN}. \quad (15.4)$$

Since all but one of the input quantities have infinite degrees of freedom supporting their standard uncertainty estimates, the Welch-Satterthwaite formula [10, expression (G.2a)] in JCGM 100:2008 yields

$$\nu_{\text{eff}} = (n-1) \left[\frac{u(\hat{F}_0)}{u(\hat{F})} \right]^4 > 10^5.$$

As a consequence, the measurement result is a breaking force $\hat{F}_0 = 10\,005.7 \text{ kN}$ with an expanded uncertainty $U(\hat{F}_0) = 19.4 \text{ kN}$ corresponding to a coverage factor k of 1.96 for a coverage probability of 95 %.

15.4.2 According to the GUF, a t -distribution with ν_{eff} degrees of freedom [10, annex G.3] is assigned as the PDF for F_0 . Since ν_{eff} is so large, the t -distribution could be taken as normal for practical purposes, as demonstrated by the coverage factor being equal to three significant decimal digits to that for a normal distribution for the same coverage probability.

15.5 Monte Carlo method (MCM)

15.5.1 To validate the results of clause 15.4, MCM was applied following the provisions of [17]. Its application in this example consists of taking 10^6 draws from the probability distributions considered in clause 15.4 and using the measurement model for each set of draws to provide a corresponding value for the measurand F_0 . For this simple measurement model, the Monte Carlo calculation to provide this sample of the probability distribution for F_0 took 0.2 s using MATLAB version 'R2023b Update 3 (23.2.0.2409890) 64-bit (win64) October 4, 2023'.

15.5.2 The mean and standard deviation of this MCM sample of F_0 are 10 005.7 kN and 10.1 kN, respectively, and the probabilistically symmetric 95 % coverage interval whose endpoints are the 2.5 % and 97.5 % quantiles of the sample ranged from 9986.0 kN to 10 025.3 kN. Since this interval is almost exactly symmetric about the mean, half the length of this interval is taken as an expanded uncertainty for 95 % coverage, namely, $U(\hat{F}_0) \approx 19.6$ kN.

15.5.3 These results were confirmed to the number of significant decimal digits quoted by re-running the Monte Carlo calculation with 10^7 trials, which took 1.6 s.

15.6 Discussion

15.6.1 The results provided by the GUF and MCM, both having estimate 10 005.7 kN and associated standard uncertainties 9.9 kN and 10.1 kN, respectively, compare well, although the GUF probability density function (PDF) is slightly more peaked than that provided by MCM: figure 15.3. Since the Monte Carlo method can be regarded as a reference [41], judging by the considerable extent of agreement with the results from applying JCGM 100:2008 it can be concluded that the latter for the above set of data is valid.

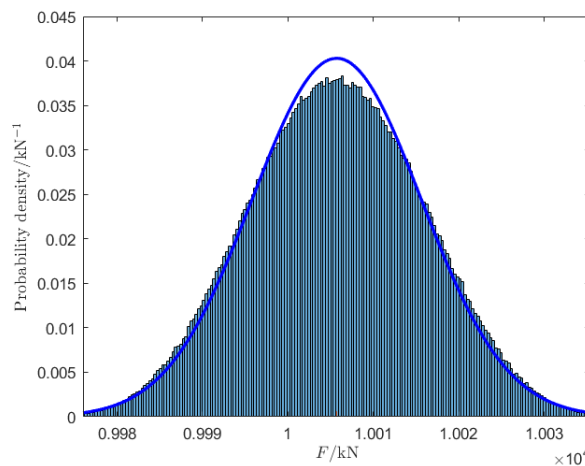


Figure 15.3: PDF for the breaking force obtained using the GUF (blue curve) and scaled histogram from MCM using information from table 15.1 and clause 15.3.1

15.6.2 A normal distribution was assigned to the breaking force F in clause 15.3.1 since it is common practice in testing laboratories when preparing uncertainty budgets in the area. However, for destructive testing, as here, a Weibull distribution is often used especially in research studies in the area [87, 96]. Accordingly, a Weibull distribution was alternatively assigned to the quantity F for this purpose. Typically, such a distribution is conventionally fitted to the data for F and the mean and standard deviation extracted, which yields in this case exactly the same mean and standard uncertainty as for the normal distribution. There are readily discernible differences between the distributions, shown in figure 15.2: for instance, whilst the normal distribution is symmetric about its mean, the Weibull distribution displays some asymmetry.

15.6.3 The question may be raised as to what are the conditions in which to use the Weibull distribution instead of the normal. For the data analyzed here, the uncertainty contribution from the six observations was virtually negligible compared with those from the correction effects. That contribution, 0.7 kN is to be compared with a the standard uncertainty, about 10 kN, associated with the estimates of the measurand. Bearing in mind that the standard uncertainties associated with the input estimates are combined in quadrature, that contribution influences the combined standard uncertainty by only 0.2 %.

15.6.4 In cases where this contribution is more meaningful, it is recommended to use the Weibull distribution, which seems state of the art in the area [96]. It is to be noted that since the means of these fitted distributions are identical, as are their standard deviations, the GUF would yield identical results for the two distributions. Conversely, MCM would deliver different results since the PDF for the first input quantity would be different. More trust could be placed in the MCM results since all conditions [17, clause 5.7] required for the GUF to apply would not hold.

Final draft

16 Comparison loss in microwave power meter calibration

16.1 Preamble

16.1.1 This example is provided by the JCGM-WG1.

16.1.2 This example illustrates the evaluation of measurement uncertainty for a power ratio that can occur during a microwave power meter calibration. Four methods for evaluating the uncertainty are considered:

1. JCGM 100:2008 uncertainty framework (GUF) with first-order terms (GUF₁);
2. GUF with higher-order terms (GUF₂);
3. Monte Carlo method (MCM);
4. analytic calculation.

The example includes scenarios where the Central Limit Theorem (CLT) [10, Annex G.2] holds from acceptably to poorly, or does not hold at all. Whilst the use of GUF yields correct estimates and standard uncertainties – provided that higher-order terms are taken into account – it does not in general provide reliable coverage intervals. MCM instead provides realistic uncertainties and coverage intervals for all scenarios. Clause 16.5 contains some information that is relevant to the analytic calculations presented in this example that may be of interest to advanced practitioners of uncertainty calculations. This information can be disregarded by other (more general) practitioners, who can therefore ignore references in the example to that clause.

16.1.3 Microwave power meter calibration involves the measurement of several voltage reflection coefficients. Voltage reflection coefficients are complex-valued quantities. Manufacturers of microwave power meters usually try to produce power meters where these voltage reflection coefficients are as close to $0 + j0$, where $j^2 = -1$, as possible. In practice, the absolute values for both the real and imaginary components of these voltage reflection coefficients are expected to be less than 0.050. Hence, this example considers absolute values for these components between zero and 0.050.

16.2 Formulation

16.2.1 During the calibration of a microwave power meter, the power meter and a standard power meter are connected in turn to a stable signal generator. The power absorbed by each meter will in general be different because their complex input voltage reflection coefficients are not identical. The ratio Y of the power P_M absorbed by the meter being calibrated and that, P_S , by the standard meter is [110]

$$Y = \frac{P_M}{P_S} = \frac{1 - |\Gamma_M|^2}{1 - |\Gamma_S|^2} \times \frac{|1 - \Gamma_S \Gamma_G|^2}{|1 - \Gamma_M \Gamma_G|^2}, \quad (16.1)$$

where Γ_G is the voltage reflection coefficient of the signal generator, Γ_M that of the meter being calibrated and Γ_S that of the standard meter. This power ratio is an instance of ‘comparison loss’ [9, 78].

16.2.2 Consider the case where the standard and the signal generator are reflectionless, that is, $\Gamma_S = \Gamma_G = 0 + j0$, and measured values are obtained of the real and imaginary parts X_1 and X_2 of $\Gamma_M = X_1 + jX_2$. Since $|\Gamma_M|^2 = X_1^2 + X_2^2$, formula (16.1) becomes

$$Y = 1 - X_1^2 - X_2^2. \quad (16.2)$$

16.2.3 The available data are the estimates x_1 and x_2 of the quantities X_1 and X_2 from measurement and the associated standard uncertainties $u(x_1)$ and $u(x_2)$. X_1 and X_2 are often not independent. Denote by $u(x_1, x_2)$ the covariance associated with x_1 and x_2 . Equivalently, $u(x_1, x_2) = r(x_1, x_2)u(x_1)u(x_2)$, where $r = r(x_1, x_2)$ denotes the associated correlation coefficient [10, clause 5.2.2].

NOTE The practitioner may sometimes have difficulty in quantifying the covariance. In such cases, the uncertainty evaluation can be repeated with various trial numerical values for the correlation coefficient in order to study its effect. This example carries out calculations using a correlation coefficient of zero and of 0.9 (compare 16.2.7).

16.2.4 On the basis of [17, clause 6.4.8], $\mathbf{X} = (X_1, X_2)^T$ is assigned a bivariate normal probability density function (PDF) in X_1 and X_2 , with expectation and covariance matrix

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \begin{bmatrix} u^2(x_1) & ru(x_1)u(x_2) \\ ru(x_2)u(x_1) & u^2(x_2) \end{bmatrix}. \quad (16.3)$$

16.2.5 Because the magnitudes of X_1 and X_2 in expression (16.2) are in practice small compared with unity, the resulting Y is close to unity. Results are accordingly expressed in terms of the quantity

$$\delta Y = 1 - Y = X_1^2 + X_2^2, \quad (16.4)$$

taken as the measurement model. For physical reasons, $0 \leq Y \leq 1$, and hence $0 \leq \delta Y \leq 1$.

16.2.6 The determination of an estimate δy of δY , the associated standard uncertainty $u(\delta y)$ and a coverage interval for δY will be considered for choices of x_1 , x_2 , $u(x_1)$, $u(x_2)$ and $r(x_1, x_2)$. All quantities have dimension 1.

