ASSESSMENT OF UNCERTAINTIES OF MEASUREMENT
for calibration & testing laboratories

R R Cook
2nd Edition
2002
# Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preface</td>
<td>6</td>
</tr>
<tr>
<td>one Introduction</td>
<td>7</td>
</tr>
<tr>
<td>two General principles</td>
<td>9</td>
</tr>
<tr>
<td>three Sources of uncertainty and their estimation</td>
<td>13</td>
</tr>
<tr>
<td>four Determining the measurement uncertainty</td>
<td>17</td>
</tr>
<tr>
<td>five A simplified approach to uncertainty estimation</td>
<td>23</td>
</tr>
<tr>
<td>six Expression of uncertainty in test reports</td>
<td>27</td>
</tr>
<tr>
<td>seven Converting from one confidence level to another</td>
<td>29</td>
</tr>
<tr>
<td>eight Acknowledgements</td>
<td>31</td>
</tr>
<tr>
<td>nine References</td>
<td>33</td>
</tr>
</tbody>
</table>

# Appendices

<table>
<thead>
<tr>
<th>Appendix</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>36</td>
</tr>
<tr>
<td>ii</td>
<td>38</td>
</tr>
<tr>
<td>iii</td>
<td>41</td>
</tr>
<tr>
<td>iv</td>
<td>43</td>
</tr>
<tr>
<td>v</td>
<td>45</td>
</tr>
<tr>
<td>vi</td>
<td>46</td>
</tr>
<tr>
<td>vii</td>
<td>47</td>
</tr>
<tr>
<td>viii</td>
<td>49</td>
</tr>
<tr>
<td>ix</td>
<td>50</td>
</tr>
<tr>
<td>x</td>
<td>51</td>
</tr>
<tr>
<td>xi</td>
<td>53</td>
</tr>
</tbody>
</table>
Preface

In the two and a half years since this book was first published there has been an encouraging demand for it. Unfortunately the ubiquitous gremlins have been at work and several typographic errors and similar defects crept into the text. In order to correct these and to improve some other sections, a second edition has been produced. The contributions by many sharp-eyed readers in detecting and reporting the defects were very much appreciated.

R R Cook, 10th October 2001.
Introduction

Every measurement made has an error associated with it, and, without a quantitative statement of the error, a measurement lacks worth. Indeed without such a statement it lacks credibility. The parameter that quantifies the boundaries of the error of a measurement is called the uncertainty of measurement. Accuracy is a general term that is subject to various interpretations, whereas uncertainty has a specific meaning. Uncertainty is defined as the parameter, associated with the result of a measurement that characterises the dispersion of the values that could reasonably be attributed to the measurand. The measurand is the particular quantity subject to measurement.

Good laboratory practice requires that all reports express results in particular ways. The results should be expressed either in the manner prescribed by the test method used, or if a numeric value is given, should be accompanied by a statement of uncertainty. The practice of reporting numbers such that the last digit in each numeric value can be deemed to be significant, unless otherwise qualified, can be misleading and is no longer what may be considered best practice. Further, the determination of significant figures can only be done objectively after an assessment of the measurement uncertainty.

This document describes the assessment of uncertainty of measurement using the principles given in the ISO Guide to the Expression of Uncertainty in Measurement, hereafter called the ISO GUM.

In 1980, a working group of the Comité International des Poids et Mesures (CIPM) made recommendations on the expression of uncertainty. Following this, ISO was asked to develop a standard. This was done by a working group that was coordinated by the ISO Technical Advisory Group on Metrology (TAG4). Several draft documents on the expression of uncertainty in measurement were produced and a final version was published in 1993. A version with minor corrections was published in 1995. The ISO GUM presents a method of assessment of uncertainties that has a sound statistical base. However, this method is built on a number of assumptions. The main assumptions are that all measurement uncertainty components can be represented by a number equivalent to a variance and that the combined uncertainty can be obtained by combining the variances.

An important aspect of the ISO GUM is the recommendation that individual uncertainty components can be estimated by either a type A evaluation or a type B evaluation. A type A evaluation involves the use of statistical methods and applies only to series of observations, for example repeated measurements of the same quantity. A type B evaluation is one using any other means. Type B evaluations may involve the application of both knowledge and experience. It should be noted that the source of the uncertainty does not determine how it is calculated. In other words, type B evaluation does not only apply to uncertainty components arising from “systematic” errors; it can be applied to uncertainty sources that arise from “random” errors. Further, the shape of the distribution which describes the error does not determine the evaluation method.

Many current Australian and international standards still use the terms “random” and “systematic” extensively when considering accuracy or uncertainty and these terms can be expected to appear in general usage for some considerable time to come. It is up to the user to distinguish between ISO GUM practice and previous practice.

Some common sources of measurement errors that give rise to the total measurement uncertainty are listed later, however, it is not possible to cover all sources for all measurements. This is beyond the scope of this book and requires knowledge specific to each measurement field.

In addition to describing the determination of the total measurement uncertainty, a recommended method of expressing uncertainty in endorsed test reports is given. The uncertainty of measurement is often a crucial element in determining compliance with specifications and this aspect is also covered. An associated topic, determination of calibration intervals and drift, is briefly discussed. Worked examples and other details are included in a comprehensive set of appendices.

This book is intended to be a practical guide to uncertainty assessment for laboratories with/seeking NATA accreditation. More detailed explanations of some of the underlying theory can be obtained from the references listed. The reader’s attention is directed to Appendix i which gives definitions for special terms used in this document.
General principles

This book does not intend to give instructions about how to make measurements. It will be assumed that appropriate tests have been made and all due care has been taken in obtaining a value for the measurand (the subject of the measurement). Any measurand will have a true value that will be approximated by the measurement value. The error of the measurement is the difference between the measured value and the true value. In general the measured value cannot be repeated exactly, so we need a measure or parameter that describes not only the range of the error but also how fuzzy or dispersed the range is. This parameter is called the uncertainty of the measured value. The uncertainty is a measure of the dispersion that may reasonably be associated with the measured value. It gives a range, centred on the measured value, within which, to a stated probability, the true value lies. It is usual, but not universal, that the range have equal positive and negative limits.

Many measuring instruments and systems require the application of scaling factors and one or more known corrections to the reading to obtain the measurement value and further corrections may be necessary to correct the result to a value corresponding to standard test conditions. For example, a measurement made at 25 °C may need to be corrected to give a value at 23 °C. The result of these calculations and corrections is the measurement value. It is essential that all applicable corrections be made before any attempt is made to assess the uncertainty of measurement.

As already stated, the measurement uncertainty statement gives a range within which the measurand’s true value is considered to lie. This range is centred on the measurement value, and while the measurand’s true value is likely to be close to the measurement value, it may lie anywhere within the range. Indeed, there is a small probability that the measurand’s true value may lie outside the range.

In Australia, legal measurements made under the National Measurement Act 1960 and calibration reports issued by CSIRO’s National Measurement Laboratory (NML) contain an uncertainty statement based on a 95% confidence level. NATA’s policy on measurement uncertainties includes a requirement that all numerical results in an endorsed report must be accompanied by the relevant statement on uncertainties, except when the measurements are conducted in strict accordance with a given documentary standard and that reported uncertainties shall generally be at the 95% confidence level.

To be meaningful, the uncertainty statement must have an associated confidence level. It is necessary to state the probability that the true value lies within the range given. A 95% confidence level means that there is a 95% probability that the true value lies within the stated range, or in other words, that there are not more than five chances in 100 that the true value lies outside the range.

For example, a calibration report for a standard resistor which gives the value as 1.000 673 Ω with an uncertainty of measurement of ± 0.000 010 Ω at a 95% confidence level. This means that this resistor has a measured value of 1.000 673 Ω and the probability of the true value being outside the range of ± 0.000 010 Ω about the measured value is estimated to be only five in one hundred.

The reasons for choosing a 95% confidence level may be summarised as follows:

1. It is established practice through much of Europe, North America and Asia.
2. The ISO Guide assumes that the combined uncertainty has a distribution that is a close approximation to a normal distribution. A 95% confidence level approximates to a range of two standard deviations. It is a widely held view that, for most measurement systems, the approximation to a normal distribution for the distribution of the combined uncertainty is reliable out to two standard deviations, but beyond that the approximation is less reliable.
3. An approximate 95% confidence interval can be simply obtained by multiplying the combined standard uncertainty by 2. Whilst the calculations required for a more accurate determination are not difficult, for most testing applications and some calibration applications, the additional effort is not justified.

To assess the uncertainty of a measurement it is recommended that the first step be the construction of a model of the measurement system followed by a list of all the factors that can contribute errors to the final result. This requires a good understanding of the measurement principles, the equipment and the environment and is often the most difficult yet crucial step.

Often corrections must be applied to readings from instruments. These corrections will have their own uncertainties that must be considered when the model is developed. Further advice on corrections is given in appendix vi.

The next step is to decide whether each component is to be evaluated by type A or type B analysis. Then it is necessary to calculate and estimate, as appropriate, the limits for each component. For type A evaluation there are well-tested mathematical techniques for calculating the limits, but, for type B evaluation, an element of judgment is often involved and consequently their estimation might be thought to be less precise. In practice this is not so, providing the recommendations of the ISO GUM are followed. That is, the estimations must be soundly based and defendable.
Regardless of whether a type A or a type B analysis is used, each uncertainty component is estimated so that it is the equivalent of a standard deviation. These components are known as standard uncertainties.

For many measurements, the units of the uncertainty components differ from the units of the measurand. For example, consider a measurand which has units of MPa and has sensitivity to the test temperature. The measured values of temperature and the associated uncertainty will have units of °C, so a conversion factor or weighting factor will be required to obtain the effect of uncertain temperature in terms of pressure units. Such conversion factors are known as sensitivity coefficients. They relate the sensitivity of the measurand to the particular influence or input quantity. Failure to apply these sensitivity coefficients will result in gross errors and nonsensical uncertainty values. Neglecting or incorrectly estimating the sensitivity coefficients are common causes of erroneous uncertainty estimates.

Finally, the standard uncertainty components must be combined to give the overall measurement uncertainty. Providing the standard uncertainties represent the equivalent of a standard deviation, they may be combined using the law of combination of variances. This combined standard uncertainty gives a range with a 68% confidence level, which for many applications is insufficient. Hence, an increased range, called the expanded uncertainty, corresponding to a higher level of confidence is usually quoted. A 95% confidence level is the most common as discussed earlier. A coverage factor is used to expand the combined uncertainty to the expanded uncertainty. Either an appropriate but convenient coverage factor, such as 2, may be used or a more rigorous calculation may be done to find the mathematically correct factor.

These steps are dealt with in more detail in the following sections.
Sources of uncertainty and their estimation

The nature of errors
It should be borne in mind that the measurement process must be “well behaved” and under control before any assessment of uncertainties is attempted. When this has been achieved, if sufficient resolution is available, it will be found that repeated measurements still yield different values.

It will be found that the measured values vary randomly, often without apparent cause, hence the term Random Errors. The values will cluster around a central value.

The best estimate of this central value is the mean or average of the measured values, but the mean may not be the true value. It may be biased away from the true value by a combination of errors that are essentially constant. If their values were able to be accurately determined then they could be applied as corrections to the mean, thereby bringing it into better agreement with the true value.

Whilst precise determination is frequently not possible, there will usually be extreme limits which can be estimated or measured which will encompass the possible value of such errors. These errors are called Systematic Errors. Repeated measurements will not reduce the size of a systematic error or, in general, improve the knowledge of its magnitude or sign.

There are established measurement procedures, such as “reduction to a mean instant”, and “randomising” which minimise systematic errors, but there will always be a residual uncompensated component that adds to the uncertainty.

Even when the effect of all known systematic errors has been accounted for, there still remain random variations whose magnitude and sign are unpredictable. These random errors cannot be eliminated, but they are amenable to statistical treatment. Averaging many repeated measurement results does improve the estimate of the mean and reduces the magnitude of the estimate of the variance associated with the mean.

The total error of the measurement thus consists of at least one systematic component and at least one random component.

Sources of uncertainty
It is not practical in this document to list all probable sources of uncertainty: what follows is the list given by the ISO GUM. The list, reproduced in Table 1, is intended to be illustrative and general, rather than an exhaustive, detailed list.

<table>
<thead>
<tr>
<th>TABLE 1 ISO GUIDE LIST OF SOURCES OF UNCERTAINTIES</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) incomplete definition of the measurand;</td>
</tr>
<tr>
<td>(b) imperfect realisation of the definition of the</td>
</tr>
<tr>
<td>measurand;</td>
</tr>
<tr>
<td>(c) nonrepresentative sampling - the sample measured</td>
</tr>
<tr>
<td>may not represent the defined measurand;</td>
</tr>
<tr>
<td>(d) inadequate knowledge of the effects of environ-</td>
</tr>
<tr>
<td>mental conditions on the measurement or imperfect</td>
</tr>
<tr>
<td>measurement of environmental conditions;</td>
</tr>
<tr>
<td>(e) personal bias in reading analogue instruments;</td>
</tr>
<tr>
<td>(f) finite instrument resolution or discrimination</td>
</tr>
<tr>
<td>threshold;</td>
</tr>
<tr>
<td>(h) inexact values of measurement standards and</td>
</tr>
<tr>
<td>reference materials;</td>
</tr>
<tr>
<td>(i) inexact values of constants and other parameters</td>
</tr>
<tr>
<td>obtained from external sources and used in the data-</td>
</tr>
<tr>
<td>reduction algorithm;</td>
</tr>
<tr>
<td>(j) approximations and assumptions incorporated in</td>
</tr>
<tr>
<td>the measurement method and procedure;</td>
</tr>
<tr>
<td>(k) variations in repeated observations of the</td>
</tr>
<tr>
<td>measurand under apparently identical conditions.</td>
</tr>
</tbody>
</table>

Random errors
The random variations in the results of measurements are called random errors. Sources of these errors are many and are inherent in the measuring instrument, the item under test, the test procedure and the test environment. Sources of random errors include varying influence parameters which cannot be adequately controlled, such as air currents and ambient temperature fluctuations, relative humidity variations, power source disturbances, microphonic components and cables exhibiting small piezoelectric effects which are subject to mechanical vibration, electromagnetic interference, minor connector contact resistance variations, lack of discrimination in setting an index to a fiducial mark on a scale or interpolation between marked points on a scale and incidental modulation (amplitude and/or frequency or phase) of signal sources. Mechanical vibrations from machinery, foot and vehicular traffic often affect instruments and sometimes the measurand.

In many cases a random component will be immediately obvious because it occurs at a rate and amplitude sufficient to make taking a particular reading very difficult. For example, the last digit of a digital readout may be varying by many units on every sample taken by the measuring instrument, so that in the space of a second, several different readings are displayed. In the case of an analogue display, a fluctuation of the index may be an indication of a random error. If random errors are not apparent after careful observation and ex-
amination of repeated measurements then it can be said that the measuring system has insufficient sensitivity to detect random errors. Of course this does not mean that the system is inadequate for its purpose.

