GUIDE 98-6

First edition
2021-02

Uncertainty of measurement —

Part 6: Developing and using measurement models

Incertitude de mesure —

Partie 6: Élaboration et utilisation de modèles de mesure

Reference number
ISO/IEC GUIDE 98-6:2021(E)

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

ISO/IEC Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives or www.iec.ch/members_experts/refdocs).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents) or the IEC list of patent declarations received (see patents.iec.ch).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT), see www.iso.org/iso/foreword.html. In the IEC, see www.iec.ch/understanding-standards.

This document was prepared by Working Group 1 of the Joint Committee for Guides in Metrology (as JCGM GUM-6:2020), and was adopted by the national bodies of ISO and IEC.

A list of all parts in the ISO/IEC Guide 98 series can be found on the ISO and IEC websites.

Given that ISO/IEC Guide 98-6 is identical in content to JCGM GUM-6, the decimal symbol is a point on the line in the English version.

Annex ZZ has been appended to provide a list of corresponding ISO/IEC Guides and JCGM guidance documents for which equivalents are not given in the text.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html and www.iec.ch/national-committees.

Joint Committee for Guides in Metrology

Guide to the expression of uncertainty in measurement — Part 6: Developing and using measurement models

Guide pour l'expression de l'incertitude de mesure — Partie 6: Élaboration et utilisation des modèles de mesure

JCGM GUM-6:2020

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

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 ISO Technical Advisory Group 4 (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 International Organization of Legal Metrology (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' (the '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 became known as the 'Guide to the expression of uncertainty in measurement' or 'GUM'. This document, previously known as JCGM 103, Supplement 3 to the GUM, is the first to be published as a part of that portfolio, and is entitled and numbered accordingly.

The present guide is concerned with the development and use of measurement models, and supports the documents in the entire suite of JCGM documents concerned with uncertainty in measurement. The guide 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

A measurement model constitutes a relationship between the output quantities or measurands (the quantities intended to be measured) and the input quantities known to be involved in the measurement. There are several reasons for modelling a measurement. Models assist in developing a quantitative understanding of the measurement and in improving the measurement. A model enables values of the output quantities to be obtained given the values of the input quantities. Additionally, a model not only allows propagation of uncertainty from the input quantities to the output quantities; it also provides an understanding of the principal contributions to uncertainty. This document is accordingly concerned with the development of a measurement model and the practical use of the model.

One of the purposes of measurement is to assist in making decisions. The reliability of these decisions and the related risks depend on the values obtained for the output quantities and the associated uncertainties. In turn, these decisions depend on a suitable measurement model and the quality of information about the input quantities.

Although the development of a measurement model crucially depends on the nature of the measurement, some generic guidance on aspects of modelling is possible. A measurement model might be a straightforward mathematical relationship, such as the ideal gas law, or, at the other extreme, involve a sophisticated numerical algorithm for its evaluation, such as the detection of peaks in a signal and the determination of peak parameters.

A measurement model may take various forms: theoretical, empirical or hybrid (part-theoretical, part-empirical). It might have a single output quantity or more than one output quantity. The output quantity may or may not be expressed directly in terms of the input quantities. The quantities in the measurement model may be real-valued or complex-valued. Measurement models may be nested or multi-stage, in the sense that input quantities in one stage are output quantities from a previous stage, as occurs, for instance, in the dissemination of measurement standards or in calibration. Measurement models might describe time series of observations, including drift, and dynamic measurement. A measurement model may also take the form of a statistical model. In this document the concept 'measurement model' is intended in this broader meaning.

In developing or using a measurement model there are important choices to be made. The selection of a model that is adequate or fit for purpose is a key issue. Particularly for empirical models, there is choice of representation (or parametrization) of the families of functions concerned (polynomials, polynomial splines or rational functions, etc.). Certain choices can be far superior to others in their numerical behaviour when the model is implemented on a computer. The uncertainty arising from the choice of model is a necessary consideration.

In many disciplines, a basic measurement model requires extension to incorporate effects such as temperature corrections arising from the measurement to enable values for output quantities and the associated uncertainties to be obtained reliably.

Following the introduction in 1993 of the Guide to the expression of uncertainty in measurement, or GUM (also known as JCGM 100:2008), the practice of uncertainty evaluation has broadened to use a wider variety of models and methods. To reflect this, this Guide includes an introduction to statistical models for measurement modelling (clause 11) and additional guidance on modelling random variation in Annex C.

Guide to the expression of uncertainty in measurement — Part 6: Developing and using measurement models

1 Scope

This document provides guidance on developing and using a measurement model and also covers the assessment of the adequacy of a measurement model. The document is of particular interest to developers of measurement procedures, working instructions and documentary standards. The model describes the relationship between the output quantity (the measurand) and the input quantities known to be involved in the measurement. The model is used to obtain a value for the measurand and an associated uncertainty. Measurement models are also used in, for example, design studies, simulation of processes, and in engineering, research and development.

This document explains how to accommodate in a measurement model the quantities involved. These quantities relate i) to the phenomenon or phenomena on which the measurement is based, that is, the measurement principle, ii) to effects arising in the specific measurement, and iii) to the interaction with the artefact or sample subject to measurement.

The guidance provided is organised in accordance with a work flow that could be contemplated when developing a measurement model from the beginning. This work flow starts with the specification of the measurand (clause 6). Then the measurement principle is modelled (clause 7) and an appropriate form of the model is chosen (clause 8). The basic model thus obtained is extended by identifying (clause 9) and adding (clause 10) effects arising from the measurement and the artefact or sample subject to measurement. Guidance on assessing the adequacy of the resulting measurement model is given in clause 12. The distinction between the basic model and the (complete) measurement model in the work flow should be helpful to those readers who already have a substantial part of the measurement model in place, but would like to verify that it contains all effects arising from the measurement so that it is fit for purpose.

Guidance on the assignment of probability distributions to the quantities appearing in the measurement model is given in JCGM 100:2008 and JCGM 101:2008. In clause 11, this guidance is supplemented by describing how statistical models can be developed and used for this purpose.

When using a measurement model, numerical problems can arise including computational effects such as rounding and numerical overflow. It is demonstrated how such problems can often be alleviated by expressing a model differently so that it performs well in calculations. It is also shown how a reformulation of the model can sometimes be used to eliminate some correlation effects among the input quantities when such dependencies exist.

Examples from a number of metrology disciplines illustrate the guidance provided in this document.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

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.

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.

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' — Extension to any number of output quantities. Joint Committee for Guides in Metrology, JCGM 102:2011.

BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. International vocabulary of metrology — Basic and general concepts and associated terms. Joint Committee for Guides in Metrology, JCGM 200:2012.

3 Terms and definitions

The terms and definitions of JCGM 100:2008, JCGM 101:2008, JCGM 102:2011 and JCGM 200:2012 apply.

ISO, IEC and IUPAC maintain terminological databases for use in standardization at the following addresses:

— IEC Electropedia: available at `http://www.electropedia.org`
— ISO Online Browsing Platform: available at `http://www.iso.org/obp`
— IUPAC Gold Book: available at `http://www.goldbook.iupac.org`

4 Conventions and notation

4.1 The conventions and notation in JCGM 100:2008, JCGM 101:2008 and JCGM 102:2011 are adopted. Principal symbols used throughout the document are explained in annex A. Other symbols and those appearing in examples are explained at first occurrence.

4.2 Most examples in this document contain numerical values rounded to a number of decimal digits appropriate to the application. Because of rounding there are often numerical inconsistencies among the values presented. An instance is the correlation coefficient of −0.817 in the example in 8.1.6. It is obtained from the *computer-held values* of two

standard uncertainties and a covariance. If it were computed from the three presented values (three significant decimal digits), its value correct to three decimal digits would be −0.822.

4.3 Links to numbered subclauses are indicated by underlining.

5 Basic principles

5.1 A *measurand* (see JCGM 200:2012, 2.3) is in many cases not measured directly, but is indirectly determined from other *quantities* (see JCGM 200:2012, 1.1) to which it is related by a *measurement model* (see JCGM 200:2012, 2.48) such as formula (1) in 5.2. The measurement model is a mathematical expression or a set of such expressions (see JCGM 100:2008, 4.1.2), comprising all the quantities known to be involved in a measurement. It enables a value (see JCGM 200:2012, 1.19) of the measurand to be provided and an associated standard uncertainty to be evaluated. The measurement model may be specified wholly or partly in the form of an algorithm. The quantities to which the measurand is related constitute the *input quantities* (see JCGM 200:2012, 2.50) in the measurement model. The measurand constitutes the *output quantity* (see JCGM 200:2012, 2.51).

5.2 Many measurements are modelled by a real functional relationship f between N real-valued input quantities $X_1, \ldots, X_N$ and a single real-valued output quantity (or measurand) Y in the form

$$Y = f(X_1, \ldots, X_N). \tag{1}$$

This simple form is called a real explicit univariate measurement model; real since all quantities involved take real (rather than complex) values, explicit because a value for Y can be computed directly given values of $X_1, \ldots, X_N$, and univariate since Y is a single, scalar quantity. However, it does not apply for all measurements. A measurement model can be complex, involving *complex-valued quantities* (see JCGM 102:2011, 3.2). It can be *implicit* where a value for Y cannot be determined directly given values of $X_1, \ldots, X_N$ (see 13.5). The measurement model can be *multivariate* where there is more than one measurand, denoted by $Y_1, \ldots, Y_m$; for further information, see 13.4 and JCGM 102:2011.

EXAMPLE *Volume of a cylinder*

The volume of a cylinder is given by the measurement model

$$V = \frac{\pi}{4} L d^2$$

in which cylinder length L and diameter d are the $N = 2$ input quantities, corresponding to X_1 and X_2, and an output quantity V corresponding to Y.

5.3 The process of building a measurement model can be subdivided into the following steps, each step being described in the indicated clause:

a) Select and specify the measurand (see clause 6).

b) Model the measurement principle, thus providing a basic model for this purpose (see clause 7), choosing an appropriate mathematical form (see clauses 8 and 11).
c) Identify effects involved in the measurement (see clause 9).
d) Extend the basic model as necessary to include terms accounting for these effects (see clauses 10 and 11).
e) Assess the resulting measurement model for adequacy (see clause 12).

In any one instance, a number of passes through the process may be required, especially following step c). It may be more efficient or effective to take the steps as listed in a different order.

5.4 The manner in which a measurement model is used to obtain a value for the measurand (or values for the measurands) and evaluate the associated standard uncertainty (or covariance matrix) depends on its mathematical form (see clause 13). JCGM 100:2008 mainly considers explicit univariate models and applies the law of propagation of uncertainty (LPU). JCGM 102:2011 gives guidance on the use of generalizations of LPU for multivariate models and implicit models. For non-linear models, the use of the Monte Carlo method of JCGM 101:2008 (univariate measurement models) and JCGM 102:2011 (multivariate models) is often more appropriate (see clause 13).

5.5 The measurement model is a mathematical relationship among quantities, and as such it is subject to the rules of quantity calculus [20]. The same symbols used for the quantities are also used for the corresponding *random variables* (see JCGM 100:2008, C.2.2), whose *probability distributions* (see JCGM 101:2008, 3.1) describe the available knowledge about the quantities. Therefore, the measurement model can also be considered to be a model involving random variables, subject to the rules of mathematical statistics. The law of propagation of uncertainty as described in JCGM 100:2008, 5.1 and 5.2 uses a simple property of the transformation of *random variables* when only expectations and variances (and, perhaps, covariances) are used, rather than the whole distributions.

EXAMPLE *Mass of a spherical weight*

The mass of a weight that has been machined in the form of a sphere from a block of material is given by

$$m = \frac{\pi}{6} d^3 \rho, \tag{2}$$

where d is the diameter of the weight and ρ is the density of the material.

Expression (2) is a simple, well-known physical model that is idealized, applying to a perfect sphere and relating the output quantity mass to the input quantities diameter and density. At the same time, d and ρ can be considered as random variables describing the available information on the corresponding physical quantities obtained, for instance, from a dimensional measurement made with a vernier caliper and from a table of reference data, respectively. Thus, expression (2) also describes how to transform this information about the input physical quantities to the mass m of the weight (the measurand).

5.6 When building a measurement model that is fit for purpose, all effects known to affect the measurement result should be considered. The omission of a contribution can lead to an unrealistically small uncertainty associated with a value of the measurand [156], and even to a wrong value of the measurand.

Considerations when building a measurement model are given in 9.3. Also see JCGM 100:2008, 3.4.

5.7 The fitness for purpose of a measurement model can encompass considerations made before measurement. Such aspects include the measurement capability [85] (see JCGM 200, 2.6) in the case of a laboratory routinely performing calibrations. Fitness for purpose can also encompass the cost of measurement at a given level of uncertainty compared with the consequent costs of incorrect decisions of conformity (see also JCGM 106:2012 [21]). The measurand, which in the terminology of conformity assessment is a 'quality characteristic of the entity', can be, as in statistics, either

— a measure of 'location', for instance, a quantity relating to an entity such as the mass of a single object, an error in mass (deviation from a nominal value), or an average mass of a batch of objects, or
— a measure of 'dispersion', for instance, the standard deviation in mass amongst a batch of objects in a manufacturing process.

5.8 When developing a measurement model, the ranges of possible values of the input quantities and output quantities should be considered. The model should be capable of providing credible estimates and associated uncertainties for all output quantities over the required ranges of the input quantities, which should be specified as appropriate. The measurement model should only be used within the ranges of all quantities for which it has been developed and assessed for adequacy. See 13.2.

5.9 One aspect of specifying the domain of validity of the measurement model (see also 5.8) is to identify any restrictions on the domains of the quantities involved in the measurement model. Some quantities are necessarily positive (or at least non-negative). Some quantities might have lower and upper limits. There can be interrelationships between two or more quantities that need to be included in the measurement model.

EXAMPLE *Quantities having restrictions*

— Positive quantities, for instance, mass and volume.
— Quantity with limits, for instance, a mass fraction can only take values between zero and one.
— Quantities having interrelationships, for instance, the relative proportions (fractions) of all components (hydrocarbons and other molecules) of a natural gas sum to a constant.

Such quantities can sometimes be re-expressed by applying transformations. For instance, denoting by θ a new real quantity that is unconstrained:

— a quantity q is positive if re-expressed as $q = \theta^2$,
— a quantity q lies between a and b if re-expressed as $q = a + (b - a)\sin^2\theta$, and
— quantities q_1 and q_2 sum to unity by the transformation $q_1 = \sin^2\theta$, $q_2 = \cos^2\theta$.

6 Specifying the measurand

6.1 The choice of the measurand depends on the purpose of the measurement and may take account of the *target measurement uncertainty* (see JCGM 200, 2.34). Other processes and measurands are possible and the appropriate choice depends on the application of the measurement result.

NOTE Some measurands may be time-dependent such as in 10.6 and annex B.

EXAMPLE *Diameter of a cylindrical component*

In dimensional metrology, the diameter of a component of cylindrical form is obtained from knowledge of the profile of a right section of the component. If the component is a cylindrical piston, which is required to fit inside the cylinder of a piston-cylinder assembly, the measurand is the diameter of the minimum circumscribing circle (MCC) for the profile. Figure 1 (left) gives the MCC for a lobed profile.

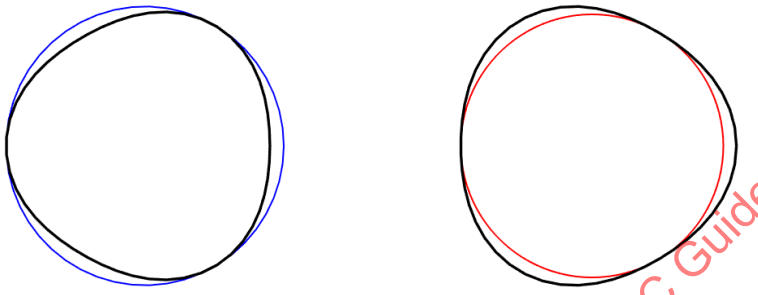

Figure 1: Minimum circumscribed circle (left, blue) and maximum inscribed circle (right, red), described by thin lines, for lobed profiles indicated by thick curved lines

If the component is a cylinder, which is required to contain a cylindrical piston in a piston-cylinder assembly, the measurand is the diameter of the maximum inscribed circle (MIC) for the profile. Figure 1 (right) gives the MIC for a lobed profile.

The profiles exhibit lobing due to the machining process used to produce the corresponding parts. For further information see reference [68] (Calibration Guide 6). Reference [6] describes how to determine MCC and MIC by expressing the problems as optimization problems with linear constraints, a 'standard form' for solution.

6.2 Taking account of any given target measurement uncertainty, the measurand should be specified sufficiently well so that its value is unique for all practical purposes related to the measurement. Regulations, legislation or contracts can contain stipulations concerning the measurand, and often these documents specify a measurand to the relevant extent, for instance, by reference to an international standard (such as ISO or IEC) or OIML recommendation. In general, an adequate specification of the measurand would often involve location in space and time of the measurement, or specification of reference conditions of, for example, temperature and pressure.

EXAMPLE *Length of a gauge block* (also see example in 13.3)

The (central) length of a gauge block is defined as the perpendicular distance between the central point of one end face and a plane in contact with the opposite end face, when the gauge block is at 20 °C.

6.3 In material testing, the measurand is often some property of an entire bulk of material under consideration. The measurement result to be obtained is required to be valid for the bulk (population) from which the sample is taken for measurement or testing. In such cases the measured value can be obtained through a process of sampling and measurement. Aspects such as the uncertainty arising from sampling, or also sample processing, are often

part of the specification of the measurand in these circumstances. Often these aspects are covered in a sampling plan (see, for instance, ISO/IEC 17025 [100]). Also see 6.1.

EXAMPLE 1 *Rockwell C hardness*

Rockwell C hardness is defined as hardness value, measured in accordance with ISO 6508, obtained using a diamond cone indenter and a force of 1 471 N [9,95]. Possible ways to specify the measurand relating to Rockwell C hardness of a material can relate to a specified point on the material (or on a sample from it) at a specified time, and mean Rockwell C hardness of the material. The latter is typically obtained as the average over designated points (in the material itself or in the sample) at a specified time. The specification of the measurand as 'Rockwell C hardness of the material' would only be adequate if the sample (and the material) were substantially homogeneous and stable for the intended use of the material.

EXAMPLE 2 *Emission monitoring*

Two possible measurands relating to a component in an effluent are its mass concentration at the time of sampling and its total mass over a calendar year. In the fields related to power production, emissions are often also qualified in terms of mass emitted per unit of energy produced.

EXAMPLE 3 *Biological material measurement*

For a consignment of biological material subject to measurement, the measurand might relate to a particular sample of the material, a set of samples, the method used to make the measurement, the laboratory performing the measurement, or a set of laboratories involved in making the measurement.

6.4 The specification of a measurand often requires describing the conditions for which the measurement result is valid.

EXAMPLE 1 *Catalytic activity*

The catalytic activity of an enzyme depends on the pH, temperature and other conditions. A complete specification of the measurand therefore requires these conditions to be specified.

EXAMPLE 2 *Calorific value of natural gas*

The calorific value (or, more correctly, the enthalpy of combustion) of natural gas is a function of temperature and pressure. When specifying the measurand, the relevant temperature and pressure are part of that specification. For example, reference conditions could be 288.15 K and 101.325 kPa [87]. Often there are contractually agreed reference conditions specified.

6.5 The measurement procedure should, as appropriate, address how the result is converted from the conditions of measurement to the conditions for which it is reported.

EXAMPLE 1 *Length of a gauge block* (also see example in 13.3)

The standard reference temperature for dimensional metrology is 20 °C [86]. This standard reference temperature is exact, so that the laboratory temperature, even in the most sophisticated systems, can only approximate it. Consequently, the measurement model for the length of the gauge block in the example in 6.2 contains a correction to that length based on i) the temperature difference between the reference temperature and the laboratory temperature (for example, 23 °C), and ii) the average coefficient of linear expansion in that range. Even when the *indicated* laboratory temperature is 20 °C and the value of the correction would be equal to zero, there would still be a non-zero associated uncertainty.

EXAMPLE 2 *Natural gas volume*

Natural gas volume is measured at the metering pressure and temperature of the gas in the transfer line. The volume is reported at reference conditions for pressure and temperature, which are agreed

between the contracting parties. The conversion of the natural gas volume at metering to reference conditions involves, among others, the compressibility factor of the natural gas at metering and reference conditions. The conversion is part of the measurement model.

6.6 In general, the same measurand can be represented by different models, depending primarily on the measurement principle chosen for its determination. Even within the same measurement principle, different models would result from different practical implementations of that principle, from the level of detail in the description of the measurement and the specific mathematical representation chosen among many that are often possible.

EXAMPLE *SI value of the Boltzmann constant*

The SI value of the Boltzmann constant [125] was obtained from three different measurement principles, to which correspond as many measurement methods: acoustic gas thermometry, dielectric constant gas thermometry and Johnson noise thermometry. A specific measurement model holds for each of these methods. In addition, differences exist even among the models of those laboratories using the same method, due to different practical implementations of the method, and perhaps to the level of detail in identifying effects.

6.7 Ultimately, the measurement model describes the realization of the measurand according to the knowledge of those involved in the measurement. The description may not fully represent the measurand because some unrecognized effects were not included. As a consequence, the uncertainty associated with the estimate of the measurand will not contain the contributions from those effects.

EXAMPLE *Mass comparison*

In modern mass comparators, the mass m_w of a standard weight W is determined (say, in vacuum) by comparing the forces $F_w = g_w m_w$ and $F_r = g_r m_r$ exerted on the comparator pan by W and by a reference mass standard R, subjected to local accelerations due to gravity g_w and g_r, respectively. The pan is held at a constant level by a servo mechanism. In most applications, it can safely be assumed that $g_w = g_r = g$, so that the model takes the simple form $m_w = m_r - \Delta m$, where Δm is the comparator indication and g has been dropped.

The assumption $g_w = g_r = g$ is correct when the centres of gravity of the two standards during weighings lie at the same elevation, and can anyway be adequate depending on the target uncertainty. If this is not the case, the model may need to be extended, usually with a correction (a further input quantity to the measurement model) to the mass m_r of the reference standard, having the form

$$m_r \frac{\Delta g}{\Delta h} \frac{1}{g} (h_r - h_w),$$

where $\Delta g/(g\Delta h) \approx -2.4 \times 10^{-7}\,\mathrm{m}^{-1}$ is the relative vertical gradient of the local acceleration due to gravity and $h_r - h_w$ is the difference in elevation between the centres of gravity of the standards being compared. Whereas the uncertainty associated with the correction is normally negligible, the correction itself can be significant. This is the case in the comparisons at the highest level of standards made of materials having very different densities (and consequently different sizes and sometimes shapes), such as silicon, stainless steel and platinum-iridium standards. Neglecting to include such a correction in the model would imply an incomplete description of the measurand, resulting in an error with respect to a correctly specified measurand. The error amounts to about $-3\,\mu g$ when comparing a stainless steel standard with a platinum-iridium prototype. This value is of the same order of magnitude of the standard uncertainty associated with the estimate of the measurand.

7 Modelling the measurement principle

7.1 General

The measurement principle (see JCGM 200, 2.4) enables a basic model, often based on a scientific law, or combination of such laws, to be established. A set of input quantities, namely, those quantities appearing in the laws underlying the measurement, follows from the measurement principle.

NOTE In many standardized test methods, as published by ISO and IEC for instance, the measurement principle is already described in the form of one or more mathematical formulæ.

EXAMPLE *Simple theoretical model of mass measurement*

The mass m of a weight is measured with a spring balance. The relationship between the restoring force F exerted by the spring and m at equilibrium is $F = mg$, where g is the acceleration due to gravity. Hooke's law relating F to the extension X of the spring and the spring constant k is $F = kX$. Thus, a theoretical measurement model relating mass to extension, spring constant and acceleration due to gravity is

$$m = \frac{k}{g}X.$$

Available knowledge about X, k and g enables knowledge about m to be inferred. Since Hooke's law only applies when the extension is sufficiently small, this measurement model should only be used within this limit. The largest mass, for a given k, for which the restoring force is proportional to the extension of the spring defines the domain of validity of the measurement model (see 5.8).

7.2 Theoretical, empirical and hybrid measurement models

7.2.1 A theoretical measurement model is based on scientific theory that describes how the output quantities in the measurement model relate to the input quantities.

EXAMPLE 1 *SI value of the Avogadro constant*

The Avogadro constant N_A is the universal proportionality constant relating in any sample the number $N(X)$ of entities X and the corresponding amount of substance $n(X)$,

$$N_A = \frac{N(X)}{n(X)}.$$

In the SI, the Avogadro constant has the exact value $N_A = 6.022\,140\,76 \times 10^{23}\,\text{mol}^{-1}$ [125]. One of the most accurate methods that was used to determine this value is the X-ray crystal density (XRCD) method [110]. In XRCD, the sample is a near-perfect sphere of mono-crystalline, isotopically enriched silicon (99.99 % ^{28}Si). The number of silicon atoms in this sample is $N(\text{Si}) = 8V/a_0^3$, V and a_0^3 being respectively the volumes of the sphere and of the unit atomic cell, containing 8 atoms. The amount of silicon is $n(\text{Si}) = m/M$, where m and M are the mass of the sphere and the molar mass of silicon. Therefore the definition of the Avogadro constant may be written as

$$N_A = \frac{N(\text{Si})}{n(\text{Si})} = \frac{8V}{a_0^3}\frac{M}{m},$$

which is the basic model describing the measurement principle for the XRCD experiment.

EXAMPLE 2 *Pressure generated by a pressure balance* (continued in 8.1.4)

The pressure p generated by a pressure balance is given implicitly by the measurement model

$$p = \frac{m\left(1 - \dfrac{\rho_a}{\rho_m}\right) g}{A_0\,(1 + \lambda p)\,[1 + \alpha\,(t - t_{\text{ref}})]}, \tag{3}$$

where m is the total applied mass, ρ_a and ρ_m are respectively the mass densities of air and the applied mass, g is the local acceleration due to gravity, A_0 is the effective area of the piston-cylinder assembly at zero pressure and a reference temperature, λ is the distortion coefficient of the piston-cylinder assembly, α is the temperature coefficient, t is Celsius temperature, and t_{ref} is a reference temperature, 20 °C, say [113]. There are eight input quantities $A_0, \lambda, \alpha, t, m, \rho_a, \rho_m, g$ and a single output quantity p.

7.2.2 Sometimes the measurement principle can only be formulated in terms of an empirical model. For such a model, the measurand would often be expressed in terms of mathematical functions such as polynomials, polynomial splines or rational functions, all of which have adjustable parameters. These models are often expressed as statistical models (see clause 11). Estimates of the parameters of these functions can be provided by reference tables (for instance, see example 2 in 8.3.3) or obtained by data analysis (least-squares regression, say), together with the associated standard uncertainties and covariances.

EXAMPLE *Industrial platinum resistance thermometer*

Platinum resistance thermometers of nominal resistance 100 Ω, also called Pt100 sensors, are used in industry for accurate measurement of temperature. The resistance R of the sensor as a function of Celsius temperature t is modelled over the interval [−200 °C, 800 °C] using the empirical Callendar-Van Dusen equation [51]

$$R(t) = \begin{cases} R_0\left[1 + At + Bt^2 + Ct^3(t - t_{\text{ref}})\right], & -200\,°\text{C} \le t < 0\,°\text{C}, \\ R_0\left(1 + At + Bt^2\right), & 0\,°\text{C} \le t \le 800\,°\text{C}, \end{cases} \tag{4}$$

where $t_{\text{ref}} = 100\,°\text{C}$, being a reference value, has no associated uncertainty, and R_0, A, B and C are determined by calibration. Typical values are $\widehat{R}_0 = 100\,\Omega$, $\widehat{A} = 3.9 \times 10^{-3}\,°\text{C}^{-1}$, $\widehat{B} = -6 \times 10^{-7}\,°\text{C}^{-2}$, and $\widehat{C} = -4 \times 10^{-12}\,°\text{C}^{-4}$.

