Joint Committee for Guides in Metrology

JCGM
104
2007-07-27
Draft

Evaluation of measurement data — An introduction to the "Guide to the expression of uncertainty in measurement" and related documents

Évaluation des données de mesure – Une introduction au "Guide pour l'expression de l'incertitude de mesure" et les documents relatifs

Contents

Page

Foreword . . . v

Introduction . . . vi

1 Scope . . . 1

2 Normative references . . . 1

3 What is measurement uncertainty? . . . 1

4 Concepts and basic principles . . . 3

5 Stages of uncertainty evaluation . . . 5

6 Formulation . . . 5

7 The calculation (propagation and summarizing) stage of uncertainty evaluation . . . 6
- **7.1** General . . . 6
- **7.2** The GUM uncertainty framework . . . 7
- **7.3** Analytic methods . . . 7
- **7.4** Monte Carlo method . . . 7
- **7.5** Models with any number of output quantities . . . 9

8 The role of measurement uncertainty in deciding conformance to specified requirements . . . 11

9 Applications of the least-squares method 12

Bibliography . . . 13

Alphabetical index . . . 14

Foreword

In 1997 a Joint Committee for Guides in Metrology (JCGM), chaired by the Director of the BIPM, was created by the seven international organizations that had originally in 1993 prepared the "Guide to the expression of uncertainty in measurement" (GUM) and the "International vocabulary of basic and general terms in metrology" (VIM). 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 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). A further organization joined these seven international organizations, namely, the International Laboratory Accreditation Cooperation (ILAC).

JCGM has two Working Groups. Working Group 1, "Expression of uncertainty in measurement", has the task to promote the use of the GUM and to prepare Supplements and other documents for its broad application. Working Group 2, "Working Group on International vocabulary of basic and general terms in metrology (VIM)", has the task to revise and promote the use of the VIM. For further information on the activity of the JCGM, see www.bipm.org.

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

Drafting note

This version of *Evaluation of measurement data — An introduction to the "Guide to the expression of uncertainty in measurement" and related documents* is not intended to be an advanced draft. Rather, it is hoped that the Member Organizations and other interested parties will be prepared to make input and provide constructive criticism, which Working Group 1 of the JCGM would be pleased to take into consideration in preparing a new draft.

The final version of this introductory document will be web-based, containing hyperlinks to appropriate clauses and sub-clauses of the GUM and related documents, including GUM Supplements, and the VIM. Moreover, it is intended that the index will, in addition to making reference to appropriate clauses and sub-clauses of this document, provide links to relevant information, including definitions, in other documents (at the moment specimen links are included to GUM Supplement 1). It will also provide links to other relevant material, including that from other organizations, for a broader understanding.

At the moment specimen links are included to

— the websites of the JCGM and its Member Organizations,

— clauses and sub-clauses of the version of GUM Supplement 1 that has been finalized for publication (it is necessary that this document and GUM Supplement 1 are contained in the same folder or directory for these links to work), and

— a web-based version of one of the cited references [1] that is regarded as particularly valuable (suggestions for other such references are welcomed).

Internal links are included in the form of cross-references to clauses and sub-clauses, to figures, tables and equations, and to items in the bibliography.

Introduction

A measurement uncertainty statement is valuable in judging the fitness for purpose of a measured value. At the grocery store the customer would be content if, when buying a kilogram of fruit, the scales gave a value within, say, 10 grams of that given by a reliable balance. But the dimensions of parts in the gyroscopes within modern aircraft guidance systems must be measured to submicrometre accuracy for correct functioning.

Measurement uncertainty can be used to help judge the consistency of experiment and theory, of different measurement procedures, and of different laboratories. As the tolerances applied in industrial production become more demanding, the role of measurement uncertainty becomes more important when assessing conformance to these tolerances. Indeed, measurement uncertainty is a concept central to quality assurance [13].

The other titles in the series of documents prepared by the JCGM that support the GUM and extend its applicability include the expression "Evaluation of measurement data", which embraces

a) concepts and basic principles [4],

b) propagation of distributions using a Monte Carlo method [5],

c) models with any number of output quantities [6],

d) modelling [7],

e) the role of measurement uncertainty in deciding conformance to specified requirements [8], and

f) applications of the least-squares method [3].

Evaluation of measurement data — An introduction to the "Guide to the expression of uncertainty in measurement" and related documents

1 Scope

The Joint Committee for Guides in Metrology (JCGM) has prepared this introductory document to promote the sound evaluation of measurement uncertainty through the use of the GUM and to provide a prelude to the GUM Supplements and other documents JCGM is producing [3, 4, 5, 6, 7, 8].

As in the GUM, this document is primarily concerned with the expression of uncertainty in the measurement of a well-defined physical quantity — the measurand — that can be characterized by an essentially unique value.

The purpose of the GUM Supplements and the other documents is to help with the interpretation of the GUM and enhance its application. The GUM Supplements and the other documents are together intended to have a scope that is considerably broader than that of the GUM.

This document introduces (a) measurement uncertainty, (b) the GUM, and (c) the GUM Supplements and other documents that support the GUM.

This introductory document is directed at (a) calibration, testing and inspection laboratories in industry, and laboratories such as those concerned with health and the environment, (b) accreditation bodies, and (c) scientific disciplines in general. It is hoped that it will also be useful to designers, because a product specification that takes better account of inspection requirements (and the associated measurement) can result in less stringent manufacturing requirements. It is also directed at academia, with the hope that more university departments will include modules on uncertainty evaluation in their courses. As a result, a new generation of students would be better armed to understand and provide uncertainty statements associated with measured values, and hence gain an improved appreciation of measurement.

This introductory document, the GUM, the GUM Supplements and the other documents should be used in conjunction with the "International Vocabulary of Metrology—Basic and General Concepts and Associated Terms" (VIM).

2 Normative references

The following referenced documents are indispensable for the application of this document.

BIPM, IEC, IFCC, ISO, IUPAC, IUPAP and OIML. Guide to the expression of uncertainty in measurement (GUM). ISBN 92-67-10188-9. International Organization for Standardization, Geneva, corrected and reprinted 1995.

BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP and OIML. International Vocabulary of Metrology—Basic and General Concepts and Associated Terms, VIM, 3rd Edition. International Organization for Standardization, Geneva, 2007.

3 What is measurement uncertainty?

3.1 The purpose of measurement is to provide knowledge of a quantity of interest—a measurand [VIM:2007 2.3]. The measurand might be the volume of a vessel, the electromagnetic radiation from a domestic appliance, or the concentration of lead in water.

3.2 No measurement is perfect in that, when a quantity is measured, the outcome depends on the measurement process comprising the instrument used, the measurement procedure, the skill of the operator, the environment, and other effects [1]. Even if the quantity were to be measured several times, in the same way and in the same circumstances, a different value each time would in general be obtained.

NOTE This statement assumes that the indication of the measurement system has sufficient resolution to distinguish between the values.

