



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

Technical Report No. 1/2007  
March 2007

Technical Report

**Measurement uncertainty revisited:  
Alternative approaches to uncertainty  
evaluation**

## **Impressum**

EUROLAB Technical Report 1/2007  
“Measurement uncertainty revisited:  
Alternative approaches to uncertainty evaluation”

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](mailto:eurolab@lne.fr)  
URL: [www.eurolab.org](http://www.eurolab.org)

## Editorial notes

This report was prepared by a Eurolab expert group on measurement uncertainty, affiliated to the Eurolab Technical Committee on Quality Assurance in Testing (TCQA), that was established for this purpose and disbanded after completing the work. The group held a series of meetings in Berlin from October 2004 to November 2006. The members of this group were:

Stephen Ellison, LGC Limited, United Kingdom

Manfred Golze, Bundesanstalt für Materialforschung und -prüfung (BAM), Germany

Werner Hässelbarth, Bundesanstalt für Materialforschung und -prüfung (BAM), Germany (Convenor)

Ulf Hammerschmidt, Physikalisch-Technische Bundesanstalt (PTB), Germany

Wilfried Hinrichs, Materialprüfanstalt für das Bauwesen, Germany

Ulrich Kurfürst, Fulda University of Applied Sciences, Germany

Bertil Magnusson, SP Technical Research Institute of Sweden

Teemu Näykki, Finnish Environment Institute, Finland

Marc Priel, Laboratoire National d'Essais (LNE), France

Burkhard Peil, Deutsche Akkreditierungsstelle Chemie GmbH (DACH), Germany

Pedro Rosario Ruiz, Real Casa de Moneda, Spain

Anita Schmidt, Bundesanstalt für Materialforschung und -prüfung (BAM), Germany

The draft technical report was approved by the EUROLAB Technical Committee for Quality Assurance in Testing (TCQA) at its meeting in Prague on 19 October 2006 and by the EUROLAB General Assembly in Bristol, United Kingdom, on 14 March 2007. □□



## Contents

Introduction.....	7
Chapter 1: Review of uncertainty evaluation.....	9
1.1 Approaches to uncertainty evaluation .....	9
1.2 Uncertainty data obtained from the various approaches .....	14
Chapter 2: Use of PT data in measurement uncertainty evaluation.....	20
2.1 Drawbacks and advantages of using PT data instead of reference materials.....	20
2.2 Use of PT data for evaluating the overall performance of laboratories .....	20
2.3 Use of PT data for estimating bias for a single laboratory.....	21
2.4 The Nordtest approach to estimating uncertainty from PT data .....	21
2.5 Other uses of PT data .....	23
Chapter 3: Verification, adjustment and comparison of uncertainty estimates.....	24
3.1 Verification of uncertainty estimates.....	24
3.2 Adjusting uncertainty estimates.....	27
3.3 Comparison of uncertainty estimates .....	28
Chapter 4: Examples.....	30
Example 1: Uncertainty evaluation for the determination of lead in a biological tissue	31
Example 2: Uncertainty evaluation for the determination of the Flakiness index of aggregates .....	33
Example 3: Uncertainty evaluation for the measurement of the aperture size of wire screen products.....	35
Example 4: Comparing and combining calculated uncertainty and experimental variability – a study of sample preparation uncertainty in chemical analysis.....	37
Example 5: Ammonium determination in water - Verification of uncertainty estimates ....	41
Example 6: Measurement uncertainty in optical emission spectrometry.....	43
Example 7: Rockwell hardness testing.....	47
Example 8: Determination of cadmium and phosphorous in agricultural top soil – comparison of evaluation methods focussed on uncertainty from sampling .....	51
Example 9: Pesticide residues in foodstuffs.....	54
Example 10: Uncertainty evaluations in the environmental sector – summary of a comprehensive study .....	57
Conclusions and recommendations .....	59
Annex: References and further reading .....	60



## Introduction

This report is the 3<sup>rd</sup> in a series of EUROLAB Technical Reports on measurement uncertainty in quantitative testing. The first in this series, (No. 1/2002), is an introductory text for newcomers, which was recently supplemented by a comprehensive technical guide for more experienced users (No. 1/2006). As a common feature, both reports emphasize the use of “empirical approaches” to uncertainty evaluation as an alternative to the modelling approach and provide guidance for this purpose. Meanwhile such alternatives have proliferated considerably and are increasingly recognised. Therefore this report is now focussed on reviewing and comparing the currently available approaches for evaluating measurement uncertainty of quantitative test results and giving some examples.

After more than ten years since publication of the 1<sup>st</sup> edition, the *Guide to the Expression of Uncertainty in Measurement*, known as the *GUM*, is acknowledged as the master document on measurement uncertainty throughout the testing community. The term “measurement uncertainty” is recognised to apply to all types of quantitative test results, and the GUM principles are fully accepted. Among others, these principles require that

- uncertainty evaluation is comprehensive, accounting for all relevant sources of measurement error;
- uncertainties arising from random and systematic effects are treated alike, i.e. are expressed and combined as variances of associated probability distributions;
- statistical evaluation of measurements (Type A) and alternative techniques, based on other data / information (Type B), are recognised and utilised as equally valid tools;
- uncertainties of final results are expressed as standard deviations (standard uncertainty) or by multiples of standard deviations (expanded uncertainty) with a specified numerical factor (coverage factor).

However, when it comes to evaluating the uncertainty of the results for a (quantitative) test procedure, the GUM is often criticised as 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 variances. 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. Alternative “empirical” approaches have only recently received greater attention. They are based on whole-method performance investigations designed and conducted so 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. Focussing on the first principle from the bullet points above, the basic requirements for any valid uncertainty evaluation are

- a clear definition of the measurand, i.e. the quantity to be measured,
- a comprehensive specification of the measurement procedure and the test items, and
- a comprehensive analysis of the effects impacting the measurement results.

Given a comprehensive list of relevant effects/uncertainty sources, uncertainty evaluation may be carried out using various different approaches. They range from individual quantification and combination of input uncertainties to collective quantification, e.g. using a reproducibility standard deviation for a standard test procedure.

Handling of uncertainty information requires due attention to the scope and the form of the data concerned. For example, the results obtained using “empirical” approaches normally refer to the typical performance of a specified test procedure on specified test objects, while uncertainty estimates obtained using the modelling approach most often refer to individual measurement results.

For ensuring that all relevant uncertainty sources are covered, error models developed in various testing fields are useful tools, e.g.

- hierarchical schemes such as the classification of measurement error according to repeatability error – run bias – laboratory bias – method bias (the “ladder of errors”), providing the basis for the definition and evaluation of various method performance characteristics: repeatability standard deviation, intermediate-precision standard deviation, reproducibility standard deviation and bias estimates (see ISO 5725 series and Thompson 2000, reference in section 1.2.2);
- schemes where measurement errors are attributed to the various parts of the measurement system, e.g. test item, measuring instrument, operator, method, environment (see e.g. Measurement Systems Analysis, reference [14] in the Annex).

Concerning empirical approaches for uncertainty evaluation, the use of the reproducibility standard deviation from an interlaboratory method validation study has recently been firmly established (see ISO/TS 21748). The use of within-laboratory data, that is, data from method validation studies and quality control carried out in the laboratory, is also widely recognised as a valid approach. Concerning the use of laboratory performance data from proficiency tests, some proposals have been published but the approach is still under debate.

Whatever approach is utilised, uncertainty evaluation is a difficult task, prone to mistakes. Several studies have shown that measurement uncertainty is often significantly underestimated. In the modelling approach e.g. major uncertainty contributions may be lacking, input uncertainties may be mis-estimated, and correlations may be overlooked. In the empirical approach, significant effects which have not been included in the experimental design for the method performance investigation, e.g. variations of test items or test conditions, will be missing in a (collaborative or within-laboratory) reproducibility standard deviation. Therefore extrapolation from the specific conditions of the performance investigation is a critical issue. Given the present lack of comparability and reliability in uncertainty evaluation in testing, the way forward is to compare uncertainty estimates obtained using different approaches. Beyond comparing data, the aim of such comparison should be to investigate whether the effects accounted for in either of the uncertainty estimates are essentially the same, or whether there are significant differences. Such comparisons, data-wise and source-wise, are by no means straight-forward, e.g. concerning the different performance characteristics in use and the systematic or random character of effects, and guidance is needed to avoid “comparing apples with pears”.

**Chapter 1** provides a summary of the current main approaches for uncertainty evaluation: (i) the modelling approach, (ii) the single-laboratory validation approach, (iii) the approach using inter-laboratory validation data, and (iv) the approach using proficiency-testing data. This chapter also includes a section discussing and comparing the uncertainty data obtained from the various approaches. Among those mentioned above, (i) – (iii) are meanwhile well covered by published papers and guides. However the use of PT data is still under debate and authoritative references are few. Therefore in **chapter 2** a full description of the approach using PT data is given. **Chapter 3** is devoted to technical issues pertinent to comparison, validation and revision of uncertainty estimates. As the core part of this report, **chapter 4** presents a range of examples. These are case studies from various testing fields, where different approaches were used to evaluate the relevant uncertainty and the results so obtained were compared. Some **conclusions and recommendations** complete the main body of the document. Finally the **Annex** presents a compilation of selected standards, guidelines, books and websites on measurement uncertainty.

## Chapter 1: Review of uncertainty evaluation

*This chapter presents a summary of the current main approaches for uncertainty evaluation: the modelling approach, the single-laboratory validation approach, the approach using inter-laboratory validation data, and the approach using proficiency-testing data. It also includes a section discussing and comparing the uncertainty data obtained from these various approaches.*

### 1.1 APPROACHES TO UNCERTAINTY EVALUATION

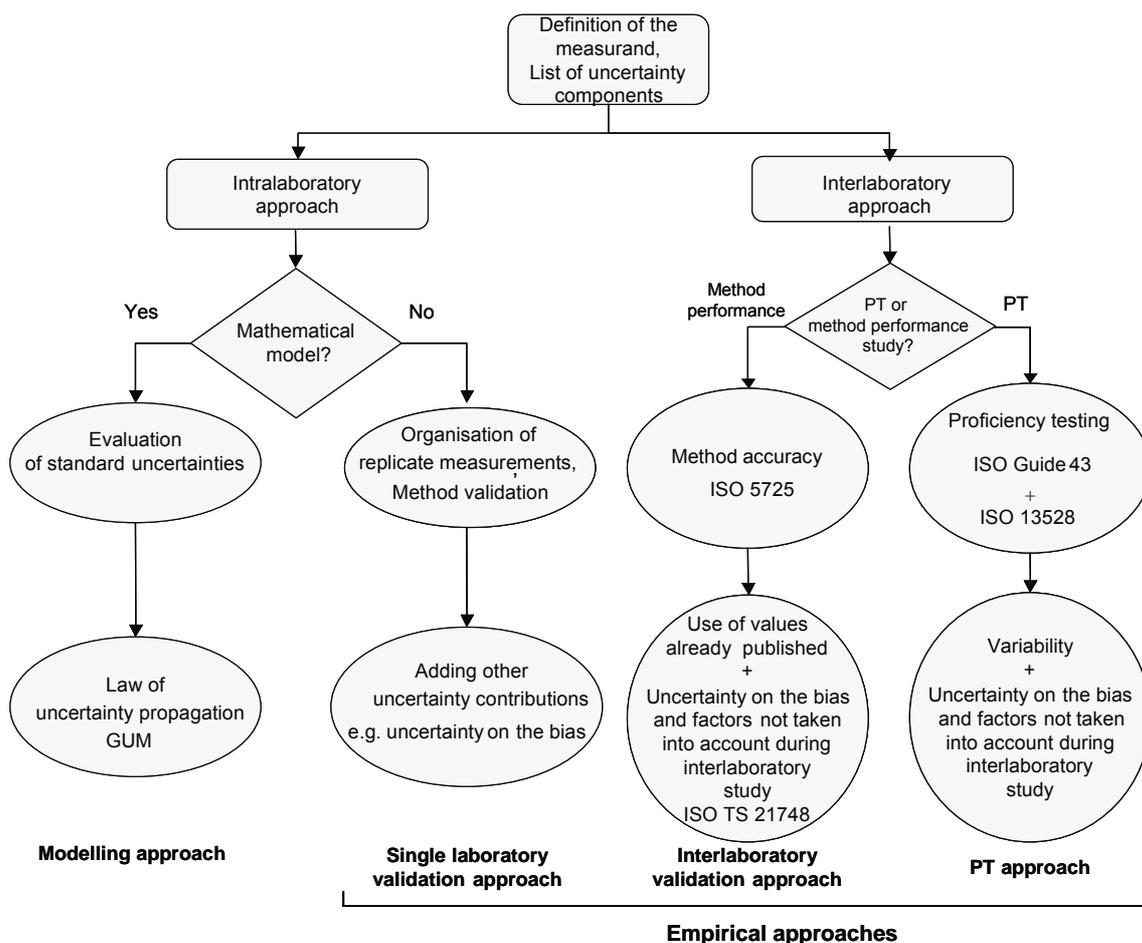
#### 1.1.1 Consistent reporting of measurement uncertainty

The concept of measurement uncertainty and the basic principles are defined in the “Guide to the Expression of Uncertainty in Measurement” (the “GUM”). The GUM is based on sound theory and provides a consistent and transferable evaluation of measurement uncertainty.

- Basic concept: The concept and definition of measurement uncertainty
- Recommendations: Expression in the form of standard uncertainty; combination of standard uncertainties; equal acceptance of Type A and Type B evaluations; treatment of degrees of freedom and expanded uncertainty
- Evaluation procedure: A detailed evaluation procedure applicable to the case where a comprehensive measurement equation can be developed, and where uncertainties are small compared to the respective values.