16.2.7 Six cases are considered, in which x_2 is taken as zero and $u(x_1) = u(x_2) = 0.005$. The first three cases correspond to $x_1 = 0, 0.010$ and 0.050 , each with $r = 0$. The other three cases correspond to taking the same x_1 , but with $r = 0.9$. The various numerical values of x_1 (comparable to those occurring in practice) are used to investigate the extent to which the results obtained using the considered approaches differ. In these six cases, x_2 is set to zero. It would be equally applicable to set x_1 to zero and use differing values for x_2 in these examples. In most practical situations, it is expected that $-0.05 \leq x_1 \leq 0.05$ and $-0.05 \leq x_2 \leq 0.05$.

16.2.8 For the cases in which $r = 0$, the covariance matrix given in formulæ (16.3) reduces to a diagonal matrix with entries $(u^2(x_1), u^2(x_2))$ and the corresponding joint distribution for X_1 and X_2 to the product of two univariate normal distributions for X_i , for $i = 1, 2$, with expectation x_i and standard deviation $u(x_i)$.

16.3 Propagation and summarizing: uncorrelated case

16.3.1 General

16.3.1.1 The evaluation of uncertainty is treated

1. analytically (for purposes of comparison),
2. using the GUF and
3. using the MCM.

NOTE These approaches do not constrain δY to be less than or equal to unity. However, for sufficiently small uncertainties $u(x_1)$ and $u(x_2)$, as here, the PDF for δY may adequately be approximated by a simpler PDF defined over all non-negative values of δY . A rigorous treatment, using Bayesian inference [126], which applies regardless of the magnitudes of $u(x_1)$ and $u(x_2)$, is possible, but beyond the scope here. Also see [17, Scope, NOTE 2].

16.3.1.2 δy and $u(\delta y)$ can generally be formed analytically as the expectation and standard deviation of δY , as characterized by the PDF for δY . See 16.5.2. The PDF for δY can be formed analytically when $x_1 = 0$ and, in particular, used to determine the endpoints of the shortest 95 % coverage interval in that case. See 16.5.3.

16.3.1.3 GUF with first-order terms (GUF_1) and with higher-order terms (GUF_2) is applied for each of the three estimates x_1 in the uncorrelated case. See 16.5.4.

16.3.1.4 GUF_1 (see clause 7.3.1) yields

$$\delta y_{\text{GUF}_1} = x_1^2 + x_2^2, \quad (16.5)$$

$$u^2(\delta y_{\text{GUF}_1}) = u^2(x_1^2) + u^2(x_2^2) = 4x_1^2 u^2(x_1) + 4x_2^2 u^2(x_2). \quad (16.6)$$

16.3.1.5 GUF_2 (see clause 7.3.2) yields

$$\delta y_{\text{GUF}_2} = x_1^2 + x_2^2 + u^2(x_1) + u^2(x_2), \quad (16.7)$$

$$u^2(\delta y_{\text{GUF}_2}) = 4x_1^2 u^2(x_1) + 4x_2^2 u^2(x_2) + 2u^4(x_1) + 2u^4(x_2). \quad (16.8)$$

Considering higher-order terms in the Taylor expansion of the measurement model impacts not only on the standard uncertainty, but on the estimate [85]. This important fact is not taken into account in JCGM 100:2008 (see [10, clause 5.1.2]).

16.3.1.6 MCM is applied in each case with $M = 10^6$ trials.

16.3.2 Input estimate $x_1 = 0$

16.3.2.1 For the input estimate $x_1 = 0$, GUF_2 and MCM only can be used, because the partial derivatives of δY with respect to X_1 and X_2 , evaluated at $X_1 = x_1$ and $X_2 = x_2$, are identically zero when $x_1 = x_2 = 0$. Accordingly, first-order terms in the Taylor expansion would be identically zero. Thus, if GUF_1 were applied, both the estimate and the associated standard uncertainty would incorrectly be computed as zero!

NOTE A similar difficulty would arise for x_1 close to zero.

16.3.2.2 Figure 16.1 shows the results for δY determined

1. analytically (the exponentially decreasing curve for $\delta Y \geq 0$ and zero elsewhere),
2. using GUF_2 (bell-shaped curve) and
3. using MCM (scaled frequency distribution shown as a histogram).

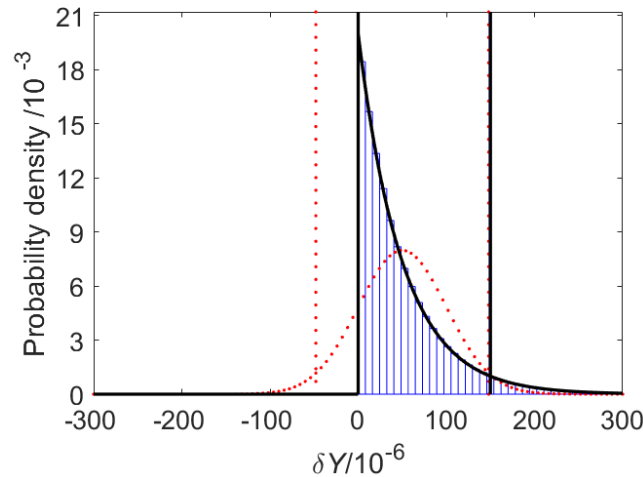


Figure 16.1: Results for the model of comparison loss in the case $x_1 = x_2 = 0$, with $u(x_1) = u(x_2) = 0.005$ and $r = 0$ (16.3.2.2, 16.3.2.5, 16.3.2.8)

NOTE In all figures, normal PDFs are depicted for the GUF results. These PDFs are those that would hold for the measurand should the CLT be applicable. As such, they are to be understood as a mere graphical tool useful to assess visually the adequacy of the CLT approximation.

16.3.2.3 Since the partial derivatives of the model function (16.4) of order higher than two are all identically zero, GUF_2 takes all Taylor-series terms, that is, the full non-linearity of the problem, into account. Thus, the GUF_2 solution is the best that is possible within the GUF.

16.3.2.4 The analytic solution is a particular chi-squared distribution—the sum of squares of two standard normal variables (see 16.5.3). It is seen in the figure that this PDF is very different from the normal PDF assumed when using GUF_2 . It can therefore be concluded that the CLT approximation is inadequate and, as a consequence, it can be expected that, although the estimate and associated standard uncertainty provided by GUF_2 are correct, coverage intervals are not.

16.3.2.5 It is also seen in figure 16.1 that the PDF provided by MCM is consistent with the analytic solution.

16.3.2.6 The estimates δy determined as the expectation of δY obtained analytically (A), using GUF_1 (G_1), GUF_2 (G_2) and MCM (M) are given in columns 2 to 5 of the row corresponding to $x_1 = 0.000$ in table 16.1. Columns 6 to 9 contain the corresponding $u(\delta y)$.

Table 16.1: Comparison loss results for $r = 0$ obtained analytically (A), using GUF₁ (G₁), GUF₂ (G₂) and MCM (M) (16.3.2.6)

x_1	Estimate $\delta y/10^{-6}$				Standard unc. $u(\delta y)/10^{-6}$				Shortest 95 % coverage interval for $\delta Y/10^{-6}$			
	A	G ₁	G ₂	M	A	G ₁	G ₂	M	A	G ₁	G ₂	M
0.000	50	0	50	50	50	0	50	50	[0, 150]	[0, 0]	[-48, 148]	[0, 150]
0.010	150	100	150	150	112	100	112	112	—	[-96, 296]	[-70, 370]	[0, 367]
0.050	2550	2500	2550	2551	502	500	502	502	—	[1520, 3480]	[1566, 3534]	[1590, 3543]

16.3.2.7 Use of GUF₁ with model (16.4) is not possible, as explained in subclause 16.3.2.1. Both GUF₂ and MCM correctly provide estimate and associated standard uncertainty in agreement with the analytic solution.

16.3.2.8 Figure 16.1 also shows the shortest 95 % coverage intervals for the corresponding approximations to the distribution function for δY . The 95 % coverage interval, indicated by dotted vertical lines, as provided by GUF₂ is infeasible: it erroneously implies there is a non-zero probability that δY is negative. However, it provides a reasonably accurate upper endpoint. The continuous vertical lines are the endpoints of the shortest 95 % coverage interval derived from the analytic solution, as described in 16.5.3. The endpoints of the shortest 95 % coverage interval determined using MCM are indistinguishable to graphical accuracy from those for the analytic solution.

16.3.2.9 The endpoints of the shortest coverage intervals for $x_1 = 0.000$ in table 16.1 are given in row 1 and columns 10 to 13 of that table.

16.3.3 Input estimate $x_1 = 0.010$

16.3.3.1 For the input estimate $x_1 = 0.010$, with correlation coefficient $r = 0$, figure 16.2 shows the results obtained using GUF₁ (higher-peaked curve), GUF₂ (lower-peaked curve) and MCM (histogram).

16.3.3.2 The PDF provided by MCM exhibits a modest left-hand flank, although it is truncated at zero, the smallest feasible value of δY . Further, compared with the results for $x_1 = 0$, it is closer in form to a normal PDF. This outcome is due to the fact that first-order terms in the Taylor expansion are no longer identically zero as for $x_1 = 0$, so that the CLT approximation starts becoming acceptable. The expectations for δY provided by GUF₁ and GUF₂ are 100×10^{-6} and 150×10^{-6} , and the standard uncertainties 100×10^{-6} and 112×10^{-6} , respectively. This closeness of the standard uncertainties indicates that contributions from second-order terms are smaller than those from first order.