Expression (4) is then used to provide values of the measurand t given values for the input quantities R_0, A, B, C and R and is thus an implicit measurement model (see 13.5).

7.2.3 Most measurement models are hybrid, that is, they combine aspects of theoretical and empirical models. Even if the complete measurement principle is based on scientific theory, there are often quantities involved in the measurement that need to be included in the measurement model and are modelled empirically (also see clauses 9 and 10).

EXAMPLE *Spectral irradiance of a lamp: hybrid model*

A tungsten filament lamp source emits light when its filament is heated by an electric current. The primary physics behind this process is that of a hot source and to first order the lamp acts in a manner similar to a blackbody source. The spectral radiance of a blackbody source can be modelled using the Planck function

$$L(\lambda, T) = \frac{c_{1,L}}{\lambda^5\{\exp[c_2/(\lambda T)] - 1\}},$$

where λ denotes wavelength, T the blackbody temperature and the coefficients $c_{1,L}$ and c_2 can be expressed in terms of fundamental physical constants [5].

Since lamps are not ideal blackbody sources, the emissivity of the tungsten filament and the transmittance of the bulb should be considered. For tungsten filament lamps used as standards, their spectral irradiance varies smoothly with wavelength and can be modelled by the product of a polynomial in λ and a Planck function. (Spectral irradiance is related to radiance by a constant factor depending on geometry.) Such a function is known as a Planck-polynomial function [41, 79]:

$$F(\lambda, T, a_1, \ldots, a_p) = L(\lambda, T) G_p(\lambda, a_1, \ldots, a_p). \tag{5}$$

The correction factor $G_p(\lambda, a_1, \ldots, a_p)$ is an empirical polynomial function of order p in λ with adjustable parameters $a_1, \ldots, a_p$. These parameters are estimated by fitting model (5) to suitable data using regression analysis.

7.3 Differential equation models

7.3.1 Many theoretical models used in science, engineering and medicine specify the rates at which some quantities vary as functions of time and other quantities, and are formulated as differential equations [152]. In some cases, the differential equation is only used as a foundation to develop simpler models.

EXAMPLE *Vapour pressure at vapour-liquid equilibrium*

The vapour pressure equations used in example 1 in 8.1.4 are solutions of the Clausius-Clapeyron relation, which is a differential equation,

$$\frac{\mathrm{d}p}{\mathrm{d}T} = \frac{\Delta s}{\Delta v}, \tag{6}$$

where p denotes the pressure, the measurand, and T the thermodynamic temperature of the vapour-liquid equilibrium, and Δs and Δv the differences in specific entropy and molar volume, respectively, between the two phases [11].

7.3.2 In measurement models that involve differential equations, the measurand is typically a function of coefficients in these equations. These coefficients are estimated based on empirical data, usually by application of a statistical method (see also clause 11). The equations are solved in the process, often a number of times, usually employing numerical methods, and typically a Monte Carlo method, similar to that described in JCGM 101:2008, is used for uncertainty evaluation [31, 55, 137, 153].

EXAMPLE *SIR model for influenza infection*

The numbers of cases reported once per week during an influenza epidemic in an isolated village with $n = 327$ inhabitants, during the first 11 weeks of their influenza season were $(1, 0)$, $(8, 0)$, $(15, 4)$, $(22, 26)$, $(29, 61)$, $(36, 67)$, $(43, 42)$, $(50, 40)$, $(57, 23)$, $(64, 13)$, $(71, 5)$. The first element of each pair is the day since the beginning of the season, and the second is the number of people known to have influenza on that day. For instance, the first pair means that on day 1 there were 0 cases, and the last pair means that on day 71 there were 5 cases.

The simple SIR model [102] for the spread of infectious diseases through a population does not contemplate births or deaths occurring during the course of the epidemic. It regards the population as partitioned into three subsets at each time instant t: those that are susceptible to the infection but not yet sick, $S(t)$, those that are infected, $I(t)$, and those that have already recovered from

the infection and are no longer contagious, $R(t)$. Their numbers add up to the population total, $S(t)+I(t)+R(t)=n$, and satisfy three simultaneous differential equations:

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -\beta I(t)S(t)/n, \qquad \frac{\mathrm{d}I}{\mathrm{d}t} = \beta I(t)S(t)/n - \gamma I(t), \qquad \frac{\mathrm{d}R}{\mathrm{d}t} = \gamma I(t), \tag{7}$$

for some specified non-negative initial values $S(0)$, $I(0)$ and $R(0)$. The model parameters are such that $1/\gamma$ is the typical duration of the infection in an infected person and the parameter of interest is the the average number of individuals directly infected by an infectious person in a population where everyone is susceptible, known as the *basic reproduction number* $R_0 = \beta/\gamma$. The model is completed by assuming that the actual number of cases at time t is a random variable $N(t)$ with a Poisson distribution with mean $I(t)$.

Computing the maximum-likelihood estimates [133] of β, γ and $S(0)$ involves finding the values of these parameters that maximize a product of Poisson probability densities (one for each of the 11 days with a reported number of cases), whose means are $I(1)$, $I(8)$, …, $I(71)$. Since equations (7) cannot be solved analytically, each time the likelihood function needs to be evaluated in the course of the maximization process, these differential equations have to be solved, also numerically. The resulting parameter estimates are $\widehat{\beta} = 0.34\,\mathrm{d}^{-1}$ and $\widehat{\gamma} = 0.14\,\mathrm{d}^{-1}$, and hence $\widehat{R}_0 = 2.4$. Figure 2 (left) depicts the data, the model as fitted to the observations, and the fitted values on the same days when the numbers of influenza cases were observed.

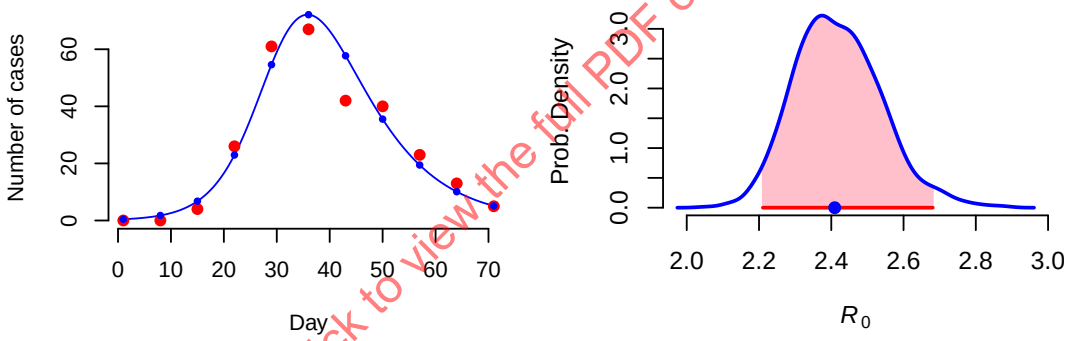

Figure 2: Observed numbers of cases (large, red dots), calibrated model $\widehat{I}$ (solid blue curve), and expected numbers of cases (small blue dots) corresponding to the observed numbers of cases and (right) probability density of R_0 based on a sample of size 1 000 drawn from its distribution by a Monte Carlo method, with the blue dot marking the maximum-likelihood estimate, the shaded area comprising 95 % of the area under the curve, and the thick, horizontal red line segment representing a 95 % coverage interval for the reproduction number

The uncertainty associated with the basic reproduction number was evaluated by application of the parametric bootstrap [53], a Monte Carlo method. This method is particularly onerous computationally, because it involves repeating the following steps a substantial number K of times: (i) draw one sample value from each of the 11 Poisson distributions whose means are the values $\widehat{I}(1)$, $\widehat{I}(8)$, …, $\widehat{I}(71)$ corresponding to the solution of equations (7) for $\beta = \widehat{\beta}$, $\gamma = \widehat{\gamma}$ and $S_0 = \widehat{S}_0$; (ii) estimate the same parameters, again via maximum likelihood, using the simulated draws, and compute the corresponding value of the basic reproduction number.

A sample of modest size $K = 1000$ drawn from the probability distribution of R_0 as just described (figure 2, right) had standard deviation 0.12, which is used as the standard uncertainty $u(R_0)$ associated with R_0. A 95 % coverage interval for R_0 deduced from the sample ranged from 2.2 to 2.7.

8 Choosing the form of the measurement model

8.1 General

8.1.1 The same measurement principle can give rise to different models. Therefore, the experimenter is often faced with the need to choose a suitable form for the model from various possible forms. This clause gives guidance on the criteria that can be used in making this choice. Some of the forms of model are considered, with examples, in clause 13.

8.1.2 A measurement model should be capable of accounting for all information that is relevant and available. The following aspects should be taken into consideration when choosing the form of the model (for further information or specific guidance, see the referenced subclauses):

Availability of a reliable theoretical form The extent to which a model can be based on the measurement principle, usually a well-established scientific law, should be considered. A model based on a reliable scientific principle is likely to require less work to check its adequacy than does an empirical or hybrid model. See clauses 7 and 12.

Target measurement uncertainty The target measurement uncertainty affects the degree of approximation that is appropriate, and will also affect the number of corrections, etc. taken into consideration (see JCGM 200, 2.34).

Simplicity It may be important to provide a model that can be implemented with minimal effort. Simpler representations can also help to avoid mistakes in implementation. The use of sub-models (see 8.4) can be useful in this respect. See 9.2.

Measurement range The model should be applicable over the whole range of values of the input and output quantities over which it is intended to be used. See 5.8. It may be appropriate to divide the range into subranges in each of which a different model applies. See the example in 7.2.2. Also see clause 12.

Dominant sources of uncertainty Where the dominant uncertainties arise from measurement of input quantities, models that associate uncertainties with input quantities are most appropriate. Where the dominant uncertainties relate to unpredictable or random variation in observations, models incorporating performance data are usually more appropriate. See annex C.5.

Relevance A model that generates coefficient or parameter values that can be interpreted in terms of physical quantities can be easier for users to apply than a model that for mathematical simplicity uses abstract functions of physical quantities.

Parsimony If a statistical model is fitted to data, the model should not include more terms than necessary. This aspect is often addressed by the statistical process of model selection. See 11.10.

Available information The availability of comprehensive data on performance of the measurement procedure can make it possible to use simplified measurement models. See 10.3 and annex C.5.

Numerical accuracy The model should be numerically as 'well-posed' as possible, that is, not adversely affected by the limitations of the available computational precision. See 8.6.

Solution stability Some models requiring numerical solution (and particularly some representations) can lead to unstable numerical performance. Models that lead to stable solutions are preferred. See 8.6.

Computational cost Models that are easier to evaluate or solve with adequate accuracy (see 13.3 and 13.4) are often preferred to those that might require considerable computational effort (see 13.5 and 13.6).

8.1.3 The aspects listed in 8.1.2 are often mutually incompatible. For instance, the simplest model is rarely the most accurate, and the most tractable mathematical form may not be easy to interpret. Choice of model is therefore a balance depending on local priorities. The overriding concern is whether the model is capable of providing a valid estimate together with an associated measurement uncertainty. When a target measurement uncertainty is available, the model should be capable of providing results consistent with it.

8.1.4 Consideration should be given to choosing the most appropriate form of model, as different models may be appropriate for different purposes. Example 1 in this subclause is concerned with two models, the choice of which is made on the basis of providing data for different physical properties. Example 2 illustrates the advisability of using an implicit model when it provides a more natural description of the measurement. It also emphasizes that both algebraically and numerically an implicit model might have advantages over an explicit model. The example on shell thickness of microscopic spherical particles in 13.5 demonstrates a case in which it is not possible to transform an implicit model into an explicit model. Such a situation is common.

EXAMPLE 1 *Saturated vapour pressure* (also see 7.3.1)

The dependence of vapour pressure p at saturation on thermodynamic temperature T is often used as the basis for determining and predicting the volumetric properties of fluids. A differential equation, the Clausius-Clapeyron equation (6) defines the first derivative $\mathrm{d}p/\mathrm{d}T$ of the saturated vapour pressure curve. A solution to this equation is the Antoine equation [7], a measurement model that takes the form

$$\ln \frac{p}{p_0} = A - \frac{B}{T + C}, \tag{8}$$

where p_0 is a reference pressure and A, B and C are coefficients, usually obtained by least-squares adjustment of vapour pressure data.

A different solution of the Clausius-Clapeyron equation, the Clarke and Glew equation [34], gives an alternative measurement model that relates vapour pressure data to thermodynamic functions:

$$R \ln \frac{p}{p_0} = -\frac{\Delta G^0(\Theta)}{\Theta} + \Delta H^0(\Theta)\left(\frac{1}{\Theta} - \frac{1}{T}\right) + \Delta C_p^0 \left(\frac{\Theta}{T} - 1 - \ln \frac{\Theta}{T}\right), \tag{9}$$

where R is the gas constant, Θ denotes a reference temperature, $\Delta G^0(\Theta)$ and $\Delta H^0(\Theta)$ are the molar free enthalpy and molar enthalpy differences between the vapour and the liquid phases at the temperature Θ, and ΔC_p^0 is the difference in molar heat capacity between the two phases.

Whereas measurement model (8), within a stated temperature range, conveniently reproduces experimental data within experimental uncertainty and interpolates vapour pressure data adequately, measurement model (9) is preferred when deriving thermodynamic properties from such data.

EXAMPLE 2 *Pressure generated by a pressure balance* (also see example 2 in 7.2.1)

Expression (3) in example 2 in 7.2.1 is a measurement model that defines implicitly the pressure p generated by a pressure balance. This model may equivalently be expressed as a quadratic equation

$$ap^2 + bp + c = 0 \tag{10}$$

in p, where

$$a = A_0\lambda[1+\alpha(T-T_{\text{ref}})], \qquad b = A_0[1+\alpha(T-T_{\text{ref}})], \qquad c = -m\left(1-\frac{\rho_a}{\rho_m}\right)g,$$

with solution in the form of the explicit model

$$p = \frac{\sqrt{b^2-4ac}-b}{2a}. \tag{11}$$

Expression (11) is one of the two roots $(-b \pm \sqrt{b^2-4ac})/(2a)$ of equation (10), and is positive since a is positive and the term $4ac$ is negative. The other root is negative, carrying no physical meaning. Also see example 4 in 8.6.5.

In some situations it might be necessary to use the measurement model

$$p = \frac{m\left(1-\frac{\rho_a}{\rho_m}\right)g}{A_0\,(1+\lambda p+\gamma p^2)[1+\alpha\,(T-T_{\text{ref}})]},$$

that is, with a polynomial of degree two (or higher) replacing the term $1+\lambda p$ in measurement model (3).

Consequently, the output quantity p corresponds to a root of a polynomial equation of degree at least 3, and it becomes more difficult (or impossible for degrees > 4) to derive an explicit measurement model for p in this case. Also see reference [40].

To an approximation that may be sufficient for the case in hand, p on the right-hand side of expression (3) can be replaced by a nominal value [113], and the measurement treated as explicit (see 13.3).

More complete measurement models can also be considered [113] that include, for instance, a correction to account for surface tension effects.

8.1.5 In calibration the relationship between the response variable and the stimulus variable can often be modelled using a polynomial of modest degree, often a straight line (polynomial of degree 1). Guidance on using a particular representation of polynomials (also see annex D), which assists in determining an appropriate polynomial degree, is available in ISO/TS 28038:2018 [99] and [37]. Also see example 2 in 8.3.3.

NOTE For many fields, specific guidance documents or international standards exist, providing advice on the selection of appropriate (polynomial) models for multipoint calibration.

8.1.6 In cases where there is more than one output quantity, the model should usually be treated as multivariate (see 13.4 and 13.6) rather than as separate univariate models (see 13.3 and 13.5) since knowledge is often required of covariances associated with the estimates of the output quantities. Instances arise in subsequent modelling and uncertainty evaluation and the determination of coverage regions (see JCGM 102:2011). Not taking such covariances into consideration generally leads to invalid statements of uncertainty.

EXAMPLE *Method of standard addition for determination of chemicals*

The method of standard addition is a popular way to estimate the concentration of various chemicals. It relies on partitioning a sample into several aliquots, adding to them known masses of a standard and measuring the analytical response from such spiked samples. A straight-line regression is determined from the data and the ratio of the intercept and the slope of this regression line estimates the amount of the analyte in the analyzed sample. Often the covariance between the intercept and the slope is neglected resulting in the uncertainty associated with the estimate from the method of standard addition being too small [121].

Data relating to the analysis of bromide in a water sample are

Mass fraction increment $x/\mathrm{mg\,g^{-1}}$	0	47.8	95.6	142.8	190.7
Instrumental response y	1.775	2.676	3.578	4.418	5.316

The measurement model for the mass fraction of bromide in the analyzed water sample is $w = a/b$, where a and b are the coefficients of the straight-line regression $y = a + bx$. Application of ordinary least squares gives estimated coefficients $\widehat{a} = 1.786$ and $\widehat{b} = 0.018\,52\,\mathrm{g\,mg^{-1}}$ with associated standard uncertainties and covariance

$$u(\widehat{a}) = 0.012\,4, \qquad u(\widehat{b}) = 0.000\,106\,\mathrm{g\,mg^{-1}}, \qquad u(\widehat{a},\widehat{b}) = -1.08 \times 10^{-6}\,\mathrm{g\,mg^{-1}}.$$

Alternatively, the correlation coefficient $r(\widehat{a},\widehat{b}) = u(\widehat{a},\widehat{b})]/[u(\widehat{a})u(\widehat{b})] = -0.817$ can be reported. If the coefficients of the straight-line regression are treated as the output quantities obtained independently from two univariate models, the estimate of w is $\widehat{w} = 96.4\,\mathrm{mg\,g^{-1}}$ with associated standard uncertainty $u(\widehat{w}) = 0.9\,\mathrm{mg\,g^{-1}}$ given by the law of propagation of uncertainty. If, however, the correlation between a and b is (correctly) taken into consideration, the standard uncertainty is appreciably larger: $u(\widehat{w}) = 1.2\,\mathrm{mg\,g^{-1}}$.

It is important to note that in this instance the data cannot be "mean-centred" (that is shifted with respect to x such that the arithmetic mean of the shifted x-values is zero), with a resulting zero correlation, before fitting the straight line. Doing so would yield a value $\widehat{a}$ different from that required.

8.2 Fitness for purpose and approximations

8.2.1 Ideally, a basic measurement model should describe a scientific law or relation known to hold for the true values of the input quantities and the output quantities. In practice, the measurement model usually represents only a practicable, fit-for-purpose approximation of the ideal relationship. This is so because, although that relationship may be known, it may be too complicated or inconvenient for practical use. Alternatively, the relationship may be unknown, but an empirical approximation to it may be sufficient for the intended purpose. In some instances, it may be possible to use a close approximation to the ideal relationship on a small number of occasions, and also work with a somewhat inferior model that would be more practical for routine use.

8.2.2 It is generally inadvisable to select a model purely for convenience, for example, because it results in a simpler analysis or software is available for its implementation. In exceptional cases, it is accepted that analysis can be difficult, or software is unavailable or hard to develop. In such a circumstance, where a simpler model is used, there will be an uncertainty associated with the estimate of the measurand resulting from that use, which should be determined to the extent possible, and a corresponding term included in the measurement model. See 11.10 concerned with model uncertainty.

8.3 Representation and transformation of models

8.3.1 Parametrization

Parametrization is the process of representing or expressing a measurement model, whether theoretical or empirical, in terms of its quantities or parameters. For some models (especially theoretical models) it is important to retain the contextual meaning of the parameters. Some parametrizations can be more convenient operationally, some more in

accordance with working practices and some more reliable numerically (see 8.6). Transformations of the parameters in a measurement model, or re-parametrizations, are frequently applied in metrology to re-express a measurement result in an alternative form for purposes of communication or to yield a simpler (or numerically more stable) computation. In the latter case, the results of the computation are transformed back to the original formulation. Great care should be taken when establishing and propagating uncertainties following transformation (see 8.3.6).

EXAMPLE 1 *Straight line in terms of Cartesian (x, y) co-ordinates*

A straight line can be represented in Cartesian co-ordinates as

$$y = a + bx, \tag{12}$$

with parameters a and b, where a is the intercept on the y-axis and b is the slope. This parametrization is appropriate when describing functional relationships between quantities.

EXAMPLE 2 *Straight line in terms of polar (r, θ) co-ordinates*

The same line can also be represented as

$$x \sin\theta + y \cos\theta = \rho,$$

with parameters θ and ρ, where θ is the angle between the line and the horizontal and ρ is its distance of nearest approach to the origin. This parametrization can be useful when modelling geometrical problems where the line may be near-vertical, since b diverges while θ remains finite.

8.3.2 Re-parametrization

Re-parametrization of a measurement model is the use of a different set of parameters that depend in some manner on the original parameters. For unchanged values of the input quantities the values of the output quantities are unchanged following re-parametrization.

EXAMPLE *Straight line in terms of shifted coordinates*

An alternative to the line (12) makes use of *shifted* coordinates:

$$y - y_0 = a_0 + b(x - x_0), \tag{13}$$

with parameters a_0 and b, where x_0 and y_0 are selected in some convenient way. The expression of the line (13) as $y = (y_0 + a_0 - bx_0) + bx$ exposes the relationship between the original and new parametrization. The slope parameter b is unchanged. It can sometimes be useful to work in terms of *shifted and scaled* coordinates such as

$$x_{\text{new}} = \frac{x - x_0}{s}$$

for suitable choices of x_0 and s.

8.3.3 Use in regression

Suppose x_i, y_i, $i = 1, \ldots, m$, represent measured coordinates of some underlying straight-line relationship. The use of representation (13) is often more desirable than the use of (12) since adverse numerical effects in the calculation can be more readily avoided.

EXAMPLE 1 *Drifting resistance standard* (also see 10.6.4)

The straight-line regression line in 10.6.4 is based on the shifted and scaled quantity

$$y = \frac{R - R_{\text{nom}}}{R_{\text{nom}}}$$

rather than the original quantity R. The values y_i of y (given in table 5 in 10.6.4) range from 17.900 8 to 18.691 6, whereas the corresponding values R_i of R range from 200.003 580 to 200.003 738. The former with dynamic range 1.04 (the ratio of the range endpoint values) are considerably more manageable than the latter (dynamic range 1.000 000 8).

EXAMPLE 2 *Thermoelectric voltage*

This example illustrates a circumstance in which the Chebyshev representation [37] of a polynomial function confers considerable advantages over the often-used monomial representation. The monomial and Chebyshev representations (annex D) of thermoelectric voltage are, respectively,

$$E = \sum_{r=0}^{n} c_r \vartheta^r, \qquad E = \sum_{r=0}^{n} a_r T_r(t), \tag{14}$$

in the reference function for Type S thermocouples, for Celsius temperatures ϑ in the interval [−50 °C, 1 064.18 °C]. The first of these expressions is given in [83] and the second is a re-parametrization of the first. The variable t in the right-hand expression (14) is given by

$$t = \frac{2\vartheta - \vartheta_{\text{min}} - \vartheta_{\text{max}}}{\vartheta_{\text{max}} - \vartheta_{\text{min}}} \tag{15}$$

and $T_r(t)$ denotes the Chebyshev polynomial of degree r in t (see annex D). There is a factor of some 10^{21} between the estimated non-zero coefficients $\widehat{c}_r$ of largest and smallest magnitude in the monomial form, which are held to 12 significant decimal digits (12S): presumably it was considered that care is needed in working with this particular representation. The $\widehat{c}_r$ are given to 5S in column 2 of table 1.

Table 1: Estimated polynomial coefficients for a Type S thermocouple

r	Raw, $\widehat{c}_r$/mV °C^{-r}	Scaled, $\widehat{d}_r$	Normalized, $\widehat{b}_r$	Chebyshev, $\widehat{a}_r$
0	0	0	4.303 6	4.639 1
1	$5.403\,1 \times 10^{-3}$	5.749 9	5.527 8	5.371 1
2	$1.259\,3 \times 10^{-5}$	14.261 8	0.478 4	0.370 6
3	$-2.324\,8 \times 10^{-8}$	−28.017 4	−0.054 3	−0.072 9
4	$3.220\,3 \times 10^{-11}$	41.300 5	0.220 6	0.037 1
5	$-3.314\,7 \times 10^{-14}$	−45.239 0	−0.163 7	−0.013 0
6	$2.557\,4 \times 10^{-17}$	37.144 7	0.021 6	0.002 2
7	$-1.250\,7 \times 10^{-20}$	−19.331 0	−0.024 9	−0.000 4
8	$2.714\,4 \times 10^{-24}$	4.464 8	0.025 2	0.000 2

A scaled variable $q = \vartheta/B$ has been used in ITS-90 in recent years, where in this instance B = 1 064.18 °C is the upper endpoint of the interval of definition. Then, $E = \sum_{r=0}^{n} d_r q^r$, with $d_r = B^r c_r$. The scaling implies that the contribution from the rth term in the sum is bounded in magnitude by $|d_r|$. Values of E in mV are typically required to 3D (three decimal places). Accordingly, estimates $\widehat{d}_r$ of the coefficients d_r are given in column 3 of table 1 to 4D (that is, including an extra digit) and are much more manageable. Alternatively, the variable can be normalized to the interval $[-1, 1]$ using the transformation (15) with $\vartheta_{\text{min}} = -50$ °C and $\vartheta_{\text{max}} = B$. Estimates $\widehat{b}_r$ of the corresponding coefficients b_r are given in column 4 and estimates $\widehat{a}_r$ of the Chebyshev coefficients a_r in column 5 of table 1, both to 4D, obtained using references [35, 158].

Figure 3 depicts the reference function (14). It curves only gently, but the non-linearity present cannot be ignored. The estimated coefficients in the monomial representation in terms of the raw or scaled variable in table 1 give no indication of the gently curved form in figure 3. However, in strong contrast, for the normalized and Chebyshev forms, the dominance of the magnitudes of the first two estimated coefficients indicate that the calibration function has an appreciable linear (straight-line) component. The Chebyshev coefficients for degrees 7 and 8 could arguably be replaced by zero, since to 3D they make little contribution, with the degree of the polynomial consequently reduced from 8 to 6. Such reasoning could not be applied to the other polynomial representations.

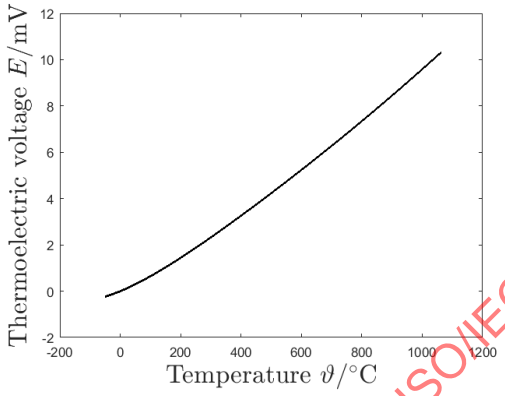

Figure 3: Relationship between temperature and thermoelectric voltage

8.3.4 Simple transformations

Some measurement models constitute simply transformation of the quantities involved ('transformation of variables').

