The Expression of Uncertainty in Testing

[UKAS Publication ref: LAB 12]

Contents

About the United Kingdom Accreditation Service

The United Kingdom Accreditation Service (UKAS) is recognised by the UK Government as the national body responsible for assessing and accrediting the competence of organisations in the fields of calibration, testing, inspection and certification of systems, products and personnel.

1 Introduction

1.1 The general requirements for the estimation and reporting of uncertainty in accredited laboratories are given in ISO/IEC 17025. General guidance on how they may be met when estimating and reporting uncertainty in testing is given in this publication. UKAS Publications, M4 lists some other publications giving guidance on evaluation of uncertainty.

1.2 The general approach to evaluating and expressing uncertainty in testing outlined here is based on the recommendations produced by the International Committee for Weights and Measures (CIPM), as described in the *Guide to the Expression of Uncertainty in Measurement*, 1993, ISO Geneva. (*The Guide*).

1.3 In the interests of keeping to a general, descriptive format, exceptions, special cases and qualifying remarks have not been dealt with in detail and equations have not been included. *The Guide* should be consulted for equations and symbols, and when needing to resolve special difficulties that may arise in specific tests.

2 Definitions

2.1 Definitions of terms are given in alphabetical order in Appendix A. Defined terms are printed in **bold** at the first appropriate occurrence in the main body of the text.

3 Reporting and evaluation of uncertainty

3.1 The requirements of ISO/IEC 17025 distinguish between the need for reporting and the need for evaluation of uncertainty of measurement.

3.2 Reporting is required when information on uncertainty is relevant to the validity or application of the test results, when the client requires it or when the uncertainty affects compliance with a specification limit.

3.3 Evaluation of uncertainty of measurement is required for calibrations, including those performed in house, and procedures for estimating uncertainty of measurement of tests are needed and need to be applied.

3.4 The complexity of tests may in some cases preclude a rigorous evaluation of uncertainty. In such cases, at least a list of the potential contributors to uncertainty should be made and should include reasonable estimates of the magnitude of each component uncertainty. These estimates may be based on previous experience and make use of data from method validation and other sources, such as quality control data, proficiency test data and round robin test results. It is recommended that an example be produced in which the overall uncertainty of the result for that test method is calculated.

3.5 In cases where a well-recognised test method specifies limits to the values of major sources of uncertainty of measurement and specifies the form of presentation of the results the requirement to estimate uncertainty of measurement can be considered to have been satisfied by following the test method and its reporting instructions.

4 Reasons for evaluating uncertainty

4.1 The expression of the uncertainty of a result allows realistic comparison of results from different laboratories, or within a laboratory or with reference values given in specifications or standards. This information can often prevent unnecessary repetition of tests.

4.2 The uncertainty of the result of a test may need to be taken into account by a customer when interpreting data. For example, comparison of results from different batches of material will not indicate real differences in properties or performance if the observed differences could simply be accounted for by the uncertainty of the results.

4.3 An evaluation [or at least a full consideration] of the components, including random effects from human operators, that contribute to the overall uncertainty of a measurement or test result provides a means of establishing that the test procedure, including the metrological characteristics of the equipment used, will allow valid measurements and results to be obtained.

4.4 A consideration of uncertainty components also indicates aspects of a test to which attention should be directed to improve procedures.

4.5 Systematic assessment of the factors influencing the result and of the uncertainty based on the understanding of the principles of the method and practical experience of its application can be a key part of method validation.

5 General principles

5.1 The objective of a measurement is to determine the value of the **measurand**, i.e. the specific quantity subject to measurement. When applied to testing, the general term measurand may cover many different quantities, e.g. the strength of a material, the concentration of an analyte, the level of emissions of noise or electromagnetic radiation, the quantity of micro-organisms. A measurement begins with an appropriate specification of the measurand, the generic method of measurement and the specific detailed measurement procedure.

5.2 In general, no measurement or test is perfect and the imperfections give rise to **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.

5.3 Errors of measurement may have two components, a random component and a systematic component. Uncertainty arises from random effects and from imperfect correction for systematic effects.