The estimation of the random components may be done in the manner described later using a type A analysis. This requires that the measurements be repeated a number of times. In general this would not be less than five if significant reduction of the random uncertainty is to be achieved. A simple manipulation of the data produces a mean value and a standard deviation. The standard deviation is a number having the same units as the mean and is a measure of the spread of results. Performing twenty or more repeated readings is often a trivial matter for modern instrumentation and can produce very much better results. To fully account for all random sources of error the measurement system should be disassembled and reassembled between each measurement. This is not usually practicable and so it is necessary to carefully assess all possible error components and make approximate allowances. Also, with proper design of the measurement system, errors due to the arrangement of the measurement system can usually be made insignificant.

Where multiple measurements on an instrument are not practicable, a component of uncertainty should be included to allow for short term instrument instabilities. The estimate of this component must be based on prior experience of instabilities of the particular instrument or other instruments of the same type. This is a type B analysis of the random error contribution.

Large random variations in a measurement process should be investigated, as this may indicate that the measurement process is not under control due to a problem such as inadequate screening, a poorly made connection, a fault in the instrumentation, a high level of external electrical interference, inadequate cleaning of the apparatus, sample contamination or insufficient isolation from other influence quantities.

**Systematic errors**

Systematic errors may be considered to be corrections that are not applied because their actual magnitude and sign are not precisely known. It is a characteristic of systematic errors that they can neither be reduced, nor our knowledge of their magnitude improved, by taking the average of repeated measurements under constant conditions.

Many systematic errors can be reduced or virtually eliminated by careful choice of measuring method, or by an extended calibration process to determine the actual error sign and magnitude or, in some cases, by altering the influence quantity over a range and performing repeated readings so as to either characterise or randomise the error. For example, repeating the measurements at many temperatures above and below the calibration temperature and reducing the result back to the calibration temperature can minimise the effect of measurements at temperatures other than the temperature at which the instrument was calibrated.

These randomising processes unfortunately require an investment in time and effort, which may not be economic.

Reducing all measurements to a common time can reduce the effect of systematic errors, whose value changes uniformly with time. For example, the effect of drift in an electronic mass balance can be minimised by making measurements in a timed symmetric sequence such as Standard, Unknown, Unknown, Standard. The average of the readings for the Unknown and the Standard then gives the readings for each at the median time.

Because systematic errors are essentially constant under fixed conditions, they can be very difficult to detect. It is possible, and indeed common, for a systematic error to be many times greater than either the random effects or the resolution and repeatability of the measurement. When it is appropriate to carefully quantify systematic errors it will be necessary to perform supplementary measurements. The decision as to whether this is done or not ultimately rests on cost and time constraints. A distinction should be drawn between known corrections that are not applied and systematic errors. We are concerned here with unknown errors. Corrections can and should always be applied.

Some systematic measurement errors are produced by influence factors such as ambient temperature, humidity, mains voltage, external electrical noise, internal electrical noise, temperature gradients, mechanical vibration and so on. Whilst some of these influences produce random errors, many also produce systematic errors, that is errors which remain constant while the factor producing them remains constant.

Ideally, the various influence quantities would be investigated by holding all constant except for one and for this to be varied in a controlled manner. Once the relationship was characterised a correction could be applied. Unfortunately it is often difficult, if not impossible, to either measure or control the influence quantity with sufficient accuracy to be able to apply comprehensive corrections, thus our imperfect knowledge of these effects gives rise to an uncertainty in the measurement due to residual or uncompensated systematic errors. Even after such corrections were applied, there would be residual errors arising from the uncertainties of the correction determination.

Other sources of systematic errors include interaction between the measuring instrument and the measurand. These errors can exceed the intrinsic capability of the test instrument. For example, an electrical instrument may draw current from the circuit, thereby changing the voltages being measured. Measuring a current with
a meter using a shunt introduces a voltage drop in the circuit being tested and reduces the current below the level that would otherwise occur. A length-measuring instrument using a contacting probe will compress the surface against which it bears, thus reducing the length measured.

Another example is the apparent difference in dimensions that result when the same measurement is made using a high-resolution contact probe instrument and one using a beam of light. This difference arises because the surface from which the light reflects differs from the surface which the probe presses against. Both measurements are correct, but they are measurements of subtly different measurands. The measurand needs to be clearly defined for these two types of measurement to be compared. Adequate definition of the measurand is important, as imprecision in the definition can lead to systematic errors.

Much instrumentation today uses electronic amplifiers and electro-optical displays or an electronic interface to a computer. These can be affected by induced currents from mains wiring, imperfections in the electronic circuits, deviation of measurement temperature from instrument calibration temperature, scale errors, self-heating, and so on. Electrical measurements are affected by lead resistance effects, signal leakage from generator to detector, mismatch of rf components, drift in signal sources and change in response with change in test frequency. Electro-chemical sensors, such as used in pH meters, suffer from contamination. Mechanical systems may have backlash in gearing and levers, imperfect knife-edges, and uncompensated expansion due to pressure or temperature changes.

Errors in the standard against which measurements are made are common to all measurements. When standards such as masses, ratio transformers or reference materials are manufactured, there remains some small deviations from the nominal value. Applying corrections determined by a higher echelon laboratory would minimise the effect of these errors. This leads to one of the most obvious systematic errors, that given as the uncertainty in the calibration certificate for the standard or measuring instrument, or both if appropriate.

In general, except for uncertainties given on calibration certificates supplied by another laboratory, systematic uncertainty components are estimated from measurements, manufacturer’s specifications and calculations based on established laws of engineering, physics and chemistry.

For example, suppose a standard resistor, which is tested at negligible load, has a temperature coefficient of 20 ppm/°C and during measurements on it, the thermometer used to measure its temperature shows readings which vary between 19.5 °C and 20.5 °C due to laboratory temperature cycling. Because of the effect of thermal time constants and other effects we cannot be sure of the exact temperature of the resistor at the time of any measurement of resistance. We can, with a high degree of certainty, say that the temperature of the resistor was always within the range 19.5 °C to 20.5 °C. In the absence of other experimental evidence we may assume an average temperature of 20.0 °C with a variation of ±0.5 °C. Thus, the range of error due to temperature effects is ±0.5 °C × 20 ppm/°C which is ±10 ppm.

Operator related errors are generally systematic, but may also be random in nature. Deftness of touch, skill in setting to scale marks, and ability to visually subdivide scales, are some of the well-known operator effects.

Systematic errors which have been overlooked are the greatest cause for concern of the metrologist, particularly when engaged in a testing or measuring procedure for the first time. Using a different measurement system on the same measurand will assist in detection of systematic errors providing that both systems have comparable resolution. Other powerful aids are proficiency tests and inter-laboratory comparisons. The use of these is highly recommended.

It should be noted that reading errors (recording a value different to that indicated on the instrument), transcription errors, computational errors, incorrect application of scale factors and corrections, and other errors in this general category are blunders. They cannot be properly treated as a systematic error and must be eliminated by good laboratory practice and careful cross-checking. Reading back from the laboratory book to the instrument after recording a reading, and having all calculations and transcriptions checked by another person, are recommended procedures.
Determining the measurement uncertainty

Estimating uncertainty

Uncertainty estimation is a straightforward process in principle. Some measurement processes are complex and hence the uncertainty calculations take on a degree of complexity. The overall principles remain the same.

A model of the measurement should be developed, either implicitly or explicitly depending on the complexity of the measurement. Next, all the uncertainty components are listed and their standard uncertainties calculated. Generally, it will be necessary to also calculate sensitivity coefficients. Sensitivity coefficients convert the components to the same units as the measurand, and also scale or weight them so that they have the proper influence on the total uncertainty. Then the components are combined and an expanded uncertainty calculated for the measurand. These steps are dealt with in detail in the following sections and summarised in Table 2 on page 22.

Modelling the measurement system

Before any calculations are attempted it is necessary to consider the measurement system and its environment. A simple sketch or an equation may be all that is required. The model should provide a simple means of describing the relationship between the input parameters and influence quantities to the measurand. This is the step that many metrologists find most difficult. The worked examples in appendix xi should assist in understanding this task. Without a model some significant uncertainties may be overlooked or it may be difficult to determine the sensitivity coefficients.

Determining standard uncertainties

For type A assessments it is necessary to choose the appropriate statistical method. For the situation where the average of several measurements is taken, there is an established method, which yields a standard uncertainty that is a measure of the dispersion of the average value. The procedure is to calculate the mean of the set of readings and a standard deviation known as the Experimental Standard Deviation of the Mean. These simple procedures are described in detail in the next section. When data is treated to find a “best fit,” such as when a curve is fitted or finding a distance when surveying land, the standard techniques will also yield a standard uncertainty. Type A assessments are perhaps the most straightforward to make.

For type B assessments the knowledge and understanding of the metrologist is crucial. Each component must be examined to determine its limits, that is the range of dispersion, and the nature of the dispersion. This means that it must be decided if the uncertainty component is expected to cluster near a central value or is evenly dispersed across its range, or if it is more likely to be at the extremes. This analysis is called determining the distribution curve of the uncertainty. Various examples are given later. Once the range and nature of the dispersion have been decided, specific formulas are applied to find the standard uncertainty component.

The ISO GUM assumes that for each uncertainty component there is a parameter, similar to a variance, which describes the magnitude of the component. While the variances are the key element from a mathematician’s view, metrologists prefer to work with standard deviations because they have the same units as the associated parameter. These standard deviations are called standard uncertainties. A measured temperature may have a value of 427 °C with a standard uncertainty of 1.7 °C. This is easier to visualise than a variance of (3 °C)^2. There are rules that may be used to calculate the standard uncertainties.

Type A assessment

When a measurement is repeated several times, the mean value and the standard deviation can be calculated. The standard deviation describes the dispersion applicable to the whole population of possible measured values. The relevant formulas are as follows.

The mean

The mean of a set of repeated measurements is the best estimate of the true mean of the population and is calculated from the formula that follows:

\[ \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \]  

where:

- \( \bar{x} \) represents the mean of \( n \) measurements
- \( x_i \) represents the individual measurements
- \( \sum_{i=1}^{n} \) represents the summation of all the measurements.

Standard deviation

The standard deviation is a measure of the dispersion of the population from which the \( n \) values were taken.

\[ s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}} \]  

where:

- \( s \) is the standard deviation
- the other symbols have the same meaning as before.
Note that most scientific calculators and computer spreadsheet programs have these functions built in, so all that is necessary is to enter the data and press the appropriate button. Care needs to be taken that the button labelled “s” or “σn-1” is used to obtain the unweighted estimate of the population standard deviation. There is also a button that computes the standard deviation using n rather than n-1 on the bottom line of equation 2. This is only correct when n is very large.

The ESDM

If after taking one set of repeated measurements we were to take a second set of measurements we could again calculate the mean. It most likely would be slightly different from the first mean. Because we are taking small samples from an infinite population of possible values, the mean of each set of measurements will differ a little from other means obtained in the same way. That is, we see some dispersion associated with the means of sets of repeated measurements.

Having taken a mean of several measurements we expect that this will be closer to the population mean or true mean than any single measurement. Also, although sets of means have a dispersion, we expect a smaller dispersion than for the original or parent population. It follows that we also expect that the mean should have its own standard deviation and that this will be less than that of the parent population. This is correct, and the associated deviation is called the Experimental Standard Deviation of the Mean, or ESDM.

When the measurand is obtained from a mean of repeated measurements, the ESDM is also the standard uncertainty for the dispersion of the measurand associated with random effects. Thus, this type A method of calculation of uncertainty components involves calculating the ESDM. The ESDM is given by the formula:

\[ ESDM = \frac{s}{\sqrt{n}} \]

where the terms have the same meaning as before.

This shows that increasing the number of repeated measurements reduces the dispersion to be associated with the mean. Unfortunately, once n is large enough to give a good estimate of s, to halve the ESDM requires that there be four times as many measurements.

It is important to note that, regardless of the shape of the distribution of the population from which the individual measurements have come, the mean belongs to a normal distribution.

Type B assessment

Amongst the skills already mentioned for type B assessment, the ISO GUM advises that there is no “substitute for critical thinking, intellectual honesty, and professional skill”. While this is intended to apply to the whole process it is of most significance when making type B estimates.

The simplest example is that of using the uncertainty of a standard which has a calibration certificate. To obtain the standard uncertainty the expanded uncertainty on the certificate is divided by the coverage factor given on the certificate. In the absence of a value for the coverage factor, a factor of 2 may be used if the expanded uncertainty has a 95% confidence level.

Thus:

\[ u(x) = \frac{U_{95}}{2} \]  

where:  

- \( u(x) \) is the standard uncertainty of the calibration values given on the certificate and  
- \( U_{95} \) is the expanded uncertainty at a 95% confidence limit given on the certificate

Equation 4 gives an approximate value for \( u(x) \). When \( k \), the coverage factor is known it should be used instead of the factor 2.

Thus:

\[ u(x) = \frac{U_{95}}{k} \]

The standard uncertainty \( u(x) \) of the calibration values given on the calibration certificate is the \( i^{th} \) component in a new uncertainty estimate.

In other cases the uncertainty can usually be given limits, \( \pm a \) and some distribution determined.

Rectangular distribution

If limits can be determined, but the value of the measurand is just as likely to be anywhere in the range, then the distribution of the uncertainty is a rectangular distribution. It is sometimes called the distribution of minimum knowledge, and is most often applied when worst case limits are calculated.

For example, consider the earlier example where the temperature was known to have a range of 19.5°C to 20.5°C. This means we have \( a = 0.5 \) °C. If we had no further information, then we can only say that it is equally likely that the temperature can have any value in that range. This is a uniform or rectangular distribution, and the formula for the standard deviation of such a distribution gives us the standard uncertainty due to the variation in temperature.

The formula is:

\[ u(x) = \frac{a}{\sqrt{3}} \]  

......6)
where:

\[ u(x_i) \] is the standard uncertainty and
\[ a \] is the semi-range of the limits of the uncertainty component

There are other formulas for all other distributions that may be appropriate. These distributions can be used if we have more knowledge of the nature of the dispersion. Some of the common distributions are dealt with later.

**Sensitivity coefficients**

This is the other aspect of uncertainty estimation that causes metrologists considerable difficulty. As already mentioned the sensitivity coefficient converts all uncertainty components to the same units as the measurand. This is a necessary precondition to combining the components as combining, say, metres and kilograms is not dimensionally correct. Common sense would also tell us that only like items should be combined.

