UNCERTAINTY ANALYSIS FOR LABORATORY ACCREDITATION

Peter B. Crisp

Fluke Precision Measurement Ltd,
52 Hurricane Way, Norwich, UK

Introduction

NAMAS Accredited laboratories are required to follow recommended methods for the estimation of uncertainties in electrical measurements. These are well documented in the NAMAS Publication M3003[1] and are consistent with the ISO/TAG4 "Guide to the Expression of Uncertainty in Measurement"[2]. This paper examines the techniques used by Fluke's UK, Test and Measurement Division for the estimation of uncertainties in support of the company's NAMAS Accredited standards and calibration facilities. The paper discusses the identification and evaluation of the uncertainty contributions and in particular, the principle of Audit Via Traceability - a mandatory requirement for accredited laboratories with capabilities better than 1ppm for DC voltage.

Accreditation Requirements

The NAMAS Accreditation Standard M10[3] requires accredited laboratories to use an acceptable method for the analysis of uncertainties. The method adopted by most of the UK laboratories is described in the NAMAS Publication M3003. This very informative document is intended primarily for electrical calibration laboratories and takes the reader through the uncertainty estimation process step by step. This document, although somewhat different to the rather weighty ISO/TAG4 document, effectively achieves the same results but with rather less statistical jargon. The NAMAS document offers several worked examples and offers sound advice that has been well tried by over 300 accredited laboratories.

Definition of Uncertainty Types

Uncertainty contributions can have different characteristics that affect the way that they should be treated. Experience has shown that uncertainty contributions can be allocated to one of two characteristic "pots" (A and B) for which a particular method can used to evaluate and combine them in order to arrive at a final estimated value.

Type A contributions may be conveniently defined as those that may be evaluated by statistical methods and are attributable to random effects. They are typically evaluated by repeating a measurement several times and calculating the arithmetic mean (of the sample) and standard deviation. The mean is reported as the measured value and the standard deviation is used as an uncertainty contribution.

Put simply, type B contributions are anything else that is not obviously type A. They are specifically evaluated by other means and may be attributable to systematic effects. An important consideration in the treatment of type B uncertainties is that all known sources of error should be corrected rather than considered as an uncertainty contribution. It is only the residual unknown contributions that are estimated in the evaluation. Such contributions may only be revealed by varying the method of measurement or by using different equipment or operators.

Uncertainty Contributions

Any estimation of uncertainties must start by identifying all significant contributions. There may be several sources of error in a measurement where the magnitude of the error could be quantified. Where these can be identified, they should be corrected such that only the residual unknown component contributes to the uncertainty of the measurement. Using the importation of DC voltage from a higher level laboratory as an example, there will be the following uncertainty components:

- Calibration Uncertainty
- Transportation
- Stability with Time
- Stability with Temperature
- Noise

Usually the measurement techniques used will ensure well defined conditions and minimize loading effects such that the main contributions listed above will be the only significant ones.

Calibration Uncertainty

Calibration uncertainty is a significant contribution and is usually reported on the certificate of calibration issued by another organization i.e. the national laboratory. Quite reasonably, the national laboratory is beyond the control of other commercial organizations and there is little that can be done by the "customer" to evaluate the uncertainty reported as a single ± value on the certificate. For this reason the calibration uncertainty is usually treated as a type B contribution. The reported uncertainty may vary slightly for each calibration and will usually be at a 95% minimum confidence level.

Transportation

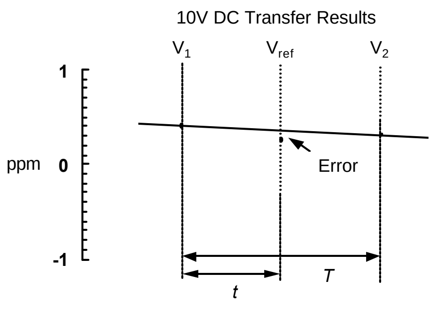
The transportation of the standard to and from another laboratory can introduce additional errors. These can be quantified (and corrected for) if knowledge of the behavior of the standard under shipment conditions is acquired. Several things can affect the importation of the Volt from another laboratory. The standard will drift by a small amount in the time it takes to transfer between laboratories. The environmental conditions may be different at the two locations. Loss of power can also affect the accuracy of the transportable standard.

The safest way to quantify changes during shipment is the "loop closure" technique. This requires the availability of second, or ideally, a group of other standards to compare the transfer standard with before and after shipment. This technique is essential in order to obtain the best possible uncertainties where transportable standards are used. The Fluke reference standards laboratory Volt is based on the mean of thirteen independent references and an additional group standard of eight references. Three different Zener technologies are involved to improve the detection of systematic behavior of any one particular reference type in response to external influences. Residual errors due to transportation (after all corrections have been applied) are considered to be type B uncertainty contributions.

Audit Via Traceability (AVT)

This is a loop closure technique that has several advantages and involves three sets of measurements. V_1 and V_2 are made by the laboratory before and after shipment to the reference laboratory. The difference between these two values is used to evaluate effects of transportation such that linear drift of the transfer standard is substantially removed from the importation process. The process is also not dependent on the timing of the measurements made by the reference laboratory.