16.3.3.3 Figure 16.2 also shows the endpoints of the shortest 95 % coverage intervals obtained by the three approaches. The continuous vertical lines denote the endpoints of the interval provided by MCM, the broken vertical lines those resulting from GUF₁ and the dotted vertical lines from GUF₂. The interval provided by GUF₁ is shifted to the left compared with that from GUF₂ and with the shortest 95 % coverage interval for MCM. Both intervals provided by GUF again include infeasible values of δY . Again, GUF₂ provides

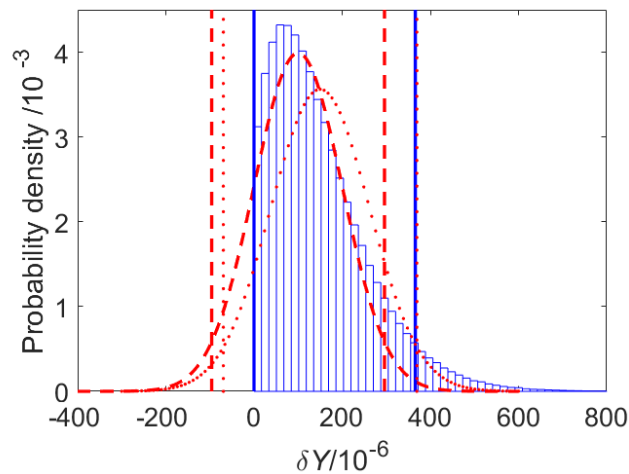


Figure 16.2: As figure 16.1 except that $x_1 = 0.010$, and the results from GUF_1 (higher-peaked curve) and GUF_2 (lower-peaked curve) (16.3.3.1, 16.3.3.3, 16.3.4.1, 16.4.3)

an accurate upper limit, graphically indistinguishable from that provided by MCM. The interval provided by MCM has its left-hand endpoint at zero, the smallest feasible value.

16.3.3.4 The corresponding results are given in the penultimate row of table 16.1.

16.3.4 Input estimate $x_1 = 0.050$

16.3.4.1 Figure 16.3 is similar to figure 16.2, but for $x_1 = 0.050$. Now, the PDF provided by MCM is much closer to a normal PDF, a clear indication that the CLT applies in this case. The reasons are that the first-order term in the Taylor expansion is dominant and the input quantities are assigned normal distributions. For the same reasons, the PDFs assumed with both variants of GUF are virtually indistinguishable from each other, – apart from a slight difference in the expectations – and from the PDF provided by MCM. The latter exhibits a slight skewness, as evidenced in the tail regions. The coverage intervals provided by the two variants of GUF are now feasible, that provided by GUF_2 being slightly closer to the MCM solution, as expected.

16.3.4.2 The corresponding results are given in the final row of table 16.1.

16.3.5 Discussion

As x_1 becomes increasingly removed from zero, the results given by GUF, with first-order and with higher-order terms, and those for MCM become closer to each other.

The numerical values $x_1 = x_2 = 0$ lie in the centre of the region of interest to the practitioner, corresponding to the so-called ‘matched’ condition for the power meter being calibrated, and thus in no sense constitute an extreme case.

Because of the symmetry of the model in X_1 and X_2 , exactly the same effect would occur were x_2 used in place of x_1 .

One reason why GUF_1 might be used in practice is that software for its implementation is readily available: results obtained from it might sometimes be accepted without question.

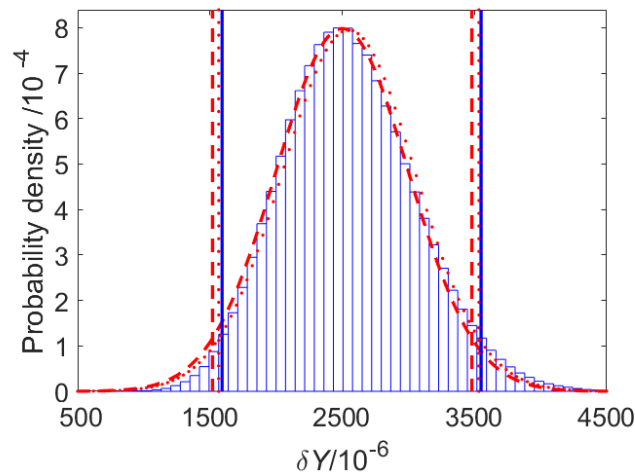


Figure 16.3: As figure 16.2 except that $x_1 = 0.050$ (16.3.4.1, 16.4.3)

For the case where $x_1 = x_2 = 0$ (figure 16.1), the danger would be apparent because the standard uncertainty $u(\delta y)$ was computed as zero. Consequently, any coverage interval for δY would be of zero length for any coverage probability. For $x_1 \neq 0$ (or $x_2 \neq 0$), $u(\delta y)$ and the length of the coverage interval for δY are both non-zero, so no such warning would be available without prior knowledge of likely values for $u(\delta y)$ and this length. Thus, a danger in implementing software based on GUF for these calculations is that checks of the software for x_1 or x_2 sufficiently far from zero would not indicate such problems, although, when used subsequently in practice for small values of x_1 or x_2 , the results would be invalid, but conceivably unwittingly accepted.

16.4 Propagation and summarizing: correlated case

16.4.1 General

16.4.1.1 The three approaches used in the cases where the X_i are uncorrelated (see 16.3) are now applied for the three cases in which they are correlated, using, as an example, $r = 0.9$. Unlike the cases where the X_i are uncorrelated, GUF₂ is not applied, no counterpart being provided in JCGM 100:2008 for the formula containing higher-order terms when the x_i have associated non-zero correlation coefficient [17, clause 5.8].

NOTE Calculations with higher-order terms are possible even in the case of correlated x_i [85].

16.4.1.2 For GUF₁, $u(\delta y)$ is evaluated as described in 16.5.5. Expression (16.17) in that subclause gives, for $x_2 = 0$,

$$u^2(\delta y) = 4x_1^2 u^2(x_1).$$

Consequently, $u(\delta y)$ does not depend on r and GUF with first-order terms gives identical results to those presented in 16.3. In particular, for the case $x_1 = 0$, $u(\delta y)$ is (incorrectly) computed as zero, as in 16.3.2.1.

16.4.1.3 MCM was implemented by sampling randomly from X characterized by a bivariate normal PDF with the given expectation and covariance matrix (expressions (16.3)). The procedure in [17, Annex C.5] was used.

NOTE Apart from the requirement to draw from a multivariate distribution, the implementation of MCM for input quantities that are correlated is no more complicated than when the input quantities are uncorrelated.

16.4.2 Input estimates $x_1 = 0, 0.010$ and 0.050

16.4.2.1 Table 16.2 contains the results obtained. Those from MCM indicate that although δy is unaffected by the correlation between the X_i , $u(\delta y)$ is so influenced, more so for small x_1 , compared with the equivalent values in Table 16.1. The 95 % coverage intervals are influenced accordingly.

Table 16.2: Comparison loss results for $r = 0.9$ obtained analytically (A), and using GUF_1 (G_1) and MCM (M) (16.4.2.1)

x_1	Estimate $\delta y/10^{-6}$			Std. uncertainty $u(\delta y)/10^{-6}$			Shortest 95 % coverage interval for $\delta Y/10^{-6}$		
	A	G_1	M	A	G_1	M	A	G_1	M
0.000	50	0	50	67	0	67	—	[0, 0]	[0, 185]
0.010	150	100	150	121	100	121	—	[-96, 296]	[13, 398]
0.050	2550	2500	2551	505	500	504	—	[1520, 3480]	[1628, 3555]

16.4.2.2 Figures 16.4 and 16.5 show the results provided by GUF_1 (bell-shaped curves) and MCM (scaled frequency distributions) in the cases $x_1 = 0.010$ and $x_1 = 0.050$, respectively. The endpoints of the shortest 95 % coverage interval provided by the two approaches are also shown, as dotted vertical lines for GUF and continuous vertical lines for MCM.

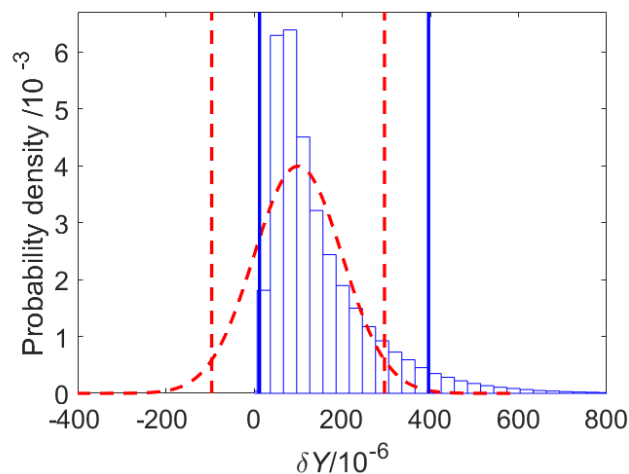


Figure 16.4: Results for the model of comparison loss in the case $x_1 = 0.010$, $x_2 = 0$, with $u(x_1) = u(x_2) = 0.005$ and $r = 0.9$ (16.4.2.2, 16.4.3)

NOTE Strictly, the conditions under which δY can be characterized by a normal PDF do not hold following an application of GUF in this circumstance (see [17, clause 5.8] and [10, Annex G.6.6]). However, this PDF and the endpoints of the corresponding 95 % coverage interval are shown because such a characterization is commonly used.