Consistency requires the basic concept to be accepted and recommendations to be followed. The procedure proposed in the GUM (see chapter 8 of GUM: Summary of procedure for evaluating and expressing uncertainty) is, however, one of several possible approaches for evaluating uncertainty. Other important and equally valid approaches include, for example, Monte Carlo simulation (see Reference 11 in the Annex), and empirical approaches based on intra- and inter-laboratory studies of method performance. Monte Carlo studies are especially important where the model is not differentiable, where the model is strongly non-linear or where distributions are strongly non-normal; empirical methods - which include inter-laboratory comparisons and method validation studies - are particularly appropriate where major contributions to uncertainty cannot readily be modelled in terms of measurable influence quantities, and where many laboratories use essentially identical test methods and equipment.

The ISO/IEC 17025 standard “General requirements for competence of testing and calibration laboratories” accordingly references ISO 5725 “Accuracy (trueness and precision) of measurement methods and results”, as well as the GUM, among its uncertainty evaluation requirements applicable to testing laboratories. It is, however, important to retain the consistency provided by adherence to the GUM concepts and recommendations. Fortunately, careful application of the different approaches can ensure that all the different approaches presented in this document remain compliant with the basic principles of the GUM.

**Figure 1: A road map for uncertainty estimation approaches**

### 1.1.2 Classifying different approaches

shows a convenient classification of uncertainty approaches. The classification is based on distinction between uncertainty evaluation carried out by the laboratory itself (called intralaboratory approach) and uncertainty evaluation based on collaborative studies (called interlaboratory approach).

The intralaboratory approach is then subdivided into:

- Use of uncertainty propagation based on a mathematical model, that is, an equation giving the quantitative relationship between the quantity measured and all the quantities on which it depends.
- Use of data from single laboratory method validation

The interlaboratory approach is then subdivided into:

- Use of data from collaborative method performance data (e.g. according to ISO 5725)
- Use of data from (interlaboratory) proficiency tests (PT)

This classification scheme will be used throughout this report.

Of course, the validation and interlaboratory approaches use statistical models as the basis for data analysis, and these could also be described as ‘mathematical models’. For convenience, however, the term ‘mathematical model’ is used only for the modelling approach in this technical report, and the term ‘statistical model’ for the other approaches.

### 1.1.3 Common points between the different approaches

Whatever uncertainty estimation approach is intended to be used, the following points are always important:

- Define clearly, with no ambiguity the measurand or the characteristic to be measured, analysed or tested.
- Analyse the measuring or testing process carefully in order to identify the major components of uncertainty and to examine if they are taken on board in the application of the law of propagation of uncertainty or if they are active during the repetition of observations organised to evaluate repeatability and reproducibility or if they are included in collaborative studies. It is also important to admit that in some situations, it is not possible to identify the individual components of the uncertainty. The symptom of this can be seen when the uncertainty evaluated by applying the modelling approach leads to a smaller uncertainty than the variation observed in laboratory intercomparisons.

Where sampling activities are performed, it is also important to define the measurand clearly. For example, do we seek information related to the test item transmitted to the laboratory for analysis or do we need information concerning the batch (the sampling target). It is obvious that the uncertainty will be different in both cases; where inferences are made about the sampling target itself, primary sampling effects become important and are often much larger than the uncertainty associated with measurement of a laboratory test item.

### 1.1.4 Presentation of the four approaches

#### (A) *The modelling approach*

The most widely understood modelling approach to evaluation of uncertainty is described in chapter 8 of the GUM. This procedure is based on a model formulated to account for the interrelation of all the influence quantities that significantly affect the measurand. Corrections are assumed to be included in the model to account for all recognised, significant systematic effects. The application of the law of propagation of uncertainty enables evaluation of the combined uncertainty on the result. The approach depends on partial derivatives for each influence quantity, so depends on either an equation for the measured result or, if the form is algorithmic, on numerical differentiation. Modelling the measuring process may be infeasible for economic or other reasons. In such cases alternative approaches may be used.

Where the assumptions apply, modelling is relatively economical compared to extensive replication and experimental study, and is particularly useful for evaluating the contribution of reference value uncertainties to the combined uncertainty associated with the final measurement result.

It is important to understand that the other approaches presented here are as valid as the modelling approach and sometimes lead to more realistic evaluation of the uncertainty. These approaches are based on long experience and reflect common practice.

#### (B) *The single laboratory validation approach*

The major sources of variability can often be assessed by method validation study. Estimates of bias, repeatability, and within laboratory reproducibility can be obtained by organising experimental work inside the laboratory. Information can also be obtained from quality control data (control chart). Combined with experimental investigation of important individual effects, this approach provides essentially all of the data required for uncertainty estimation.

The most important issues are:

- First, to vary, during the repetition of the experiment, a majority of influence quantities that can affect the result.
- Second, to assess the bias (or trueness) of the method. The use of certified reference materials (CRMs) and/or comparison with definitive or reference methods can help to evaluate the component of uncertainty related to the trueness.

**(C) The interlaboratory validation approach**

The major sources of variability can often be assessed by interlaboratory studies as stated in ISO 5725 “Accuracy (trueness and precision) of measurement methods and results” which provide estimates of repeatability (repeatability standard deviation  $s_r$ ), reproducibility (reproducibility standard deviation  $s_R$  and (sometimes) trueness of the method (measured as a bias with respect to a known reference value).

Uncertainty estimation based on precision and trueness data acquired in compliance with ISO 5725 is fully described in ISO/TS 21748 “Guidance for the use of repeatability, reproducibility and trueness estimates in measurement uncertainty estimation”.

**(D) The use of Proficiency Testing (EQA) data - the “PT approach”**

Proficiency tests (EQA) are intended to check periodically the overall performance of a laboratory. The laboratory’s results from its participation in proficiency testing can accordingly be used to check the evaluated uncertainty, since that uncertainty should be compatible with the spread of results obtained by that laboratory over a number of proficiency test rounds.

The “PT approach” can also be used to evaluate the uncertainty. For example, if the same method is used by all the participants to the PT scheme, the standard deviation is equivalent to an estimate of interlaboratory reproducibility and can, in principle, be used in the same way as the reproducibility standard deviation obtained from collaborative study (section 1.1.3 above). Further, over several rounds, the deviations of laboratory results from the assigned value can provide a preliminary evaluation of the measurement uncertainty for that laboratory. Eurolab Technical Report No. 1/2002 “Measurement Uncertainty in Testing” [3] provides an example of uncertainty assessment of the measurement of 100 mg sulphate in waste water determined with ion-chromatography from proficiency test results, and a Nordtest guide [6] provides a general approach.

The use of a single laboratory’s deviations from the assigned value is equally applicable when the laboratories are free to choose any appropriate method.

Little guidance is currently available on the use of PT data for uncertainty estimation; the approach is accordingly discussed in greater detail in chapter 2.

**1.1.5 Combination of different approaches**

Very often, a combination of the different approaches needs to be used to assess the uncertainty. For example:

- When a laboratory decides to use the modelling approach, the model invariably includes terms associated with random variation; such contributions are usually best assessed using quality control data or other replication.
- Uncertainty estimates may be based on a model including only those effects considered systematic over a long period, together with a single term accounting for random variation on the same timescale; control chart or other intra-laboratory reproducibility data are then used to estimate the contribution due to random variation.
- The use of the interlaboratory validation approach can require the application (by the CRM supplier) of the modelling approach to evaluate the uncertainty on the reference value of the CRMs used to estimate the trueness of the method.
- It is often necessary to apply some elements of the modelling approach to estimate contributions that cannot be obtained by experimental variation. For example, it is not usually possible to arrange for deliberate variations from nominal values of measurement standards (etalons). Simple modelling is often the best way of estimating the contribution of reference value uncertainties, even when the general approach is based on inter- or intra-laboratory empirical studies.
- If the measurand includes sampling, empirical methods are very likely to be essential for estimating the contribution of sampling variability, whether the final measurement process is investigated by modelling or by empirical methods.

In general, therefore, most practical uncertainty estimates involve elements of both modelling and 'empirical' approaches.

## References

The table contains a compilation of references (guidelines and standards) for the various approaches. As far as these documents are also referenced in the Annex, the relevant number is given under "Reference". The other columns indicate which uncertainty evaluation approaches are addressed in the respective document.

Document	Reference	General	Modelling	Single Laboratory	Inter-laboratory	PT
ISO (1993/1995), <i>Guide to the expression of uncertainty in measurement (GUM)</i>	1	X	X			
EURACHEM / CITAC (2000), <i>Quantifying uncertainty in analytical measurement, 2<sup>nd</sup> edition</i>	2	X	X	X		
EUROLAB Technical Report No. 1/2002, <i>Measurement Uncertainty in Testing</i>	3	X				
EUROLAB Technical Report No. 1/2006, <i>Guide to the Evaluation of Measurement Uncertainty for Quantitative Test Results</i>	4	X	X	X		X
EA 4/16 (2004), <i>Guidelines on the expression of uncertainty in quantitative testing</i>	5	X	X	X	X	X
NORDTEST Technical Report 537 (2003), <i>Handbook for calculation of measurement uncertainty in environmental laboratories</i>	6			X		X
EA-4/02 (1999), <i>Expression of the Uncertainty of Measurement in Calibration</i>	7		X			
ISO 5725 Accuracy (trueness and precision) of measurement methods and results (6 parts)	8				X	
ISO 5725-3 Accuracy (trueness and precision) of measurement methods and results – Part 3: Intermediate measures of the precision of a standard measurement method	8			X		
ISO/TS 21748 <i>Guide to the use of repeatability, reproducibility and trueness estimates in measurement uncertainty estimation</i>	9				X	
AFNOR FD X 07-021, <i>Fundamental standards – Metrology and statistical applications – Aid in the procedure for estimating and using uncertainty in measurements and test results</i>	10		X		X	
Supplement No. 1 to the GUM, <i>Propagation of distributions using a Monte Carlo method</i> (publication expected 2007)	11		X			
ISO 13528 <i>Statistical methods for use in proficiency testing by interlaboratory comparison</i>	12					X
ISO/TS 21749 <i>Measurement uncertainty for metrological applications – Repeated measurements and nested experiments</i>	13			X		

## 1.2 UNCERTAINTY DATA OBTAINED FROM THE VARIOUS APPROACHES

The modelling approach for the evaluation of measurement uncertainty, as featured in the GUM, has been designed to deliver an estimate of the uncertainty of a measurement result in the format of a combined standard uncertainty, i.e. a standard deviation comprising all effects impacting on the measurement results, random as well as systematic. Multiplication with an appropriate coverage factor (most often  $k = 2$  for approximately normally distributed data) yields an expanded uncertainty which can be used to construct a 95 % coverage interval (an analogue of a confidence interval). According to GUM principles the evaluation of measurement uncertainty is valid for a specified measurement result, and the concept of method performance characteristics determined during method validation and monitored by quality control is not deeply rooted in metrology.

The results obtained using “empirical approaches” for the evaluation of measurement uncertainty most often have a different scope, usually referring to a specified test procedure rather than to a specified measurement result. The form of the data - usually separate estimates of precision and bias - also differs from one approach to another. Handling of uncertainty data obtained using different approaches therefore requires some basic information concerning scope and form of the output obtained from the various approaches for evaluating measurement uncertainty.

### 1.2.1 Modelling approach

#### *Form of uncertainty data*

The typical output of the modelling approach is an “uncertainty budget” summarising the evaluation of the combined standard uncertainty of the measurement result from the uncertainties attributed to the various data (measured and other) used in evaluating the measurement. The uncertainty budget comprises data for each such “input quantity” to the measurement result, and data for the measurement result itself, as follows.

<b>Input quantities</b>	<b>Measurement result</b>
value $x_i$	value $y$
standard uncertainty $u(x_i)$	(combined) standard uncertainty $u(y)$
sensitivity coefficient $c_i = (\partial y / \partial x_i)$	coverage factor $k$
uncertainty contribution $u_i(y) = c_i \times u(x_i)$	expanded uncertainty $U(y) = k \times u(y)$

Optional data may include or concern degrees of freedom, type of distribution used, proportions of uncertainty contributions and others.

Unless correlation among input quantities has to be taken into account, the standard uncertainty  $u(y)$  is given by the root sum of squares of the uncertainty contributions  $u_i$ .

$$(1.1) \quad u(y) = \sqrt{\sum u_i^2(y)}$$

By default in an uncertainty budget absolute uncertainties are used. Conversion to relative uncertainties is always possible but requires due care (other sensitivity coefficients).

#### *Scope of uncertainty data*

An uncertainty budget refers to a specified measurement. But the algorithm behind the uncertainty budget applies to all measurements made using the same measurement system and procedure on comparable test items. For any new measurement, the (combined) standard uncertainty  $u(y)$  is obtained by plugging the input data  $x_i$  and  $u(x_i)$  for this measurement into the algorithm, which then will return  $y$  and  $u(y)$ . Of course, if the input data are close to those for a previous measurement, the standard uncertainty  $u(y)$  will be about the same as obtained before.

As an obvious benefit, an uncertainty budget provides information about the relative magnitude of the various uncertainty contributions. This information is particularly useful when planning improvements of the measurement procedure.

### 1.2.2 Single laboratory validation approach

#### ***Form of uncertainty data***

The basic principle behind this approach is the synthesis of uncertainty estimates from estimates of precision and estimates of bias:

*Measurement accuracy = precision & trueness*

*Measurement uncertainty = within-laboratory reproducibility & uncertainty on bias*

Measurement uncertainty is estimated as a root sum of squares of a standard deviation  $s$  characterising the (im)precision of the measurement and an estimate  $b$  accounting for measurement bias, which gives the standard uncertainty  $u$  according to the schematic equation

$$(1.2) \quad u = \sqrt{s^2 + b^2}$$

Here it is understood that measurement bias is investigated, and corrective actions are taken to remove/reduce such bias to the greatest possible extent. The bias-related uncertainty estimate accounts for the potential bias left after correction. In practice, however, it happens quite often that significant bias is found, but the data are not sufficient for deriving a sound correction. For example, it may be doubtful whether a single-level correction, based on measurements of a single standard, is applicable to the entire measuring range. Then additional measurements, e.g. including another standard, should be made in order to characterise the bias to an appropriate degree. If this is not possible or not practical, a pragmatic alternative is to increase the uncertainty to account for the observed bias instead of attempting any correction.