5.4 Random errors arise from random variations of the observations (random effects). Every time a measurement is taken under the same conditions, random effects from various sources affect the measured value. A series of measurements produces a scatter around a mean value. A number of sources may contribute to variability each time a measurement is taken, and their

influence may be continually changing. They cannot be eliminated but increasing the number of observations and applying statistical analysis may reduce the uncertainty due to their effect.

5.5   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 influences the result. These remain unchanged when a measurement is repeated under the same conditions, and their effect is to introduce a displacement between the value of the measurand and the experimentally determined mean value. They 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.

5.6   *The Guide* has adopted the approach of grouping uncertainty components into two categories based on their method of evaluation. 'Type A' evaluation is done by calculation from a series of repeated observations, using statistical methods. Type B' evaluation is done by means other than that used for 'Type A'. For example, by judgement based on data in calibration certificates, previous measurement data, experience with the behaviour of the instruments, manufacturers' specifications and all other relevant information.

5.7   Components of uncertainty are evaluated by the appropriate method and each is expressed as a **standard deviation** and is referred to as a **standard uncertainty**.

5.8   The standard uncertainty components are combined to produce an overall value of uncertainty, known as the **combined standard uncertainty**.

5.9   An **expanded uncertainty** is usually required to meet the needs of industrial, commercial, health and safety, or other applications. It is intended to provide a greater interval about the result of a measurement than the standard uncertainty with, consequently, a higher probability that it encompasses the value of the measurand. It is obtained by multiplying the combined standard uncertainty by a **coverage factor**, *k*. The choice of factor is based on the coverage probability or **level of confidenc**e required (see paragraph 7.4).

6   Sources of uncertainty

6.1   There are many possible sources of uncertainty in testing, including:

(a)   Incomplete definition of the test; the requirement is not clearly described, eg the temperature of a test may be given as 'room temperature';

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

(c)   Sampling - the sample may not be fully representative;

(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;

(f)   Instrument resolution or discrimination threshold, or errors in 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: short-term fluctuations in local environment, eg temperature, humidity and air pressure; variability in the performance of the tester.

6.2   These sources are not necessarily independent and, in addition, unrecognised systematic effects may exist that cannot be taken into account but contribute to error. (The existence of such effects may sometimes be evident from examination of the results of an inter-laboratory comparison programme.)

7   Estimation of uncertainty

7.1   GENERAL APPROACH

7.1.1   This Section describes the basic stages in the estimation of uncertainty. Where available, sector-specific publications may provide more detailed guidance, for example *Quantifying Uncertainty in Analytical Measurement*, Second Edition, 2000, EURACHEM/CITAC.

7.1.2   The total uncertainty of a measurement is a combination of a number of uncertainty components. Even a single instrument reading may be influenced by several factors. 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 measuring equipment, the principles and practice of the test and the influence of environment.

7.1.3   The next step is to *quantify* uncertainty components by appropriate means. An initial approximate quantification may be valuable in enabling some components to be shown to be negligible and not worthy of more rigorous evaluation. In most cases, a practical rule would be that a component is negligible if it is not more than a fifth of the magnitude of the largest component. Some components may be quantified by calculation of the standard deviation from a set of repeated measurements (Type A) as detailed in *The Guide*. Quantification of others will require the exercise of judgement, using all relevant information on the possible variability of each factor (Type B). For 'Type B' estimations, the pool of information may include:

(a)   Previous measurement data;

(b)   Manufacturer's specifications;

(c) Data provided in calibration certificates;

(d) Uncertainty assigned to reference data taken from handbooks;

(e) Experience with or general knowledge of the behaviour and properties of relevant materials and instruments; estimations made under this heading are quite common in many fields of testing.

7.1.4 Subsequent calculations will be made simpler if, wherever possible, all components are expressed in the same way, e.g. either as a proportion (percent, parts per million) or in the same units as used for the reported result.

7.2 *STANDARD UNCERTAINTY*

7.2.1 The standard uncertainty is defined as one standard deviation. The potential for mistakes at a later stage of the evaluation may be minimised by expressing all uncertainty components as one standard deviation. This may require adjustment of some uncertainty values, such as those obtained from calibration certificates and other sources, which often will have been expressed to a higher level of confidence, involving a multiple of the standard deviation.

