eurolab

European Federation of National Associations of Measurement, Testing and Analytical Laboratories

Technical Report No. 1/2006
August 2006

Guide to the Evaluation of Measurement Uncertainty for Quantitative Test Results

Technical Report

Impressum

EUROLAB Technical Report 1/2006
“Guide to the Evaluation of Measurement Uncertainty for Quantitative Test Results”

August 2006

EUROLAB Technical Secretariat - EUROLAB
1 rue Gaston Boissier
75724 PARIS Cedex 15
FRANCE

Phone: +33 1 40 43 39 45
Fax: +33 1 40 43 37 37
e-Mail: eurolab@lne.fr
URL: www.eurolab.org

Guide to the Evaluation of Measurement Uncertainty for Quantitative Test Results

Editorial note

This document is based on the *BAM-Leitfaden zur Ermittlung von Messunsicherheiten bei quantitativen Prüfergebnissen* published by the Federal Institute for Materials Research and Testing (BAM), Germany [Forschungsbericht 266, 2004], which was jointly agreed by BAM and the EUROLAB Board of Administrators to be translated into English and published as a EUROLAB Technical Report.

The original document was prepared on behalf of the BAM Quality Management Committee (AQM) as a technical guidance document supporting the BAM Directive for estimation and specification of the uncertainty of test results.

Author:	Werner Hässelbarth
Contributions:	Manfred Golze, Siegfried Noack, Andreas Subaric-Leitis
Translation:	Nigel Pye
Editing:	Werner Hässelbarth and Manfred Golze

Comments are acknowledged from Bertil Magnusson (SP, Sweden), Pascal Launey (LNE, France) on behalf of the measurement uncertainty expert group of EUROLAB-France, and Vitor Ramos (RELACRE, Portugal).

The draft technical report was approved by the EUROLAB Technical Committee for Quality Assurance in Testing (TCQA) at its meeting on 8 May 2006 and by the EUROLAB General Assembly in Borås, Sweden, on 16 May 2006.

Contents

Foreword

After more than ten years since the first arrival of measurement uncertainty in the EUROLAB community (where the GUM was presented at the Eurolab Symposium in Florence 1994), evaluation of measurement uncertainty for test results is still an issue of major concern – not in principle but in daily practice.

Concerning principal issues, the *Guide to the Expression of Uncertainty in Measurement*, known as the *GUM*, is acknowledged as the master document on measurement uncertainty throughout testing. The term “measurement uncertainty” is recognised to apply to all types of quantitative test results, and the GUM principles are fully accepted.

However, when it comes to evaluating the uncertainty of the results for a (quantitative) test procedure, the GUM is often criticised to be inapplicable. This impression is due to the fact that the GUM almost exclusively treats a single approach for uncertainty evaluation: the “modelling approach” based on a comprehensive mathematical model of the measurement procedure, where every uncertainty contribution is associated with a dedicated input quantity, the uncertainty contributions are evaluated individually and combined as a root sum of squares. This is therefore often (mis)conceived as being “the GUM approach” for uncertainty evaluation. Actually the GUM principles admit a variety of approaches, but this fact was buried under a plethora of papers and lectures celebrating the “modelling approach” as a new paradigm in measurement quality assurance. Only recently alternative “empirical approaches” received greater attention. They are based on whole-method performance investigations designed and conducted as to comprise the effects from as many relevant uncertainty sources as possible. The data utilised in these approaches are typically precision and bias data obtained from within-laboratory validation studies, quality control, interlaboratory method validation studies, or proficiency tests. Such approaches are fully compliant with the GUM, provided that the GUM principles are observed.

Eurolab has consistently advocated the use of empirical approaches as a valid, and often more practical alternative to the modelling approach, a.o. by publication of Technical Reports on measurement uncertainty in testing. The first in this series (No. 1/2002) is an introductory text for newcomers. This is now supplemented by a comprehensive technical guide for more experienced users. It provides a survey both of the modelling or “bottom-up” approach, which presupposes a complete mathematical model of the measurement process, and empirical or “top-down” approaches based on whole-method performance data. A further Eurolab TR under development by a dedicated expert group will deal with comparison and combination of uncertainty estimates obtained from the major approaches currently available. That report will include a collection of examples from different testing fields, where the results obtained using different approaches are compared, and the conclusions drawn from the comparison are discussed.

Guide to the Evaluation of Measurement Uncertainty for Quantitative Test Results

This document offers techical guidance on the evaluation of measurement uncertainty for quantitative test results. While fully compliant with the principles of the *Guide to the Expression of Uncertainty in Measurement* (GUM), the document also includes alternative approaches to the “bottom-up” approach, based on a comprehensive mathematical model of the measurement process, as emphasized in the GUM. These are “top-down” approaches utilising whole-method performance data from inter-laboratory comparisons (collaborative method validation, proficiency testing) and from within-laboratory validation and quality control data (precision, bias). Supplementary information concerning frequently occurring uncertainty sources and addressing data evaluation problems arising in uncertainty evaluation is given in various annexes.

1 Definitions

In these guidelines the terms “quantitative test“ and “measurement“ will be used as synonyms. In the same token and in accordance with relevant standards, the terms measurement, measurand, measuring object, measurement result and measurement uncertainty will predominantly be used. Without changing the basic content, these terms could be replaced with test, test quantity, test item, test result and uncertainty of the result.

Throughout this document the term “measurement procedure” is used to designate what is often called a “measurement method”: a protocol based on a specified measurement technique, developed and validated for specified measuring objects and measuring conditions. Only a well-defined measurement procedure allows for an associated measurement uncertainty which is applicable to within-specification measurements.

1.1 Terms of Measurement Uncertainty

The objective of a measurement (or any other quantitative investigation) is to determine an estimate for the true value of the measurand. This estimate, i.e. the measurement result, may be an individual measured value. Often however, the measurement result is obtained from a number of measured values by a statistical evaluation procedure, e.g. as a mean value. For each measurement procedure the expression of the measurement result and the data evaluation must be unambiguously defined.

As a rule the use of measurement results requires knowledge of the accuracy, i.e. the extent of the potential deviation of the measurement result from the true value of the measurand must be known. In metrology, “uncertainty of measurement“ is used as a quantitative measure of accuracy. This term is also used for the uncertainty of quantitative test results.

In the following, three definitions from basic terminology documents are reproduced (in extracts) which emphasise different aspects of uncertainty while their meaning is essentially the same.

Uncertainty (of measurement)

Parameter, associated with the result of a measurement, that characterises the dispersion of the values that could reasonably be attributed to the measurand.

(Source: International Vocabulary of Basic and General Terms in Metrology)

Uncertainty

Parameter obtained from measurements, which serves, together with the measurement result, to characterise a range of values for the true value of the measurand.

(Source: DIN 1319-1)

Uncertainty of the result

Estimated quantity intended to characterise a range of values which contains the reference value, where the latter may be either the true value or the expectation, depending on definition or agreement.

(Source: DIN 55350-13)

The following terms make up the "**G**uide to the Expression of **U**ncertainty in **M**easurement" (GUM) system of terms. These terms and their symbols defined by GUM will be used throughout this document.

Standard uncertainty (u)

Uncertainty of the result of a measurement expressed as a standard deviation.

Combined standard uncertainty (u)

Standard uncertainty of the result of a measurement when that result is obtained from the values of a number of other quantities, equal to the positive square root of a sum of terms, the terms being the variances or covariances of these other quantities weighted according to how the measurement result varies with changing these quantities.

> *Note*: In GUM, combined standard uncertainties are marked by an index with u_c. This marking will not be used here, since a distinction between combined and non-combined standard uncertainties has no practical relevance in testing.

Expanded uncertainty (U)

Quantity defining an interval about the result of a measurement that may be expected to encompass a large fraction of the distribution of values that could reasonably be attributed to the measurand.

Coverage factor (k)

Numerical factor used as a multiplier of the (combined) standard uncertainty in order to obtain an expanded uncertainty.

1.2 Terms of Test Accuracy

The terms in the previous section are predominantly new creations from the field of metrology, opposed to a generally accepted system of terms used in the field of testing and chemical analysis. Since this system of terms is in common use there, the principal terms are compiled in this section, taken from the fundamental terminology standard ISO 3534-1. The relationship of these two systems of terms will be discussed in Section 2.

Accuracy

The closeness of agreement between a test result and the accepted reference value.

Trueness

The closeness of agreement between the average value obtained from a large series of test results and an accepted reference value.

Precision

The closeness of agreement between independent test results obtained under stipulated conditions.

2 Basics

2.1 Basic Metrological Terms and Concepts

Terms in bold are defined in the relevant standards. Unless stated otherwise, the terms are based on the *International Vocabulary of Basic and General Terms in Metrology* (VIM), 1994, 2nd Edition. In accordance with relevant standards, the terms measurement, measurand, measurement result and measurement uncertainty will exclusively be used in this section. Without changing the basic content, these terms could be replaced with test, test quantity, test result and uncertainty of the result.

In the simplest case of a measurement one has only a single **measurand**, i.e. the subject of the measurement is only one particular quantity. For instance, this may be the vapour pressure of a given water sample at 20 ºC. It is crucial that the measurement task is accurately defined by specifying all relevant parameters, e.g. time, temperature or pressure. If the measurand of a measurement task is accurately defined in this way, then an unambiguous value, the so-called **true value**, can be attributed to it. An ideal measurement would yield this true value.

However, since one always has to work with real measurements, there is an (unknown) difference, called **error** between the measurement result and the true value. Repeated measurements generally fail to yield the same value each time, rather, they produce values more or less close to each other. Repeating a measurement many times and plotting the frequency with which a value x occurs as a function of x, one would obtain a bell-shaped curve, which can be approximated in many cases by the so-called normal distribution (see Figure 2.1). A normal distribution is characterised by two parameters: the location parameter μ, which indicates the position of the maximum, and the standard deviation σ, which describes the width of the curve.

Because of this dispersion of measured values, if possible and efforts are justified, measurements are carried out several times (n times) and the arithmetic mean $\overline{x}$ of n individual values x_i is calculated according to Eq. (2.1).

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad (2.1)$$

($\overline{x}$: arithmetic mean; x_i: i-th measured value; n: number of measurements, n > 1)

The (experimental) standard deviation s calculated by Eq. (2.2) is a measure of the dispersion of the individual values, i.e. the width of the curve in Figure 2.1.

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2} \qquad (2.2)$$

(s: experimental standard deviation; $\overline{x}$: arithmetic mean; x_i: i-th measured value; n: number of measurements, n > 1)

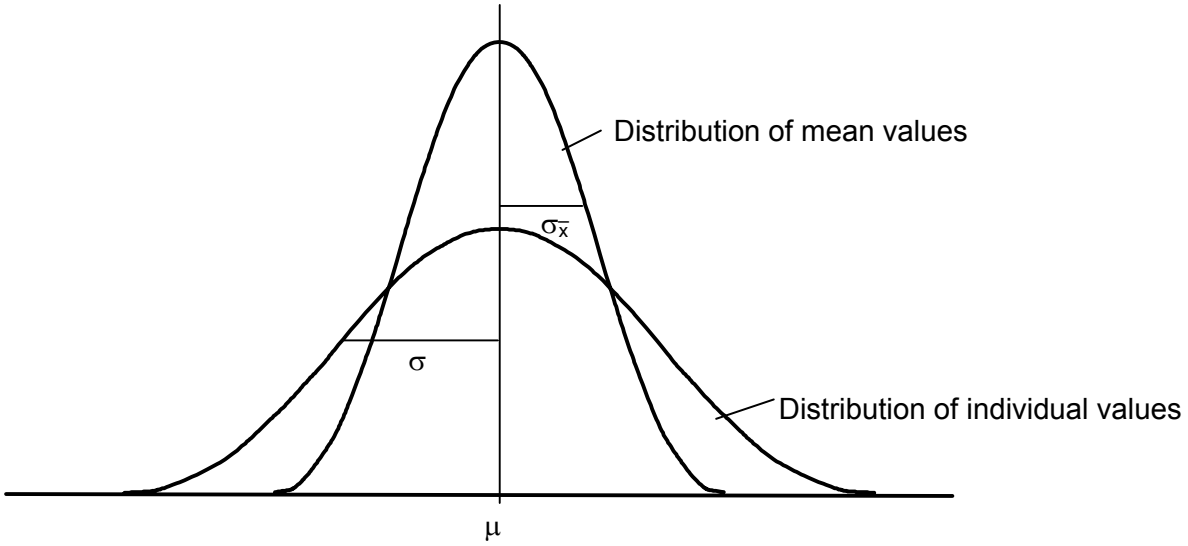


Figure 2.1 Distributions for individual measured values x with parameters μ and σ and for the mean values $\bar{x}$ from n measurements each with parameters μ and $\sigma_{\bar{x}}$.

If such series of measurements, each comprising n individual measurements are repeated many times, and the means calculated and plotted analogously to the diagram of the individual values, another normal distribution is obtained with the same position parameter μ, but a smaller width (see Figure 2.1). The standard deviation $\sigma_{\bar{x}}$ of this distribution is given by:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \tag{2.3}$$

($\sigma_{\bar{x}}$: standard deviation of the means; σ: standard deviation of the individual values; n: number of measured values used for the calculation of means).

The dispersion of the measured values obtained under apparently identical conditions is the result of a multiplicity of influences beyond control of the measuring conditions, whose effect changes when measurements are repeated. The deviations of the measured values from the central value μ, varying between positive and negative, are designated as **random errors**. If only random errors are present, μ equals the true value of the measurand. One would obtain μ as the mean $\bar{x}$ if one could repeat the measurement for an unlimited number of times, because the standard deviation of the mean would then shrink towards zero.

In practice however, only a limited number of repetitions of the measurement is possible, so a certain dispersion of the means and thus a certain lack of knowledge of the measurand remains, which one seeks to estimate by **measurement uncertainty**. Measurement uncertainty is defined according to DIN 1319-1 as a "parameter which is obtained from measurements and, together with the measurement result, serves to characterise a range of values for the true value of the measurand".

In addition to these random errors, one also usually has to deal with so-called **systematic errors.** They result in the centre of the distribution being shifted away from the true value, even in the case of infinite repetitions (see Figure 2.2). Potential causes for random and systematic errors are listed in Annex A.1. Systematic errors detected should be eliminated, as far as possible, or minimized by applying suitable corrections, accounting for the uncertainty of the correction in the uncertainty budget.