Note: The GUM appears to rather discourage such procedure, stating in the note to clause 6.3.1 “Occasionally one may find that a known correction for a systematic effect has not been applied to the reported result of a measurement, but instead an attempt is made to take the effect into account by enlarging the “uncertainty” assigned to the result. This should be avoided; only in very special circumstances should corrections for known systematic effects not be applied to the results of a measurement. Evaluating the uncertainty of a measurement result should not be confused with assigning a safety limit to some quantity.” In appreciating this guidance, a key phrase to recognise is that of a “known correction”. Certainly systematic effects (i.e. bias) that have been characterised to a degree that the applicable corrections can be considered as known, should be corrected, unless this entails unacceptable expenses. In practice, however, it will often be the expenses for deriving rather than for applying a “known correction” that are prohibitive. Then increasing measurement uncertainty to account for significant bias is most certainly better than applying a doubtful correction or, even worse, ignoring the bias.

Handling of uncorrected bias is a contentious issue, requiring informed judgement. A range of different approaches have been proposed, but a generally accepted procedure has not yet emerged.

#### ***Data on precision***

The precision of a measurement procedure is investigated during method validation, monitored in quality control, and quantified by standard deviations obtained from replicate measurements on appropriate test items. Depending on the conditions for replicate measurements, two different standard deviations are obtained:

- $s_{rw}$  the within-laboratory repeatability standard deviation, obtained under repeatability conditions: same operator, same equipment, short-time repetition.
- $s_{Rw}$  the within-laboratory reproducibility standard deviation, obtained under within-laboratory reproducibility conditions (often called “intermediate conditions”): different operators (if applicable), different equipment (if applicable), long-time repetition.

For the purpose of estimating measurement uncertainty, the within-laboratory reproducibility standard deviation  $s_{Rw}$  will be used. The repeatability standard deviation  $s_{rw}$  is not normally a suitable uncertainty estimate, since it excludes major uncertainty contributions.

Often different reproducibility data, obtained from different measurement series will be available. Then these data should be compared and combined into a joint precision estimate  $s_{Rw}$ , preferably as a function of the measurand level.

### **Data on bias**

It is understood that measurement bias is eliminated to the greatest possible extent. Residual bias is investigated during method validation, monitored in quality control, and quantified by deviations of measurement results on appropriate test items from corresponding reference values. Most often reference materials are used for this purpose, but alternatively a reference measurement procedure may be used.

Typical data obtained from bias investigation and control are

- $\Delta$  the mean deviation of replicate measurement results from the corresponding reference value

In addition, an uncertainty estimate  $u_{ref}$  for the reference value should be available.

The bias contribution to measurement uncertainty is obtained from the mean deviation, the uncertainty of the reference value, and the (im)precision of the mean value of the replicate measurements made in the bias investigation:

$$(1.3) \quad b = \sqrt{\Delta^2 + u_{ref}^2 + \frac{s^2}{n}}$$

Here it should be noted that the standard deviation in eq. (1.3), accounting for the variability of measurements made in the bias investigation, may not be the same as the standard deviation in eq. (1.2). With major numbers of replicate measurements the term  $s^2/n$  in eq. (1.3) can normally be neglected.

Often different data on bias, obtained from different measurement series, will be available. Then these data should be compared and combined into a joint estimate for the uncertainty on bias, preferably as a function of the measurand level. In absence of within-laboratory bias investigations the PT approach (see section 1.2.4 and chapter 2) may be used. In this case bias estimates are obtained from PT data (deviation of the laboratory's result from the assigned value) while the within-laboratory reproducibility standard deviation is used as precision estimate.

If bias estimates are not available at all, a pragmatic approach would be to expand the within-laboratory standard deviation using a rule-of-thumb factor. For the chemical field, e.g., average proportions between various within-laboratory and interlaboratory precision data ("ladder of errors") were published in [M Thompson, "Towards a unified model of errors in analytical measurement", *Analyst* (2000), 125, 2020-2025]. Considering that a factor of two is quite commonly observed in such studies,  $u \approx 2 s_{Rw}$  could be used as a preliminary estimate of measurement uncertainty in absence of bias data.

### **Scope of uncertainty data**

Precision and bias estimates obtained using the within-laboratory validation approach are designed as to cover all effects impacting the measurement that would occur under normal operation conditions for the measurement procedure. Therefore, provided that the measurements are under statistical control, uncertainty estimates obtained using this approach are applicable for all measurements within the scope of the measurement procedure. The application range of the uncertainty estimates is determined by the range covered in the validation study and the on-going quality control. Therefore these investigations should include appropriate within-scope variations, e.g. different levels of the measurand and different types of test items.

### 1.2.3 Interlaboratory validation approach

#### **Form of uncertainty data**

For standard test procedures, trueness and precision are usually determined by an inter-laboratory comparison (see ISO 5725-2). The main performance characteristics obtained in such studies are

- $s_r$  the repeatability standard deviation
- $s_R$  the interlaboratory reproducibility standard deviation

For the purpose of estimating measurement uncertainty, the reproducibility standard deviation  $s_R$  will be used. The repeatability standard deviation  $s_r$  is not normally a suitable uncertainty estimate, since it excludes major uncertainty contributions.

Often precision data are determined for different levels of the quantity concerned, and interpolation of these data between the different levels is specified.

When suitable reference test objects are available, the interlaboratory validation study may also include an investigation of bias. However, since the (interlaboratory) reproducibility standard deviation already comprises systematic effects due to different ways of operation in the laboratories involved (laboratory bias), such study will only address method bias. Most often method bias is not significant or not relevant and is not specified as a separate performance characteristic.

Therefore the default uncertainty estimate from an interlaboratory validation study is, as a standard uncertainty  $u$ :

$$(1.4) \quad u = s_R$$

#### **Scope of uncertainty data**

Interlaboratory validation studies of standard test procedures are designed and evaluated as to obtain estimates of precision and bias which are typical for the performance of the test procedure when operated by an experienced laboratory under proper quality control. Therefore, in principle, these data may be utilised by any test laboratory. In the International Technical Specification ISO/TS 21748 *Guide to the use of repeatability, reproducibility and trueness estimates in measurement uncertainty estimation* 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 test items are consistent 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 its own trueness and precision (see section 1.2.2) for compatibility with the interlaboratory comparison data  $s_r$  and  $s_R$ .

Provided that the measurements are under statistical control, the reproducibility standard deviation  $s_R$  is applicable for all measurements within the scope of the standard procedure.

For out-of scope applications, i.e. if the test conditions or the test objects substantially deviate from those in the interlaboratory validation study, the effect of these deviations has to be estimated and combined with the reproducibility standard deviation. For this purpose the following schematic equation applies:

$$(1.5) \quad u = \sqrt{s_R^2 + \sum u_{other}^2}$$

### 1.2.4 Approach using PT data

If a laboratory has successfully participated in an inter-laboratory proficiency test, it may utilise the results for estimating the measurement uncertainty for the measurement procedure used. Chapter 2 provides a full description how this can be done. Therefore this section is restricted to a brief summary.

Alternatively proficiency tests offer the possibility to check the validity of measurement uncertainty estimates obtained otherwise. This topic is addressed in chapter 3.

#### **Form of uncertainty data**

In an (inter-laboratory) proficiency test each participating laboratory measures a specified quantity (or several specified quantities) on a sample or specimen received and submits its result to the organiser. In return the laboratory receives a score, based upon the deviation of its result from the assigned value, e.g. a z-score  $z = (x - x_{\text{ass}}) / \sigma_{\text{PT}}$  where  $x$  is the result submitted by the laboratory,  $x_{\text{ass}}$  is the assigned value of the PT sample or specimen, and  $\sigma_{\text{PT}}$  is the standard deviation for proficiency assessment. Rather than the score, the deviation  $\Delta = (x - x_{\text{ass}})$  may be used for uncertainty estimation as a bias estimate, similar to the single-laboratory validation approach where deviations  $\Delta = (x - x_{\text{ref}})$  obtained on reference samples or specimens are used for estimating the bias contribution to measurement uncertainty.

Like in the single-laboratory validation approach (section 1.2.2), measurement uncertainty may be estimated as a root sum of squares of a standard deviation characterising the (im)precision of the measurement and an estimate accounting for measurement bias according to the schematic equation

$$(1.6) \quad u = \sqrt{s^2 + b^2}$$

In the simplest case, where a single PT result is available, an initial estimate of uncertainty can be formed from the observed deviation and the within-laboratory precision; this can usefully be improved based on additional PT data.

#### **Data on precision**

For the purpose of an initial estimate of measurement uncertainty using the PT approach, the within-laboratory reproducibility standard deviation  $s_{RW}$  obtained from single-laboratory validation of the measurement procedure (see section 1.2.2) may be used. The standard deviation obtained exclusively from replicate measurements of the PT sample or specimen is not normally a suitable uncertainty estimate, since it excludes major uncertainty contributions.

#### **Data on bias**

Typical data obtained from PT participation are

- $\Delta$  the deviation of the laboratory's result (or the average of several results from replicate measurements) from the assigned value of the PT sample or specimen

In addition, an uncertainty estimate  $u_{\text{ass}}$  for the assigned value should be available.

The bias contribution to measurement uncertainty,  $b$ , is obtained from the deviation  $\Delta$ , the uncertainty of the assigned value, and the (im)precision of the measurement on the PT sample:

$$(1.7) \quad b = \sqrt{\Delta^2 + u_{\text{ass}}^2 + \frac{s^2}{n}}$$

Equation (1.7) refers to the case where the laboratory's result is a mean value of  $n$  replicates. The standard deviation is obtained from these replicates, but the within-laboratory reproducibility standard deviation  $s_{RW}$  may also be used.

If results from several PT rounds (for comparable measurements) are available, the estimates for uncertainty on bias obtained from individual rounds should be compared and

combined, if compatible, by taking mean squares on an absolute or relative basis, as appropriate.

***Scope of uncertainty data***

In the PT approach the two basic components of measurement uncertainty are obtained from different investigations: precision is determined by in-house method validation while bias is estimated from PT results. Most often the application range of a bias estimate from a single PT will be rather restricted, and this will carry over to the application range of the entire uncertainty estimate obtained. If results from several PT rounds, covering a wider measuring range and a wider range of test items are available, the application range of uncertainty estimates obtained using the PT approach may be enlarged significantly. Further information is given in chapter 2.

## Chapter 2: Use of PT data in measurement uncertainty evaluation

*This chapter describes the use of PT data either for estimating the overall performance of laboratories or for estimating bias for a single laboratory to be used in measurement uncertainty evaluation. The use of PT data for a single laboratory for checking an uncertainty estimate is described in section 3.1.*

### 2.1 DRAWBACKS AND ADVANTAGES OF USING PT DATA INSTEAD OF REFERENCE MATERIALS

For estimating measurement uncertainty for the test results one would like to have several CRMs similar to the test items; that is, of similar type, with similar homogeneity and measurand range. This is seldom the case in practice. Data from proficiency testing (PT) can provide useful supplementary information. The advantage of using PT data is that, while principally a test of laboratories performance, a single laboratory will, over time, test a range of well-characterised materials chosen for their relevance to the particular field of measurement. Further, PT test items may be more similar to a routine test item than a CRM since the demands on stability and homogeneity are frequently less stringent.

The relative disadvantage of PT samples is the lack of traceable reference values similar to those for certified reference materials. Consensus values in particular are prone to occasional error. This certainly demands due care in their use for uncertainty estimation, as indeed due care is recommended in recent PT protocols [M Thompson, S L R Ellison, R Wood; "The International Harmonized Protocol for the proficiency testing of analytical chemistry laboratories (IUPAC Technical Report)"; Pure Appl. Chem. 78(1) 145-196 (2006)]. However, appreciable bias in consensus values is relatively infrequent as a proportion of all materials circulated, and substantial protection is provided by the extended timescale common in proficiency testing. PT assigned values, including those assigned by consensus of participants' results, may therefore be regarded as sufficiently reliable for most practical purposes.

The data obtained from a laboratory's participation in PT can therefore be a good basis for uncertainty estimates provided the following conditions are fulfilled:

- The test items in PT should be reasonably representative of the routine test items. For example the type of material and range of values of the measurand should be appropriate.
- The assigned values have an appropriate uncertainty.
- The number of PT rounds is appropriate; a minimum of 6 different trials over an appropriate period of time is recommended in order to get a reliable estimate.

Where consensus values are used, the number of laboratories participating should be sufficient for reliable characterisation of the material.

### 2.2 USE OF PT DATA FOR EVALUATION OF THE OVERALL PERFORMANCE OF LABORATORIES

Proficiency Tests are becoming more and more widely used within the testing community and, providing the PT study is relevant, the interlaboratory standard deviation (between laboratories) from on-going studies can be used in several ways. For any customer using the analytical service PT data is a key indication of the overall performance of laboratories. For a laboratory the interlaboratory standard deviation can be used as a first estimation of the standard uncertainty (see Eurolab TR 1/2002 and Nordtest TR 537, references [3] and [6] in the Annex). For example, in the Eurolab technical report [3] the interlaboratory standard deviation from a PT round is used as an estimate of the standard uncertainty for determination of sulphate in water with Ion Chromatography.

## The SWEDAC compilation of PT data

For the environmental sector an extensive set of PT data for water and sludge matrices has been compiled over several years. The interlaboratory relative standard deviation is taken as standard uncertainty,  $u_{rel}$  (in % of the measured value). The compilation is split into different analytical techniques for each measurand and sample preparation method, and measurement uncertainty is estimated from the reproducibility standard deviation for each parameter within a given concentration range. A detailed compilation (in Swedish) is given on the SWEDAC website. Example 10 in chapter 4 presents an English summary and example data. The relative standard uncertainty is either constant or varying within the given concentration range and is given in % according to equation (2.1).

$$(2.1) \quad u_{rel} = K \cdot \frac{1}{x} + L$$

The concentration ( $x$ ) dependence of  $u_{rel}$  according to equation (2.1) can thus be estimated within a given concentration range by the parameters  $K$  and  $L$ . The parameters  $K$  and  $L$  and the lower limit of the concentration range are given in a table with one row for each combination of analytical technique and sample preparation. When the parameter  $K$  is equal to zero the relative uncertainty is constant within the given concentration range. An example is ammonium determination in water and waste water for concentrations over 0.3 mg/l using different procedures (analytical techniques/ sample preparation). The relative standard uncertainty is in the range 7 to 14 % dependent on which procedure was used. More illustrations of this method of evaluation are given in chapter 4, example 10.