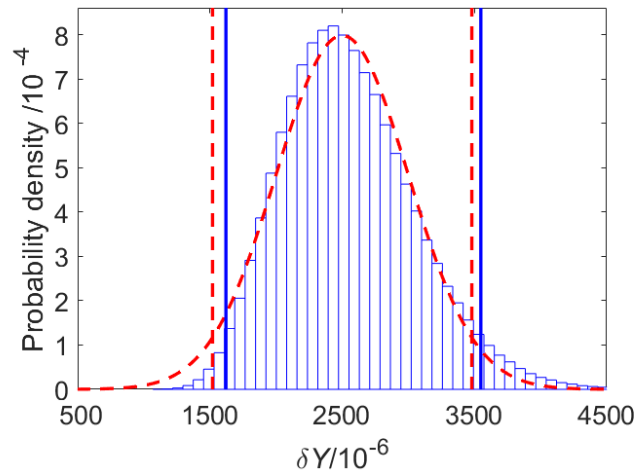


Figure 16.5: As figure 16.4 except that $x_1 = 0.050$ (16.4.2.2,16.4.3)

16.4.3 Discussion

In the case $x_1 = 0.010$ (figure 16.4), the effect of the non-zero correlation coefficient has been to change noticeably the results returned by MCM (compare with figure 16.2). Not only has the shape of (the approximation to) the PDF changed, but the corresponding coverage interval no longer has its left-hand endpoint at zero. In the case $x_1 = 0.050$ (figure 16.5), the differences between the results for the cases where the input quantities are uncorrelated and correlated (compare with figure 16.3) are less obvious.

16.5 Analytic calculations

16.5.1 General

This clause provides some analytic results, partly based on the considerations in [38]. Subclause 16.5.2 provides expectation and standard deviation of δY (16.3.1.2). Subclause 16.5.3 provides the PDF for δY when $x_1 = x_2 = r(x_1, x_2) = 0$ (16.3.1.2). Subclause 16.5.4 applies GUF for uncorrelated input quantities (16.3.1.3) and correlated input quantities (16.4.1.1).

16.5.2 Expectation and standard deviation

16.5.2.1 The variance of a random variable X can be expressed in terms of expectations as [109, page 124]

$$V(X) = E(X^2) - [E(X)]^2.$$

Thus,

$$E(X^2) = [E(X)]^2 + V(X) = x^2 + u^2(x),$$

where x is the estimate of X and $u(x)$ the associated standard uncertainty.

For model (16.4), i.e. $\delta Y = X_1^2 + X_2^2$,

$$\delta y = E(\delta Y) = x_1^2 + x_2^2 + u^2(x_1) + u^2(x_2).$$

This result applies

1. regardless of the PDFs assigned to X_1 and X_2 , and
2. whether X_1 and X_2 are independent or not.

The standard uncertainty associated with δy can be obtained from

$$u^2(\delta y) = u^2(x_1^2) + u^2(x_2^2) + 2u(x_1^2, x_2^2),$$

where, for $i = 1$ and $i = 2$, $u^2(x_i^2) = V(X_i^2)$, and $u(x_1^2, x_2^2) = \text{Cov}(X_1^2, X_2^2)$. Then, applying Price's Theorem for normal distributions [99, 104],

$$u^2(\delta y) = 4u^2(x_1)x_1^2 + 4u^2(x_2)x_2^2 + 2u^4(x_1) + 2u^4(x_2) + 4u^2(x_1, x_2) + 8u(x_1, x_2)x_1x_2. \quad (16.9)$$

When $x_2 = 0$ and $u(x_2) = u(x_1)$, and replacing $u(x_1, x_2)$ by $r(x_1, x_2)u^2(x_1)$,

$$u(\delta y) = 2 \{x_1^2 + [1 + r^2(x_1, x_2)]u^2(x_1)\}^{1/2} u(x_1).$$

16.5.2.2 When X_1 and X_2 are uncorrelated, that is, $u(x_1, x_2) = 0$, expression (16.9) becomes

$$u^2(\delta y) = 4u^2(x_1)x_1^2 + 4u^2(x_2)x_2^2 + 2u^4(x_1) + 2u^4(x_2). \quad (16.10)$$

Expression (16.10) can be verified by applying formula (10) in [10, clause 5.1.2] and the immediately following note.

16.5.3 Analytic solution for zero estimate of the voltage reflection coefficient having associated zero covariance

16.5.3.1 For the case $x_1 = x_2 = r(x_1, x_2) = 0$ and $u(x_1) = u(x_2)$, the PDF $g_Y(\eta)$ for Y can be obtained analytically. It is valuable to have such a solution for further validation purposes. In the above circumstances,

$$\delta Y = u^2(x_1) \left[\frac{X_1^2}{u^2(x_1)} + \frac{X_2^2}{u^2(x_2)} \right].$$

16.5.3.2 The term in square brackets is the sum, Z , say, of the squares of two independent quantities, each of which is distributed as a standard normal PDF. Thus, the sum is distributed as chi-squared with two degrees of freedom [109, page 177], so that

$$\delta Y = u^2(x_1)Z,$$

where Z has PDF

$$g_Z(\zeta) = \chi_2^2(\zeta) = e^{-\zeta/2}/2.$$

16.5.3.3 The application of a general formula [109, pages 57–61] for the PDF $g_Y(\eta)$ of a differentiable and strictly decreasing function of a variable (here Z) with a specified PDF yields

$$g_Y(\eta) = \frac{1}{u^2(x_1)} \chi_2^2\left(\frac{\eta}{u^2(x_1)}\right) = \frac{1}{2u^2(x_1)} \exp\left(-\frac{\eta}{2u^2(x_1)}\right), \quad \eta \geq 0.$$

16.5.3.4 The expectation of δY is

$$\delta y = E(\delta Y) = \int_0^{\infty} \eta g_Y(\eta) d\eta = 2u^2(x_1)$$

and the variance

$$u^2(\delta y) = V(\delta Y) = \int_0^{\infty} (\eta - \delta y)^2 g_Y(\eta) d\eta = 4u^4(x_1),$$

that is, the standard deviation is $2u^2(x_1)$, results that are consistent with those in 16.5.2.

16.5.3.5 By integration, the corresponding distribution function is

$$G_Y(\eta) = 1 - \exp\left(-\frac{\eta}{2u^2(x_1)}\right), \quad \eta \geq 0. \quad (16.11)$$

16.5.3.6 Let δy_α be that η in expression (16.11) corresponding to $G_Y(\eta) = \alpha$ for any α satisfying $0 \leq \alpha \leq 1 - p$. Then

$$\delta y_\alpha = -2u^2(x_1) \ln(1 - \alpha)$$

and a $100p$ % coverage interval for δY [17, clause 7.7] is

$$[\delta y_\alpha, \delta y_{p+\alpha}] \equiv -2u^2(x_1)[\ln(1 - \alpha), \ln(1 - p - \alpha)] \quad (16.12)$$

with length

$$H(\alpha) = -2u^2(x_1) \ln\left(1 - \frac{p}{1 - \alpha}\right).$$

16.5.3.7 The shortest $100p$ % coverage interval is given by determining α to minimize $H(\alpha)$ [17, clause 5.3.4]. Since $H(\alpha)$ is a strictly increasing function of α for $0 \leq \alpha \leq 1 - p$, $H(\alpha)$ is minimized when $\alpha = 0$. Thus, the shortest $100p$ % coverage interval for δY is

$$[0, -2u^2(x_1) \ln(1 - p)].$$

For $u(x_1) = 0.005$, the shortest 95 % coverage interval is

$$[0, 0.000\ 149\ 8].$$

16.5.3.8 The 95 % probabilistically symmetric coverage interval for δY is given by setting $\alpha = (1 - p)/2$ [17, clause 5.3.3]:

$$-2u^2(x_1)[\ln 0.975, \ln 0.025] = [0.000\ 001\ 3, 0.000\ 184\ 4],$$

which is 20 % longer than the shortest 95 % coverage interval.

NOTE The above analysis is indicative of an approach that can be applied to some problems of this type. In this case, the results could have been obtained more directly since $g_Y(\eta)$ is strictly increasing and unimodal, and so the shortest coverage interval is in the region of highest density.

16.5.4 GUF — Uncorrelated input quantities

16.5.4.1 In model 16.4, X_1 and X_2 are assigned normal PDFs having expectations x_1 and x_2 and variances $u^2(x_1)$ and $u^2(x_2)$, respectively. The corresponding estimate of δY is thus

$$\delta y = x_1^2 + x_2^2. \quad (16.13)$$

The only non-trivially non-zero partial derivatives of the model are, for $i = 1, 2$,

$$\frac{\partial f}{\partial X_i} = 2X_i, \quad \frac{\partial^2 f}{\partial X_i^2} = 2.$$

16.5.4.2 Hence the application of [10, clause 5.1.2] gives, for the standard uncertainty $u(\delta y)$,

$$u^2(\delta y) = \left[\left(\frac{\partial f}{\partial X_1} \right)^2 u^2(x_1) + \left(\frac{\partial f}{\partial X_2} \right)^2 u^2(x_2) \right] \Bigg|_{X=x} = 4x_1^2 u^2(x_1) + 4x_2^2 u^2(x_2), \quad (16.14)$$

based on a first-order Taylor series approximation of $f(\mathbf{X})$. If the non-linearity of f is significant [10, 5.1.2 note], the terms

$$\frac{1}{2} \left(\frac{\partial^2 f}{\partial X_1^2} \right) u^2(x_1) + \frac{1}{2} \left(\frac{\partial^2 f}{\partial X_2^2} \right) u^2(x_2)$$

and

$$\frac{1}{2} \left(\frac{\partial^2 f}{\partial X_1^2} \right)^2 u^4(x_1) + \frac{1}{2} \left(\frac{\partial^2 f}{\partial X_2^2} \right)^2 u^4(x_2)$$

need to be appended to formulæ (16.13) and (16.14), respectively. These formulæ thus become

$$\delta y = E(\delta Y) = x_1^2 + x_2^2 + u^2(x_1) + u^2(x_2) \quad (16.15)$$

and

$$u^2(\delta y) = 4x_1^2 u^2(x_1) + 4x_2^2 u^2(x_2) + 2u^4(x_1) + 2u^4(x_2). \quad (16.16)$$

16.5.4.3 A 95% coverage interval for δY is given by

$$\delta y \pm 2u(\delta y),$$

as a consequence of δY having a normal PDF.

16.5.5 GUF — Correlated input quantities

16.5.5.1 When the input quantities are correlated, the covariance matrix associated with the estimates of the input quantities is given in formulæ (16.3).