3.3 The scatter of the measured values would relate to the quality of the measurement. The average of the set of measured values would provide an estimate of the quantity that generally would be more reliable than each individual measured value. The scatter and the number of measured values would provide information relating to the quality of the average value as an estimate of the quantity. It would not furnish all the information of this type, however.

3.4 The measuring system [VIM:2007 3.2] may provide values that are not scattered about the value of the quantity, but about some value offset from it. Take the domestic bathroom scales. If they are not set to show zero when there is nobody on the scales, the indicated mass of a person would not be correct. No matter how many times the person's mass were taken, the effect of this offset would be inherently present in the average of the measured values.

NOTE 1 There are several ways of taking an average, but the choice made does not affect the argument.

NOTE 2 In areas such as biochemistry there is sometimes no natural counterpart of setting the scales to zero, considerable effort being required to quantify the effect.

3.5 There are thus two main effects, in this example and in general. The first is a random effect [VIM:2007 2.19] associated with the fact that when a measurement is repeated it will generally provide a value that is different from the previous value. It is random in that the next measured value cannot be predicted from previous measured values. (If a prediction were possible, allowance for the effect could be made!) The second effect is a systematic effect [VIM:2007 2.17] (an estimate of which is known as a bias [VIM:2007 2.18]) associated with the fact that all the measured values contain an offset. In general, there can be a number of contributions to each effect. Depending on the application, either the random effect or systematic effect might dominate.

3.6 The above discussion concerns the direct measurement of a particular quantity. However, the physical quantity actually measured by the instrument is rarely that required in practice. For instance, the display (indication [VIM:2007 4.1]) on the bathroom scales does not correspond to the physical measurement. The latter in that case might be the extension of a spring whose length varies according to the load (the mass of the person on the scales). The raw measured value is therefore converted into an estimate of the measurand. For a perfect (linear) spring, the conversion is straightforward, being based on the fact that the required mass is proportional to the extension of the spring. The particular relationship between the extension of the spring and the displayed mass constitutes the calibration [VIM:2007 2.39] of the scales.

3.7 A relationship such as that in subclause 3.6 constitutes a rule for converting a measured value into that required. The rule is usually known as a measurement model [VIM:2007 2.48] or simply a model. There are very many different types of measurement in practice and therefore different rules or models. Even for one particular type of measurement there may well be more than one model. A simple model (for example a proportional rule, as above) might be sufficient for everyday domestic use. Alternatively, a sophisticated model involving more complicated calculations that is capable of delivering better results might be necessary for industrial or laboratory purposes.

3.8 As stated in subclause 3.3, the indications corresponding to repeated measurement may be averaged to obtain a more reliable value. The situation is frequently more general in another way. There are often measurements of several different quantities that contribute to the measurand. Here, the concern is not repetition of measurement, but measurement of intrinsically different quantities, for example of temperature, humidity and displacement.

3.9 In addition to the model being based on a physical understanding, the model is often augmented by correction terms [VIM:2007 2.53] to account for the fact that the conditions of measurement are not exactly as stipulated. These terms contain quantities (often additional quantities) that influence the measurement. Instances are the measurement of a height, when the attitude of the measuring instrument is not exactly vertical, and the ambient temperature is different from that specified. The attitude of the instrument and the ambient temperature are not known exactly, but information concerning them is available, for example that the ambient temperature at the time of measurement differed from that stipulated by no more than 2 °C.

3.10 As well as raw data representing measured quantities, there is another form of data that is frequently needed to apply a model. Some such data relates to quantities representing physical constants, each of which is known imperfectly. Examples are material constants such as modulus of elasticity and specific heat. There is often other relevant data given in text books, calibration certificates, etc., regarded as estimates of further quantities.

3.11 The items required by a model to define a measurand are known as *input quantities* [VIM:2007 2.50]. The rule or model is often referred to as a *functional relationship* [GUM:1995 4.1], because it is the use of physical modelling of a measurement, with correction and other terms, as necessary, that enables the model to be established. The model *output quantity* [VIM:2007 2.51] is the measurand.

3.12 Formally, the output quantity, denoted by Y, about which information is required, is related functionally to input quantities, denoted by $X_1, \ldots, X_N$, about which information is available, by a model

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

where the vector of the X_i is denoted by $\boldsymbol{X}$.

NOTE 1 A more general form of model can be expressed as

$$h(Y, X_1, \ldots, X_N) = 0. \qquad (2)$$

In this case, a value of Y is given by solving the model given values of $X_1, \ldots, X_N$, whereas expression (1) constitutes a formula that can be evaluated given values of $X_1, \ldots, X_N$.

NOTE 2 Model (1) is known as *explicit* and model (2) as *implicit*.

3.13 Consider estimates $x_1, \ldots, x_N$, respectively, of $X_1, \ldots, X_N$, obtained from the analysis of measurement data, certificates and reports, manufacturers' specifications, and so on. These estimates are particular values of $X_1, \ldots, X_N$ regarded as random variables. These random variables are characterized by probability distributions that describe the relative likelihood of their values lying in different intervals. These probability distributions are deduced from available knowledge concerning $X_1, \ldots, X_N$.

3.14 For each X_i, the corresponding probability distribution has the property that the estimate x_i of X_i is the expectation [GUMS1:2007 3.6] (mean) of X_i. Moreover, the uncertainty of measurement, or, more precisely, standard (measurement) uncertainty [VIM:2007 2.30], which is given the symbol $u(x_i)$, is defined as the standard deviation [GUMS1:2007 3.8] of X_i. This standard uncertainty is said to be *associated* with the (corresponding) estimate.

NOTE 1 The rationale for using the expectation of X_i is that x_i is the best estimate of X_i in the sense that the standard uncertainty associated with any other estimate is larger than $u(x_i)$, the standard deviation of X_i.

NOTE 2 For a quantity characterized by an asymmetric probability distribution, the expectation is generally not equal to the mode, and so the estimate x_i of X_i is not the most probable value, although it has the smallest possible associated standard uncertainty.

3.15 The use of available knowledge to establish a probability distribution to characterize each quantity of interest applies to the X_i and also to Y. In the latter case, the characterizing probability distribution for Y is determined by the functional relationship (1) together with the probability distributions for the X_i. The determination of the probability distribution for Y from this information is known as the *propagation of distributions* [GUMS1:2007 5.2].

3.16 Other (prior) knowledge about Y, or about any other non-observable quantities that appear in the model, if available, can also be taken into account. Returning again to the domestic bathroom scales, the fact that the person's mass must be positive, or that it is the mass of a person, rather than that of a motor car, that is being measured, both constitute prior knowledge about the possible values of the measurand in this example. Bayes' theorem provides the means to combine prior information about Y, expressed as a probability distribution for Y, with the information that the measurement data provides about Y [2, 11].

4 Concepts and basic principles

4.1 A document [4] provides a brief introduction to the fundamental concepts and principles of probability theory that underlie the approach to evaluation and expression of measurement uncertainty.