EXAMPLE *Logarithmic and linear formats*

In radio-frequency and microwave electrical metrology, transformations are often made between logarithmic and linear representations of dimensionless reflection and transmission coefficients (scattering parameters or S-parameters). For an n-port electrical device, the expression is

$$\alpha = -20 \log_{10} |S| ,$$

where $|S|$ is the modulus of a (complex-valued) linear S-parameter and α is the corresponding logarithmic loss (that is, return loss or attenuation), measured in dB.

$|S|$ is defined as

$$|S| = (x^2 + y^2)^{1/2},$$

where x and y are the real and imaginary components of the linear S-parameter.

In this example, measured values of x and y are $\widehat{x} = 0.006\,0$ and $\widehat{y} = 0.008\,0$ and their stated associated standard uncertainties are $u(\widehat{x}) = u(\widehat{y}) = 0.010\,0$. An estimate $|\widehat{S}|$ of $|S|$ and thence $|\widehat{\alpha}|$ of α are obtained:

$$|\widehat{S}| = (\widehat{x}^2 + \widehat{y}^2)^{1/2} = 0.010\,0, \qquad \widehat{\alpha} = -20 \log_{10} |\widehat{S}| = 40.0\,\text{dB}.$$

These values typify the measurement of the linear reflection coefficient of a near-matched (that is, low-reflecting) device.

Applying the law of propagation of uncertainty (LPU) (JCGM 100:2008, 5.1.2) gives

$$u(|\widehat{S}|) = \frac{[\widehat{x}^2 u^2(\widehat{x}) + \widehat{y}^2 u^2(\widehat{y})]^{1/2}}{|\widehat{S}|} = 0.010\,0, \qquad \widehat{u}(\alpha) = \frac{20}{\ln 10}\,\frac{u(|\widehat{S}|)}{|\widehat{S}|} = 8.7\,\text{dB}. \tag{16}$$

The appropriateness of formula (16) for the standard uncertainty $u(|\widehat{S}|)$ is considered by also evaluating this uncertainty by applying the Monte Carlo method given in JCGM 101:2008. Application of this method with $M = 10^8$ trials, assuming x and y are independently normally distributed with means $\widehat{x}$ and $\widehat{y}$, standard deviations $u(\widehat{x})$ and $u(\widehat{y})$, and no correlation between x and y, gives the entries in column 3 of table 2, which are not in close agreement with the entries in column 2.

Table 2: Data and results for a logarithmic transformation using the law of propagation of uncertainty (LPU) and a Monte Carlo method (MC)

	$u(\widehat{x}) = u(\widehat{y}) = 0.01$		$u(\widehat{x}) = u(\widehat{y}) = 0.005$	
	LPU	MC	LPU	MC
$\lvert\widehat{S}\rvert$	0.010 0	0.015 5	0.010 0	0.011 4
$u(\lvert\widehat{S}\rvert)$	0.010 0	0.007 8	0.005 0	0.004 6
$\widehat{\alpha}$/dB	40.0	37.6	40.0	39.8
$u(\widehat{\alpha})$/dB	8.7	5.4	4.3	4.4

The M values of $|\widehat{S}|$ given by these trials can be used to construct a histogram, which when scaled appropriately, as described in JCGM 101:2008, 7.5.2, would give an approximation (which would improve as M is increased) to the probability density function (PDF) for $|\widehat{S}|$. This PDF is asymmetric, unlike the normal distribution assumed by the GUM. MC is expected to be more reliable in such circumstances.

Consider now that $u(\widehat{x}) = u(\widehat{y}) = 0.005\,0$ (rather than 0.010 0). The results are given in columns 4 and 5 of table 2. The difference between the outcomes from LPU and JCGM 101:2008 is now much less since the degree of approximation in the first-order Taylor expansion of the measurement function, which underpins LPU, is better over a shorter interval. Moreover, the PDF is less skewed.

In general, it is not recommended to perform uncertainty calculations using logarithmic quantities in, for example, dB. Logarithmic quantities should be converted to the equivalent linear format before performing uncertainty calculations.

8.3.5 Non-linear relationships

When a plot of data representing a relationship between two variables suggests an underlying non-linear behaviour, it is often possible to transform the raw data to make it 'more linear'. Doing so allows linear regression techniques to be used more effectively with such data. Re-parametrization in such an instance relates to the fact that the straight-line representation of the transformed data implies a different representation from that for the raw data. Such transformations generally change the statistical structure of the problem (see 8.3.6).

8.3.6 Impact on uncertainties

For uncertainty propagation, not all algebraically equivalent models behave necessarily in the same way because they are not equivalent statistically. An algebraic transformation should be accompanied by a corresponding statistical transformation to ensure the

integrity of the results obtained. The use of the law of propagation of uncertainty (JCGM 100:2008, 5.1.2) for this purpose is inadequate in some situations. Since the law of propagation of uncertainty is based on a linearization of the model, if errors originally followed a certain distribution, in the transformed problem they would generally follow a different distribution. Transformations have an impact on the uncertainties associated with the estimates of the quantities involved or on the weights in a statistical regression analysis involving such a model. See the example in 8.3.4.

NOTE Transformations also have an impact on the probability distributions used to characterize the quantities involved [129].

EXAMPLE *Re-parametrization of a model for lead adsorption by nano-particles*

The Langmuir adsorption model [111] describes the theoretical relation between the amount concentration Γ of a solute on the surface of an adsorbent to the amount concentration c of the solute in the liquid with which it is in contact:

$$\Gamma = \Gamma_{\max} \frac{Kc}{1 + Kc}, \tag{17}$$

where K is the Langmuir binding-strength coefficient and $\Gamma_{\max}$ (the measurand) is the maximum sorption capacity of the adsorbent. A reparametrized form of the Langmuir model is often used based on taking the reciprocal of both sides of model (17), which expresses $1/\Gamma$ as a linear function of $1/c$:

$$\frac{1}{\Gamma} = \frac{1}{\Gamma_{\max}} + \frac{1}{\Gamma_{\max} K} \times \frac{1}{c}. \tag{18}$$

The model parameters $1/\Gamma_{\max}$ and $1/(\Gamma_{\max} K)$ can be obtained by linear regression. Although models (17) and (18) are mathematically equivalent, different results are obtained for $\Gamma_{\max}$.

Consider measured values of c equal to

$$0.20\,\text{mmol}\,\text{L}^{-1}, \quad 0.25\,\text{mmol}\,\text{L}^{-1}, \quad 0.33\,\text{mmol}\,\text{L}^{-1}, \quad 0.60\,\text{mmol}\,\text{L}^{-1}, \quad 2.00\,\text{mmol}\,\text{L}^{-1},$$

and corresponding measured values of Γ equal to

$$9.4\,\text{mmol}\,\text{g}^{-1}, \quad 11.1\,\text{mmol}\,\text{g}^{-1}, \quad 13.9\,\text{mmol}\,\text{g}^{-1}, \quad 20.0\,\text{mmol}\,\text{g}^{-1}, \quad 33.3\,\text{mmol}\,\text{g}^{-1}.$$

Assuming c and Γ are independent, and not accounting for measurement uncertainties associated with Γ and c in order to illustrate the point, $\widehat{\Gamma}_{\max} = 46.4\,\text{mmol}\,\text{g}^{-1}$ with associated standard uncertainty $u(\widehat{\Gamma}_{\max}) = 0.3\,\text{mmol}\,\text{g}^{-1}$ is obtained by non-linear least-squares regression using expression (17).

Ordinary least-squares (OLS) fitting using expression (18) gives $1/\widehat{\Gamma}_{\max} = 0.021\,4\,\text{g}\,\text{mmol}^{-1}$ with associated standard uncertainty $u(1/\widehat{\Gamma}_{\max}) = 0.000\,6\,\text{g}\,\text{mmol}^{-1}$, which following reciprocation corresponds to $\widehat{\Gamma}_{\max} = 46.7\,\text{mmol}\,\text{g}^{-1}$ with associated standard uncertainty $u(\widehat{\Gamma}_{\max}) = 1.4\,\text{mmol}\,\text{g}^{-1}$.

Several other linearized versions of model (17) exist; it is emphasized that re-parametrization can have a significant effect on the results if the transformed uncertainties are obtained using the law of propagation of uncertainty, which assumes sufficiently accurate linearizability. Although both models are *algebraically* equivalent, OLS treatment of these models differs in its assumptions. The OLS treatment of (17) implies normality in the dispersion of Γ values whereas OLS treatment of (18) implies normality in the dispersion of $1/\Gamma$ values. These assumptions are inconsistent.

Although the values of $\widehat{\Gamma}_{\max}$ obtained by the two methods are consistent with respect to their associated standard uncertainties, this may not be the case in other circumstances.

8.3.7 Explicit and implicit forms of measurement model

When it is possible to transform an implicit expression to an explicit expression, there can be a preferred choice that depends on the particular circumstances. Sometimes an implicit form is preferable because it can have better numerical behaviour even when an explicit form can be deduced from it.

EXAMPLE *Representation of a circle*

In the assessment of geometric form in dimensional metrology [6] the equation for the co-ordinates (x, y) of a point on a circle centred at (a, b) with radius r is written implicitly as expression $(x - a)^2 + (y - b)^2 = r^2$. Then, y expressed explicitly in terms of the other quantities is given by expression $y = b \pm [r^2 - (x - a)^2]^{1/2}$. Because of subtractive cancellation the use of the latter expression becomes problematic when $|x - a| \approx r$. Alternatively, a convenient way to represent a circle is through the use of a parameter θ:

$$x = a + r\cos\theta, \quad y = b + r\sin\theta,$$

which is appropriate and numerically stable for values of θ between 0 and 2π and all co-ordinates (x, y).

8.4 Multi-stage measurement models

8.4.1 In many stepwise processes in metrology, quantities from intermediate measurements are naturally used in a subsequent measurement. Each stage in the process can be described by a measurement model with input quantities and output quantities. This set of measurement models constitutes a multi-stage measurement model and can be used as such. Equivalently, the complete process can sometimes conveniently be described by a single-stage measurement model, perhaps involving a different number of quantities.

8.4.2 Multi-stage models also arise as a natural consequence of the dissemination chain connecting ideally all kinds of measurements to the relevant SI units. Sometimes, to evaluate correctly the uncertainty associated with an estimate of the output quantities at a given stage, it is necessary to consider the previous stages in connection with the current stage. By implication, it can be necessary to consider the corresponding measurement models as a single multi-stage measurement model. An instance arises in mass and force measurements, in which covariances between volume and mass at the various stages of the dissemination chain must be taken into account to evaluate correctly the contribution of the buoyancy correction to the uncertainty associated with the output estimate of the current calibration [118].

EXAMPLE 1 *How long is a piece of string?*

Consider the problem of establishing a model for the measurement of the length L of a piece of string using a steel tape. Inevitably, idealizations and approximations are necessary. A simpler treatment involving fewer factors is available [15]. The measurand is the length L. It depends on several input quantities representing the observed string length and corrections to that length. Most of these quantities are expressed in terms of other quantities. The measurement model takes the form

String length = Observed string length (A)
+ Steel tape length correction (B)
+ String length correction (C)
+ Measurement process correction (D),

where

(A) Observed string length = Average of a number of repeated observations (Q_{11})

(B) Steel tape length correction = [Length deviation due to tape calibration imperfections (Q_{21})
+ Extension in tape due to stretching (this correction is negative if there is shrinking rather than stretching) (Q_{22})
+ Reduction in effective length of tape due to bending of the tape (Q_{23})]
× Multiplicative correction for steel tape thermal expansion (Q_{24})

(C) String length correction = Reduction in effective string length due to string departing from a straight line (Q_{31})
+ Reduction in string length as a result of shrinking (negative if there is stretching rather than shrinking) (Q_{32})

(D) Measurement process correction = Length deviation due to inability to align end of tape with end of string due to fraying of the string ends (Q_{41})
+ Length deviation due to tape and string not being parallel (Q_{42})
+ Length deviation due to reading a numerical value from the tape (Q_{43}).

The measurement model is thus expressed symbolically as

$$L = Q_{11} + (Q_{21} + Q_{22} + Q_{23})Q_{24} + (Q_{31} + Q_{32}) + (Q_{41} + Q_{42} + Q_{43}).$$

Each input quantity Q_{ij} would have an estimate and associated standard uncertainty, or be characterized by an appropriate probability distribution depending on the knowledge available. In the latter case, for instance, (a) a scaled and shifted t distribution is used for the quantity Q_{11} corresponding to the observed string length, based on repeated observations (see JCGM 101:2008, 6.4.9), and (b) a distribution based on a circular-arc model is used for the quantity Q_{23}. This example is continued in example 1 in 10.4.3.

The distribution for Q_{23} is related to the chi-squared distribution and does not have zero expectation, since the *minimum* effect of tape bending on the output quantity is zero. This degree of sophistication would not be warranted when measuring the length of a piece of string, but relevant in some other applications.

This example can be interpreted as a multi-stage model with the quantities representing the steel-tape-length correction, string-length correction and measurement-process correction described by sub-models.

EXAMPLE 2 *Peak area determined from spectral data*

A requirement in spectroscopy is the detection of peaks in a signal and the determination of peak parameters such as area and location. The effects of noise in the signal are reduced by first smoothing the data. Peak area can be considered in terms of a two-stage model or a single-stage model.

Two-stage model In a first stage data, with associated standard uncertainties, are filtered using a windowing function, of rectangular or triangular form, for instance. The result of the filtering is a set of smoothed values, with associated standard uncertainties and covariances. At least

some of these covariances are non-zero, since the filtered data depend on the original data. This stage involves an explicit multivariate model (see 13.4).

In a second stage a linear function of the filtered data is obtained such as peak area. To evaluate the standard uncertainty associated with peak area, explicit use is made of the standard uncertainties and covariances associated with the filtered data established in the first stage. This stage involves an explicit univariate model (see 13.3).

Single-stage model In a single-stage model, peak area is expressed directly as a function of the unfiltered data by combining explicitly the above two stages. As a consequence, the covariances generated in the above first stage require no consideration. This single-stage model is univariate and explicit (see 13.3).

The mathematics of these two approaches is as follows:

Two-stage model In the first stage data $\boldsymbol{y}$ representing a spectral signal are filtered using a windowing function. The result of the filtering is a vector $\boldsymbol{z} = \boldsymbol{C}\boldsymbol{y}$, where $\boldsymbol{C}$ is a matrix depending on the windowing function. If the covariance matrix associated with $\boldsymbol{y}$ is $\boldsymbol{U}_y$, the covariance matrix associated with $\boldsymbol{z}$ is $\boldsymbol{U}_z = \boldsymbol{C}\boldsymbol{U}_y\boldsymbol{C}^\top$ (see JCGM 102:2011, 6.2.1.3). In the second stage a linear function $p = \boldsymbol{w}^\top\boldsymbol{z}$ of the filtered data $\boldsymbol{z}$ is obtained such as peak area. The variance associated with p is $u^2(p) = \boldsymbol{w}^\top\boldsymbol{U}_z\boldsymbol{w}$.

Single-stage model The two stages when combined into a single stage become $p = \boldsymbol{w}^\top\boldsymbol{C}\boldsymbol{y}$ and $u^2(p) = \boldsymbol{w}^\top\boldsymbol{C}\boldsymbol{U}_y\boldsymbol{C}^\top\boldsymbol{w}$. The covariance matrix $\boldsymbol{U}_z$ is not involved in the combined calculation.

The signal is often superimposed on a background that is estimated by a constant, inclined or curved baseline and subtracted from the original signal before peak processing. This subtraction introduces further correlation.

8.4.3 Calibration (see 11.4) is an important instance of a multi-stage model, where the determination of a calibration function is followed by its use. Subdivision into stages is necessary when one laboratory determines a calibration function and another uses it. When there are several calibration parameters (calibration coefficients, say), the parameter covariances as well as their standard uncertainties have to be made available to the next stage. Combining stages, when that is possible, avoids the need to work with these covariances. An instance is example 2 in 8.4.2 in which many covariances would be passed from the first stage to the second. A further instance is the example in 8.1.6. Sometimes a change of variables can give an alternative representation of a model, which avoids the use of covariances; see, for instance JCGM 100:2008, annex H.3.

8.5 Uncertainty associated with choice of model

The choice of measurement model should be examined if it is considered it will have an effect on the estimate of the measurand that is significant compared with the uncertainty associated with that estimate. For some classes of model, such as those that naturally include terms the number of which depends on the available data as in polynomial calibration, guidance is available on choosing a model [99]. See 11.10 for model uncertainty in the context of statistical models. Often, simply an awareness of the issue and its possible repercussions is the main consideration. Also see 8.2.2.

8.6 Loss of numerical accuracy and its compensation

8.6.1 Inaccuracy can occur in numerical computation as a result of using formulæ or algorithms which, although mathematically correct, are not suitable for the task in hand.

Such inaccuracy can arise in many areas such as calculating basic statistics (for instance, mean and standard deviation of a set of data), fitting polynomials to data and simply evaluating a formula. Specific causes of such behaviour include numerical errors due to loss of significant digits arising from subtractive cancellation, build up of rounding error, and computer underflow or overflow.

8.6.2 The need to evaluate measurement functions given as formulæ is widespread in metrology. Formulæ are provided in papers, reports, data sheets, procedures, specifications, guides and documentary standards, and elsewhere. They are used to express both theoretical and empirical relationships between quantities. Metrologists arguably have the right to expect that a formula provided in a reputable source is fit for the purpose for which it was developed. In other words, the formula can be directly evaluated manually or implemented on a computer with confidence. It is not always the case that a faithful implementation of a given formula yields a result with the numerical accuracy that might be expected, especially compared with the use of alternative mathematically equivalent formulæ. In some instances the loss of numerical accuracy may not be acceptable.

8.6.3 Since a computer will be used for most calculations of the type considered in this guide, computations will naturally be carried out using all digits available (typically about 16 decimal digits in a computer wordlength). This number of digits is usually more than adequate for providing measurement results to the numerical accuracy required in practice. Certain formulæ or algorithms, however, are such that the computed measurement results might not have sufficient numerical accuracy due to such causes as in 8.6.1. A reasonable choice of parametrization can help to avoid such a possibility.

8.6.4 If the measurement results are to be used in a further calculation, sufficient numerical accuracy should be retained in the values to be passed to that calculation. In accordance with 8.6.3, whenever possible calculations should proceed with the numerical values represented on the computer rather than the rounded values that may be used for reporting purposes. A common occurrence is when a calibration function is used in a first step to obtain calibration coefficient estimates, uncertainties and covariances, and, in a second step, this information is used to provide stimulus values and associated uncertainties corresponding to given response values and associated uncertainties. The use of rounded values from the first step may provide insufficient numerical accuracy in the required stimulus values and in the associated uncertainties.

8.6.5 Compensation can often be made for loss of numerical accuracy, especially if it is anticipated that the computed results might be adversely affected. Rearranging, simplifying or re-parametrizing a mathematical expression can in some instances improve substantially the numerical accuracy of the evaluated result. Although it is beyond the scope of this guide to give extensive advice (that is the province of numerical analysis [162]), the examples below are indicative of numerical difficulties that can arise and the manner in which they can be overcome. The use of good quality mathematical software libraries is recommended.

EXAMPLE 1 *Subtractive cancellation in calculating mean and standard deviation*

For a set of values x_i, $i = 1, \ldots, n$, which may be repeated observations, obtained independently, of some quantity, formulæ for their mean $\bar{x}$ and variance (squared standard deviation) s^2 are

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2. \tag{19}$$

These formulæ are numerically stable. An alternative, mathematically equivalent formula is

$$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right), \tag{20}$$

a formula that should not be used for high-precision metrological data since it is potentially numerically unstable [39]. For data having r leading decimal digits in common, typically $2r$ decimal digits are lost by applying equation (20).

Mass data, in grams, obtained under repeatability conditions of measurement using a comparison method [17] are

1 000.000 712 2 1 000.000 712 3 1 000.000 712 6 1 000.000 712 7
1 000.000 712 8 1 000.000 712 8 1 000.000 712 9

s^2 given by formula (19) is $7.1 \times 10^{-14}\ \mathrm{g}^2$ and that given by formula (20) is $3.1 \times 10^{-10}\ \mathrm{g}^2$, using 16 significant decimal digits in calculations [84]. These values differ by some four orders of magnitude. The corresponding values of s are 0.27 µg and 18 µg. The latter value is clearly incorrect since the data values on which s is based differ from each other by less than 1 µg. The reason for this failure is that catastrophic subtractive cancellation takes place in formula (20) due to minute numerical rounding errors: the two contributory terms in parentheses in formula (20) are respectively $7.000\,009\,976\,603\,555 \times 10^6\ \mathrm{g}^2$ and $7.000\,009\,976\,603\,553 \times 10^6\ \mathrm{g}^2$, differing by 2 in the 15th significant decimal place.

A second instance of mass measurement data, again in grams, is

1 000.000 094 8 1 000.000 095 0 1 000.000 095 0 1 000.000 095 0
1 000.000 095 0 1 000.000 095 1 1 000.000 095 0

For this data, formula (19) gives $s^2 = 8.1 \times 10^{-15}\ \mathrm{g}^2$ and formula (20) gives $-3.1 \times 10^{-10}\ \mathrm{g}^2$. Again, catastrophic subtractive cancellation takes place in formula (20), so much so that the variance is computed as negative, an impossible value. The two contributory terms in parentheses in formula (20) are $7.000\,001\,329\,800\,062 \times 10^6\ \mathrm{g}^2$ and $7.000\,001\,329\,800\,064 \times 10^6\ \mathrm{g}^2$, respectively, which differ by -2 in the 15th significant decimal place.

As both sets of data have some 9 leading decimal digits in common, about 18 digits would be expected to be lost by unstable formula (20). Since about 16 decimal digits are available, no correct digit in the computed s^2, and hence s, can be anticipated.

Consider n, the number of values of x_i, to be very large, 10^6 or 10^7, say, as arises, for instance, in a Monte Carlo implementation of the propagation of distributions (see JCGM 101:2008). The direct use of formulæ (19) not only requires all the x_i to be available, but also necessitates two passes through the data, first to calculate $\bar{x}$ and then s^2. Incremental, 'one-pass', forms of expressions (19) are available [76, chapter 4]. They provide numerical accuracy comparable to that for expressions (19), but use minimal memory.

EXAMPLE 2 *Inadequate form of polynomial for evaluation purposes*

Consider the evaluation of a particular polynomial $p(x)$ of degree six [146, p125] in the region of a minimum of p. The function is convex, with essentially flat behaviour in the middle of the interval. Figure 4 shows p evaluated at 101 equispaced x-values in the interval $[0.99, 1.01]$ with successive points joined by straight lines. The behaviour is highly erratic, and nothing like the smooth behaviour expected of a polynomial of modest degree. The irregularities correspond to large relative numerical errors in the computed values of the polynomial.

The polynomial in monomial form is

$$p(x) = 1 - 6x + 15x^2 - 20x^3 + 15x^4 - 6x^5 + x^6, \tag{21}$$

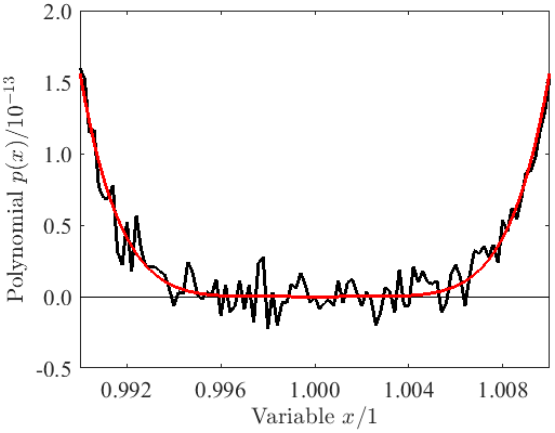

Figure 4: Polynomial evaluated using a numerically stable formula and (irregular curve) an unstable formula with successive evaluation points joined by straight lines

which was the form that was evaluated in creating the 'jagged' line. The evaluation of the form (21) near $x = 1$ suffers appreciable subtractive cancellation: individual terms in the polynomial (for instance, $15x^2$ and $-20x^3$) have magnitudes of between 1 and 20 near $x = 1$, yet the value of p is very much smaller, for instance, of the order of 10^{-13} at $x = 0.99$. So many (some 14) digits are lost for this value of x.

The polynomial $p(x)$ is in fact the expansion of $(1-x)^6$, a form that is perfectly suitable for direct evaluation. The smooth curve in figure 4 was formed using this representation.

Imagine using the form (21) for the determination of a root of $p(x) = 0$, as would be required when working with an implicit model. The graph shows many spurious 'zeros' (the points of intersection of the irregular curve and the horizontal line) induced by the noisy function values. More strictly, there are many pairs of adjacent points when the computed value of the function takes opposite signs.

A similar situation would arise when determining the minimum of p. The noise induced into the values of the function by the use of representation (21) would cause a poor estimate of the minimum to be returned. Indeed, there are many local minima. It would also be likely to introduce difficulties for the minimization routine, which is probably designed to operate effectively for smooth continuously differentiable functions. This function has all these properties mathematically, but not numerically in the field of floating-point arithmetic.

EXAMPLE 3 *Underflow and overflow in the evaluation of the geometric mean*

The geometric mean x [38] of n positive values $x_1, \ldots, x_n$ is required. For each x_i approximately equal to 0.5 and $n = 2\,000$, the geometric mean obtained using the 'natural' formula $x = (x_1 \times \cdots \times x_n)^{1/n}$ is computed as zero in IEEE arithmetic [84], as implemented on most computers, because of underflow. The correct result is approximately 0.5. The difficulty can be overcome by forming $(1/n)\sum_{i=1}^{n} \ln x_i$ and exponentiating the result. If, on the other hand, each x_i is now greater than unity, approximately 2, say, and $n = 2\,000$, say, the 'natural' calculation will overflow, but the modification will provide a reliable value.

EXAMPLE 4 *Roots of a quadratic equation*

The textbook formula for the roots of a quadratic equation is not always numerically ideal. Consider example 2 in 8.1.4. For certain values of a, b and c (for instance, given by $T \approx T_{\text{ref}}$, $\rho_a \ll \rho_m$, $\lambda \ll mg$), a value of p obtained from expression (11) may be numerically inaccurate, as a result of subtractive cancellation, and thus not be fit for purpose. A recommended approach to forming the output estimates given input estimates is as follows. Let the two roots of the quadratic equation (10) be z_1 and z_2 with $|z_1| \geq |z_2|$. Then, since the coefficients a and b in this equation are positive for this application, z_1 and z_2 are evaluated (without subtractive cancellation) from

$z_1 = (-b - \sqrt{b^2 - 4ac})/(2a)$ and $z_2 = c/(z_1 a)$, using a property for the product of the roots of a quadratic equation.

9 Identifying effects arising from the measurement

9.1 The basic model describing the measurement principle [step b) in 5.3] holds in ideal conditions. That model should usually be extended to cover effects arising in the practical implementation of the measurement. These effects should be identified and considered for inclusion in the model to obtain a measurement model adequate for the purpose [steps c) and d) in 5.3].

EXAMPLE *Local acceleration due to gravity*

The local acceleration due to gravity g (the measurand) is determined in absolute gravimeters by a free-fall experiment in which a body initially at rest falls a distance h under the effect of gravity. The measurement principle is based on the model $h = gt^2/2$ or, equivalently, $g = 2h/t^2$, where t is the time for the fall. This model is approximate, strictly holding only in a vacuum. In reality, an additive correction needs to be introduced in the model to accommodate the fact that air friction alters the law of motion of the body. The correction is proportional to the air density, the cross-sectional area of the body and the square of its velocity. These quantities, be they determined experimentally or estimated by other means, are themselves uncertain. Their uncertainties contribute to the uncertainty associated with the correction and in turn to the uncertainty associated with the measurand. Also, since the model is approximate in that g is not strictly constant over the distance h, a more elaborate model may be constructed to account for that effect.