16.5.5.2 The application of JCGM 100:2008 clause 5.2.2 gives

$$\begin{aligned} u^2(\delta y) &= \left[\left(\frac{\partial f}{\partial X_1} \right)^2 u^2(x_1) + \left(\frac{\partial f}{\partial X_2} \right)^2 u^2(x_2) + 2 \frac{\partial f}{\partial X_1} \frac{\partial f}{\partial X_2} r(x_1, x_2) u(x_1) u(x_2) \right] \Bigg|_{X=x} \\ &= 4x_1^2 u^2(x_1) + 4x_2^2 u^2(x_2) + 8r(x_1, x_2) x_1 x_2 u(x_1) u(x_2). \end{aligned} \quad (16.17)$$

16.6 Conclusion

This example shows that in microwave power meter calibration the JCGM 100:2008 uncertainty framework (GUF), either with (GUF₂) or without (GUF₁) the inclusion of higher-order terms, can yield unrealistic estimates, standard uncertainties and coverage intervals, as demonstrated by the results presented in the figures and tables in this example. It is therefore recommended that, for situations such as these, a Monte Carlo method [17] is used.

Final draft

Annex A Conventions and notation

A.1 A generic quantity is denoted by an upper case letter and a measured value (estimate) by a lower case letter [17, Clause 4.8].

EXAMPLE Measured values of the measurand Y and of the generic input quantity X_i are denoted by y and x_i , respectively.

A.2 A specific quantity is denoted by its agreed symbol. The measured value (estimate) is denoted by the same symbol adorned with a hat [17, Clause 4.8].

EXAMPLE 1 The thermodynamic temperature is denoted by T and a measured value of this temperature by \hat{T} , for example $\hat{T} = 295.4$ K.

EXAMPLE 2 The mass of a weight is denoted by m and the measured value by \hat{m} .

NOTE 1 In JCGM 106:2012, a measured value is denoted by a subscript 'm'.

NOTE 2 ISO/IEC 80000 and the IUPAC Green Book [27] contain information about symbols for specific physical quantities.

A.3 The state of knowledge about a quantity is described by the probability density function (PDF) of a random variable denoted by the same symbol used for the quantity [17, Clause 4.1].

A.4 The state of knowledge about an input quantity X is described by a PDF $g_X(\xi)$, where the subscript X denotes a random variable with possible values ξ [17, Clause 4.3].

NOTE The Greek letter ξ is used here instead of its Latin counterpart x , common in statistical literature, because the latter is used in the GUM suite for the estimate of the quantity X .

A.5 Similarly, the state of knowledge about the measurand Y is described by a PDF $g_Y(\eta)$ and the corresponding distribution function by $G_Y(\eta)$. Y denotes a random variable with possible values η [17, Clause 4.7].

A.6 A PDF for more than one input quantity is often called joint even if all input quantities are mutually independent.

A.7 The PDF provided by the JCGM 100:2008 uncertainty framework (GUF) is either a normal distribution with mean y and standard deviation $u(y)$ or a shifted and scaled Student t -distribution with mean y , scale $u(y)$ and effective degrees of freedom ν .

A.8 The law of propagation of uncertainty (LPU) applies to the use of a first-order Taylor series approximation to the model. The term is qualified accordingly when a higher-order approximation is used [17, Clause 4.9].

A.9 The notation $u(y)$ is to be read as the standard uncertainty associated with y . It represents the standard uncertainty about the true value of Y .

A.10 Noting that a covariance matrix U is symmetric, only the upper triangular part is shown. The same applies for the correlation matrix R .

A.11 The terms “coverage interval” and “coverage probability” are used throughout this document. JCGM 100:2008 uses the term “level of confidence” as a synonym for coverage probability, drawing a distinction between “level of confidence” and “confidence level” [10, 6.2.2], because the latter has a specific definition in statistics. Since, in some languages, the translation from English of these two terms yields the same expression, the use of these terms is avoided here [17, Clause 4.11].

A.12 The subscript “c” [10, Clause 5.1.1] in $u_c(y)$ for the combined standard uncertainty is deemed redundant and not used. Thus, the standard uncertainty associated with y is denoted generally as $u(y)$. Likewise, the qualifier “combined” is also regarded as superfluous [17, Clause 4.10].

A.13 Unless otherwise qualified, values are expressed in a manner that indicates the number of meaningful significant decimal digits [17, Clause 4.13].

A.14 Some symbols have more than one meaning in this document. See Annex B. The context clarifies the usage.

A.15 According to Resolution 10 of the 22nd CGPM (2003) “... the symbol for the decimal marker shall be either the point on the line or the comma on the line ...”. The Joint Committee for Guides in Metrology (JCGM) has decided to adopt, in its documents in English, the point on the line as decimal marker [17, Clause 4.12].

A.16 When referencing rows in a table by number, the convention used is to treat the first row following the table heading as row 1. Similarly, when the first column serves as a heading for the rows, the numbering of columns begins with the first column after the heading, which is referred to as column 1.

Annex B Glossary of principal symbols

B.1 Generic variables and functions

C	sensitivity matrix
Cov	covariance
E	expectation
f	measurement model of the form $Y = f(X)$
\mathbf{f}	measurement model having multiple output quantities of the form $Y = \mathbf{f}(X)$
h	measurement model of the form $h(Y, X) = 0$
\mathbf{h}	measurement model having multiple output quantities of the form $\mathbf{h}(Y, X) = \mathbf{0}$
k	coverage factor
N	number of input quantities
U	expanded uncertainty
u	standard uncertainty
r	coefficient of linear correlation
R	correlation matrix
U	covariance matrix
V, σ^2	variance
X	input quantity
\mathbf{X}	column vector of input quantities X_1, \dots, X_N
x	measured value (estimate) of input quantity X
Y	measurand (output quantity of a measurement model)
\mathbf{Y}	column vector of output quantities Y_1, \dots, Y_N
y	measured value (estimate) of measurand Y

B.2 Probability distributions

Cauchy	folded scaled Cauchy distribution
CTrap	curvilinear trapezoidal distribution
Gamma	gamma distribution
LN	lognormal distribution
N	normal distribution
R	rectangular distribution (uniform distribution)
U	U-shaped distribution (arc sine distribution)
t	Student t -distribution

References

- [1] EMPIR project 17NRM05 EMUE – Examples of measurement uncertainty evaluation. Accessed 2021-06-08. URL: http://empir.npl.co.uk/emue/wp-content/uploads/sites/49/2021/07/Compendium_M36.pdf.
- [2] ISO 6142:2015/Amd 1:2020 Amendment 1: Cross reference list to ISO Guide 31:2015 and ISO/IEC 17025:2017. ISO, International Organization for Standardization, Geneva, Switzerland, 2020.
- [3] ISO 33405 Reference materials — Approaches for characterization and assessment of homogeneity and stability. ISO, International Organization for Standardization, Geneva, Switzerland, 2024. First edition.
- [4] J. Aitchison. *The Statistical Analysis of Compositional Data*. The Blackburn Press, 2003.
- [5] A Alink and A M H van der Veen. Uncertainty calculations for the preparation of primary gas mixtures. Part 1: Gravimetry. *Metrologia*, 37(6):641–650, dec 2000. doi:10.1088/0026-1394/37/6/1.
- [6] Kyle Armour, Piers Forster, Trude Storelvmo, William Collins, Jean-Louis Dufresne, Dave Frame, Dan Lunt, Thorsten Mauritsen, Matthew Palmer, Masahiro Watanabe, et al. The Earth’s energy budget, climate feedbacks, and climate sensitivity. In *AGU Fall Meeting Abstracts*, volume 2021, pages U13B–07, 2021.
- [7] 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 CF₄ and NF₃ emissions in East Asia. *Atmospheric Chemistry and Physics*, 18(18):13305–13320, 2018. URL: <https://acp.copernicus.org/articles/18/13305/2018/>, doi:10.5194/acp-18-13305-2018.
- [8] 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.
- [9] R.W. Beatty. Insertion loss concepts. *Proceedings of the IEEE*, 52(6):663–671, 1964. doi:10.1109/PROC.1964.3047.
- [10] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. Evaluation of measurement data — Guide to the expression of uncertainty in measurement. Joint Committee for Guides in Metrology, JCGM 100:2008. URL: https://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf.
- [11] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. Evaluation of measurement data — Guide to the expression of uncertainty in measurement. Joint Committee for Guides in Metrology, JCGM 100:2008. URL: https://www.bipm.org/documents/20126/2071204/JCGM_100_2008_E.pdf/cb0ef43f-baa5-11cf-3f85-4dcd86f77bd6.
- [12] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. Evaluation of measurement data — Supplement 2 to the “Guide to the expression of uncertainty in measurement” — Models with any number of output quantities. Joint Committee for Guides in Metrology, JCGM 102:2011. URL: https://www.bipm.org/utils/common/documents/jcgm/JCGM_102_2011_E.pdf.