4.2 Uncertainty of measurement is defined [VIM:2007 2.26] as

> *non-negative parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used*

This definition is consistent with the considerations of subclauses 3.13–3.16.

4.3 Two forms for a probability distribution [GUMS1:2007 3.1] are used in uncertainty evaluation: the *probability density function* (PDF) [GUMS1:2007 3.3] and the *distribution function* [GUMS1:2007 3.2].

4.4 Knowledge of each input quantity X_i is often summarized by x_i and $u(x_i)$. Some of the input quantities may be related to each other, in which case the summary information will also include covariances (subclause 7.5.3). $\mathrm{cov}(x_i, x_j)$ denotes the covariance associated with x_i and x_j, a measure of the strength of the relationship between X_i and X_j. If X_i and X_j are independent, $\mathrm{cov}(x_i, x_j) = 0$.

4.5 The *evaluation of measurement data*, in the context of the model (1), is the use of knowledge available concerning the input quantities $X_1, \ldots, X_N$ to deduce properties of the output quantity Y.

4.6 In general, knowledge about an input quantity is inferred from repeated indications (Type A evaluation) [GUM:1995 4.2], or scientific judgement or knowledge concerning the possible values of the quantity (Type B evaluation) [GUM:1995 4.3].

4.7 For a Type A evaluation of uncertainty [VIM:2007 2.28], it is frequently asserted that the quantity of which the average of repeated indications (obtained independently) is a particular value follows a Gaussian distribution (figure 1, continuous curve). The assertion can be justified on the basis that, for a large number of indications, the central limit theorem of statistics applies [GUM:1995 G.2]. Other considerations would apply when the number of indications is small (a t-distribution can be used), or when the indications are not obtained independently.

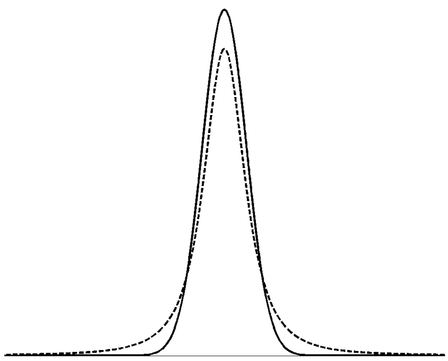

Figure 1 — A Gaussian distribution and (broken curve) a t-distribution. The particular distributions shown here have the same expectation and the same standard deviation.

4.8 For a Type B evaluation of uncertainty [VIM:2007 2.29], it is commonly the case that the only available information is that the quantity lies in a specified interval $[a, \ b]$. In such circumstances, the quantity can be characterized by a rectangular PDF [GUM:1995 4.3.7] with limits a and b (figure 2) [14]. If different information were available, a PDF consistent with that information would be used.

NOTE The principle of maximum (information) entropy (PME) provides a way to assign a PDF to a quantity that agrees with what is known (or reasonably believed to be true), but otherwise avoids any unjustified assumptions with respect to knowledge about the quantity.

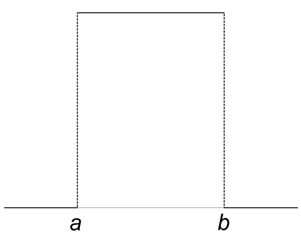

Figure 2 — Rectangular PDF with limits a and b.

4.9 Once the input quantities X_i have been characterized by appropriate PDFs, and the model has been developed, the PDF for the measurand Y is fully specified in terms of this information. In particular, the expectation of Y, as described by its PDF, can be used as the estimate y of Y, and the standard deviation of Y as the standard uncertainty $u(y)$ associated with y. Often a coverage interval for Y is also required. Such an interval is one in which a specified fraction of the possible values of Y can be expected to lie. A coverage interval can also be deduced from the PDF for Y.

4.10 As a simple example, figure 3 depicts the model $Y = X_1 + X_2$ in the case where X_1 and X_2 are each characterized by a (different) rectangular PDF. Y has a trapezoidal PDF in this case.

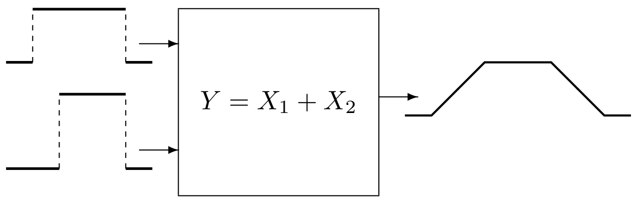

Figure 3 — Illustration of an additive model with two input quantities X_1 and X_2, each of which is characterized by a (different) rectangular PDF. The PDF for the measurand Y is a symmetric trapezoidal PDF.

4.11 A coverage interval [VIM:2007 2.36] is an interval containing the value of a quantity with a stated coverage probability [VIM:2007 2.37]. It is not unique, two important coverage intervals being

a) the *probabilistically symmetric coverage interval* [GUMS1:2007 3.15], namely the coverage interval for a quantity such that the probability that the quantity is less than the smallest value in the interval is equal to the probability that the quantity is greater than the largest value in the interval, and

b) the *shortest coverage interval* [GUMS1:2007 3.16], namely the coverage interval for a quantity with shortest length among all coverage intervals for that quantity having the same coverage probability.

4.12 Figure 4 shows the chi-squared distribution with two degrees of freedom (exponentially decreasing curve) with the endpoints of the shortest (continuous vertical lines) and those of the probabilistically symmetric (broken vertical lines) 80 % coverage intervals for a quantity characterized by this distribution. The distribution is asymmetric and the two coverage intervals are different (most notably their right-hand endpoints), with the length of the former interval about three quarters of that of the latter. The shortest coverage interval has its left-hand endpoint at zero, the smallest possible value for the quantity.

NOTE 1 The probability distribution in figure 4 would be the PDF for Y for the model $Y = X_1^2 + X_2^2$, where X_1 and X_2

are independent and characterized by standard Gaussian (normal) distributions.

NOTE 2 The coverage probability is chosen as 80 % for visual purposes. For higher coverage probabilities the left-hand endpoints of the two coverage intervals are almost indistinguishable to graphical resolution.

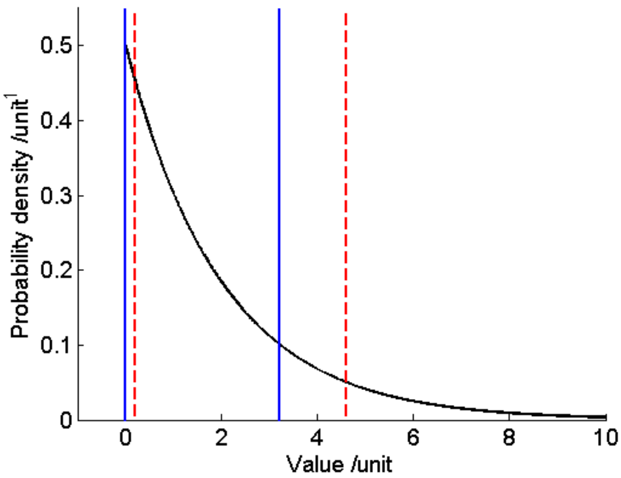

Figure 4 — The shortest (continuous vertical lines) and probabilistically symmetric (broken vertical lines) 80 % coverage intervals for a quantity characterized by a chi-squared distribution with two degrees of freedom (exponentially decreasing curve).