The sensitivity coefficient also performs a scaling or weighting function. This is to be expected from the earlier description where it was stated that the sensitivity coefficient was a measure of the sensitivity of the measurand to a particular input or influence parameter. For example, a steel tape has roughly twelve times the linear temperature expansion coefficient of an invar tape (11.5 ppm/°C compared to 1 ppm/°C). Thus, for measurements involving a steel tape, the measured length is more sensitive to variations in the tape temperature than for the invar one. So the sensitivity of the uncertainty of the measured length to temperature is greater for steel tapes than for invar tapes.

It may often be a trivial and obvious matter to apply the sensitivity coefficients. For example, the temperature coefficient of linear expansion converts uncertainties in measured temperature with units of degrees into uncertainties with units of length. In the example just given the sensitivity coefficient for the measured length for temperature uncertainties is 11.5 ppm/°C for the steel tape or 1 ppm/°C for the invar tape.

Evaluation of the sensitivity coefficients can always be done by algebraic differentiation of the equation that models the measurement or, in more complex cases, by numerical calculation which approximates the differentiation process.

The general formula for the sensitivity coefficient is:

\[ c_i = \frac{\partial y}{\partial x_i} \]  

\[ ......7 \]

where:

\[ c_i \] is the sensitivity coefficient for component \( x_i \) and
\[ y \] the measurand, is a function of \( x_i \) and
\[ \frac{\partial y}{\partial x_i} \] is the partial derivative of \( y \) with respect to \( x_i \)

The partial derivative gives the slope of the curve that results when the function \( y \), the measurand, is plotted for the appropriate range of \( x_i \) values. The slope or derivative is the sensitivity of the measurand to a particular component of the function. It gives a numerical value that describes how the measurand is influenced by changes in the particular factor.

Consider an example where the value of a resistance, \( R_t \) at a test temperature, \( t \), is given by the equation

\[ R_t = R_0 (1 + \alpha t) \]

where:

\[ \alpha \] is the temperature coefficient of the resistor, ohms/°C
\[ t \] is the temperature of the resistor in °C and
\[ R_0 \] is the resistance in ohms at the reference temperature, 0 °C.

We can differentiate the equation for \( R_t \) with respect to \( t \) as follows

\[ \frac{\partial R_t}{\partial t} = R_0 \alpha = \alpha R_0 \]

This is the result we would expect by applying logic. We could reason that if the temperature changes from 0 °C to temperature \( t \) °C, then the resistance changes by an amount \( \alpha R_0 \). So the sensitivity coefficient of \( R_t \) with respect to changes in temperature \( t \), is \( \alpha R_0 \).

A similar process is used to determine all sensitivity coefficients.

It will be found to be convenient in some cases to calculate the measurand in terms of proportional parts (percent or parts per million) deviation from a nominal value. For example, a standard resistor might be found to be 73 ppm low compared to its nominal value. It would be appropriate to express the uncertainty in the same units, for example ±2 ppm. Often, but not always, this procedure allows the use of the factor 1 as the sensitivity coefficients. The influences all need to be calculated in terms of ppm/(unit of influence) for this to work.
Next to forming a model, determining sensitivity coefficients is the area where difficulty is most likely to be encountered. It is often best to use the base SI units for all inputs as errors of three orders of magnitude are easily made when calculating sensitivity coefficients. A “reality check” should be made when doing an estimate for the first time. For example, the numeric value could be compared with that expected from experience.

**Combined uncertainty**

Once the individual components and their associated sensitivity coefficients have been evaluated, they may be combined to produce the combined standard uncertainty. This is done by using equation (10) of the ISO GUM. This equation must be modified if any of the components are correlated or there are higher order effects. These conditions are dealt with separately later in this book.

The ISO GUM equation (10) can be written as:

\[
uc(y) = \sqrt{\sum_{i} c_{i} u(x_{i})^{2}}
\]

where:
- \(uc(y)\) is the combined standard uncertainty of the measurand,
- \(c_{i}\) is the sensitivity coefficient for \(u(x_{i})\) which is the \(i^{th}\) standard uncertainty of the input estimates and
- \(\sum_{i}^{n}\) denotes the summation of all the terms, of which there are \(n\).

This means that the uncertainty components are converted to the same units as the measurand using sensitivity coefficients then these products are squared and summed. The combined uncertainty is the square root of the sum. This is often called the “root sum of the squares” (RSS).

**Expanded uncertainty**

In order to have an adequate probability that the value of the measurand lies within the range given by the uncertainty the combined uncertainty is multiplied by a coverage factor. This coverage factor may be selected or it may be calculated so as to reflect a stated confidence level. For example, a factor of 2 gives an expanded uncertainty \((U)\) with an approximate 95% confidence level.

When the uncertainty has been estimated from data with poor reliability, a larger coverage factor may be required to maintain a 95% confidence interval. One measure of reliability is the number of degrees of freedom, a parameter which is dealt with in more detail shortly.

The expanded uncertainty is calculated from

\[
U_{95} = k \cdot uc(y)
\]

where:
- \(U_{95}\) is the expanded uncertainty at a 95% confidence limit
- \(k\) is the coverage factor, and
- \(uc(y)\) is the combined standard uncertainty

By assuming that the combined uncertainty has essentially a normal distribution we can use Student’s \(t\) factor as the coverage factor \(k\). This is justified by invoking the Central Limit Theorem, which in essence states that if many distributions are combined, irrespective of their own shape, the combined distribution will approximate a normal distribution. Hence the combined uncertainty will tend towards a normal distribution as more and more components are included.

We can find values of \(t\) for any desired level of confidence provided we have the number of degrees of freedom. The question is how do we find the number of degrees of freedom for the combined standard uncertainty?

**Effective degrees of freedom**

Whilst the reason for determining the number of degrees of freedom associated with an uncertainty component is to allow the correct selection of the value of Student’s \(t\), it also gives an indication of how well a component may be relied upon. A high number of degrees of freedom is associated with a large number of measurements or a value with a low variance or low dispersion associated with it. A low number of degrees of freedom corresponds to a large dispersion or poorer confidence in the value.

Every component of uncertainty can have an appropriate number of degrees of freedom, \(\nu\), assigned to it. For the mean, \(\bar{x}\), for example, \(\nu = n - 1\), where \(n\) is the number of repeated measurements. For other type A assessments, the process is also quite straightforward. For example, most spreadsheets provide the standard deviation of the fit when data is fitted to a curve. This standard deviation may be used as the uncertainty in the fitted value due to the scatter of the measured values. The question is how to assign components evaluated by type B processes.

For some distributions, the limits may be determined so that we have complete confidence in their value. In such instances the number of degrees of freedom is effectively
The assigning of limits which are worst case leads to this instance, namely infinite degrees of freedom, and simplifies the calculation of the effective degrees of freedom of the combined uncertainty.

If the limits themselves have some uncertainty, then a lesser number of degrees of freedom must be assigned. The ISO GUM gives a formula that is applicable to all distributions. It is equation G.3 that is:

$$\nu \approx 1 - \frac{1}{2} \left[ \frac{\Delta u(x_j)}{u(x_j)} \right]^2$$

where:

- $\Delta u(x_j)$ is the relative uncertainty in the uncertainty $u(x_j)$

This is a number less than 1, but may for convenience be thought of as a percentage or a fraction. The smaller the number, the better defined is the magnitude of the uncertainty.

For example, if the relative uncertainty is 10%, ie:

$$\frac{\Delta u(x_j)}{u(x_j)} = 0.1$$

then it can be shown that the number of degrees of freedom is 50. For a relative uncertainty of 25% then $\nu = 8$ and for a relative uncertainty of 50%, $\nu$ is only 2.

Rather than become seduced by the elegance of the mathematics, it is better to try to determine the limits more definitely, particularly if the uncertainty is a major one.

It is of interest to note that equation (10) tells us that when we have made 51 measurements and taken the mean, the relative uncertainty in the uncertainty of the mean is 10%. This shows that even when many measurements are taken, the reliability of the uncertainty is not necessarily any better than when a type B assessment is made. Indeed, it is usually better to rely on prior knowledge rather than using an uncertainty based on two or three measurements. It also shows why we restrict the uncertainty to two digits. The value is usually not reliable enough to quote to better than 1% resolution.

Once the uncertainty components have been combined, it remains to find the number of degrees of freedom in the combined uncertainty. The degrees of freedom for each component must also be combined to find the effective number of degrees of freedom to be associated with the combined uncertainty. This is calculated using the Welch-Satterthwaite equation, which is:

$$\nu_{eff} = \frac{n \sum_i u_i^4(y)}{\sum_i u_i^4(y) \nu_i}$$

where:

- $\nu_{eff}$ is the effective number of degrees of freedom for $u_i$, the combined uncertainty and
- $\nu_i$ is the number of degrees of freedom for $u_i$, the $ith$ uncertainty term.
- $u_i(y)$ is the product $c_i u(x_i)$, with the sign of $c_i$ being neglected.

The other terms have their usual meaning.

**TABLE 2 THE SIX STEPS TO DETERMINING UNCERTAINTY OF MEASUREMENT**

1. Make a model of the measurement system.
2. List all the sources of uncertainties.
3. Calculate the standard uncertainties for each component using type A analysis for those with repeated measurements and type B for others.
4. Calculate the sensitivity coefficients.
5. Calculate the combined uncertainty, and, if appropriate its effective degrees of freedom.
6. Calculate the expanded uncertainty. Use a nominal or a calculated coverage factor. Round the measured value and the uncertainty to obtain the reported values.
A simplified approach to uncertainty estimation

It is possible in many testing and measuring situations to make some broad assumptions and to simplify the process of uncertainty analysis. This is not applicable to all cases because of the broadness of the underlying assumptions.

The assumptions are:

- The effects of correlation, higher order terms and non-linearity are negligible.
- When uncertainty components are not available from a source, such as a calibration certificate and must be estimated, it is sufficient to take worst case limits and use rectangular distributions.
- A nominal coverage factor may be used. (Using 2 gives an approximate 95% confidence interval)

The approach may best be explained by an illustrative example. Assume that a bright rolled steel bar is to be measured to see if it meets a manufacturing specification. To ensure that a marginal sized bar is not accepted, the uncertainty of the measured diameter is to be estimated. The bar diameter is measured with a digital micrometer at five places along its length and the following results, with units of mm, obtained.

19.003, 18.998, 19.005, 19.007, 19.006

The model is that the measurand is the average measured diameter plus the micrometer correction.

We now consider the contributions to the total uncertainty. Firstly, there is some variation in the measured diameter. The rod may not be round or its diameter may vary along its length. The micrometer is not perfect so it must contribute to the error. As it is a digital instrument, parallax and scale interpolation do not apply. It has a calibration certificate that states that it has a correction of zero, but the uncertainty of the correction is 4 µm at a 95% confidence level.

From experience we know that the effect of environmental factors can be neglected when measurements are made on steel. (The reader might like to check that even under extreme conditions the uncertainty due to temperature effects would not be expected to exceed ±0.3 µm). The micrometer has a mechanism to provide a constant measuring force, so operator influence is minimised. It has a resolution of 1 µm that we will take into account. Assume that we cannot think of any other sources of uncertainty, so we proceed with the calculations.

We have some repeat measurements so we can make a type A assessment of the variation of the diameter. Using a calculator we find the following:

The average indicated diameter is 19.003 8 mm. As the correction is zero this is our measured value.

The standard deviation of the measured values is 0.003 563 71 mm.

The \( ESDM \) is

\[
\frac{s}{\sqrt{n}} = \frac{0.00356371}{\sqrt{5}} = 0.00159374 \text{ mm}
\]

Thus the first of the standard uncertainties is

\[ u(x_1) = 0.00159374 \text{ mm} \]

It might be argued that it is not necessary to record to greater resolution than 0.001 mm. Retaining extra digits involves very little extra effort and ensures that additional errors due to rounding are avoided.

The next term is from the calibration certificate. If the expanded uncertainty of the correction does not have the coverage factor given, then we may assume that the factor was 2. To find the standard uncertainty we divide the expanded uncertainty by the coverage factor. Thus we have the second standard uncertainty:

\[ u(x_2) = \frac{4}{2} = 2.00 \text{ µm or 0.002 mm} \]

Remember, we must convert to the same units as the measurand.

The third term is the resolution. We will assume that the reading could be in error by as much as 0.5 µm or ± 0.000 5 mm. Assuming a rectangular distribution, then the semi-range \( a \) is:

\[ a = 0.000 5 \text{ mm} \]

and so

\[
u(x_3) = \frac{a}{\sqrt{3}} = \frac{0.0005}{\sqrt{3}} = 0.00028868 \text{ mm}
\]

We now use the RSS combination equation.

\[
u_r(y) = \sqrt{\sum_{i=1}^{n} u_i^2(y)}
\]

As the sensitivity coefficients are 1, the formula can be written as:

\[
u_r(y) = \sqrt{u^2(x_1) + u^2(x_2) + u^2(x_3)}
\]
which gives
\[ u(y) = 0.002\,574\,\text{mm} \]

To get the expanded uncertainty we will use a coverage factor of 2, so:
\[ U = 0.005\,147\,\text{mm} \]

This can be rounded to 0.005 2 mm or even 0.005 mm.

Thus the bar diameter is:
\[ 19.004\,\text{mm} \]

The uncertainty of the diameter is:
\[ \pm 0.005\,\text{mm} \]

with a confidence level of approximately 95%

We are now in a position to decide whether the bar should be accepted or rejected.

**Discussion**

No formal mathematical model was derived. The model used was that the measurand was the average measured diameter plus the micrometer correction. While it is not necessary to write equations for such a simple case, it would be dangerous to assume that this approach can be used for all testing situations.

There was a systematic listing of all the sources of uncertainty and some were not included in the calculation as prior work or current consideration had shown this to be unnecessary.

The sensitivity coefficients were obtained by logical analysis. While they were not all unity, their application was trivial. The values of each of the three uncertainty components were converted to mm, the units of the measurand, so all the sensitivity coefficients were then 1.

Whether the bar is accepted or rejected depends on the specified tolerance.
Expression of uncertainty in test reports

Usually endorsed test reports providing numerical test results must also give a statement of uncertainty of measurement. To prevent rounding errors it is recommended that all calculations are made with several more significant figures than will be retained in the final report. As a rough rule of thumb this should be two digits more than is anticipated will be reported. For some work it may be necessary to use all available significant figures in calculations, for example where the difference of two large numbers is involved. Indeed, with modern application programs, such as spreadsheets, all figures are carried through and the user may determine how many are to be visible. Rounding of the value to be reported should be left until the end.

The uncertainty value should consist of one or two significant figures only because the nature of its estimation does not support more precision. It will therefore be necessary to round off the calculated uncertainty. To avoid giving an optimistic figure it is recommended that the uncertainty always be rounded up. Thus an uncertainty calculated as 0.0185% may be written as 0.019% or 0.02% depending on the degree of resolution considered necessary.