9.2 Effects arising from the measurement (see 9.3) are generally accounted for in the measurement model depending on the available knowledge about them. Subclauses 9.3 and 9.5 and clause 10 provide guidance on extending the measurement model accordingly.

9.3 The identification of the effects arising in the practical implementation of the measurement is one of the most demanding tasks in the construction of a fit-for-purpose model. Areas of consideration include, but are not limited to, contributions from the following effects (compare reference [15]):

— realization of the specification of the measurand,
— approximations and assumptions in modelling the measurement,
— conditions of measurement,
— external influences (including environmental effects),
— drift of the measuring system,
— measurement standards and certified reference materials,
— artefacts, objects or materials to which the measurement result is applicable,
— instability or inhomogeneity of the artefact or material subject to measurement,
— sampling,
— sample preparation,
— calibration,
— reading analogue and digital measuring systems,
— system resolution or discrimination threshold,
— zeroing of an instrument, and
— variations in and correlations between observations.

NOTE 1 The effects are not necessarily independent.

NOTE 2 It is often more practicable to assess and include in the model combinations of these effects, rather than each effect separately.

9.4 During the execution of a measurement procedure, many factors need to be kept within specified limits. In principle, all these factors should be included in the measurement model.

9.5 Identifying influences on a measurement can be facilitated by the use of a cause-and-effect diagram, also known as an Ishikawa or 'fishbone' diagram [61–63]. A brief description of cause-and-effect diagrams and their use for measurement modelling is given in Annex E.

10 Extending the basic model

10.1 General

10.1.1 There are various ways of extending the basic model to account for the effects identified in 9.3. The purpose of elaborating the basic model is to complement it to become a measurement model that includes all quantities and effects that influence the measurement result. Two important choices should be made when including an effect in the model:

a) whether to model the effect as an effect on the observed values of an input quantity in the basic model, or as an effect on the observed values of the output quantity, and
b) whether the effect can be described sufficiently well that a correction can be made for it.

10.1.2 The effects arising from the measurement can often be described as an effect on one or more input quantities. Consequently, it is appropriate to model such effects in relation to the input quantities being affected. Another frequently occurring situation arises when an effect cannot be easily described as an effect on an input quantity. Such effects should be incorporated in the measurement model as an effect on the output quantity. Such situations arise in all kinds of product and material testing, and also in the production of reference materials when evaluating batch homogeneity or stability [97]. This modelling choice is further elaborated in 10.2.

10.1.3 In considering whether an effect can be described sufficiently well to allow correction, two broad classes can usefully be distinguished:

Well-understood effects Many effects can usefully be measured or estimated with sufficiently small uncertainty to improve the measurement. Others may be known to have small uncertainties compared with the target measurement uncertainty. These well-known or readily measurable systematic effects will be considered here as 'well-understood effects'.

Poorly understood effects Other effects are known to exist, but little can be said about their magnitude or even their sign in a particular measurement. These include random effects, which generally give rise to a different deviation for every instance of a particular random effect. Similarly, some systematic effects cannot readily be measured or characterized sufficiently for correction in a particular measurement, for example, because they depend heavily on unknown characteristics of each different test item. It is not possible to make a useful correction for either of these effects. They can, however, contribute perhaps appreciably to the measurement uncertainty. These will be considered as 'poorly understood effects'.

Subclauses (10.3 and 10.4) give guidance on extending a basic measurement model to include these two broad types of effect.

NOTE The classification above is for convenience in discussion. Because it relies on the degree to which a particular effect is understood in a specific case, an effect that is 'poorly understood' in one circumstance can, for example, because of more extensive study of the measurement, be modelled as a 'well-understood' effect in different circumstances.

10.2 Adding effects to the basic model

10.2.1 An input quantity can often be described as a function of several other input quantities, as demonstrated in the following example.

NOTE This form of modelling is also known as a 'bottom-up' approach to uncertainty evaluation [3].

EXAMPLE *Mass concentration computed from preparation data*

Suppose the mass concentration of substance in a solution is calculated from the mass m of the substance and the volume V of the solution. As is common in analytical chemistry, a balance is used to determine m and a volumetric flask is used to determine V. The measurement model takes the form

$$\gamma = \frac{m}{V}, \tag{22}$$

where γ denotes the mass concentration. The volume of the solution is affected by the repeatability of filling and the temperature, and largely determined by the volume of the flask V_{flask}. Hence, incorporating terms δV_{r} and δV_T to represent these respective effects,

$$V = V_{\text{flask}} + \delta V_{\text{r}} + \delta V_T.$$

This expression can be substituted in the measurement model (22) to account for these effects.

The effect of the purity of the substance can be modelled by developing an expression for m that incorporates a correction for purity, w. If this correction is available, then

$$m = m_{\text{gross}} w,$$

where m_{gross} is the mass calculated from weighing the substance. For m_{gross}, a measurement model is also constructed that typically includes the readings of the balance when weighing the empty container and when weighing the container with the substance, among others. A fully worked example is given in the Eurachem/CITAC Guide [63, appendix A, example A1].

10.2.2 Another frequently occurring situation arises when an effect of, for example, sampling, sample preparation or the measuring system, is more conveniently evaluated in terms of the output quantity.

NOTE This form of modelling is also known as a 'top-down' approach to uncertainty evaluation [3].

EXAMPLE *Ash content of coal*

The ash content of coal (expressed as mass fraction) is a function of temperature. However, it is not possible to describe this relationship mathematically, due to the complexity of the chemical reactions involved in converting the coal sample into ash. Depending on the conditions of ashing, ash is produced with a different composition and hence a different mass.

A laboratory evaluates the temperature effect by repeating under repeatability and (within-laboratory) reproducibility conditions the test on a set of laboratory samples from the same coal, over a number of days, to assess repeatability and reproducibility. The standard deviations thus obtained relate to the ash content of the coal. The temperature effect, which cannot be described as a relationship between the ash content and the (fluctuations of the) ashing temperature, would be incorporated in the measurement model as part of a precision term associated with the output quantity. Whether a bias component needs to be included would usually be assessed by using a certified reference material or retained sample from a proficiency test.

10.2.3 The mathematical form in which an effect is modelled depends primarily on whether this effect is in magnitude constant over the range of values concerned for the corresponding quantity, or whether some other relationship is determined. In cases where there exists an established functional relationship, such a relationship should be used, or a simplification of such a relationship if it can be demonstrated that it is fit for purpose.

EXAMPLE 1 *Volume of a solution*

The contraction or expansion of a liquid can be described as a function of the temperature by using the cubic expansion coefficient [63]. Over large temperature ranges, this coefficient is generally also a function of temperature, so that a different mathematical expression can be needed depending on the range of temperatures contemplated when preparing solutions.

EXAMPLE 2 *Air buoyancy*

Air buoyancy can be described as a function of the volume of the object being weighed and the air density. The air density can for example be computed from the CIPM 2007 formula, which relates the air density to pressure, temperature, relative humidity and the amount fraction of carbon dioxide [135]. A simpler form of the correction would use a constant value for the air density, with a suitably large uncertainty that reflects the variability of the air density the year round.

10.3 Modelling well-understood effects

10.3.1 Well-understood effects exert an influence on the estimate of the measurand whose magnitude can be determined and for which a correction can usefully be considered. They are included in the measurement model by introducing input quantities as corrections or correction factors (see JCGM 200, 2.53) to compensate, respectively, for additive effects (absolute) or multiplicative effects (proportional to the measurand). The associated uncertainties describe the residual doubt about how well the corrections or correction factors are known. Some input quantities might themselves be known to depend on further quantities, and so the measurement model would be expressed in terms of an enlarged set of input quantities. Equivalently, these input quantities can be regarded as output quantities from further measurement models in a multi-stage measurement model (see 8.4).

10.3.2 In many cases the corrections or correction factors can be described by mathematical expressions that represent the physical phenomena causing the effects.

EXAMPLE *Buoyancy in mass measurement*

If the mass measurement in the example in 7.1 is carried out in air, an additive correction due to air buoyancy of the form $\rho_a V$ is introduced, where ρ_a is the air density and V the volume of the weight. The model becomes

$$m = \frac{k}{g} X + \rho_a V.$$

In turn, the volume is typically obtained from a known value V_{ref} determined at a given reference temperature t_{ref} ($t_{ref} = 0\,°C$ or $t_{ref} = 20\,°C$), and corrected by taking into account the temperature t at which the measurement is made and the average coefficient of thermal expansion α of the material that constitutes the weight in the temperature range between t and t_{ref}. Therefore,

$$V = V_{ref}[1 + \alpha(t - t_{ref})].$$

This expression can be introduced in the model, or be considered as a sub-model in a multi-stage measurement model (see also 8.4). Similar considerations can be applied to other input quantities.

10.3.3 If there is no known relationship between the effect and the quantity concerned, then often the dependence is determined experimentally, for example, in method evaluation or method verification studies [100]. Also, consideration should be given to the possibility that the uncertainty component arising from the effect studied is not necessarily constant over the range of values considered for the quantity causing the effect. A well-established experiment provides data over this range, so that it can be assessed whether the uncertainty component is best approximated by a contribution that is independent of or proportional to the value of the measurand or has some other relationship to the measurand. In the first situation, the effect should be modelled as an additive term, whereas in the second situation a multiplicative factor is the appropriate choice.

10.3.4 The correction corresponding to a well-understood effect should always be included in the measurement model in the form of a mathematical expression. How to determine the (unknown) value of the correction is a matter of choice, largely depending on the target uncertainty. The value can be determined experimentally during the measurement, or be estimated from other information, or be taken equal to zero, if additive, or to one, if multiplicative, in which cases the correction does not impact on the estimate of the measurand. Whichever its value, a correction in a measurement model contributes to the uncertainty associated with the estimate of the measurand.

10.3.5 The practice of not including in the measurement model a correction having known value and applying instead an inflated uncertainty is scientifically unsound and is deprecated (see JCGM 100:2008, F.2.4.5).

10.3.6 The greater the effort that is devoted to evaluating a correction, the better is the available knowledge about it, and the more reliable is the resulting associated uncertainty. Simplifying a measurement model by neglecting corrections, or by including only a simple approximate correction for a more complex effect, can significantly reduce the effort required to characterise a measurement and evaluate its uncertainty. On the other hand, the additional uncertainties associated with the neglected effects or with the approximations employed are generally larger and less reliable than if the effects were characterised and corrected as necessary.

EXAMPLE *Mass measurement (continued)*

In the example in 10.3.2, the temperature t can be measured or, based on prior knowledge about the laboratory conditions, it can be taken as t_{ref}. Depending on the decision made, $t - t_{\text{ref}}$ has either a non-zero value with an uncertainty equal to that associated with the estimate of t, or is equal to zero with a larger uncertainty depending on the known temperature fluctuations in the laboratory. Also the air density can be measured, with state-of-the-art relative uncertainty of, say, some parts in 10^5 or can be taken equal to an average value with an associated relative standard uncertainty typically between 0.1 % and 1 %.

10.3.7 The discussion above explains an apparent paradox arising when the same type of measurement is carried out at the highest accuracy and at the industrial level. In the former case, a complete model is used, containing all the corrections, which are determined experimentally. In the latter case, in which the target uncertainty is much greater, most of the corrections are typically estimated from other information (Type B evaluation) or simply 'neglected', so that the model contains just a few terms having, however, a much larger uncertainty. As a result, for the same kind of measurement a model containing a large number of terms yields an estimate having an uncertainty smaller than that provided by a simpler model containing a smaller number of terms. This somehow counter-intuitive fact is explained by the difference in the effort devoted to evaluate corrections in the two cases.

10.3.8 There are circumstances in which an effect known to affect the measurement result can be represented only as a correction (or correction factor) that cannot be determined experimentally during the measurement. An estimate of the correction and the associated uncertainty are instead obtained from prior knowledge generated, for example, by method evaluation or method verification studies [100, 117]. In some cases, where the measurement is carried out for different values of a quantity such as concentration, the estimated correction and the associated uncertainty may depend on the particular value of the quantity.

EXAMPLE *Recovery in chemical measurement*

In chemistry the component of interest (the analyte) is often embedded in a complex mixture of other molecules (the matrix). Many methods of analysis to determine the concentration (the measurand) of the analyte require the analyte to be separated from the matrix. The quantity actually measured is the concentration of the analyte following removal. A systematic error arises if the measured value is obtained using a measurement model that ignores the effect of incomplete removal [52]. Correction for incomplete removal is usually described by a multiplicative recovery factor, which is estimated in an experiment where a reference material, having a known concentration of the analyte in the matrix, is measured. The recovery factor is included in a measurement model for the concentration of the analyte, and account is taken of the uncertainty associated with the estimate of the factor, which includes the uncertainty associated with the value carried by the reference material.

10.4 Modelling poorly understood effects

10.4.1 Poorly understood effects, in contrast with the well-understood effects considered in 10.3, are known to exist, but little can be said in a specific measurement about their magnitude or even their sign. Although a correction cannot usefully be made for these effects, it is important to include them in the measurement model because they contribute to the uncertainty associated with the estimate of the measurand.

EXAMPLE Instances of poorly characterized effects include environmental conditions, effects between samples, mismatch between the calibration standards and the samples subject to measurement, treatment of test items prior to measurement, and the effects of sampling, sample transformation and sub-sampling.

10.4.2 Further instances of poorly understood effects include effects due to different operators, measurement conditions kept within certain limits, different artefacts and subsamples. These effects, which give rise to the dispersion of the values obtained for the measurand, should be included in the measurement model. It is often appropriate to evaluate under such conditions the standard uncertainty associated with the estimate of an input quantity, unless method validation, repeated observations on retained samples or artefacts, or quality control data show relevant between-run or between-measurement effects (see also JCGM 100:2008, annex H.5). In particular, long-term effects tend to be more significant and appropriate techniques, such as analysis of variance, should be used for evaluating such reproducibility effects. It is crucial to include random variation in a measurement model in order to obtain a realistic evaluation of the uncertainty arising from these effects. The topic of modelling random variation is considered in annex C.

10.4.3 The available knowledge about poorly understood effects is formally encapsulated in a random variable (see 5.5) having expectation zero or one for additive or multiplicative effects, respectively, and standard deviation providing a reasonable representation of the uncertainty associated with the effect. Such a random variable does not modify the estimate of the measurand, but contributes to the uncertainty associated with that estimate (also see 9.3 note 2).

NOTE Reproducibility studies can be a suitable way to characterise poorly understood effects if reproducibility can be established in related measurements. Guidance on the use of precision data is given in annex C.5.

EXAMPLE 1 *How long is a piece of string?* (continued from the example in 8.4.2)

Apart from Q_{11}, the quantities in the measurement model given there all constitute correction terms. All these correction terms apart from Q_{23} and Q_{24} are additive and would have an estimate of zero. Q_{23} would have an estimate equal to the (non-zero) expectation of the chi-squared distribution assigned to it, and Q_{24} would have an estimate of unity since it is the only multiplicative correction in the model.

EXAMPLE 2 *Release of cadmium from ceramic ware* (also see the example in annex F)

The amount of cadmium released from ceramic ware is typically determined using atomic absorption spectrometry. A measurement model for r, the mass of cadmium leached per unit area of ceramic ware, is [63, appendix A, example A.5]

$$r = \frac{\gamma_0 V_\mathrm{L}}{a_\mathrm{V}} d,$$

where γ_0 is the mass concentration of cadmium in the extraction solution, V_L is the volume of the leachate, a_V is the surface area of the vessel and d is the dilution factor. A more complete model is

$$r = \frac{\gamma_0 V_\mathrm{L}}{a_\mathrm{V}} d\, f_\mathrm{acid} f_\mathrm{time} f_\mathrm{temp}, \tag{23}$$

where f_acid, f_time and f_temp are multiplicative correction factors modelling the effects of acid concentration, duration of digestion and temperature. Estimates of the input quantities and the associated

relative standard uncertainties are given in table 4. The relative standard uncertainties associated with the estimates

$$\widehat{f}_{\text{acid}} = \widehat{f}_{\text{time}} = \widehat{f}_{\text{temp}} = 1$$

are provided by, for instance, method validation data [63, Example A.5]. Since no dilution was applied in the present example, d is exactly one. Using the relative standard uncertainties in table 4, given in [63, appendix A, example A.5] for this example, yields $u_{\text{rel}}(\widehat{r}) = 0.097$.

Table 4: Estimates and relative standard uncertainties for cadmium release measurement model

$\widehat{\gamma}_0$	$\widehat{V}_{\text{L}}$	$\widehat{a}_{\text{V}}$	$\widehat{d}$	$\widehat{f}_{\text{acid}}$	$\widehat{f}_{\text{time}}$	$\widehat{f}_{\text{temp}}$
$0.26\,\text{mg}\,\text{L}^{-1}$	$0.332\,\text{L}$	$5.73\,\text{dm}^2$	1	1	1	1
$u_{\text{rel}}(\widehat{\gamma}_0)$	$u_{\text{rel}}(\widehat{V}_{\text{L}})$	$u_{\text{rel}}(\widehat{a}_{\text{V}})$	$u_{\text{rel}}(\widehat{d})$	$u_{\text{rel}}(\widehat{f}_{\text{acid}})$	$u_{\text{rel}}(\widehat{f}_{\text{time}})$	$u_{\text{rel}}(\widehat{f}_{\text{temp}})$
0.069	0.005 4	0.033	0	0.000 8	0.001	0.06

10.5 Shared effects

10.5.1 In many cases, the same effect acts on two or more input quantities in a measurement model, thus introducing correlation. Such an effect is due, for instance, to the same measuring system, physical measurement standard or reference datum.

EXAMPLE *Shared temperature effect*

A thermometer is used to determine a temperature correction required in the estimation of input quantity X_i. The same thermometer is used to determine a similar temperature correction needed in the estimation of input quantity X_j. The quantities X_i and X_j (strictly, the associated random variables) may be significantly correlated.

10.5.2 The original set of input quantities $X_1, \ldots, X_N$ upon which the measurand Y depends can sometimes be redefined in such a way as to include as additional independent input quantities those quantities Q_ℓ that are common to two or more of the original X_i. In the case of the example in 10.5.1, the quantities that define the calibration function for the thermometer (used to determine the temperature corrections) may be included as additional independent input quantities. Then the measurement model can be expressed in terms of independent quantities. Nonetheless, in some situations it may be more convenient to retain covariances rather than to increase the number of input quantities. In many cases it is not possible to remove the correlation because of insufficient information. Such a situation is likely to occur whenever measurement standards are calibrated in a laboratory and used in another laboratory for subsequent calibrations (also see 8.4.3).

10.5.3 The correlation from shared effects has implications in the context of building an appropriate measurement model and impacts on the evaluation of the uncertainty associated with an estimate of the measurand. It is convenient to describe these common effects in the framework of multi-stage models (see 8.4.1 and 8.4.3). In this context, two input quantities X_i and X_j depend on quantities $Q_1, \ldots, Q_L$ (the common systematic effects). Thus, $X_i = F_i(Q_1, \ldots, Q_L)$ and $X_j = F_j(Q_1, \ldots, Q_L)$, although some of $Q_1, \ldots, Q_L$ may actually appear only in one function and not in the other.

EXAMPLE *Calibrated resistors in series*

Ten resistors, each of nominal resistance $R_i = 1\,\mathrm{k\Omega}$, are calibrated with a negligible uncertainty of comparison in terms of the same $1\,\mathrm{k\Omega}$ standard resistor R_S characterized by a standard uncertainty $u(R_S) = 100\,\mathrm{m\Omega}$ as given in its calibration certificate. The resistors are connected in series with wires having negligible resistance in order to obtain a reference resistance R_{ref} of nominal value $10\,\mathrm{k\Omega}$. Thus, the measurement model is

$$R_{\mathrm{ref}} = \sum_{i=1}^{10} R_i. \tag{24}$$

Let the calibration of resistor i be represented by

$$R_i = \alpha_i R_S, \tag{25}$$

with α_i a resistance ratio and $u(\alpha_i)$ the associated standard uncertainty as obtained from repeated observations. It follows from the law of propagation of uncertainty [JCGM 100:2008 5.2] that the variance associated with R_i and the covariance associated with R_i and R_j are

$$u^2(R_i) = R_S^2 u^2(\alpha_i) + \alpha_i^2 u^2(R_S), \qquad u(R_i, R_j) = \alpha_i \alpha_j u^2(R_S), \tag{26}$$

respectively. When, as occurs in practice, the α_i take values very close to a common value α with very small standard uncertainties, expressions (26) reduce to

$$u^2(R_i) = u(R_i, R_j) \approx \alpha^2 u^2(R_S),$$

for all i and j, from which the correlation coefficient

$$\frac{u(R_i, R_j)}{u(R_i)u(R_j)} \approx 1.$$

The formula in JCGM 100:2008 5.2.2 note 1 then yields for the standard uncertainty associated with R_{ref},

$$u(R_{\mathrm{ref}}) = \sum_{i=1}^{10} u(R_s) = 10 \times (100\,\mathrm{m\Omega}) = 1\,\Omega.$$

Taking no account of correlation, namely, applying the law of propagation of uncertainty for independent quantities, JCGM 100:2008, equation (10), yields

$$u(R_{\mathrm{ref}}) = \left[\sum_{i=1}^{10} u^2(R_s)\right]^{1/2} = 0.32\,\Omega,$$

which is incorrect, being too small by a factor of 3.

Since the dominant uncertainty is that associated with R_S, which is common to all the resistor measurements, the errors in determining the different R_i are almost perfectly correlated and the final resistance R_{ref} is effectively a known multiple ($\sum_i \alpha_i \approx 10$) of R_S with a corresponding 10-times larger uncertainty. Neglecting correlations leads to the incorrect result $\sqrt{10}u(R_i)$ because it assumes that the errors associated with individual resistance measurements vary independently.

The expressions (24) and (25) constitute a simple instance of a two-stage model, the first stage in which R_S and the α_i are input quantities and the R_i are output quantities, and the second stage in which the R_i are the input quantities and R_{ref} is the output quantity. Substitution of expression (25) into (24) gives a single-stage model.

10.6 Drift and other time-dependent effects

10.6.1 Many measurement standards show drift over time. A measurement standard used as a reference is typically not calibrated immediately before use. Whenever the standard is used, it is necessary to update its value and standard uncertainty reflecting the possible change in the measurement standard since it was last calibrated. A recorded history of calibration results provides information about changes in the standard up to its most recent calibration. By fitting a suitable model to the recorded values of the standard as a function of time, the value of the measurement standard and the associated standard uncertainty at the time of use can be inferred.

10.6.2 By plotting the time history of the quantity of interest, it can be seen whether the standard is fluctuating at random, or if there is a trend whether it appears to be linear or not. For a well-maintained measurement standard, a linear drift model such as that described in the example in 10.6.4 is often appropriate.

10.6.3 Any explanatory model should include terms to represent the time effect. The terms can describe corrections to the observations so that they relate to a single quantity measured at a convenient reference time. Alternatively, the terms can constitute parameters in a model of the observations as a function of time (as in 10.6.1). The corrections (in the former case) and the function parameters (in the latter case) are characterized by estimates and associated uncertainties and covariances (or, more generally, by a joint probability distribution).

10.6.4 A suitable model will enable, under appropriate assumptions, a value of the measurement standard to be estimated at the time of use and the associated uncertainty evaluated. Care should always be taken since such estimation is a prediction or extrapolation beyond the span of existing data. See 13.2. See also reference [54], which provides additional detail, particularly including approaches for dealing with the effects of such time-dependent changes as a part of traceability statements.

NOTE Current guidance on the production and certification of reference materials [89,97] requires that the uncertainty associated with the property value includes long-term effects.

EXAMPLE *Drifting resistance standard*

A national metrology institute (NMI) provides calibration services for resistance standards based on a secondary working standard that is calibrated using a primary 200 Ω standard approximately every six months. The 200 Ω standard exhibits drift due to recognized physical effects. Although no theoretical model of such effects is used, experience of a wide selection of high-quality standards suggests the drift can be explained by an empirical straight-line model. The value R of a standard resistor is most conveniently expressed as a relative deviation $y = \Delta R/R_{\text{nom}} = (R - R_{\text{nom}})/R_{\text{nom}}$ from its nominal value R_{nom}.

The observed relative deviations for a 200 Ω secondary resistance standard over the last 20 calibrations with time values t_i and relative deviations y_i is shown in figure 5. The calibration points (t_i, y_i) are given as crosses in the figure and provided in table 5. The constant standard uncertainty 0.025×10^{-6} was assigned to all y_i based on knowledge of the measurement. The t_i were taken as exact.

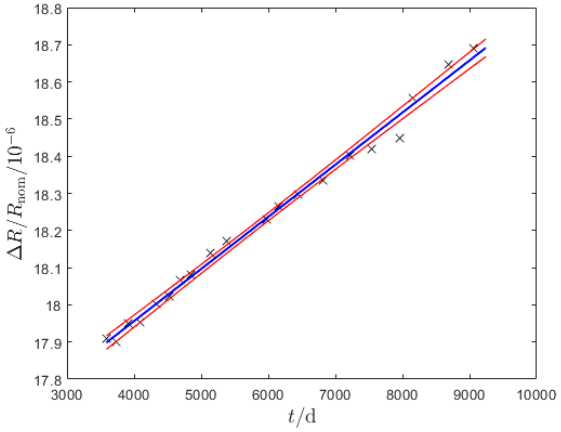

Figure 5: Relative deviation from nominal of a resistance standard measured at approximately 6-monthly intervals, fitted drift line and 95 % coverage corridor

Table 5: Resistance drift data

t/d	3 588	3 731	3 916	4 089	4 325
$y/10^{-6}$	17.909 3	17.900 8	17.949 3	17.953 9	18.003 2
t/d	4 520	4 686	4 843	5 143	5 381
$y/10^{-6}$	18.022 0	18.065 9	18.082 1	18.138 4	18.171 9
t/d	5 982	6 157	6 445	6 817	7 219
$y/10^{-6}$	18.229 4	18.265 5	18.295 9	18.334 8	18.400 9
t/d	7 533	7 966	8 149	8 683	9 059
$y/10^{-6}$	18.420 5	18.449 8	18.558 0	18.647 0	18.691 6

The central blue line in figure 5 shows the time-dependent estimate of the relative deviation obtained by fitting a straight line to the calibration points using least-squares regression respecting the given uncertainties. The lower and upper red lines indicate a 95 % coverage corridor for y, as functions of time. The evaluated standard deviation of the dispersion of the data about the straight line is 0.023×10^{-6}, which compares well with the standard uncertainty of 0.025×10^{-6} associated with the y_i.

A customer calibration against the NMI's working standard can take place at any time t_{C}, say, typically within a 6 month period, since the most recent re-calibration. The value of the NMI's standard at that time is based on the value of the straight line at $t = t_{\mathrm{C}}$.

For the value $t_{\mathrm{C}} = 9\,242\,\mathrm{d}$, approximately 6 months after the most recent t-value in the table, the extrapolated value using the fitted straight line is $\widehat{y}_{\mathrm{C}} = 18.691\,7 \times 10^{-6}$. (Extrapolation can in general be dangerous: see 13.2.) The standard uncertainty associated with $\widehat{y}_{\mathrm{C}}$, obtained by propagating the standard uncertainties and covariance associated with the regression coefficients, is $u(\widehat{y}_{\mathrm{C}}) = 0.012\,3 \times 10^{-6}$. The expanded uncertainty for 95 % coverage is $0.025\,8 \times 10^{-6}$ based on the assumption that the measurement errors in the y_i can be regarded as draws from a normal distribution.