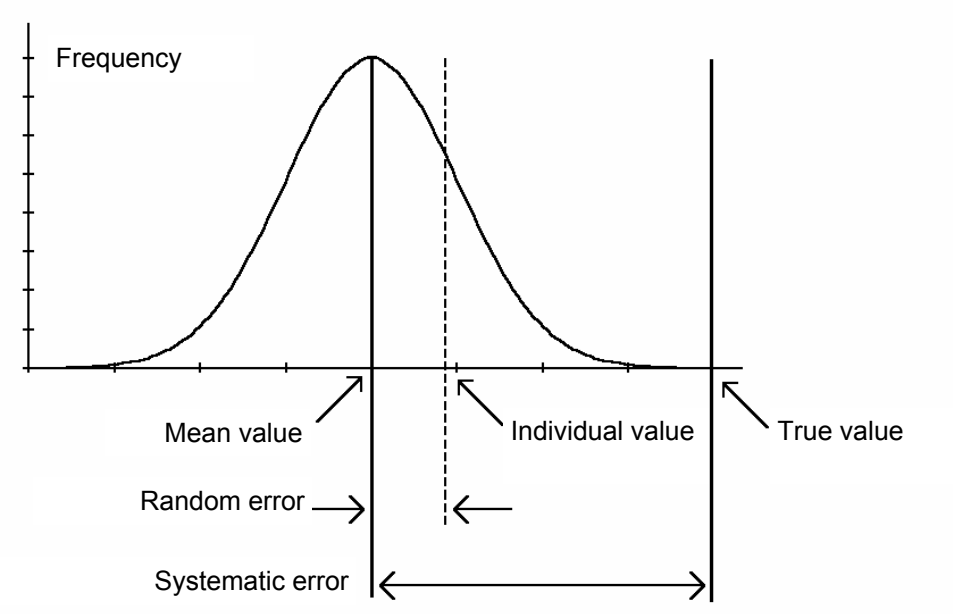


Figure 2.2 Measured values for simultaneously occurring random and systematic errors

Figure 2.3 shows how the different types of measurement error enter into the result of a measurement and the associated uncertainty.

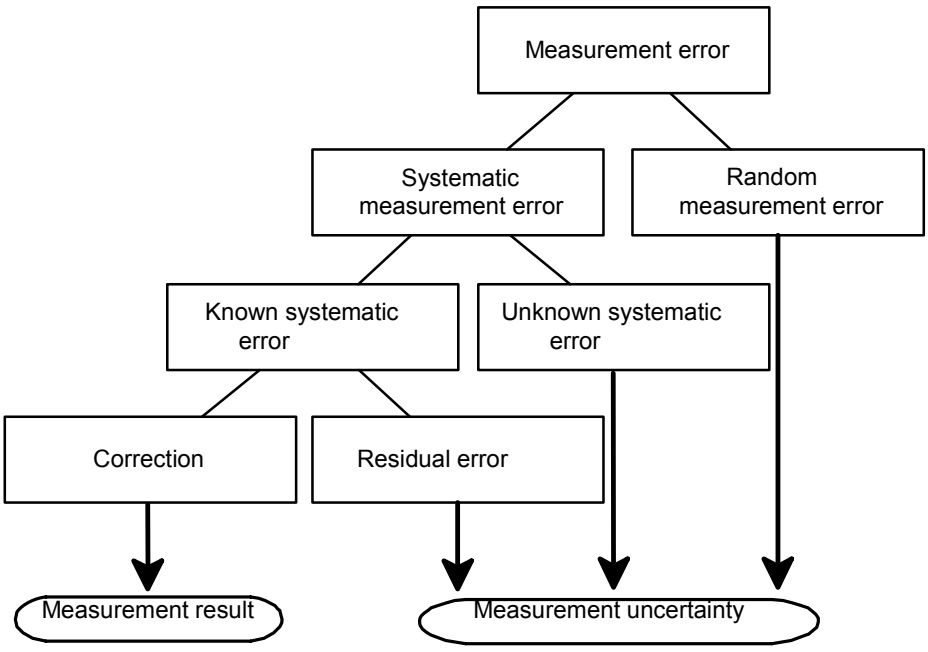


Figure 2.3 Types of measurement error and their consideration in determining the result of a measurement and the associated uncertainty (Figure according to M. Hernla, QZ **41** (1996), 1156)

2.2 Accuracy, Trueness and Precision; Target Model

The terms **accuracy**, **trueness** and **precision** from ISO 3534-1 as defined in Section 1 of this guide can be used to characterise a measurement procedure with respect to the associated uncertainty.

Accuracy as an umbrella term characterises the closeness of agreement between a measurement result and the true value. If several measurement results are available for the same measurand from a series of measurements, accuracy may be split up into trueness and precision, where trueness accounts for the closenes of agreement between the mean value and the true value, while precision accounts for the closeness of agreement between the individual values among themselves (see Figure 2.4).

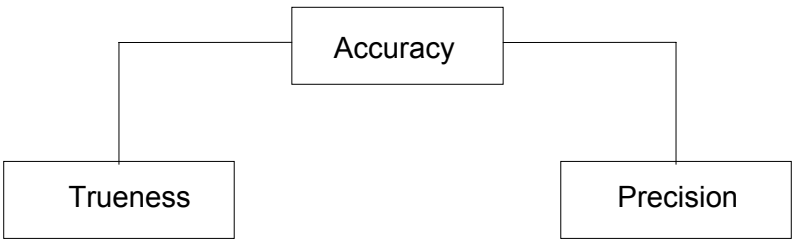


Figure 2.4 Accuracy as an umbrella term for trueness and precision

The different possible combinations which result from true or wrong and precise or imprecise results, can be best described using the target model (Figure 2.5).

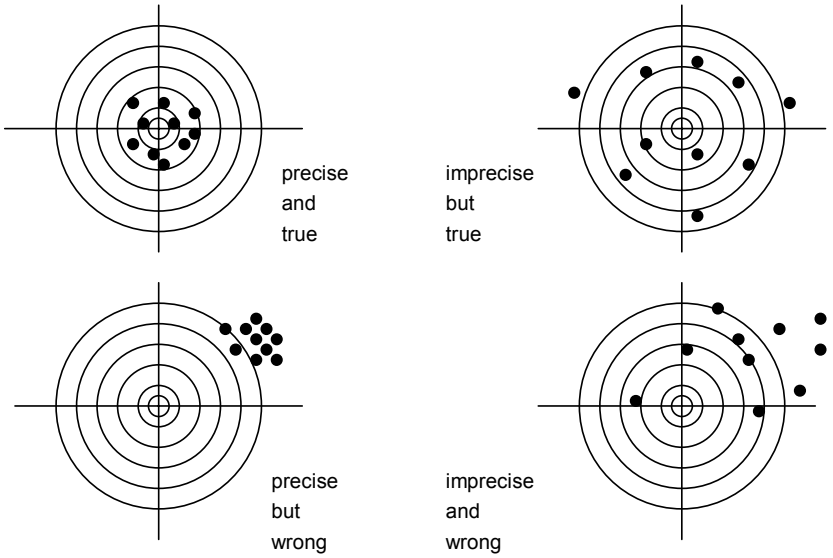


Figure 2.5 Target model to illustrate trueness and precision. The centre of the target symbolises the (unknown) true value.

Estimates of precision are strongly dependant on the conditions under which precision is investigated. Therefore **repeatability precision**, **reproducibility precision** and **intermediate precision** are distinguished, referring to **repeatability conditions**, **reproducibility conditions** and **intermediate conditions**.

Repeatability conditions include:

- the same measurement procedure,
- the same laboratory
- the same operator,
- the same equipment,
- repetition within short intervals of time.

Reproducibility conditions include:

- the same measurement procedure,
- different laboratories,
- different operators,
- different equipment.

Repeatability conditions and reproducibility conditions represent the cases of minimum and maximum variability in conditions for repeated measurements. Conditions between these extreme cases are called intermediate conditions. When using intermediate conditions, it must be specified exactly, which factors are varied and which are constant. For the within-laboratory characterisation of precision of measurement procedures e.g. the following conditions are used:

- the same measurement procedure,
- the same laboratory,
- different operators,
- the same equipment (alternatively: different equipment),
- repetition within long intervals of time.

This special case of intermediate conditions is often called “within-laboratory reproducibility conditions“.

While determining the precision of a measurement procedure is fairly straightforward, it is much more difficult to investigate the trueness of a measurement procedure since the true value of the measurand is in principle unknown. One approach is to apply the measurement procedure to suitable reference objects (standards, material measures, reference materials). Alternatively, a reference procedure is applied in parallel with the measurement procedure to suitable measuring objects. The value attributed to a reference object or the result obtained by the reference procedure, respectively, are then used as a reference value, i.e. as an estimate of the unknown true value, whose uncertainty is known and sufficiently small for the intended purpose. Trueness is then referred to this reference value.

2.3 New Aspects in "Guide to the Expression of Uncertainty in Measurement"

The ***G****uide to the Expression of* ***U****ncertainty in* ***M****easurement* (GUM) takes a somewhat different view compared with the traditional approach and promotes a unified and pragmatic procedure for determining the uncertainty, which will be explained in the following.

2.3.1 New Definition of Measurement Uncertainty

Since the true value is an ideal quantity, which is in principle unknown, a new definition has been developed for the term **measurement uncertainty** when drafting the GUM, which does not refer to the true value any more:

Uncertainty of Measurement

Parameter, associated with the result of a measurement, that characterises the dispersion of the values that could reasonably be attributed to the measurand.

This definition, explained in Annex D of GUM in great detail, has also been included in the 2nd edition of the *International Vocabulary of Basic and General Terms in Metrology* (VIM).

The range of values for the measurand typically includes values obtained under repeatability conditions (see above), and may also include values obtained under reproducibility conditions, e.g. by another operator, in another laboratory or by another measurement procedure, accounting for biases between operators, laboratories and measurement procedures. In addition, differences with respect to data processing (e.g. correction of identified bias, cf. Figure 2.3) may also contribute to this dispersion. Also the measurand may not be defined so exactly that a single true value can be attributed to it.

Unless some of the various values obtained experimentally or derived by calculation or theory turn out to be wrong, all of them must be assigned to the measurand. Uncertainty is a measure for the width of the range derived from these data and, together with an appropriate mean value as the result of measurement, describes the level of knowledge about the measurand. Because of our limited knowledge it is quite possible that the uncertainty is underestimated due to missing uncertainty components.

Despite this somewhat different view, there is no fundamental disagreement between GUM and traditional uncertainty standards.

2.3.2 Determining Type A and Type B Uncertainty Components

GUM classifies uncertainty components according to their method of determination into type A and type B:

Type A: Evaluation using statistical analysis of measurement series,

Type B: Evaluation using means other than statistical analysis of measurement series.

This classification will be explained in Section 3.2. It has some relation to the distinction between uncertainty components arising from random effects and uncertainty components arising from systematic effects, but there are essential differences.

With respect to the suggested methodology, GUM does not differentiate between uncertainty components due to systematic effects and uncertainty components due to random effects. It is however assumed that, as far as possible, recognised systematic errors are either eliminated by technical means or corrected by calculation. For the uncertainty budget a component remains which accounts for the uncertainty arising from any such action.

GUM proposes a uniform treatment for all uncertainty components (see Section 2.3.3). The reason being that for the error related to an uncertainty component the systematic or random character is not unambiguously defined, but depends on the actual case. Thus an error based on random effects becomes a systematic error, if the measurement result is entered as input into a further measurement.

Example: The concentration of a radioactive isotope in a reference standard was determined by activity measurements. For simplicity, assume that exclusively random deviations occur in this measurement.

If the unknown content of a sample is then determined in further measurements through comparison with this reference standard, its error affects all of these measurement results in the same way, thus effecting a systematic error.

Conversely, systematic errors made by a laboratory in performing specified measurements become random errors if the results of a large number of laboratories, exhibiting different systematic errors, are collected in an inter-laboratory comparison and summarized in a reproducibility standard deviation for the measurement procedure under investigation.

2.3.3 Equal Treatment of All Uncertainty Components

When calculating the combined standard uncertainty, all uncertainty components are treated equally. A comprehensive discussion and justification of this procedure is given in Annex E of GUM.

Traditionally uncertainty components arising from random effects and uncertainty components arising from systematic effects (for short: random components and systematic components) were most often treated in a different fashion as follows: random components were added as a root sum of squares (according to Eq. (3.12) in Section 3.3) while the systematic components were added linearly (by Eq. (3.16) in Section 3.4). These two sums were then added linearly. The intention in doing so was to obtain a "conservative" estimate for the uncertainty, i.e. to avoid underestimating uncertainty under any circumstances. A possibly excessive uncertainty was then accepted.

Example: The uncertainty budget for a measurement has yielded the values 3 and 2 as random components and 2 and 4 as systematic components (in arbitrary units).

Under GUM all these components are added as a root sum of squares (cf. Section 3.3, Eq. (3.12)):

$$u = \sqrt{3^2 + 2^2 + 2^2 + 4^2} = \sqrt{9 + 4 + 4 + 16} = \sqrt{33} = 5.74$$

However, if the systematic components are added linearly, one gets

$$u' = \sqrt{3^2 + 2^2} + 2 + 4 = \sqrt{9 + 4} + 6 = 9.61$$

i.e. a considerably larger value for the uncertainty.

Within the framework of GUM protection against underestimation of uncertainty is achieved by choosing a suitable coverage factor k for the expanded uncertainty (see Section 3.3). In addition, worst-case estimations of standard uncertainty are also reasonable under certain circumstances, e.g. comparison with a specification limit (see the remarks in Sections 2.4 and 3.5).

2.3.4 Expanded Uncertainty

One option for specifying measurement uncertainty under GUM is the expanded uncertainty

$$U(y) = k \times u(y)$$

i.e. the product of the standard uncertainty u(y) and an appropriate coverage factor k. This yields an interval, the so-called confidence interval

$$y - U(y) \leq Y \leq y + U(y)$$

(y: measurement result; Y: true value of measurand; U: expanded uncertainty)

which can be expected to include the true value Y of the measurand at a defined probability p (e.g. p = 95 %). From the viewpoint of GUM, this interval contains the proportion p of all values that can be reasonably attributed to the measurand.

The calculation of the confidence interval assumes the knowledge of the probability distribution for the measured values. Since this condition is usually only very imperfectly fulfilled, GUM reasonably suggests the choice of a coverage factor k between 2 and 3. A default value of k = 2 is recommended, which roughly corresponds to a confidence level p of 95%. In each case the factor k must be explicitly stated, so that the standard uncertainty u can be recovered.

A statistically more substantially founded procedure for the determination of the coverage factor can be found in the Annex G of GUM.