### 2.3 USE OF PT DATA FOR ESTIMATING BIAS FOR A SINGLE LABORATORY

Estimates of bias using PT data are obtained from a statistical evaluation based on a series of PT results. The bias is calculated from a series of PT results either as an absolute value  $\Delta$

$$(2.2) \quad \Delta = \frac{1}{n} \sum (x_i - x_{ref})$$

or as a relative value  $\Delta'$

$$(2.3) \quad \Delta' = \frac{1}{n} \sum \frac{x_i - x_{ref}}{x_{ref}}$$

In many cases not many PT rounds are available and then the bias estimate may be less reliable. But even a single PT study can be very informative in the absence of other data and may be utilised for initial uncertainty estimation (see Eurolab TR 1/2006, reference [4] in the Annex).

The bias estimate from PT studies should not normally be used for any correction of the results. If the observed bias is regarded as unacceptable the laboratory has to take action and resolve this issue.

### 2.4 THE NORDTEST APPROACH TO ESTIMATING UNCERTAINTY FROM PT DATA

In this approach, the within-laboratory reproducibility standard deviation is combined with estimates of the method and laboratory bias using PT data. The details are described in the Nordtest Handbook (Annex, ref. [6]). The uncertainty estimate is constructed using the following formula (using the symbols and terminology of reference [6]):

$$(2.4) \quad U = k \cdot u = k \cdot \sqrt{u(R_w)^2 + u(bias)^2}$$

where  $U$  is the expanded uncertainty,  $k$  is the coverage factor,  $u$  is the combined standard uncertainty,  $u(R_w)$  is the within-laboratory reproducibility standard deviation obtained from QC data and  $u(bias)$  is the uncertainty component arising from method and laboratory bias, estimated from PT data in the example given below. For extended measuring ranges it is important to decide whether absolute or relative uncertainty is most appropriate.

In the examples below relative uncertainties (in %) are used. The uncertainty component arising from method and laboratory bias is calculated using the following formula.

$$(2.5) \quad u(\text{bias}) = \sqrt{RMS_{\text{bias}}^2 + u(C_{\text{ref}})^2}$$

where  $RMS_{\text{bias}}$  is the root mean square of the bias values and  $u(C_{\text{ref}})$  is the average uncertainty of the assigned values; that is,  $u(C_{\text{ref}})$  is, in the Nordtest approach, an estimate of an average over several rounds (often relative standard uncertainties averaged as variances).

Bias (lab result *minus* assigned value) can be both positive and negative. Even if the results appear to give positive biases on certain occasions and negative on others, all bias values should be used to estimate the uncertainty component,  $RMS_{\text{bias}}$  (*root mean square of the bias values*)<sup>1</sup>. One could also use the standard deviation when the mean bias is small. However it is recommended to use the  $RMS$  value since it is appropriate with both small and high mean bias values. In the case of a mean bias of zero the  $RMS$  value is the same as the standard deviation.

The way forward is thus very similar to that for use of multiple certified reference materials (CRMs). However, the estimation of the bias from PT generally has large uncertainty, and the resulting uncertainty estimate is generally higher than if CRMs are used. This is partly due to the fact that the certified value of a CRM is normally better defined than a nominal or assigned value in a PT exercise. In some cases, too, the calculated uncertainty of the assigned value  $u(C_{\text{ref}})$  from PT studies may be too high for the purpose, and is not appropriate for estimating the uncertainty component arising from method and laboratory bias,  $u(\text{bias})$ .

#### 2.4.1 Estimating $u(\text{bias})$ – the uncertainty component from method & lab bias

<b>Uncertainty component for the uncertainty of the assigned value</b>	
PT providers increasingly report uncertainties for their assigned values. If the provider reports an uncertainty, use the uncertainty estimate supplied.	
Step	Example
Find the interlaboratory standard deviations, $s_R$ , for all laboratories participating in the exercises.	The $s_R$ has been on average 9 % in six exercises. Mean number of participants = 12
Calculate the uncertainty of the assigned value, $u(C_{\text{ref}})$ <sup>Note 1</sup>	$u(C_{\text{ref}}) = \frac{s_R}{\sqrt{n}} = \frac{9}{\sqrt{12}} = 2.6 \%$

Note 1: If the assigned value is a median value the equation will, following the principles of ISO 13528 [12], be  $u(C_{\text{ref}}) = 1.253 \cdot s_R / \sqrt{n}$ .

<b>Uncertainty component for laboratory and method bias for a specific laboratory</b>	
Step	Example
Obtain the laboratory's deviations from the assigned value for at least six PT rounds	The relative bias has been 2 %, 7 %, -2 %, 3 %, 6 % and 5 %.
Quantify the components	$RMS_{\text{bias}} = 4.6 \%$ , $u(C_{\text{ref}}) = 2.6 \%$
Calculate uncertainty component arising from method and laboratory bias, $u(\text{bias})$	$u(\text{bias}) = \sqrt{RMS_{\text{bias}}^2 + u(C_{\text{ref}})^2} =$ $= \sqrt{4.6^2 + 2.6^2} = 5.3\%$

<sup>1</sup> The use of an RMS value is equivalent to an estimated standard deviation around an assumed value of bias equal to zero. This implies that the RMS value takes into account both the bias and the variation of bias.

The uncertainty arising from method and laboratory bias,  $u(bias)$  which in this case is 5.3 % then has to be combined with the within-laboratory reproducibility  $u(R_w)$  according to equation (2.4) to estimate the combined standard uncertainty and the expanded uncertainty, respectively – see also Nordtest Handbook, section 3.

There will of course be some double counting of uncertainty components since e.g. the repeatability component is included both in  $u(bias)$  and  $u(R_w)$ . If the repeatability component is small compared with the other components this is little problem since this double counting will not influence the (combined) standard uncertainty significantly. However at lower levels close to the detection limit, where run-to-run variations are often the main uncertainty component, this approach may have to be modified.

## **2.5 OTHER USES OF PT DATA**

PT data may also be used to verify uncertainty estimates obtained by other methods. This is discussed further in section 3.1.2.

## Chapter 3: Verification, adjustment and comparison of uncertainty estimates

*This chapter is devoted to technical issues pertinent to comparison, validation and revision of uncertainty estimates.*

### 3.1 VERIFICATION OF UNCERTAINTY ESTIMATES

#### 3.1.1 Principle of verification

The estimated uncertainty is a range and in order to verify that the estimated uncertainty is valid for the laboratory over time several checks have to be performed. It is assumed here that the measurement procedure is validated and normal QC is in place.

The basis of an experimental check on the validity of an uncertainty estimate is simple; is the result or (better) series of results consistent with the reported uncertainty? Ideally, such a test will be done on new test items for which the value is unknown to the laboratory, or at least, to the individual scientist at the time of the measurement(s). The test items should also be similar to those normally measured in the laboratory. For example, in chemical analysis, the concentration range and matrix should be similar to those of normal test materials.

A variety of verification procedures may be used, including:

- i) Checks using observed within-laboratory precision
- ii) Checks using certified reference materials or suitable test materials
- iii) Checks using reference methods
- iv) Checks based on the results of proficiency testing (including EQA data or measurement audits)
- v) Checks based on comparison of results with other laboratories
- vi) Comparison with other uncertainty estimates based on different approaches or different data

These approaches are described in more detail below.

#### 3.1.2 Verification procedures

##### **Checks using within-laboratory precision (i)**

Compare the estimated standard uncertainty with the standard deviation of a series of results on an appropriate test item over a period of time (60 results keeps the likely variation of observed standard deviation within approximately [ $\pm 15\%$ ]). Observations on a routine quality control material are appropriate. An *F*-test may be used in order to protect against random variations in in-house precision. The standard uncertainty for a routine test method should never be smaller than the long-term precision for the same method and test material; if the standard uncertainty is significantly smaller than the observed within-laboratory standard deviation, the uncertainty estimate should be reviewed immediately.

A standard uncertainty larger than the observed within-laboratory standard deviation might arise from a variety of causes. Unless uncertainty contributions other than random variability are known to be small or negligible, there is no immediate cause for action unless the in-house precision is substantially smaller than the precision terms used in the uncertainty estimate.

##### **Checks based on certified reference materials (CRM) or suitable test materials (ii)**

A certified reference material may be used in the in-house checks above. A typical procedure is as follows:

Measure a suitable test material or CRM of known assigned value  $x_{ref}$  with small uncertainty. Check the difference  $d$  between observed value  $x$  and reference value  $x_{ref}$  against the expanded uncertainty  $U(x)$ . If the uncertainty on the reference value is not small\*, calculate the uncertainty of the difference from  $u(d)^2 = u(x_{ref})^2 + u(x)^2$ , expand appropriately and compare  $d$  with  $U(d)$ . If the difference  $d$  is equal to or greater than the expanded uncertainty  $U(d)$ , it should be concluded that the uncertainty fails to account for the observed bias on the material. The uncertainty estimate should be reviewed and appropriate steps taken to identify the source of the bias.

### **Checks based on reference methods (iii)**

Reference methods provide independent reference values. A single such value can be used to check an uncertainty estimate in the same way as using a single CRM value (above). A series of such values can be used in the same way as a series of proficiency test results (below).

### **Checking an uncertainty estimate against proficiency test results (iv)**

Appropriate proficiency testing (PT) provides a series of relevant test items and therefore gives a suitable check. PT may even improve on tests based on certified reference materials; in contrast to a CRM the samples can sometimes better reflect normal samples since the stability is a less important issue. For example, in the clinical area fresh whole blood or serum samples can be used – samples which the laboratory analyses every day. If PT is not available, comparison with other competent laboratories can be very useful. At least 6 separate samples analysed over a period of some months are recommended for this assessment.

The assessment of the uncertainty estimates is performed using the zeta score  $\zeta$ .

$$(3.1) \quad \zeta = \frac{x - x_a}{u(x)^2 + u(x_a)^2}$$

or the  $E_n$  number

$$(3.2) \quad E_n = \frac{x - x_a}{U(x)^2 + U(x_a)^2}$$

where  $x_a$  is the assigned value,  $x$  the laboratory result,  $u(\cdot)$  a standard uncertainty and  $U(\cdot) = k \cdot u(\cdot)$  the corresponding expanded uncertainty with coverage factor  $k$ .

The zeta score is most appropriate for checking the standard uncertainty  $u$ ;  $E_n$  provides a check on the expanded uncertainty  $U = k \cdot u$  and therefore additionally checks the validity of the coverage factor  $k$ .

If the estimated uncertainty is correct the zeta score should be in the range -2 to 2, and the  $E_n$  value should be in the range -1 to 1.

Note: When  $k$  is 2.0 for both assigned value expanded uncertainty  $U(x_a)$  and the laboratory uncertainty  $U(x)$ ,  $E_n$  and zeta scores give equivalent information (that is, although the numerical scores differ by a factor of 2, a given laboratory deviation  $x - x_a$  will lead to the same conclusion about the validity of the uncertainty estimate).

There are three clear cases

- Case 1: Uncertainty overestimated –  $|E_n|$  is always significantly less than 1 or  $|\zeta|$  always significantly less than 2
- Case 2: Correct – most values of  $|E_n|$  are in the range 0 to 1, or  $|\zeta|$  in the range 0 to 2.
- Case 3: Uncertainty underestimated –  $|E_n|$  is frequently over 1 or  $|\zeta|$  frequently over 2

---

\* The standard uncertainty of the CRM value should ideally be smaller than  $u/5$ , where  $u$  is the uncertainty estimated for routine application of the method.

**Case 1: Overestimate** – The estimated uncertainty is clearly higher than the laboratory performance suggests. This could be acceptable, especially if the reported uncertainty is lower than or equal to the target value of uncertainty (that is, within the customer's requirements). However if there is a need for lower uncertainty, a new estimate has to be made. An example of an approach where a laboratory can make an overestimate is when a particularly good laboratory uses a reproducibility estimate from PT as a first, conservative, estimate of uncertainty. Here a new estimate based on the laboratory's method performance data or using the modelling approach would probably give a lower estimate of the uncertainty.

**Case 2: Correct** – Here one could think that all is clear-cut but we have to bear in mind that there are many sources that are not always tested in a PT scheme, including sampling, analyte stability, sample inhomogeneity in real samples, and other concentration levels. This should of course be dealt with in the validation but there may be more, new, data to take into account.

**Case 3: Underestimate** – The estimated uncertainty is clearly lower than the laboratory is performing. The uncertainty estimate should be revised to obtain a more realistic estimate (see section 3.2).

### **Checks based on comparison of results with other laboratories (v)**

The same principles used for checks based on proficiency testing can be used for comparison with other laboratories after collaborative measurement of several test items.

### **Comparison with other uncertainty estimates (vi)**

#### **a) Comparison of standard uncertainties**

For the purpose of comparison, uncertainty estimates are expressed as a standard uncertainty  $u$  (or as a standard deviation  $s$ , respectively). Any other uncertainty estimates (e.g. expressed as an expanded uncertainty or the width of a confidence interval) are converted into a standard uncertainty (standard deviation) first.

When checking whether two uncertainty estimates agree or disagree, one should keep in mind that the precision of uncertainty estimates is often very limited. For example, for an empirical standard deviation determined from 10 repeated measurements, the coefficient of variation is 24 %, and  $F$ -tests on two such standard deviations would not be considered significant with standard deviations differing by less than a factor of about 1.8. It would therefore be unreasonable to expect different uncertainty estimates to agree very closely. For estimates based on a very small number of data, the expected level of agreement is even more limited. When a formal statistical tool is needed, the  $F$ -test may be used for examining whether two variances (squared standard deviations) are significantly different.