It is recommended that the reported value consist of only truly significant figures. That is, the measured value should not be reported to a resolution greater than the uncertainty will sustain. A resolution of about 10% of the uncertainty would generally be appropriate. Thus, for an item measured to a resolution of 0.001% and with a measurement uncertainty of 0.1%, the reported value would normally be rounded to 0.01%.

(In some work where systematic errors are known to be dominant it may be useful to give one extra figure to aid detection of drifts and changes in the test system.)

It is recommended that the test results be reported in a format similar to the following example.

<table>
<thead>
<tr>
<th>Measured value:</th>
<th>100.01 kPa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uncertainty of measurement:</td>
<td>± 0.1 kPa</td>
</tr>
<tr>
<td>Confidence Level:</td>
<td>95%</td>
</tr>
<tr>
<td>Coverage factor:</td>
<td>2.5</td>
</tr>
</tbody>
</table>

Although it has been customary for some laboratories to omit the “±” sign from the uncertainty statement, this practice is discouraged. The uncertainty statement may be followed by a statement similar to: “Uncertainties given in this report have been estimated on the basis of there being only five chances in one hundred that the reported value differs from the true value by more than the stated uncertainty”.

This of course applies at the time of measurement and so the uncertainty of measurement statement should not include a component for projected drift. In some applications the user may require a value that will apply for the period corresponding to the duration of the calibration interval. A supplementary statement would then be required. This is covered briefly in appendix x.

If a 99% confidence interval were used then the statement would refer to “... one chance in one hundred...”. This will be seen on many old reports.

It may also be helpful to indicate whether the uncertainty has been derived by using the principles described by the ISO GUM or by some other method. It should be noted that the ISO GUM offers the only generic method accepted internationally.
Converting from one confidence level to another

Instruments and standards calibrated in the past may have the measurement uncertainty expressed in terms of a 99% confidence level instead of the 95% confidence level commonly in use now, thus necessitating a different conversion to be made before the certificate can be properly used. That is, a different $k$ factor is required to convert the expanded uncertainty to the combined standard uncertainty.

If the uncertainty is a good approximation to a normal distribution, then dividing the 99% confidence uncertainty by 2.6 will yield the combined standard uncertainty. This is equivalent to dividing the 99% confidence interval by 1.3 to get a 95% confidence interval. Although it will rarely be practical to ascertain the goodness of approximation to a normal distribution, there is usually no other practical approach than to make this assumption.

Thus:

$$u_c(y) = \frac{u_{99}}{2.6}$$

In those cases where the uncertainty estimate was conservative, using a divisor of 3 is more appropriate.

Other confidence intervals may be converted back to a combined uncertainty by dividing by an appropriate coverage factor. In the absence of other information, Student’s $t$ tables may be used to find the coverage factor.

There are unfortunately two further complications. Firstly, the preceding argument relates only to conversion of uncertainties whose distribution approximates a normal distribution. If the uncertainty contains a dominant systematic component that has essentially a rectangular distribution, then division by 2.6 (for a claimed 99% confidence interval) will underestimate the combined uncertainty limits.

Secondly, some laboratories still use algebraic addition of estimated absolute limits of the error components, thus arriving at a figure which more closely represents a 100% confidence level rather than the nominal 95% confidence level claimed. A divisor of 3 is more appropriate here.

In the absence of detailed information of the composition of the uncertainty, the application of the “divide by $k$ rule”, with $k$ having values between 2 and 3, is the only practical approach.
Acknowledgements

The kind permission of Standards Australia to reproduce some definitions in appendix i is acknowledged.

During several years of presenting courses on the topic of uncertainty I have been taught much by the students and, in particular, by my co-presenter Walter Giardini, who lifted the veil on correlation and other aspects of uncertainty analysis.

The following have also made significant contributions: J R Miles, R G Rigby, M Purss, and J L Gardiner.
References

- AS 1514, 1980, *Glossary of terms used in metrology, Part 1 - General terms and definitions*

- AS 2706, 1984, *Numeric values - Rounding and interpretation of limiting values*

- AS 2833, 1985, *Metrology - Symbols for expressing uncertainty of measurement*


- *General Requirements for Registration*, National Association of Testing Authorities, Australia, 1992


- Hooper, M R (1986), *Properties of Laboratory Data*, Good Laboratory Management Seminar, AOQC (SA) and NATA, Adelaide, May 1996

- ISO 14253-1:1998, *Geometrical Product Specifications (GPS) - Inspection by measurement of workpieces and measuring equipment - Part 1: Decision rules for proving conformance or non-conformance with specifications*


- Miles, J R (1980), *An Introduction to Errors and Uncertainty Calculation*, CSIRO, Division of Applied Physics Melbourne Branch, Technical Memorandum M 10,

- Somlo, P I and Hunter, J D (1985), *Microwave Impedance Measurement*, Peter Peregrinus


Definitions and formulas

Definitions

The following are abbreviated versions of the definitions given in the ISO GUM. They are intended to quickly give the reader a grasp of the essential element(s) of the definitions. Once this has been understood, the reader should refer to the complete text of the definitions in Appendix B of the ISO GUM.

Cross-references to the ISO GUM are given in the square brackets.

- quantity (measurable quantity) [B.2.1]
  Attribute of a phenomenon, body or substance that may be distinguished qualitatively and determined quantitatively.
  e.g. particular quantity: length of a piece of string.

- value (of a quantity) [B.2.2]
  Magnitude of a particular quantity generally expressed as a unit of measurement multiplied by a number.
  e.g. length of a piece of string: 153 mm.

- true value (of a quantity) [B.2.3]
  Value consistent with the definition of a given particular quantity.
  Note: The true value cannot be determined by measurement as all measurements have uncertainties. Further, the definition of any measurand is imperfect, therefore the true value is a hypothetical quantity.

- conventional true value (of a quantity) [B.2.4]
  Value attributed to a particular quantity and accepted, sometimes by convention, as having an uncertainty appropriate for a given purpose.
  Note: This may be a value obtained from a number of measurements taken to establish a conventional true value.

- measurement [B.2.5]
  Set of operations having the objective of determining a value of a quantity.

- measurand [B.2.9]
  Particular quantity subject to measurement.
  e.g. diameter of a particular rod under conditions of standardised temperature and pressure.

- influence quantity [B.2.10]
  Quantity that is not the measurand, but that affects the result of the measurement.
  e.g. temperature of a micrometer used to measure a length.

- result of a measurement [B.2.11]
  Value attributed to a measurand, obtained by measurement.
  Note: The value should be accompanied by additional information, including its uncertainty.

- uncorrected result [B.2.12]
  Result of a measurement before correction for systematic error.

- corrected result [B.2.13]
  Result of a measurement after correction for systematic error.

- accuracy of measurement [B.2.14]
  Closeness of agreement between the result of a measurement and a true value of the measurand.
  Note: This is a qualitative term and is not the same as precision.

- repeatability (of results of a measurement) [B.2.15]
  Closeness of the agreement between the results of successive measurements of the same measurand carried out under the same conditions of measurement.
  Note: These conditions must be specified (e.g. time over which tests are made).

- reproducibility (of results of a measurement) [B.2.16]
  Closeness of agreement between the results of measurements of the same measurand carried out under changed conditions of measurement.
  Note: These changed conditions must be specified.

- uncertainty (of measurement) [B.2.18]
  A parameter, associated with the result of a measurement, that characterises the dispersion of the values that could reasonably be attributed to the measurand.

- error (of measurement) [B.2.19]
  Result of a measurement minus the true value of the measurand.

- random error [B.2.21]
  Result of a measurement minus the mean that would result from an infinite number of measurements of the same measurand carried out under repeatable conditions.

- systematic error [B.2.22]
  Mean that would result from an infinite number of measurements of the same measurand carried out under repeatable conditions minus a true value of the measurand.

- correction [B.2.23]
  Value added algebraically to the uncorrected result of a measurement to compensate for systematic error.
  Note: It has the same size, but opposite sign, to the systematic error.
Assessment of Uncertainties of Measurement

- **probability** [C.2.1]
  A number in the range 0 to 1 attached to a random event.

  Note: A number close to 1 indicates a high degree of belief that an event will occur, while a value near 0 indicates a belief that the occurrence of the event is unlikely. A probability of 0.5 means a belief that the event is equally likely to occur or not occur. An example of probability = 0.5 is the probability that a coin tossed into the air will land as “tails”.

- **confidence interval** [C.2.27, 28]
  A range within which a value is believed to lie with a stated probability or level of confidence. The level of confidence may be expressed as a percentage in the range 0 to 100% and is equal to the probability multiplied by 100.

- **confidence level** [C.2.29]
  The level of confidence associated with an interval within which a value is expected to lie. It is expressed as a percentage and is given by 100 multiplied by the probability that the value will fall within the specified interval. For uncertainty estimates a confidence level of 95% is commonly used.

**Common formulas**

**Type A evaluation**

If the number of measurements is \( n \), and \( x_i \) is the \( i \)th measurement (and \( x \) is the mean) then:

- **Mean**
  \[
  x = \frac{1}{n} \sum_{i=1}^{n} x_i
  \]

- **Variance**
  \[
  \text{var} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - x)^2
  \]

- **Standard deviation**
  \[
  s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - x)^2}
  \]

- **Experimental Standard Deviation of the Mean**
  \[
  ESDM = \frac{s}{\sqrt{n}}
  \]

- **Degrees of freedom for ESDM**
  \[
  \nu = n - 1
  \]

- **Standard uncertainty, type A evaluation**
  \[
  u = ESDM
  \]

**Type B evaluation**

Degrees of freedom, from relative uncertainty

\[
\nu \approx \frac{1}{2} \left( \frac{\Delta u(x_i)}{u(x_i)} \right)^2
\]

if:

\[
R = \frac{\Delta u(x_i)}{u(x_i)} \times 100\%
\]

then:

\[
\nu \approx \frac{1}{2} \left( \frac{100}{R} \right)^2
\]

**Rectangular distribution**

If the semi-range is \( a \), then the standard uncertainty, \( u \), is given by:

\[
u = \frac{a}{\sqrt{3}}
\]

The degrees of freedom (\( \nu \)) for a rectangular distribution are infinite if the semi-range represents absolute limits.

- **Sensitivity coefficient, \( c_i \)**

  If \( y \) is a function of \( x \), then:

  \[
  c_i = \frac{\partial y}{\partial x_i}
  \]

- **Combined standard uncertainty, \( u_c(y) \)**

  \[
  u_c(y) = \sqrt{\sum_{i=1}^{n} [c_i u(x_i)]^2}
  \]

- **Effective degrees of freedom, \( \nu_{eff} \)**

  \[
  \nu_{eff} = \frac{u_c^4(y)}{\sum_{i=1}^{n} u_i^4(y) / \nu_i}
  \]

- **Coverage factor, \( k \)**

  \[
  k = \text{Student’s } t\text{-factor},
  \]

- **Expanded uncertainty, \( U \)**

  \[
  U = k u_c(y)
  \]
Type A evaluations of uncertainty components

A type A evaluation is an evaluation of a standard uncertainty by statistical calculations. This is only possible when a series of observations such as repeated measurements are feasible. The most common reason for taking repeated measurements is to minimise the effect of random errors or noise. This section looks more closely at random errors.

Although different sorts of measurements may each produce a normal distribution curve, some will give broad low curves and some will give narrow peaked curves as is illustrated in Fig A2.3. Thus, apart from the mean, another parameter is required to describe different distributions. This parameter is called the standard deviation and is given the symbol \( \sigma \). The standard deviation is a measure of the spread of individual values around the mean.

If we refer back to Fig A2.2, and imagine a line drawn vertically through the point representing the mean, \( \mu \), then clearly half the area under the graph lies to the left of this line and half to the right. Obviously half the measured values are less than the mean and half are greater. Further, if we randomly select any measured value or take one new measured value, the probability of it exceeding the mean is 0.5 or 50%, and the probability that it will be less than the mean is also 50%. For an infinite number of readings, the probability of a reading being exactly the mean value is infinitesimally small. If we were now to draw vertical lines on this graph passing through the \( \pm \sigma \) points, then the area under the graph between these lines is 0.683. This means that 68.3% of the readings lie in the range \( \mu \pm \sigma \). If the lines were drawn at \( \pm 2 \sigma \) then the area enclosed would be 0.954, and for lines at \( \pm 3 \sigma \) the area enclosed is 0.997. To obtain an area of 0.95 we would draw the lines at \( \pm 1.96 \sigma \).

We can now say that, with a probability of 95%, any value will lie within the range \( \mu \pm 1.96 \sigma \). The interval spanning this range is called the 95% confidence interval and the range limits are called the 95% confidence limits.

The values of \( \mu \) and \( \sigma \) can only be found from an infinite number of measurements. Clearly we must make do with an estimate of these parameters based on a realistic number of measurements.

Analysis of random errors has shown that they have the characteristics of a random process with a tendency for values near the mean to occur more frequently. Figure A2.1 shows a bar graph, called a histogram, in which the frequency of occurrence of 50 readings from a measuring instrument is plotted for a number of equal intervals. As more and more readings are taken and smaller intervals are used, and, if the source of variation is truly random, then the histogram will take on a more clearly defined smooth bell shaped curve. Such a curve represents what is called the normal or Gaussian distribution. This is illustrated in Fig A2.2. While a plot of a very large set of measurements with random errors will not exhibit infinite length tails, the approximation to a normal distribution curve is a good one.

The mean value is given the symbol \( \mu \). For comparative purposes it is usual to normalise, or scale the curve so that the area between the curve and the horizontal axis is one unit.
The scatter of the means obtained from a series of repeated measurements will be found to have its own normal distribution. The standard deviation of this distribution is called the **Standard Error of the Mean (SEOM)** and this is calculated from:

\[
SEOM = \frac{\sigma}{\sqrt{n}} 
\]

The SEOM applies to any set of \(n\) measurements from an infinite population and is a measure of the spread of the distribution of the means in the same way that \(\sigma\) is a measure of the spread of infinite set of statistics from which the means are obtained. This equation assumes knowledge of \(\sigma\) and so it is the number of samples, \(n\), that is important, not the number of repeated sets of measurements.

Unfortunately, we do not know the value of \(\sigma\), but we do have an estimate of its value, \(s\). We can therefore calculate a term that is applicable to our experimentally determined mean. This is called the **Experimental Standard Deviation of the Mean (ESDM)**. The formula is:

\[
ESDM = \frac{s}{\sqrt{n}} 
\]

If in doubt, some test calculations should be done. Indeed, all software should be trialed using data for which the calculations have been independently done.

The variance produces inconvenient units. In order to obtain the same units as the population, the square root of the population variance is taken. This is given the symbol \(\sigma\) and is applicable only when the sample size, \(n\), is very large. When \(n\) is not very large and the equation A2.2 is used to find an estimate of the variance, we can calculate an estimate of the standard deviation, that is denoted by \(s\).