11 Statistical models used in metrology

11.1 General

11.1.1 Statistical models use probability distributions to reconcile the variability of replicated observations of the same property that depends on the measurand, or the dispersion of multiple measured values obtained for the same measurand, with the unknown, single true value of the measurand [44, 127].

11.1.2 Models of this kind underlie all Type A evaluations of uncertainty, even if they are not articulated explicitly: in table H.2 of JCGM 100:2008, five replicated readings of electric current are combined into their arithmetic average, and the experimental standard deviation of this average is computed according to equation (5) in the GUM. The choice of the average as presumably 'best' (in the sense of minimum mean squared error) to summarize those observations suggests a measurement model where the observations are equal to the true value of the current plus normally distributed errors, all of the same precision. Reliance on equation (5) suggests that these observations are believed to be uncorrelated.

EXAMPLE *Force transducer*

When calibrating a force transducer [12, 13], and essentially the same force is applied repeatedly to it, the corresponding values of the deflection caused by the force typically will differ from one instance of application to another. This measurement may be modelled by saying that each observation of the deflection D is a value of a normal random variable whose mean is a linear function of the force F, with an unknown standard deviation σ that characterizes the dispersion of the values of the deflection obtained under conditions of repeatability. In this case the measurement model may be written as

$$D = \alpha + \beta F + \varepsilon,$$

where α and β denote coefficients to be estimated using least squares, for example, from the data gathered in the calibration experiment, and ε is assumed to be a normal random variable with mean 0 N and unknown standard deviation σ.

11.1.3 Statistical models are either mathematical or algorithmic descriptions of the relation between empirical observations and the value of some property that is of substantive interest. The characteristic trait of statistical models, which distinguish them from other kinds of mathematical models, is that probability distributions are used to accommodate the fact that empirical observations made repeatedly of the same property typically vary, even when the value of this property remains constant.

11.1.4 Statistical models employ probability distributions to describe sampling variability, or uncertainty more generally, which render details of the observations (empirical data) unpredictable. Such uncertainty clouds the relationship between true values of observable properties and true values of properties of interest that are not accessible to direct observation, whose values need to be inferred from the experimental data.

EXAMPLE *Newtonian constant of gravitation*

The Newtonian constant of gravitation G has been measured in many experiments implementing a wide variety of approaches that rely on different physical laws describing how observable phenomena are informative about its value. From early measurements of the density of the Earth, in particular by Henry Cavendish [27], to interferometry using cold atoms [147], and including a great variety of experiments involving balances of different kinds, a collection of measurement results has been assembled in the course of more than 200 years, yet without succeeding in achieving relative measurement uncertainty comparable to those that are associated with most other fundamental constants, even though sophisticated statistical models and methods of data reduction have been employed to derive a consensus value from the accumulated data [122].

11.1.5 Many different types of statistical models may be used as measurement models. The simplest, and most commonly used such model for scalar observations made under conditions of repeatability is illustrated in example 1 in 11.3. Subclauses 11.4 to 11.8 review several models that are particularly useful in measurement science.

11.2 Observation equations

Statistical models link measurement inputs (observations) to the true value of the measurand probabilistically. For example, a so-called *observation equation* [69, 142] expresses the observations as realized values of random variables whose probability distribution depends on the true value of the measurand: more precisely, where such true value is a known function of one or several parameters that specify the probability distribution of the observations. The expression 'observation equation', or 'observational equation', has been in use in metrology for a very long time [22, 60, 136, 155], even though the precise and specific meaning just described is fairly recent.

NOTE 1 The inputs for a measurement model (statistical or of any other kind) are observed values $x_1, \ldots, x_N$ of properties (which may be quantitative or qualitative) that are informative about the measurand in the sense that, once suitably combined, possibly also taking into account information about the measurand originating from other sources, produce an estimate of the true value of the measurand.

NOTE 2 For statistical models, the observations x_i may be considered as being made under conditions of repeatability (see JCGM 200:2012, 2.20) or under other conditions (for example, reproducibility conditions (see JCGM 200:2012, 2.24), depending on modelling choice and circumstances.

EXAMPLE *Newtonian constant of gravitation (continued from 11.1.4)*

A statistical model for N independent measurement results obtained for G expresses them as outcomes of observable random variables $G_i = G + \lambda_i + E_i$, for $i = 1, \ldots, N$, where λ_i denotes an effect specific to experiment i, and E_i denotes measurement error. The specification of the model includes a characterization of the probability distributions of the experiment effects and of the measurement errors: for example, that $\lambda_1, \ldots, \lambda_N$ are like a sample from a normal distribution with mean $0\,\mathrm{m}^3\,\mathrm{kg}^{-1}\,\mathrm{s}^{-2}$ and unknown standard deviation τ, and that the measurement errors are like outcomes of independent random variables with distributions centered at $0\,\mathrm{m}^3\,\mathrm{kg}^{-1}\,\mathrm{s}^{-2}$ but with possibly different unknown standard deviations.

11.3 Specification of statistical models

11.3.1 Statistical models can often be specified by providing (i) as many simultaneous equations as there are observations made in the course of measurement, whose left sides have random variables representing the observations, and whose right sides have functions of several random variables whose realized values determine the observations, (ii) a description of the joint probability distribution of the random variables that appear on the right side of these equations, and (iii) an equation or algorithm that describes the relation between the parameters of this joint distribution and the true value of the measurand. A simple instance is provided as example 1 in 11.3.3.

11.3.2 Since the specification of a statistical model includes a statement about the probability distribution(s) assumed, it is important to recognize that the results obtained will generally be different under different assumptions. No general guidance can be offered since knowledge of the application almost wholly dictates the assumptions to be made.

11.3.3 When describing statistical models, particular observed values are usually distinguished from the corresponding random variables by using lower case letters for the former and upper case letters for the latter. However, the same symbols are very often used to denote either, their particular meaning being clear from the context. This simplification is motivated by the informal understanding of a random variable as a property value that is clouded by uncertainty, and that has a probability distribution as an attribute that characterizes and quantifies this uncertainty precisely.

EXAMPLE 1 *Electric current — model*

Table H.2 in JCGM 100:2008 lists the following values of electric current determined under conditions of repeatability:

$$19.663\,\text{mA}, \quad 19.639\,\text{mA}, \quad 19.640\,\text{mA}, \quad 19.685\,\text{mA}, \quad 19.678\,\text{mA}.$$

Assuming that the current I_0 (the measurand) remained constant as these observations were made, a particularly simple statistical model that explains how they may be consistent with the true value of the measurand expresses each observed value as

$$I_j = I_0 + E_j, \qquad j = 1, \ldots, 5,$$

where $E_1, \ldots, E_5$ denote (non-observable) measurement errors assumed to be independent, normal random variables with mean 0 mA and unknown standard deviation σ. In this case, the statistical model asserts that $I_1, \ldots, I_5$ are independent, normal random variables whose common mean is the measurand I_0, all with the same standard deviation σ. As is often the case with statistical models, the measurand is a parameter (the mean in this case) of the probability distribution of the observations.

EXAMPLE 2 *Lifetime of ball bearings — model*

Table B-5 of [114] lists the following results of an experiment to measure the lifetime of a batch of ball bearings, expressed as millions of revolutions at the time of failure: 6.0, 8.6, 17.8, 18.0, 27.5, 33.5, 50.5, 51.5, 69.0, 74.0, 74.0, 89.0, 109.0, 118.0, 119.0, 138.0, 141.0, 144.0, 146.0, 150+, 151, 153+, 153+, 153+. Four of these (marked with a plus sign) represent censored observations. For instance, '150+' means that the corresponding ball bearing had not yet failed when the experiment was terminated and the ball had already undergone 150 million revolutions. These numbers of revolutions are assumed to have been observed under conditions of repeatability, and, in the

original study [114] they are modelled as a sample from a Weibull probability distribution with unknown values of its scale S and shape A. The measurand is the mean (or expected) lifetime,

$$T = S\,\Gamma(1 + 1/A),$$

where Γ denotes the gamma function of mathematical analysis [8]. In this case, the measurand is a function of the parameters (S and A) of the probability distribution of the observations. Considering the fact that if R is a random variable with a Weibull distribution with scale S and shape A, then $(R/S)^A$ has an exponential distribution with mean 1, the statistical model may also be written as an explicit observation equation that represents each observed number of revolutions as outcome of the random variable R_i such that

$$\ln R_i = \ln S + (1/A)\ln E_i, \qquad i = 1, \ldots, 24,$$

where the E_i denote non-observable measurement errors modelled as independent random variables whose common probability distribution is exponential with mean 1.

EXAMPLE 3 *Mass fraction of tin — model*

The value assigned to the mass fraction of tin in a reference solution used for spectrometric calibrations is a consensus value resulting from combining measurement results obtained gravimetrically and via inductively coupled plasma optical emission spectroscopy (ICP-OES) [138]. For one particular version of this reference solution, the measured values were $w_G = 10.000\,07\,\mathrm{mg\,g^{-1}}$ from gravimetry and $w_I = 10.022\,39\,\mathrm{mg\,g^{-1}}$ from ICP-OES. The corresponding associated standard uncertainties were $u(w_G) = 0.000\,31\,\mathrm{mg\,g^{-1}}$ based on $\nu_G = 278$ degrees of freedom, and $u(w_I) = 0.002\,64\,\mathrm{mg\,g^{-1}}$ based on $\nu_I = 11$ degrees of freedom. The measurand is ω in the *random effects model* (see 11.5.3) that comprises two simultaneous observation equations,

$$w_G = \omega + \lambda_G + \varepsilon_G, \qquad w_I = \omega + \lambda_I + \varepsilon_I,$$

where λ_G and λ_I denote method effects, and ε_G and ε_I denote method-specific measurement errors. The specification of the model is completed by assuming that the method effects are independent normal random variables with mean $0\,\mathrm{mg\,g^{-1}}$ and unknown standard deviation τ, and that the method-specific measurement errors are independent random variables with scaled Student's t distributions centered at $0\,\mathrm{mg\,g^{-1}}$, with ν_G and ν_I numbers of degrees of freedom, and unknown standard deviations $u(w_G)$ and $u(w_I)$.

11.4 Models for calibration and analysis

11.4.1 The calibration of a measuring instrument serves to characterize how the instrument responds when presented with a measurement standard. This characterization involves establishing a relationship (*calibration function*) between values of calibration standards presented to the instrument, and the indications that the instrument produces in response (that is, the calibration function maps values of standards to instrumental indications), and evaluating the uncertainty associated with the resulting calibration function.

11.4.2 To use the instrument in practice the inverse relation is needed, which maps an instrumental indication into a value of the measurand: for example, the inverse of the calibration function is called *analysis function* in gas metrology [73], and *measurement function* in force metrology [13]. For example, the calibration of an electrical load cell involves applying specified forces to it, reading the corresponding values of electromotive force (voltages) produced by the cell, and then building a function, most commonly a polynomial of low degree, that maps values of force to voltages or ratios of voltages [13].

11.4.3 Since both the values of the measurement standards used in calibrations, and the corresponding instrumental indications, typically have non-negligible associated uncertainties, the statistical model usually most appropriate for calibrations is *errors-in-variables regression* [70], which may be employed to produce an analysis function, in the sense of ISO 6143 [94], directly, not by inversion of a calibration function. When the uncertainties associated with the values of the standards are negligible by comparison with the uncertainties associated with the instrumental indications, ordinary regression (where the fitted calibration curve minimizes a measure of the distances between observed and predicted values of the instrumental indications) may suffice. In some cases, the analysis (or measurement) function is determined directly during calibration, not via mathematical inversion of the corresponding calibration function.

EXAMPLE *Amount fraction of* N_2O *in southern oceanic air — model*

Figure 6 shows the data listed in table 6 for the primary standard mixtures that were used to build an analysis function A to assign values to the amount fraction of N_2O in a reference material comprising southern oceanic air collected at Baring Head, New Zealand [145]. The analysis function A produces a value of the amount fraction of N_2O, $x(N_2O) = A(r)$, that corresponds to a ratio r of the instrumental indications obtained contemporaneously for a calibration standard (or for a gas sample whose amount fraction of N_2O is to be determined) and for the air in a particular cylinder designated as the lot standard. The statistical model comprises three relations for each standard:

$$x = \xi + \varepsilon, \qquad r = \rho + \delta, \qquad \xi = A(\rho).$$

The first two indicate that the reported amount fraction in the standard x, and the corresponding instrumental reading r, are both affected by measurement errors, with ξ and ρ denoting the corresponding true values. The third states that the analysis function A relates corresponding true values. In gas metrology, A usually is a polynomial of low degree that is fitted to the calibration data using errors-in-variables regression because the values of both x and r have non-negligible associated uncertainties, whose relative sizes are comparable in this case (last two columns of table 6). Example 4 in 11.9.8 describes the estimation and uncertainty evaluation for A.

Table 6: Analysis of the primary standard mixtures used to assign values to the amount fraction x of N_2O in a southern oceanic air reference material [145]

Standard mixture	x/ (nmol/mol)	$u(x)$/ (nmol/mol)	r	$u(r)$	$(u(x)/x)$/ %	$(u(r)/r)$/ %
FF22270	314.952	0.047	0.972 355	0.000 099	0.015	0.010
FF22225	319.646	0.048	0.987 466	0.000 155	0.015	0.016
CAL016773	326.325	0.057	1.006 460	0.000 078	0.018	0.007 8
FF22145	331.135	0.057	1.021 310	0.000 132	0.017	0.013
FF22181	338.390	0.058	1.043 740	0.000 131	0.017	0.013
FF22146	344.378	0.060	1.062 970	0.000 071	0.017	0.006 7

11.5 Models for homogeneity studies

11.5.1 The preparation of reference materials in batches intended to be subdivided into separately packaged items (or units) that are assigned the same value and associated uncertainty, typically involves a study of the homogeneity of the batch, based on replicated measurements made in different units of the material.

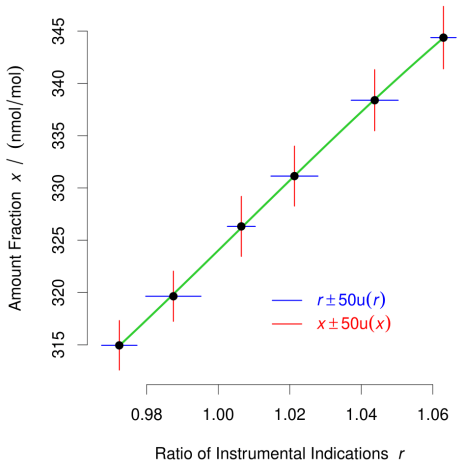

Figure 6: Analysis function used to assign values to the amount fraction of N_2O in a southern oceanic air reference material. The coordinates of the black dots are ratios and amount fractions for the primary calibration standards

11.5.2 ISO Guide 35 [97] provides guidance for characterizing the homogeneity of reference materials. This Guide recognizes that homogeneity studies traditionally have included a statistical test to assess whether between-unit differences are significantly larger than within-unit differences, or larger than the uncertainty that the producer of the reference material achieved when measuring a single unit. ISO Guide 35 chooses to focus instead on whether the between-unit standard deviation is sufficiently small for the purpose that the candidate reference material is intended to serve, and in its clause 7.10 suggests four criteria for checking whether this is the case: criteria (a)-(c) are accepted, practical conventions that ignore the sampling uncertainty surrounding the estimate of the between-unit standard deviation; criterion (d) recommends that a statistical test be carried out to ascertain 'that the between-unit term is not statistically significant at the 95 % level of confidence', suggesting the Snedecor-Fisher F test as one possibility.

11.5.3 The Snedecor-Fisher F test is usually carried out using the output from a conventional analysis of variance (ANOVA). The statistical model underlying such ANOVA expresses each measured value w_{ij} as a sum of the value ω of the measurand, an item effect λ_j, and a measurement error ε_{ij}, for replicate determination i in item j: $w_{ij} = \omega + \lambda_j + \varepsilon_{ij}$. The item effects λ_j are modelled as a sample from a normal probability distribution with mean 0 and unknown standard deviation τ, and the ε_{ij} are modelled as a sample from a normal probability distribution with mean 0 and unknown standard deviation σ. Because the item effects are modelled as random variables, this model is a *random effects model*. In this setting, the Snedecor-Fisher F test is equivalent to testing the hypothesis that $\tau = 0$ (with the appropriate units). In practice, and for many reference materials, a modicum of heterogeneity is acceptable, and it is recognized and expressed in the uncertainty associated with the estimate of ω. When a candidate reference material is found to be excessively heterogeneous, in some cases it may be possible to split it into portions that are sufficiently homogeneous: this is often the case when the reference material is a gas mixture, and the items are cylinders that have been filled by drawing from several batches of the mixture, nominally of the same composition, yet whose actual compositions are too different for the intended purpose.

11.5.4 The Snedecor-Fisher F test is very sensitive to violations of the assumptions just stated, that validate the underlying statistical model. Therefore, the validity of these assumptions needs to be checked, at a minimum by examining QQ-plots [28] of the estimates of the item effects, and of the measurement errors. Neither the Kruskal-Wallis test [78, Section 6.1] (implemented in function `kruskal.test`, which is included in 'base' R, available at `https://www.r-project.org/`) nor the Fisher-Pitman permutation test [14] (implemented in function `oneway_test` of R package `coin`, available at `https://www.r-project.org/`) require the assumption of normality, and both can be used to assess homogeneity because they test whether the true values of the measurand in different items are significantly different, when determinations of the measurand can be made in multiple aliquots drawn from each item.

EXAMPLE *Barium in clay soil — model*

Table 7 lists measured values of the mass fraction of barium in aliquots of a clay soil used in a homogeneity study [159]. The model for this data is the additive model described in 11.5.3. The results of fitting this model to the data in the table are presented in Example 5 under 11.9.8. Here the focus is on statistical tests of the hypothesis of homogeneity. The p-values from the conventional ANOVA F-test, the Kruskal-Wallis test and the Fisher-Pitman permutation test above are 0.09, 0.15, and 0.09, respectively. A p-value is the probability of observing a value of the test criterion at least as extreme as was observed, under the assumption of homogeneity: since none of these p-values is alarmingly small, none of the tests rejects the hypothesis of homogeneity.

Table 7: Values of the mass fraction ($\mathrm{mg\,kg^{-1}}$) of barium in a clay soil measured for a homogeneity study — reproduced from Table 1 of [159]

Sample	Replicates			Sample	Replicates		
0118	323	301	310	1133	310	328	312
0201	340	334	316	1249	314	314	302
0383	320	321	309	1464	329	300	299
0442	315	338	321	1581	320	329	311
0557	326	338	325	1607	322	312	311
0666	325	302	304	1799	332	317	299
0791	324	331	317	1877	313	294	293
0918	310	310	331	1996	324	314	335
1026	336	321	328	2000	321	342	316

11.6 Models for the adjustment of observations

11.6.1 In geodetic and astronomical measurement, as well as in conventional dimensional metrology and in the calibration of weights, the *adjustment of observations* consists of applying a measurement model and data reduction procedure to produce mutually consistent estimates of the true values of the angles or distances, or of the masses, that are the measurands. Applications include 'aligning' point clouds measured by different coordinate measuring machines and expressed relative to different reference frames;, assigning values to distances between landmarks based on redundant, mutually inconsistent triangulations in geodetic surveys, civil engineering, and industrial assembly of large structures (for example, in aeroplane manufacturing); and calibrating weights. In all cases, it is required to qualify the results with evaluations of uncertainty.

11.6.2 The model that is most commonly used for the adjustment of observations in this sense expresses each observation as the sum of an unknown true value and a non-observable error (which may be a scalar or a vector) with mean zero, where the true values are assumed to satisfy particular 'equations of condition'. The adjustment is typically carried out by application of the method of least squares, which is equivalent to the method of maximum likelihood when the errors are assumed to be outcomes of independent, normally distributed random variables.

EXAMPLE 1 *Calibrating a geodetic base line — model*

Figure 7 shows the segments with lengths $L_1, \ldots, L_6$ that were surveyed to obtain an estimate of the length of the straight base line between geodetic markers A and D, computed as the sum of the distances from A to B (λ_1), from B to C (λ_2), and from C to D (λ_3). The observations were $L_1 = 27.19\,\text{m}$, $L_2 = 33.08\,\text{m}$, $L_3 = 25.33\,\text{m}$, $L_4 = 60.33\,\text{m}$, $L_5 = 58.35\,\text{m}$, and $L_6 = 85.59\,\text{m}$. The observed lengths are mutually inconsistent because, for example, $L_2 + L_3 \neq L_5$.

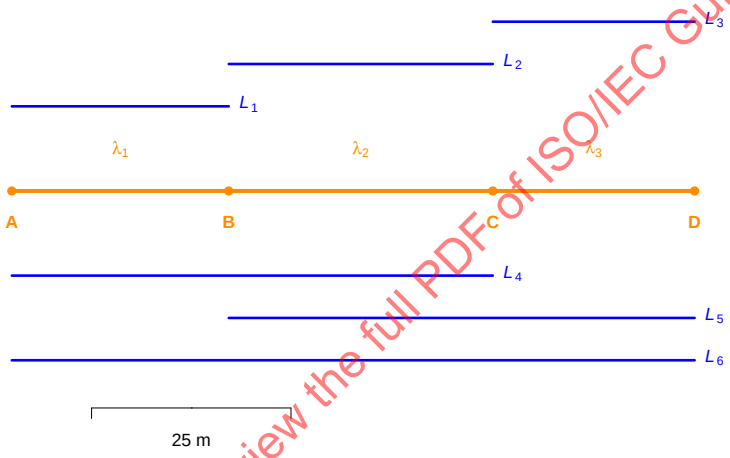

Figure 7: Observations made to calibrate the length of base line A-D

The models for the observations are $L_1 = \lambda_1 + E_1$, $L_2 = \lambda_2 + E_2$, $L_3 = \lambda_3 + E_3$, $L_4 = \lambda_1 + \lambda_2 + E_4$, $L_5 = \lambda_2 + \lambda_3 + E_5$, and $L_6 = \lambda_1 + \lambda_2 + \lambda_3 + E_6$, where $E_1, \ldots, E_6$ denote measurement errors. In geodesy, the additional assumption is commonly made that the measurement errors are like a sample from a normal distribution with mean 0 m and unknown standard deviation σ.

EXAMPLE 2 *Standard relative atomic mass of tellurium*

Before mass spectrometric measurements were common, relative atomic masses (also called atomic weights) of the elements were determined from sets of measurements linking the chemical equivalents of various elements. For instance, the relevant data involving tellurium are as follows [45]:

$$\begin{aligned}
A_\text{r}(\text{TeBr}_4)/(4A_\text{r}(\text{Ag})) &= 1.036\,49, & u(A_\text{r}(\text{TeBr}_4)/(4A_\text{r}(\text{Ag}))) &= 0.000\,04,\\
A_\text{r}(\text{TeBr}_4)/(4A_\text{r}(\text{AgBr})) &= 0.595\,426, & u(A_\text{r}(\text{TeBr}_4)/(4A_\text{r}(\text{AgBr}))) &= 0.000\,025,\\
A_\text{r}(\text{TeCl}_4)/(4A_\text{r}(\text{Ag})) &= 0.624\,427, & u(A_\text{r}(\text{TeCl}_4)/(4A_\text{r}(\text{Ag}))) &= 0.000\,034,\\
A_\text{r}(\text{TeCl}_4)/(4A_\text{r}(\text{AgCl})) &= 0.469\,962, & u(A_\text{r}(\text{TeCl}_4)/(4A_\text{r}(\text{AgCl}))) &= 0.000\,027,\\
A_\text{r}(\text{Ag}_2\text{Te})/(4A_\text{r}(\text{Ag})) &= 1.591\,44, & u(A_\text{r}(\text{Ag}_2\text{Te})/(4A_\text{r}(\text{Ag}))) &= 0.000\,03.
\end{aligned}$$

Treating the relative atomic mass of silver, $A_\text{r}(\text{Ag})$, as an additional input variable these observations provide a set of equations whose least-squares solution estimates the relative atomic mass of tellurium (along with bromine and chlorine).

11.7 Models for time series

11.7.1 Observations made repeatedly of the same phenomenon or object, over time or along a transect in space, often equispaced in time or equidistant in space, tend to be interrelated, the more so the greater their proximity. The models used to study such series of observations involve correlated random variables [23, 134].

11.7.2 A particularly rich and widely used class of models for series of observations $x_{t_1}, x_{t_2}, \ldots$ made at equispaced epochs $t_1, t_2, \ldots$ are the so-called ARIMA (auto-regressive, integrated, moving average) models [23]. These models are applicable to series whose level and pattern of oscillations remain constant over time, as well as to series whose level (for example, the mean) drifts with passing time. For the latter, the model focuses on differences, for example on first differences $x_{t_i} - x_{t_{i-1}}$.

11.7.3 ARIMA models express the value observed at a particular epoch (possibly after differencing the series), as a linear combination of values observed at a finite number of previous epochs, plus an 'error' that itself is a linear combination of measurement errors affecting the current observation and a finite number of previous observations. In the following example, an ARIMA model is used for a time series of values of temperature of a thermal bath measured at regular intervals. It captures the fact that values measured at nearby epochs tend to be more similar than values at more widely separated epochs. Recognizing the structure of the resulting correlations, and modelling them accurately, may impact on uncertainty evaluations considerably, as illustrated in example 9 in 11.9.8.

EXAMPLE *Temperature of a thermal bath — model*

The readings of temperature listed in Table 8 and depicted in Figure 8 were taken every minute with a thermocouple immersed in a thermal bath during a period of 100 min. These observations serve (i) to estimate the temperature of the bath assuming that it is in thermal equilibrium, and (ii) to ascertain that the bath indeed is in thermal equilibrium. Owing to uncontrolled exchanges of thermal energy with the environment external to the bath, which are counterbalanced by controlled heating and circulation of the fluid, even if the temperature of interest does not drift, it will still oscillate over time around its mean value τ, which is the primary measurand. The secondary measurand is the pattern of such oscillations. The inputs are readings of temperature $t_1, \ldots, t_m$, where t_i denotes the reading made i minutes after the beginning of the experiment, for $i = 1, \ldots, m = 100$. A time series model for the t_i describes how consecutive readings of temperature are related to one another and to τ. The structure of this model, and estimates of its adjustable parameters, characterize the pattern of oscillations observed when the bath is in a steady thermal state. The following auto-regression of order 2 [23] has been found to be adequate for these data [141, V.B.1]:

$$t_i = \tau + \varphi_1(t_{i-1} - \tau) + \varphi_2(t_{i-2} - \tau) + \varepsilon_i,$$

where the ε_i are assumed to be uncorrelated normal random variables with mean 0 °C and unknown standard deviation σ. The specification of this model comprises four parameters, τ, φ_1, φ_2 and σ, of which τ is the primary measurand, and the other three characterize the state of thermal equilibrium.

This example has also been analyzed using a slightly different model in reference [139].