**EXAMPLE 1** – For a specified measurement procedure for the analysis of water, the uncertainty is estimated using data from in-house investigations of precision and bias. Precision has been monitored on a routine basis (control chart), giving a long-term standard deviation of 2.8 % relative. The bias was investigated using a reference water sample. For these measurements ( $n = 10$ ) the standard deviation is 3.6 % relative. Are these estimates compatible, or is this difference significant?

Considering the imprecision of empirical standard deviations (see previous paragraph), the difference is clearly insignificant. Employing an  $F$ -test (for 10 and 100 measurements, respectively) would allow for a current standard deviation of  $2.8 \% \times \sqrt{1.97} = 3.9 \%$  at a significance level of 95 %. The difference would be rated as significant only for substantially larger values of  $F$ .

**EXAMPLE 2** – An in-house precision study of a standard test method, carefully designed to cover all variations in measurement conditions to be expected in routine use, provided a standard deviation of 7.5 % relative. The standard specifies a repeatability standard deviation of  $s_r = 5 \%$  and a reproducibility standard deviation of  $s_R = 12 \%$ . How does the result of the in-house study compare with the performance data specified in the standard?

The 7.5 % obtained in-house constitutes an "intermediate precision" between the precision at repeatability conditions and the precision at reproducibility conditions. As such, 7.5 % fits very well between the limiting precision data  $s_r = 5 \%$  and  $s_R = 12 \%$ . In addition, the laboratory could determine the in-house repeatability standard deviation and check whether this is less or equal to  $s_r = 5 \%$ .

## b) Comparison of expanded uncertainties at a specified level of confidence

For the purpose of such a comparison, uncertainty estimates are expressed as an expanded uncertainty  $U = k \times u$ , where the coverage factor  $k$  is chosen according to the specified level of confidence (or as the half-width  $t \times s$  of a confidence interval for the specified level of confidence). The GUM provides guidance on how to determine the coverage factor for a given level of confidence, considering the relevant degrees of freedom.

As a pragmatic recipe for the common confidence level of 95 %, the GUM recommends a coverage factor of  $k = 2$ , independently of the number of measurements involved. This may appear questionable for standard deviations obtained from less than 10 measurements, but the actual uncertainty estimate will often be based on additional supporting information.

NOTE - Comparison of expanded uncertainties  $U = k \times u$  with the same  $k$  is of course equivalent to comparison of standard uncertainties, while this is not the case when the coverage factors happen to be different. An even more subtle comparison could be made using probability distributions generated by Monte-Carlo simulation instead of standard uncertainties and coverage factors.

Occasionally it may be difficult to identify the intended level of confidence. Then a default level of 95 % may be stipulated for common technical applications, and 99 % for safety-related applications. In any such case the stipulated confidence level must be specified.

EXAMPLE 3 – Two measurement procedures for workplace air measurement (VOC) are investigated as to whether they conform to regulatory requirements. The relevant regulation (EN 482) specifies that the overall uncertainty of screening methods shall not exceed 30 % relative. Taking this target uncertainty to be an expanded uncertainty at a confidence level of 95 %, the task then is to estimate the 95 % expanded uncertainty for each of the two procedures and to compare them with the target uncertainty.

The first method employs activated test tubes for single use. For the type of test tubes concerned, a validation study carried out by a research institute reported a maximum indication error of 20 % and a between-items indication reproducibility  $R$  of 20 %. Since no other information is available, the first statement is taken to refer to a confidence level of 95 %. The second statement is converted into an expanded uncertainty utilising the information that  $R = 2\sqrt{2} \times s_R = 20$  %. This gives  $U = 2 \times s_R = R / \sqrt{2} = 14$  %. Combining these two estimates (same confidence level) directly gives an expanded (95 %) overall uncertainty of  $(20^2 + 14^2)^{1/2} = 24$  %. The second method operates by adsorption on charcoal tubes, thermal desorption and quantification by gas chromatography with flame ionisation detection (FID). An in-house validation study gave a combined standard uncertainty of  $(10^2 + 4^2)^{1/2} = 11$  %. Applying a coverage factor of  $k = 2$  gives an expanded (95 %) uncertainty of 22 %. So both methods comply with the requirement and deliver about the same measurement uncertainty.

## **3.2 ADJUSTING UNCERTAINTY ESTIMATES**

### **3.2.1 Introduction**

It is not unusual to find that uncertainty estimates based on within-laboratory validation data prove too small when compared with interlaboratory data, such as PT data. In these circumstances, it is generally necessary to increase the laboratory's estimate of uncertainty. Conversely, new measurement equipment, improvements in measurement quality, or improved validation data may justify a reduced uncertainty estimate. This section describes some simple approaches to revising an uncertainty estimate.

In addition to the guidance below, there is a good short paper *Is my uncertainty estimate realistic* (AMC Brief 15) from the UK Royal Society of Chemistry (reference [27] in the Annex) which discusses this issue in detail.

### **3.2.2 Principles**

In general, an uncertainty estimate includes one or more terms associated with random variation (random effects), and one or more terms associated with effects which vary little within the laboratory (systematic effects). Some approaches, particularly in testing, base the entire estimate on two contributions - one associated with random effects and one associated with systematic effects - with additional contributions only if required (e.g. Nordtest report 537 [6], ISO TS 21748 [9], see Annex).

Where the budget includes more than five or six significant contributions, or demands application of the uncertainty propagation approach described in detail in the GUM, it is recommended that the uncertainty estimate should be reviewed and recalculated in detail.

However, where uncertainty budgets are constructed from two or three substantial contributions, particularly those derived from studies of method performance, and those contributions are simply combined 'in quadrature', it is often simple to replace one of these contributions with an improved estimate. This approach is described below.

### 3.2.3 Model for simple adjustments

The model assumed for simple updates to a basic uncertainty budget is based on that of ISO/TS 21748 [9]. It will be assumed that the uncertainty estimate is of the form:

$$(3.3) \quad u_c^2 = s^2 + u^2(\hat{\delta}) + \sum_i c_i^2 u^2(x_i)$$

where  $u_c$  is the combined standard uncertainty,  $s$  an estimate of precision based, for example, on inter-laboratory or long-term within-laboratory precision,  $u(\hat{\delta})$  an estimate of uncertainty associated with bias for the implementation of the method, and the summation is over a series of additional contributions necessary for the particular material or circumstance.

Note that in many cases, the uncertainty  $u_c$  and the terms on the right of the model expression above may be relative to the value found. That is, the entire expression may be presented in terms of relative standard deviations (with coefficients  $c_i$  adjusted accordingly). This does not affect the following approach.

### 3.2.4 Adjusting the precision term

If an independent and improved estimate of the precision is obtained and shown to be valid by (among others) the checks described above, the term  $s^2$  may be replaced directly by the revised contribution.

The term should, where necessary, be estimated so as to allow for variation with level of response; for example, precision contributions often increase in absolute magnitude as the response increases, and show a limiting precision at lower response. It is therefore often useful and convenient to construct a suitable predictive model for the precision term. ISO/TS 21748 [9] and the Eurachem/CITAC guide [2] provide appropriate models.

### 3.2.5 Adjusting the bias term

As above, if an independent improved estimate of the uncertainty associated with bias becomes available, this may replace the appropriate term above.

Typically, uncertainties associated with bias will be derived from studies on well-characterised test items, often by some simple expedient such as taking the standard deviation of observed relative differences (see the Eurachem/CITAC guide [2], Example A4 for an example). Where this is the case, it is often possible to recalculate the contribution with new test items included, rather than simply replace one study with another of similar size. Where this approach is proposed, it is essential that suitable tests be undertaken to show that the new data are commensurate with the original data; in particular, that the dispersion of differences has not changed significantly.

## 3.3 COMPARISON OF UNCERTAINTY ESTIMATES

Estimation of measurement uncertainty is a difficult and laborious exercise. But having finally completed this task is often not the end of the road. For an in-house uncertainty estimate for the results obtained by a specified measurement method on specified samples or specimens, additional tasks could be to compare this estimate with, for example:

- *a previous uncertainty estimate, obtained using the same approach*  
e.g. when an uncertainty estimate obtained in an initial in-house method validation study is re-evaluated to confirm its validity;
- *an independent uncertainty estimate, obtained using a different approach*  
e.g. when a preliminary uncertainty estimate is checked using another approach and/or other data;

- *an uncertainty estimate for a similar measurement method, obtained in another laboratory*  
e.g. when benchmarking an uncertainty estimate against results published by another recognised laboratory;
- *precision data obtained in an inter-laboratory method validation study*  
e.g. when benchmarking an uncertainty estimate for a standardised method against data published in the respective standard;
- *an uncertainty estimate for an alternative measurement method*  
e.g. when comparing the performance of several methods for the same measurement;
- *a target uncertainty specified by a customer or in a regulatory document*  
e.g. when examining the fitness of a method for a specified purpose.

In most of these cases two different levels of comparison are relevant:

- (1) Comparison of data
- (2) Comparison of sources

This section mainly addresses the comparison of complete uncertainty estimates intended to comprise the contributions from all relevant uncertainty sources. However it is also applicable to incomplete estimates, covering only part of the relevant uncertainty sources.

### 3.3.1 Comparison of data

"Comparison of data" implies the comparison of numerical estimates of uncertainty, such as standard uncertainty or expanded uncertainty. This is covered in detail in section 3.1.2 (item vi), and is not discussed further here.

### 3.3.2 Comparison of sources

When comparing two uncertainty estimates, an essential piece of information is an overview of the uncertainty sources covered by these estimates. Here "coverage of an uncertainty source" means that the contribution of this source under routine measurement conditions is included. If the uncertainty was estimated using the modelling approach, this implies that the effect of the uncertainty source is included in the model, and the uncertainty attributed to the associated input quantity is realistic. If the uncertainty was estimated using data from a whole-method performance investigation, coverage implies that in that investigation the effect associated with the uncertainty source was varied to the same extent as operative in routine measurements.

If the major uncertainty sources of a measurement procedure are known, and uncertainty sources are either covered or not included at all, uncertainty estimates may be compared source-wise. In principle, this would also be possible for partial coverage of uncertainty sources, but such uncertainty estimates would be of limited value and should rather be improved.

Given two incomplete uncertainty estimates for the same measurement procedure, source-wise comparison provides a sound basis for deriving an improved estimate (see Chapter 3). To this end, the two estimates would be added, and the common part be subtracted, in quadrature (see Example 4 in Chapter 4). Often it will be difficult to estimate the common part of two uncertainty estimates. The subtraction must then be omitted, noting that the uncertainty may be over estimated due to double counting of some contributions.

Uncertainty estimates based on whole-method performance investigations (see Chapter 1) include data on precision and bias of the measurement method. In comparing such estimates with estimates obtained using the analytical approach (detailed uncertainty budget, based on a mathematical model), it is important to know whether an uncertainty source contributes a random effect or a systematic effect. The comparison of sources covered should then be carried out separately, for precision against "random" uncertainty contributions, and bias against "systematic" uncertainty contributions.

## Chapter 4: Examples

*This chapter presents a range of examples. These are case studies from various testing fields, where different approaches were used to evaluate the relevant uncertainty and the results so obtained were compared.*

- 1 – Uncertainty evaluation for the determination of lead in a biological tissue
- 2 – Uncertainty evaluation for the determination of the Flakiness index of aggregates
- 3 – Uncertainty evaluation for the measurement of the aperture size of wire screen products
- 4 – Comparing and combining calculated uncertainty and experimental variability – a study of sample preparation uncertainty in chemical analysis
- 5 – Ammonium determination in water – Verification of uncertainty estimates
- 6 – Measurement uncertainty in optical emission spectrometry
- 7 – Rockwell hardness testing
- 8 – Determination of Cadmium and Phosphorous in agricultural top soil – comparison of evaluation methods focussed on uncertainty from sampling
- 9 – Pesticide residues in foodstuffs
- 10 – Uncertainty evaluations in the environmental sector – summary of a comprehensive study

In the examples, the comparison of uncertainty estimates is focussed on *standard uncertainties*. The reason for this is that comparison of *expanded uncertainties* either carries over trivially from standard uncertainties, if the same coverage factor is used in all cases, or adds considerable complications, if different coverage factors, tailored to the relevant degrees of freedom are used, see section 3.1.2 (item vi).

References specific to the example are given at the end of each example.

Where required, numbering of equations starts with (1) in each example.

The decimal point is used for numerical data.

**EXAMPLE 1: UNCERTAINTY EVALUATION FOR THE DETERMINATION OF LEAD IN A BIOLOGICAL TISSUE**

Sector	Measurand / Matrix	Technique	Approaches for uncertainty evaluation	
Food	Lead content / Biological tissue	ICP-MS	Modelling	Interlaboratory validation

**1 Scenario and specification of the measurand**

This example presents an extract from a study published in [1], where the goal was to compare different approaches in estimating uncertainty when determining very low levels of trace elements in a biological sample. This type of analysis provides a good example because many factors, such as sample heterogeneity and/or stability or sample preparation (acid ashing or digestion), can be invoked. Another interesting feature of this study is that expert laboratories were involved, all of them having already implemented quality control systems for many years. This means that many causes of variations due to the environment, personnel or equipment were correctly controlled and, supposedly, minimized.

The example focuses on the determination of lead in a dried mussel tissue. The concentration of lead is expressed as a mass fraction (in mg/kg) in the dried mussel tissue.

**2 Measurement procedure**

Eleven European reference laboratories participated in the interlaboratory comparison. Each participating laboratory applied its own method and made six independent replicate determinations on two different bottles and on two different days. The techniques used are summarized below :

- Inductively couple plasma - mass spectrometry (ICP-MS) (external calibration: 2 labs; calibration by standard addition: 2 labs; isotope dilution (ID-ICPMS): 2 labs)
- Thermal ionisation mass spectrometry (TIMS) (isotope dilution: 1 lab)
- Atomic absorption spectrometry (AAS) (standard addition: 1 lab)
- Voltammetry (2 labs)

*Note: One laboratory did not report the technique used.*

**3 Measurement uncertainty evaluation*****Modelling approach***

The modelling approach was used to evaluate the measurement uncertainty of one of the participants in the interlaboratory comparison which used ID-ICPMS.