4.13 Sensitivity coefficients $c_1, \ldots, c_N$ describe how the estimate y of Y would be influenced by small changes in the estimates $x_1, \ldots, x_N$, respectively, of the input quantities $X_1, \ldots, X_N$. For the model (1), c_i is taken as the partial derivative of first order of f with respect to X_i evaluated at $X_1 = x_1, \ldots, X_N = x_N$. For the linear model

$$Y = c_1X_1 + \cdots + c_NX_N, \tag{3}$$

in which the X_i are independent, a change in x_i equal to $u(x_i)$ would give a change $c_iu(x_i)$ in y. This statement would generally be approximate for the model (1). The relative magnitudes of the terms $|c_i|u(x_i)$ are useful in assessing the respective contributions to $u(y)$ from each input quantity.

5 Stages of uncertainty evaluation

5.1 The main stages of uncertainty evaluation constitute formulation and calculation, the latter consisting of propagation and summarizing.

5.2 The formulation stage (clause 6) constitutes

a) defining the output quantity Y (the measurand),

b) determining the input quantities $X_1, \ldots, X_N$ on which Y depends,

c) developing a model relating Y to the X_i, and

d) on the basis of available knowledge, assigning PDFs—Gaussian, rectangular, etc.—to the X_i (or a joint PDF to those X_i that are not independent).

5.3 The calculation stage (clause 7) consists of propagating the PDFs for the X_i through the model to obtain the PDF for Y, and summarizing by using the PDF for Y to obtain

a) the expectation of Y, taken as an estimate y of Y,

b) the standard deviation of Y, taken as the standard uncertainty $u(y)$ associated with y [GUM:1995 E.3.2], and

c) a coverage interval containing Y with a specified probability (the coverage probability).

6 Formulation

6.1 The formulation stage of uncertainty evaluation involves building a mathematical model of the measurement, incorporating correction and other effects as necessary. It also involves using available knowledge to characterize the model input quantities by PDFs. Model-building aspects are considered in this clause. GUM Supplement 3 [7] provides guidance on developing and working with a model of measurement. The assignment of PDFs is considered in GUM Supplement 1 [5].

6.2 A model relating the input quantities $\boldsymbol{X}$ to the output quantity Y is initially built. In many cases, there might be more than one output quantity, $Y_1, \ldots, Y_m$, collectively denoted by $\boldsymbol{Y}$. The model is formed on physical or empirical grounds, or a hybrid model is developed. Such a model generally depends strongly on the metrology discipline, electrical, dimensional, thermal, etc. The model is then augmented by additional terms, constituting further input quantities, which describe various effects that influence the measurement. These effects may be categorized into random and systematic effects, or in other ways. Guidance is provided on handling these additional effects.

6.3 Attention is paid to discrete and continuous models, the former including algebraic models such as those addressed explicitly in the GUM, and the latter models such as finite element models used in solving partial differential equations.

6.4 The model is categorised according to whether (a) the quantities involved are real or complex, (b) the model is explicit or implicit (subclauses 3.12, note 2 and 7.5.5), and (c) $\boldsymbol{Y}$ is univariate or multivariate (subclause 7.5.1). With regard to category (a), complex quantities occur especially in electrical metrology, but also in acoustical and optical metrology and elsewhere. In category (b), the model is classified as explicit if $\boldsymbol{Y}$ can be expressed directly as formulae in terms of $\boldsymbol{X}$, and implicit otherwise, that is as an equation to be solved for $\boldsymbol{Y}$ in terms of $\boldsymbol{X}$ (subclause 7.5.1). Regarding category (c), a univariate model, the model primarily treated in the GUM, is a model having a single (scalar) output quantity. A multivariate model has any number of output quantities, that is a vector quantity.

6.5 Examples from a range of metrology disciplines illustrate the various aspects of GUM Supplement 3 [7]. Guidance on numerical analysis issues that arise, for example appropriate matrix computation, in treating these examples is given. This guidance includes the use of model re-parametrization, viz., the use of changes of variables so that all or some of the resulting quantities are uncorrelated or only weakly correlated.

6.6 Multistage models, where the output quantities from previous stages become the input quantities to subsequent stages, are also treated. Arguably the commonest example of a multistage model is a calibration curve, in which the following operations are typically carried out (see figure 5):

a) The determination of a calibration curve by modelling measured stimulus and response values by a suitable mathematical form. The uncertainties associated with the measured data values give rise to uncertainties associated with the estimates of the coefficients of the curve;

b) The evaluation of the calibration curve at subsequently measured response values to provide stimulus value predictions. The uncertainties associated with the estimates of the coefficients of the curve, together with the uncertainties associated with the new measured response values, give rise to uncertainties associated with the stimulus value predictions.

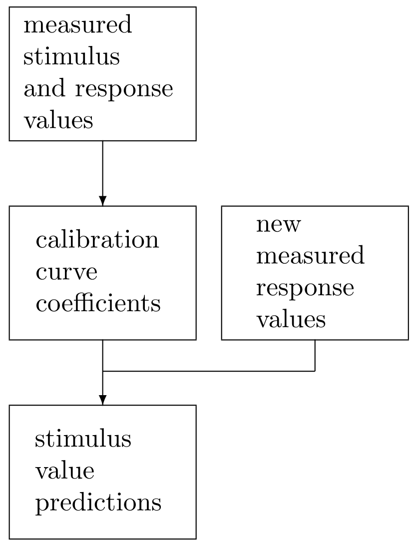

Figure 5 — A multistage model for calibration curves. The measurement data is used to establish the coefficients of the calibration curve. The functional form for the calibration curve, containing these coefficients, together with subsequently measured response values, is used to provide stimulus value predictions from the curve corresponding to these response values.

7 The calculation (propagation and summarizing) stage of uncertainty evaluation

7.1 General

7.1.1 The propagation stage of uncertainty evaluation is known as the *propagation of distributions* [GUMS1:2007 5.2].

7.1.2 Various approaches for the propagation stage are available, including

a) the GUM uncertainty framework, constituting the application of the law of propagation of uncertainty, and the characterization of Y by a Gaussian or t-distribution (subclause 7.2),

b) analytic methods, in which mathematical analysis is used to derive the PDF for Y (subclause 7.3), and

c) a Monte Carlo method, in which an approximation to the PDF or the distribution function for Y is established numerically by making random draws from the PDFs for the X_i (subclause 7.4).