Example: In a test report an expanded uncertainty U = 11.48 is given with a coverage factor k = 2. From this the standard uncertainty u is obtained as:

$$u = \frac{U}{k} = \frac{11.48}{2} = 5.74 .$$

2.4 Worst-case Estimation of Measurement Uncertainty

Worst-case estimates of measurement uncertainty can be of interest if for instance the extent of measurement uncertainty plays only a subordinate role to further investigations or if compliance with certain limit values or specifications has to be checked. In this case, unlike the principle of squared addition, the contributions to the uncertainty of the result are summed up linearly and if necessary, maximum errors can also be used instead of standard uncertainties (see Section 3.5, Eqs. (3.16) and (3.17)) which leads to a simplified determination of measurement uncertainty.

3 Analytical-Computational Determination of Measurement Uncertainties

3.1 Overview

The analytical-computational determination of measurement uncertainty is generally a complex procedure, including many steps to be followed and requiring consideration of many aspects. The main ingredients will be summarized below to provide an overview of the procedural steps and the aspects involved as described in the next sections.

Prerequisite: systematic effects – as far as known – are eliminated or corrected.

- All relevant uncertainty sources are identified and listed.
- The contributions of the individual uncertainty sources to the uncertainty of the result are estimated and sorted according to significant/insignificant. Insignificant uncertainty contributions are neglected.
- The remaining (significant) uncertainty contributions are quantified as standard uncertainties (standard deviations). The following methods are equivalent: statistical evaluation of measurement series (type A evaluation) and estimation based on an alternative procedure (type B evaluation).
- The uncertainty contributions are examined for correlations. If necessary, correlations are quantified as covariances.
- The uncertainty contributions are combined using squared addition; if necessary covariances are included.
- To specify the result, the combined standard uncertainty is multiplied by a suitable coverage factor (usually k = 2).
- When determining worst-case uncertainty estimates, the uncertainty contributions are added linearly; covariances are omitted.

Usually measurement uncertainty is not determined individually for single measurement results, but as a parameter for a measurement procedure. It then applies to all measuring objects and all measuring conditions which were considered when the measurement uncertainty was determined. Therefore before use, each case must be checked as to whether the measuring object and the measuring conditions comply with the specification for the determination of measurement uncertainty. If significant uncertainty components of the application are not accounted for in the “procedural uncertainty”, then it is often appropriate

to adopt the procedural uncertainty as a component of the measurement uncertainty and to supplement the missing uncertainty contributions.

When determining measurement uncertainty, the cost/benefit ratio has to be considered. It is for example better to determine all significant uncertainty contributions with an acceptable accuracy instead of determining individual uncertainty contributions with extreme accuracy while others are only estimated roughly or ignored completely.

3.2 Classification of Measurement Uncertainty According to Evaluation Type

According to GUM (see Section 2.3) all uncertainties are expressed by standard deviations, independently of whether they are based on random or systematic effects. There are essentially two different procedures for the determination of this standard deviation. The conventional procedure (type A evaluation) is based on the assumption of a probability distribution for the random variation of measurement results. Estimates of the standard deviation of this distribution are obtained by replicate measurements and statistical analysis of the measured values (measurements series).The alternative procedure (type B evaluation) is predominantly used for the estimation of uncertainties, which are due to systematic effects. It uses reasonably assumed probability distributions, which account for the available information about the quantities concerned, and the standard deviation of these distributions. The two classes of uncertainty evaluation are defined in GUM as follows.

Type A: Evaluation using statistical analysis of measurement series;

Type B: Evaluation using means other than statistical analysis of measurement series.

A typical example for a **type A evaluation** is the determination of an estimate of the standard deviation σ of an assumed normal distribution. If $x_1, x_2, ..., x_n$ are the results of repeated measurements of the quantity concerned, then the experimental standard deviation s of the measurement series $\{x_1, x_2, ..., x_n\}$ can be used as an estimate of the standard deviation σ of this normal distribution.

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}} \tag{3.1}$$

where

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{3.2}$$

In absence of systematic errors the (arithmetic) mean $\bar{x}$ of the measurement series $\{x_1, x_2, ..., x_n\}$ is a suitable estimate for the value of the measurand. The standard uncertainty $u(\bar{x})$ of this result is given by

$$u(\bar{x}) = \frac{s}{\sqrt{n}} \tag{3.3}$$

If it can be taken for granted that in the measuring range concerned the measurement procedure operates free of biases and with a constant statistical spread, then the experimental standard deviation of the measurement series $\{x_1, x_2, ... , x_n\}$ can also be used for the estimation of the standard uncertainty of the results of other measurements within this measuring range. Here it has to be considered whether the result is a single measured value or the mean of several independently measured values. For a single value the standard uncertainty is equal to s, while for an (arithmetic) mean of m values the standard uncertainty equals $s/\sqrt{m}$.

Note: The factor $1/\sqrt{n}$ for the standard deviation of a mean of n individual values applies only to individual values independent of each other. The gain in precision is smaller for individual values dependent on each other (due to correlated errors), see Annex A.5.

A typical example of a **type B evaluation** is the transformation of a maximum/minimum specification into a standard uncertainty. Suppose that only a minimum value x_{min} and a maximum value x_{max} are known for the characteristic value (reference value) attributed to a reference material. If all values in this interval are equally likely candidates of the true value, the mean and the standard deviation of a rectangular distribution with boundaries x_{min} and x_{max} can be used for the reference value x and its standard uncertainty u(x).

$$x = \frac{(x_{max} + x_{min})}{2} \tag{3.4}$$

$$u(x) = \frac{(x_{max} - x_{min})}{\sqrt{12}} \tag{3.5}$$

However, if there is reason to believe that values in the centre of the interval are more likely than values at the boundaries, then e.g. a symmetrical triangular distribution with boundaries x_{min} and x_{max} can be chosen instead of the rectangular distribution (uniform distribution). This gives

$$x = \frac{(x_{max} + x_{min})}{2} \tag{3.6}$$

$$u(x) = \frac{(x_{max} - x_{min})}{\sqrt{24}} \tag{3.7}$$

These and other examples of type B evaluation are contained in GUM, Sections 4.3 and 4.4.

Note: Until recently, almost exclusively type A procedures have been used for the evaluation of uncertainties. Since these procedures are not universally applicable, often significant uncertainty components were either not taken into account properly or even not at all. The introduction of type B procedures serves the purpose to remedy this deficiency and to facilitate the use of expert knowledge for the estimation of uncertainty components.

In general, the uncertainty of a measurement result is made up from several components, part of which was evaluated through type A, the other part through type B procedures. Therefore, the classification according to type A or type B is usually only applicable for individual uncertainty components.

3.3 General Method for Determination of Uncertainty

The uncertainty of a measurement result usually consists of several components. Accordingly, the determination of the measurement uncertainty is also usually a complex procedure, which comprises several steps.

In this section a generally applicable succession of steps will be described. The formulation selected here refers to measurement procedures, but it can easily be transferred to test procedures and analytical procedures.

Frequently occurring uncertainty sources and data evaluation techniques are described in the Annex.

Step 1: Specification of the measurand and the measurement procedure

In this step the quantity to be measured, y, and the procedure for the determination of its value are specified. In addition to the actual measurement, this procedure comprises all preparative steps, e.g. sampling and sample preparation, the conditions which have to be maintained during preparation and measurement as well as data processing.

Step 2: Definition of the input quantities, identification of the uncertainty sources

In this step the input quantities x_i (i = 1, 2…, N), are defined, on which the result depends. As a basis for this, all potential uncertainty sources for the result are identified (see Annex A.1). The input quantities are defined in such a way that the effect of all relevant uncertainty sources is covered.

The input quantities may be:

- Constitutive measurands of the target quantity, e. g. mass and volume, when density as a quotient of these two is determined;
- Parameters, i.e. quantities that are not the subject of the measurement, but have an effect on the result, e.g. pressure and temperature of the sample in a volume measurement;
- Reference quantities, i.e. quantities used for calibration or for correction of systematic errors, e.g. values embodied by standards or reference materials;
- Characteristics for the input/output behaviour of individual steps of the complete measurement procedure, e.g. efficiencies of sample preparation procedures, correction factors for biases observed, parameters of a calibration curve etc.;
- Other quantities used during evaluation, for which data are taken from the literature, e.g. natural constants or materials characteristics.

Uncertainties of the input quantities are sources for the uncertainty of the measurement result. Conversely, the effect of each uncertainty source can be described by means of suitable input quantities (e.g. efficiencies or correction factors). Such a description is assumed in the following. For this purpose the input quantities must be defined in such a way that the effect of all potential uncertainty sources is accounted for. Here the use of flow charts is recommended. The use of efficiencies, correction factors or the like as input quantities for the modelling of process steps is dealt with in Annex A.3.

Summing up, the task within this step is to develop a mathematical model for the complete measurement procedure, $y = F(x_1, x_2, ..., x_N)$, i.e. an equation or an algorithm, which describe the measurement result as a function of all relevant input quantities.

Step 3: Determination of significant uncertainty sources

In this step the identified sources of uncertainty are assessed as to whether their contribution to the uncertainty of the result is significant. To this end, the uncertainty contribution of an input quantity is approximately calculated as a product of a rough estimate of the standard uncertainty associated with this quantity (e.g. accounting for the variability to be expected under given conditions) and the sensitivity with which the result depends on the input quantity.

If two contributions differ by a factor of 1/5, then the smaller contribution can usually be neglected in relation to the larger one.

> *Note:* Because of the squared addition, a standard uncertainty which is smaller by a factor of 1/p, contributes a proportion of approx. $1/(2p^2)$ of the larger standard uncertainty to the combined uncertainty of the result. For p = 5 this proportion amounts to approximately 2 %. However, small uncertainty contributions cannot be neglected if

they occur in large numbers or if correlations are present which entail a linear addition of uncertainty contributions instead of the squared addition.

Step 4: Quantifying significant uncertainty sources

In this step the contributions of significant uncertainty sources are quantified by means of the associated input quantities x_i (i = 1, 2…, N). For each of these, the standard uncertainty $u(x_i)$ is determined (depending upon available experimental data) either as the standard deviation of the values of a measurement series (type A evaluation) or as the standard deviation of a reasonably assumed probability distribution (type B evaluation), e.g. a rectangular distribution between experimentally established extreme values.

Note: Type A evaluations apparently have the advantage of greater objectivity. However, experimental standard deviations of very short measurement series, which are quite common in practice, provide such inaccurate estimates of standard uncertainties that an experience-based expert estimate (type B evaluation) may be preferred. For example the relative standard deviation of an experimental standard deviation from 5 values amounts to approximately 36 % and for 10 values it is still 24 %. This holds for values from a normal distribution; in the case of deviations from the normal distribution the “uncertainty of uncertainty estimates” may be even worse.

Furthermore, the sensitivity coefficients c_i for the input quantities x_i are determined. These coefficients specify how the result $y = F(x_1, x_2, ..., x_N)$ changes with variations of x_i. They are given by the derivatives

$$c_i = \frac{\partial F}{\partial x_i} \tag{3.8}$$

In the case of simple model functions for the result y (sums, products etc.) the derivatives can be obtained by differential calculus. If model functions are more complicated, numerically calculated difference quotients can be used instead of derivatives (see Annex A.4). If the influence of an input quantity x_i on the result cannot be described by a model, experimentally determined difference quotients will be used instead

$$c_i = \frac{\Delta y}{\Delta x_i} \tag{3.9}$$

The contribution of the uncertainty of an input quantity x_i to the combined standard uncertainty of the result y is obtained as a product $u_i = c_i \times u(x_i)$ of the standard uncertainty $u(x_i)$ and the sensitivity coefficient c_i.

Step 5: Consideration of correlations

This step first checks whether there are correlations between uncertainty contributions. Such correlations arise when errors of two input quantities x_i and x_k are dependent on one another and behave either sympathetically or antagonistically. Correlations can be expected if the input quantities concerned depend on each other or both depend on a third quantity. This can refer to the quantities themselves, or to the procedures for the determination of their values.

Example: A correlation exists, if the same standard is used for the calibration of two different measurements, or if two volumetric solutions are prepared by dilution from the same parent solution. An error of the standard then affects the results of the two measurements in the same direction. Likewise, an error of the parent solution formulation similarly affects the concentration of the two volumetric solutions.

In principle, correlations should be avoided to the largest possible extent. That is, preferably independent input quantities and independent procedures for the determination of their values should be used. If this is not possible, the correlations must be quantified by

appropriate covariances and taken into account in the calculation of the combined standard uncertainty of the result.

Correlations contribute as products $u_{ik} = c_i \times c_k \times u(x_i, x_k)$ of the covariance $u(x_i, x_k)$ and the relevant sensitivity coefficients c_i and c_k to the combined standard uncertainty of the result y.

The determination of covariances is briefly discussed in Annex A.6 and in greater detail in the GUM and DIN 1319-4. The following case is easy to handle and sufficient for many purposes: Two input quantities x_i and x_k depend on the same quantity z. The covariance of x_i and x_k is then $u(x_i,x_k) = (\partial x_i/\partial z)(\partial x_k/\partial z)u(z)^2$. Here u(z) is the standard uncertainty of z while $(\partial x_i/\partial z)$ and $(\partial x_k/\partial z)$ are the sensitivity coefficients for the dependence of the quantities x_i and x_k on z. If two input quantities depend on several common quantities, then the covariance is the sum of the respective products.

Step 6: Calculating the combined standard uncertainty

In this step the contributions determined in the preceding steps are combined into the standard uncertainty of the result. In the most general version, i.e. when correlations between all input quantities are considered, this combination is performed according to:

$$u(y)^2 = \sum_{i=1}^{N} u_i^2 + 2\sum_{i=1}^{N-1}\sum_{k=i+1}^{N} u_{ik} \tag{3.10}$$

In greater detail, this equation reads

$$u(y)^2 = \sum_{i=1}^{N}\left(\frac{\partial F}{\partial x_i}\right)^2 u(x_i)^2 + 2\sum_{i=1}^{N-1}\sum_{k=i+1}^{N}\left(\frac{\partial F}{\partial x_i}\right)\left(\frac{\partial F}{\partial x_k}\right)\cdot u(x_i, x_k) \tag{3.11}$$

In most applications there are no correlations between the input quantities or the contribution of the correlations can be neglected. Then Eq. (3.11) reduces to

$$u(y)^2 = \sum_{i=1}^{N}\left(\frac{\partial F}{\partial x_i}\right)^2 u(x_i)^2 \tag{3.12}$$

The standard uncertainty of the result u(y) is obtained as the positive square root of the sum calculated by equation (3.11) or (3.12).

> *Note*: Equation (3.12) is the usual form of the Gauss "error propagation law" for uncorrelated errors. Eq. (3.11) is its generalisation, accounting for correlations. Both equations are based on a series expansion of the result in terms of powers of the deviations of the input quantities from their stipulated values, which is truncated after the linear term. If there are pronounced non-linearities, this approximation may be insufficient. In this case either consecutive terms of the series expansion (higher powers of the deviations) must be included, or other evaluation methods (numerical simulation etc., see Annex A.4.2) must be used.