7.3 *COMBINED STANDARD UNCERTAINTY*

7.3.1 The uncertainty components have to be combined to produce an overall uncertainty using the procedure set out in *The Guide.* In most cases, this reduces to taking the square root of the sum of the squares of the component standard uncertainties (the root sum square method). However, some components may be interdependent and could, for example, cancel each other out or could reinforce each other. In many cases this is easily seen and the interdependent components may be added algebraically to give a net value. However, in more complex cases more rigorous mathematical methods may be required for such 'correlated' components and *The Guide* should be consulted.

7.4 *EXPANDED UNCERTAINTY*

7.4.1 It is usually necessary to quote an expanded uncertainty and the combined standard uncertainty therefore needs to be multiplied by the appropriate coverage factor (*k*). This reflects the level of confidence required and is dictated by the details of the probability distribution characterised by the measurement result and its combined standard uncertainty. However, the extent and reliability of the available information do not always justify the extensive computation required to combine probability distributions. In many cases, an approximation is acceptable, viz. that the probability distribution can be assumed to be normal and that a value of *k*=2 defines an interval having a level of confidence of approximately 95%. For more critical applications, a value of *k*=3 defines an interval having a level of confidence of approximately 99%.

7.4.2 Exceptions to these cases would need to be dealt with on an individual basis by consulting more detailed guidance to determine the appropriate value of *k*. This situation would be characterised by one or more of the following:

(a) A random contribution to uncertainty that is relatively large compared with other contributions and only a small number of repeat readings. In this case there is the possibility that the probability distribution will not be normal in form and a value of *k*=2 will give a level of confidence of less than 95%. [This would not usually arise if the uncertainty assessment involved only one Type A evaluation and the number of readings is greater than 2 and the combined standard uncertainty is more than twice the Type A uncertainty].

(b) The absence of a significant number of uncertainty components having well-behaved probability distributions, such as normal and rectangular;

(c) Domination of the combined value by one component with an unknown probability distribution. There is not a clear-cut definition of a dominant component but a practical guide would be where one component was more than five times greater than any other.

8 Summary of the steps in estimating uncertainty

8.1 The following is a short, simplified summary of the general route to evaluation of uncertainty and is applicable in most circumstances. The identification of sources of uncertainty is the most important part of the process. Quantification of uncertainty in testing normally involves a large element of estimation of Type B uncertainty components. Consequently, suitably experienced personnel who apply their knowledge in a critical manner and base their estimates on quantitative data to the maximum extent should do this.

8.2 The steps involved are as follows:

(a) List all factors that may influence the measured values;

(b) Make a preliminary estimate of the values of the uncertainty components, and eliminate insignificant values;

(c) Estimate the values that are to be attributed to each significant component uncertainty. Express in the same way at the one standard deviation level (see paragraph 7.1.4 and Sub-Section 7.2);

(d) Consider the components and decide which, if any, are interdependent and whether a dominant component exists;

(e) Add any interdependent components algebraically, i.e. take account of whether they act in unison or in opposition and thereby derive a net value (see sub-Section 7.3);

(f) Take the independent components and the value(s) of any derived net components and, in the absence of a dominant component, calculate the square root of the sum of their squares to produce a combined standard uncertainty (see sub-Section 7.3);

(g) Except when only the standard uncertainty (ie one standard deviation) is required, multiply the combined standard uncertainty by a coverage factor *k*, selected on the basis of the level of confidence required, to produce an expanded uncertainty. In the absence of a particular level of confidence being specified in the standard or by the client, the coverage factor should normally be 2, giving a level of confidence of approximately 95% (see sub-Section 7.4).

9 Method of stating results

9.1 GENERAL APPROACH

9.1.1 The extent of the information given when reporting the result of a test and its uncertainty should be related to the requirements of the client, the specification and the intended use of the result. The methods used to calculate the result and its uncertainty should be available either in the report or in the records of the test including:

(a) Sufficient documentation of the steps and calculations in the data analysis to enable a repeat of the calculation if necessary

(b) All corrections and constants used in the analysis, and their sources.

(c) Sufficient documentation to show how the uncertainty is calculated.

9.1.2 When reporting the result and its uncertainty, the use of excessive numbers of digits should be avoided. In most cases the uncertainty need be expressed to no more than two significant figures (although at least one more figure should be used during the stages of estimation and combination of uncertainty components in order to minimise rounding errors).