7.1.3 In terms of the uncertainty evaluation problem defined by the model (1) and the PDFs for the input quantities X_i, approach a) in subclause 7.1.2 is generally approximate, approach b) is exact, and approach c) provides a solution with a numerical accuracy that can be controlled.

7.1.4 For any particular uncertainty evaluation problem, approach a), b) or c), or some other approach, is selected.

7.1.5 The application of approaches a) and c) to models with any number of output quantities, and models that define the output quantities implicitly in terms of the input quantities, is considered in subclause 7.5.

7.2 The GUM uncertainty framework

7.2.1 The GUM uncertainty framework [GUM:1995 5.1.2] (figure 6) uses

a) the expectations of the X_i as best estimates x_i of the X_i,

b) the standard deviations of the X_i as the standard uncertainties $u(x_i)$ associated with the x_i, and

c) the sensitivity coefficients c_i (subclause 4.13)

to form an estimate y of Y and the associated standard uncertainty $u(y)$ [GUM:1995 5.1]. Variants [GUM:1995 5.2] apply to instances where the X_i are mutually dependent (not indicated in figure 6) and the degrees of freedom relating to the $u(x_i)$ are finite. By regarding the PDF for Y as Gaussian, a coverage interval for Y corresponding to a specified coverage probability is also determined [GUM:1995 G.2]. When the degrees of freedom relating to the $u(x_i)$ are finite, an (effective) degrees of freedom relating to $u(y)$ is accordingly determined, and the PDF for Y is regarded as a scaled and shifted t-distribution.

7.2.2 There are many circumstances where the GUM uncertainty framework [GUM:1995 5.1.2] can be applied and leads to valid statements of uncertainty.

7.2.3 If the model is linear in the X_i and the PDFs for the X_i are Gaussian, the GUM uncertainty framework provides exact results. Even when these conditions do not hold, the approach can often work sufficiently well for practical purposes.

7.2.4 There are circumstances where the GUM uncertainty framework might not be satisfactory, including (a) the uncertainty contributions $|c_i|u(x_i)$ (subclause 4.13) are not of approximately the same magnitude [GUM:1995 G.2.2], (b) the PDF for Y is either asymmetric, or not a Gaussian or a scaled and shifted t-distribution, (c) the models are far from linear, and (d) the PDFs for the X_i are asymmetric, for example when dealing with the magnitudes of complex variables in electrical metrology. It can sometimes be difficult to establish in advance that the circumstances hold for the GUM uncertainty framework to apply.

7.2.5 The use of the GUM uncertainty framework becomes more difficult when it is necessary to form partial derivatives (or numerical approximations to them) of a model that is complicated, as needed by the law of propagation of uncertainty (possibly with higher-order terms) [GUM:1995 5]. A valid and sometimes more readily applicable treatment is obtained by applying a suitable Monte Carlo implementation of the propagation of distributions.

7.3 Analytic methods

7.3.1 Analytic methods to obtain the PDF for Y are preferable in that they do not introduce any approximation. However, they can be applied in relatively simple cases only. A treatment of such methods is available [10]. Instances that can be so handled for general N are linear models (3), where all X_i are Gaussian, or all are rectangular with the same semi-width.

7.3.2 Cases where $N = 1$ can often be treated analytically. Such cases are important in the context of transformation of measurement units, for example from linear to logarithmic units, or the reverse process.

7.3.3 As an example with $N = 2$, the form of the trapezoidal distribution in figure 3, in particular the breakpoints (where adjacent straight-line pieces intersect), can be determined explicitly in this case from the expectations and semi-widths (or standard deviations) of the input quantities [9].

7.3.4 An advantage of an analytic solution is that it provides insight through displaying the explicit dependence of the PDF for Y on any physically meaningful parameters.

7.4 Monte Carlo method

7.4.1 A Monte Carlo method (MCM) involves the following steps (figure 7) for the propagation and summarizing stages of uncertainty evaluation:

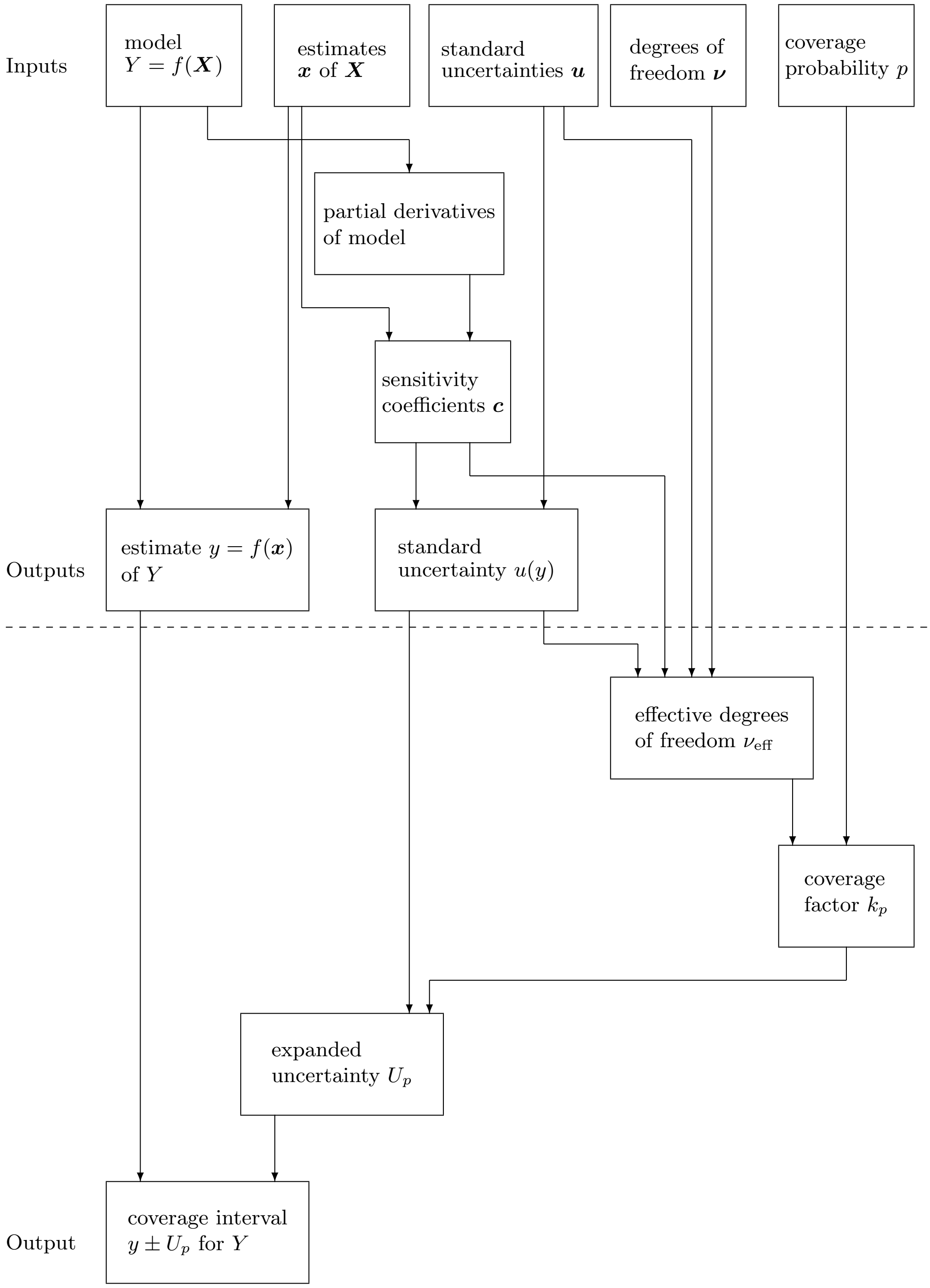

Figure 6 — Uncertainty evaluation using the GUM uncertainty framework. The part of the figure above the broken horizontal line relates to obtaining an estimate y of the output quantity Y and the associated standard uncertainty $u(y)$. The part below the line relates to the determination of a coverage interval for Y. In the figure, $\boldsymbol{x}$ denotes the vector containing the x_i, $\boldsymbol{u}$ similarly for the $u(x_i)$, $\boldsymbol{\nu}$ for the corresponding degrees of freedom ν_i (some or all of which might be infinite), and $\boldsymbol{c}$ for the c_i. Other terms are as defined in the text.

a) Select a number M of Monte Carlo trials;

b) Carry out the following steps for each value of r from 1 to M:

   1) Make a random draw from each of the PDFs for the X_i, for $i = 1, \ldots, N$;

   2) Evaluate the model (1) at the values of the X_i so obtained to give the model value y_r;

c) Sort the values y_r, for $r = 1, \ldots, M$, into non-decreasing order, to provide a discrete representation $\boldsymbol{G}$ of the probability distribution for Y. Denote the sorted values by $y_{(r)}$, for $r = 1, \ldots, M$;

d) Approximate y by the average of the y_r;

e) Approximate $u(y)$ by the standard deviation of the y_r;

f) Form a coverage interval $[y_{\text{low}}, y_{\text{high}}]$ for Y, selecting the endpoints y_{low} and y_{high} such that $100p$ % of the values of $y_{(r)}$ lie between them. Depending on how the endpoints are selected, the probabilistically symmetric coverage interval or the shortest coverage interval (subclause 4.11) or some other interval can be formed.

7.4.2 GUM Supplement 1 [5] provides detailed information on MCM as an implementation of the propagation of distributions. It gives examples to compare it with the use of the GUM uncertainty framework. It also provides an adaptive MCM procedure, in which the number M of Monte Carlo trials is determined automatically.

7.4.3 MCM has relatively few conditions associated with its use.

7.4.4 Sometimes it is unclear whether the application of the GUM uncertainty framework in any particular case is valid. A validation procedure can be used to check its adequacy. The procedure assesses the extent to which the GUM uncertainty framework provides y, $u(y)$ and a coverage interval for Y that are compatible (to within a stipulated numerical tolerance) with the corresponding results provided by MCM.

7.4.5 If the comparison is favourable it can reasonably be concluded that the GUM uncertainty framework is valid in that circumstance. It does not imply that it would be valid for (a) a different coverage probability, (b) the same model with different estimates of the input quantities, or (c) a modified model. If the comparison is unfavourable, the use of MCM could be considered for such uncertainty evaluations in the future.

7.5 Models with any number of output quantities

7.5.1 The GUM and GUM Supplement 1 [5] concentrate on measurement models having a single output quantity Y, that is univariate (scalar) models. Many measurement problems arise, however, for which there is more than one output quantity, depending on a common set of input quantities. These output quantities are denoted by $Y_1, \ldots, Y_m$ and the vector containing these quantities by $\boldsymbol{Y}$. The models of measurement that apply in such circumstances are referred to as *multivariate models*. Instances include (a) an output quantity that is complex, and represented in terms of its real and imaginary components (or magnitude and phase), (b) the coefficients of a calibration curve, and (c) parameters describing the geometry of the surface of an artefact. The GUM does not directly address such models, although examples are given concerning simultaneous resistance and reactance measurement [GUM:1995 H.2] and thermometer calibration [GUM:1995 H.3].

7.5.2 The counterpart of the model (1) is

$$\begin{aligned} Y_1 &= f_1(X_1, \ldots, X_N), \\ Y_2 &= f_2(X_1, \ldots, X_N), \\ &\vdots \\ Y_m &= f_m(X_1, \ldots, X_N), \end{aligned} \tag{4}$$

in which there are m model functions $f_1, \ldots, f_m$. Figure 8 illustrates a model with three input quantities X_1, X_2 and X_3, two model functions f_1 and f_2, and hence two measurands Y_1 and Y_2.

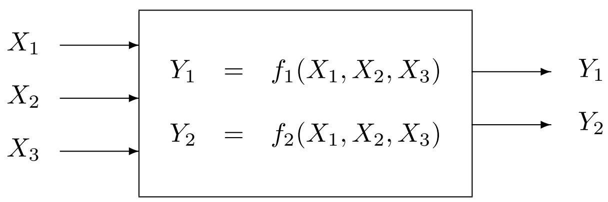

Figure 8 — Illustration of a model with three input quantities X_1, X_2 and X_3, two model functions f_1 and f_2, and hence two measurands Y_1 and Y_2.

7.5.3 The formulation phase of uncertainty evaluation is consistent with that for a model with a single measurand: it comprises developing a multivariate model of the form (4), and assigning PDFs to the input quantities based on available knowledge. However, for each $i = 1, \ldots, m$, there is now an estimate y_i of Y_i and a standard uncertainty $u(y_i)$ associated with y_i. Furthermore, since in general each Y_i depends on all $X_1, \ldots, X_N$, there will typically be non-zero