Table 8: Time series of temperature readings made every minute, expressed as deviations from 50 °C. The first reading was 50.102 4 °C, the tenth 50.107 2 °C, the eleventh 50.105 4 °C, etc. [48]

0.1024	0.1054	0.1026	0.1042	0.1026	0.1039	0.1065	0.1052	0.1067	0.1072
0.1054	0.1049	0.1082	0.1039	0.1052	0.1085	0.1088	0.1075	0.1085	0.1098
0.1070	0.1060	0.1067	0.1065	0.1072	0.1062	0.1085	0.1062	0.1034	0.1049
0.1044	0.1057	0.1060	0.1082	0.1052	0.1060	0.1057	0.1072	0.1072	0.1077
0.1103	0.1090	0.1077	0.1082	0.1067	0.1098	0.1057	0.1060	0.1019	0.1021
0.0993	0.1014	0.0965	0.1014	0.0996	0.0993	0.1003	0.1006	0.1026	0.1014
0.1039	0.1044	0.1024	0.1037	0.1060	0.1024	0.1039	0.1070	0.1054	0.1065
0.1072	0.1065	0.1085	0.1080	0.1093	0.1090	0.1128	0.1080	0.1108	0.1085
0.1080	0.1100	0.1065	0.1062	0.1057	0.1052	0.1057	0.1034	0.1037	0.1009
0.1009	0.1044	0.1021	0.1021	0.1029	0.1037	0.1049	0.1082	0.1044	0.1067

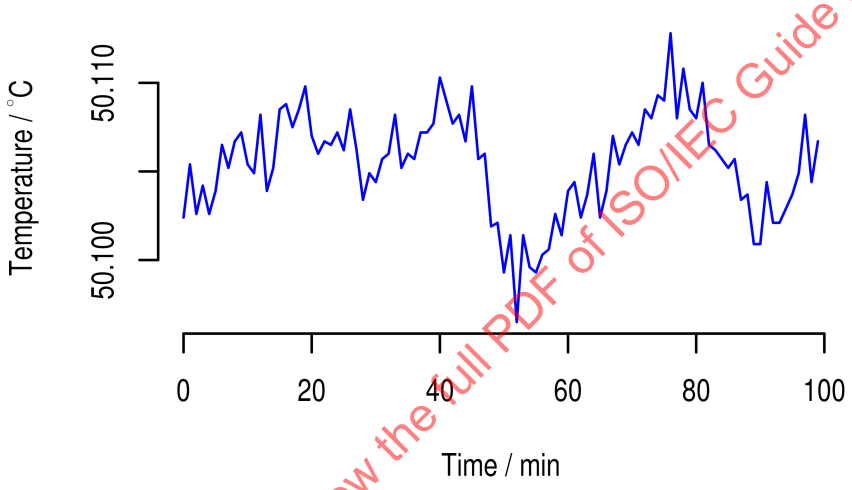

Figure 8: Time series of the temperature readings listed in table 8

11.7.4 In some applications, where an observed time series is used as input for subsequent data reduction steps, an *ad hoc* smoothing procedure, as an alternative to fitting an explicit model, suffices to abate the extent of the uncertainty that is propagated downstream. Moving averages (in particular, exponentially weighted moving averages) and moving medians are widely used for this purpose [126, §6.4.3].

11.8 Bayesian statistical models

11.8.1 Bayesian statistical models reflect an understanding of uncertainties associated with both inputs and outputs as characterizing states of incomplete knowledge about the true value of the input quantities and of the measurand. They are particularly useful when there exists information about the true value of the measurand prior to obtaining the results of a new measurement, by providing the means to update such information with the fresh data that will be acquired in the new measurement.

11.8.2 The distinctive features of a Bayesian treatment are these: (i) experimental data are modelled as realized values, or outcomes, of random variables with probability dis-

tributions involving parameters whose values are either unknown or known only incompletely; (ii) these parameters are modelled as random variables, hence so is the measurand because it is a function of those parameters — however, none of these random variables is accessible to direct observation: knowledge of their values is obtained only through the empirical data; (iii) estimates and uncertainty evaluations for the parameters are derived from the conditional probability distribution of the parameters given the data (the so-called *posterior distribution*).

11.8.3 Accepting (i) and (ii) involve specifying probability distributions for all quantities involved (data and unknown parameters), and (iii) involves the application of Bayes' rule [46, section 2.3, theorem 7.2.1], typically using Markov Chain Monte Carlo (MCMC) sampling to produce an arbitrarily large sample from the posterior distribution [71], which may then be reduced to produce an estimate of the measurand and the associated uncertainty.

11.8.4 The distribution assigned to the measurand (and to any other quantities whose values are unknown) is called a *prior distribution*. It conveys pre-existing knowledge about the measurand, before any fresh data are acquired in the new measurement experiment. This knowledge may be derived from historical data [138] describes an instance of this case for value assignment to reference solutions used in spectrometric calibrations), or it may reflect a subjective, expert assessment of what is known about the value of the measurand [130]. When no such information is available, the application of Bayesian methods requires that a choice be made of a so-called *non-informative* prior distribution for the measurand.

11.8.5 Developing a Bayesian model in practice involves choosing a prior distribution. This choice typically is influential and challenging [16, 72]. It is recommended that, when Bayesian methods are employed, the sensitivity of the conclusions to the choice of prior be assessed and documented, regardless of whether the prior results from an elicitation exercise [128] or from application of some prior selection principle [101]. Furthermore, implementing and using the Bayesian model in practice requires familiarity with suitable, specialized software for statistical computing [141].

EXAMPLE 1 *Pitot tube — model*

The airspeed v of an aircraft may be measured using a Pitot tube and the measurement model

$$v = \sqrt{2DR_sT/p},$$

where D denotes the difference between the total and static pressures, p and T denote air pressure and temperature, and the specific gas constant for air $R_s = 287.058\,\mathrm{J\,kg^{-1}\,K^{-1}}$. Reference [140] discusses this example using data from [104]: $D = 1.993\,\mathrm{kPa}$, $p = 101.4\,\mathrm{kPa}$, and $T = 292.8\,\mathrm{K}$, with associated standard uncertainties $u(D) = 0.012\,5\,\mathrm{kPa}$, $u(p) = 1.05\,\mathrm{kPa}$, and $u(T) = 0.11\,\mathrm{K}$. Suppose that the estimate of D is the average of five determinations

$$1.975\,\mathrm{kPa}, \quad 2.039\,\mathrm{kPa}, \quad 1.972\,\mathrm{kPa}, \quad 1.979\,\mathrm{kPa}, \quad 2.000\,\mathrm{kPa}$$

made under repeatability conditions, and that long experience in making these measurements suggest that the standard uncertainty associated with the estimate of D is 0.03 kPa. Suppose also that the prior knowledge about v is expressed by a normal distribution with mean $55\,\mathrm{m\,s^{-1}}$ and standard deviation $10\,\mathrm{m\,s^{-1}}$. A Bayesian model can be defined by 'inverting' the measurement equation

and reformulating it as a statistical model (cf. [69]): the determinations of D are now regarded as a sample from a normal distribution with mean $pv^2/(2R_sT)$ and standard deviation 0.03 kPa; R_s is treated as a known constant. The pressure p and the temperature T are assigned distributions with means equal to their measured values and standard deviations equal to their standard uncertainties, the distribution selected for p being normal, and the distribution selected for T being lognormal (which avoids division by 0 in the mean for D, while being practically indistinguishable from a normal distribution given that the coefficient of variation $u(T)/T$ is 0.04 %.)

EXAMPLE 2 *Copper in wholemeal flour — model*

The Analytical Methods Committee of the Royal Society of Chemistry lists the following observations of the mass fraction of copper (expressed in $\mu g\,g^{-1}$) in wholemeal flour obtained under conditions of repeatability [2]:

2.9,	3.1,	3.4,	3.4,	3.7,	3.7,	2.8,	2.5,	2.4,	2.4,	2.7,	2.2,
5.28,	3.37,	3.03,	3.03,	28.95,	3.77,	3.4,	2.2,	3.5,	3.6,	3.7,	3.7.

The statistical model for these data expresses the observations as

$$w_i = \omega + \varepsilon_i,$$

where ω denotes the true value of the mass fraction of copper in the flour, and the ε_i denote measurement errors. A cursory examination of these data suffices to conclude that it is inappropriate to assume that the ε_i are like a sample from a normal distribution. The Anderson-Darling test [4] (for example, as implemented in function `ad.test` of R package `nortest`, available at `https://www.r-project.org/`), corroborates this impression by roundly rejecting the hypothesis of normal shape for the w_i. In this case, the culprit is the single outlying observation of 28.95 $\mu g\,g^{-1}$. That Committee recommended that a Huber M-estimator of location [80] be used instead of the simple arithmetic average when the observations do not appear to be a sample from a normal distribution, as alternative to discarding observations that appear to be outliers. The Bayesian model defined in [109] (and implemented in R package BEST, available at `https://www.r-project.org/`) implements a Bayesian alternative that effectively can model and adaptively down-weights potential outliers: (a) the data are modelled as a sample from a Student's t distribution with ν degrees of freedom, re-scaled to have standard deviation σ, and shifted to have mean ω; (b) the prior distribution for ω is normal with mean equal to the average of the observations, and standard deviation 1000s, where s denotes the sample standard deviation of the observations; (c) the number of degrees of freedom ν is modelled *a priori* as $E+1$, where E is a random variable with an exponential distribution with mean 29; (d) the standard deviation σ of the ε_i has a uniform (or rectangular) prior distribution concentrated between $s/1000$ and 1000s; and (e) ω, σ, and ν are assumed to be independent *a priori*.

11.9 Estimation and uncertainty evaluation for statistical models

11.9.1 Statistical models enable recasting the measurement problem as a problem in statistical estimation, thereby consolidating, and executing simultaneously, the two key tasks involving reductions of measurement data: the assignment of a value to the measurand, and the evaluation of the associated measurement uncertainty. The generic, statistical paradigm is as follows: (i) given measurement data that will serve as inputs, a statistical model (as defined in 11.1.1) is developed that relates the probability distribution underlying the data to the value of the measurand; (ii) a criterion is selected that determines how the information about the measurand should be extracted from the data; (iii) the device used to extract such information is called an *estimator*, and its probability distribution fully characterizes the uncertainty associated with the estimate of the measurand that it produces.

11.9.2 The statistical model in (i) above may be expressed by saying that the inputs $x_1, \ldots, x_m$ are outcomes of random variables whose joint probability distribution depends on a (possibly vectorial) parameter θ, and the corresponding probability density is written as g_θ. In Bayesian models, θ is regarded as the realized value of a non-observable random variable whose distribution (*prior distribution*) also needs to be specified. Finally, a function φ is required that maps that parameter into the true value of the measurand, $\eta = \varphi(\theta)$. In example 2 of 11.3.3, the inputs are the numbers of revolutions until failure of the $m = 24$ ball bearings, $R_1, \ldots, R_m$. Their probability distribution is Weibull, and $\theta = (S, A)$ is a vector with two components: the scale S and shape A of the Weibull distribution. The measurand is the expected lifetime of the ball bearings, $\eta = \varphi(\theta) = S\Gamma(1+1/A)$.

11.9.3 Many criteria are available that determine how the information should be extracted that the data provide about the measurand. Reference [46] describes several of them, including the method of moments, least squares, maximum likelihood, and minimum Bayes risk: these will be variously illustrated in the following examples. Application of a criterion to the statistical model determines an *estimator*, which is a function ψ that takes the measurement data as inputs and produces an estimate $\widehat{\theta} = \psi(x_1, \ldots, x_m)$ of the parameter θ, whence a corresponding estimate of the true value of the measurand may be derived as $y = \varphi(\widehat{\theta})$. For all statistical models, y is the value of a random variable whose probability distribution characterizes measurement uncertainty fully.

11.9.4 In some cases it is not possible to define ψ explicitly, and $\psi(x_1, \ldots, x_m)$ is evaluated by executing the steps of an algorithm. In Example 4 in this subclause, estimating the coefficients of the analysis function used for value assignment for a reference material involves the numerical optimization of a non-linear function. In general, no 'rearrangement' of a statistical model is possible that usefully renders it in the form of a measurement model that could be dealt with as contemplated in JCGM 100:2008. This is obvious for the model discussed in example 2 of 11.3.3, owing to the presence of censored observations. In some cases, such 'rearrangement' is possible, but it depends on the underlying statistical model and on the selection of a principle of estimation, as the following example illustrates.

EXAMPLE 1 *Method of moments for Weibull lifetime* Suppose that, in example 2 of 11.3.3, none of the observations $R_1, \ldots, R_m$ (numbers of revolutions until failure of the $m = 24$ ball bearings) were censored. In these circumstances, the method of moments [18, example 2.1.2] can be employed to estimate the scale S and shape A of the Weibull distribution selected as a model for the lifetime of the ball bearings. The estimates are the solution of the following two, simultaneous, non-linear equations:

$$S\Gamma(1+1/A) = (R_1 + \cdots + R_m)/m, \qquad S^2\Gamma(1+2/A) = (R_1^2 + \cdots + R_m^2)/m.$$

This system of equations is an implicit measurement model for the vector output quantity (S, A), with the R_i as input quantities. Notice that the measurement function is determined by the underlying statistical model and by the choice of method of estimation. S and A may be estimated by evaluating the measurement function at the observed values of the inputs, and the associated uncertainties may be evaluated accordingly. In this case, the solution is $\widehat{S} = 103$ million revolutions, and $\widehat{A} = 1.89$; hence the corresponding estimate of the mean lifetime is

$$\widehat{T} = 103\Gamma(1 + 1/1.89) \text{ million revolutions} = 91.4 \text{ million revolutions}.$$

The Monte Carlo method of JCGM 101:2008 produces $u(\widehat{T}) = 10.7$ million revolutions, and the non-parametric bootstrap [53] produces $u(\widehat{T}) = 10.5$ million revolutions.

EXAMPLE 2 *Electric current — rearrangement*

Even for the simplest of statistical models, as in example 1 of clause 11.3.3, the trivial 'rearrangement'

$$\widetilde{I}_0 = (I_1 + I_2 + I_3 + I_4 + I_5)/5$$

is justified only if it is assumed that the non-observable measurement errors in the I_j are like a sample from a normal distribution, and that the estimator should have minimum mean squared error. Different assumptions would lead to estimators different from this arithmetic average.

11.9.5 Statistical models call for a varied assortment of techniques for uncertainty evaluation. In some cases, including when the method of moments is used to estimate the parameters of the distribution underlying the measured values, the approximate technique described in JCGM 100:2008 [using the law of propagation of uncertainty as in JCGM 100:2008, equation (10) or (13)] is applicable and adequate for the intended purpose.

11.9.6 When maximum likelihood is used to estimate the vector θ of parameters introduced in 11.9.2, and the conditions detailed in [18, theorem 5.4.3] are satisfied (as they will be in most cases likely to be encountered in practice), then $H^{-1}(\widehat{\theta})$ provides an approximation either to the squared standard uncertainty (when θ is a scalar), or to the covariance matrix (when θ is a vector), associated with the maximum likelihood estimate, $\widehat{\theta}$. When θ is a scalar, $H^{-1}(\widehat{\theta})$ is the reciprocal of the second derivative of the negative loglikelihood function corresponding to the probability distribution underlying the inputs, evaluated at $\widehat{\theta}$. When θ is a vector, $H^{-1}(\widehat{\theta})$ is the inverse of the matrix of the second-order partial derivatives of the negative loglikelihood function, evaluated at $\widehat{\theta}$. In this case, the square roots of the main diagonal elements of $H^{-1}(\widehat{\theta})$ are approximate evaluations of the standard uncertainties associated with the estimates of the parameters, and the covariances between them are given by the off-diagonal elements of the same matrix. Since the measurand is a known function of these parameters (in the aforementioned example 2 of 11.3.3 the measurand is $T = S\Gamma(1 + 1/A)$), the uncertainty associated with the corresponding estimate can then be computed using equation (13) in JCGM 100:2008.

11.9.7 Bayesian procedures produce a probability distribution for the parameters of the distribution underlying the inputs (or an arbitrarily large sample drawn from this distribution), from which the probability distribution (or a sample from the distribution) of the measurand may be derived straightforwardly. This distribution depends not only on the model used for the observations, but also on the prior distribution chosen for the parameters that appear in the model for the observations.

11.9.8 In many other cases, methods other than those just mentioned will have to be employed, for example Monte Carlo methods for propagation of distributions (JCGM 101:2008 and [123]) or bootstrap methods [53]. Also see JCGM 102:2011.

EXAMPLE 1 *Electric current — estimation and uncertainty evaluation*

The maximum likelihood estimates (MLEs) of I_0 and σ that correspond to the model described in Example 1 of 11.3.3 are as follows [46, Example 7.5.6]:

$$\overline{I} = \frac{I_1 + \cdots + I_5}{5} = 19.661\,\text{mA}, \qquad \widehat{\sigma} = \sqrt{\frac{(I_1 - \overline{I})^2 + \cdots + (I_5 - \overline{I})^2}{5}} = 0.019\,\text{mA}.$$

For *unbiased* estimates (that is, estimates whose mathematical expectations are equal to the respective measurands), the estimate of I_0 continues to be $\bar{I}$, but the estimate of σ is as follows [47]:

$$\widehat{\sigma} = \sqrt{\frac{2}{5-1}} \cdot \frac{\Gamma(\frac{5}{2})}{\Gamma(\frac{5-1}{2})} \cdot \sqrt{\frac{(I_1-\bar{I})^2+\cdots+(I_5-\bar{I})^2}{5-1}} = 0.020\,\text{mA}.$$

If the measurement errors are best modelled as a sample from a Laplace distribution, then the MLE of I_0 is the median of the observations, $\widetilde{I}_0 = 19.663\,\text{mA}$, and the MLE of σ is the average of the absolute values of the deviations of the observations from the median, multiplied by $\sqrt{2}$, which is $\widetilde{\sigma} = 0.024\,\text{mA}$. If the estimate of I_0 is $\bar{I}$, then the law of propagation of uncertainty provides an exact evaluation of $u(\bar{I})$, which is the same as the evaluation that corresponds to maximum likelihood estimation, $u(\bar{I}_0) = \widehat{\sigma}/\sqrt{5} = 0.009\,\text{mA}$. However, if the estimate should be

$$\widetilde{I}_0 = \text{median}\{I_1,\ldots,I_5\}$$

instead, then the techniques described in JCGM 100:2008 are unable to evaluate $u(\widetilde{I}_0)$. The evaluation corresponding to the MLE for Laplace errors is $u(\widetilde{I}_0) = 0.023\,8\,\text{mA}/\sqrt{5} = 0.011\,\text{mA}$. When the median still seems to be the best choice as estimate of I_0, but no specific distributional shape can confidently be assumed for the measurement errors other than their distribution is symmetrical and centered at 0, then an evaluation of the associated uncertainty based on inversion of the Wilcoxon test is recommended [141]: for the data considered in this example this produces $0.011\,5\,\text{mA}/\sqrt{5} = 0.005\,\text{mA}$.

EXAMPLE 2 *Lifetime of ball bearings — estimation and uncertainty evaluation*

In example 2 of 11.3.3, estimating the expected lifetime, $T = S\Gamma(1+1/A)$, reduces to estimating the scale S and shape A of the Weibull model chosen for the inputs. The method of maximum likelihood (and Bayesian estimation procedures) are able to take into account the information provided by the four censored observations. The MLEs of S and A are the values that maximize

$$\sum_{i\in\mathcal{U}} \ln w_{S,A}(R_i) + \sum_{i\in\mathcal{C}} \ln(1-W_{S,A}(R_i))$$

with respect to S and A, where $w_{S,A}$ and $W_{S,A}$ denote the probability density and the probability cumulative distribution function of the Weibull distribution with scale S and shape A, $\mathcal{U}$ denotes the set of indices corresponding to uncensored observations, and $\mathcal{C}$ denotes the set of indices corresponding to censored observations. The MLEs are $\widehat{S} = 111$ million revolutions and $\widehat{A} = 1.34$; hence $\widehat{T} = 102$ million revolutions. The approximate uncertainty evaluations associated with the MLEs are $u(\widehat{S}) = 19$ million revolutions and $u(\widehat{A}) = 0.26$, and the correlation between them is $r(\widehat{S},\widehat{A}) = 0.096\,6$, computed as described in 11.9.6. Application of formula (13) in JCGM 100:2008 then yields $u(\widehat{T}) = 17$ million revolutions. The parametric statistical bootstrap evaluates the same uncertainty at 20 million revolutions.

EXAMPLE 3 *Mass fraction of tin — estimation and uncertainty evaluation*

The DerSimonian-Laird estimate of the mass fraction of tin produced by the *NIST Consensus Builder* [106, 107] with the inputs specified in example 3 of 11.3.3 is $10.01\,\text{mg}\,\text{g}^{-1}$. The hierarchical Bayesian procedure also implemented in the *NIST Consensus Builder* produced the same estimate. The evaluation of the standard uncertainty associated with the DerSimonian-Laird estimate is $0.011\,\text{mg}\,\text{g}^{-1}$. A 95 % coverage interval for the mass fraction of tin ranges from $9.99\,\text{mg}\,\text{g}^{-1}$ to $10.04\,\text{mg}\,\text{g}^{-1}$. The corresponding values produced by the hierarchical Bayesian procedure are $0.025\,\text{mg}\,\text{g}^{-1}$ for the standard uncertainty and $9.96\,\text{mg}\,\text{g}^{-1}$ to $10.06\,\text{mg}\,\text{g}^{-1}$ for a 95 % coverage interval. The Bayesian procedure yields considerably larger uncertainty evaluations than the DerSimonian-Laird procedure because it captures more accurately, and expresses more realistically, the fact that the component of uncertainty attributable to differences between analytical methods is evaluated based on a single degree of freedom [106].

EXAMPLE 4 *Amount fraction of* N_2O *in southern oceanic air — estimation and uncertainty evaluation*

The estimates of the coefficients of the cubic polynomial

$$A(r) = \beta_0 + \beta_1 r + \beta_2 r^2 + \beta_3 r^3$$

that was selected as a model for the analysis function in the example of 11.4.3 are the values of the coefficients $\beta_0, \ldots, \beta_3$ and of the true ratios $\rho_1, \ldots, \rho_6$ that maximize

$$\sum_{i=1}^{6} [\ln \phi((x_i - A(\rho_i))/u(x_i)) - \ln u(x_i) + \ln \psi_9((r_i - \rho_i)/u(r_i)) - \ln u(r_i)],$$

where $x_1, \ldots, x_6$ denote the amount fractions of N_2O in the calibration standards, ϕ denotes the probability density of the normal distribution with mean 0 and standard deviation 1, and ψ_9 denotes the probability density of the Student's t distribution with 9 degrees of freedom (because the ratios for the standards are averages of 10 replicates). The estimates obtained by numerical optimization using the Nelder-Mead algorithm as implemented in R function `optim` are

$$\widehat{\beta}_0 = 329.13\,\text{nmol/mol}, \quad \widehat{\beta}_1 = 24.928\,\text{nmol/mol},$$
$$\widehat{\beta}_2 = -0.296\,52\,\text{nmol/mol}, \quad \widehat{\beta}_3 = -0.196\,12\,\text{nmol/mol},$$
$$\widehat{\rho}_1 = 0.972\,39, \quad \widehat{\rho}_2 = 0.987\,08, \quad \widehat{\rho}_3 = 1.006\,5, \quad \widehat{\rho}_4 = 1.021\,3, \quad \widehat{\rho}_5 = 1.043\,7,$$
$$\widehat{\rho}_6 = 1.063.$$

The uncertainty associated with $A(r)$ was evaluated using the Monte Carlo method described in [73].

EXAMPLE 5 *Barium in clay soil — estimation and uncertainty evaluation*

The random effects model described in the example of 11.5.4 may be fitted to the data by any one of several different methods. The conventional ANOVA approach to estimate τ and σ uses the method of moments [18, example 2.1.2], equating observed and expected mean squares, and then solving for τ^2 and σ^2. Since this approach may produce a negative estimate for τ^2 — compare formula (5) in [159] — it is preferable to fit the model by restricted maximum likelihood (REML) [150] (for example, as implemented in function `lmer` of R package `lme4`, available at `https://www.r-project.org/`). For this data set, the REML estimates of τ and σ reproduce the results of the conventional ANOVA, $\widehat{\tau} = 5.3\,\text{mg kg}^{-1}$ and $\widehat{\sigma} = 11\,\text{mg kg}^{-1}$, which are denoted by s_{bb} and s_{within} in [159]. The REML procedure also produces an estimate $\widehat{\omega} = 318\,\text{mg kg}^{-1}$ of the measurand, and an evaluation of the associated standard uncertainty $u(\widehat{\omega}) = 2\,\text{mg kg}^{-1}$. An approximate, profile-likelihood 95 % coverage interval for τ ranges from $0\,\text{mg kg}^{-1}$ to $9.8\,\text{mg kg}^{-1}$, thus confirming the conclusion of the conventional F-test enabled by the ANOVA, whose p-value is 0.09, as already noted in the example under 11.5.4. (The profile likelihood for τ is its marginal probability density evaluated at the maximum likelihood estimates of the other parameters in the model (see [133], page 61).)

EXAMPLE 6 *Calibrating a geodetic base line — estimation and uncertainty evaluation*

The method of least squares amounts to choosing values for λ_1, λ_2, and λ_3 that minimize $S(\lambda_1, \lambda_2, \lambda_3) = (L_1 - \lambda_1)^2 + (L_2 - \lambda_2)^2 + (L_3 - \lambda_3)^2 + (L_4 - \lambda_1 - \lambda_2)^2 + (L_5 - \lambda_2 - \lambda_3)^2 + (L_6 - \lambda_1 - \lambda_2 - \lambda_3)^2$ with respect to λ_1, λ_2, and λ_3. Computing $\partial S(\lambda_1, \lambda_2, \lambda_3)/\partial \lambda_1 = 6\lambda_1 + 4\lambda_2 + 2\lambda_3 - 2L_1 - 2L_4 - 2L_6$, and $\partial S(\lambda_1, \lambda_2 \lambda_3)/\partial \lambda_2$ and $\partial S(\lambda_1, \lambda_2 \lambda_3)/\partial \lambda_3$ similarly, and equating all to 0 leads to the following system of three simultaneous equations in three unknowns:

$$3\lambda_1 + 2\lambda_2 + \lambda_3 = L_1 + L_4 + L_6,$$
$$2\lambda_1 + 4\lambda_2 + 2\lambda_3 = L_2 + L_4 + L_5 + L_6,$$
$$\lambda_1 + 2\lambda_2 + 3\lambda_3 = L_3 + L_5 + L_6.$$

The solution is $\widehat{\lambda}_1 = 27.22\,\text{m}$, $\widehat{\lambda}_2 = 33.08\,\text{m}$, and $\widehat{\lambda}_3 = 25.30\,\text{m}$. Therefore, the distance from A to D is estimated as $\widehat{\Delta} = \widehat{\lambda}_1 + \widehat{\lambda}_2 + \widehat{\lambda}_3 = 85.59\,\text{m}$. Since the eigenvalues of the matrix of second derivatives of S as defined above, evaluated at $(\widehat{\lambda}_1, \widehat{\lambda}_2, \widehat{\lambda}_3)$, all are positive, the solution indeed corresponds to a minimum. The adjusted observations (or, fitted lengths) are $\widehat{L}_1 = 27.22\,\text{m}$, $\widehat{L}_2 = 33.08\,\text{m}$, $\widehat{L}_3 = 25.30\,\text{m}$, $\widehat{L}_4 = 60.30\,\text{m}$, $\widehat{L}_5 = 58.38\,\text{m}$, and $\widehat{L}_6 = 85.59\,\text{m}$.