9.1.3 Unless otherwise specified, the result of the measurement should be reported, together with the expanded uncertainty appropriate to the 95% level of confidence, in the following manner:

Measured value   100.1 (units)
Uncertainty of measurement   ± 0.1 (units)

The reported expanded uncertainty is based on a standard uncertainty multiplied by a coverage factor of k=2, providing a level of confidence of approximately 95%.

9.2 SPECIAL CASES

9.2.1 In exceptional cases, where a particular factor or factors can influence the results, but where the magnitude cannot be either measured or reasonably assessed, the statement will need to include reference to that fact, for example:

The reported expanded uncertainty is based on a standard uncertainty multiplied by a coverage factor of k=2, providing a level of confidence of approximately 95%, but excluding the effects of . . . . . .

9.2.2   Any uncertainty that results from the test sample not being fully representative of the whole should normally be identified separately in the evaluation of uncertainty. However, there may be insufficient information to enable this to be done, in which case this should be stated in the report of uncertainty.

10   Assessment of compliance with specification

10.1   When the client or the specification requires a statement of compliance, there are a number of possible cases where the uncertainty has a bearing on the compliance statement and these are examined below.

10.1.1   The simplest case is where the specification clearly states that the measured result, extended by the uncertainty at a given level of confidence, shall not fall outside a defined limit or limits. In these (rare) cases, assessment of compliance would be straightforward.

10.1.2   More often, the specification requires a compliance statement in the certificate or report but makes no reference to taking into account the effect of uncertainty on the assessment of compliance. In such cases it may be appropriate for the user to make a judgement of compliance, based on whether the result is within the specified limits with no account taken of the uncertainty. This is often referred to as 'shared risk', since the end-user takes some of the risk that the product may not meet the specification. In this case there is an implicit assumption that the magnitude of the uncertainty is acceptable and it is important that the laboratory should be in a position to determine the uncertainty.

10.1.3   In the absence of any specified criteria, eg sector-specific guides, test specifications, client's requirements, or codes of practice, the following approach is recommended:

   (a) If the limits are not breached by the measured result, extended by the expanded uncertainty interval at a level of confidence of 95%, then compliance with the specification can be stated, (Case A, Fig 1 and Case E, Fig 2);

   (b) Where an upper specification limit is exceeded by the result even when it is decreased by half of the expanded uncertainty interval, then non-compliance with the specification can be stated, (Case D, Fig 1);

   (c) If a lower specification limit is breached even when the measured result is extended upwards by half of the expanded uncertainty interval, then non-compliance with the specification can be stated (Case H, Fig 2);

   (d) If the measured value falls sufficiently close to a limit such that half of the expanded uncertainty interval overlaps the limit, it is not possible to confirm compliance or non-compliance at the stated level of confidence. The test result and expanded uncertainty should be reported together
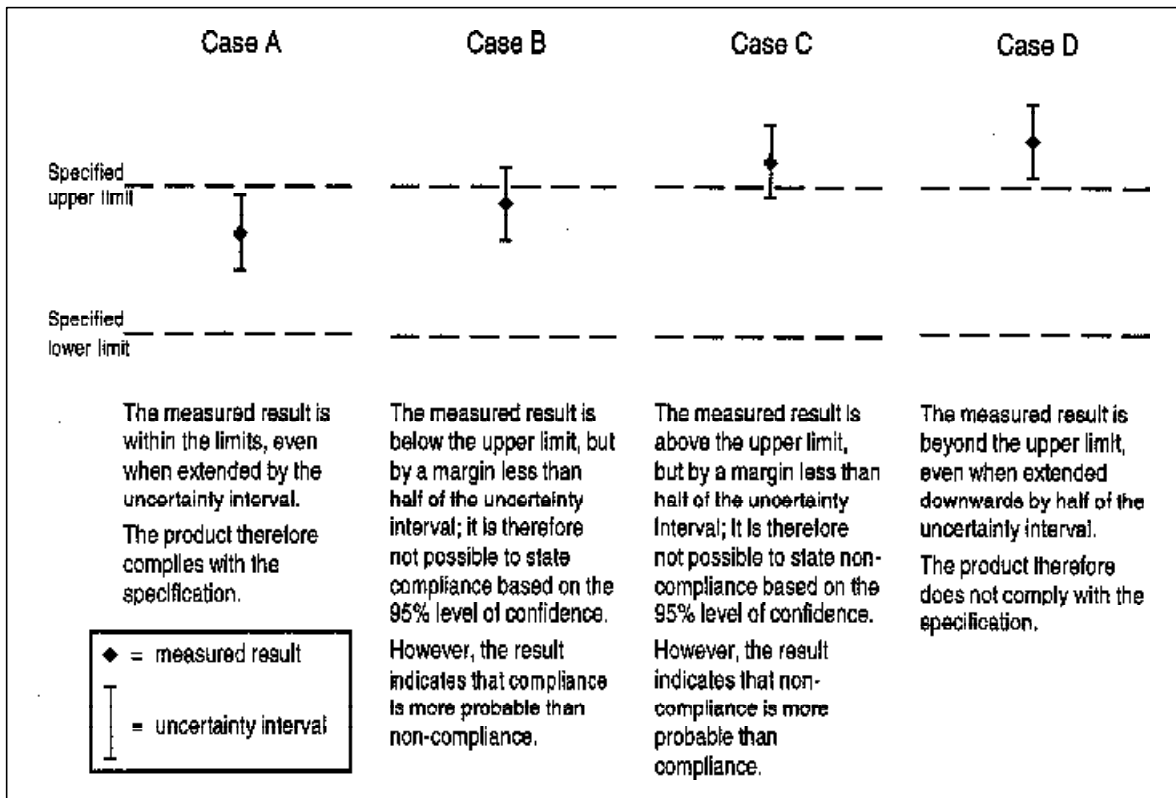
Case A   Case B   Case C   Case D