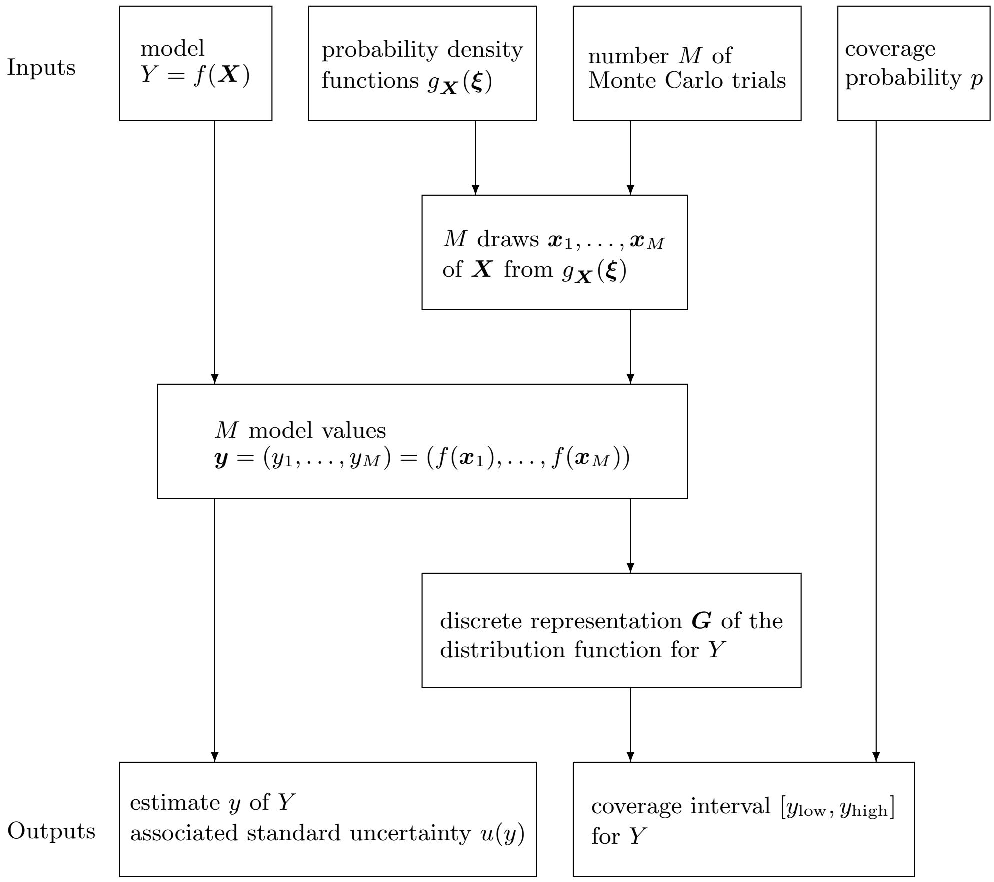

Figure 7 — Uncertainty evaluation using a Monte Carlo method. In the figure, $g_{\boldsymbol{X}}(\boldsymbol{\xi})$ denotes the (joint) PDF for the input quantities $\boldsymbol{X}$. Other terms are as defined in the caption to figure 6 and the text.

covariances $\mathrm{cov}(y_i, y_j)$ for all values of i and j from 1 to m.

7.5.4 In order to evaluate the uncertainties associated with a (vector) estimate $\boldsymbol{y}$ of the multivariate output quantity $\boldsymbol{Y}$ for such models, the GUM uncertainty framework and MCM for undertaking the propagation of distributions, as treated in GUM Supplement 1 [5], require extension. The GUM outlines such an extension, but considers it further only in examples.

7.5.5 In GUM Supplement 2 [6], a first step is to demonstrate that the law of propagation of uncertainty, a main constituent of the GUM uncertainty framework, can succinctly be expressed in an equivalent matrix form when applied to a univariate model. The matrix expression has the advantage of being particularly suitable (a) as the basis for implementation in software, and (b) for extension to more general types of measurement model. The extensions to a bivariate model (as a particular instance of a multivariate model) and a general multivariate model are provided. Furthermore, consideration is given to measurement models that are distinguished as explicit, that is in which the measurement model takes the form of formulae for $\boldsymbol{Y}$ in terms of the input quantities $\boldsymbol{X}$, or implicit, that is in which $\boldsymbol{X}$ and $\boldsymbol{Y}$ are related by equations or other relationships and such that $\boldsymbol{Y}$ cannot conveniently be expressed explicitly in terms of $\boldsymbol{X}$.

7.5.6 GUM Supplement 2 [6] also applies MCM to multivariate models. The counterpart of the discrete representation $\boldsymbol{G}$ of the probability distribution for the output quantity in the univariate case is provided. Expressions are given for the estimate $\boldsymbol{y}$ of $\boldsymbol{Y}$, and the variances and covariances associated with the components of $\boldsymbol{y}$, in terms of that representation.

7.5.7 The GUM uncertainty framework provides a coverage interval (in terms of an expanded uncertainty) for a univariate output quantity Y, given a coverage probability, by characterizing Y by a Gaussian distribution (or a scaled and shifted t-distribution) (subclause 7.2.1). GUM Supplement 1 [5] generalizes this concept to any distribution for a univariate output quantity, considering two coverage intervals in particular: the probabilistically symmetric coverage interval and the shortest coverage interval (subclause 7.4.1). There is no natural counterpart for multivariate models in the form of a coverage region of the former coverage interval, whereas there is for a shortest coverage interval. However, in addition to the determination of a smallest coverage region being a much more difficult task in the multivariate case, it is expected that, even if it were provided, the metrologist would not generally find it particularly useful. GUM Supplement 2 [6] provides probabilistically symmetric or shortest coverage intervals for individual output quantities or linear combinations of the output quantities. Such coverage intervals are often expected to be more practical than coverage regions.

7.5.8 In some circumstances, it is reasonable to provide an approximate coverage region having simple geometric shape. Two particular forms of coverage region are considered in this regard. One form results from assigning a multivariate Gaussian (multi-normal) distribution to $\boldsymbol{Y}$, for example, on the basis of the central limit theorem, in which case the smallest coverage region is bounded by a hyper-ellipsoid. The other form constitutes a hyper-rectangular coverage region.

8 The role of measurement uncertainty in deciding conformance to specified requirements

8.1 Conformance assessment is an area of importance to manufacturing industry and health and safety. In the industrial inspection of manufactured parts, decisions are made concerning the compatibility of the parts with the design specification. Similar issues arise in the context of regulation (relating to emissions, radiation, dope testing, etc.) concerning whether stipulated limits have been surpassed.

8.2 Measurement is intrinsic to conformance assessment in deciding whether a quantity of interest (the output quantity, or measurand) conforms to a specified requirement. For a single (scalar) quantity, such a requirement typically takes the form of specification limits that define an interval of permissible values of the quantity. A quantity lying within this interval is said to be conforming, and non-conforming otherwise. The influence of measurement uncertainty on the inspection process necessitates a balance of risks between producers and consumers.

8.3 The possible values of a quantity Y of interest are represented by a PDF, in many cases summarized by its expectation and standard deviation, taken to be an estimate y of Y and an associated standard uncertainty $u(y)$, respectively (subclause 4.9). The probability that the quantity conforms to specification can be calculated, given this PDF and the specification limits.

8.4 Because of the incomplete knowledge of Y (as encoded in the distribution for Y), there is a risk of error in deciding conformance to specification. Such errors are of two types: a quantity accepted as conforming may actually be non-conforming, and a quantity rejected as non-conforming may actually conform.

8.5 By defining an acceptance zone (or 'region of permissible values' [12]) of acceptable values of Y, the risks of accept/reject decision errors can be balanced in such a way as to minimize the costs associated with these errors. A document [8] addresses the problem of calculating the conformance probability and the probabilities of the two types of error, given the PDF, the specification limits, and the limits of the acceptance zone. The problem of choosing the acceptance zone limits is a decision that depends on the implications of accept/reject decision errors.

