Manual of Codes of Practice for the Determination of Uncertainties in Mechanical Tests on Metallic Materials

SECTION 1

Introduction to the evaluation of uncertainty

F A Kandil

National Physical Laboratory
Queens Road
Teddington, Middlesex TW11 0LW
UNITED KINGDOM

Issue 1

September 2000

CONTENTS

1.1 Introduction

This Section outlines the principles and gives guidance for the estimation of measurement uncertainty that were used in developing this "*UNCERT Manual of Codes of Practice for the Determination of Uncertainties in Mechanical Tests on Metallic Materials*", hereafter referred to as *The Manual*. These principles and guidelines are based on the recommendations of the "*Guide to the Expression of Uncertainty in Measurement",* hereafter referred to as The GUM [Ref. 1]. For those who are new to this field, a good starting point is Ref. [2], "*A Beginner's Guide* , a copy of which is included in this publication. Other suggested readings, giving more detail of the principles and practice of estimating uncertainties in testing are given in Ref. 3 - 6. Lists of symbols and definitions of terms are given in Section 2 of *The Manual*.

1.2 General principles

1.2.1 The objective of a measurement is to determine the value of the **measurand**, i.e. the specific quantity subject to measurement. A measurement begins with an appropriate specification of the measurand, the generic method of measurement and the specific detailed **measurement procedure**.

1.2.2 In general, no measurement or test is perfect and the imperfections give rise to an **error of measurement** in the result. Consequently, the result of a measurement is only an approximation to the value of the measurand and is only complete when accompanied by a statement of the **uncertainty** of that approximation. Indeed, because of measurement uncertainty, a 'true value' can never be known [4].

1.2.3 It is essential to distinguish the term 'error' (in a measurement result) from the term 'uncertainty'. Error is the measurement result minus the **true** value of the measurand. Whenever possible, a correction equal and of opposite to an error is applied to the result. Because the true values are never known exactly, corrections are always approximate and a residual error will remain. The uncertainty in this residual error will contribute to the uncertainty of the reported result [5].

1.2.4 Errors of measurement may have two components, a random component and a systematic component. Uncertainties arise from random effects and from incomplete correction for systematic effects [1].

1.2.5 **Random errors** arise from random effects. Each time a measurement is taken under identical conditions, random effects from various sources affect the value of the measurand. A series of measurements produce a scatter around a mean value. A number of sources may contribute to the variability each time a measurement is taken, and their influence may be continually changing. Random errors cannot be eliminated but the uncertainty due to their effect may be reduced by increasing the number of observations and applying statistical analysis [4].

1.2.6 **Systematic errors** arise from systematic effects, i.e. an effect on a measurement result of a quantity that is not included in the specification of the measurand but which influences the result. These remain unchanged when a measurement is repeated under identical conditions, and their effect is to introduce an offset between the value of the measurand and the experimentally determined mean value. Systematic errors cannot be eliminated but may be reduced, e.g. a correction may be made for the known extent of an error due to a recognised systematic effect [4].

1.2.7 Rather than using the concept of random and systematic errors, the GUM has adopted the approach of grouping uncertainty components into two categories based on how they are evaluated, '*Type A*' and '*Type B*', where:

> '***Type A***' evaluation is by calculation from a series of repeated observations, using statistical methods [1], and
>
> '***Type B***' evaluation is by means other than those used for a *Type A* evaluation. For example, by using data from: calibration certificates, manufacturers' specifications, previous measurement data, experience with the behaviour of the instruments, and all other relevant information [1].

1.2.8 Whether the components of uncertainty are classified as 'random' or 'systematic' in relation to a specific measurement, or described as ' *Type B*', they are modelled by probability distributions and quantified by the variance or standard deviation. Therefore any convention as to how they are classified does not affect the estimation of total uncertainty. However, it should always be remembered that, in this document, when the terms 'random' and 'systematic' are used they refer to the effects of the uncertainty on a specific measurement process. It is the usual case that random components require *Type A* evaluations and systematic components require *Type B* evaluations, but there are some exceptions [5].

1.2.9 Component uncertainties are evaluated by the appropriate method and each is expressed as a **standard deviation** and is referred to as a **standard uncertainty**.

1.2.10 The component standard uncertainties are combined to produce an overall value of uncertainty, known as the **combined standard uncertainty**.

1.2.11 **An expanded uncertainty** is usually required to meet the needs of the industrial, commercial, health and safety, or other requirements/regulations. The expanded uncertainty provides a greater interval about the result of a measurement than the standard uncertainty with, a subsequent higher probability that it encompasses the value of the measurand. The expanded uncertainty is obtained by multiplying the combined standard uncertainty by a **coverage factor**, *k*. The choice of factor is based on the coverage probability or **confidence level** required.