Specified upper limit

Specified lower limit

The measured result is within the limits, even when extended by the uncertainty interval.

The product therefore complies with the specification.

◆ = measured result

| = uncertainty interval

The measured result is below the upper limit, but by a margin less than half of the uncertainty interval; it is therefore not possible to state compliance based on the 95% level of confidence.

However, the result indicates that compliance is more probable than non-compliance.

The measured result is above the upper limit, but by a margin less than half of the uncertainty interval; it is therefore not possible to state non-compliance based on the 95% level of confidence.

However, the result indicates that non-compliance is more probable than compliance.

The measured result is beyond the upper limit, even when extended downwards by half of the uncertainty interval.

The product therefore does not comply with the specification.

Fig 1   Assessing compliance where the result is close to an upper limit

Case E   Case F   Case G   Case H

Specified upper limit

Specified lower limit

The measured result is within the limits, even when extended by the uncertainty interval.

The product therefore complies with the specification.

◆ = measured result

| = uncertainty interval

The measured result is above the lower limit, but by a margin less than half of the uncertainty interval; it is therefore not possible to state compliance based on the 95% level of confidence.

However, the result indicates that compliance is more probable than non-compliance.

The measured result is below the lower limit, but by a margin less than half of the uncertainty interval; it is therefore not possible to state non-compliance based on the 95% level of confidence.

However, the result indicates that non-compliance is more probable than compliance.

The measured result is beyond the lower limit, even when extended upwards by half of the uncertainty interval.