If we perform several repeated series of measurements on the same item using the same equipment under the same conditions and each set of measurements consists of \(n\) readings, even when \(n\) is large we will obtain slightly different values for each mean and different variances for each set. This is to be expected as \(x\) and \(s\) are estimates of \(\mu\) and \(\sigma\) that apply to the parent population of an infinite number of measurements.
We now calculate the mean, standard deviation and standard uncertainty arising from the variation in repeated readings.

Mean:

\[
\bar{x} = \frac{\sum x_i}{n}
\]

\[
= 999.527 \text{ kΩ}
\]

and the estimate of the standard deviation, \(s\), thus:

\[
s = 0.0279 \text{ kΩ}
\]

and:

\[
ESDM = \frac{s}{\sqrt{n}}
\]

\[
= 0.00883 \text{ kΩ}
\]

The standard uncertainty, \(u\), is therefore also 0.00883 kΩ.

This would be combined with other terms as will be shown later.

Sometimes it is necessary to know the number of degrees of freedom, \(\nu\). For a set of \(n\) measurements for which we obtain a mean the formula is:

\[
\nu = n - 1
\]

There are many methods of determining standard uncertainty by a type A evaluation, however the \(ESDM\) is the most common calculation. The next most common type A evaluation is determination of standard uncertainties from fitted curves.

Again, as with means and standard deviations, there are standard formulas but it is rarely necessary to use them as modern calculators and spreadsheets provide these calculations literally at a touch of a button. The number of degrees of freedom is the number of measurements minus the number of constants fitted by the regression calculations.

Thus, for a straight line of the form:

\[
y = a + bx
\]

there are two constants, \(a, b\), so then \(\nu\) is given by \(n-2\).

The standard uncertainty of a fitted curve, with \(y = f(x)\) may be taken as the square root of the variance of \(y\). The in-built programs will usually also give the variances of all constants calculated by the regression calculation (except for the zeroth order coefficient and the number of degrees of freedom).
appendix iii

Student’s $t$ distribution

If we have calculated a mean value of a set of readings we have a better estimate of the true mean $\mu$ than we would have with a single reading. The sample standard deviation $s$ is a measure of the dispersion of the population and the ESDM is a measure of the dispersion of sample means. It can be shown that the population formed by the ESDM’s is a normal distribution when $n$ and the number of means are sufficiently large. Even for small $n$ the assumption of a normal distribution is reasonable.

It would be useful if another parameter could be found which measured the difference between the mean of the sample, $x$, and the population mean $\mu$. Because the mean is taken from a distribution, the difference can only be estimated on a statistical basis. This means that we need to decide how sure we want to be to have the difference covered by the new parameter. Suppose this parameter were determined to have a 68% probability of being large enough to cover the difference between the sample mean and the population mean. If we later wanted to have a 95% probability of encompassing the difference, then we would need a larger value for the parameter.

We call the range given by the parameter a confidence interval (CI).

A brewer named Gosset solved the problem of determining the correct confidence interval for a set with a finite number of samples. He published a paper on a suitable distribution under the pseudonym “Student” and this distribution has become known as Student’s $t$ distribution. The factor $t$ is a variable coefficient which, when multiplied by the ESDM, defines the confidence interval for the mean of a finite set of samples or measurements. Tables A3.1 and A3.2 contain values of $t$.

To apply this distribution we need to determine the number of degrees of freedom for our set of measurements. When only a mean of a set of numbers is involved the number of degrees of freedom is one less than the number of measurements, that is $(n - 1)$.

The confidence interval, $U$, is given by:

$$U = \frac{ts}{\sqrt{n}}$$  	......A3.1

Application of Student’s $t$

By following the process previously described we can find the individual uncertainty components and combine them. The combined standard uncertainty is equivalent to the standard deviation of the measured value and has a confidence interval of 68%. We would normally like to have a larger confidence interval.

For most purposes, a 5% chance of the true mean being beyond the range is acceptable, and indeed this is the level of risk used by the National Standards Commission and National Measurement Laboratory. This is a 95% confidence level and it is recommended that this be adopted by all laboratories as standard practice.

If the combined standard uncertainty were normally distributed then we would multiply it by 1.96 to obtain a 95% CI.

In practice, the standard uncertainty often applies to a measurement based on a small or relatively small number of repeated measurements. In some cases it applies to a single measurement. The important indicator is the number of degrees of freedom, $v$. A large number of degrees of freedom implies a good knowledge of the normal distribution associated with the uncertainty and a small number indicates that we have less confidence in the parameters.

Student’s $t$ tables (A3.1 and A3.2) show that once we have selected the confidence level required we need to know the number of degrees of freedom to obtain a value of $t$. This is the only reason for requiring $v$.

As stated elsewhere, we prefer to use an expanded uncertainty, $U$, rather than the combined standard uncertainty, $uc(y)$. We multiply the combined uncertainty by a coverage factor, $k$, to obtain $U$, thus $U = ku(y)$. In some applications, it is sufficient to select a value of $k$ equal to 2. If we want to be as precise as possible then the value of $k$ is equal to the value of $t$ obtained for the selected CI and the effective number of degrees of freedom of the combined standard uncertainty.
Example

Taking the previous example of ten measurements of a resistor we have:

\[ ESDM = 0.008 \, 83 \, k\Omega \]

There were 10 measurements, so:

\[ \nu = 10 - 1 \]
\[ \nu = 9. \]

We look up the Student’s \( t \) table and for a 95% CI, we find that \( t = 2.26 \). This is our coverage factor, \( k \).

Therefore, if there were no other components of uncertainty, then the combined uncertainty would be \( u_c \).

The expanded uncertainty \( U \) is given by:

\[ U = ku_c(y) \]
\[ = 2.26 \times 0.008 \, 83 \, k\Omega \]
\[ = 0.019 \, 96 \, k\Omega . \]

We would round this back, giving \( U_{95} = \pm 0.020 \, k\Omega \).

It is instructive to consider the effect of taking only the first two readings. Then we would obtain:

Mean:

\[ \bar{x} = 999.535 \, k\Omega \]

and:

\[ s = 0.035 \, 4 \, k\Omega \]

The ESDM then is \( \frac{0.0354}{\sqrt{2}} = 0.025 \, 0 \, k\Omega \)

From Table A3.1:

For degrees of freedom = 1, we see \( t = 12.7 \)

The expanded uncertainty, \( U = ku_c(y) \)
\[ = 12.7 \times 0.025 \, 0 \]
\[ = 0.317 \, k\Omega \]

The expanded uncertainty is now 0.32 k\Omega

It is obvious that taking only two readings in the above example yields a result which is inadequate compared to that of which the system is capable. A type B assessment may be more appropriate if only two measurements are made.

---

<table>
<thead>
<tr>
<th>Degrees of Freedom (( \nu ))</th>
<th>( t )</th>
<th>Degrees of Freedom (( \nu ))</th>
<th>( t )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12.7</td>
<td>18</td>
<td>2.10</td>
</tr>
<tr>
<td>2</td>
<td>4.30</td>
<td>19</td>
<td>2.09</td>
</tr>
<tr>
<td>3</td>
<td>3.18</td>
<td>20</td>
<td>2.09</td>
</tr>
<tr>
<td>4</td>
<td>2.78</td>
<td>25</td>
<td>2.06</td>
</tr>
<tr>
<td>5</td>
<td>2.57</td>
<td>30</td>
<td>2.04</td>
</tr>
<tr>
<td>6</td>
<td>2.45</td>
<td>35</td>
<td>2.03</td>
</tr>
<tr>
<td>7</td>
<td>2.37</td>
<td>40</td>
<td>2.02</td>
</tr>
<tr>
<td>8</td>
<td>2.31</td>
<td>45</td>
<td>2.02</td>
</tr>
<tr>
<td>9</td>
<td>2.26</td>
<td>50</td>
<td>2.01</td>
</tr>
<tr>
<td>10</td>
<td>2.23</td>
<td>60</td>
<td>2.00</td>
</tr>
<tr>
<td>11</td>
<td>2.20</td>
<td>70</td>
<td>2.00</td>
</tr>
<tr>
<td>12</td>
<td>2.18</td>
<td>80</td>
<td>1.99</td>
</tr>
<tr>
<td>13</td>
<td>2.16</td>
<td>90</td>
<td>1.99</td>
</tr>
<tr>
<td>14</td>
<td>2.15</td>
<td>100</td>
<td>1.98</td>
</tr>
<tr>
<td>15</td>
<td>2.13</td>
<td>110</td>
<td>1.98</td>
</tr>
<tr>
<td>16</td>
<td>2.12</td>
<td>120</td>
<td>1.98</td>
</tr>
<tr>
<td>17</td>
<td>2.11</td>
<td>( \infty )</td>
<td>1.96</td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th>Degrees of Freedom (( \nu ))</th>
<th>( t )</th>
<th>Degrees of Freedom (( \nu ))</th>
<th>( t )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>63.7</td>
<td>18</td>
<td>2.88</td>
</tr>
<tr>
<td>2</td>
<td>9.93</td>
<td>19</td>
<td>2.86</td>
</tr>
<tr>
<td>3</td>
<td>5.84</td>
<td>20</td>
<td>2.85</td>
</tr>
<tr>
<td>4</td>
<td>4.60</td>
<td>25</td>
<td>2.79</td>
</tr>
<tr>
<td>5</td>
<td>4.03</td>
<td>30</td>
<td>2.75</td>
</tr>
<tr>
<td>6</td>
<td>3.71</td>
<td>35</td>
<td>2.72</td>
</tr>
<tr>
<td>7</td>
<td>3.50</td>
<td>40</td>
<td>2.70</td>
</tr>
<tr>
<td>8</td>
<td>3.36</td>
<td>45</td>
<td>2.69</td>
</tr>
<tr>
<td>9</td>
<td>3.25</td>
<td>50</td>
<td>2.68</td>
</tr>
<tr>
<td>10</td>
<td>3.17</td>
<td>60</td>
<td>2.66</td>
</tr>
<tr>
<td>11</td>
<td>3.11</td>
<td>70</td>
<td>2.65</td>
</tr>
<tr>
<td>12</td>
<td>3.06</td>
<td>80</td>
<td>2.64</td>
</tr>
<tr>
<td>13</td>
<td>3.01</td>
<td>90</td>
<td>2.63</td>
</tr>
<tr>
<td>14</td>
<td>2.98</td>
<td>100</td>
<td>2.63</td>
</tr>
<tr>
<td>15</td>
<td>2.95</td>
<td>110</td>
<td>2.62</td>
</tr>
<tr>
<td>16</td>
<td>2.92</td>
<td>120</td>
<td>2.62</td>
</tr>
<tr>
<td>17</td>
<td>2.90</td>
<td>( \infty )</td>
<td>2.58</td>
</tr>
</tbody>
</table>
Type B evaluation of uncertainty components

Type B determination of uncertainties is determination of uncertainties by means other than by established statistical methods. It involves using data provided by others, such as uncertainties on a calibration certificate or from the manufacturer’s specifications, reference tables and books. It involves sound scientific and engineering calculations to determine worst case limits. It involves estimation of limits by suitably experienced people.

In many instances, type B evaluation involves estimation of systematic errors. This type of error will be considered in more detail here.

More on systematic errors

One important type B analysis is that of taking an expanded uncertainty from a calibration certificate and calculating a standard uncertainty. We divide the expanded uncertainty by the coverage factor. That is:

\[ u(x) = \frac{U}{k} \quad \ldots A4.1 \]

For many systematic errors the magnitude of the error is not known, only the limits within which it lies. The next step involves making a decision about the distribution that describes the dispersion of the errors.

Rectangular distribution

It may often be assumed that the errors have equal probability of having a value anywhere in that range and, unlike random errors, are not more likely to be near the mean value than the extremes. This is a rectangular distribution, which is illustrated in Fig A4.1. The maximum and minimum limits are assumed to be equal and often are denoted by \( a \), which is called the semi-range.

Such a distribution is capable of mathematical analysis and it can be shown (Dietrich) that the standard deviation and hence the standard uncertainty, \( u(x) \), is given by:

\[ u(x) = \frac{a}{\sqrt{3}} \quad \ldots A4.2 \]

where \( a \) is the semi-range

While it is usual to make the semi-range symmetric about the mean and while this will often occur naturally there will be instances where there is a considerable lack of symmetry. In such cases it is recommended that the semi-range be made symmetrical and equal to the larger value. Sometimes a change of units, such as dBm (a logarithmic scale) to mW (a linear scale), will yield a symmetric distribution.

For example, consider a test where the ambient temperature readings ranged from 17 °C to 21 °C but there was an average temperature of 20 °C. We now have a range of (-3, +1)°C. If we take the semi-range as 3 °C and we have a temperature coefficient of 20 ppm/°C then we have 60 ppm as the estimate for \( a \). This is conservative, but a more rigorous approach is usually not warranted.

Type B evaluations are applicable to all types of distributions. The rectangular distribution may be considered the distribution of minimal knowledge. This is because all we know are the limits and, in the absence of further knowledge, assume that any value is equally probable.

The assumption of rectangular distributions when the individual probability density curves are not known can be justified on the basis that the estimates of the variances will tend to be pessimistic, the labour involved is reduced and the final result will, for practical purposes, be an adequate estimate.

Normal distribution

In addition to the rectangular distribution, the next most common distribution encountered in type B analysis is the normal distribution. We already know its characteristics (see appendix ii), so will not dwell long on it. It should be noted that if we know the uncertainty component to be normally distributed and can estimate approximate 100% limits, ±\( a \), then the uncertainty can be estimated to be given by:

\[ u(x) = \frac{a}{3} \quad \ldots A4.3 \]

This is based on 99.7% of the area under the curve being within ± three times the standard deviation.

Triangular distribution

Another commonly used distribution is the triangular distribution. The limits may be estimated in the same
way as for a rectangular distribution. It is used when there is evidence that the values near the mean are the most probable and, as the limits are approached, the probability decreases to zero. It represents less knowledge than in the normal distribution case, but more knowledge than the rectangular distribution case. For a triangular distribution, the uncertainty is calculated by:

\[ u(x_i) = \frac{a}{\sqrt{6}} \]  
......A4.4)

**The U distribution**

This distribution occurs in several areas of metrology. An example is the distribution for uncertainties arising from radio frequency connector reflections. It may also be applicable to air temperature variations where the temperature control produces regular temperature excursions between limits. If the temperature variation produced in a test object has exponential variations rather than linear ramp variations, then more time is spent near the extremes than at the average temperature, resulting in a U-shaped distribution.

The uncertainty is calculated from:

\[ u(x_i) = \frac{a}{\sqrt{2}} \]  
......A4.5)

\( u(x_i) \) is the standard uncertainty and
\( a \) is the semi-range of the limits of the uncertainty component

For other distributions the reader is referred to texts on statistics for methods of determining the standard deviation. The need for this should be very, very rare.