Isotope dilution consists in adding a 'spike' (an additional quantity of material), enriched with an isotope of the element to be analysed, to the sample. The original concentration of the element in the sample is calculated from the measurements of the isotope ratios in the sample before and after treatment, according to a detailed equation from which the uncertainty budget was established.

The combined standard uncertainty was evaluated to 0.033 mg/kg for a concentration of lead of 2.010 mg/kg. This corresponds to a relative standard uncertainty of 1.7%.

***Interlaboratory comparison approach***

The data from the interlaboratory comparison were evaluated using the tools of ISO 5725-2, and the values of the repeatability standard deviation  $s_r$ , the between-laboratories standard deviation  $s_L$  and the reproducibility standard deviation  $s_R$  were obtained ( $s_R^2 = s_r^2 + s_L^2$ ).

The uncertainty for the laboratory for which the modelling approach was performed, was calculated as a root sum of squares of two components: the standard deviation of the results of this laboratory and the between-laboratory standard deviation.

This calculation gave a standard deviation of 0.087 mg/kg, corresponding to a relative standard uncertainty of 4.4%.

#### 4 Comparison of uncertainty estimates

Uncertainty estimates obtained from the approaches	
Modelling	Interlaboratory study
1.7 %	4.4 %

All data are given as relative standard uncertainties in %.

#### 5 Conclusions

The previous table shows that the uncertainties differ by a factor of about 3. Obviously, they do not take the same sources of uncertainty into account and it is important to try to understand these discrepancies. The difference in uncertainty estimates between the two approaches mainly comes from the uncertainty sources such as method, environment, operators, sample handling which are much more influential in the interlaboratory comparison than in the evaluation based on the modelling approach where only one laboratory was included (one method, one operator and small variations of the effects from other uncertainties sources compared to the interlaboratory comparison). Despite the well-recognized skill of the personnel of the participating laboratories, the differences in sample digestion techniques, or the variations due to the different instruments used in the collaborative study can change the uncertainty by a factor of 3 as illustrated when comparing the two approaches.

It also has to be mentioned that ISO 5725 normally applies to one method, which was not the case for this study. The different analytical techniques that were used have different performance in terms of precision. This can also explain the discrepancies between the modelling approach performed on one single method and the interlaboratory comparison approach performed on different methods.

#### 6 References

- [1] Feinberg M. et al., *Accred. Qual. Assurance* (2002), 7, 403 – 408

## EXAMPLE 2: UNCERTAINTY EVALUATION FOR THE DETERMINATION OF THE FLAKINESS INDEX OF AGGREGATES

Sector	Measurand / Matrix	Technique	Approaches for uncertainty evaluation	
Construction	Particle shape/ Pile of aggregates	Geometrical / Mass method	Modelling	Interlaboratory validation

### 1 Scenario and specification of the measurand

Concrete and road construction industries handle aggregates in large quantities. The properties of these materials are determined by geometrical, mechanical and chemical tests. One of the geometrical criteria for their use is the shape of the fragments. Particles with a more or less cubic shape are in many cases easier and more cost-effective to process than those with a flat or longish form. Moreover, in concrete construction the need for binders increases when particles with larger surface areas are used which in turn involve higher material costs. The measurand is the mass fraction of flaky particles in an aggregate material.

### 2 Measurement procedure

A test method for quantitative determination of the proportion of unfavourably shaped particles is described in EN 933-3. The material is washed, dried and weighed ( $M_1$ ) and then sieved on metal plates with square holes. The masses of the different fractions are noted. After that the particular fractions are put on bar sieves that are specified for the given sizes. Particles with a clearly higher length  $l$  in comparison to the width  $w$ , i.e. flat or longish aggregates with a size of about  $l : w > 3 : 1$ , pass the bar sieve ( $M_2$ ). The average of percentage by mass of the fractions gives the Flakiness index  $FI$  as

$$FI = \frac{M_2}{M_1} \cdot 100\%$$

$FI$  is traditionally and in the standard expressed in  $M$ -%.

### 3 Measurement uncertainty evaluation

#### 3.1 Modelling approach

The calculation was based on an analysis of uncertainty sources according to GUM for  $FI = 9 M$ -%. Based on a simple mathematical model the test process was subdivided into nine steps:

No.	Step of test procedure
1	Sampling and reduction of samples
2	Drying of test sample
3	Weighing of test sample
4	Sieving on square holes in defined size fractions
5	Weighing of single size fractions
6	Sieving of each size fraction on bar sieves
7	Weighing of passing masses of each size fraction
8	Calculation of results for each size fraction
9	Statement of results

These uncertainty components were used for the calculation of the combined standard uncertainty. The uncertainty was not a function of  $F_l$  within the range considered. More detailed information can be found in [2].

### 3.2 Interlaboratory validation approach

EN 933-3 provides data for the reproducibility limit. For values between 8  $M\%$  and 20  $M\%$  the constant figure  $R = 5 M\%$  (i.e.  $s_R \approx 1.8 M\%$ ) is given.  $M\%$  is weight-% according to the standard.

## 4 Comparison of uncertainty estimates

Uncertainty estimates obtained from the approaches	
Modelling	Interlaboratory validation
2.6 %	1.8 %

All values are given in  $M\%$ .

## 5 Conclusions

The presumptions for the testing conditions in the modelling approach are valid for a day-to-day laboratory situation. Both the sampling and the measures to create a very high homogeneity through reducing the material for the interlaboratory comparison are rather different from that situation. Precision data usually only partly encompass the influences from sampling and reducing. The standard deviation should therefore be considerably smaller than a combined standard uncertainty stemming from the above modelling approach. The figure of  $u_c \approx 2.6 \%$  can therefore be taken as a realistic answer to the initial question.

## 6 References

- [1] EN 933-3:1997/2003, *Test for geometrical properties of aggregates – Part 3: Determination of particle shape – Flakiness index*
- [2] Hinrichs, W. (January 2003), *Example for estimating the measurement uncertainty in building materials testing*, [http://141.63.4.16/docs/EL\\_04-11\\_03\\_419.pdf](http://141.63.4.16/docs/EL_04-11_03_419.pdf)



The left photograph shows piles of aggregates used for road construction. On the right are the instruments for the determination of particle shape. The bar sieve is necessary for testing according to EN 933-3, the special version of a calliper is necessary for an alternative test according to EN 933-4.

### EXAMPLE 3: UNCERTAINTY EVALUATION FOR THE MEASUREMENT OF THE APERTURE SIZE OF WIRE SCREEN PRODUCTS

Sector	Measurand / Matrix	Technique	Approaches for uncertainty evaluation	
Industrial product	Mesh / Wire screen	Geometrical method	Modelling	Interlaboratory validation

#### 1 Scenario and specification of the measurand

Wire screen and wire cloth products are used in many technical fields. They are important in filtration and separation processes as well as in press printing. The range of apertures usually lies between 20  $\mu\text{m}$  and 125 mm.

The test procedure for the aperture  $w$  in product standards does not go into detail. According to ISO 3310-1 one has to “measure the aperture sizes using appropriate equipment having a precision of reading of 1  $\mu\text{m}$  or one fourth of the tolerance for average aperture size, whichever is the greater.” Additional information includes recommendations for the magnification when optical methods are used and provisions for spot checks of apertures for measurements of warp and weft dimensions.

#### 2 Measurement procedure

In practice the measurement conditions and instruments are very different due to the wide range of apertures. In this example results are considered for measurements of steel wire products with a digital vernier calliper with a readability of 0.01 mm in a range of aperture size from 2 mm to 32 mm.

#### 3 Measurement uncertainty evaluation

##### 3.1 Modelling approach

The analysis according to the GUM included the following uncertainty sources:

- Calibration of the vernier calliper
- Handling of the vernier calliper
- Personnel
- Test procedure
- Documentation (i.e. transcription errors)

It is a well-known fact that the uncertainty arising from variation of aperture sizes within samples strongly contributes to the overall uncertainty of the test result. The interlaboratory comparison was therefore tailor-made in such a way that an influence of the uncertainty of the sample was negligible. The uncertainty was determined, using the modelling approach, both for favourable and unfavourable measurement conditions.

##### 3.2 Interlaboratory validation approach

Five samples of metallic woven wire with aperture sizes of approximately 2 mm, 4 mm, 8 mm, 16 mm and 32 mm had been stabilized and sent to 18 international participants for an interlaboratory comparison test. All mesh in a sample had to be measured in order to make the influence of the inhomogeneity of the material negligible. The repeatability and reproducibility standard deviations of measurements were evaluated according to ISO 5725-2.

#### 4 Comparison of uncertainty estimates

Uncertainty estimates obtained from the approaches	
Modelling	Interlaboratory PT
0.034 / 0.107	0.03 / 0.09

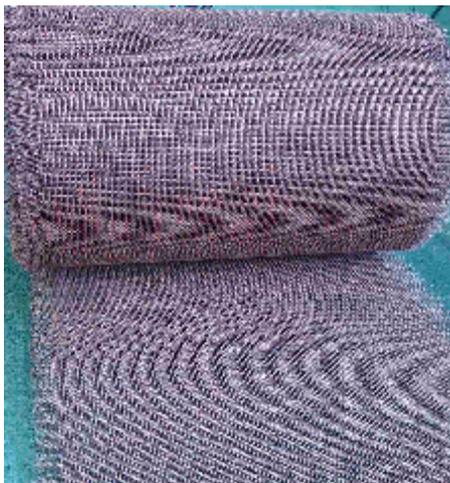
In each case, the lower figures are valid for favourable conditions (precise instrumentation, skilled personnel familiar with measurements on wire screens etc.), and the higher figures are valid for unfavourable conditions (unknown calibration standard of the instruments used, restricted readability of the measuring device etc.). All values are given in mm.

## 5 Conclusions

The figures for the uncertainty under favourable conditions from the modelling approach ( $\approx 34 \mu\text{m}$ ) and the standard deviation under repeatability conditions ( $\approx 30 \mu\text{m}$ ) from the interlaboratory comparison are similar. It is plausible that favourable conditions are generally given when a measurement is repeated. The mathematic model according to GUM can therefore be considered to produce a correct combined standard uncertainty.

The combined standard uncertainty for unfavourable conditions ( $\approx 107 \mu\text{m}$ ) is about 20% larger than the standard deviation under reproducibility conditions ( $\approx 90 \mu\text{m}$ ). In this field, performance in interlaboratory comparison tests is usually better than routine laboratory testing. It is therefore likely that the reproducibility standard deviation is smaller than it would be under unfavourable conditions. Moreover the presumptions for the calculation include systematic effects such as uncalibrated callipers. But as not all participants used uncalibrated callipers in the interlaboratory comparison it is realistic that the uncertainty is closer to a maximum than to a minimum limit.

Additional remark: The factor of 3 difference between the standard deviations of repeated and reproduced measurements is considerable. That is no surprise: While in-house test results have proven to be a good basis for maintaining consistency in production control, third-party testing of products is known to show wider dispersion in this sector.



Left photograph: A batch of industrial wire screen

Right photograph: A sample used in the interlaboratory comparison test

## 6 References

- [1] ISO 3310-1:2000, *Test sieves – Technical requirements and testing – Part 1: Test sieves of metal wire cloth*
- [2] ISO 9044:1999, *Industrial woven wire cloth – Technical requirements and testing*
- [3] ISO 14315:1997, *Industrial wire screens – Technical requirements and testing*
- [4] Hinrichs, W. (2005), *Characterization of a unified measurement method and validation of specific procedures for the determination of the mesh size of steel wire screens* (in German); Schriftenreihe des Instituts für Baustoffe, Massivbau und Brandschutz der TU Braunschweig, Heft 184, March 2005
- [5] DIN ISO 5725-2:2002-12, *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*

#### EXAMPLE 4: COMPARING AND COMBINING CALCULATED UNCERTAINTY AND EXPERIMENTAL VARIABILITY – A STUDY OF SAMPLE PREPARATION UNCERTAINTY IN CHEMICAL ANALYSIS

Sector	Measurand / Matrix	Technique	Approaches for uncertainty evaluation	
Chemical analysis	Trace metals / Various	Isotope dilution MS	Modelling	Single laboratory validation

### 1 The problem

In chemical analysis, mathematical models of the measurement process are often incomplete. That is, the input quantities of the equation (or the algorithm) used to calculate the final result from the various data obtained, fail to cover all relevant uncertainty sources. As a consequence the combined uncertainty calculated from the uncertainty attributed to the input quantities underestimates the real uncertainty. A first hint of an incomplete uncertainty budget may be obtained by comparing the combined standard uncertainty calculated from the identified uncertainty contributions with the standard deviation obtained from a series of independent replicate determinations. If the calculated standard uncertainty is smaller than the empirical standard deviation, there is reason to believe that there are uncertainty contributions missing in the uncertainty budget. Often this will be the case with the uncertainty contributions from sample preparation and sample inhomogeneity. In principle, these effects could be included in the measurement equation (or algorithm) as correction factors, but often the associated uncertainty cannot be estimated directly. Therefore a pragmatic approach has been developed for estimating the combined uncertainty contribution from sample preparation and related effects from the standard deviation of results obtained on parallel sub-samples.

### 2 Principle

For the measurement concerned, two uncertainty estimates are determined as follows:

$u_{model}$  the combined standard uncertainty calculated from a mathematical model of the measurement process, i.e. from the uncertainty budget of the measurement equation using the uncertainty of the input data;

$s_{rep}$  the empirical standard deviation obtained from replicate measurements.

If  $u_{model}$  compares well with  $s_{rep}$ , this may be taken to confirm the modelling estimate. However if  $u_{model}$  is smaller than  $s_{rep}$ , the modelling estimate is obviously deficient. In this case a revised uncertainty estimate is obtained by combination.