Next, we compute a Type A evaluation of the uncertainty associated with $\widehat{\Delta}$, the estimate of the distance from A to D. First one computes the covariance matrix of $(\widehat{\lambda}_1, \widehat{\lambda}_2, \widehat{\lambda}_3)$ using conventional methods of linear regression [50] (for example, as implemented in function `lm` of 'base' R, available at `https://www.r-project.org/`). Second, considering that $\Delta = \lambda_1 + \lambda_2 + \lambda_3$ is a linear measurement model of the sort considered in JCGM 100:2008, equation (13) may be used to obtain $u(\Delta) = 0.025\,\text{m}$ with 3 degrees of freedom.

EXAMPLE 7 *Pitot tube — estimation and uncertainty evaluation*

Application of MCMC sampling (for example, as implemented in function `metrop` of R package `mcmc`, available at `https://www.r-project.org/`), to the Bayesian model defined in example 1 of 11.8.5, produces a sample from the posterior distribution of v with mean 57.5 m/s and standard deviation 0.4 m/s — a measurement result that, for this particular choice of prior distribution, is statistically indistinguishable from the result obtained via the conventional analysis according to JCGM 100:2008: $v = 57.48\,\text{m/s}$ and $u(v) = 0.35\,\text{m/s}$ [140, A.6].

EXAMPLE 8 *Copper in wholemeal flour — estimation and uncertainty evaluation*

The Huber M-estimator of location [80] (for example as implemented in function `huberM` of R package `robustbase`, available at `https://www.r-project.org/`) provides a robust alternative to the arithmetic average as an estimate of ω that yields both an estimate $3.21\,\mu\text{g}\,\text{g}^{-1}$ of that mass fraction, and an evaluation $0.14\,\mu\text{g}\,\text{g}^{-1}$ of the associated standard uncertainty. The Bayesian model described in example 2 of 11.8.5 produces a posterior distribution for the true value of the mass fraction whose mean, $3.22\,\mu\text{g}\,\text{g}^{-1}$, is an estimate of the measurand ω, and whose standard deviation, $0.15\,\mu\text{g}\,\text{g}^{-1}$, is an evaluation of the associated standard uncertainty. These results are in close agreement with the results of the classical robust analysis using the Huber M-estimator discussed above. Coverage intervals can also be derived from the posterior distribution: for example, $2.92\,\mu\text{g}\,\text{g}^{-1}$ to $3.51\,\mu\text{g}\,\text{g}^{-1}$ is a 95 % Bayesian coverage interval for the mass fraction of copper.

EXAMPLE 9 *Temperature of a thermal bath — estimation and uncertainty evaluation* The MLEs of the parameters in the auto-regressive model described in the example of 11.7.3 are

$$\widehat{\tau} = 50.105\,^{\circ}\text{C}, \qquad \widehat{\varphi}_1 = 0.41, \qquad \widehat{\varphi}_2 = 0.42, \qquad \sigma = 0.002\,^{\circ}\text{C}.$$

They were computed using function `Arima` defined in package `forecast` (available at `https://www.r-project.org/`) [82]. The same function also provides evaluations of the corresponding standard uncertainties

$$u(\tau) = 0.001\,^{\circ}\text{C}, \qquad u(\varphi_1) = u(\varphi_2) = 0.09,$$

computed as described in 11.9.6. It should be noted that $u(\tau)$ is about three times larger than the naive (and incorrect) uncertainty evaluation $s/\sqrt{100} = 0.000\,3\,^{\circ}\text{C}$ that would have been obtained neglecting the auto-correlations, where s denotes the standard deviation of the 100 readings of temperature listed in Table 8 of 11.7.3. According to the auto-regression used as a model for these data, the correlations between the observations arise owing to a memory effect: the temperature at time T_i is determined by the values of temperature measured during the previous two minutes, plus a volatile (unpredictable, or 'random') measurement error whose standard deviation is 0.002 °C. The state of thermal equilibrium is a *bona fide* measurand, characterized by the coefficients of the auto-regression: in this case, the estimates of these coefficients correspond to a *stationary* process, meaning that the level, magnitude, and pattern of the oscillations of temperature do not depend on the time of the first observation.

11.10 Model selection and model uncertainty

11.10.1 Whenever a statistical model is considered, a choice needs to be made about the particular model to use (*model selection*), because in most cases several models can be reasonable alternatives. In the example of 11.7, a normal auto-regression of order 2 was selected for the time series of observations of temperature, but several other ARIMA models would be similarly reasonable.

11.10.2 Model selection typically involves making a compromise between the goodness of fit of the model to the data, and the trustworthiness and fitness for purpose of the model in the application where the results will be used. For example, in the context of calibration, as discussed in the example of 11.4.3, a model that tracks all the values pertaining to the standards very closely may be too sensitive to their peculiarities and not serve as a reliable summary and useful generalization of the relationship that it strives to capture. The choice of degree for the polynomial used for the analysis function illustrates the compromise between achieving good fit and selecting a parsimonious model (usually a polynomial of low degree).

11.10.3 Commonly used model selection criteria [25] include (i) the Akaike Information Criterion (AIC) [1]; (ii) its version AICc [81] 'corrected' for the finite size of the data set used to build the model; and (iii) the Bayesian Information Criterion (BIC) [149]. None of these should be the sole guide used for model selection. Model selection criteria suggest a reasonable compromise between goodness of fit and simplicity of the model. They should always be supplemented with statistical examinations of residuals (differences between observations and model predictions), for instance QQ-plots [160], and cross-validation (to determine how well alternative models can predict observations that were not used in the process of building them) [75, 124].

EXAMPLE *Temperature of a thermal bath — model selection*

According to BIC, a normal auto-regression of order 2, often denoted by ARIMA$(2,0,0)$, is 'best' for the time series of observations of temperature mentioned in the example in 11.7, among all auto-regressive moving-average models with up to 5 coefficients besides a non-zero mean. (In general, the notation ARIMA(p,d,q) denotes a time series model for the data after computing d differences, comprising p auto-regressive terms and q moving average terms.) On the other hand, according to AICc, the 'best' model is an ARIMA$(1,0,2)$ with non-zero mean [139].

11.10.4 Instead of selecting a particular model, several models may be used and the corresponding results combined into some sort of average (*model averaging*). This averaging may be as simple as illustrated in the following example, or it may take into account how likely *a priori* the different models are judged to be [33, 65]. Furthermore, the uncertainty surrounding model selection (*model uncertainty*) should be evaluated and propagated to the results recognizing the several alternatives [30, 36], and become just another contribution to the uncertainty to be associated with the estimate of the measurand.

EXAMPLE *Copper in wholemeal flour — model averaging*

In example 2 of 11.8, the normal model was discredited by the Anderson-Darling test, and a Bayesian model was used instead that offers built-in protection against potential outliers. However,

several other models would do likewise, and instead of selecting one, several may be employed and the corresponding results averaged. For example: the Laplace model suggests the median, 3.39 µg/g with associated standard uncertainty 0.19 µg/g (computed using the non-parametric, statistical bootstrap [53]); the Huber M-estimate of location is 3.21 µg/g with associated standard uncertainty 0.14 µg/g; the Hodges-Lehmann estimate [77] is 3.23 µg/g with associated standard uncertainty 0.17 µg/g (derived from the 68 % coverage interval obtained by inversion of the Wilcoxon rank sum test [78]); and the aforementioned Bayesian model yields 3.22 µg/g with associated standard uncertainty 0.15 µg/g. The weighted average of these four estimates, with weights proportional to the reciprocals of their squared standard uncertainties, is 3.25 µg/g, with associated standard uncertainty 0.17 µg/g (computed using equation (9) in [24]).

12 Assessing the adequacy of the measurement model

12.1 A measurement model is adequate if the estimate of the measurand obtained by using it is corrected for all known effects, and the associated uncertainty reflects all factors that could reasonably affect the estimate. If this is the case, values for the same measurand obtained using alternative models (see 8.5) should agree within reported uncertainties. In addition, the measurement model should be capable of providing estimates and associated uncertainties for the entire range(s) of measurand values and values of the input quantities for which it has been developed.

12.2 Experimental evidence of measuring system performance is the most general check of adequacy of the model chosen. Such evidence is usually obtained when validating the measurement procedure, or performing measurements in support of quality control and quality assurance. There are different ways to assess by experiment the output of a measurement model, such as

- participation in an interlaboratory comparison and use of the estimate and uncertainty from the measurement model in the performance evaluation,
- use of a certified reference material (CRM) or a reference method to assess the performance of the measurement procedure,
- comparison of the output of the measurement model with literature or reference data.

By showing agreement between the measured and reference values within the respective uncertainties, support is provided for the adequacy of that model. Similarly, unexpected disagreement normally calls for further investigation and should result in improvements to the preferred measurement model.

EXAMPLE *Performance evaluation in an interlaboratory comparison*

Computing an E_n or ζ score [88] from participation in a proficiency test facilitates assessing the performance evaluation taking into account both the measured value and the stated measurement uncertainty. Similarly, in key comparisons such an evaluation is facilitated by calculating degrees of equivalence [19]. An unsatisfactory score can indicate an issue in the measurement model.

NOTE 1 Guidance on using CRMs is given in ISO Guide 33 [96].

NOTE 2 Depending on the kind of CRM used, adaptations to the measurement model can be necessary, for example, because the form of the CRM can be different from the items subject to measurement for which the measurement model has been developed.

12.3 Supplementary to assessing experimentally the outputs of a measurement model for adequacy, it can be helpful to assess targeted parts of the measurement procedure separately, to assess aspects of the procedure, such as, but not limited to,

— the calibration function and (deviations from) linearity of the instrument,
— sampling,
— sample preparation,
— reproducibility effects.

The use of validation studies and quality control data can be helpful in making these assessments.

12.4 Reference data (exact or otherwise) for comparison purposes is widely used [26]. 'Exact' data can often be provided by inverting the proposed model to generate example data from prescribed values for the measurand. This process provides reassurance that a model is correctly implemented. Where it is possible to generate sample data from a more complete model, or from data on well-characterized test items, demonstration that the expected result is returned will provide additional confidence in the model actually used. In many cases, example data and calculations are published in relevant measurement or testing methods; comparison of calculated results with those in the respective standards then provides evidence of model adequacy.

12.5 Assessing the adequacy over the ranges of input and output values requires verifying the output using one or more of the methods described in this clause. For instance, where a calibration function is to be used for measuring pH in the range 4 to 10, it should be ensured that either a single function performs sufficiently well over this range, or that alternative functions perform adequately (with appropriate calibration) over sub-ranges. In assessing adequacy across a range or sub-range, it is usually sufficient to explore the intended extremes and a small number of intermediate cases. Close attention should be given to regions where modelling assumptions change. If the number of terms in a computational model is adjusted depending on a measured response, the adequacy of the model should be confirmed near each such decision point. Consideration should be given to the sensitivity of a measurement result to the choice of model [143], particularly in critical applications. Also see 11.10.

NOTE It is not generally necessary for testing laboratories to confirm the adequacy of a calculation procedure given in a standard test or measurement method. Such a procedure would imply an underlying measurement model. Such a confirmation would have normally been undertaken by the expert committee in developing a relevant documentary standard or have been established by prior research.

12.6 Supplementary to the checks on the output of the measurement model, there are several others that are especially helpful in verifying the appropriateness of the measurement model. These activities include the following, which are discussed further in the subsequent subclauses:

— Reviewing the contributions to the measurement model;
— Checking that a simplified model adequately reproduces the results of a more complete model it approximates using, for instance, the target measurement uncertainty for this purpose;

— Confirming that the computer's numerical accuracy is sufficient for the application;
— For models fitted to data, verifying the validity of the model and of the distribution assumptions made;
— Obtaining evidence of model applicability from published research literature;
— Obtaining experimental evidence that the model performs adequately for its intended use; and
— Confirming that uncertainties predicted by the model are consistent with the observed performance of the measurement procedure.

12.7 The contributions to the measurement model should be reviewed. They include contributions obtained from established scientific principles (see 7.1) and effects derived from metrological considerations (see 9.2, clause 9 and annex C). The developed model should be checked against the compiled list of contributory effects to assess that all meaningful contributions have been taken into account (see 9.3 and E.1.1). In this review, account should be taken of any prescribed target measurement uncertainty. Also the possibility of understated values for uncertainty contributions should be part of this review.

12.8 Where the proposed model is an intentional simplification of a more complicated model that can also be implemented, comparison of results obtained by the two models will provide an indication of the agreement between the two. Where measurement uncertainties and differences in calculated values are small compared with the target measurement uncertainty, the simplified model may be accepted as adequate. See 11.10.

12.9 Different mathematical formulations or solution methods can lead to results of different numerical accuracy due to the limitations of computer arithmetic, premature termination of iterative calculations or for other reasons. Confirming sufficient numerical accuracy, often using reference or simulated data, can therefore be important when developing new implementations. Model choice for adequate numerical accuracy is discussed in more detail in 8.6.

12.10 Research literature can provide evidence of the adequacy of a given model, sometimes over a particular domain. Literature may also be useful in establishing the limitations of a suggested model.

13 Using the measurement model

13.1 General

13.1.1 The main use of a measurement model is to calculate an estimate of the measurand and the associated uncertainty. Depending on the requirements, the uncertainty can be provided as a standard uncertainty, a coverage interval for a stipulated coverage probability, or both. Especially when using Monte Carlo or Bayesian methods, it can be more appropriate to provide a (representation of) the probability distribution for the measurand. The latter is especially true when the output of the measurement model is used in a subsequent uncertainty calculation taking a probability distribution as input.

13.1.2 Given values for the input quantities in the measurement model, the corresponding value of the measurand is obtained by evaluating a formula or by solving an equation. Some quantities can take complex values. There can be a single (scalar) measurand or multiple measurands (vector measurand). The mathematical, statistical or numerical tools applied for obtaining the estimate of the measurand and uncertainty evaluation depend on the form of the measurement model.

13.1.3 The law of propagation of uncertainty holds exactly for a linear measurement model, that is, a model that is linear in its input quantities, but not generally otherwise. For a single output quantity, a linear measurement model takes the form

$$Y = c_0 + c_1 X_1 + \cdots + c_N X_N, \tag{27}$$

where $c_0, \ldots, c_N$ are exactly known constants. This expression is straightforwardly generalized to multivariate models, and to those involving complex quantities. Guidance on how to use the law of propagation of uncertainty is given in JCGM 100:2008 for explicit univariate measurement models [as in expression (1)] and in JCGM 102:2011 for other forms, including multivariate measurement models.

13.1.4 A non-linear measurement model is a model that cannot exactly be expressed in the form (27). With some degree of approximation, a non-linear model can be expressed locally in linear form in the neighbourhood of estimates $x_1, \ldots, x_N$ of the input quantities. On the basis of this linear form the law of propagation of uncertainty can be applied to yield an approximation to the standard uncertainty associated with the estimate $y = f(x_1, \ldots, x_N)$ of the output quantity. The adequacy of this approximation can be checked in any particular instance by applying the Monte Carlo method of JCGM 101:2008, 7, or JCGM 102:2011, 7 in a multivariate case. Annex F gives further details of model linearization and provides a simple method (that is not foolproof) of checking the approximation.

EXAMPLE *Failure of the law of propagation of uncertainty to propagate uncertainty in a non-linear model*

Bromine has two stable isotopes, bromine-79 and bromine-81. Consider a dibromine molecule (Br_2) whose most abundant molecule is $^{79}Br^{81}Br$ (the other molecules being $^{79}Br^{79}Br$ and $^{81}Br^{81}Br$). The abundance of this molecule, having a mass of 160 Da, with respect to all molecules of Br_2 is given by the expression [120]

$$x_{160} = 2x_{79}(1 - x_{79}).$$

The isotopic abundance x_{79} of bromine-79 varies naturally from 0.505 to 0.508. Consider that Br_2 has been made from a slightly altered isotopic bromine with $x_{79} = 0.500$ and $u(x_{79}) = 0.005$, and the abundance x_{160} of $^{79}Br^{81}Br$ is required. In such a case, the application of the law of propagation of uncertainty for the above measurement model yields

$$u^2(x_{160}) = \left(\frac{\mathrm{d}x_{160}}{\mathrm{d}x_{79}}\right)^2 u^2(x_{79}) = (2 - 4x_{79})^2 u^2(x_{79}).$$

In this instance, the law of propagation of uncertainty provides invalid results since it yields $u(x_{160}) = 0$ when $x_{79} = 0.500$, regardless of the uncertainty associated with x_{79}. The inadequate linearization of this model as used by the law of propagation of uncertainty can be remedied by using

the Monte Carlo method: by modelling x_{79} by a normal random variable, N(0.500, 0.005^2), for instance, $x_{160} = 0.499\,95$ with $u(x_{180}) = 0.000\,07$ is obtained.

The issue can also be overcome by calculating non-symmetric approximations for the sensitivity coefficients or by including higher-order terms [120].

13.2 Use of a model beyond the range for which it has been validated

The use of a measurement model for extrapolation beyond an interval over which it is deemed valid is generally inadvisable (also see 5.8 and the example in 10.6.4). It is generally unsafe to fit an empirical or hybrid model to data and use the model beyond the span of the fitted data. Extrapolation with purely empirical functions such as polynomials and polynomial splines is particularly dangerous. Such functions can exhibit spurious behaviour beyond the span of the data they explain and should not be used for this purpose. In certain circumstances, a theoretical model may be used with caution for extrapolation beyond the span of the data that defines it when the model is judged to describe adequately the phenomenon of interest.

EXAMPLE *Isotope-based quantitation: use of theoretical and empirical models*

In isotope-based quantitation, routinely employed in chemical measurement, the measurand is the isotope amount ratio of a substance of interest (the analyte) in an 'unknown' sample. This ratio is specified by a calibration function giving the relationship between isotope amount ratio and mass ratio. The calibration function is provided by regression given a data set consisting of several pairs of values. Each pair corresponds to a 'known' sample, for which a value of R_{AB}, the isotope ratio of the mixture, is observed (measured) corresponding to a value of q, the mass ratio of the sample (analyte) A and the isotopically-labelled internal standard B. See the data points indicated by filled circles in figure 9.

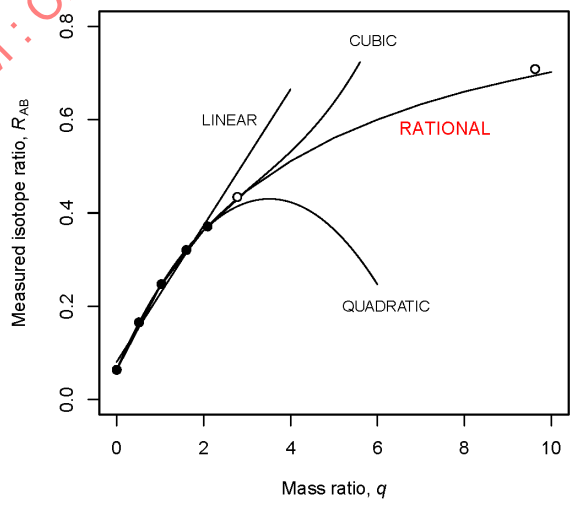

Figure 9: Isotope dilution theoretical (rational) model and empirical models fitted to calibration data and used for prediction

A feature that distinguishes isotope dilution from many other methods of analytical chemistry is the inherently non-linear theoretical calibration function,

$$R_{AB} = \frac{a_0 + a_1 q}{1 + a_2 q}, \tag{28}$$

which is a rational function having polynomial degree 1 in the numerator and in the denominator, where a_0, a_1 and a_2 are parameters calculated by regression from the provided data. Then, for a sample having a specified value of mass ratio q, the corresponding value of R_{AB} is determined from the function (28). Accordingly, the calibration function given by expression (28) is an explicit univariate model (see 13.3) in which R_{AB} is the output quantity or measurand, and the input quantities are q and the calibration parameters a_0, a_1 and a_2.

Many current implementations of isotope dilution calibration either ignore the curvature of the function (due to the presence of the term involving a_2) or use instead an empirical model such as a polynomial of degree 2 or 3 (see annex D). Such empirical models do not generally extrapolate well beyond the interval spanned by the calibration data. Figure 9 shows the use of a linear polynomial or straight line (corresponding to ignoring the term $a_2 q$), and quadratic and cubic polynomials fitted to five calibration points (filled circles). It also shows the function (28) fitted to the same data. The open circles correspond to two additional samples outside the span of the calibration data. The advantage of the theoretical model over the candidate empirical models is evident in this case. Further details are given in reference [132].

13.3 Explicit univariate measurement model

A measurement model taking the form (1) relating a single output quantity Y to the input quantities $X_1, \ldots, X_N$ is termed univariate and explicit. Given values for the input quantities $X_1, \ldots, X_N$, an explicit model enables the corresponding value of the output quantity Y to be determined directly. This form of model is the simplest and can be handled by the provisions of JCGM 100:2008 or JCGM 101:2008.

EXAMPLE *Gauge-block measurement* (also see the example in 6.2 and example 1 in 6.5)

The length of a nominally 50 mm gauge block is determined by comparing it with a known gauge-block measurement standard of the same nominal length using a gauge block comparator. An expression for the direct output of the comparison of the two gauge blocks is the difference

$$d = [1 + (\alpha_S + \delta\alpha)\theta]\ell - [1 + \alpha_S(\theta - \delta\theta)]\ell_S \tag{29}$$

in their lengths, where

- ℓ is the length at 20 °C of the gauge block being calibrated,
- ℓ_S is the length of the gauge-block measurement standard at 20 °C as given on its calibration certificate,
- α_S is the coefficient of thermal expansion of the gauge-block measurement standard,
- $\delta\alpha = \alpha - \alpha_s$, where α is the coefficient of thermal expansion of the gauge block being calibrated,
- θ is the deviation in Celsius temperature from the 20 °C reference temperature of the gauge block being calibrated, and
- $\delta\theta = \theta - \theta_S$, where θ_S is the deviation in Celsius temperature from the 20 °C reference temperature of the gauge-block measurement standard.

From expression (29) the output quantity ℓ can be expressed in terms of the $N = 6$ input quantities d, ℓ_S, α_S, $\delta\alpha$, θ and $\delta\theta$ as the explicit univariate measurement model

$$\ell = \frac{[1 + \alpha_S(\theta - \delta\theta)]\ell_S + d}{1 + (\alpha_S + \delta\alpha)\theta}, \tag{30}$$

the right-hand side of model (30) constituting the measurement function.

13.4 Explicit multivariate measurement model

A measurement model taking the form

$$Y_1 = f_1(X_1, \ldots, X_N), \quad \ldots, \quad Y_m = f_m(X_1, \ldots, X_N). \tag{31}$$

where $Y_1, \ldots, Y_m$ are m output quantities, and $f_1, \ldots, f_m$ denote the m measurement functions, is termed multivariate and explicit. This form of model can be handled by the provisions of JCGM 102:2011.

EXAMPLE *Resistance and reactance of a circuit element*

The resistance R and reactance X of a circuit element are given by JCGM 100:2008, example H.2:

$$R = \frac{V}{I}\cos\phi, \qquad X = \frac{V}{I}\sin\phi. \tag{32}$$

Here expression (32) is the explicit multivariate measurement model for resistance and reactance in which V denotes potential difference, I denotes current and ϕ denotes phase angle. There are $N = 3$ input quantities, V, I and ϕ, and $m = 2$ output quantities R and X. The symbol X, conventionally denoting reactance, is not to be confused with X_1, X_2, etc., the generic symbols for the input quantities.

13.5 Implicit univariate measurement model

A measurement model taking the form

$$h(Y, X_1, \ldots, X_N) = 0, \tag{33}$$

where Y is a scalar output quantity, and h denotes a function of Y and the input quantities $X_1, \ldots, X_N$, is termed univariate and implicit. Given values for the input quantities $X_1, \ldots, X_N$, the corresponding value of the output quantity Y cannot be calculated directly from an implicit model. Specifically, equation (33) cannot readily or numerically stably be represented in terms of a measurement function. It is generally solved numerically, often by using an iterative procedure, to obtain a value of Y corresponding to given values of $X_1, \ldots, X_N$. This form of model can be handled by the provisions of JCGM 102:2011, which covers a generalization of the univariate form of the law of propagation of uncertainty given in JCGM 100:2008, 5.1 and 5.2 and a Monte Carlo method.

NOTE 1 There is not a unique way to write an implicit measurement model. For instance, in place of equation (3) in example 2 in 7.2.1 the measurement model given by equating to zero the difference between the left- and right-hand sides of expression (3) could be used, which will then take the form (33). The numerical efficiency of the solution of the measurement model equation depends on the choice made.

NOTE 2 It is assumed that the implicit univariate measurement model (33) can be solved uniquely for the value of Y corresponding to given values of $X_1, \ldots, X_N$ close to their estimates.

EXAMPLE *Shell thickness of microscopic spherical particles*

A practical approach to determining the overlayer thickness of samples of microscopic spherical particles is to calculate an equivalent planar thickness T and multiply it by a geometrical correction

term [151]. An implicit measurement model of the relationship between the measurand T and experimentally measured independent quantities A and B is [42]

$$h(T,A,B) \equiv e^{-T} - Ae^{-BT} - 1 = 0. \tag{34}$$

In equation (34), A and B are the input quantities and T is the output quantity or measurand. It is not possible to re-express T explicitly in terms of A and B. Using the provisions of JCGM 102:2011, given estimates $\widehat{A}$ and $\widehat{B}$ of A of B, this equation can be solved numerically for the corresponding estimate $\widehat{T}$ of T. Moreover, the standard uncertainty $u(\widehat{T})$ can be obtained by propagating the standard uncertainties associated with $\widehat{A}$ and $\widehat{B}$.

13.6 Implicit multivariate measurement model

A measurement model taking the form

$$h_1(Y_1,X_1,\ldots,X_N) = 0, \quad \ldots, \quad h_m(Y_m,X_1,\ldots,X_N) = 0, \tag{35}$$

implicitly defining the output quantities $Y_1,\ldots,Y_m$ in terms of the input quantities $X_1,\ldots,X_N$, is termed multivariate and implicit. Typically, equation (35) cannot readily or numerically stably be represented in terms of measurement functions $f_1,\ldots,f_m$ as in the explicit multivariate case in 13.4. It is generally solved numerically, typically by using an iterative procedure, to obtain values of $Y_1,\ldots,Y_m$ corresponding to values of $X_1,\ldots,X_N$. This form of model can be handled by the provisions of JCGM 102:2011, which covers a multivariate generalization of the form of the law of propagation of uncertainty given in JCGM 100:2008 and a Monte Carlo method.

NOTE It is assumed that the implicit multivariate measurement model (35) can be solved uniquely for values of $Y_1,\ldots,Y_m$ given values of $X_1,\ldots,X_N$ close to their estimates.

EXAMPLE *Pressures generated by a pressure balance accounting for correlation*

A generalization of the example of the pressure balance in example 2 in 7.2.1 is as follows. For $i = 1,\ldots,m$, let p_i denote the generated pressure for applied mass m_i and temperature t_i, with A_0, λ, α, ρ_a, ρ_w and g as before. An estimate of each p_i is obtained by solving an equation of the form (3) given estimates of $A_0,\lambda,\alpha,t_i,m_i,\rho_a,\rho_w$ and g. However, the quantities representing the generated pressures are not independent because they all depend on the quantities A_0, λ, α, ρ_a, ρ_w and g. The measurement is described by a multivariate measurement model in which $A_0,\lambda,\alpha,t_1,\ldots,t_m,m_1,\ldots,m_m,\rho_a,\rho_w,g$ are $2m+6$ input quantities, $p_1,\ldots,p_m$ are m output quantities, and the measurement model takes the form

$$h_i(Y_1,\ldots,Y_m,X_1,\ldots,X_N) = A_0 p_i(1+\lambda p_i)[1+\alpha(t_i - t_{\text{ref}})] - m_i\left(1 - \frac{\rho_a}{\rho_w}\right)g = 0,\ i = 1,\ldots,m,$$

where t_{ref} is a reference temperature, 20 °C, say [113].