The product therefore does not comply with the specification.
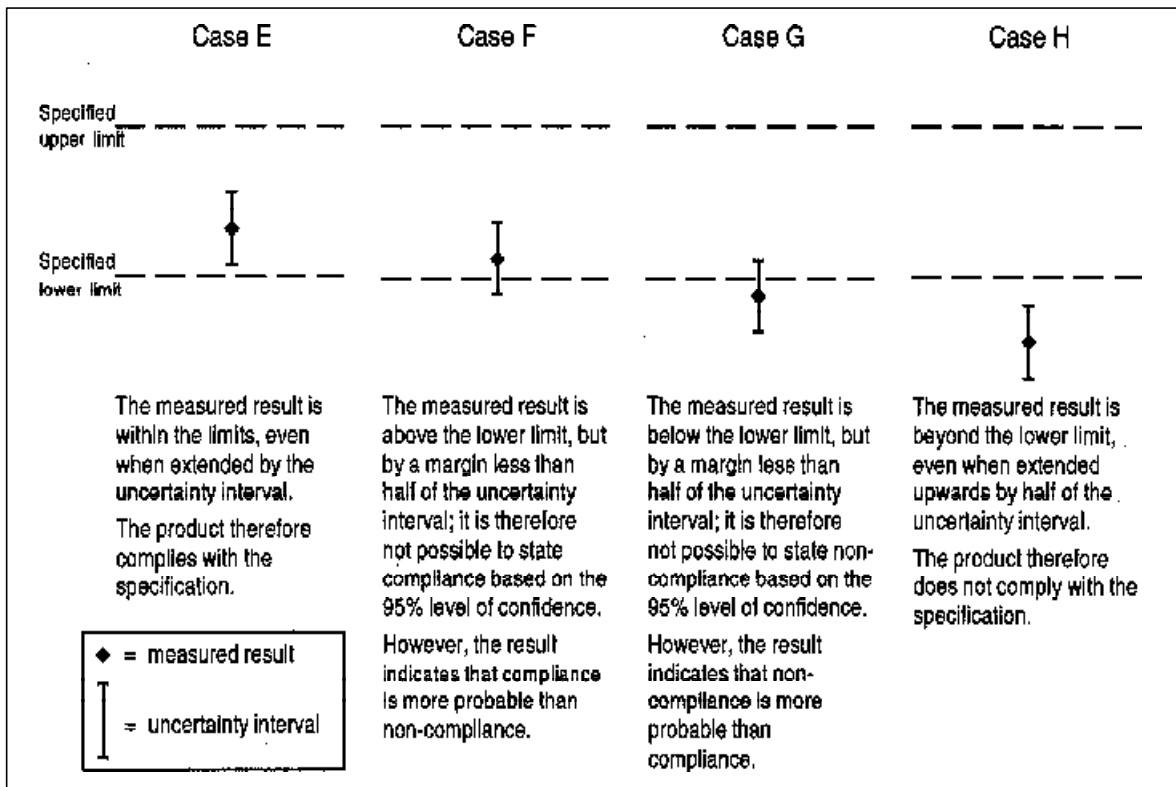
Fig 2   Assessing compliance where the result is close to a lower limit

with a statement indicating that compliance was not demonstrated. A suitable statement to cover these situations (Cases B and C, Fig 1 and Cases F and G, Fig 2) would be, for example:

The measured result is above (below) the specification limit by a margin less than the measurement uncertainty; it is therefore not possible to state compliance based on the 95% level of confidence. However, the result indicates that compliance (non-compliance) is more probable than non-compliance (compliance) with the specification limit.

Note: In these circumstances if a confidence limit of less than 95% is acceptable, a statement of compliance/non-compliance may be possible.

11 Proficiency testing and measurement audit

11.1   The estimation of the uncertainty of a test result may, in some cases, be confirmed by the results of a proficiency test or an inter-laboratory comparison. However, the evaluation of uncertainty for the test result obtained in the laboratory should ensure that all potential contributions to uncertainty have been considered.

11.2   Calibration laboratories in most fields take part in measurement audits. The results of the audits are used to confirm that the claimed Best Measurement Capabilities (BMCs), expressed as uncertainty, are being achieved by the calibration methods normally used. The BMCs are as stated on the schedule of an accredited calibration laboratory. However, where the information is available, the test laboratory should always use the value(s) of uncertainty quoted on the calibration certificate for the instrument(s) used in the test, since this may be significantly different to the BMC for the calibration laboratory.

Appendix A   Definitions of terms

Coverage factor

      A number that, when multiplied by the combined standard uncertainty, produces an interval (the expanded uncertainty) about the measurement result that may be expected to encompass a large, specified fraction (eg 95%) of the distribution of values that could be reasonably attributed to the measurand.

Error of measurement

      The result of a measurement minus the value of the measurand (not precisely quantifiable because the true value lies somewhere unknown within the range of uncertainty).

Level of confidence

      The probability that the value of the measurand lies within the quoted range of uncertainty.

Measurand

      The specific quantity subject to measurement.

Standard deviation

      The positive square root of the variance.

Uncertainty

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

Standard uncertainty

      The estimated standard deviation.

Combined standard uncertainty

      The result of the combination of standard uncertainty components.

Expanded uncertainty

      Obtained by multiplying the combined standard uncertainty by a coverage factor.

Variance

      A measure of the dispersion of a set of measurements; the sum of the squared deviations of the observations from their average, divided by one less than the number of observations.