When combining the two estimates, care has to be taken to avoid double counting. To this end, the combined contribution from all those uncertainty sources which are common to both,  $u_{model}$  and  $s_{rep}$  has to be determined and accounted for. This is done as follows:

$$(1) \quad u_{revised} = \sqrt{u_{model}^2 + s_{rep}^2 - u_{var}^2}$$

In this equation,  $u_{revised}$  is the combined (and hopefully complete) standard uncertainty, covering the uncertainty sources contributing to  $u_{model}$  and  $s_{rep}$ . The third term in the root sum,  $u_{var}$  is the combined contribution from all the common uncertainty sources of  $u_{model}$  and  $s_{rep}$ . This common uncertainty is obtained from a modified uncertainty calculation, restricted to those input data in the uncertainty budget whose measurement error contributes to the variability of replicate measurements.

### 3 Procedure

The procedure below is used for cases where it is known in advance that the uncertainty budget is incomplete – sample preparation is not included. The objective is (i) to investigate

whether the contribution from sample preparation is relevant, (ii) to estimate the size of this contribution and (iii) to include the contribution in a revised uncertainty estimate.

**Step 1:** A standard uncertainty  $u_{\text{model}}$  is calculated by propagation of the uncertainties attributed to the input data in the measurement algorithm. In this uncertainty budget, the contribution from sample preparation is not included.

**Step 2:** A standard deviation  $s_{\text{rep}}$  is determined from the results obtained on parallel sub-samples. Here the uncertainty from sample preparation is included among other contributions to the variability.

**Step 3:** The data obtained in step 1 and step 2 are compared. If  $u_{\text{model}} \geq s_{\text{rep}}$ , there is no indication that the contribution from sample preparation is relevant, and the procedure may be terminated (but see note below). If  $u_{\text{model}} < s_{\text{rep}}$ , the modelling estimate is deficient, indicating that sample preparation contributes significantly.

**Step 4:** Using the uncertainty budget of the measurement algorithm from step 1, a restricted uncertainty calculation is performed, including only the uncertainties of those input data whose measurement error contributes to the variability of replicate results recorded in step 2.

This is done by putting the other uncertainties equal to zero, giving a combined standard uncertainty  $u_{\text{var}}$ .

**Step 5:** A reduced standard deviation  $s_{\text{samp}}$  accounting for the variability due to sample preparation is calculated according to

$$(2) \quad s_{\text{samp}} = \sqrt{s_{\text{rep}}^2 - u_{\text{var}}^2}$$

**Step 6:** The revised standard uncertainty  $u_{\text{revised}}$  is calculated by combination of  $u_{\text{model}}$  as determined in step 1 and  $s_{\text{samp}}$  as determined in step 5

$$(3) \quad u_{\text{revised}} = \sqrt{u_{\text{model}}^2 + s_{\text{samp}}^2} = \sqrt{u_{\text{model}}^2 - u_{\text{var}}^2 + s_{\text{rep}}^2}$$

If the standard uncertainty for a mean of  $n$  replicate measurements is required, the uncertainty calculation is modified as follows:

$$(4) \quad u_{\text{revised}} = \sqrt{u_{\text{model}}^2 - u_{\text{var}}^2 + \frac{s_{\text{rep}}^2}{n}}$$

The procedure above may be refined by changing the sequence of step 3 and step 4 and comparing  $s_{\text{rep}}$  with  $u_{\text{var}}$ , the modelling estimate accounting for measurement variability. Even if  $u_{\text{model}} > s_{\text{rep}}$ ,  $u_{\text{var}} < s_{\text{rep}}$  indicates that the uncertainty budget is deficient and the contribution from sample preparation is relevant.

#### 4 Discussion

The standard deviation  $s_{\text{samp}}$  combines contributions from all random effects which are not accounted for in the measurement equation. These may comprise other effects than sample preparation. The composition of  $s_{\text{samp}}$  is, however, irrelevant for most practical purposes. Decomposition of  $s_{\text{samp}}$  could be achieved using an appropriate replication design and Analysis of Variance. For example, the variability due to sample preparation and measurement variability can, in principle, be separated by carrying out replicate measurements on each prepared sample. In the particular case of IDMS measurements, one of the measurement operations, addition of a measured amount of isotopically enriched material, is carried out before sample preparation and measurements therefore cannot be carried out independently of sample preparation.

The input uncertainties for calculating  $u_{\text{var}}$  have to be restricted to random effects. For example, if weighings take place with each sample preparation, only the weighing reproducibility should be included in the calculation, because the systematic effect from balance calibration does not vary between replicate preparations. However, in practice a

subdivision of input uncertainties into random and systematic parts is not necessary, because either the random or the systematic effects are dominant.

For the example of isotope dilution analysis shown below, the approach has two objectives: (a) to obtain a comprehensive uncertainty, starting from an incomplete uncertainty budget, and (b) to estimate the sample preparation uncertainty which was missing. If (a) is the only objective, a simpler approach would be to combine the empirical standard deviation from replicate determinations with the combined standard uncertainty obtained from a reduced uncertainty budget, restricted to contributions from systematic effects.

## 5 Application: Determination of Cadmium by isotope dilution mass spectrometry with thermal ionisation (ID-TIMS)

This case of isotope dilution mass spectrometry allows for a comparatively simple measurement equation. The main variables are measured isotope ratios and weighing results, with uncertainties dominated by random variation and determined as type A uncertainties. Further variables are tabular values, for which type B uncertainty estimation applies.

The measurement result is calculated as follows:

$$C_x = C_y \cdot \left[ \frac{M_x \cdot m_y}{M_b \cdot m_x \cdot h_{x;b}} \cdot \frac{R_y - R_{xy}}{R_{xy} - R_x} \right] - \text{blank}$$

where

$$m_y = w_y \cdot f_{\text{buo};y} \cdot f_{\text{lin}} \cdot f_{\text{temp}} \cdot f_{\text{cal}} \quad ; \quad m_x = w_x \cdot f_{\text{buo};x} \cdot f_{\text{lin}} \cdot f_{\text{temp}} \cdot f_{\text{cal}} \cdot f_{\text{lod}}$$

$$f_{\text{buo};x} = \frac{1 - \delta_{\text{air}} / \delta_{\text{ref}}}{1 - \delta_{\text{air}} / \delta_{\text{samp}}} \quad ; \quad f_{\text{buo};y} = \frac{1 - \delta_{\text{air}} / \delta_{\text{ref}}}{1 - \delta_{\text{air}} / \delta_{\text{spike.sol}}} \quad ; \quad f_{\text{lod}} = \frac{100 - \text{lod}}{100}$$

Quantity	Definition
$f_{\text{buo};x}, f_{\text{buo};y}$	buoyancy correction factors for the weights of sample and spike
<i>blank</i>	result blank
$h_{x;b}$	isotope abundance of the spike isotope in the sample
$M_b, M_x$	molar mass of the spike isotope and analyte, respectively
$m_x, m_y$	masses of sample and spike solution in the blend
$w_x, w_y$	weights of sample and spike solution in the blend
$R_{xy}, R_y, R_x$	observed isotope ratios (a/b) of the blend, the spike and the sample
$f_{\text{lod}}$	correction factor for the dry mass of the sample (loss on drying)
<i>lod</i>	moisture content of the sample (loss on drying) in mass %
$f_{\text{evap}}$	evaporation correction factor for the spike solution
$\delta_{\text{air}}, \delta_{\text{ref}}, \delta_{\text{sample}}, \delta_{\text{spike.sol}}$	densities of air, reference mass piece, sample and spike solution
$C_y$	amount content of the spike isotope in the spike solution
$f_{\text{lin}}$	correction factor for balance linearity
$f_{\text{temp}}$	correction factor for temperature coefficient of balance sensitivity
$f_{\text{cal}}$	correction factor for reference mass piece for balance calibration
$C_x$	result - mass content of the analyte in the sample

The isotopes utilised are a = Cd<sup>112</sup> and b = Cd<sup>113</sup>, and Cd<sup>113</sup> is used as spike isotope.

The balance correction factors  $f_{\text{lin}}$ ,  $f_{\text{temp}}$  and  $f_{\text{cal}}$  are only introduced to enable accounting for the associated uncertainties. Their default value is unity, but they have non-zero uncertainty.

In this application Cadmium was measured in sewage sludge in the framework of an international interlaboratory comparison. The measurement uncertainty was evaluated as follows.

1. The standard uncertainty  $u_{\text{model}}$  was calculated (using numerical derivatives) from the uncertainty of all variables in the IDMS equation using the computer program "GUM workbench".

$$u_{\text{model}} = 0.017 \text{ } \mu\text{g/g}$$

2. The standard deviation  $s_{\text{rep}}$  was calculated from the results obtained on 5 independently processed sub-samples (standard deviation of the 5 values).

$$s_{\text{rep}} = 0.041 \text{ } \mu\text{g/g}$$

3. As the standard deviation  $s_{\text{rep}}$  was larger than the standard uncertainty  $u_{\text{model}}$ , the modelling estimate was clearly deficient, indicating that sample preparation contributed significantly. Therefore the procedure was continued.
4. The uncertainty  $u_{\text{var}}$  was calculated like in step 1 using the uncertainty budget for the IDMS equation restricted to those variables whose values are determined independently for each sub-sample. These are: the weight of the sub-sample ( $w_x$ ), the weight of the  $\text{Cd}^{113}$ -spike solution ( $w_y$ ) and the measured isotope ratio of the blend ( $R_{xy}$ ). The uncertainties of all the other variables (which are the same for all blends) were put to zero.

$$u_{\text{var}} = 0.013 \text{ } \mu\text{g/g}$$

5. The reduced standard deviation ( $s_{\text{samp}}$ ) was obtained by difference from  $s_{\text{rep}}$  and  $u_{\text{var}}$

$$s_{\text{samp}} = (0.041^2 - 0.013^2)^{1/2} = 0.039 \text{ } \mu\text{g/g}.$$

6. The revised uncertainty  $u_{\text{revised}}$  was calculated by combination of  $u_{\text{model}}$  and  $s_{\text{samp}}$

$$u_{\text{revised}} = (0.039^2 + 0.017^2)^{1/2} = 0.042 \text{ } \mu\text{g/g}.$$

The uncertainty calculation above refers to the result of a single determination. For a mean value of  $n$  replicate determinations the standard deviation  $s_{\text{rep}}$  would be replaced by  $s_{\text{rep}}/\sqrt{n}$ . This gives  $u_{\text{revised}} = (0.041^2/5 + 0.017^2 - 0.013^2)^{1/2} = 0.021 \text{ } \mu\text{g/g}$ .

### EXAMPLE 5: AMMONIUM DETERMINATION IN WATER – VERIFICATION OF UNCERTAINTY ESTIMATES

Sector	Measurand / Matrix	Technique	Approaches for		
			estimation	verification	
Environmental	Ammonium / Seawater	Photometry Indophenol method	Single laboratory validation data and PT data	Internal QC	PT

Here we present some examples of different ways in which a laboratory may verify its uncertainty estimates.

#### 1 Approaches for verifying uncertainty estimates

The example is from chemical measurement – ammonium in water in the range 70 to 300 µg/l using the photometric indophenol method specified in ISO 11732. The measurement uncertainty is required to be 10 % of the reported ammonium level or better in the stated range; at lower levels of ammonium 20 % is sufficient. Using the Nordtest approach (see section 2.4), the laboratory has previously estimated the measurement uncertainty in the required concentration range in fresh water, finding a standard uncertainty of 3.2 % and corresponding expanded uncertainty ( $k = 2$ ) of 6.4 %. This is smaller than the reproducibility standard deviation  $s_R$  given in the ISO method, which equates to a reproducibility standard deviation of 10 % relative at 200 µg/l; the laboratory therefore wishes to verify the smaller estimated uncertainty using independent evidence.

(a) Checks using observed within-laboratory reproducibility,  $s_{RW}$

The  $s_{RW}$  for a control sample covering the whole analytical process is 1.5 %. This is close to a factor of 2 smaller than the standard uncertainty, a common situation.

(b) Checks based on the results of PT

The laboratory has participated in several PT rounds over recent years. The PT test materials are for seawater, a somewhat different sample type, and the concentrations generally lower. The results for samples within or near the range of interest are shown in Table 1; all are at the very lower limits of the range of interest. Since it is known that the relative measurement uncertainty increases at lower ammonium levels, this PT data forms a 'worst case' check. The four results show a root-mean-square relative error of 4.5 %, which at first sight is larger than the expected 3.2 %. However, a chi-squared test returns a  $p$ -value of 0.09, which is not sufficient to rule out consistency with a 3.2 % uncertainty. Further, the RMS error is at least partly due to a modest positive bias of about 4 %; this can be accounted for by the high sodium concentration in seawater, which is known to cause positive bias for the method. Fresh water can therefore be expected to perform significantly better.

PT No.	Assigned value $x_{ref}$	Lab. result $x_i$	Lab. bias $(x_i - x_{ref}) / x_{ref}$
	µg / l	µg / l	%
1	61	63	2.5
2	57	62	7.3
3	102	105	3.2
4	84	87	3.5

Table 1 – PT results for ammonium (µg N/l) in seawater  
(source: Haarvard Hovind, NIVA Norway, personal communication, Feb 2006)

## **2 Conclusions**

The overall indication is accordingly that there is no compelling reason to alter the standard uncertainty of 3.2 % for the normal range of interest in fresh waters. A prudent laboratory manager would, however, review the typical range of ammonium levels for their routine samples; if many were at the low end of the range, or had high sodium content, it would be sensible to gather additional evidence and perhaps increase the declared uncertainty for the lower part of the concentration range until stronger confirmation for fresh water became available.

## **3 References**

ISO 11732:2005, *Water quality – Determination of ammonium nitrogen – Method by flow analysis (CFA and FIA) and spectrometric determination*

## EXAMPLE 6: MEASUREMENT UNCERTAINTY IN OPTICAL EMISSION SPECTROMETRY

Sector	Measurand / matrix	Technique	Approaches for uncertainty evaluation	
Analysis of metals	Zinc content / metal alloys	Optical emission spectrometry	Single laboratory validation	Interlaboratory data

### 1 Objectives of the study

We explain some pragmatic approaches for estimating uncertainty in the analysis of metals by OES. The aims are to:

- Define some practical procedures for MU in analysis of metals by OES.
- Set up the main contributions for uncertainty.
- Reduce time and costs.
- Confirm uncertainty estimates by interlaboratory data.