13.7 Measurement models involving complex-valued quantities

A complex measurement model involves at least one complex-valued quantity. Complex models can also be classified as univariate and explicit [of the form (1)], multivariate and explicit [of the form (31)], univariate and implicit [of the form (33)], and multivariate and implicit [of the form (35)]. For guidance on working with complex numbers in the context of measurement models see JCGM 102:2011, 6.4.

EXAMPLE *Reflection coefficient measured by a calibrated microwave reflectometer*

The (complex) reflection coefficient Γ measured by a calibrated microwave reflectometer, such as an automatic network analyser (ANA), is given by

$$\Gamma = \frac{aW + b}{cW + 1}, \tag{36}$$

where W is the (complex-valued) quantity representing the uncorrected reflection coefficient and a, b and c are (complex-valued) calibration coefficients characterizing the reflectometer [67, 103, 154].

There are four (complex-valued) input quantities a, b, c, W and a single (complex-valued) output quantity Γ. Expression (36) defines an explicit univariate complex measurement model, the right-hand side constituting the measurement function.

A Glossary of principal symbols

The following principal symbols are used throughout the document:

c	sensitivity coefficient (first partial derivative)
f	measurement model (explicit form)
h	measurement model (implicit form)
m	number of output quantities
N	number of input quantities
Q	input quantity, used to calculate another input quantity (X)
s	standard deviation
u	standard uncertainty
X	input quantity
x	estimate of an input quantity
Y	output quantity
y	estimate of an output quantity

Other symbols, particularly those appearing in the examples, are introduced as they occur.

B Modelling of dynamic measurements by linear time-invariant systems

B.1 General

B.1.1 Dynamic measurements are measurements in which the measurand and at least one of the input quantities are dynamic quantities, that is, are time-dependent, and the behaviour of the measuring system depends on its dynamic characteristics. An important aspect is that the relation between dynamic quantities is usually described by methods from signal processing [131] and systems theory [112] rather than by an algebraic equation. By means of transforming the continuous-time models to discrete time, some models of dynamic measurements can, however, be cast as multivariate models that can be handled with JCGM 102:2011. This clause provides introductory material for continuous and discrete time dynamic models and refers to relevant literature resources where applicable.

B.1.2 Dynamic measurements can be found in many fields of metrology, such as acceleration measurement [116], dynamic force measurement [105], dynamic pressure measurement [49], hydrophone measurements [161] or ultra-fast oscilloscope measurements [74]. In all these applications, the relation between the input and the output of the measuring system is modelled by a linear time-invariant (LTI) system with linearity and time-invariance as defined in B.2.2 and B.2.3, respectively. In a sense, LTI systems can be seen as the simplest extension of the measurement model of JCGM 100:2008 to dynamic input and output quantities. For general non-linear dynamic systems, see, for instance, [148].

B.2 Continuous-time models

B.2.1 The relation between the input signal $Y(t)$ and output signal $X(t)$ of a dynamic system is denoted mathematically as

$$X(t) = \mathcal{H}[Y(t)],$$

where $\mathcal{H}$ denotes the model of the dynamic system.

B.2.2 A dynamic system model is called a *linear system* if it is linear in its dynamic inputs. That is, for dynamic quantities $Y_1(t)$ and $Y_2(t)$ with real-valued scaling factors c_1 and c_2 it holds that

$$\mathcal{H}[c_1 Y_1(t) + c_2 Y_2(t)] = c_1 \mathcal{H}[Y_1(t)] + c_2 \mathcal{H}[Y_2(t)].$$

The linearity of the system in the dynamic quantities should not be confused with linearity in the non-dynamic system parameters.

B.2.3 A system is called *time-invariant* when the dynamic system model does not change with time. That is, provided that

$$\mathcal{H}[Y(t)] = X(t),$$

a time shift in $Y(t)$ results in the same time shift in $X(t)$:

$$\mathcal{H}[Y(t - t_0)] = X(t - t_0).$$

B.2.4 For LTI systems the relation between the system input $Y(t)$ and system output $X(t)$ is given by the convolution equation [131]

$$X(t) = (H * Y)(t) = \int_{-\infty}^{\infty} H(t-\tau)Y(\tau)\,\mathrm{d}\tau,$$

where $H(t)$ denotes the impulse response of the dynamic system $\mathcal{H}(\cdot)$.

B.2.5 Alternatively, the relation between system input and system output can be modelled by a linear state-space system model with system matrices $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ and $\boldsymbol{D}$:

$$\frac{\mathrm{d}Z(t)}{\mathrm{d}t} = A\boldsymbol{Z}(t) + BY(t),$$
$$\boldsymbol{X}(t) = \boldsymbol{C}Z(t) + \boldsymbol{D}\boldsymbol{Y}(t).$$

B.2.6 Mathematically, an LTI system $\mathcal{H}$ can be represented in different equivalent forms (for instance, [112, 131]). The basic form is the system's transfer function $H(s)$ in the Laplace domain with complex-valued s from which all other representations can be derived. For instance, the system's representation in the frequency domain by its frequency response $H(\mathrm{j}\omega)$ is obtained by setting $s = \mathrm{j}\omega$ in the transfer function. For its representation in the time domain by its impulse response $H(t)$, the inverse Laplace transform is applied to the transfer function.

NOTE 1 Consequently, the result of a calibration of an LTI system can be a transfer function, frequency response or impulse response. The result can be parametric or non-parametric.

NOTE 2 The imaginary unit i as used elsewhere in this document is usually denoted by j in signal analysis and related fields.

B.3 Discrete-time models

B.3.1 In this subclause it is assumed that an analogue-to-digital conversion of the dynamic system output $X(t)$ results in an equidistant discrete-time dynamic quantity $X = (X(t_1), \ldots, X(t_M))^\top$. The corresponding dynamic measurand is the discrete-time dynamic quantity $\boldsymbol{Y} = (Y(t_1), \ldots, Y(t_M))^\top$.

B.3.2 An implicit measurement model in the time domain is given by

$$X(t_j) - (H * Y)(t_j) = 0, \qquad j = 1, \ldots, M,$$

where H denotes the dynamic system model's impulse response.

An implicit measurement model in the frequency domain is given by

$$X(f_j) - H(f_j)Y(f_j) = 0, \qquad j = 1, \ldots, M, \tag{37}$$

where H in this context denotes the dynamic system model's frequency response.

B.3.3 An explicit model in the time domain can be derived by constructing a digital deconvolution filter designed such that the input of the filter is the observed system output X and the filter output is an approximation to the dynamic measurand $\boldsymbol{Y}$ [56]. Therefore, the filter's frequency response equals the reciprocal of the system's frequency response up to a chosen frequency. Above that frequency the filter resembles a low-pass behaviour, attenuating high-frequency noise. The explicit measurement model is

$$Y(nT_s) = \sum_{r=0}^{N_b} B_r X\left((n-r)T_s\right) - \sum_{r=1}^{N_a} A_r Y\left((n-r)T_s\right) + \Delta(nT_s), \tag{38}$$

where T_s denotes the duration of the sampling interval. The filter coefficients $\boldsymbol{A} = (A_1, \ldots, A_{N_a})^\top$ and $B = (B_0, \ldots, B_{N_b})^\top$ are determined from knowledge about the system model using, for instance, methods such as least squares adjustment [56]. The additional term $\Delta(nT_s)$ on the right-hand side of expression (38) denotes the correction for a time-dependent error caused by the deconvolution filter used [59]. For the propagation of uncertainty for such measurement models, the methodology of JCGM 102:2011 can be applied [57, 64, 115].

EXAMPLE For an LTI measuring system calibrated in terms of its frequency response $H(f_r)$ for frequencies $f_1, \ldots, f_M$, an Lth order finite-impulse-response (FIR) filter with filter coefficients $\boldsymbol{A} = (A_0, \ldots, A_L)^\top$ is designed such that with radial frequency $\omega_r = 2\pi f_r$ its frequency response $F_{\boldsymbol{A}}(e^{j\omega_r/T_s})$ satisfies [64]

$$H(j\omega_r)F_{\boldsymbol{A}}(e^{j\omega_r/T_s}) \approx 1.$$

for all $\omega_r \leq \bar{\omega}_{\text{up}}$ and zero beyond that frequency. Thus, the digital filter approximates the inverse of the system model in the frequency region $[0, \bar{\omega}_{up}]$ and attenuates frequency components for larger frequencies. The attenuation of high-frequency components is necessary in order to avoid otherwise strong noise amplification. Such attenuation can be achieved, for instance, by designing an FIR filter that mimics the inverse of the system model and which is then applied in cascade with an FIR low-pass filter. The measurement model is

$$Y(t_n) = \sum_{r=0}^{L} A_r X(t_{n-r}).$$

With $\boldsymbol{a} = (a_0, \ldots, a_L)^\top$ denoting the estimate of the FIR filter coefficients with associated covariance matrix $\boldsymbol{U}_{\boldsymbol{a}}$, the standard uncertainty $u(y_n)$ associated with the estimate y_n of the measurand at time instant t_n is given by

$$u^2(y_n) = \boldsymbol{a}^\top \boldsymbol{U}_{\boldsymbol{x}(n)} \boldsymbol{a} + \boldsymbol{x}^\top(n) \boldsymbol{U}_{\boldsymbol{a}} \boldsymbol{x}(n) + \text{Tr}(\boldsymbol{U}_{\boldsymbol{x}(n)} \boldsymbol{U}_{\boldsymbol{a}}),$$

where $\boldsymbol{U}_{\boldsymbol{x}(n)}$ denotes the covariance matrix associated with $\boldsymbol{x}(n) = (x(t_n), \ldots, x(t_{n-M}))^\top$ and Tr() denotes the trace of a matrix.

B.3.4 An explicit model in the frequency domain can be derived by transforming the implicit model (37) and multiplying by the frequency response of a chosen low-pass filter $H_L(f)$ for the attenuation of high-frequency noise:

$$X(f) = \frac{Y(f)}{H(f)} H_L(f),$$

or by carrying out a Tikhonov regularization approach [157]. For the propagation of uncertainty for such measurement models, the methodology of JCGM 102:2011 can be applied [58, 74].

EXAMPLE For the characterization of medical ultrasound devices, definite ultrasound pulse shapes are measured using hydrophones. Due to the imperfect dynamic behaviour of most hydrophones, a correction for the influence of the measuring device has to be made [161]. For instance, based on the calibrated frequency response of the hydrophone, a deconvolution in the frequency domain can be carried out resulting in the measurement model

$$Y(f_r) = \frac{X(f_r)}{H(f_r)} H_L(f_r) = \frac{X(f_r)}{\widetilde{H}(f_r)}, \qquad r = 1, \ldots, M, \tag{39}$$

with $X(f_r)$ and $Y(f_r)$ at frequencies $f_1, \ldots, f_M$ denoting the discrete Fourier transform of $\boldsymbol{X} = (X_1, \ldots, X_M)^\top$ and $\boldsymbol{Y} = (Y_1, \ldots, Y_M)^\top$, respectively, and $H_L(f_r)$ the frequency response of a low-pass filter chosen for noise attenuation. $Y(f_r)$, $X(f_r)$ and $\widetilde{H}(f_r)$ are complex-valued quantities. With $\Re$ and $\Im$ denoting real and imaginary parts, respectively, the measurement model (39) can be written as

$$Y(f_r) = \frac{[\Re_X(f_r)\Re_{\widetilde{H}}(f_r) + \Im_X(f_r)\Im_{\widetilde{H}}(f_r)] + \mathrm{j}[-\Re_X(f_r)\Im_{\widetilde{H}}(f_r) + \Im_X(f_r)\Re_{\widetilde{H}}(f_r)]}{[\Re_{\widetilde{H}}(f_r)]^2 + [\Im_{\widetilde{H}}(f_r)]^2}, \qquad r = 1, \ldots, M.$$

The measurement model is thus multivariate and JCGM 102:2011 can be applied for the propagation of uncertainty [58]. Figure 10 shows the resulting estimate $\boldsymbol{y} = (y_1, \ldots, y_M)^\top$ and its associated point-wise standard uncertainties $u(y_1), \ldots, u(y_M)$ in the time domain after application of the inverse discrete Fourier transform with propagation of uncertainty. Note that as for equation (38) account has to be taken of the regularization error [59].

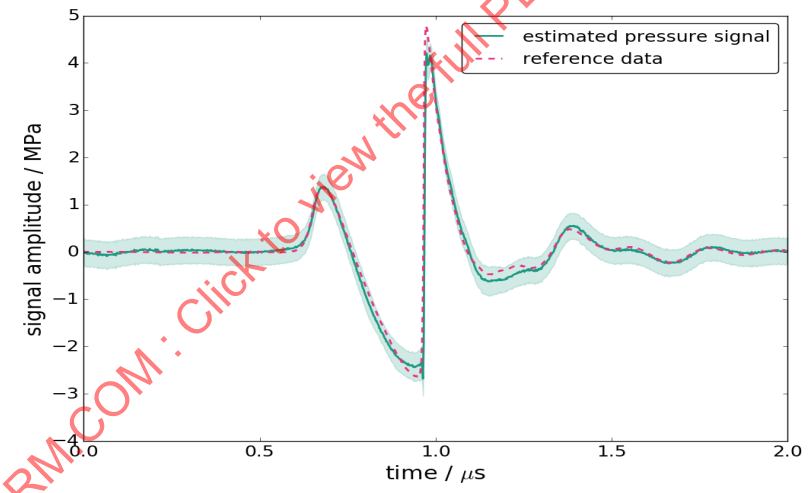

Figure 10: Estimate and associated point-wise standard uncertainties for the hydrophone pressure input signal (solid line) and the values obtained with a reference instrument (broken line)

C Modelling random variation

C.1 General

C.1.1 Random variation

Variation is present in most sets of observations. Some variation arises from effects that can be measured and a correction made. Some variation, however, is unpredictable and cannot be corrected. Variation sometimes arises only from the accumulation of small fluctuations in measurement conditions, test material or other phenomena. In other cases, such as radiation counting, it can be a fundamental feature of the system under consideration. This unpredictable variation is usually regarded, and modelled, as random variation. Because of the presence of random variation, all complete measurement models should include at least some terms representing random variation.

C.1.2 Considerations in modelling random variation

Considerations in modelling random variation include:

— Random variation in observations of an input quantity necessarily leads to random variation in corresponding values of the measurand. Random variation can therefore often be modelled as effects associated with individual input quantities or as an effect associated with the measurand, leading to an important choice in modelling;
— Random variation in repeated observations will be influenced by any random variation that occurs in the input quantities during the particular set of measurements. However, not all input quantities necessarily vary during a given set of measurements; some may even be constant. The variation in a particular set of observations will therefore usually include the variation in only some of the input quantities. Considerable care should therefore be taken to avoid 'double-counting' or omission of some random effects;
— Random variation often varies in scale depending on the time period over which variation is observed. Models may need to provide for multiple random effects in order to model the effect of variation over different time scales;
— Random effects need not be independent, making it necessary to allow for covariance; and
— The number of instances of each random effect depends on the particular experimental design used for the measurement. A complete model for random variation should therefore be specific to a particular experimental design.

C.2 Including random variation in a measurement model

C.2.1 Options for including random variation

The simplest method of including random variation explicity in a measurement model is to introduce a term representing the effect of each source of random variation or embodying a combination of sources of random variation. There are two broad options for including random variation (see C.1.2), which can be regarded as extremes of a continuum. One option is to associate random variation with each individual input quantity (see C.2.2). The other extreme is to associate all random variation with the measurand (see C.2.3).

C.2.2 Random variation associated with an existing input quantity

When random variation is associated with an existing input quantity in the model, it often suffices to include the random effect only in the probability distribution for that quantity. This extension can be as simple as increasing the variance associated with an existing probability distribution; the new random effect is then treated as an additional contribution to the uncertainty in a particular input quantity. More generally, a new random effect can be included as a new term to the model. For instance, a simple approach is to include an additive term with an expected value of zero and specify an appropriate probability distribution and its parameters. The examples in this subclause illustrate these two options, starting with the explicit identification of a new random effect.

EXAMPLE 1 *Adding a random effect to a simple calculation of concentration*

A simple model for the concentration x of a contaminant in drinking water is

$$x = \frac{I_x}{I_{\text{ref}}} x_{\text{ref}}, \tag{40}$$

where I_x is the chromatographic peak area observed for a known volume of the test material, x_{ref} is the (known) concentration of a reference solution and I_{ref} is the chromatographic peak area for the same volume of the reference solution. To allow explicitly for random variation in, say, the observed peak area for the test item, one simple extension to the model might be:

$$x = \frac{(I_x + \varepsilon_x)}{I_{\text{ref}}} x_{\text{ref}} \tag{41}$$

with

$$\varepsilon_x \sim \text{N}(0, \sigma_I^2). \tag{42}$$

This model indicates that random variation in the observed peak area is modelled as an additive effect on peak area with a normal distribution, centred at zero and with variance σ_I^2. Note that the complete statistical model, (41) together with (42), includes a statement about the probability distribution.

EXAMPLE 2 *Simplified inclusion of a random effect for an existing input quantity*

Example 1 in this subclause can be simplified. Instead of including the new, zero-centred term ε_x, the probability distribution can be associated directly with the peak area I_x. The model then becomes

$$x = \frac{I_x}{I_{\text{ref}}} x_{\text{ref}}$$

(which appears identical to the unmodified model) with

$$I_x \sim \text{N}(\mu_I, \sigma_I^2),$$

where μ_I is the (true) mean value for the peak area. In simple cases the observed value can be inserted into the calculation in place of the true value μ_I, making this representation essentially equivalent to the statistical model (41) and (42).

The model has been simplified for clarity. In practice, the true value for a peak area is of little interest. A more complete statistical model would show that the peak area I_x is proportional to the true concentration μ_x of interest. If the model is rewritten in this way, it can be seen as an implicit statistical model requiring a solution for μ_x, from which the model (40) follows to a good approximation when the variances are small.

C.2.3 Random variation as an effect associated with the measurand

C.2.3.1 It can be convenient to treat random variation as a direct effect on observed values, instead of as a set of effects on individual input quantities. Sometimes, proceeding in this way is a matter of modelling convenience, for instance, because it is more straightforward to study the dispersion of the observations than to examine each input quantity separately. In other circumstances, the dispersion of the observations is found to be greater than can be accounted for by the known variation in input quantities, and this excess variation needs to be taken into account in evaluating uncertainty. Both situations can be addressed by associating one or more sources of random variation with the measurand, rather than with an individual input quantity.

EXAMPLE *The effect of random variation in ambient laboratory temperature on a measured value*

Ambient laboratory temperature can influence a measurement in a number of ways including the introduction of a random effect. This effect could be due to random fluctuations in the laboratory temperature affecting the thermal properties of electrical detectors (noise floor, etc.) used by a measuring system. Such a random effect cannot easily be represented mathematically. However, laboratory experiments can be used to evaluate the uncertainty due to these random ambient temperature variations. For instance, if several repeated determinations are made over a suitable period of time (during which the laboratory temperature fluctuates in its usual way), then the set of determinations can be analyzed statistically and included as a component to the overall uncertainty in the measurement. Further, approximations to sensitivity coefficients can be determined empirically (see JCGM 100:2008, 5.1.4).

C.2.3.2 The simplest extension to a model is, as before, to include an additive effect with mean zero. This extension is most useful where the dispersion of observations is largely independent of the value of the measurand over the range of interest.

EXAMPLE *Random variation associated with an existing input quantity: alternative representation*

In example 1 in C.2.2, an alternative representation is

$$x = \frac{I_x}{I_{\text{ref}}} x_{\text{ref}} + \varepsilon_x$$

with

$$\varepsilon_x \sim \mathrm{N}(0, \sigma_x^2)$$

in which σ_x includes all sources of random variation under a particular set of conditions of measurement.

C.2.3.3 When the standard deviation is known to be approximately proportional to the measured value over the relevant range of values, then it can be useful to choose a multiplicative term rather than an additive term. The new term is given a mean of 1 and a standard deviation equal to the relative standard deviation of the observations. Also see 10.3.4. Treating random effects as additive or multiplicative leads to different models. The output of these models is nearly equivalent for small dispersion (for instance, relative standard deviation much less than 0.1). In general, however, different calculations are required to obtain unbiased estimates and, further, the probability distributions will be different for the two models.

EXAMPLE *Model including random effect with approximately constant relative standard deviation*

Continuing the example in C.2.3.2, a simple model that includes a random effect with approximately constant relative standard deviation is

$$x = \frac{I_x}{I_{\text{ref}}} x_{\text{ref}} F_\sigma$$

with

$$F_\sigma \sim \text{N}\left(1, (\sigma_x/\mu_x)^2\right)$$

in which, again, the dispersion includes all sources of random variation under particular conditions of measurement.

C.2.3.4 As indicated in C.1.2, when random variation is modelled as an effect associated directly with the measurand, considerable care should be taken in the treatment of uncertainties associated with individual input quantities to ensure that effects are neither counted twice nor omitted. Particular issues that should be considered include:

— When an input quantity, say A, is measured once (including a single average of several observations) for every measured value of the measurand, the term representing variation in observations of A under repeatability conditions can be associated *either* with the input quantity *or* with the output quantity, but not with both; and
— When the random variation associated with the output quantity relates to reproducibility conditions, it is important to be aware that some, or even all, contributions to the standard uncertainties associated with individual input quantities can appear in the observed reproducibility standard deviation. For instance, variation in ambient temperature (an effect that can be important in evaluating the uncertainty associated with measured volumes, for instance) will appear in a reproducibility standard deviation over sufficient time. In such circumstances, those contributions should not be included in the uncertainty associated with those input quantities as well as appearing implicitly in a random effect associated with the observation.

NOTE The fact that reproducibility conditions can lead to random variation that includes most effects on the measurand value underlies the concept of the 'top-down' approach to uncertainty evaluation, considered further in C.5.2.

C.3 Multiple sources of random variation

Frequently, more than one cause of random variation can be identified; for instance, variation in environmental conditions over time and change in (say) operator might both be modelled as random effects. It is often useful to separate short-term variation (typically within a measurement run) from longer-term variation. Such separation ensures that the effect of averaging is properly reflected in the uncertainty evaluation. These separately identified sources of variation are common features of statistical models, which are straightforwardly extended to include the effects (even multiple sources) of random variation as further terms in the model.

EXAMPLE *Separating 'within-run' and 'between-run' effects*

It can be useful to recognize that observations obtained over a short period (a single measurement run or day) typically vary much less than observations obtained from different runs. A simple extension to the model allows for a second such 'random effect':

$$X = \mu + \delta_{\text{run}} + \epsilon, \tag{43}$$

where δ_{run} is a 'between-run' term, with its own probability distribution, having mean (usually zero) and variance that can be characterized by statistical methods such as analysis of variance.

NOTE 1 Often, a model such as (43) can be considered 'hierarchical' in the sense that all measured values within one run are subject to the same 'run' error; the error term ϵ is 'nested' within the run. Although it is perfectly possible to have other structures for random effects (for instance, the same two instruments could be used to measure the same material on two different days, forming a 'crossed' design), uncertainties (and associated degrees of freedom [91, definition 2.54]) can be simpler to handle in the case of a hierarchical model.

NOTE 2 ISO 21748:2016 [90] gives some guidance on the use of hierarchical (nested) models for uncertainty evaluation; some simple 'crossed' designs are addressed by ISO/TS 17503 [98]. Both give information on the associated degrees of freedom.

C.4 Asymmetrically distributed effects

Additional care should be taken where a probability distribution for a random effect is asymmetric or where the measurement model is non-linear in a quantity associated with an otherwise symmetric random effect. In these cases, random variation often leads to bias in the measured value and the very simple model extension given in clause C.3 can be inappropriate. It is then much better to express the model as a statistical model for the data-generating process (see clause 11) and consider how best to obtain estimates of the parameters of the probability distribution that represent the desired measurement result. Sometimes the bias resulting from an asymmetric distribution can conveniently be addressed by including a correction (with associated uncertainty) based on the dispersion of the random effect; the principle is identical to the treatment of an asymmetric effect given in JCGM 100:2008, F.2.4.4. A more general alternative is to solve the resulting implicit model directly, using (for instance) maximum-likelihood [133] or Bayesian methods [71].

C.5 Use of reproducibility studies

C.5.1 A particular case of models that combine essentially all random variation effects into a small number of random effects associated with the measurand is the set of models used for examining the repeatability and reproducibility of measurement procedures. These models focus on random variation in the measurand without separating the effects associated with individual input quantities.

C.5.2 One application of models of this type has been described as the 'top-down approach', a concept first introduced in a paper by the Analytical Methods Committee of the UK Royal Society of Chemistry [3] (also see E.1.1). The top-down approach is based on the principle that the reproducibility standard deviation obtained in a collaborative study is a valid basis for measurement uncertainty evaluation [90, annex A.2.1]. See C.5.3.

This approach is widely used in measurement and testing, where it is accepted that a suitably representative standard uncertainty can be obtained by that approach. Measurements made in different laboratories, at different times, with different instruments and personnel are carried out under a well-defined reproducibility condition of measurement. This approach is sometimes known as the 'pure top-down approach'. An instance of the use of the top-down approach in food analysis is available [119], and recommendations for relevant terminology in measurement uncertainty in chemistry have been discussed [43].

C.5.3 In the implementation of the top-down approach given in ISO 21748:2016, the basic statistical model [90,92] for a measurement is (following the nomenclature of references [90, 92, 93])

$$Y = m + B + E,$$

where Y is the quantity measured by a laboratory, m is the 'level' of the measurement (or the 'general mean' of Y), B is the laboratory systematic error under repeatability conditions, assumed to be normally distributed with expectation zero and unknown standard deviation σ_{L}, and E is the random error under repeatability conditions, assumed to be normally distributed with expectation zero and unknown standard deviation σ_{W}.

NOTE 1 The value of the laboratory systematic error B is considered to be constant during any set of observations obtained under repeatability conditions, but to differ in general for observations obtained under other conditions. The variance of B is called the between-laboratory variance.

NOTE 2 The value of the random error E is considered to be different for every observation of Y. The variance of E is called the within-laboratory variance.

NOTE 3 Estimates s_{L}^2 and s_{r}^2 of, respectively, σ_{L}^2 and σ_{W}^2 are obtained in an inter-laboratory or collaborative study. s_{r}^2 is called the repeatability variance.

NOTE 4 The standard uncertainty associated with a value y of Y is then given by

$$u^2(y) = s_{\mathrm{L}}^2 + s_{\mathrm{r}}^2 \equiv s_{\mathrm{R}}^2.$$

s_{R}^2 is called the reproducibility variance.

C.5.4 In some collaborative studies the measurement model is extended to

$$Y = \mu + \delta + B + E$$

with $\widehat{\mu}$, the certified value of the quantity being measured, constituting an estimate of μ with associated standard uncertainty $u(\widehat{\mu})$, and $\widehat{\delta}$, the bias of the measurement method, constituting an estimate of δ with associated standard uncertainty $u(\widehat{\delta})$.

NOTE The standard uncertainty associated with a value y of Y is then given by

$$u^2(y) = u^2(\widehat{\mu}) + u^2(\widehat{\delta}) + s_{\mathrm{R}}^2.$$

C.5.5 In practice s_{R} and $u(\widehat{\delta})$ might not describe the variability of the measurand arising from all effects that influence the measurement. Some important effects might be missing from the collaborative study. If X_i, $i = 1, \ldots, n$, with estimates x_i and associated standard