1.3. Reasons for estimating uncertainty

1.3.1 There are requirements by accreditation authorities for test laboratories to evaluate and report the uncertainty associated with their test results.

1.3.2 Customers may also demand to know the limits within which the reported result may be reasonably assumed to lie.

1.3.3 The laboratory itself may want to gain a better understanding of which aspects of the test procedure have the greatest effect on results so that this may be monitored more closely or improved.

1.3.4 The uncertainty of the result of a test needs to be taken into account when interpreting data. For example, a comparison of results from different batches of material will not indicate real differences in the properties or performance if the observed differences fall within the range of the inherent variation associated with the test procedure [4].

1.3.5 In some cases the uncertainty in a measurement or test result may be considered to be so small as to be not worth formal evaluation. However, without a formal estimate, this consideration remains intuitive and, when challenged, a convincing response is not possible [4].

1.3.5 The results of some types of test are subject to large uncertainty, for example, where tests are carried out on samples that are themselves inconsistent in their properties. In such a case it may be asserted that even a relatively large uncertainty in the measurement may be ignored compared with the uncertainty associated with the sample variation. Once again, however, unless an estimate of the uncertainty of the measurement itself is made, the validity of this assertion cannot be supported [4].

1.4 Sources of uncertainty

Uncertainties can come from:

- The test instrument
- The item being tested
- The test procedure
- The test environment
- Operator skill
- Sampling issues

Possible sources of uncertainties include [1]:

(a) Incomplete definition of the measurand or the test procedure. The requirement is not clearly described, e.g. the temperature of a test may be given as ‘room

(b) Imperfect realisation of the test procedure; even when the test conditions are clearly defined it may not be possible to produce the required conditions.

(c) Non representative sampling (for quantifying repeatability and effects attributed to material variability.)

(d) Inadequate knowledge of the effects of environmental conditions on the measurement process, or imperfect measurement of environmental conditions.

(e) Personal bias in reading analogue instruments, judging colour, or reacting in time.

(f) Instrument resolution or discrimination threshold or errors in the graduation of a scale.

(g) Values assigned to measurement standards (both reference and working) and reference materials.

(h) Changes in the characteristics or performance of a measuring instrument since the last calibration.

(i) Values of constants and other parameters used in data evaluation.

(j) Approximations and assumptions incorporated in the measurement method and procedure.

(k) Variations in repeated observations made under apparently identical conditions - such random effects may be caused by, for example: (i) short-term fluctuations in local environment, e.g. temperature, humidity and air pressure and/or (ii) variability in the performance of the person carrying out the test.

The above sources are not necessarily independent. Unrecognised systematic effects may also exist which cannot be taken into account but still contribute to the error. (The existence of such effects may sometimes be deduced, for example, from the results of crosschecking exercises such as inter-laboratory comparisons or proficiency testing [5].)

1.5 Estimation of uncertainties

1.5.1 General approach

The overall uncertainty of a measurement is a combination of a number of uncertainties. Careful consideration of each measurement involved in the test is required to *identify* and list all the factors that contribute to the overall uncertainty. This is a very important step and requires a good understanding of the measurement equipment, the principles and practice of the test and the influence of environment.

The next step is to *quantify* these uncertainties. An initial approximate quantification may be valuable in identifying if some components are negligible and not worthy of more rigorous evaluation. In most cases, a practical definition of negligible would be a component that is not more than a fifth of the magnitude of the largest component [5]. Some uncertainties may be quantified by calculating the standard deviation from a set of repeated measurements (*Type A*). Others will require some judgement by the operator, using all relevant information on the possible variability of each factor (*Type B*).

Figure 1 summarises the main steps for estimating uncertainty.