If the relation between the result y and the input quantities x_i, can be expressed by simple formulas, then the sensitivity coefficients can be determined by differential calculus.

Example: For sums $y = ax_1 + bx_2$ and differences $y = ax_1 - bx_2$

$$u(y)^2 = a^2 u(x_1)^2 + b^2 u(x_2)^2$$

For products $y = cx_1x_2$ and and quotients $y = cx_1/x_2$

$$\left(\frac{u(y)}{y}\right)^2 = \left(\frac{u(x_1)}{x_1}\right)^2 + \left(\frac{u(x_2)}{x_2}\right)^2$$

In all other cases it is more appropriate to approximate the sensitivity coefficients by finite differences. Both these calculations and the combination by squared addition can be carried out very convenienty using a spreadsheet program, see Section A.4.1.

The covariances $u(x_i,x_k)$ in Eqs. (3.11) are closely connected with the standard uncertainties of the input quantities concerned as follows:

$$u(x_i, x_k) = r(x_i, x_k) \cdot u(x_i) \cdot u(x_k) \tag{3.13}$$

Here $r(x_i,x_k)$ is the so-called coefficient of correlation; its value is between -1 and 1. The value 1 indicates that the input quantities vary sympathetically while -1 indicates that they vary antagonistically, the value 0 indicates the absence of correlation. If all input quantities are totally correlated (r = 1), the combined standard uncertainty results as a linear sum $u(y) = \Sigma u_i$ of uncertainty contributions. In the case of completely uncorrelated input quantities the uncertainty contributions are added quadratically as $u(y)^2 = \Sigma u_i^2$. The squared addition usually results in considerably smaller values for the combined standard uncertainty u(y) than the linear addition. Therefore, linear addition can be used for worst-case estimation of combined standard uncertainties, without checking for correlations. Linear addition is not suitable as a procedure for the determination of uncertainties to be used as input data for the determination of the uncertainty of other quantities, because it usually overestimates the combined standard uncertainty.

Step 7: Definition of coverage factors

The uncertainty of the result can be specified alternatively either as a standard uncertainty u(y) or as an expanded uncertainty $U(y) = k \times u(y)$, i.e. as a product of the standard uncertainty and a suitably chosen coverage factor.

The expanded uncertainty is selected in order to define a range which is expected to contain the true value of the result with high probability.

If there are no comprehensible reasons for another choice, a value between 2 and 3 should be selected for k; k = 2 is recommended as a default value. If sufficient knowledge about the probability distribution of the result is available, then k can be calculated as a "confidence factor" at a fixed confidence level. For this purpose a confidence level of 0.95 (95 %) is recommended.

For such calculations the number of effective degrees of freedom is needed. It can be calculated from the standard uncertainties and the degrees of freedom of the distributions for the values of the input quantities, see GUM, Section G.4.

3.4 Instructions on the Use of Uncertainty Budgets

The analytical-computational determination of measurement uncertainty based on a detailed uncertainty budget is particularly suitable for measurement procedures with a broad application range, i.e. with a substantial range of variation of the measuring objects and the measuring conditions. Then it is worth effort to compile a detailed uncertainty budget, in which the measurement uncertainty is calculated as a function of the relevant influence quantities – in particular the property of the measuring objects and the measuring conditions.

For measurement procedures with a narrow range of application – measuring objects with a small range of variation, standardised measuring conditions – procedures described in Sections 4 and 5 for the estimation of measurement uncertainty using within-laboratory validation data and inter-laboratory comparison data offer a good alternative.

Uncertainty budgets are valuable diagnostic tools in the development and optimisation of measurement procedures. For this purpose the following form of Eq. (3.10) is particularly well suited (for simplification without correlations):

$$1 = \sum_{i=1}^{N} \frac{u_i^2}{u(y)^2} \tag{3.14}$$

Variance terms $u_i^2 / u(y)^2$ indicate which influence quantities contribute significantly to the combined uncertainty of the measurement result, and which influence quantities contribute only marginally. It is only worth expending effort to improve the accuracy for quantities with significant influence, while this would be a wasted effort for quantities with marginal influence.

Another useful form of the principal equation for uncertainty propagation (for simplification without correlations) is:

$$\left(\frac{u(y)}{y}\right)^2 = \sum_{i=1}^{N} d_i^2 \left(\frac{u(x_i)}{x_i}\right)^2 \qquad \text{with} \qquad d_i = \frac{c_i \cdot x_i}{y} \tag{3.15}$$

The coefficients d_i indicate, how strongly the relative uncertainty of an influence quantity affects the relative uncertainty of the result.

3.5 Worst-case Estimation

The procedure described in Section 3.3 aims to determine the measurement uncertainty with appropriate accuracy. In individual cases, however, a worst-case estimate (i.e. an upper bound) can be of interest instead of an exact value for the measurement uncertainty, e.g. if the magnitude of the measurement uncertainty plays only a subordinated role in further use of the result, or if compliance is to be ensured with given specifications or limit values.

For worst-case estimation of measurement uncertainty the procedure described in Section 3.3 can be simplified as follows:

- The uncertainty contributions $u_i = c_i \times u(x_i)$ of the input quantities are added linearly; the correlation contributions $u_{ik} = c_i \times c_k \times u(x_i,x_k)$ are omitted.
- In the uncertainty contributions $u_i = c_i \times u(x_i)$ of the input quantities, maximum values of the possible errors $|\Delta x_i|_{max}$ may be used instead of the standard uncertainties $u(x_i)$.

With these simplifications the following two equations are obtained, which can be used alternatively for the calculation of worst-case estimates of measurement uncertainty.

$$|\Delta y|_{max} = \sum_{i=1}^{N} \left|\frac{\partial F}{\partial x_i}\right| \cdot u(x_i) \tag{3.16}$$

$$|\Delta y|_{max} = \sum_{i=1}^{N} \left|\frac{\partial F}{\partial x_i}\right| \cdot |\Delta x_i|_{max} \tag{3.17}$$

4 Estimation of Measurement Uncertainties using Within-laboratory Validation and Quality Control Data

4.1 General

A direct method for measurement uncertainty determination is to apply the measurement procedure concerned to appropriate reference objects (standards, material measures, reference materials) and compare the results obtained under within-laboratory reproducibility conditions (see Section 2.2) with the known reference values. A variant, which follows the same principle to a large extent, consists of applying the measurement procedure in parallel with a reference procedure to suitable measuring objects and comparing the results of the procedure to be evaluated with those of the reference procedure. In both variants the measurement uncertainty is determined according to the basic principle *accuracy* = *trueness* + *precision* from characteristic values of trueness (estimates of bias) and characteristic values of precision (estimates of random variability).

The protocol described below consists of the following steps:

- Investigation of precision;
- Investigation of trueness (bias);
- Correction of bias (if significant);
- Determination of measurement uncertainty (including correction terms).

In Section 4.2 the protocol using one single reference object is described as the simplest case. If more than one reference object is needed for technical reasons, e.g. for the determination of uncertainty over a wide measuring range, then the protocol described here should be extended accordingly. Protocols suitable for this purpose are described in Section 4.3.

Normally, investigations of precision and bias of a measurement procedure are carried out on a regular basis (and additionally if necessary). It is important to ensure that the data of a current investigation are comparable with the data from preceding investigations:

- If the data are compatible with one another, they can be combined in order to improve the statistical basis of the estimated values concerned (average deviations, average recovery rates and their standard deviation).
- Otherwise the data comparison may be used as a diagnostic tool to resolve the observed discrepancies.

For this reason, measurements on reference objects should always be carried out and be evaluated in the same manner – without considering corrections determined beforehand.

4.2 One-Point Protocol

The protocol described here is applicable only if it can be taken for granted that the result obtained on the reference object can be considered as a representative point for the entire measuring range (in other words for all of the measuring objects and/or measuring tasks). Otherwise either the measuring range must be restricted accordingly or a multi-point protocol described in Section 4.3 should be used.

The reference object will be measured repeatedly (at least n = 6 times) under appropriate within-laboratory reproducibility conditions (see Section 2.2), which correspond to those employed in normal operation. Concerning these measurements, the following quantities will be used later on:

x_{ref}	Reference value of the measurand;
$u(x_{ref})$	Standard uncertainty of the reference value;
x_{meas}	Measured value obtained by the measurement procedure under investigation;
$\overline{x}_{meas}$	Mean value of n measured values x_{meas};
s_{meas}	Standard deviation of n measured values x_{meas};
Δ	Mean deviation ($\Delta = \overline{x}_{meas} - x_{ref}$) from the reference value;
Q	Mean recovery rate ($Q = \overline{x}_{meas} / x_{ref}$) of the reference value.

The first step is to investigate whether the standard deviation of the measurement series is compatible with the previously determined and monitored standard deviation of the measurement procedure (Section 4.2.1). Subsequently, the mean value of the measurement results is compared with the reference value in order to investigate potential bias. The bias observed will be assessed as "inacceptable", "significant but acceptable" or "insignificant" (Section 4.2.2). Appropriate actions will be taken (Section 4.2.3) depending on the results of the assessment:

Result	Inacceptable	Significant but acceptable	Insignificant
Action	Examine and amend the measurement procedure to remove/reduce bias	Apply a correction for bias or Introduce an additional uncertainty contribution to account for uncorrected bias	Introduce an additional uncertainty contribution to account for uncorrected bias

Table 3.1 Results and actions with regard to bias

As a result of the investigation one obtains an estimate of the uncertainty of the measurement procedure (including corrections, where applicable) (Section 4.2.3).

4.2.1 Investigation of Precision

As a preliminary investigation the precision of the measurement procedure is determined under within-laboratory reproducibility conditions (see Section 2.2), which should correspond to the normal operation conditions. This may be done using the standard deviation obtained from regular measurements on a suitable measuring object (precision control chart), or an appropriate pooled standard deviation when several measuring objects or several measuring instruments are utilised. This precision will thereafter be called "procedural precision". For the associated standard deviation the designation "procedural standard deviation" and the symbol s_V will be used.

Note: Combination (pooling) of two standard deviations is performed as follows:

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Here n_1 and n_2 are the numbers of measured values, from which s_1 and s_2 were calculated.

The standard deviation s_{meas} for the measurement series on the reference object should agree with the standard deviation s_V of the procedure, or at least s_{meas} should not be significantly larger than s_V. In case of doubt this can be checked using the F-test.

Note: The F-test examines whether two standard deviations differ significantly. To this end the squared quotient $(s_{>} / s_{<})^2$ of the larger and the smaller of the two standard deviations is compared with the tabulated value of the F-distribution for the respective degrees of freedom and the desired level of significance. The F-test is described in almost every text book on statistical data evaluation, see e.g. A. Bowker, G. Liebermann, *Engineering Statistics*, 2nd edition, Prentice Hall Inc. (1972).

4.2.2 Investigation of Bias

If the precision at the reference object is compatible with the procedural precision as determined before, then the deviation of the measured values obtained on the reference object from the reference value is examined and assessed. In principle this can be done for each individual measured value. However, for the sake of simplicity the mean deviation, i.e. the deviation of the mean value will be examined here. It will be first checked whether the deviation of the mean value is acceptable or not.

An inacceptable deviation indicates serious deficiencies of the measurement procedure, which require a detailed investigation of all process steps and devices concerning sources of error, and appropriate corrective actions to eliminate or at least reduce the observed bias.

An acceptable deviation matches the expectation regarding the trueness of the procedure and does not require any revision of the measurement procedure.

If the deviation of the mean value is acceptable, then it will be tested for (statistical) significance.

A deviation is considered to be significant, indicating significant bias, if the magnitude (absolute value) of the deviation of the mean value $\bar{x}_{meas}$ from the reference value x_{ref} is larger than twice the standard uncertainty of this deviation,

$$\left|\bar{x}_{meas} - x_{ref}\right| > 2\sqrt{\frac{s_{meas}^{\,2}}{n} + u(x_{ref})^2} \qquad (4.1)$$

Otherwise the deviation is insignificant.

Note: Instead of checking whether the mean deviation Δ is significantly different from zero, it can be checked whether the mean recovery rate Q is significantly different from unity. These two tests are in principle equivalent.

4.2.3 Dealing with Observed Bias

Depending on the data, significant deviations will be either corrected, or taken into account in the uncertainty. Insignificant bias is not corrected but accounted for in the uncertainty.

Correction for Bias

In the case of significant bias a one-point correction is only reasonable if it can be taken for granted that the (absolute or relative) bias is constant over the entire measuring range.

If a constant absolute deviation can be expected, the mean deviation $\Delta = \bar{x}_{meas} - x_{ref}$ observed will be subtracted from the measurement result.

$$y_{corr} = y_{meas} - \Delta \qquad (4.2)$$

Here y_{meas} is the measurement result on a test object and y_{corr} is the corrected measurement result.

If a constant relative deviation can be expected, the correction will be carried out using the mean recovery $Q = \bar{x}_{meas} / x_{ref}$ as follows.

$$y_{corr} = \frac{y_{meas}}{Q} \tag{4.3}$$

The correction can be made either by adjustment of the measuring instrument (alignment of zero point and/or sensitivity) or by calculation.

The standard uncertainty of the corrected measurement result is calculated acording to the rules of uncertainty propagation (see Section 3.3). For a correction according to Eq. (4.2) this gives

$$u(y_{corr})^2 = s(y_{meas})^2 + u(\Delta)^2 = s_V^2 + \frac{s_{meas}^2}{n} + u(x_{ref})^2 \tag{4.4}$$

Here the procedural standard deviation s_V is used for the standard deviation of the uncorrected measurement result. For simplification s_{meas} can also be replaced by s_V.

For a correction according to Eq. (4.3) a corresponding equation applies for the relative standard uncertainty, $u_{rel}(y) = u(y)/|y|$.

$$u_{rel}(y_{corr})^2 = s_{rel}(y_{meas})^2 + u_{rel}(Q)^2 = s_{rel\_V}^2 + \frac{s_{rel\_meas}^2}{n} + u_{rel}(x_{ref})^2 \tag{4.5}$$

Here s_{rel_V} is the relative procedural standard deviation and $s_{rel\_meas} = s_{mess} / \overline{x}_{meas}$ the relative standard deviation of the measurement results obtained on the reference object. Also here s_{rel_meas} can be replaced by s_{rel_V} for simplification.

If the measurement result y_{meas} is a mean value of m individual measured values, then s_V^2 has to be replaced with s_V^2 / m in Eq. (4.4). The same applies to Eq. (4.5).