**Instabilities**

Some instruments and standards exhibit apparent good short-term stability, but are subject to occasional step changes. Electronic voltage references are one type of standard which exhibit such behaviour. Experienced metrologists are alert to such possibilities and will assess a component, in addition to the instabilities observed at the time of test, based on their personal knowledge, manufacturers specifications and any previous measurements with the particular, or similar, device. This component may be much larger than might be calculated from the test results alone for the first calibration, particularly when little data is available on stability.

Previously, the recommendation was that the component of uncertainty due to instability should be algebraically added to the combined systematic components. The ISO GUM states that an effort should be made to treat this uncertainty component in the same way as any other component.
**Combination of distributions**

The Central Limit Theorem states that when a number of distributions are combined, the resultant distribution will approximate a normal distribution. This applies regardless of the shape of each distribution. The more distributions are combined the better the approximation. This allows the use of Student’s $t$ as the coverage factor for the expanded uncertainty. So, although we may have only rectangular distributions for our uncertainty components, the combined standard uncertainty will have a distribution that approximates a normal distribution and we can apply the same rules that apply to normal distributions.

It is useful to see what happens when different distributions are combined. Wagner describes how the combination (or convolution) of two rectangular distributions with similar variances combine to form a triangular distribution and shows that if three such rectangular distributions are combined the result is bell-shaped and approximates to a normal distribution. Dietrich gives a complete derivation of these results in his book.

Given the degree of accuracy associated with the estimation of systematic uncertainties there is no practical difference between the convolution of three rectangular distributions and a normal distribution. The probability of a value falling in the range $-2\sigma$ to $+2\sigma$ is 95.8% for the combination of two identical rectangular distributions as compared to 95.4% for a normal distribution. Indeed the difference between a triangular distribution and a normal distribution in terms of area under the curve between any specified number of standard deviations is small. Thus, once the variances of several similar rectangular distributions have been combined, such a combination may be treated as if it were a normal distribution. So to obtain the 95% confidence limits the square root of the variance is multiplied by the same factor as would be used for a normal distribution, namely 1.96.

If the variances are not close to the same magnitude, then it will require more than three to be combined to get the same degree of approximation to a normal distribution. In an extreme case, one component may be so dominant that the combination remains essentially rectangular. This may occur if one component has a variance that exceeds about three times that of any other. In that case the lack of approximation to a normal distribution results in an overestimation of up to 15% in the confidence interval when the usual procedure is applied.

<table>
<thead>
<tr>
<th>DISTRIBUTION TYPE</th>
<th>STANDARD DEVIATION</th>
<th>$\pm \sigma$</th>
<th>$\pm 1.5\sigma$</th>
<th>$\pm 1.73\sigma$</th>
<th>$\pm 2\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rectangular</td>
<td>$\frac{a}{\sqrt{3}}$</td>
<td>57.7</td>
<td>86.6</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Combination of two equal rectangular</td>
<td>$\frac{2a}{\sqrt{6}}$</td>
<td>65.0</td>
<td>85.0</td>
<td>91.4</td>
<td>96.6</td>
</tr>
<tr>
<td>Combination of three equal rectangular</td>
<td>$a$</td>
<td>66.7</td>
<td>85.9</td>
<td>91.5</td>
<td>95.8</td>
</tr>
<tr>
<td>Normal</td>
<td>$\sigma$</td>
<td>68.3</td>
<td>86.6</td>
<td>91.7</td>
<td>95.4</td>
</tr>
</tbody>
</table>

Adapted from Wagner. The semi-range is taken to be equal to $a$. 

---

Assessment of Uncertainties of Measurement
Corrections

Often, past practice has been that small corrections were not applied but were incorporated in the uncertainty statement. The assumption was that the correction could be replaced by a rectangular distribution with a semi-range equal to the magnitude of the correction. This practice is not recommended.

Firstly, corrections do not have a rectangular distribution. They have a discrete value and sign. If they are negligible then they should be neglected. If not, they should be applied. The uncertainty of the corrections must be considered for inclusion in the assessment of the measurement uncertainty. It can be shown that, in some circumstances, the uncertainty will not be appreciably increased by including small corrections, however the measured value may then be in error by an amount approaching the uncertainty due to this sloppy procedure alone.

For example, consider a measurement where we have neglected four small corrections estimated to be +0.02%. Assume that the combined standard uncertainty based on all other terms is 0.3%. If we do a RSS combination of the corrections we get a new combined uncertainty of 0.300 9%, not significantly different from the initial value. However, the combination of the four corrections is 0.08%. If our expanded uncertainty is 0.6%, then we have an unaccounted bias of around 13% in our measurand, something we would rather not knowingly introduce.
If two or more systematic errors are correlated, that is they are not independent, then the RSS combination is not appropriate. It is recommended that the correlated uncertainty components be added algebraically and this value then combined with the other components in the normal way. This is a simplified approach that may sometimes overestimate the uncertainty. It is justified on the basis that, in many cases where there are correlated components of uncertainty, they are in fact completely correlated. In more complex cases, the statistically rigorous calculations should be made.

For example, consider the calibration of a 2 kg mass. We start with our 1 kg standard, calibrated externally, and calibrate two working standard 1 kg masses. The two 1 kg masses are then used to make a one to one comparison with the 2 kg mass. Suppose the 1 kg standard had a systematic error of -20 µg. This error would be transferred to both working standards when they are calibrated. They would both have an error of -20 µg. The error in their combined mass is, of course, -40 µg. We can say that the error due to the reference is correlated in the two working standards. That is, if the error in the reference were to change, the associated errors in the working standards would change in sympathy.

Similarly, the uncertainty in the 1 kg standard should be doubled when estimating the uncertainty in the combination of the two 1 kg working standards.

There are analogies in other fields of measurement where consideration of the measurement system leads to a determination that some correlation exists. How do we know if apparently independent parameters are correlated? A simple test for correlation is to graph the two parameters against measurement number. If they can be seen to change together, say both increase for the second measurement, both stay the same for the third and both reduce for the fourth, then there is clearly correlation between them.

The ISO GUM equation 13, reproduced below, shows that the effect of correlation is to add a second term to the basic equation 10.

\[
u_i^2(y) = \sum_{i=1}^{n} \left[ \frac{\partial f}{\partial x_i} \right]^2 u_i^2(x_i)\]

ISO GUM (10)

\[
u_i^2(y) = \sum_{i=1}^{n} \left[ \frac{\partial f}{\partial x_i} \right]^2 u_i^2(x_i) + 2 \sum_{i=1}^{n} \sum_{j=i+1}^{n} \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} \nu(x_i, x_j)\]

ISO GUM (13)

This term may be negative or positive, so correlation may reduce the uncertainty. This may be appreciated by considering the compensation of pendulum clocks. The mass at the end of the pendulum arm can be made so that a retaining washer at the end of the arm holds it in place. As the temperature rises, the pendulum grows longer. The mass also grows in diameter, and its centre of gravity is thus lifted relative to the end of the pendulum. By having a low coefficient of expansion for the arm, the extension of the arm can be made the same as the movement of the centre of the mass. So the effective length of the pendulum is maintained constant regardless of temperature changes. There is negative correlation between the motion of the centre of the mass with respect to the motion of the end of the arm.

Covariances

A covariance is a type of variance that depends on the variances of each correlated term and their correlation coefficient. Covariances can be calculated using standard formulas or determined experimentally. The estimated covariance is \(u(q, r)\), where \(q_i\) and \(r_i\) are the estimates (or measured values) of the input parameters. Refer to Annex C.3.4 of the ISO GUM for more information on the formula.

If \(q_i\) and \(r_i\) are independent pairs of simultaneous observations, the estimate of the covariance, \(u(q_i, r_i)\) is:

\[
u(q_i, r_i) = \frac{1}{n(n-1)} \sum_{i=1}^{n} (q_i - \bar{q})(r_i - \bar{r})\]

Note the similarity between this equation and that for the variance of a distribution.

If we want the correlation between the means \(\bar{q}\), \(\bar{r}\) of two independent pairs of simultaneous observations \(q_i\), \(r_i\), then the equation is:

\[
u(q, r) = \frac{1}{n(n-1)} \sum_{i=1}^{n} (q_i - \bar{q})(r_i - \bar{r})\]

Note the similarity between this equation and that for the ESDM.

Correlation coefficient

The degree of correlation is given by the correlation coefficient, \(r\). Parameters with very high correlation have correlation coefficients approaching +1 or -1, while those that are completely uncorrelated have a correlation coefficient of 0.

The degree of correlation is given by the estimated correlation coefficient \(r(x_i, x_j)\):

\[
r(x_i, x_j) = \frac{u(x_i, x_j)}{u(x_i)u(x_j)}\]
where:

\[ u(x_i) \text{ and } u(x_j) \]

are the standard uncertainties of \( x_i \) and \( x_j \) respectively.

Examination of ISO GUM equation (13) shows that the last part of the additional term is the product of the standard uncertainties and the correlation coefficient. Hence, when there is no correlation, the correlation coefficient is zero so the additional term vanishes. When the correlation coefficient is +1 then the uncertainty is a maximum. When the correlation coefficient is negative the uncertainty actually reduces to a minimum when the correlation coefficient is -1.

Considerable effort can be saved if we neglect correlation when it is very weak and assume complete correlation when there is significant correlation. The difference in the estimate of the uncertainty values as compared to a rigorous calculation is usually small. Either condition can be determined from experience or consideration of the model of the measurement.

In the case of complete correlation, the worst condition, ISO GUM equation (13) can be simplified. It translates to combining the two components by simply adding their standard uncertainties to arrive at a combined standard uncertainty.

Thus:

\[ u^2_e(y) = \left[ \sum_{i=1}^{N} \left( \frac{\partial f}{\partial x_i} u(x_i) \right)^2 \right] \]

or

\[ u_e(y) = \sum_{i=1}^{N} \frac{\partial f}{\partial x_i} u(x_i) = c_1 u_1 + c_2 u_2 + c_3 u_3 + \ldots + c_N u_N \]

where the symbols have their usual meaning.

Returning to our mass example, suppose we wish to calibrate a weighbridge for a 20 tonne load. Suppose that we have a calibrated 1 tonne mass and twenty 1 tonne working masses. We can calibrate each of the working masses by comparison against the 1 tonne calibrated mass and then place these twenty masses on the platform of the weighbridge.

The uncertainty of the value of each mass in the set of twenty has a component arising from the uncertainty in the standard. Because all the masses are calibrated against the one standard, there is correlation between this part of the uncertainty of each mass. When the components are combined to find the uncertainty of the 20 tonne mass, the correlation must be taken into account.

Suppose that the uncertainty in the calibrated one tonne mass is ±0.075 kg at a 95% confidence interval with a coverage factor of 2.3.

The standard uncertainty is therefore:

\[ u_0 = \frac{0.075}{2.3} \]

\[ = 0.032 61 \text{ kg} \]

The uncertainty arising from the standard, for the value of the 20 tonne mass of the combined set, is found by applying the formula:

\[ u_e(y) = \sum_{i=1}^{N} c_i u(x_i) \]

\[ = 20 \times 0.032 61 \]

\[ = 0.652 2 \text{ kg} \]

This component can now be combined in the usual way with other components such as the standard deviation of the scale readings to arrive at the combined uncertainty.

The ISO GUM also gives an example, H.2, where correlated components occur and shows how the calculations should be done. The example H.3 shows how modifying the model eliminates correlation.
Higher order terms

The ISO GUM states that the combined standard uncertainty is the positive square root of the combined variance. This is given in mathematical form as equation 10 which is:

$$u^2(y) = \sum _{1}^{n} \left[ \frac{\partial f}{\partial x_i} \right] ^2 u^2(x_i)$$

ISO GUM (10)

This is derived from a Taylor series and neglects the higher order terms. If the next highest order term is considered then the uncertainty is increased significantly if the second order effects cannot be neglected. The additional term is:

$$\sum _{i=1}^{n} \sum _{j=1}^{n} \left[ \frac{1}{2} \left( \frac{\partial^2 f}{\partial x_i \partial x_j} \right)^2 + \frac{\partial f}{\partial x_i} \frac{\partial^3 f}{\partial x_j \partial x_i \partial x_j} \right] u^2(x_i) u^2(x_j)$$

Those with well-developed mathematical skills will be able to follow the method given in the ISO GUM. Example H.1 illustrates a typical calculation. It is instructive nevertheless to consider a simpler representation.

Graphical illustration of higher order terms

Consider the case of an area determined by measuring lengths $x$ and $y$.

The area $A$, is therefore $xy$

Let there be increments in $x$ and $y$, $\delta x$, $\delta y$ respectively.

We can draw a diagram that illustrates this point:

![Diagram of area measurement](image)

Fig A8.1 Illustration of the measurand, $A=xy$, the uncertainties $\delta x$, $\delta y$ and the higher order uncertainty $\delta x \delta y$

The incremental area is:

$$A = yx + xy + \delta x \delta y$$
Assessment of compliance with specifications

Often it is appropriate to include in a report a statement of compliance or otherwise with a nominated specification. For those instances where there is a simple go/no-go test the statement can be made readily, but when the compliance depends on measured values falling within a specified range then some care is required in assessment. It is necessary to determine the “region of uncertainty” before making statements about compliance when the specification defines numerical limits. As shown in Fig. A9.1 these regions are the bands centred on the specification limits with a width equal to the total measurement uncertainty. Measured results, which fall between the nominal value and the region of uncertainty, constitute compliant results and those beyond the outer regions constitute non-compliant results. When a measured value falls sufficiently close to the specification limit, such that the uncertainty band overlaps the limit, then it is not possible to be confident about compliance or non-compliance. In such circumstances the endorsed report must state the measured value and the test uncertainty. It may be helpful to the client to draw attention to those measurements for which it is not possible to state compliance or non-compliance, but no statement of opinion as to whether the item is fit for a particular purpose should appear in the report.

Fig. A9.1 Measured values falling within the region of uncertainty do not conclusively support the conformity or lack of conformity with a specification.

In Fig. A9.1, small circles numbered 1 to 9 represent nine measurements. The centres of the circles are placed on the values of the measurements. Measurements 1, 2 and 7 have values which fall not only between the high and low specification limits, but also within the space between the zones of uncertainty. The length of the bars attached to each measured value represents the uncertainty of measurement. Measurements 1, 2 and 7 represent uncontestable passes.

Measurements 3 and 6 represent uncontestable fails. The other measurements (4, 5, 8 and 9) have values that fall within the zones of uncertainty and cannot uncontestably be given as evidence of either a pass or a fail. They are inconclusive results.

To limit the number of inconclusive results, test equipment should be chosen such that the zone of uncertainty is small compared to the specification limits.

Clearly, the measurement uncertainty must be determined to establish how large the zone of uncertainty is.