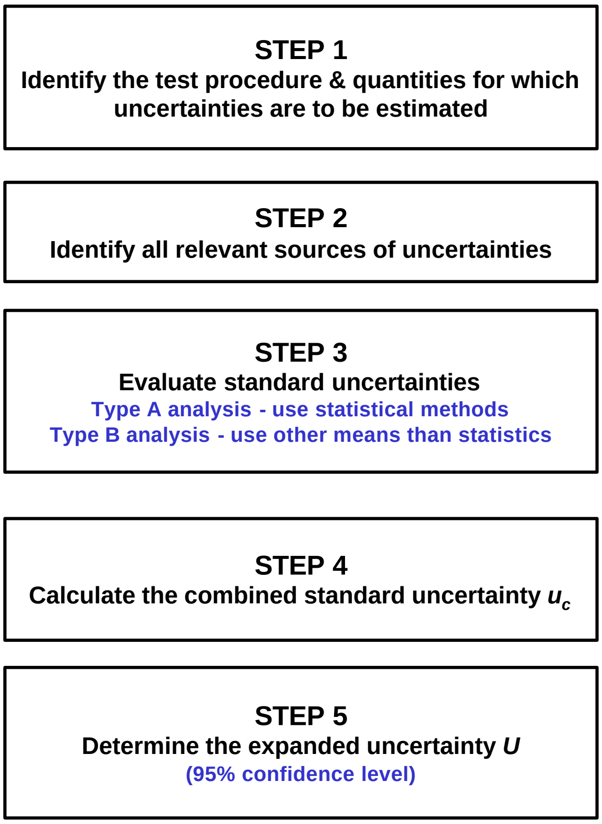

Fig.1 Steps of estimating a measurement or testing uncertainty

In general a measurement process can be regarded as having estimated input quantities, given the symbol x, which contribute to the value of the measurand or output quantity, y. Where there are several input quantities they can be represented by x_i, and the standard uncertainty associated with the estimated value of each input quantity is represented by $u(x_i)$.

The measurement process can usually be modelled by a functional relationship between the estimated input quantities and the output in the form:

$$y = f\,(x_1, x_2, \ldots\ldots x_m) \tag{1}$$

For example, if Young's modulus E is measured in terms of the stress s and the strain e then the relationship is $E = f\,(s,e) = s/e$. The mathematical model is used to identify the input quantities that need to be considered in the uncertainty budget and their relationship to the total uncertainty for the measurement. In some cases the input quantities are not in the same units as the output quantity and each input uncertainty will need to be multiplied by an appropriate factor c_i before it is combined with other uncertainties.

Subsequent calculations will be made clearer if, wherever possible, all components are expressed in the same way, e.g. either in the same units as used for the reported result or in relative terms (i.e. in percent.)

1.5.2 *Type A evaluation of standard uncertainty*

For a series of n repeated readings, the estimated standard uncertainty, u, of the arithmetic mean $\bar{x}$ is calculated from:

$$u = \frac{s}{\sqrt{n}} \tag{2}$$

where s is the estimated standard deviation:

$$s = \sqrt{\frac{1}{(n-1)} \sum_{i=1}^{n} (x_i - \bar{x})^2} \tag{3}$$

It may not always be practical to repeat a measurement many times. In these cases a more reliable estimate of the standard deviation of the measurement can be obtained from an earlier *Type A* evaluation, based on a larger number of readings [1, 5].

Although no correction can be made for a random component of uncertainty, Eq. (2) shows the benefit of increasing the number of measurements. However, the benefit becomes progressively less as the number is increased, and it is usually not necessary to make more than about 10 measurements. Often 5 measurements are sufficient, provided that the required level of confidence is maintained.

1.5.3 Type B evaluation of standard uncertainty

The standard uncertainty of an input quantity that has not been obtained from repeated measurements should be evaluated by scientific judgment based on all of the available information on the possible contributing factors. The information may include:

- Data provided in calibration and other certificates.
- Manufacturer's specification.
- Previous measurement data.
- Experience with or general knowledge of the behaviour of the relevant materials and instruments.
- Uncertainties assigned to reference materials.
- Uncertainties assigned to reference data taken from handbooks.

For most *Type B* evaluations, one might only be able to: (i) estimate the upper and lower limits of uncertainty and (ii) assume a rectangular probability distribution (i.e. the value is equally likely to fall anywhere in between the upper and lower limits). The standard uncertainty for a rectangular distribution is:

$$u = \frac{a}{\sqrt{3}} \tag{4}$$

where a is the mid point value between the upper and lower limits. Rectangular distributions are quite common but other distributions can occur. For example, the uncertainty often stated on an instruments calibration certificate is usually a normal distribution. In this case, the standard uncertainty is:

$$u = \frac{\text{expanded uncertainty}\, U}{k} \tag{5}$$

where k is the covering factor.

1.5.4 Combined standard uncertainty u_c

Once the standard uncertainties $u(x_i)$ of the input quantities x_i have been derived from both *Type A* and *Type B* evaluations, the standard uncertainty of the output quantity $y = f(x_i, x_2, \ldots, x_m)$, also called the combined standard uncertainty, can be calculated using the root sum squares as follows:

$$u_c(y) = \sqrt{\sum_{i=1}^{m} [c_i u(x_i)]^2} \tag{6}$$

where c_i is the partial derivative, $\frac{\partial f}{\partial x_i}$, or in some cases a known sensitivity coefficient associated with the input quantity x_i. A typical evaluation of the combined standard uncertainty is as follows:

$$u_c(y) = \sqrt{\frac{[c_1 U_1]^2}{k} + \frac{c_2^2 a_2^2 + c_3^2 a_3^2}{3} + c_4^2 u_4{}^2(x_4)} \tag{7}$$

where U_1 has a normal probability distribution, a_2 and a_3 are limits with rectangular probability distributions, all obtained from a *Type B* evaluation, and $u(x_4)$ is obtained from a *Type A* evaluation.