Allowance for Bias in the Uncertainty Budget

If direct transfer of the (absolute or relative) bias determined on the reference object to the test objects under investigation is doubtful, then one should rather refrain from a correction. Instead the (mean) deviation determined on the reference object should be included in the uncertainty budget. To this end the procedure of Lira and Wöger is recommended [reference at the end of this section]. This gives

$$u(y_{uncorr})^2 = u(y_{corr})^2 + (y_{corr} - y_{meas})^2 \tag{4.6}$$

In the equation above the first term accounts for the uncertainty that would be obtained if the correction were carried out, while the second term accounts for the observed deviation. This procedure also applies when the (mean) deviation on the reference object is insignificant and therefore no correction is carried out.

Based on Eq. (4.2) the following uncertainty budget is obtained:

$$u(y_{uncorr})^2 = s(y_{meas})^2 + u(\Delta)^2 + \Delta^2 = s_V^2 + \frac{s_{meas}^2}{n} + u(x_{ref})^2 + (x_{meas} - x_{ref})^2 \tag{4.7}$$

Eq. (4.7) assumes that the uncertainty is approximately constant in the measuring range considered. However, generally uncertainty increases with increasing values of the measurand. Assuming proportional growth, an analogue of Eq. (4.7) can be derived from Eq. (4.3) which is applicable when relative uncertainty is approximately constant. Often however both approaches will be doubtful. Then the following estimate, extrapolating the uncertainty determined on the reference object both to smaller and larger values of the measurand may be used.

$$u(y_{uncorr})^2 = s_V^2 + \frac{s_{meas}^2}{n} + u(x_{ref})^2 + (\overline{x}_{meas} - x_{ref})^2 \qquad \text{for } y_{meas} \le \overline{x}_{meas} \tag{4.8}$$

$$u(y_{uncorr})^2 = \left(\frac{y_{meas}}{\overline{x}_{meas}}\right)^2 \left(s_V^2 + \frac{s_{meas}^2}{n} + u(x_{ref})^2 + (\overline{x}_{meas} - x_{ref})^2\right) \qquad \text{for } y_{meas} > \overline{x}_{meas}$$

Here s_V is the procedural standard deviation valid for $y_{meas} \approx \overline{x}_{meas}$.

The extrapolation of the uncertainty determined on the reference object according to Eqs. (4.8) is based on the following empirical fact: With decreasing values of the measurand, uncertainty also decreases or remains constant at most. With increasing values of the measurand, uncertainty also increases, at most proportionally to the value of the measurand. However this rule does not apply without any exceptions. Therefore the applicability of the extrapolation using Eqs. (4.8) must be checked in cases of doubt.

Reference: Lira, I.H. und Wöger, W., *Evaluation of the uncertainty associated with a measurement result not corrected for systematic effects*, Meas Sci Technol 1998, **9**, p. 1010-1011.

4.3 N-Point Protocol (N ≥ 2)

4.3.1 Interpolation

If the conditions for the application of the one-point protocol are not fulfilled, several reference objects must be used for investigating bias and, if necessary, for determining a correction. If a linear relationship can be assumed between measurement errors and values of the measurand, two reference objects are sufficient for these purposes.

The quantities and symbols are the same as in Section 4.2, up to an additional index A and B to characterise quantities for the respective reference objects A and B. For simplicity it is assumed that the same number n of measured values is available for both reference objects.

Investigation of Precision

The test will be performed on both reference objects based on the protocol described in 4.2.1. If the reference values x_{Aref} and x_{Bref} are close to each other, the same procedural standard deviation s_V can be used at both reference points. Otherwise two appropriate standard deviations s_{AV} and s_{BV} must be determined and used in the protocol. If the procedural standard deviation increases proportionally to the value of the measurand, the test will be carried out using relative standard deviations. Then a single value s_{rel_V} of the relative procedural standard deviation is sufficient.

Investigation of Bias

The test is carried out on both reference objects under the protocol described in 4.2.2. If an inacceptable bias is found on a reference object, a detailed investigation of all process steps and devices is necessary for sources of error and appropriate corrective actions.

If the bias found on both reference objects is acceptable, it will be tested for significance. In case of significance on one or both reference objects, depending on available data, either a correction is carried out or the deviation determined is taken into account in the uncertainty budget. Insignificant bias on both reference objects is not corrected but accounted for in the uncertainty budget.

Correction for Bias

Correction of observed bias is carried out using two correction parameters p and q according to the following equation.

$$y_{corr} = p + q \cdot y_{meas} \tag{4.9}$$

The correction parameters are determined in such a way that y_{corr} for the two reference objects agrees with the corresponding reference values. The calculation gives

$$q = \frac{x_{Bref} - x_{Aref}}{\bar{x}_{Bmeas} - \bar{x}_{Ameas}} \tag{4.10}$$

$$p = \frac{1}{2}\left[\left(x_{Aref} + x_{Bref}\right) - q\left(\bar{x}_{Ameas} + \bar{x}_{Bmeas}\right)\right] \tag{4.11}$$

The standard uncertainty of corrected measurement results is calculated according to the rules of uncertainty propagation from the standard uncertainties and standard deviations $s(y_{meas})$, $u(x_{Aref})$, $u(x_{Bref})$, $s(\bar{x}_{Ameas}) = s_{Ameas}/\sqrt{n}$, $s(\bar{x}_{Bmeas}) = s_{Bmeas}/\sqrt{n}$ of the quantities involved as follows.

$$u(y_{corr})^2 = q^2 s(y_{meas})^2 + \left(\frac{\bar{x}_{Bmeas} - y_{meas}}{\bar{x}_{Bmeas} - \bar{x}_{Ameas}}\right)^2 \left(u(x_{Aref})^2 + p^2 \cdot \frac{s_{Ameas}^2}{n}\right)$$

$$+ \left(\frac{y_{meas} - \bar{x}_{Ameas}}{\bar{x}_{Bmeas} - \bar{x}_{Ameas}}\right)^2 \left(u(x_{Bref})^2 + p^2 \cdot \frac{s_{Bmeas}^2}{n}\right) \tag{4.12}$$

For the standard deviation $s(y_{meas})$ of uncorrected measurement results the appropriate procedural standard deviation will be used. Here it has to be considered whether the result is defined as a single measured value or as the mean of a specified number of measured values. If the measurement result is a single value, $s(y_{meas}) = s_V$ holds; if the measurement result is the mean of m values, $s(y_{meas}) = s_V / \sqrt{m}$ holds.

Allowance for Bias in the Uncertainty Budget

If a correction using Eqs. (4.9) - (4.11) is doubtful, or if the determined deviations are insignificant, then no correction is made. Instead, the deviations observed on the reference objects are included in the uncertainty budget. As in Section 4.2, Lira and Wöger's procedure is recommended:

$$u(y_{uncorr})^2 = u(y_{corr})^2 + (y_{corr} - y_{meas})^2 \tag{4.13}$$

For the purpose of this calculation Eqs. (4.9) - (4.12) have to be substituted into Eq. (4.13). Due to its complexity, the resulting equation for the uncertainty $u(y_{uncorr})$ will not be presented here.

4.3.2 Least Squares Fit

For a wide measuring range and a large variety of measuring objects, two reference objects will often not be enough. Already in a linear case, reliable estimation of bias and appropriate correction require at least three reference objects. The evaluation of such a multiple comparison follows a different pattern than in the preceding cases because the number of reference objects is larger than the number of correction parameters to be determined (least squares fit instead of interpolation).

In the following the determination a "correction line" is dealt with as the most frequent application. However for extended measuring ranges nonlinear "correction curves" may be necessary. Their computational treatment requires methods of nonlinear regression, e.g. using polynomials.

Following the N-point protocol, N reference objects are measured and the resulting set of data, consisting of the reference values and the corresponding measured values, is evaluated by means of linear regression. For this evaluation most often the standard form of the method of least squares can be used. The conditions for this are as follows:

- The uncertainty of the reference values is significantly smaller than the dispersion of the measured values.

- The dispersion of the measured values is approximately constant in the measuring range considered.
- Replicate measured values on the same reference object approximately follow a normal distribution.

For the calculation, the reference values are selected as values of the independent variable x_{ref} and the measured values as values of the dependent variable x_{meas}. The result is a best fit straight line, i.e. a linear function

$$x_{meas} = \alpha + \beta \cdot x_{ref} \tag{4.14}$$

If there is no bias, $\alpha = 0$ and $\beta = 1$ hold. If $\alpha \neq 0$, there is an additive bias, $\beta \neq 1$ indicates a multiplicative bias. If so required, bias correction of measurement results obtained on comparable measuring objects is carried our according to the following equation:

$$y_{corr} = \frac{y_{meas} - \alpha}{\beta} \tag{4.15}$$

The uncertainty of corrected measurement results is calculated according to the procedures described in Annex A.2 from the standard deviation $s(y_{meas})$ and the uncertainty of the parameters of the best-fit line, $u(\alpha)$, $u(\beta)$ and the covariance $u(\alpha\ \beta)$.

5 Estimation of Measurement Uncertainties using Inter-laboratory Comparison Data

5.1 Inter-laboratory Comparisons for Method Validation

For standard test procedures, trueness and precision are usually determined by an inter-laboratory comparison (see ISO 5725-2). Among the performance characteristics obtained in this manner, the so-called "reproducibility standard deviation" (s_R) is a suitable estimate for the measurement uncertainty. Since this already comprises systematic effects due to different ways of operation in the laboratories involved, an additional uncertainty contribution accounting for systematic effects is normally not necessary.

> *Note*: A “repeatability standard deviation“ alone, determined by an inter-laboratory comparison (s_r), or determined within the laboratory by replicate measurements at the same conditions, is not normally a suitable uncertainty estimate, since it excludes major uncertainty contributions.

In the International Technical Specification ISO/TS 21748 *Guide to the use of repeatability, reproducibility and trueness estimates in measurement uncertainty estimation* of May 2003 the exact conditions are identified under which a laboratory can use the reproducibility standard deviation s_R assigned to a standard test procedure as an estimate for the measurement uncertainty of results obtained using this procedure. Essentially the laboratory must prove

(a) that the tests are carried out in conformity with the standard, and in particular

(b) that the measuring conditions and measuring objects agree with those in the inter-laboratory comparison, and

(c) that for its implementation of the test procedure, trueness and precision are compatible with the inter-laboratory comparison data.

Requirement (c) means that the laboratory has to check trueness and precision (see Section 4) for compatibility with the inter-laboratory comparison data s_r and s_R. For this purpose for instance, the laboratory can perform replicate measurements on a suitable reference object.

If n_{lab} is the number of measurements, s_{lab} the standard deviation of the measurement series and $\Delta = \bar{x}_{lab} - x_{ref}$ the deviation of the mean value of the measurement series from the reference value, then compatibility is given if:

$$s_{lab} \approx s_r \qquad \text{and}$$

$$|\Delta| \le 2\sqrt{\frac{s_r^2}{n_{lab}} + \left(s_R^2 - s_r^2\right)}.$$

5.2 Inter-laboratory Comparisons for Proficiency Testing

If the laboratory has successfully participated in an inter-laboratory proficiency test, it may utilise the results for estimating the uncertainty for the test procedure used. Some simple approaches how this could be done are described below. A similar approach is given in the NORDTEST Technical Report 537 *Handbook for calculation of measurement uncertainty in environmental laboratories*.

Alternatively, proficiency tests offer the possibility to check the validity of measurement uncertainty estimates obtained otherwise.

a) *Reference value given*

A reference value (target value) x_{ref} is given by the organizer of the inter-laboratory comparison including the standard uncertainty $u(x_{ref})$. The laboratory has submitted n_{lab} measured values with a mean value $\bar{x}_{lab}$ and standard deviation s_{lab}. From these data, the following performance parameters are calculated, analogously to the calculation in Section 4.2:

- the difference Δ_{RV} between the mean value of the measured values of the laboratory and the reference value

  $$\Delta_{RV} = \bar{x}_{lab} - x_{ref}$$

- the standard uncertainty $u(\Delta_{RV})$ of this difference

  $$u(\Delta_{RV}) = \sqrt{\frac{s_{lab}^2}{n_{lab}} + u(x_{ref})^2}$$

In principle, these parameters could be utilised for a correction in accordance with Section 4.2, Eqs. (4.2) or (4.3). However, often the transfer of a bias determined in the "snapshot" of a single inter-laboratory comparison to other measurements will be doubtful. Then one should rather refrain from a correction. Instead, the (mean) deviation from the reference value should be included in the uncertainty budget as in Section 4.2 by Eqs. (4.7) or (4.8).

> *Note*: If a laboratory has already determined the measurement uncertainty for the measurement procedure under consideration, then the results of inter-laboratory comparisons can be used for checking this measurement uncertainty. To this end, the difference Δ_{RV} between the mean value of the laboratory values and the reference value is tested for significance. For this purpose the measurement uncertainty $u(\bar{x}_{lab})$ of this mean value is needed. Most often this will include uncertainty contributions due to systematic effects, with the consequence that the notorious factor $1/\sqrt{n}$ is usually <u>not</u> applicable, see Annex A.5.

If the laboratory has participated in repeated PT rounds (for comparable tests), the uncertainty estimates obtained from the individual rounds should be compared and combined, if compatible. Combination would be done by pooling (taking mean squares) of the uncertainty estimates concerned. This approach is used in the NORDTEST Handbook, where results from at least 6 repeated PT rounds are required.

b) Reference value not given

In this case, as a substitute, a reference value is determined from the results of the participants by the organiser of the inter-laboratory comparison. Usually this will be the mean value $\langle x_{RV} \rangle$ of the results of all participants (outliers removed if necessary) with an uncertainty given by the standard deviation $s_{RV} / \sqrt{n_{RV}}$ of this mean value. Here s_{RV} is the standard deviation of the laboratory means which contribute to the reference value, and n_{RV} is their number. From these data the following parameters of the (limited) trueness control can be calculated:

- the difference Δ_{RV} between the mean value of the measured values of the laboratory and the (substitute) reference value

$$\Delta_{RV} = \bar{x}_{lab} - \langle x_{RV} \rangle$$

- the standard uncertainty $u(\Delta_{RV})$ of this difference

$$u(\Delta_{RV}) = \sqrt{\frac{s_{lab}^2}{n_{lab}} + \frac{s_{RV}^2}{n_{RV}}}$$

Based on these data, the same approach is applied as in case a), but due to lack of assurance of the trueness of the (substitute) reference value, a correction should not be made. Instead, the (mean) difference between the results of the laboratory and the reference value should be included in the uncertainty budget as in Section 4.2 according to Eqs. (4.7) and (4.8).

c) *Procedure-specific Inter-laboratory Comparison*

Procedure-specific inter-laboratory comparisons are often evaluated according to ISO 5725-2. Then, under suitable conditions (see Section 5.1), the reproducibility standard deviation can be used directly as an estimate of the measurement uncertainty. It accounts for both random and systematic influences, as far as these are due to different ways of operation of the laboratories involved, but not for “method bias” conditional upon the measurement procedure.