Certainly it is worthwhile giving some attention to the anticipated measurement uncertainty before performing tests, so that the number of results that fall in the region of uncertainty are minimised. The traditional rule of thumb employed is to use a measurement system capable of measuring with an uncertainty of 1/10 of the specification limit. More recently, a ratio of 1 to 4 has been used. This ratio is usually called the Test Uncertainty Ratio (TUR). Its principle use has been in providing a rationale for selection of test equipment without undertaking a complete analysis of the measurement system.

A major objection to the use of a TUR is that it is too simplistic and does not consider the effects of influence quantities on the instrument or the measurand.

Further information is contained in:
Determination of calibration intervals and drift

Drift

Drift is a form of instability and is the determining factor for calibration intervals. In well-designed instruments and standards drift is monotonic. That is, the rate of change is uniform with respect to time or, especially with newly manufactured standards, it is decreasing in an exponential manner. For example, commercially supplied standard resistors typically drift more in the first two years than in any subsequent five-year period. Those resistors subject to a more comprehensive “ageing” process have lower initial drift rates, but the costs involved are significant and are reflected in the purchase price.

Calibration reports usually give only an uncertainty of measurement; some may include a stability statement. The stability statement combines both the uncertainty of the calibration and an allowance for the variation in the assessed drift. Most testing laboratories would not be required to attempt this. The methodology is similar to that involved in assessing uncertainties of measurement.

Some devices, such as those using fixed turns ratio transformers, would be expected to have no drift yet they do exhibit some instabilities which may be considered to be drift. A decision must be made as to whether the allowance for these changes should be incorporated in the uncertainty of measurement or in a stability statement if one is given. When the test extends over a period great enough for drift to affect the results an allowance must be made, but as a general rule it is recommended that allowances for drift be kept separate from the uncertainties of test. If an allowance for drift has been incorporated in the uncertainty statement then this should be made clear in the measurement report or calibration certificate.

The crucial step is determination of the drift between successive calibrations and a control chart, as shown in Fig. A10.1, is recommended for this purpose. Indeed, it is good practice to record the history of all the basic standards of the laboratory in such a manner. It becomes obvious that, even when the drift is monotonic, and that is not always the case, the prediction of the region of future values has its own uncertainty attached. Further, the greater the period of time a projection is made into the future the greater the chance that the prediction will be wrong. The nature of the instrument and its usage also determine its stability.

The most important source of information, apart from the results of two, three or more calibrations, is the manufacturer’s data.

For instruments for which no prior history, manufacturer’s data or experience is available, drift can only be determined after at least three sets of measurements spaced over a time frame which should be at least 20% of the anticipated recalibration interval. Two measurements give an average drift, but at least three are required to show whether the drift rate is increasing, steady or decreasing. Even with apparently identical devices, quite different drift characteristics may be obtained. Some devices will also exhibit a degree of randomness in their drift and others will exhibit periods of uniform drift interspersed with step changes. For these reasons, initial estimates of drift should be conservative.

Recalibration interval

Once the anticipated drift has been established, it is simply a matter of deciding how long it will take before the uncertainty in the latest calibration certificate has expanded to the maximum acceptable for the testing being done.

The main premise is one of fair usage and freedom from breakdown. Immediately after recalibration the results should be compared with the previous results. If the secular change produces a drift greater than the desired “in-use” uncertainty then the calibration interval should be shortened. While it is preferable to use non-adjustable standards, it is recognised that an increasing number of laboratories will be relying on electronic instruments which are adjusted, either by the use of trimmers or by
software, so as to minimise the reading error at certain cardinal points. It is essential that the errors prior to adjustment be recorded to determine the drift that has occurred.

It is also necessary to establish whether the uncertainty of measurement attributed to an instrument has already incorporated an allowance for secular change. Most modern digital instruments have three levels of accuracy statements to cover different periods of time after adjustment. Many other instruments and standards will not have such a statement and it will be necessary to allow for secular change when determining the measurement uncertainty. This point should not be overlooked. As an example, an electronic standard of voltage may be measured with an uncertainty of ±0.3 µV, but may drift 1 µV/year. Therefore, if calibrated annually, it may be relied upon to ±0.3 µV immediately after calibration and to ±1.3 µV at the end of that year.

Electronic instruments are now more reliable, but a 12 month calibration interval is generally appropriate. Simpler, well established devices, such as standard cells and mercury-in-glass thermometers, can have recalibration intervals ranging from two to ten years.

As a general rule, the recalibration interval must be less than the time for the item to drift more than the maximum uncertainty that it is desired to attribute to that item. For example, consider a standard which is given an uncertainty of, say, ±1 ppm at the time of measurement and has an estimated drift rate of less than 1.8 ppm/year. If an uncertainty of ±10 ppm is acceptable then it will take nearly six years to drift that far. However, taking the uncertainty of the calibration into account, a maximum recalibration interval of five years is obtained.
Worked examples

The two examples that follow illustrate two different approaches. The first involves some laboratory measurements with the main calculations being made in a table that follows the format that might be used in a spreadsheet. An explanation of the derivation of each component precedes the table.

The second is in free text format and no actual measurements are involved.

Both these examples and those loose leaf ones that are provided as supplements to this book should be regarded as illustrative and not universally definitive.

Example 1
Calculation of the uncertainty of measurement of a voltage standard

In this example we have taken a set of measured values from a laboratory notebook and using other available data calculated the uncertainty of the measurand, the value of a voltage standard. This example is intended to be illustrative rather than prescriptive and some simplifying assumptions have been made. In other voltage standard measurements some of the uncertainty components used here may be quite negligible, while some shown as small or even omitted here may become significant. During the working a resolution of 1 nV or 3 significant figures are shown although in the calculations more digits were used.

The measurement

A 1 volt zener diode based voltage standard was calibrated by comparison to a 1 volt calibrated voltage standard. The measurement procedure involved measuring the voltage difference between the two standards with a digital multimeter (DMM). The DMM was used on a 10 millivolt range with 10 nanovolt resolution. Eight measurements were made over several days and the values corrected to a value at 20.00 °C. These results are given in Table 1. The measurand was the corrected mean of these values. Care was taken to minimise errors arising from any thermal emf that may have been present. The ambient temperature in the laboratory was maintained within ± 1.0 °C of 21.0 °C. A calibrated thermometer was used to measure the temperature adjacent to the standard being calibrated. The internal temperature of the reference under test was measured by an internal thermistor element. The calibrated standard was in an enclosure that maintained its temperature to within ± 0.001 °C.

The measured values are given in Table 1. The known temperature coefficient of the standard being calibrated has been applied to obtain values for an ambient of 20.00 °C and these are given in Table 1.

<table>
<thead>
<tr>
<th>DATE</th>
<th>8/7</th>
<th>5/7</th>
<th>9/7</th>
<th>12/7</th>
<th>16/7</th>
<th>18/7</th>
<th>20/7</th>
<th>23/7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meas'd diff. [µV]</td>
<td>546.52</td>
<td>546.00</td>
<td>543.30</td>
<td>544.71</td>
<td>542.50</td>
<td>550.57</td>
<td>546.35</td>
<td>544.81</td>
</tr>
<tr>
<td>Temp [°C]</td>
<td>20.40</td>
<td>20.90</td>
<td>21.10</td>
<td>21.00</td>
<td>21.10</td>
<td>20.05</td>
<td>20.50</td>
<td>20.95</td>
</tr>
<tr>
<td>Corr'n [µV]</td>
<td>2.04</td>
<td>4.59</td>
<td>5.61</td>
<td>5.10</td>
<td>5.61</td>
<td>0.26</td>
<td>2.55</td>
<td>4.85</td>
</tr>
<tr>
<td>Value [µV] at 20 °C</td>
<td>548.56</td>
<td>550.59</td>
<td>548.91</td>
<td>549.81</td>
<td>548.11</td>
<td>549.11</td>
<td>550.82</td>
<td>548.90</td>
</tr>
</tbody>
</table>

Table 1: Measured voltage differences and temperature corrected voltage differences.

The temperature coefficient for the standard under test, taken from measurements reported by the manufacturer, is - 5.10 µV/°C. The uncertainty in this value is stated to be 0.25 µV/°C. The mean temperature corrected voltage difference is 549.42 µV.

The model

The model of the measurement is represented by the following equation.

\[ V_R = V_S + V_{DMM} + C_{DMM} + C_T \]

where:

- \( V_R \) is the voltage of the standard being calibrated after correction to 20.00 °C and
- \( V_S \) is the value of the calibrated standard and
- \( V_{DMM} \) is the reading of the DMM (the uncorrected voltage difference between the standard under test and the calibrated standard) and
- \( C_{DMM} \) is the correction for the DMM and
- \( C_T \) is the correction to the reference under test from the test temperature to 20.00 °C.

The measurand is the corrected mean of all measured values of \( V_R \).
We can see by inspection of the model that if any term on the right hand side of the equation varied by say 1µV, then the measured value varies by 1 µV. That is the sensitivity coefficients will be 1 if we use the same units for all terms. For convenience we will work in microvolts.

Uncertainty components and their evaluation

Terms associated with \( V_s \), the calibrated standard

Standard calibration

The first term in the model is the voltage of the calibrated standard. We can find the value of the voltage and its uncertainty from the latest calibration certificate. In this example we find that the report gives a value of 1.000 075 volts with an expanded uncertainty (95% confidence level) of 1.5 µV. The coverage factor is stated to be 2.1.

The standard uncertainty is then:

\[
    u_s = \frac{U}{k} = \frac{1.5}{2.1} = 0.714 \ \mu V
\]

Using the Student’s \( t \) table we find that the degrees of freedom is 18 for \( k = 2.1 \).

Terms associated with \( V_{DMM} \), the DMM reading

Repeatability

We can now calculate the measurand and perform a Type A analysis on the uncertainty component arising from random errors in the measured values.

The mean \( \bar{x} = 549.42 \ \mu V \).

The experimental standard deviation, \( s \), is 0.965 µV

The standard uncertainty in the mean is the ESDM.

\[
    ESDM = \frac{s}{\sqrt{n}} = \frac{0.965}{\sqrt{8}} = 0.341 \ \mu V
\]

Thus \( u_{DMM1} = 0.341 \ \mu V \)

The associated degrees of freedom is \( (n-1) = 7 \).

Resolution

We also have the resolution of the DMM to consider. This is 10 nV or 0.01 µV.

If we take the semi range of the associated error to be 0.005 µV and follow convention and assign a rectangular distribution, we get a standard uncertainty of

\[
    u_{DMM2} = 0.005/\sqrt{3} \ \mu V = 0.0029 \ \mu V
\]

This is clearly much smaller than the other components to date and although we could neglect it we will include it for the sake of the exercise. By convention the number of degrees of freedom for this term is infinite and we can use any large number in our calculations.

Nanovoltmeter loading

There was a loading error due to the input resistance of the DMM. The input resistance is 10 MΩ, and both references have 1000 Ω source resistances. Taking the source voltage equal to the average voltage difference between the two references, the loading error would be:

\[
    \text{Error} = -549 \frac{2000}{(10^7 + 2000)} = -0.110 \ \mu V
\]

This gives a correction of 0.110 µV to be added to the measured value. The variations in actual voltage and the deviations from nominal resistance are estimated to give an uncertainty of no more than 10% in this value. Taking a worst-case view, we will assume a rectangular distribution with a semi-range = 0.0110 µV, giving a standard uncertainty, \( u_{DMM3} \) of:

\[
    u_{DMM3} = \frac{0.011}{\sqrt{3}} = 0.0064 \ \mu V.
\]

We estimate that the relative uncertainty is 20%. We calculate the approximate degrees of freedom as follows:

\[
    \nu = \frac{1}{2} \left( \frac{100}{R} \right)^2
\]

\[
    \nu = \frac{1}{2} \left( \frac{100}{20} \right)^2 = 12.5
\]

Terms associated with \( C_{DMM} \), the DMM correction.

Correction

The correction for the DMM can be obtained from the calibration certificate. In this example we will assume the certificate states the correction to be 0.0 µV and the uncertainty in this value has been given as ±(0.05% of reading + 5 least significant digits) with a coverage factor of 2. Taking the nominal reading as 549 µV, we get:
Expanded uncertainty
\[
= \left( \frac{0.05}{100} \right) \times 549 + 5 \times 0.01 \ \mu V
\]
\[
= 0.325 \ \mu V
\]
As this is an expanded uncertainty with a coverage factor of 2.0, then the standard uncertainty is:
\[
u_{\text{DMM}} = 0.325/2 \ \mu V
\]
\[
= 0.163 \ \mu V
\]
The associated degrees of freedom for \( t = 2.0 \) is about 60.

Zero and thermal errors
The DMM zero or offset error effect was minimised by using the auto-zero function of the DMM. There are unavoidable thermal emfs that are generated externally to the DMM. The effect of these can also be minimised by taking a reading and then reversing the connection to the DMM and taking a 'reverse' reading. Both techniques were applied in this calibration. Any residual offset in the DMM has been included in the uncertainty of the correction for the DMM. There remains some uncompensated thermal emfs to account for.

Uncompensated thermal emfs
The uncompensated thermal emfs have been estimated from prior tests to be no more than 0.25 \( \mu V \). As this is also a worse case estimate a rectangular distribution will be assigned, giving a standard uncertainty, \( u_{\text{DMM5}} \).
\[
u_{\text{DMM5}} = 0.25/\sqrt{3}
\]
\[
= 0.144 \ \mu V
\]
The relative uncertainty of this is estimated to be 10%. We calculate the approximate degrees of freedom as:
\[
\nu = \left( \frac{100}{2} \right)^2
\]
\[
= 100/10
\]
\[
= 50
\]
Terms associated with \( \zeta \), temperature correction to 20°C.

Temperature correction, temperature coefficient.
The model includes a term for the correction of measured values to the standard temperature of 20.0 °C. The temperature coefficient for the standard under test, taken from measurements reported by the manufacturer, is - 5.10 \( \mu V/\degree C \). The uncertainty in this value is taken from the manufacturers data and is 0.25 \( \mu V/\degree C \). To estimate the effect on the average value we make a worst-case calculation.

Maximum difference from 20.00 = 1.1 °C

Worst case correction error = 1.1 x 0.25 \( \mu V \)
\[
= 0.275 \ \mu V
\]
We will take this to be the semi range of a rectangular distribution.

The standard uncertainty for the temperature correction is then:
\[
u_{\text{Ct}} = 0.275/\sqrt{3} \ \mu V
\]
\[
= 0.159 \ \mu V
\]
The relative uncertainty is estimated to be 10%, thus degrees of freedom is approximately,
\[
\nu = \left( \frac{100}{2} \right)^2
\]
\[
= \frac{100}{10}
\]
\[
= 50
\]
Temperature correction, thermometer correction
The thermometer correction is given in a calibration certificate. It also gives the expanded uncertainty of the correction as 0.02 °C with a coverage factor of 2.0. This gives an additional uncertainty in the correction for the temperature correction for the standard being calibrated.