8.6 The general treatment is particularized to the most important case in practice, viz., when the PDF is Gaussian. Explicit formulae are provided in that case.

9 Applications of the least-squares method

9.1 A document [3] provides guidance on the application of the least-squares method, or least-squares adjustment (LSA), to data evaluation problems in metrology. In such problems there is often an underlying relationship between a stimulus variable and a response variable. This relationship constitutes the basic model of the parameter adjustment or curve-fitting problem.

9.2 In the context of calibration, a stimulus value would typically be that of a certified reference standard, and the response value that returned by the measuring system for that stimulus value.

9.3 In the curve-fitting context, the stimulus would correspond to an independent variable and the response to a dependent variable. The adjustment procedure used in the document is a generalized version of the usual least-squares procedure.

9.4 The model contains not only the defined stimulus and response quantities, but also other, unknown quantities, parameters that do not vary with the stimulus. The task is to estimate the parameters (and sometimes even their number) from measured values of the stimulus and the corresponding response quantity. These measured values yield the input data to the adjustment including the associated input uncertainty matrix. Because of possible correlations, this matrix (which contains the squared standard uncertainties and the covariances associated with the measured values) need not be diagonal.

NOTE The document [3] treats the case of imperfectly known stimulus values, corresponding, for example, to the use of standards specified in terms of certified reference values and their associated standard uncertainties.

9.5 The application of the document [3] is not restricted to curve-fitting problems in the strictest sense. It can also be used to treat, for instance, unfolding problems, the multivariate adjustment of fundamental constants, and key comparison data evaluation.

9.6 Typical measurement problems to which the document [3] can be applied include (a) linear or non-linear curve-fitting problems, including the case of imperfectly known stimulus, and (b) fitting of general models to identify physical parameters.

9.7 For problems of type (a) in subclause 9.6, once the least-squares method has been used to estimate the parameters of a calibration curve and evaluate the associated uncertainty matrix, the corresponding instrument will subsequently be used for measurement. Information about the parameters of the calibration curve and a particular response value is then used to infer the value of the stimulus, which assumes the role of the measurand. There is one equation, given by the inverse use of the calibration curve, for one output quantity. This is a clearly defined problem of uncertainty evaluation, where the output quantities of the adjustment problem enter as input quantities to a measurement model. Guidance on such evaluation tasks is given, with examples.

9.8 For problems of type (b) in subclause 9.6, or in terms of determining the parameters in problems of type (a), the adjustment problem is rarely a univariate problem, namely involving the evaluation of only one output quantity. Rather, the problem is generally multivariate in which the mathematical formulation can conveniently be expressed in terms of matrices. The document [3] makes extensive use of matrix formalism, which is well adapted to numerical solution using a computer, as usually required in practice (cf. clause 7.5).

Bibliography

[1] Bell, S. Measurement Good Practice Guide No. 11. A Beginner's Guide to Uncertainty of Measurement. Tech. rep., National Physical Laboratory, Teddington, UK, 1999.

[2] Bernardo, J., and Smith, A. *Bayesian Theory.* John Wiley & Sons, New York, 2000.

[3] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. Evaluation of measurement data — Applications of the least-squares method. Joint Committee for Guides in Metrology, Bureau International des Poids et Mesures, in preparation.

[4] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. Evaluation of measurement data — Concepts and basic principles. Joint Committee for Guides in Metrology, Bureau International des Poids et Mesures, JCGM 105, in preparation.

[5] 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, Bureau International des Poids et Mesures, JCGM 101, in preparation.

[6] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. Evaluation of measurement data — Supplement 2 to the "Guide to the expression of uncertainty in measurement" — Models with any number of output quantities. Joint Committee for Guides in Metrology, Bureau International des Poids et Mesures, JCGM 102, in preparation.

[7] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. Evaluation of measurement data — Supplement 3 to the "Guide to the expression of uncertainty in measurement" — Modelling. Joint Committee for Guides in Metrology, Bureau International des Poids et Mesures, JCGM 103, in preparation.

[8] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. Evaluation of measurement data — The role of measurement uncertainty in deciding conformance to specified requirements. Joint Committee for Guides in Metrology, Bureau International des Poids et Mesures, in preparation.

[9] Cox, M. G., and Harris, P. M. SS*f*M Best Practice Guide No. 6, Uncertainty evaluation. Tech. Rep. DEM-ES-011, National Physical Laboratory, Teddington, UK, 2006.

[10] Dietrich, C. F. *Uncertainty, Calibration and Probability.* Adam Hilger, Bristol, UK, 1991.

[11] Elster, C. Calculation of uncertainty in the presence of prior knowledge. *Metrologia 44* (2007), 111–116.

[12] ISO. ISO 10576-1. Statistical methods — Guidelines for the evaluation of conformity with specified requirements — Part 1: General principles, 2003. International Standards Organization, Geneva.

[13] ISO/IEC. ISO/IEC 17025, General Requirements for the Competence of Testing and Calibration Laboratories, 1999.

[14] Weise, K., and Wöger, W. A Bayesian theory of measurement uncertainty. *Measurement Sci. Technol. 3* (1992), 1–11.

Alphabetical index

A

analytic methods 7.3

B

bias 3.5

C

calculation . . . . . *see* stages of uncertainty evaluation
calibration 3.6
correction 3.9
coverage interval 4.11
    probabilistically symmetric 4.11
    shortest 4.11
coverage probability 4.11

E

expectation 3.14

F

formulation . . . . . *see* stages of uncertainty evaluation

G

GUM uncertainty framework [GUMS1:2007 5.6] . . . 7.2
    conditions on use
        linear models [GUMS1:2007 5.7] 7.2.2
        non-linear models [GUMS1:2007 5.8] 7.2.2
    validation using MCM [GUMS1:2007 8] 7.4.4

I

indication 3.6
input quantities 3.11, 3.12

M

measurand 3.1
measurement model 3.7
measurement uncertainty 4.2
measuring system 3.4
model of measurement 3.11, 3.12
Monte Carlo method [GUMS1:2007 5.9] 7.4
    adaptive [GUMS1:2007 7.9] 7.4.2
    comparison with the GUM uncertainty framework [GUMS1:2007 5.11] 7.4.2
    conditions on use [GUMS1:2007 5.10] 7.4.3
    examples [GUMS1:2007 9] 7.4.2
    implementation of [GUMS1:2007 7] 7.4.1

O

output quantity 3.11, 3.12

P

probability distribution 4.3
    assigned to input quantities [GUMS1:2007 6] . . 6.1
    distribution function 4.3
    probability density function 4.3
propagation of distributions [GUMS1:2007 5.2] . . 3.15, 7.1.1

R

random effect 3.5

S

stages of uncertainty evaluation [GUMS1:2007 5.1] . . 5
    calculation 7
    formulation 6
standard deviation 3.14
standard uncertainty 3.14
systematic effect 3.5

T

Type A evaluation of uncertainty 4.7
Type B evaluation of uncertainty 4.8