Other evaluations may require the standard deviation of the individual measured values of all participants (outliers removed if necessary) instead of the reproducibility standard deviation. If only the standard deviation of the participant mean values is available, this can be combined with the standard deviation determined in the laboratory under within-laboratory reproducibility conditions.

5.3 Inter-laboratory Comparisons for Reference Material Certification Inter-laboratory Comparisons for Reference Material Certification

If the laboratory has participated successfully with the procedure concerned in an inter-laboratory certification study, the difference between the result of the laboratory – usually the mean value obtained from a number of replicate measurements – and the certified value can be used for the evaluation of trueness.

The procedure is in principle the same as in Section 5.2, whereby the certified value and its uncertainty take over the role of the reference value and its uncertainty. The deviation from the certified value is considered acceptable if the result of the laboratory is included in the calculation of the certified value, but larger deviations may also be acceptable in view of routine applications. As to how the deviation should be taken into account, both correction and inclusion in the uncertainty budget may be considered, but in case of doubt the approach to include an additional allowance in the uncertainty budget should be preferred.

6 Hybrid Strategies for Evaluation of Measurement Uncertainties

If influential factors occur in a measurement that were not considered during validation of the measurement procedure (within the laboratory or in the inter-laboratory comparison), the estimate of the uncertainty determined in validation must be supplemented accordingly. This amounts to a hybrid strategy for uncertainty evaluation where, as far as possible, the combined effect of influential factors is determined using data from validation studies, and necessary additions are made by modelling the effect of residual factors on the result and using uncertainty propagation (squared addition). This strategy combines the use of existing data from validation studies with the flexibility of the model-based evaluation of individual uncertainty contributions.

If the reproducibility standard deviation s_R determined in an inter-laboratory comparison is used as a basis for estimating the uncertainty of the results obtained using a standard test procedure (see Section 5.1), and if the test conditions or the test objects substantially deviate from those in the inter-laboratory comparison, the effect of these deviations must be estimated and combined with the reproducibility standard deviation. For this purpose the following schematic equation applies:

$$u_{comb} = \sqrt{s_R^2 + \sum u_{other}^2}$$

Conversely, an uncertainty budget can be checked for completeness by comparing the calculated uncertainty with the dispersion of the results of replicate measurements, also including bias if suitable reference objects are available, and supplemented as necessary.

In any such uncertainty evaluation, an essential step is a detailed analysis of the influence quantities (sources of uncertainty) with the aim of differentiating those influence quantities, whose uncertainty contributions are included in a given performance characteristic, e.g. a reproducibility standard deviation s_R, from those influence quantities which are not included and therefore must be accounted for in some other way.

> *Note*: The approach utilised in several examples of the EURACHEM/CITAC Guide *Quantifying Uncertainty in Analytical Measurement* to include a standard deviation accounting for the overall “procedural dispersion” in a detailed uncertainty budget is counterproductive in view of a comparison between uncertainty budget and precision data, and tends to overestimate the uncertainty by double counting.

7 Specification and Documentation of Measurement Uncertainty

Calculated or estimated measurement uncertainties are stated, together with the value y of the measurand, either as a standard uncertainty u(y) or as an expanded uncertainty $U(y) = k \times u(y)$. If the expanded uncertainty is given, then the coverage factor used also has to be stated and, as far as possible, the respective estimated level of confidence.

Both standard uncertainty and expanded uncertainty can be stated as absolute values or, divided by the absolute value of the measurand, as a relative value (e.g. in per cent).

Worst-case uncertainty estimates are stated, together with the value y of the measurand, as a numerical value $|\Delta y|_{max}$ or in the form of an inequality for the measurement error $|\Delta y|$. The specification must exclude any confusion with standard uncertainties or expanded uncertainties.

The evaluation of measurement uncertainty must be documented comprehensively and in sufficient detail as to enable traceability of all major steps. If substantial uncertainty contributions were not taken into account, this circumstance must be stated and explained.

References

This list specifies the standards and guides referred to in the main text. References for other publications are given directly in the text.

[1] *Guide to the Expression of Uncertainty in Measurement*
1st corr. Edition, ISO, Geneva 1995, ISBN 92-67-10188-9

[2] *International Vocabulary of Basic and General Terms in Metrology*
2nd Edition, ISO, Geneva 1993, ISBN 92-67-10188-9

[3] DIN 1319-1, *Grundlagen der Messtechnik – Teil 1:Grundbegriffe*
Fundamentals of metrology – Part 1: Basic terminology (German and English)

[4] DIN 1319-4, *Grundlagen der Messtechnik – Teil 4: Auswertung von Messungen, Messunsicherheit*
Fundamentals of metrology – Part 4: Evaluation of measurements, uncertainty of measurement (German)

[5] DIN 55350-13, *Begriffe der Qualitätssicherung und Statistik – Teil 13: Begriffe zur Genauigkeit von Ermittlungsverfahren und Ermittlungsergebnissen*
Concepts in quality and statistics – Part 13: Concepts relating to the accuracy of methods of determination and results of determination (German)

[6] ISO 3534-1, *Statistics – Vocabulary and symbols – Part 1: Probability and general statistical terms*

[7] ISO 5725-2, *Accuracy (trueness and precision) of measurement methods and results – Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method*

[8] ISO/TS 21748, *Guide to the use of repeatability, reproducibility and trueness estimates in measurement uncertainty estimation*

[9] EURACHEM/CITAC GUIDE: *Quantifying Uncertainty in Analytical Measurement*
2nd Edition, EURACHEM / CITAC 2000

[10] NORDTEST Technical Report 537, *Handbook for calculation of measurement uncertainty in environmental laboratories*, 2nd Edition, NORDTEST 2004

Annex

This annex describes common sources of uncertainty (A.1) and common data evaluation methods (A.2 – A.6).

In the Sections A.2 to A.6 a distinction will be made in the mathematical description between quantities/variables X, Y, etc. and their values x, y, etc.

A.1 Frequently Occurring Sources of Uncertainty

As a rule, various uncertainty sources and their contributions have to be considered when evaluating the measurement uncertainty for a measurement or test procedure. Common sources of uncertainty are listed in this annex. They can be divided into four groups:

1. Uncertainties, which depend on sampling / sample preparation
    a) Taking samples which only represent the measuring object to a limited extent
    b) Contamination / degradation of samples during sampling
    c) Contamination / degradation of samples during their physical processing
    d) Homogenisation (incompleteness)
    e) Contamination / degradation of samples during their storage
    f) Chemical sample digestion (incompleteness, contamination, interferences)
    g) Chemical sample preparation / separation techniques (incompleteness, contamination)
2. Uncertainty contributions, which depend on the properties of the investigated object
    a) Noise, instability of the investigated object (temporal change of relevant quantities)
    b) Degradation or ageing of the investigated object (temporal change of relevant quantities)
    c) Inhomogeneity / non-uniformity of the investigated object (spatial change of relevant quantities)
    d) Matrix effects / interactions
3. Uncertainty contributions, which depend on the measurement / test methods used
    a) Inadequate implementation or definition of the measurand (approximations, idealisations, hypotheses)
    b) Uncertainty of process parameters (e. g. ambient conditions) and of relevant influence quantities
    c) Neglected influence quantities (e. g. ambient temperature, ambient pressure, magnetic field strength)
    d) Limited instrument resolution, fuzziness or uncertainty of the position of discriminator thresholds etc.
    e) Limits of detection, limited sensitivity
    f) Instrumental noise and drift
    g) Random interferences (such as interfering fields etc.)
    h) Inadequate impedance matching and transmission / transduction of the measurand
    i) Instrumental dead time (error due to coincidences)

j) Instrumental dynamics (frequency response / overshooting / resonance)

k) Different perception / visualisation of the measurands

l) Data evaluation, numerical accuracy etc.

m) Uncertainty obtained from inter-laboratory comparisons

4. Uncertainty of reference values on which the measurement / test is based

   a) Uncertainty of certified values / calibration values

   b) Drift / degradation of reference values / reference materials

   c) Uncertainty of imported values from the literature (data compilations, scientific publications etc.)

   d) Uncertainty obtained from inter-laboratory comparisons

The various uncertainty contributions are not necessarily independent from each other. They are partly of random and partly of systematic character. Random effects contribute to the variation of individual results in replicate measurements. Associated uncertainties can be evaluated using statistical methods, e.g. as the experimental standard deviation of a mean value (type-A evaluation). Uncertainties due to systematic effects must be estimated using other suitable approaches or, in the case of imported values, determined from information given in the references concerned (type-B evaluation).

A.2 Uncertainty in Linear Calibration

A.2.1 General

In linear calibration in the sense of this section, a linear relationship is established between the values of a quantity X and the values of a second quantity Y, e.g. between

- Tensile strength and hardness of steel materials;
- Thermoelectric voltage and temperature in thermocouples;
- Response and analyte content in an instrumental analytical method;
- Measurement results and corresponding reference values in the validation of a test method.

The relationship established by calibration is used to calculate the values of one quantity (output quantity, dependent variable) from the corresponding values of the other quantity (input value, independent variable). Depending on whether X or Y is chosen as input quantity, a linear relationship can be expressed in two different forms, $Y = \alpha + \beta X$ and $X = \gamma + \delta Y$, respectively.

In the following it will be assumed that Y is determined as a function of X according to

$$Y = \alpha + \beta X \qquad \text{(A.2.1)}$$

i. e. that the task of calibration is to determine the parameters α (intercept) and β (slope) of such function.

When using the linear relationship established by calibration, two cases must be distinguished:

Direct calibration:	Y is the target quantity, whose uncertainty has to be determined, X is an input quantity of Y.
Indirect calibration:	X is the target quantity, whose uncertainty has to be determined, Y is an input quantity of X.

In the case of direct calibration the relationship $Y = \alpha + \beta X$ is used directly to calculate the target quantity Y. According to the rules of uncertainty propagation [see Section 3.3, Eq. (3.11)] the uncertainty u(Y) of the target quantity is obtained from the uncertainty u(X) of the input quantity X, the uncertainties $u(\alpha)$, $u(\beta)$ of the parameters α, β and their covariance $u(\alpha,\beta)$ as follows:

$$u(Y)^2 = \beta^2 u(X)^2 + u(\alpha)^2 + X^2 u(\beta)^2 + 2X u(\alpha,\beta) \qquad \text{(A.2.2)}$$

In the case of indirect calibration the relationship $Y = \alpha + \beta X$ is not used directly. Instead, the inverse relationship is used to calculate the target quantity X.

$$X = \frac{Y - \alpha}{\beta} \qquad \text{(A.2.3)}$$

The uncertainty u(X) of the target quantity is obtained as follows:

$$u(X)^2 = \frac{u(Y)^2 + u(\alpha)^2 + X^2 u(\beta)^2 + 2X u(\alpha,\beta)}{\beta^2} \qquad \text{(A.2.4)}$$

Note 1: If the parameters α and β are determined jointly (as usually done), then these quantities are correlated, because their uncertainty sources are the same: the calibration data (x_1,y_1), (x_2,y_2), …, (x_K,y_K). The covariance $u(\alpha,\beta)$ generally contributes significantly to the final uncertainty and therefore must not be disregarded.

Note 2: The uncertainty determined according to Eqs. (A.2.2) or (A.2.4) is reliable only if the validity of the regression model has been checked and confirmed. This can be done by statistical analysis of the residual dispersion of the calibration points around the best-fit line (e.g. F test).

The following sections deal with the determination of the parameters α, β and the evaluation of their uncertainties, i.e. the quantities $u(\alpha)$, $u(\beta)$ and $u(\alpha,\beta)$.

A.2.2 Determination of Intercept and Slope

The parameters α (intercept) and β (slope) are obtained by statistical evaluation of appropriate calibration data $\{(x_1, y_1), (x_2, y_2) \ldots, (x_K, y_K)\}$. In principle, two points (x_1, y_1), (x_2, y_2) are sufficient for the determination of a straight line. In order to detect potential measurement errors and as far as possible compensate for them, usually more than this minimum number of calibration points (x_i, y_i) (i = 1, 2, …, K) are used, and the intercept and slope of the calibration line are determined by linear regression. In most cases the standard version of the least squares method is used for this purpose. According to this method the parameters α and β are determined in such a way that the sum of squared deviations, $S = \Sigma [y_i - (\alpha + \beta x_i)]^2$ is a minimum. The solution of this optimisation problem is given by the following, well-known formulae for the estimates of the intercept α and the slope β.

$$\beta = \frac{Q_{xy}}{Q_{xx}} \qquad \text{(A.2.5)}$$

$$\alpha = \bar{y} - \beta \bar{x} \qquad \text{(A.2.6)}$$

The symbols in these equations are:

$\bar{x} = [\Sigma x_i]/K$	Mean value of $x_1, x_2, ..., x_K$;
$\bar{y} = [\Sigma y_i]/K$	Mean value of $y_1, y_2, ..., y_K$;
$Q_{xx} = \Sigma(x_i - \bar{x})^2$	Sum of squared deviations in x;
$Q_{xy} = \Sigma(x_i - \bar{x})(y_i - \bar{y})$	Sum of deviation products in x and y.

The sums run over i = 1, 2, ..., K.

The standard version of the least squares method is based upon the following assumptions:

- The uncertainty of the values of the independent variable X is negligible in relation to the dispersion of the values of the dependent variable Y.
- The dispersion of the Y values is constant in the calibration range.
- The values of Y for a fixed X follow a normal distribution.

Under these conditions the method provides optimum estimates for the parameters α and β. However the method is also applicable in cases of moderate deviations from these conditions (see standard statistical text books).

If the above conditions are not met even approximately, other regression methods have to be used. For many applications, modifications of the standard least squares method are suitable, e.g.

- The method of weighted least squares, if the dispersion of the Y values varies strongly in the calibration range. Here the squared deviations are weighted with the inverse variances of the calibration values.
- The generalised least squares method, if the uncertainty of the X values and the dispersion of the Y values are comparable. Here the sum of the squared deviations in X and Y, if necessary weighted with the inverse variances of the calibration values, is minimised.
- Robust regression methods, if there are substantial deviations from the normal distribution or calibration data are contaminated by outliers. Here medians and quantiles are used instead of mean values and standard deviations.

For detailed information on the above mentioned regression methods, reference should be made to specialist literature on regression techniques, e. g. the Section "Modelling of data" in: Press, W. H. et al.: *Numerical Recipes in Fortran*, 2. edition, Cambridge University Press (1992).

A.2.3 Evaluating the Uncertainty of Intercept and Slope

Two basically different methods are available for evaluating the uncertainty of the parameters α und β:

- Statistical analysis of the dispersion of the calibration points;
- Propagation of uncertainty of calibration data.