### 2 Procedures for uncertainty estimation

The procedures described in this example include:

- The single laboratory validation approach, based on verification of traceability using CRM (certified reference materials) [1].
- Estimating uncertainty from interlaboratory comparisons or PT data.

#### 2.1 Single laboratory validation approach, based on verification of traceability using Certified Reference Materials (CRMs)

The aim of this procedure is to verify traceability (i.e. absence of significant bias) using CRMs and calculate the uncertainty of the analytical procedure with these data. This would be valid for samples with concentration values and matrix similar to the CRM and with the analytical procedure under statistical control.

The analysis should be carried out under intermediate precision conditions (different days, analysts, etc.). The uncertainty calculation follows the procedure given in [1].

##### 2.1.1 Principles

The approach involves the following steps, based on validation practice.

##### Step 1: Verification of traceability

For verifying traceability (i.e. absence of significant bias) we compare the observed mean value  $\bar{x}$  with the CRM certified value  $x_{CRM}$  using a Student  $t$ -test. Given a narrow measuring range, we assume that the (absolute) bias is constant. The relevant equation is:

$$(1) \quad t_{cal} = \frac{|x_{CRM} - \bar{x}|}{\sqrt{u_{CRM}^2 + \frac{s^2}{n}}}$$

where:

$u_{CRM}$  = CRM standard uncertainty ( $U_{CRM} / k$ ),

$s$  = Standard deviation of test results.

If no significant bias is found we continue with step (2).

**Step 2: Uncertainty of verification of traceability**

If the bias is not significant, the analytical method is traceable, but we nonetheless introduce a component of uncertainty associated with the process used to verify the traceability. The equation is:

$$(2) \quad u_{trac} = \sqrt{u_{CRM}^2 + \frac{s^2}{n}}$$

**Step 3: Uncertainty of the analytical procedure**

As before, this component is estimated from the standard deviation (random errors) of the analytical procedure, and it is assumed that samples will be similar in concentration to the CRM and the method is under effective quality control.

The standard uncertainty is calculated as follows:

$$(3) \quad u_{proc} = s$$

**Step 4: Calculating the combined standard uncertainty**

The combined standard uncertainty  $u_{COM}$  is calculated as follows:

$$(4) \quad u_{COM} = \sqrt{u_{trac}^2 + u_{proc}^2} = \sqrt{\left(u_{CRM}^2 + \frac{s^2}{n}\right) + s^2}$$

**2.1.2 Numerical Example**

During routine analysis of zinc in metal alloy samples (by optical emission spectrometry), a certified reference material was regularly analysed (20 times) over a period of two months. The certified zinc concentration in the reference material is assumed to be representative for the narrow working range of the method:  $20 \pm 1$  % mass fraction of zinc.

All data below are mass fractions of zinc in per cent.

**Step 1: Verification of traceability**

The CRM data are as follows:

- Certified concentration of zinc in the reference material,  $x_{CRM} = 20.225$  %
- Standard uncertainty of the certified zinc concentration,  $u_{CRM} = 0.27$  %

From the results of replicate analyses of the CRM the following values were obtained:

- Mean of replicate analyses of the CRM,  $\bar{x} = 20.253$  %
- Standard deviation of the results from the replicate analyses of the CRM,  $s = 0.1355$  %

For verifying the traceability of the results of the analytical procedure equation (1) is used:

$$t_{cal} = \frac{|20.225 - 20.53|}{\sqrt{(0.27)^2 + \frac{(0.1355)^2}{20}}} = 0.10$$

The tabulated value of the  $t$ -distribution for a significance level of 95 % and 19 degrees of freedom is  $t_{tab}(19; 95\%) = 2.093$ . Thus we have:

$$t_{cal} < t_{tab} \rightarrow \text{The analytical procedure is traceable,} \\ \text{i.e. there is no indication of significant analytical bias.}$$

**Step 2: Uncertainty of verification of traceability**

For calculating the uncertainty associated with the verification of traceability, equation (2) is used:

$$u_{\text{trac}} = \sqrt{(0.27)^2 + \frac{(0.1355)^2}{20}} = 0.27$$

**Step 3: Uncertainty of the analytical procedure**

To calculate the standard uncertainty of the analytical procedure, equation (3) is used:

$$u_{\text{proc}} = 0.1355$$

**Step 4: Calculating the combined standard uncertainty**

The combined standard uncertainty is calculated using equation (4):

$$u_{\text{COM}} = \sqrt{\left( (0.27)^2 + \frac{(0.1355)^2}{20} \right) + (0.1355)^2} = 0.30$$

**Result: The combined standard uncertainty is 0.30 % (Zn)**

## 2.2 Estimation of measurement uncertainty from interlaboratory reproducibility information derived from Proficiency Testing Scheme data

In our example we have data from several interlaboratory comparisons for the element of interest (Zn). There are two options:

- Using the laboratory's own deviations from the assigned values as described in the Nordtest approach (Section 2.4 of the main text)
- Using the reproducibility observed in the PT rounds as an estimate of interlaboratory reproducibility, which is then applied in a manner analogous to the use of collaborative study data described in ISO TS 21748.

In this case, the second of these is used. This involves the following steps: 1) obtaining estimates of the repeatability and reproducibility standard deviations; 2) verifying that the data are applicable to the particular laboratory; 3) checking and adjusting for differences in test item type; 4) estimating the uncertainty, taking account of any additional effects.

Note: All data below are mass fractions of zinc in per cent.

**Step 1: Estimating the reproducibility standard deviation**

The values of  $s_R$  obtained for the last interlaboratory comparisons (9-10 participating laboratories during 9 rounds) were:

$s_R$	0.2683	0.2572	0.4879	0.3745	0.3387	0.2842	0.2511	0.2034	0.1897
-------	--------	--------	--------	--------	--------	--------	--------	--------	--------

The mean value of  $s_R$  (calculated as the root of the mean squares) was: 0.308.

The data below are the individual standard deviations of the laboratories participating in the last interlaboratory comparison.

	Lab1	Lab2	Lab3	Lab4	Lab5	Lab6	Lab7	Lab8	Lab9
$s_r$	0.0650	0.0679	0.1824	0.0755	0.0802	0.0911	0.2321	0.0687	0.0712

The average within-laboratory standard deviation (using the root mean square) was 0.120, which can be used as an estimate of within-laboratory repeatability similar to the repeatability standard deviation obtained in a collaborative study according to ISO 5725.

**Step 2: Verifying that the data are applicable to the particular laboratory**

ISO TS 21748 requires that the laboratory demonstrate that its bias and repeatability are consistent with the population of laboratories used to estimate the reproducibility. Here, this can be demonstrated in two ways. First, the in-house validation data show an insignificant bias. Second, the laboratory's PT results are consistently acceptable over the rounds in question, providing good evidence of control of bias and precision (the dispersion of successive PT results includes both the laboratory bias and repeatability). Finally, the laboratory has participated in the PT and its within-laboratory precision is close to the estimated repeatability standard deviation above.

**Step 3: Establishing relevance to the material**

The PT data are obtained on realistic test materials closely similar to the laboratory's routine sample type and zinc concentration; there is accordingly no need to adjust for concentration or other differences in test material type.

**Step 4: Estimation of the standard uncertainty**

The root mean square of the  $s_R$  values provides a good estimate of the typical performance of laboratories undertaking this type of analysis. No other effects are considered significant; for example, the PT materials are in the same form as routine samples, so there is no need to adjust for sample preparation differences. Following the principles of ISO TS 21748, the estimated  $s_R$  value can therefore be used directly as an estimate of the standard uncertainty:

$$u_{INT} = \text{root mean square } (s_R) = 0.308$$

**Result: The standard uncertainty is 0.31 % (Zn)**

**3 Summary**

<b>APPROACH</b>	<b>STANDARD UNCERTAINTY [mass fraction of Zn in %]</b>
<b>Single laboratory validation</b>	<b>0.30</b>
<b>Interlaboratory comparison</b>	<b>0.31</b>

**4 Conclusions**

The two approaches provide similar values of measurement uncertainty. Since the interlaboratory data are likely to cover most, if not all, the major sources of uncertainty for this type of measurement, this provides a confirmation of the results obtained using the single-laboratory approach based on a single CRM.

The single laboratory approach uses information generated in the process of assessing the performance of a given analytical procedure. Uncertainty evaluation with this approach has the advantage that the effort made for checking accuracy can be used to calculate the uncertainty of future routine measurements. Nevertheless, it should be taken into account that the single laboratory validation approach has potential drawbacks, due to limitations in the validation of the accuracy of the measurement procedure using a single material. Therefore confirmation by or combination with interlaboratory data, if available, is beneficial.

**5 References**

- [1] Maroto, A., Boqué, R., Riu, J, F.X. Rius (1999), *Evaluating uncertainty in routine analysis*, Trends Anal Chem **18**, (9-10), 577-584.

**EXAMPLE 7: ROCKWELL HARDNESS TESTING**

Sector	Measurand / Matrix	Technique	Approaches for	
			estimation	verification
Materials testing	Rockwell hardness / Metal specimens	Measurement of indentation depth	Single laboratory validation data	PT performance

**1 Hardness testing – general principles**

Hardness tests are performed according to different conventional methods. Thus the results are not expressed in SI units but in terms of specific hardness scales depending on the method applied.

Several methods are standardised, e.g. Brinell [1], Vickers [2] and Rockwell [3], each of them defining its own hardness scale. There are no general procedures for relating the results from one scale to another.

In each case a specifically shaped indenter is pressed into the material under test with a defined force for a defined dwell time. The dimensions of the indent produced in the material (depth, diameter or diagonals) are measured.

A hardness scale is defined, e.g. in International Standards [1-3], by the description of the method, including the specification of relevant tolerances of the quantities involved and the limiting ambient conditions. The scale is realised at the national level by primary standard machines which are used for the calibration of primary reference hardness blocks. These primary reference hardness blocks are transfer standards which are used to calibrate the hardness calibration machines owned by calibration laboratories. By means of such hardness calibration machines the calibration laboratories themselves produce hardness reference blocks [5] which are used as secondary transfer standards to calibrate the hardness testing machines at the user level.

In addition to the so called indirect calibration based on the (primary) reference hardness blocks as described above, direct calibration of the hardness testing machines is performed involving e.g.

- force,
- shape of indenter,
- indentation measurement system,
- test cycle.

While the indirect calibration is performed before a measurement series is started, the direct calibration is only needed at longer intervals, such as annually.

The measurement uncertainty of hardness testing can in principle be evaluated from the direct calibration, as described in reference [4]. However, this example is focussed on measurement evaluation for the indirect calibration.

**2 Rockwell hardness testing (scale C)**

According to [3] for Rockwell hardness testing (scale C) a conical diamond indenter with specified dimensions [6] is pressed into the test sample in two steps. In the first step a preliminary test force  $F_0 = 98.07$  N is applied for less than 3 s. Within a period of 1 s to 8 s the force is increased by an additional force  $F_1 = 1373$  N, resulting in a total force  $F = 1471$  N, which is applied for  $4 \pm 2$  s. Then  $F_1$  is released and with the remaining preliminary test force  $F_0$  the permanent indentation depth  $h$  is measured. The Rockwell hardness of scale C is defined as:

$$(1) \quad \text{Rockwell hardness} = 100 - \frac{h}{0.002}$$

For example a Rockwell hardness of 60 HRC would result from an indentation depth of  $h = 0.08$  mm.

## 2.1 Measurement uncertainty evaluations derived from indirect calibration

In appendix G of ref. [3] two methods are given for the evaluation of measurement uncertainty based on indirect calibration, i.e. on the use of a certified reference hardness block. In both cases 5 measurements are performed on the reference hardness block and 5 measurements on the test sample.

### Procedure without bias (method 1)

The expanded measurement uncertainty  $U$  is given by:

$$(2) \quad U = k \cdot \sqrt{u_E^2 + u_{CRM}^2 + u_{\bar{H}}^2 + u_{\bar{x}}^2 + u_{ms}^2}$$

with:

- $u_E$  uncertainty from maximum tolerable bias according to [6],
- $u_{CRM}$  uncertainty of the certified value,
- $u_{\bar{H}}$  uncertainty from the 5 measurements of the reference block,
- $u_{\bar{x}}$  uncertainty from the 5 measurements of the sample,
- $u_{ms}$  uncertainty from the resolution of the measuring system.

The result of the measurement can be reported as:

$$(3) \quad \bar{X} = \bar{x} \pm U$$

### Procedure with bias (method 2)

The second method is applicable e.g. for laboratories that use control charts. At least two series of measurements ( $n = 5$ ) are needed of the reference hardness block. From these two series a mean bias (eq. 5) and its standard uncertainty  $u_b$  can be derived.

$$(4) \quad b_i = \bar{H}_i - \bar{X}_{CRM,i}$$

$$(5) \quad \bar{b} = 0.5 \cdot (b_1 + b_2)$$

For this method the expanded uncertainty is given by:

$$(6) \quad U_{corr} = k \cdot \sqrt{u_b^2 + u_{CRM}^2 + u_{\bar{H}}^2 + u_{\bar{x}}^2 + u_{ms}^2}$$

The result of the measurement can either be corrected for the bias<sup>2</sup>:

$$(7) \quad \bar{X}_{corr} = (\bar{x} - \bar{b}) \pm U_{corr}$$

or the bias can be added to the expanded uncertainty:

$$(8) \quad \bar{X}_{ucorr} = \bar{x} \pm (U_{corr} + |\bar{b}|)$$

NOTE - It seems that the equations 2 and 6 both overestimate the expanded uncertainties. Since the term  $(u_{CRM}^2 + u_{\bar{H}}^2)$  reflects the uncertainty of the measured difference between the mean value measured from the reference hardness block and the certified value, the inclusion of the additional terms  $u_E$  or  $u_b$  respectively causes some double counting of the same effect. But on the other hand  $u_b$  in eq. 6 introduces an element of intermediate conditions, while the other precision terms are derived under repeatability conditions.

<sup>2</sup> In the corresponding equation in [3] the bias is not subtracted, but erroneously added.



