- [13] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. Evaluation of measurement data — The role of measurement uncertainty in conformity assessment. Joint Committee for Guides in Metrology, JCGM 106:2012. URL: https://www.bipm.org/documents/20126/2071204/JCGM_106_2012_E.pdf.
- [14] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. Guide to the expression of uncertainty in measurement — Part 1: Introduction. Joint Committee for Guides in Metrology, JCGM GUM-1:2023. doi:10.59161/jcgm-gum-1-2023.
- [15] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. Guide to the expression of uncertainty in measurement — Part 6: Developing and using measurement models. Joint Committee for Guides in Metrology, JCGM GUM-6:2020. URL: https://www.bipm.org/documents/20126/2071204/JCGM_GUM_6_2020.pdf.
- [16] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. International vocabulary of metrology — Basic and general concepts and associated terms (VIM). Joint Committee for Guides in Metrology, JCGM 200:2012. (3rd edition). URL: https://www.bipm.org/documents/20126/2071204/JCGM_200_2012.pdf.
- [17] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. Evaluation of measurement data — Supplement 1 to the “Guide to the expression of uncertainty in measurement” — Propagation of distributions using a Monte Carlo method. Joint Committee for Guides in Metrology, JCGM 101:2008, 2008. URL: https://www.bipm.org/documents/20126/2071204/JCGM_101_2008_E.pdf.
- [18] 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.
- [19] Rebekah Bramwell, Dom Ingledeew, Eirini Karagianni, Joe London, Joanna MacCarthy, Peter Brown, Paddy Mullen, Charles Walker, Judith Bates, Nik Hill, Dan Willis, Jason Wong, Harper Robertson, Holly Zhang, Lucas Bennett, and Billy Harris. 2025 government greenhouse gas conversion factors for company reporting: Methodology paper. Methodology paper, Department for Energy Security and Net Zero (DESNZ), 2025. Produced by Ricardo Energy & Environment, Aether Ltd, and WRAP. URL: <https://assets.publishing.service.gov.uk/media/6846b0870392ed9b784c0187/2025-GHG-CF-methodology-paper.pdf>.
- [20] Willi A. Brand. Atomic weights: not so constant after all. *Anal. Bioanal. Chem.*, 405(9):2755–2761, Mar 2013. doi:10.1007/s00216-012-6608-0.
- [21] R. P. Buck, S. Rondinini, A. K. Covington, F. G. K. Bauke, C. M. A. Brett, M. F. Camões, 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. URL: <https://dx.doi.org/10.1351/pac200274112169>.
- [22] K. P. Burnham and D. R. Anderson. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach 2nd edn*. New York: Springer, 2002.
- [23] 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.
URL: <https://www.theccc.org.uk/wp-content/uploads/2017/06/Understanding-the-UK-Greenhouse-Gas-Inventory.pdf>.
- [24] M.-M. Bé, V. Chisté, C. Dulieu, X. Mougeot, E. Browne, V. Chechev, N. Kuzmenko, F. Kondev, A. Luca, M. Galán, A.L. Nichols, A. Arinc, and X. Huang. Table of radionuclides. Monographie BIPM-5, Vol 5, 2010. URL: [https://www.bipm.org/documents/20126/53814638/MonographieBIPM-5-Volume5\(2010\).pdf](https://www.bipm.org/documents/20126/53814638/MonographieBIPM-5-Volume5(2010).pdf).
- [25] Maria Filomena Camões. A century of pH measurements. *Chemistry International – News magazine for IUPAC*, 32(2), 2010. doi:10.1515/ci.2010.32.2.3.
- [26] 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.
- [27] 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.
- [28] E Richard Cohen. *Quantities, units and symbols in physical chemistry*. Royal Society of Chemistry, 2007.
- [29] 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(5):308, Sep 2007. doi:10.1088/0026-1394/44/5/007.
- [30] T. B. Coplen, J. K. Böhlke, P. de Bièvre, T. Ding, N. E. Holden, J. A. Hopple, H. R. Krouse, A. Lamberty, H. S. Peiser, K. M. Révész, S. E. Rieder, K. J. R. Rosman, E. Roth, P. D. P Taylor, R. D. Vocke Jr, and Y. K. Xiao. Isotope-abundance variations of selected elements. *Pure Appl. Chem.*, 74:1987–2017, 2002. URL: <https://publications.iupac.org/pac/74/10/1987/index.html>.
- [31] 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.
- [32] M. Cox and P. Harris. Polynomial calibration functions revisited: numerical and statistical issues. In F. Pavese, A.B. Forbes, A.G. Chunovkina, W. Bremser, and N. Fischer, editors, *Advanced Mathematical and Computational Tools in Metrology and Testing X*, Series on Advances in Mathematics for Applied Sciences, pages 9–16. World Scientific, 2015. doi:10.1142/9789814678629_0002.
- [33] M. G. Cox. The numerical evaluation of B-splines. *J. Inst. Math. Appl.*, 10:134–149, 1972. doi:10.1093/imamat/10.2.134.
- [34] 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. URL: <https://cir.nii.ac.jp/crid/1570009750618094848>.

- [35] M G Cox. The area under a curve specified by measured values. *Metrologia*, 44(5):365, Sep 2007. doi : 10.1088/0026-1394/44/5/013.
- [36] 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.
- [37] M. G. Cox, C. Eiø, G. Mana, and F. Pennechi. The generalized weighted mean of correlated quantities. *Metrologia*, 43:S268–S275, 2006. doi : 10.1088/0026-1394/43/4/S14.
- [38] M. G. Cox, P. M. Harris, G. Iuculano, and G. Pellegrini. The evaluation of the uncertainty associated with comparison loss in microwave power meter calibration. In *Proceedings of AMUEM 2006, the IEEE International Workshop on Advanced Methods for Uncertainty Estimation in Measurement*, pages 34–39. IEEE, 2006.
- [39] 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. URL: <https://eprintspublications.npl.co.uk/2717/>.
- [40] 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. URL: <https://www.bipm.org/documents/20126/53814638/Monographie%20BIPM-7%20%282007%29.pdf>.
- [41] Maurice Cox and Anthony O’Hagan. Meaningful expression of uncertainty in measurement. *Accreditation and Quality Assurance*, 27(1):19–37, February 2022. URL: <https://doi.org/10.1007/s00769-021-01485-5>.
- [42] Maurice G. Cox. Modelling clinical decay data using exponential functions. In Emmanuil H Georgoulis, Armin Iske, and Jeremy Levesley, editors, *Approximation Algorithms for Complex Systems*, pages 183–203, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg. doi : 10.1007/978-3-642-16876-5_8.
- [43] 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. URL: <https://www.imeko.org/publications/tc7-2004/IMEKO-TC7-2004-028.pdf>.
- [44] Qin Dahe, Thomas Stocker, et al. Highlights of the IPCC working group I fifth assessment report. *Advances in Climate Change Research*, 10(1):1, 2014.
- [45] 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.
- [46] P Dierckx. Algorithms for smoothing data with periodic and parametric splines. *Computer Graphics and Image Processing*, 20(2):171–184, 1982. URL: <https://www.sciencedirect.com/science/article/pii/0146664X82900430>, doi : 10.1016/0146-664X(82)90043-0.

- [47] C. F. Dietrich. *Uncertainty, Calibration and Probability*. Adam Hilger, Bristol, UK, 1991.
- [48] S. L. R. Ellison and A. Williams, editors. *Quantifying Uncertainty in Analytical Measurement*. EuraChem/CITAC, 2012. Third edition, QUAM:2012.P1.
- [49] 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.
- [50] EPA method IO-2.1 (1999) Sampling of ambient air for total suspended particulate matter (SPM) and PM10 using high volume (HV) sampler. <https://www.epa.gov/sites/default/files/2019-11/documents/mthd-2-1.pdf>, 1999.
- [51] 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>.
- [52] 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.
- [53] 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, 2012. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3387884/>.
- [54] Michael S Floater and Atgeirr F Rasmussen. Point-based methods for estimating the length of a parametric curve. *Journal of Computational and Applied mathematics*, 196(2):512–522, 2006.
- [55] Jonathan I. Gear, Maurice G. Cox, Johan Gustafsson, Katarina Sjögren 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.
- [56] A. Gelman. Prior distributions for variance parameters in hierarchical models (Comment on article by Browne and Draper). *Bayesian Analysis*, 1:515–534, 2006. doi:10.1214/06-ba117a.
- [57] 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.
- [58] A. Genz and F. Bretz. *Computation of Multivariate Normal and t Probabilities*. Lecture Notes in Statistics. Springer-Verlag, Heidelberg, 2009. doi:10.1007/978-3-642-01689-9.
- [59] 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.

- [60] J. V. Hajnal and D. L. G. Hill. *Medical Image Registration*. Taylor & Francis Ltd, 2019. URL: <https://www.taylorfrancis.com/books/mono/10.1201/9781420042474/medical-image-registration-joseph-hajnal-derek-hill>.
- [61] 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.
- [62] David Brynn Hibbert, Ernst-Heiner Korte, and Ulf Örnemark. Metrological and quality concepts in analytical chemistry (iupac recommendations 2021). *Pure and Applied Chemistry*, 93(9):997–1048, 2021. URL: <https://doi.org/10.1515/pac-2019-0819> [cited 2025-09-29], doi:doi:10.1515/pac-2019-0819.
- [63] 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*, Sep 2018. arXiv:1809.09505.
- [64] IAEA. *Dosimetry for Radiopharmaceutical Therapy*. International Atomic Energy Agency, Vienna, 2024. doi:doi:10.61092/iaea.xlzb-6h67.
- [65] IPCC. Good Practice Guidance and Uncertainty Management in National Greenhouse Gas Inventories, 2000.
- [66] IPCC. IPCC Guidelines for National Greenhouse Gas Inventories Volume 1 General Guidance and Reporting, 2006.
- [67] IPCC. UK Greenhouse Gas Inventory 1990 to 2016: Annual Report for submission under the Framework Convention on Climate Change, 2018.
- [68] ISO 1 Geometrical product specifications (GPS) — Standard reference temperature for the specification of geometrical and dimensional properties. ISO, International Organization for Standardization, Geneva, Switzerland, 2022. Fourth edition.
- [69] ISO 13528 Statistical methods for use in proficiency testing by interlaboratory comparison. ISO, International Organization for Standardization, Geneva, Switzerland, 2022. Third edition.
- [70] ISO 14167 Gas analysis – General quality aspects and metrological traceability of calibration gas mixtures. ISO, International Organization for Standardization, Geneva, Switzerland, 2018. First edition.
- [71] ISO/IEC 17043 Conformity assessment – General requirements for proficiency testing providers. ISO, International Organization for Standardization, Geneva, Switzerland, 2023. Second edition.
- [72] ISO 19229 Gas analysis – Purity analysis and the treatment of purity data. ISO, International Organization for Standardization, Geneva, Switzerland, 2019. Second edition.
- [73] ISO/TS 28037 Determination and use of straight-line calibration functions. ISO, International Organization for Standardization, Geneva, Switzerland, 2010.
- [74] ISO/TS 28037:2010 Determination and use of straight-line calibration functions. ISO, International Organization for Standardization, Geneva, Switzerland, 2010. First edition.