The calculations necessary to obtain the sensitivity coefficients by partial differentiation can be a lengthy process, particularly when there are many individual contributions and the uncertainty estimates are needed for a range of values. If the functional relationship for a particular measurement is not known, the sensitivity coefficients may be obtained experimentally.

All uncertainty contributions must be in the same units before they are combined. In many cases, however, the input quantity may not be in the same units as the output quantity. For example, in high temperature fatigue testing, one contribution to fatigue life is the test temperature. In this case the input quantity is temperature, but the output quantity is the number of cycles to failure, N_f, which is dimensionless. In such cases, a sensitivity coefficient, c_T (corresponding to the partial derivative of the N_f / test temperature relationship), is used to convert from temperature to the number of cycles to failure (see example in CoP02 in *The Manual*).

The combined uncertainty corresponds to plus or minus one standard deviation (i.e. has an associated confidence level of 68.27%.)

1.5.5 Expanded uncertainty U and level of confidence

The expanded uncertainty U is defined in the GUM [1] as "the interval about the result of a measurement that may be expected to encompass a large fraction of the distribution of values that could **reasonably** be attributed to the measurand".

It is obtained by multiplying the combined uncertainty u_c calculated in 1.5.4, by a coverage factor, k, which is selected on the basis of the level of confidence required. For a normal probability distribution, a coverage factor of 2 is most commonly used and this corresponds to a confidence interval of 95.45% (effectively 95% for most practical purposes). Where a higher confidence level is demanded by the customer, (such as for particular measurements in the aerospace and electronics industries), a coverage factor of 3 or more is sometimes used. Table 1 below gives commonly used levels of confidence and their associated coverage factors for a normal distribution.

Table 1 Value of the coverage factor k_p that produces an interval having level of confidence p (assuming a normal distribution).

Level of confidence p (percent)	Coverage factor k_p
68.27	1
90	1.645
95	1.960
95.45	2
99	2.576
99.73	3

In cases where the probability distribution of u_c is not normal or where the number of data points used in a *Type A* analysis is small, a coverage factor k_p should be determined according the degrees of freedom given by the Welsh-Satterthwaite method (see Annex G in the GUM [1] for details).

References

1. BIPM, IEC, IFCC, ISO, IUPAC, OIML, “*Guide to the expression of uncertainty in measurement*”. International Organisation for Standardisation, Geneva, Switzerland, ISBN 92-67-10188-9, First Edition, 1993. (This *Guide* is often referred to as the GUM or the ISO TAG4 document after the ISO Technical Advisory Group that produced it.)

   Identical documents:
   - ENV 13005:1999 (English)
   - NF ENV 13005:1999 (French)
   - NEN NVN ENV 13005:1999 (Dutch)
   - “*Vocabulary of metrology, Part 3. Guide to the expression of uncertainty in measurement*”, PD 6461: Part 3: 1995, British Standards Institution.

2. S Bell, “*A beginner’s guide to uncertainty of Measurement*”, Measurement Good Practice Guide No.11, August 1999. National Physical Laboratory, Teddington, Middlesex, TW11 0LW, England.

3. VIM, ISO, “*International Vocabulary of Basic and General Terms in Metrology*”, Second Edition, 1993. ISBN 92-67-01075-1. International Organisation for Standardization, Geneva, Switzerland.

4. NIS 80, “*Guide to the expression of uncertainties in testing*”, Edition 1, September 1994. NAMAS Executive, National Physical Laboratory, Teddington, Middlesex, TW11 0LW, England.

5. M 3003, “*The expression of uncertainty and confidence in measurement for calibrations*”, Edition 1, December 1997. United Kingdom Accreditation Service, UKAS, 21-47 High Street, Feltham, Middlesex TW11 4UN, England.

6. ISO 5725, “*Accuracy (trueness and precision) of measurement methods and results*”
   Part 1:1994 “*General principles and definitions*”
   Part 2:1994 “*Basic method for the determination of repeatability and reproducibility of standard measurement method*
   Part 3:1994, “*Intermediate measures of the precision of a standard measurement method*”
   Part 4:1994, “*Basic methods for the determination of the trueness of a standard*

   Part 5:1998, “*Alternative methods for the determination of the precision of a standard measurement method*”
   Part 6:1994, “*Use in practice of accuracy values*”