Worst error = 0.02 x 5.1 \( \mu V \)
\[
= 0.110 \ \mu V
\]
As the temperature correction uncertainty comes from a calibration certificate with a coverage factor of 2, we have a standard uncertainty of:
\[
u_{\text{Ct}} = 0.110/2
\]

The degrees of freedom are approximately 60.

Rounding of reported value
The reported value is to be rounded to the nearest tenth of a microvolt. This introduces an uncertainty that in all respects is a resolution uncertainty with a semirange of 0.05 \( \mu V \). Thus the rounding uncertainty is:
\[
u_{\text{round}} = 0.05/\sqrt{3} \ \mu V
\]
\[
= 0.028 \ 87 \ \mu V
\]
The number of degrees of freedom is infinite and may be represented by a large number.
Calculating the Combined Standard Uncertainty and the Expanded Uncertainty

There are no other terms to be considered, so the combined and expanded uncertainties can now be calculated. Table 2 shows these calculations. The standard uncertainties are squared, added together and the square root of this sum found. This is the combined standard uncertainty. To get the associated effective degrees of freedom requires the application of the Welch-Satterthwaite equation. This has the combined standard uncertainty raised to the fourth power as the numerator and the denominator is the sum of the individual standard uncertainties to the fourth power divided by their degrees of freedom. The last column of the table has the individual standard uncertainties to the fourth power divided by the degrees of freedom and the sum of this column. This sum is divided into the combined standard uncertainty raised to the fourth power to give the effective number of degrees of freedom. Student’s t table (or the TINV EXCEL© spreadsheet function) is then used to find the value of the coverage factor \( k \), and hence the expanded uncertainty.

### TABLE 2 Uncertainty Calculations

We could now round the expanded uncertainty and report it as ± 1.7 µV.

The reported value.

We calculate the measured value as per the model equation. That is the value of the standard, plus the temperature corrected difference, plus other corrections. Except for the loading correction of 0.11 µV, the other corrections are zero.

Thus:

\[
\text{Standard's voltage} = 1.000\ 075 + 0.000\ 549\ 42 + 0.000\ 000\ 11\ \text{V} = 1.000\ 624\ 53\ \text{V}
\]

The report would state the value of the standard as 1.000 625 V at 20 °C with an uncertainty of ±1.7 µV at a 95% confidence level and a coverage factor of 2.04.

**NOTES.**

1. Although it was neglected in this example, another term that should be considered is the drift of the standard. At least two and preferable three or more calibrations are required to find a drift rate. The drift rate uncertainty will need to be estimated, based on the uncertainties of the measured values and the probable drift rate. It will be necessary to study the device in question and obtain information on similar devices in order to have some confidence about the estimate.

2. Higher order terms are insignificant and have been neglected.

3. In this model, any drift in the unit being calibrated has been neglected. The reported value may be taken as applying to the median day of the tests for estimation of longer term drift.
Explanation of the Calculations in the Table.

For reasons of speed, accuracy and convenience, a spreadsheet has been used for the uncertainty calculations in Table 2. This table is actually a copy of the spreadsheet. Even calculations using a calculator are made easier if a table is laid out in a similar way to Table 2. An explanation of the contents and calculation process follows.

Each horizontal row contains calculations for the components named in the first (left-most) vertical column.

The second column contains the units to be used, as a reminder that if they are not the same as the measurand then the sensitivity coefficient certainly may not be unity.

The third column contains the name of the distribution chosen for each component. This assists in selecting the expanded uncertainty, \( U \), standard uncertainty, \( u(x) \), or semi-range, \( a \), to be entered in the fourth column.

The fifth column contains the divisor for the values in the fourth column. The value depends on the distribution type assigned and whether the value in the third column is an expanded uncertainty, a standard uncertainty or a semi-range.

The sixth column contains the number of degrees of freedom for each uncertainty component, obtained by calculation or estimate.

The seventh column contains the first of the calculations, namely the value in column four divided by the number in column five. These values are all standard uncertainties of the input components, \( u(x) \).

The eighth column contains the sensitivity coefficients, \( c_i \). Due to prior manipulation of the data in this example the values are all 1.

Column nine contains uncertainties of the measurand. That is the input uncertainties, \( u(x) \), have been multiplied by the sensitivity coefficients to transform them to measurand uncertainties, \( u(y) \).

In order to evaluate the combined uncertainty we need to sum the squares of \( u(y) \). The squares of \( u(y) \) appear in column ten and their sum at the bottom of this column. Immediately below this sum is the square root of this sum which is the standard combined uncertainty, \( u(y) \).

To find the effective degrees of freedom we use the Welch-Satterthwaite equation. The numerator is the standard uncertainty raised to the fourth power. The denominator is the sum of the numbers in the eleventh column. The result of the division, \( v_{\text{eff}} \) is placed under the box or cell containing the standard uncertainty.

The Student’s \( t \) table or the appropriate spreadsheet function can be used to find the coverage factor, \( k \), which is placed in the next cell down. A 95% confidence level was used for the selection.

Finally the expanded uncertainty, \( U \), is found by multiplying together the contents of the cells for \( k \) and \( u(y) \).
Example 2
Uncertainty of calibration of an external micrometer
(Provided by J R Miles)

The following is a sample calculation for the uncertainty of measurement associated with calibrating an external micrometer. Additional explanatory text is included.

Important:
Rigorous uncertainty calculations estimate the number of degrees of freedom for each uncertainty component and then calculate the total number of degrees of freedom using the Welch-Satterthwaite equation. The total number of degrees of freedom is determined so that the coverage factor $k$ may be calculated. The calculations below do not include those for degrees of freedom. Instead a nominal coverage factor of $k = 2$ is used for a 95% confidence interval.

Reference Documents:
- AS 2102, 1989 - Micrometer callipers for external measurement
- AS 1457, 1989 - Gauge blocks and rectangular length bars and their accessories

Measurand:
The deviation of traverse of the micrometer screw for an external micrometer at 20 °C. The deviation of traverse is defined as: the maximum difference between the ordinates of the curve for the deviation of the readings obtained along the complete traverse of the screw (AS 2102, Section 3.3).

The graph shows an example of the results obtained when testing the traverse of a 0-25 mm micrometer using gauge blocks. The sizes of the gauge blocks used are shown on the graph and are chosen to test the micrometer screw both at complete turns and at intermediate positions of the micrometer spindle. The deviation of traverse here would be 0.6 - (-0.4) = 1.0 µm.

Mathematical model
The deviation of traverse (DT) of the spindle at 20 °C is as follows:

$$DT_{(1)} = E_{\text{maximum}} - E_{\text{minimum}} \quad \ldots A6.1$$

$$E = I - T \quad \ldots A6.2$$

where $E$ is the micrometer error, $I$ is the micrometer reading and $T$ is the true value of gauge block length. It follows that:

$$u^2(DT) = u^2(E_{\text{maximum}}) + u^2(E_{\text{minimum}}) = 2 u^2(E) \quad \ldots A6.3$$

where:

$$u^2(E) = u^2(I) + u^2(T) + 3u^2(L_i) \quad \ldots A6.4$$

and the $L_i$ are correction terms that are assumed to be zero but do not have zero uncertainty (see following).

Standard uncertainty of micrometer reading, $u(I)$
The standard uncertainty associated with the reading of the micrometer has several components:

Resolution
Assume the smallest scale interval on the micrometer is 10 µm and that the use of the vernier allows subdivision of this scale interval to one tenth of this scale interval. The associated distribution is taken to be rectangular, so the standard uncertainty associated with the micrometer resolution is $1/\sqrt{3}$ µm.

Repeatability/random effects
Random influences such as small variations in the measuring force, thermal fluctuations, etc. will generate a standard uncertainty in the micrometer reading. One way to determine this standard uncertainty would be to take many repeat readings at each point of the micrometer measuring range and calculate the associated standard deviation. This is a type A evaluation. A type B evaluation consists of estimating the rectangular bounds on random effects based on experience.

Both methods are valid. A type B estimate is used for this example. It is estimated that the scatter of repeat readings made using this type of micrometer always lies within the bounds of ±1 µm. This is a rectangular distribution with an associated standard uncertainty of $1/\sqrt{3}$ µm.

Note: Although a type B estimate is made of the uncertainty, it is still important to take several repeat readings at each point of the micrometer range to check that the measurement system is behaving properly. These readings should not be used to determine the uncertainty as they are usually insufficient in number.
Combining the two contributions to the standard uncertainty in the micrometer reading gives:

\[ u^2(I) = (1/\sqrt{3})^2 + (1/\sqrt{3})^2 = 2/3 \, \mu m^2 \]

\[ u(I) = 0.816 \, \mu m \]

**Standard uncertainty of length of gauge block, \( u(T) \)**

Assume that the report on the gauge block set states that the gauge blocks satisfy the requirements for Accuracy Class 4 gauge blocks. No actual departures are given for each gauge block, and even if they are, they are not normally applied for a micrometer calibration.

Let the measuring length be 25 mm. The tolerance on a 25 mm Accuracy Class 4 gauge block is ±0.3 μm. The standard (AS 1457-1989) states that the permitted difference between the measured gauge length and the nominal length cannot exceed ±0.3 μm by more than the maximum permissible measurement uncertainty, which for a 25 mm Accuracy Class 4 block is ±0.24 μm. All we know is that the length of the gauge block could be anywhere within the range ±0.54 μm. The distribution is therefore rectangular with a standard uncertainty of \( u(T) = 0.54/\sqrt{3} \, \mu m \).

For micrometers that require the wringing of gauge blocks to give longer reference lengths, an estimate of the standard uncertainty associated with the wringing would have to be made. The compression of the gauge block length due to the forces exerted by the micrometer anvils is considered negligible.

The assumptions discussed above are equivalent to assuming that \( \delta \theta = 0 \) and \( \delta \alpha = 0 \). The value for \( u_2 (L_{thermal}) \) may now be estimated as follows:

\[ u(\delta \alpha) \]

Assume both micrometer and gauge block are made of steel and that the bounds on the expansion coefficient of steel are ±1 x 10⁻⁶ °C⁻¹. Therefore, the standard uncertainty for each expansion coefficient is 1 x 10⁻⁶ /√3 °C⁻¹ and the standard uncertainty for the difference between the two coefficients is \( u(\delta \alpha) = \sqrt{2} \times 1 \times 10^{-6} /\sqrt{3} \, °C^{-1} \).

\[ u(\delta \theta) \]

Let the gauge block and the micrometer “soak” for a time sufficient to ensure that the difference between the two is estimated to be zero within the range ±0.2 °C. Handling techniques and fixturing allow this uncertainty to be maintained during calibration. This is a rectangular distribution, so that \( u(\delta \theta) = 0.2/\sqrt{3} \, °C \).

It follows from equation (A6.5) that the value for \( u^2 (L_{thermal}) \) for a 25mm gauge block is given by:

\[ u^2 (L_{thermal}) = (25^2)(1)^2 (\sqrt{2} \times 1 \times 10^{-6} /\sqrt{3})^2 + (25)^2 (11.5 \times 10^{-6})^2 (2/3)^2 + (25)^2 (0.2)^2 (\sqrt{2} \times 1 \times 10^{-6}/\sqrt{3})^2 \]

\[ = (0.66+1.76 + 0.027) \times 6.25 \times 10^{-10} \, mm^2 \]

\[ = 15.31 \times 10^{-10} \, mm^2 \]

\[ \text{or} \]

\[ u (L_{thermal}) = 0.039 \, \mu m. \]

For a 25mm gauge block with good soaking, the thermal contribution is relatively small. However, checking the zero error of a 100-125 mm micrometer using a 100mm gauge block in a laboratory at 25 °C and with poor handling techniques so that \( \delta \theta \) is estimated to be in the range ±2 °C, giving \( u(\delta \theta) = 2/\sqrt{3} \, °C \), the result is quite different:

\[ u^2 (L_{thermal}) = (100^2)(5)^2 (\sqrt{2} \times 1 \times 10^{-6}/\sqrt{3})^2 + (100)^2 (11.5 \times 10^{-6})^2 (2/3)^2 + (100)^2 (0.2)^2 (\sqrt{2} \times 1 \times 10^{-6}/\sqrt{3})^2 \]

\[ = (16.67 + 176.33 + 0.027) \times 10^{-8} \, mm^2 \]

\[ = 193.0 \times 10^{-8} \, mm^2 \]
or \( u(L_{\text{thermal}}) = 1.39 \mu m \). This value is comparable with the tolerance on zero setting for a 100-125 mm micrometer, namely \( \pm 4 \mu m \).

**Geometric effect**

The geometric effect \( L_{\text{geometric}} \) arises from the parallelism and flatness of the micrometer anvils. This effect is estimated to be zero, but there is an associated uncertainty with this estimate. The permitted flatness and parallelism of micrometer anvils are 1 \( \mu m \) and 2 \( \mu m \) respectively. The bounds on \( L_{\text{geometric}} = 0 \) are estimated to be \( \pm 0.25 \mu m \), giving \( u(L_{\text{geometric}}) = 0.25/\sqrt{3} \mu m \).

**Total standard uncertainty**

Combining the values determined above using Equation (A6.4) for a 25mm gauge block gives:

\[
\begin{align*}
u^2(E) &= 2/3 + (0.54/\sqrt{3})^2 + (0.039)^2 + (0.25/\sqrt{3})^2 \\
&= 0.786 \mu m^2
\end{align*}
\]

The dominant contribution comes from \( u(I) \), the uncertainty in reading the micrometer.

From equation (A6.3) it follows that:

\[
\begin{align*}
u^2(DT) &= 2 \times 0.786 = 1.571, \\
giving \ u(DT) &= 1.25 \mu m
\end{align*}
\]

The value of \( k \) for a 95% confidence interval is 2 so that the expanded uncertainty is:

\[
\begin{align*}
u(DT) &= 2.00 \times 1.25 = \pm 2.50 \mu m \\
at the 95\% confidence level
\end{align*}
\]

This is the uncertainty for the deviation of the traverse for a calibration of a 0-25 mm micrometer. The tolerance for all external micrometers on the deviation of traverse is 3 \( \mu m \).

Note that if the uncertainty in the measured error at a given point in the traverse is required, then:

\[
u^2(\text{Error}) = 0.786, \ giving \ u(\text{Error}) = 0.887 \mu m.
\]

The value of \( k \) for a 95% confidence interval is again 2.00 so that the expanded uncertainty is:

\[
\begin{align*}
u(DT) &= 2.00 \times 0.887 = \pm 1.77 \mu m \\
at the 95\% confidence level.
\end{align*}
\]

The uncertainty of measurement for calibration of the micrometer is \( \pm 1.8 \mu m \) for an approximate 95% confidence level.