These two methods are described in the following.

Statistical Analysis

Statistical analysis makes only implicit use of the uncertainty of the calibration data. The first step is to check whether the dispersion of the calibration points around the calibration line is approximately constant. If this is the case, the so-called residual standard deviation s_R of the measured values y_i is determined according to the following equation:

$$s_R^2 = \sum \frac{[y_i - (\alpha + \beta x_i)]^2}{K - 2} \qquad \text{(A.2.7)}$$

Using this key quantity, the standard uncertainties $u(\alpha)$, $u(\beta)$ and the covariance $u(\alpha,\beta)$ are obtained as follows.

$$u(\beta)^2 = \frac{s_R^2}{Q_{xx}} \qquad \text{(A.2.8)}$$

$$u(\alpha)^2 = s_R^2\left[\frac{1}{K} + \frac{\bar{x}^2}{Q_{xx}}\right] \tag{A.2.9}$$

$$u(\alpha,\beta) = -s_R^2\,\frac{\bar{x}}{Q_{xx}} \tag{A.2.10}$$

For calculating the uncertainty of the target quantity in direct calibration according to Eq. (A.2.2)

$$u(Y)^2 = \beta^2 u(X)^2 + u(\alpha)^2 + X^2 u(\beta)^2 + 2Xu(\alpha,\beta)$$

or for indirect calibration according to Eq. (A.2.4)

$$u(X)^2 = \frac{u(Y)^2 + u(\alpha)^2 + X^2 u(\beta)^2 + 2Xu(\alpha,\beta)}{\beta^2}$$

the standard uncertainty of the respective input quantity, u(X) or u(Y), is required in addition.

In direct calibration the uncertainty u(X) of the input quantity is usually assumed to be negligibly small. Thus the standard uncertainty of the target quantity Y becomes

$$u(Y)^2 = s_R^2\left[\frac{1}{K} + \frac{(X-\bar{x})^2}{Q_{xx}}\right] \tag{A.2.11}$$

In indirect calibration, however, Y is the input quantity. The standard uncertainty u(Y) can be estimated by the residual standard deviation s_R of the calibration points. In doing so one has to take into account whether the input quantity Y is defined as a single value or as the mean of a specified number of values:

$u(Y) = s_R$ if Y is a single value;

$u(Y) = s_R/\sqrt{m}$ if Y is the mean of m independent values.

Thus in the first case (single value) the standard uncertainty of the target quantity X is

$$u(X)^2 = \frac{s_R^2}{\beta^2}\left[1 + \frac{1}{K} + \frac{(X-\bar{x})^2}{Q_{xx}}\right] \tag{A.2.12}$$

In the second case (mean value) the standard uncertainty is

$$u(X)^2 = \frac{s_R^2}{\beta^2}\left[\frac{1}{m} + \frac{1}{K} + \frac{(X-\bar{x})^2}{Q_{xx}}\right] \tag{A.2.13}$$

The uncertainty u(Y) according to Eq. (A.2.11), or u(X) according to Eqs. (A.2.12) or (A.2.13), reaches its maximum at the limits of the calibration range. If the calibration values x_1, x_2, …, x_K are approximately equidistant, this maximum value can be estimated as follows:

$$u(Y)^2 < s_R^2 \cdot \frac{4}{K} \tag{A.2.11a}$$

$$u(X)^2 < \frac{s_R^2}{\beta^2}\left(1 + \frac{4}{K}\right) \tag{A.2.12a}$$

$$u(X)^2 < \frac{s_R^2}{\beta^2}\left(\frac{1}{m} + \frac{4}{K}\right) \tag{A.2.13a}$$

These maxima can be used in order to estimate the contribution of calibration to the overall uncertainty for a measurement procedure, or for determining an upper bound of this overall uncertainty.

Analogous expressions apply for the variants of the standard least squares method as discussed in Section A.2.2.

With regard to literature research it should be noted that the symbols used in this section deviate from standard statistical symbols. Instead of $u(x)^2$ mainly var(x) and occasionally $s(x)^2$ are used, instead of u(x,y) mainly cov(x,y), and occasionally s(x,y).

Uncertainty Propagation

In the procedure discussed in the previous section, the calibration uncertainty (i. e. the uncertainty of intercept and slope) is derived from the dispersion of the calibration points around the calibration line. The uncertainty of the calibration data is not accounted for in this calculation. In contrast to that, in the procedure described in this section, the calibration uncertainty is traced back to the uncertainty of the calibration data.

In the uncertainty propagation approach, the uncertainties of the intercept and the slope of the calibration line and the covariance of these two parameters are calculated from the uncertainties $u(x_i)$, $u(y_i)$ of the coordinates of the calibration points. In addition, covariances $u(x_i,x_j)$ may have to be included if there are correlations between calibration standards. These calculations are performed using the following equations.

$$u(\alpha)^2 = \sum_i \left(\frac{\partial \alpha}{\partial x_i}\right)^2 u(x_i)^2 + \sum_i \left(\frac{\partial \alpha}{\partial y_i}\right)^2 u(y_i)^2 + \sum_{i \neq j} \left(\frac{\partial \alpha}{\partial x_i}\right)\left(\frac{\partial \alpha}{\partial x_j}\right) u(x_i, x_j) \qquad \text{(A.2.14)}$$

with an analogous equation for $u(\beta)$ and

$$u(\alpha,\beta) = \sum_i \left(\frac{\partial \alpha}{\partial x_i}\right)\left(\frac{\partial \beta}{\partial x_i}\right) u(x_i)^2 + \sum_i \left(\frac{\partial \alpha}{\partial y_i}\right)\left(\frac{\partial \beta}{\partial y_i}\right) u(y_i)^2 + \sum_{i \neq j} \left(\frac{\partial \alpha}{\partial x_i}\right)\left(\frac{\partial \beta}{\partial x_j}\right) u(x_i, x_j) \qquad \text{(A.2.15)}$$

Most often the third term in these equations can be put to zero, because the calibration values x_i of the independent variables are usually determined independently from each other. However, there are cases of practical relevance where this independence does not hold, e.g. in a dilution series of calibration solutions. If two "daughter solutions" descend from the same "parent solution", then an error in the composition of the parent solution propagates to the composition of the daughter solutions in a (positively) correlated fashion. This correlation must be taken into account by appropriate covariances.

Sensitivity coefficients $(\partial\alpha/\partial x_i)$, $(\partial\alpha/\partial y_i)$, $(\partial\beta/\partial x_i)$, $(\partial\beta/\partial y_i)$ in Eqs. (A.2.14) and (A.2.15) can be determined using differential calculus only in exceptional cases. Therefore these equations must usually be evaluated using numerical differentiation. For this purpose the procedure described in Section A.4 can be applied.

A.3 Modelling of Process Steps by Efficiencies and Increments

For the application of the uncertainty propagation law

$$u(Y)^2 = \sum_{i=1}^{N} c_i^2 u(X_i)^2 + 2\sum_{i=1}^{N-1} \sum_{k=i+1}^{N} c_i \cdot c_k \cdot u(X_i, X_k) \qquad \text{(A.3.1)}$$

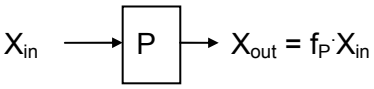
$$\text{with} \quad c_i = \left(\frac{\partial Y}{\partial X_i}\right) \qquad \text{(A.3.2)}$$

it is necessary to describe the result Y as a function of the relevant input quantities X_1, X_2, ..., X_N. This functional description is needed to determine the sensitivity coefficients $c_i = (\partial Y/\partial X_i)$.

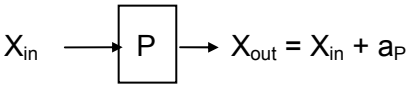
If the input quantities X_i are components (constitutive quantities) of the measurand – e.g. sample mass and sample volume in case of density – and if Y is given as a mathematical function of the input quantities, the derivatives $(\partial Y/\partial X_i)$ can in principle be calculated without

any problem. If the relation between the measurand and the input quantities is given by an algorithm, the derivatives can be calculated by means of numerical procedures (see Section A.4). However, the sources of uncertainty are often process steps – e.g. sampling, sample preparation but also corrections for observed bias – where it is unclear, how a functional description in terms of input quantities can be achieved. In the following, a description of process steps by efficiencies and increments is presented. This description is suitable for such process steps, where the input and output quantity are the same and differ only in their values, e.g. the analyte content of a sample before and after sample preparation.

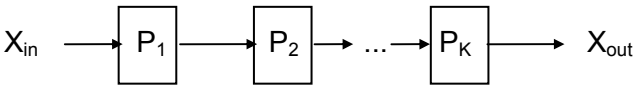
The description using efficiencies characterises the effect of the process step P by multiplying the input quantity with a dimensionless numerical factor f_P.



The description by increments characterises the effect of the process step P by adding an auxiliary term a_P (of the same dimension as the input quantity).



In simple cases the entire procedure can be described by a chain of process steps P_1, P_2, …, P_K :



Then, using efficiencies, the input/output relationship is given by

$$X_{out} = f_1 \cdot f_2 \cdot \ldots \cdot f_K \cdot X_{in} \tag{A.3.3}$$

or using increments

$$X_{out} = X_{in} + a_1 + a_2 + \ldots + a_K \tag{A.3.4}$$

Here X_{out} is the observed quantity while X_{in} is the actual quantity, whose value has to be determined. Thus the result $Y = X_{in}$ is obtained as a function of the quantity observed at the end of the chain, $X_{meas} = X_{out}$, and the characteristics for the individual process steps as follows

$$Y = \frac{X_{meas}}{f_1 \cdot f_2 \cdot \ldots \cdot f_K} \tag{A.3.5}$$

or

$$Y = X_{mess} - (a_1 + a_2 + \ldots + a_K) \tag{A.3.6}$$

The uncertainty of the result is obtained according to

$$\left(\frac{u(Y)}{Y}\right)^2 = \left(\frac{u(X_{meas})}{X_{meas}}\right)^2 + \left(\frac{u(f_1)}{f_1}\right)^2 + \ldots + \left(\frac{u(f_K)}{f_K}\right)^2 \tag{A.3.7}$$

or

$$u(Y)^2 = u(X_{meas})^2 + u(a_1)^2 + \ldots + u(a_K)^2 \tag{A.3.8}$$

These uncertainty budgets are also valid if the efficiencies f_i are equal to unity, or if the increments a_i are equal to zero, because the values unity and zero may also carry an uncertainty. Frequently the information available is not sufficient to specify a value $a_i \neq 0$ or $f_i \neq 1$, but a rough estimate of an applicable range can be given. Then putting $a_i = 0 \pm u(0)$ or $f_i = 1 \pm u(1)$ is appropriate.

In more complicated cases it may be better to work both with efficiencies and increments. Then the uncertainty budget has to be constructed accordingly.

A.4 Numerical Methods for Uncertainty Propagation

A.4.1 Finite Difference Calculation

The standard uncertainty u(Y) of a result Y, which depends on several input quantities X_1, X_2, ..., X_N, is made up from the standard uncertainties $u(X_1)$, $u(X_2)$, ..., $u(X_N)$ of the input quantities and, if relevant, the covariances $u(X_i,X_k)$ between correlated input quantities as follows.

$$u(Y)^2 = \sum_{i=1}^{N} u_i^2 + 2\sum_{i=1}^{N-1} \sum_{k=i+1}^{N} u_{ik} \tag{A.4.1}$$

Here the quantities u_i and u_{ik} are given by

$$u_i = \left(\frac{\partial Y}{\partial X_i} \right) \cdot u(X_i) \tag{A.4.2}$$

and

$$u_{ik} = \left(\frac{\partial Y}{\partial X_i} \right) \cdot \left(\frac{\partial Y}{\partial X_k} \right) \cdot u(X_i, X_k) \tag{A.4.3}$$

Application of these equations by hand calculation may encounter problems in the following steps.

(1) Determination of sensitivity coefficients ($\partial Y/\partial X_i$) using differential calculus;

(2) Determination and combination of uncertainty contributions for large numbers of influence quantities.

Sensitivity coefficients ($\partial Y/\partial X_i$) can only be determined explicitly by means of differential calculus, if the result Y is a simple mathematical function of the input quantities X_i, e.g. a sum or a product. For complicated functions the calculation of derivatives is laborious and error-prone. If the relation between the result Y and the input quantities is not given as a mathematical function, but by a computer program, then the calculation of the derivatives is not possible.

In such cases, the sensitivity coefficients ($\partial Y/\partial X_i$) can be approximated by quotients of finite differences. For the purpose of uncertainty propagation, this approach can be conveniently implemented by calculating an approximate value Δ_i for the product $u_i = (\partial Y/\partial X_i)u(X_i)$ in a single step according to

$$\Delta_i = Y\left(X_1,..., X_i + \frac{u(X_i)}{2},..., X_N \right) - Y\left(X_1,..., X_i - \frac{u(X_i)}{2},..., X_N \right) \tag{A.4.4}$$

Using these quantities, one obtains an approximate value for the standard uncertainty of the result as follows:

$$u(Y)^2 = \sum_{i=1}^{N} \Delta_i^2 + 2\sum_{i=1}^{N-1} \sum_{k=i+1}^{N} \Delta_i \cdot \Delta_k \cdot r(X_i, X_k) \tag{A.4.5}$$

Here $r(X_i,X_k)$ is the correlation coefficient of the quantities X_i and X_k (a quantity related to the respective covariance and standard uncertainties, see Section A.6).

$$u(X_i, X_k) = r(X_i, X_k) \cdot u(X_i) \cdot u(X_k) \qquad \text{(A.4.6)}$$

Unless input quantities are significantly correlated, the second sum on the right-hand side of Eq. (A.4.5) can be omitted.

The calculation described above can be carried out conveniently using a spreadsheet program. The procedure is described in detail in the EURACHEM/CITAC Guide *Quantifying Uncertainty in Analytical Measurement*, Annex E, Section E.2.

The determination of the correlation coefficients $r(X_i, X_k)$ between correlated input quantities, if required, is described in Section A.6.

A.4.2 Monte Carlo Simulation

Using the procedure described in the previous section the combined standard uncertainty can be calculated in a linear approximation by Eq. (A.4.1) for all practical applications. However, as already mentioned in the main text, the linear approximation can lead to significant errors if the relation between the result Y and an input quantity X_i is non-linear. Moreover, the probability distribution of the result Y can considerably deviate from a normal distribution, with the consequence that k = 2 substantially underestimates the coverage factor for 95 % coverage.

These problems can be avoided if, instead of the standard uncertainties, the probability distributions attributed to the input quantities are combined (propagated). In the Monte Carlo technique a suitable distribution (usually a normal distribution, a rectangular distribution or a triangular distribution) is attributed to each input quantity.