- [75] ISO/TS 28038 Determination and use of polynomial calibration functions. ISO, International Organization for Standardization, Geneva, Switzerland, 2018. First edition.
- [76] 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.
- [77] ISO 80000-3 Quantities and units — Part 3: Space and time. ISO, International Organization for Standardization, Geneva, Switzerland, 2020. Second edition.
- [78] D. M. Kerns and R. W. Beatty. *Basic Theory of Waveguide Junctions and Introductory Microwave Network Analysis*. Pergamon Press, London, 1967.
- [79] M. J. Korczynski, M. G. Cox, and P. M. Harris. Convolution and uncertainty evaluation. In P. Ciarlina, E. Felipe, A. B. Forbes, and F. Pavese, editors, *Advanced Mathematical and Computational Tools in Metrology VII*, Series on Advances in Mathematics for Applied Sciences, pages 188–195, Singapore, 2006. World Scientific. doi:10.1142/9789812774187_0018.
- [80] I. Kuselman, F. R. Pennechi, 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.
- [81] I. Kuselman, S. Shpitzer, F. Pennechi, 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.
- [82] T. P. J. Linsinger, J. Pauwels, A. M. H. van der Veen, H. Schimmel, and A. Lamberty. Homogeneity and stability of reference materials. *Accreditation and Quality Assurance*, 6(1):20–25, jan 2001. URL: <http://dx.doi.org/10.1007/s007690000261>, doi:10.1007/s007690000261.
- [83] Thomas P.J. Linsinger. Evaluation of CRM homogeneity in cases of insufficient method repeatability: comparison of bayesian analysis with substitutes for ANOVA based estimates. *Analytica Chimica Acta: X*, page 100049, apr 2020. doi:10.1016/j.acax.2020.100049.
- [84] R. Lundblad. *Handbook of biochemistry and molecular biology*. CRC Press, Boca Raton, FL, 2010.
- [85] Giovanni Mana and Francesca Pennechi. Uncertainty propagation in non-linear measurement equations. *Metrologia*, 44(3):246–251, 2007-05. doi:10.1088/0026-1394/44/3/012.
- [86] 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.

- [87] Pawel Mazurek. A comprehensive review of steel wire rope degradation mechanisms and recent damage detection methods. *Sustainability*, 15(6):5441, 2023-03. doi:10.3390/su15065441.
- [88] 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>.
- [89] C. Michotte, A. K. Pearce, M. G. Cox, and J.-J. Gostely. An approach based on the SIR measurement model for determining the ionization chamber efficiency curves, and a study of ^{65}Zn and ^{201}Tl photon emission intensities. *Applied Radiation and Isotopes*, 64:1147–1155, 2006. Proceedings of the 15th International Conference on Radionuclide Metrology and its Applications. doi:10.1016/j.apradiso.2006.02.014.
- [90] 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.
- [91] M J T Milton, G M Vargha, and A S Brown. Gravimetric methods for the preparation of standard gas mixtures. *Metrologia*, 48(5):R1, Sep 2011. doi:10.1088/0026-1394/48/5/R01.
- [92] Samarendra Kumar Mitra. On the Probability Distribution of the Sum of Uniformly Distributed Random Variables. *SIAM Journal on Applied Mathematics*, 20(2):195–198, 1971. doi:10.1137/0120026.
- [93] Andrea Murari, Emmanuele Peluso, Francesco Cianfrani, Pasquale Gaudio, and Michele Lungaroni. On the use of entropy to improve model selection criteria. *Entropy*, 21(4):394, 2019. doi:10.3390/e21040394.
- [94] Balasubramanian Narasimhan, Steven G. Johnson, Thomas Hahn, Annie Bouvier, and Kiên Kiêu. *cubature: Adaptive Multivariate Integration over Hypercubes*, 2023. R package version 2.1.0. URL: <https://CRAN.R-project.org/package=cubature>.
- [95] NIST. Certificate of Analysis, Standard Reference Material186g, pH Standards, Potassium Dihydrogen Phosphate (186-I-g), Disodium Hydrogen Phosphate (186-II-g). URL: https://shop.nist.gov/ccrz__ProductDetails?sku=186g&cclcl=en_US.
- [96] Prerita Odeyar, Derek B. Apel, Robert Hall, Brett Zon, and Krzysztof Skrzypkowski. A review of reliability and fault analysis methods for heavy equipment and their components used in mining. *Energies*, 15(17):6263, 2022-08. doi:10.3390/en15176263.
- [97] OIML. OIML D 28 Conventional value of the result of weighing in air. Technical report, Organisation International de Métrologie Légale, Paris, 2004. URL: https://www.oiml.org/en/files/pdf_d/d028-e04.pdf.
- [98] B. Osgood. *Lectures on the Fourier transform and its applications*. American Mathematical Society, Providence, Rhode Island, 2019.
- [99] A. Papoulis. Comments on 'an extension of Price's theorem' by McMahan, E. L. *IEEE Transactions on Information Theory*, 11(1):154–154, 1965. doi:10.1109/TIT.1965.1053722.

- [100] F. Pennechi, F. Rolle, A. Allard, and S. L. R. Ellison. Emue-d2-3-TSPConcentration, November 2020. doi:10.5281/zenodo.4242988.
- [101] 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.
- [102] 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.
- [103] Antonio Possolo. Copulas for uncertainty analysis. *Metrologia*, 47(3):262–271, Apr 2010. doi:10.1088/0026-1394/47/3/017.
- [104] R. Price. A useful theorem for nonlinear devices having gaussian inputs. *IRE Transactions on Information Theory*, 4(2):69–72, 1958. doi:10.1109/TIT.1958.1057444.
- [105] Thomas Prohaska, Johanna Irrgeher, Jacqueline Benefield, John K. Böhlke, Lesley A. Chesson, Tyler B. Coplen, Tipping Ding, Philip J. H. Dunn, Manfred Gröning, Norman E. Holden, Harro A. J. Meijer, Heiko Moossen, Antonio Possolo, Yoshio Takahashi, Jochen Vogl, Thomas Walczyk, Jun Wang, Michael E. Wieser, Shigekazu Yoneda, Xiang-Kun Zhu, and Juris Meija. Standard atomic weights of the elements 2021 (IUPAC technical report). *Pure and Applied Chemistry*, 94(5):573–600, May 2022. doi:10.1515/pac-2019-0603.
- [106] Sumitra Purkayastha. Simple proofs of two results on convolutions of unimodal distributions. *Statistics & Probability Letters*, 39(2):97–100, 1998. doi:10.1016/s0167-7152(98)00013-3.
- [107] 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/>.
- [108] Ricardo-AEA. An Introduction to the UK's Greenhouse Gas Inventory, 2014. URL: https://assets.publishing.service.gov.uk/media/5a7e013a40f0b62305b802ec/IntroToTheGHGI_2014_Final.pdf.
- [109] J. R. Rice. *Mathematical Statistics and Data Analysis*. Duxbury Press, Belmont, Ca., USA, second edition, 1995.
- [110] N. M. Ridler and M. J. Salter. Propagating S-parameter uncertainties to other measurement quantities. In *58th ARFTG (Automatic RF Techniques Group) Conference Digest*, 2001. doi:10.1109/ARFTG.2001.327487.
- [111] 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.
- [112] S. Eggleston and L. Buendia and Kyoko Miwa and Todd Ngara and Kiyoto Tanabe. Guidelines for National Greenhouse Gas Inventories, 2000.
- [113] 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.

- [114] I.J Schoenberg. Cardinal interpolation and spline functions. *Journal of Approximation Theory*, 2(2):167–206, 1969. doi:10.1016/0021-9045(69)90040-9.
- [115] 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. doi:10.1016/j.aca.2016.03.032.
- [116] Robert R. Sokal and F. James Rohlf. *Biometry*. W. H. Freeman, 2011.
- [117] Stan Development Team. RStan: the R interface to Stan, 2018. R package version 2.17.3. URL: <http://mc-stan.org/>.
- [118] 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.
- [119] 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.
- [120] 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. doi:10.1007/s007690000238.
- [121] 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. doi:10.1007/s007690000237.
- [122] Adriaan M H van der Veen and Katarina Hafner. Atomic weights in gas analysis. *Metrologia*, 51(1):80, Jan 2014. doi:10.1088/0026-1394/51/1/80.
- [123] Adriaan M. H. van der Veen, Juris Meija, Antonio Possolo, and David Brynn Hibbert. Interpretation and use of standard atomic weights (IUPAC technical report). *Pure and Applied Chemistry*, 93(5):629–646, April 2021. doi:10.1515/pac-2017-1002.
- [124] Adriaan M.H. van der Veen and Maurice G. Cox. Good practice in evaluating measurement uncertainty. https://empir.npl.co.uk/emue/wp-content/uploads/sites/49/2021/07/Compendium_M36.pdf.
- [125] VDI/VDE 2600 Part 2 Inspection process management – Determination of the measurement uncertainty of complex inspection processes. VDI-Richtlinie, Verband Deutscher Ingenieure, Verband der Elektrotechnik, Elektronik und Informationstechnik, Beuth-Verlag, Berlin, Germany, 2019.
- [126] K Weise and W Woger. A Bayesian theory of measurement uncertainty. *Measurement Science and Technology*, 4(1):1, Jan 1993. doi:10.1088/0957-0233/4/1/001.
- [127] 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.
- [128] Michael E. Wieser and Michael Berglund. Atomic weights of the elements 2007 (IUPAC Technical Report). *Pure and Applied Chemistry*, 81(11):2131–2156, 2009. doi:10.1351/PAC-REP-09-08-03.

- [129] Michael E. Wieser and Tyler B. Coplen. Atomic weights of the elements 2009 (IUPAC Technical Report). *Pure and Applied Chemistry*, 83(2):359–396, 2010. doi:doi:10.1351/PAC-REP-10-09-14.
- [130] Michael E. Wieser, Norman Holden, Tyler B. Coplen, John K. Böhlke, Michael Berglund, Willi A. Brand, Paul De Bièvre, Manfred Groening, Robert D. Loss, Juris Meija, Takafumi Hirata, Thomas Prohaska, Ronny Schoenberg, Glenda O'Connor, Thomas Walczyk, Shige Yoneda, and Xiang-Kun Zhu. Atomic weights of the elements 2011 (IUPAC Technical Report). *Pure and Applied Chemistry*, 85(5):1047–1078, 2013. doi:10.1351/PAC-REP-13-03-02.
- [131] R Willink. A generalization of the Welch–Satterthwaite formula for use with correlated uncertainty components. *Metrologia*, 44(5):340–349, Sep 2007. doi:10.1088/0026-1394/44/5/010.
- [132] 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, 2016. doi:10.1007/s00769-015-1183-7.
- [133] 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.
- [134] Y. C. Wu and W. F. Koch and R. A. Durst. Standard Reference Materials: Standardization of pH Measurements. Spec. Publ. 260-53, NBS, 1988.

Alphabetical index

- acceptance limit, 69
- ambient air monitoring, 13
- amount fraction, 23
- arc sine distribution
 - assignment to a quantity, 56
- atomic composition coefficient, 21
- atomic weight, 21
- atomic-weight interval, 20
- average
 - of a set of indications, 54, 56
- Bates-Guggenheim convention, 11
- Bayesian hierarchical model, 82, 83
- best estimate
 - of an input quantity, 42, 44, 49
- between-bottle homogeneity, 82
- between-bottle homogeneity standard deviation, 82
- BIPM, ii, vii
- calibration, 59
 - certificate, 55, 56
 - correlation, 2
 - gauge block, 53
 - interpolation, 2
 - mass, 48
 - microwave power meter, 104
 - multipoint, 2
 - one-point, 3
 - two-point, 2, 3
- calibration function, 2, 5
 - interpolation, 2
 - parameters, 3
 - straight-line, 4, 5
- calibration gas mixture, 23
- Cauchy distribution, 84
- Central Limit Theorem, 47, 104, 107–109
- certified reference material, 13, 82
- chemical composition, 23
- chi-squared distribution, 107, 113
- CO₂ equivalent, 34
- conformity assessment, 67
- correlated input quantities, 110, 115
- correlation, 3, 10
- correlation coefficient, 105, 108
- covariance, 105, 110
- covariance matrix, 105, 110, 115
 - block structure, 30
- input, 30
- coverage factor, 42, 55
- coverage interval, 1, 114, 115
 - probabilistically symmetric, 42–46, 114
 - shortest, 50, 57, 88, 106, 108, 111, 114
- coverage region, 1
- credible interval, 63
- degrees of freedom, 56, 57, 113
 - effective, 15, 55
- destructive testing, 99
- distribution function
 - for output quantity, 50, 108
- estimate
 - of an input quantity, 105
 - estimate of output quantity, 51, 107
 - from JCGM 100:2008 uncertainty framework, 115
 - expanded uncertainty, 55
 - expectation, 45, 49, 105–108, 110, 112, 114, 115
- false decisions, 68
- function
 - piecewise linear, 5
- gas
 - purity, 23
- gas chromatography, 13
- Gaussian distribution, 105, 108
 - assignment to a quantity, 49, 115
- GHG inventory, 33
 - bottom-up calculation, 33
- glass electrode, 2
- global risk, 68
 - consumer, 69
 - producer, 69
- glucose, 20
- gravimetry, 23
- Harned cell, 2
- hoisting rope, 99
 - breaking force, 99
 - homogeneity, 99
- IEC, ii
- IFCC, ii

- ILAC, ii
- indications, analysis of a series of, 54, 56
- interpolation
 forward, 4
 inverse, 4
- ISO, ii
- isotopic composition, 20
- IUPAC, ii
- IUPAP, ii
- JCGM, ii, vii
- JCGM 100:2008 uncertainty framework (GUF), viii, 104, 115
 example, 42, 53, 104
 validation of, 44, 46, 47, 50
 with higher-order terms, 50, 106
- laboratory reagent, 20
- law of propagation of uncertainty (LPU),
 1, 10, 15, 20, 21, 35, 37, 38, 41,
 42, 90, 96–98
 multivariate, 2, 23, 26, 27, 32, 90
- least squares regression
 polynomial, 5
- least-squares estimation, 60
- least-squares regression, 3, 5
- likelihood, 69
- likelihood function, 61
- mass spectrometry, 13
- measurement model, 60
 calibration of mass standards, 48
 extended, 6
 implicit, multivariate, 90
 implicit, univariate, 90
 multivariate, 24
 one-stage, 3, 4
 two-stage, 3, 4
- measurement uncertainty, 59
- model of measurement
 calibration of gauge block, 54
 calibration of microwave power meter, 105
 linear, 42
- molar mass, 24
 mixture, 24
- molecular weight, 20
- Monte Carlo method (MCM), 13, 15–18, 41
 example, 42, 53, 104
- multivariate Gaussian distribution, 110
- multivariate normal distribution
 assignment to a vector quantity, 105
- Normal distribution, 42
- normal distribution
 assignment to a quantity, 56
- normal material, 20
- numerical tolerance, 42, 44, 46, 47, 50, 57
- OIML, ii
- one-point calibration, 14
- partial derivative, 51, 106, 107, 115
- peak area, 14
- pH, 2
 meter, 2
 primary measurement standard, 2
 primary reference measurement procedure, 2
 scale, 2
- Polycyclic Aromatic Hydrocarbon, 13
- prior
 informative, 63
 weakly informative, 64
- prior knowledge, 60
- probability, 57, 108
- probability density function
 joint prior, 84
 likelihood, 84
 posterior, 62, 71, 87
 prior, 69
- probability density function (PDF)
 joint, 105
 sampling from, 42
- proficiency test material, 82
- propagation of distributions
 analytical, 106, 107
- rectangular distribution, 44, 46
 assignment to a quantity, 49, 56
- rectangular distribution with inexactly prescribed limits
 assignment to a quantity, 57
- reference electrode, 2
- reference material
 certified, 23
- relative atomic mass, 21
- relative molecular mass, 20

- reporting, viii
- risk
 - Bayesian framework, 69
 - false decision, 67
 - global, 67
 - specific, 67
- safety, 99
- sample
 - normal distribution, 100
 - Weibull distribution, 100
- sensitivity coefficients, 51
- significant digits, viii
- skewness, 109
- specific risk, 68
- spline
 - first-degree, 5
- standard atomic weight, 20
- standard deviation
 - between-group, 83
 - pooled, 84
 - within, 83
- statistical model, 60, 83
- straight-line regression, 59
- t-distribution
 - assignment to a quantity, 55, 56
- Taylor series approximation, 115
- torque measuring sensor, 59
- Total Suspended Particulate Matter, 67
- Type B evaluation, 21
- uncertainty budget, 51
- VIM, vii
- volumetry, 23
- Weibull distribution, 103
- Welch-Satterthwaite formula, 15, 55
 - correlated quantities, 16

Abbreviations

- AIC** Akaike's Information Criterion. 76–78
- AICc** corrected Akaike's Information Criterion. 76, 77
- ANOVA** analysis of variance. 83, 84, 87
- BIC** Bayesian Information Criterion. 76–78
- BIPM** Bureau International des Poids et Mesures. vii
- CIAAW** Commission on Isotopic Abundances and Atomic Weights. 20
- CLT** Central Limit Theorem. 43, 47, 104, 107–109
- CRM** certified reference material. 2, 13, 14, 82
- CSIR** Council for Scientific and Industrial Research, South Africa. 99
- EPA** Environmental Protection Agency. 67
- GC** gas chromatograph. 84
- GHG** greenhouse gas. 33, 34, 36, 38, 40, 41
- GUM** JCGM 100:2008 uncertainty framework. iii, v, viii, 13, 15, 17, 18, 41–50, 52, 53, 55, 57, 58, 64, 101–104, 106–112, 116, 117
- GUM** Guide to the expression of uncertainty in measurement. 1, 33, 41
- IPCC** Intergovernmental Panel on Climate Change. 33, 35, 36
- IUPAC** International Union of Pure and Applied Chemistry. 2, 10, 20, 22, 28
- JCGM** Joint Committee for Guides in Metrology. 33, 41, 118
- LPU** law of propagation of uncertainty. 1, 2, 10, 15, 20, 21, 23, 25–27, 30, 32, 35–38, 41, 42, 59, 90, 96–98, 101, 117
- MCM** Monte Carlo method. iii, v, viii, 1, 10, 13, 15–18, 41–44, 46–48, 50–55, 57, 58, 80, 101–104, 106–112
- MCMC** Markov Chain Monte Carlo method. 64, 85–89
- NIST** National Institute of Standards and Technology. 6, 10, 12
- OLS** ordinary least squares. 61, 62, 64
- PDF** probability density function. 18, 28, 32, 42–47, 49–51, 53, 57, 58, 62–64, 69–71, 84, 87, 102, 105–113, 115, 117

RM reference material. 82

SI International System of Units. 2, 11

SRM Standard reference material. 6

TSPM total suspended particulate matter. 67–69, 71

UNFCCC United Nations Framework Convention on Climate Change. 33

VIM International vocabulary of basic and general terms in metrology. vii

WLS weighted least squares. 60, 62, 64, 65

Final draft