From these distributions a "random value" for each is simulated and a value of the target quantity is calculated from this set of input data. This procedure is repeated many times, so that a set of data is obtained for the target quantity which represents a random sample from the "potential" values of the target quantity as a function of variations in the input quantities according to their distribution. The mean value and the standard deviation of this random sample are estimates for the value of the target quantity and its standard uncertainty. In order to achieve reliable estimates, a high number of replicates is necessary (from 10^3); the required order of magnitude must usually be determined by trial.

The Monte Carlo technique however provides far more than an estimate for the target quantity and its standard uncertainty: an estimated distribution of values which are attributed to the target quantity, based upon the available information about the input quantities. In case of significant deviations from a normal distribution, the simulated distribution provides a more realistic confidence interval than $x \pm 2u(x)$, e.g. as the smallest interval, which contains 95 % of the distribution.

Application of the Monte Carlo technique for uncertainty evaluation is described in a number of publications, e.g. in Cox, M., Dainton, M., Harris, P., *Software Specifications for Uncertainty Calculation*, NPL Report CMSC 40/04 (2004), and commercial software for this purpose is available.

A.4.3 Software

Various computer codes are available for calculating measurement uncertainty by uncertainty propagation.

The Nordtest Technical Report 430 *Tools for the test laboratory to implement measurement uncertainty budgets* (1999) contains an overview of the generally available software at that time.

For up-to-date information internet search is recommended.

A.5 Uncertainty of Mean Values

Averaging, i.e. taking a mean value, is by far the most frequent operation in the evaluation of experimental data. Therefore uncertainty propagation in averaging deserves special attention. Most commonly however, unconsidered use is made of the well-known $(1/\sqrt{n})$ law, according to which the standard deviation of a mean value of n individual values is $(1/\sqrt{n})$ times the standard deviation of the individual values. However, this is only valid for uncorrelated (i.e. statistically independent) individual values. For correlated individual values the covariances between the individual values must be taken into account.

Correlations between the individual values of a data set (more exactly: between the errors of the individual values) occur whenever components of measurement error do not vary at random between the individual values of the data set but are constant or vary systematically. This can be investigated by a careful analysis of the components of measurement error or by statistical evaluation of suitable measurement series (see Section A.6).

A.5.1 General

The standard uncertainty (standard deviation) of the mean $\overline{y}$ of n independent individual values $y_1, y_2, ..., y_n$ is given by

$$u(\overline{y})^2 = \frac{1}{n^2}\sum_{i=1}^{n} u(y_i)^2 \qquad \text{(A.5.1)}$$

Here $u(y_i)$ are the standard uncertainties (standard deviations) of the individual values y_i. If all individual values originate from the same statistical distribution with a standard deviation of σ, and if this random dispersion is the only source of uncertainty, $u(y_i) = \sigma$ holds, giving

$$u(\overline{y})^2 = \frac{\sigma^2}{n} \qquad \text{(A.5.2)}$$

Thus the well-known $(1/\sqrt{n})$ law results:

$$u(\overline{y}) = \frac{\sigma}{\sqrt{n}} \qquad \text{(A.5.3)}$$

Consequently, the standard deviation of a mean value of n independent individual values approaches zero with an increasing sample size n.

However, if the individual values are correlated – e.g. if they have one or more uncertainty sources in common – the standard uncertainty of a mean value is given by

$$u(\overline{y})^2 = \frac{1}{n^2}\left[\sum_{i=1}^{n} u(y_i)^2 + 2\sum_{i=1}^{n-1}\sum_{k=i+1}^{n} u(y_i, y_k)\right] \qquad \text{(A.5.4)}$$

Here $u(y_i,y_k)$ are the covariances between the individual values.

If all values are equally correlated with each other, with $u(y_i) = \sigma$ and a correlation coefficient of r, the uncertainty of the mean becomes

$$u(\overline{y})^2 = \left(r + \frac{1-r}{n}\right)\cdot\sigma^2 \qquad \text{(A.5.5)}$$

$u(\overline{y})$ is considerably underestimated by $\sigma / \sqrt{n}$ even for moderate correlation: if r = 0.5, then $u(\overline{y}) \geq \sigma / \sqrt{2}$.

Correlations a. o. occur when, in addition to random variations, the individual values are affected by a bias, e.g. due to using the same calibration. Then the random deviations cancel out (the larger the sample size n, the more perfectly). The bias will however not be cancelled out by averaging and, with increasing sample size n, becomes the dominating part of the uncertainty of the mean value.

Example: A laboratory wants to use the reproducibility standard deviation s_R determined in the inter-laboratory comparison for a standardised measurement procedure. However, deviating from the standard in order to increase the accuracy for a special application, replicate measurements were carried out (n) and averaged. Then $u(\overline{y}) = s_R / \sqrt{n}$ may not be assumed because the reproducibilty standard deviation s_R consists of two components according to $s_R^2 = s_r^2 + s_L^2$. When averaging several measurement results of a single laboratory, only the repeatability component of the measurement error varies at random, while the laboratory deviation component remains constant. Therefore the standard uncertainty of a within-laboratory mean has to be calculated according to $u(\overline{y})^2 = s_r^2 / n + s_L^2$.

Summing up, it can be stated that Eqs. (A.5.1) and/or (A.5.2) and (A.5.3) may only be used if it is ensured that the individual values are uncorrelated, or if it can be proved that the contribution of the correlations is negligible. Otherwise the correlations must be investigated and taken into account quantitatively, if relevant. For this purpose the following two alternative procedures can be used:

(1) Backtracking to uncorrelated input quantities, uncertainty propagation with regard to those input quantities;

(2) Uncertainty propagation according to Eq. (A.5.4), with covariances evaluated according to Sections A.5.2 or A.6.

A.5.2 Correlation within a Measurement Series

This section deals with the problem of how to evaluate the uncertainty of a mean value $\overline{y}$ of a measurement series $y_1, y_2, ..., y_n$, where the individual values were obtained using the same measurement procedure on the same measuring object (or essentially identical different measuring objects).

Under these circumstances the standard uncertainty is the same for all individual values, i.e. $u(y_i) = u(y)$ holds for all y_i. However, apart from purely random variations from measurement to measurement, there are usually influences on the measurement which remain unchanged during a measurement series: e.g. calibration, measuring conditions, characteristics of the measuring objects. The measurement uncertainty u(y) thus consists of two components according to

$$u(y)^2 = u_{var}(y)^2 + u_{inv}(y)^2 \qquad \text{(A.5.6)}$$

Here $u_{var}(y)$ expresses the contributions of the influences varying from measurement to measurement, while $u_{inv}(y)$ stands for the influences which are steady (invariant) from measurement to measurement.

For $u_{var}(y)$ the procedural standard deviation s_V, i.e. the standard deviation under within-laboratory reproducibility conditions, is an appropriate estimate; as a substitute the standard deviation of the individual values of the measurement series can be used instead. If no separate estimate is available for $u_{inv}(y)$, but an estimate for the overall uncertainty u(y) (e.g. a reproducibility standard deviation s_R), the square root of the difference $u(y)^2 - s_V^2$ can be used for $u_{inv}(y)$:

$$u_{inv}(y) = \sqrt{u(y)^2 - s_V^2} \qquad \text{(A.5.7)}$$

The covariance u(y,y') between two individual values is obtained as

$$u(y, y') = u_{inv}(y)^2 = u(y)^2 - s_V^2 \tag{A.5.8}$$

Thus the uncertainty of the mean becomes

$$u(\bar{y})^2 = \frac{u_{var}(y)^2}{n} + u_{inv}(y)^2 = \frac{s_V^2}{n} + \left(u(y)^2 - s_V^2\right) \tag{A.5.9}$$

This equation shows that averaging will only reduce the uncertainty arising from random effects by a factor of 1 / √n, while the uncertainty accounting for systematic effects is left unchanged.

Correlation not only affects the uncertainty of mean values, but also of other combinations of measured values, which are obtained using the same measurement procedure on the same measuring object or comparable measuring objects. Thus e.g. for a difference

$$u(y - y')^2 = u(y)^2 + u(y')^2 - 2u(y, y') \tag{A.5.10}$$

If the measurement uncertainty is the same for y and y', Eq. (A.5.8) yields

$$u(y - y')^2 = 2s_V^2 \tag{A.5.11}$$

instead of $2u(y)^2$ as would be obtained without accounting for correlations. The reason for this gain of accuracy is that biases cancel out in the case of differences (and equally for quotients), like e.g. in difference weighing.

A.6 Evaluation of Covariances and Correlation Coefficients

A.6.1 General

For calculating the standard uncertainty u(Y) of a result Y, which depends on several input quantities X_1, X_2, ..., X_N, apart from the standard uncertainties $u(X_1)$, $u(X_2)$, ..., $u(X_N)$ of the input quantities, the covariances $u(X_i,X_j)$ between correlated input quantities X_i, X_j are needed.

Correlations have to be considered whenever two input quantities depend on each other or on a common third (maybe hidden) quantity or on several such quantities. This dependence can refer directly to the physical quantities themselves. Thus the mass fractions of the components of a mixture of several substances depend on each other, because their sum is equal to unity. More frequently however, the physical quantities concerned are independent from each other, but their values are not determined independently. That is the case when two quantities are determined in the same experiment – e.g. intercept and slope of a calibration line – or when the same standard is used in different measurements. Then the determined quantities depend on common quantities: the calibration data and/or the value of the measurement standard etc.

The covariance $u(X_i,X_j)$ between two input quantities X_i and X_j can be put to zero if

- X_i and X_j are independent physical quantities and their values have been determined independently from each other, or
- at least one of the two uncertainties $u(X_i)$, $u(X_j)$ is negligibly small.

If the information required for the determination of a covariance is not available and cannot be acquired at justifiable expenditure, then one has to fall back on rough estimations. The basis for this is the fact that $u(X_i, X_j)$ lies between $u(X_i)u(X_j)$ and $-u(X_i)u(X_j)$. If there are no further indications available for the magnitude of the covariance and not even for the sign, then $u(X_i, X_j) = 0$ can be assumed if overestimation and underestimation are to be equally avoided. However, if it is essential to avoid underestimation at the expense of potential

overestimation, then $u(X_i, X_j) = u(X_i)u(X_j)$ can be assumed. Less crude estimations are possible if ideas about the correlation mechanism are available.

Often correlation is due to a change of variables, where uncorrelated original variables were substituted by a more convenient set of derived variables (smaller number, more appropriate system characterisation etc.). This advantage is however often paid for by the occurrence of correlations. If this leads to difficulties in the uncertainty evaluation, it is advisable to revert from the correlated input quantities to the original uncorrelated ones, or to seek for alternative uncorrelated quantities.

If the determination of covariances cannot be avoided, two basic approaches are available:

- Uncertainty propagation with regard to the common variables;
- Experimental determination from parallel measurements.

A.6.2 Uncertainty Propagation

If two quantities X_i and X_j depend on the same uncorrelated quantities Z_1, Z_2, ..., Z_K, the covariance between them is given by

$$u(X_i, X_j) = \sum_{k=1}^{K} \left(\frac{\partial X_i}{\partial Z_k} \right) \cdot \left(\frac{\partial X_j}{\partial Z_k} \right) \cdot u(Z_k)^2 \qquad \text{(A.6.1)}$$

If the quantities Z_1, Z_2, ..., Z_K are correlated too, their covariances also have to be taken into account. In this case the calculation is performed according to

$$u(X_i, X_j) = \sum_{k=1}^{K} \left(\frac{\partial X_i}{\partial Z_k} \right) \cdot \left(\frac{\partial X_j}{\partial Z_k} \right) \cdot u(Z_k)^2 + 2 \sum_{k=1}^{K-1} \sum_{l=k+1}^{K} \left(\frac{\partial X_i}{\partial Z_k} \right) \cdot \left(\frac{\partial X_j}{\partial Z_l} \right) \cdot u(Z_k, Z_l) \qquad \text{(A.6.2)}$$

A.6.3 Parallel Measurements

If replicate mesurements of two quantities X_i and X_j were carried out jointly under specified conditions, yielding n paired values (x_{i1}, x_{j1}), (x_{i2}, x_{j2}), ..., (x_{in}, x_{jn}) , the values of X_i and X_j may be estimated by the mean values $\overline{x}_i$ and $\overline{x}_j$ of x_{i1}, x_{i2}, ..., x_{in} and of x_{j1}, x_{j2}, ..., x_{jn}, respectively. The covariance $u(X_i, X_j)$ between X_i and X_j can be estimated correspondingly by the experimental covariance of the mean values, $s(\overline{x}_i, \overline{x}_j)$. It is given by

$$s(\overline{x}_i, \overline{x}_j) = \frac{1}{n(n-1)} \cdot \sum_{p=1}^{n} (x_{ip} - \overline{x}_i)(x_{jp} - \overline{x}_j) \qquad \text{(A.6.3)}$$

The standard uncertainties $u(X_i)$ and $u(X_j)$ can then be estimated by the experimental standard deviations $s(\overline{x}_i)$ and $s(\overline{x}_j)$ of the mean values, which are given as follows:

$$s(\overline{x}_i)^2 = \frac{1}{n(n-1)} \cdot \sum_{p=1}^{n} (x_{ip} - \overline{x}_i)^2 \qquad \text{(A.6.4)}$$

with an analogous equation for $s(\overline{x}_j)$.

If the values for $u(X_i)$, $u(X_j)$ and $u(X_i,X_j)$ are used as procedural uncertainties for future measurements of the quantities X_i and X_j, then one has to take into account whether single values or mean values are determined. For single values, the estimates for $u(X_i,X_j)$ and $u(X_i)^2$ and $u(X_j)^2$ determined according to (A.6.3) and (A.6.4) have to be multiplied by a factor n, for mean values of m individual values by a factor n/m.

> *Note:* If the number n of measured values is small, the estimates obtained from Eqs. (A.6.3) and (A.6.4) are highly inaccurate. Then it may be better to use estimates based on experience or simple models.

A.6.4 Correlation Coefficients

Correlation coefficients $r(X_i,X_j)$ can be obtained from the covariances $u(X_i,X_j)$ by normalisation with the respective standard uncertainties:

$$r(X_i, X_j) = \frac{u(X_i, X_j)}{u(X_i) \cdot u(X_j)} \qquad \text{(A.6.5)}$$

If needed, estimates of correlation coefficients can be determined from estimates of covariances and standard uncertainties, e.g. by combining Eqs. (A.6.3) and (A.6.4). If information about the sign and the magnitude of the correlation is available, correlation coefficients can be estimated directly on this basis.

Example: Correlations between individual values of a measurement series (see Section A.5) can only be positive, i.e. $0 < r \leq 1$. In absence of more detailed information, $r = 0.5$ can be assumed.