Fig 1



There is a benefit to the accreditation body as well. The laboratory is expected to declare the value of their standard <u>before</u> shipping it to the reference laboratory. The value V_1 is determined from known drift rates and comparisons with other standards in the group. The reference laboratory does not report the results (V_{ref}) of its measurements until the other laboratory has made its second set of measurements (V_2). In this way the capabilities of the laboratory are audited using its own standard. A scheme that could only be invented by Civil servants!

AVT Calculations

These determine the drift (equation 1) in the transfer standard over the importation period (ignoring temperature effects which are dealt with separately) and the error (equation 2) in the laboratory's Volt. The national laboratory measures the standard ten times over a two week period. Similar measurements are made before and after shipment by the Fluke laboratory.

$$Drift = V_2 - V_1 \qquad (1)$$

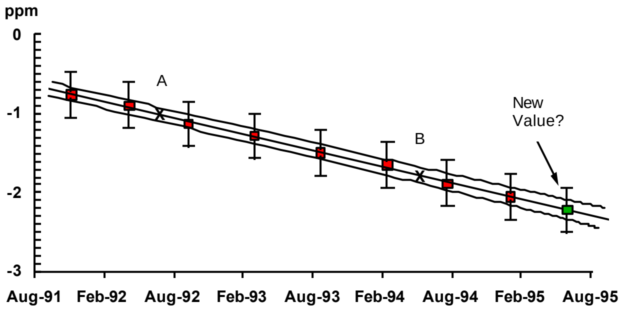
$$Error = V_1 + \{[(t / T) \times (V_2 - V_1)] - V_{ref}\} \qquad (2)$$

The calculation is not dependent upon the exact timing of the calibrations, but it is considered to be good practice to keep the intervals between sets of measurements approximately equal, as it allows time for the standard to recover from any short term perturbations as a result of transportation. At the end of the process, the calculated error may be used to correct the house Volt. It is possible for things to go wrong during shipment or even at the national laboratory. For this reason unexpectedly large errors are not corrected without due consideration of their cause.

Stability with Time

This is potentially the most significant uncertainty contribution. In lower level laboratories, it is normal practice to have a standard calibrated yearly without worrying too much about how it may perform in between calibrations. In a high level laboratory (with a capability better than 1ppm), the calibration history will be used to determine a drift rate from which predictions can be made about how the standard will behave between calibrations. With such data it is quite normal to make weekly or monthly corrections, depending on the actual drift rates. Correcting in this way can make a vast difference to the overall uncertainty achieved. For example, a standard that changes by 1ppm per year may be predictable to within 0.2ppm per year - a five-fold improvement if drift corrections are applied. Systematically correcting for the effects of drift is one of the most valuable ways of improving the company Volt.

Fig 2



Using Historical Data

The drift rate can be determined in several different ways. If the drift is essentially linear, the process is very simple. Figure 2 shows calibration results at six monthly intervals over three years. The data points can be conveniently split into two groups. The average values for each group in terms of ppm and time can be plotted with a "X" as points "A" and "B". A straight line drawn between the two points is the best possible fit (assuming linear drift). It is the same as would be achieved by computation using Linear Regression or a "Least Squares Fit". Because of its simplicity and visibility it is a valuable evaluation tool. The latest (new) value will be expected to be on, or close to, the best fit line. The difference between the predicted value and the actual value reported by the national laboratory is treated as a type B uncertainty contribution.

There are other methods that are used for drift rate analysis such as Microsoft's Excel®. This particular spreadsheet program not only offers a variety of curve-fitting options, but will also report the actual coefficients used as well as providing very high quality reports and charts.

It is also important to estimate the drift rate uncertainty. This can be obtained from the distribution of the individual points around the best-fit line. The standard deviation or standard error of these points can be multiplied by a value from the Student's "t" table and the required confidence level and number of points. Sometimes the drift rate uncertainty is plotted as additional curves above and below the "best fit". These lines are known as the confidence prediction interval.

Use of Spreadsheets

Definitely the most valuable software tool in the laboratory, a spreadsheet is particularly suitable where complex calculations are required on a regular basis. We have several spreadsheets included as part of our accreditation. They are used for both calculating measurement results and for estimating uncertainties. However, before a new spreadsheet is used, it is tested. The calculations are evaluated by entering known values, if formulae are copied to other cells or ranges they are also tested. It is very easy to have power of ten errors or cell address errors that are not immediately obvious. Once the spreadsheet has been evaluated and proved to work as designed, it is released with an issue level and date and the name and organization of the originator. Spreadsheets are subjected to configuration control just like any other software. No unauthorized modifications are allowed.

The potential of such software when coupled with software for automation offers some fantastic opportunities for automated measurement and data analysis. Unfortunately, accreditation bodies do not always have the necessary computer skills to evaluate such systems with confidence. This must be borne in mind when deciding on the adoption of complex methods.

Stability with Temperature

Most electrical standards are affected by temperature. Knowledge of the temperature coefficient of the standard allows a correction to be made. Most instruments will have a temperature coefficient specification - usually expressed in ppm/°C. However, the sign and actual magnitude may not be known. Where the best accuracy is required, the temperature coefficient must be measured and a temperature dependent correction applied. The residual uncertainty remaining after correction will be that of determining the actual temperature and the uncertainty of the temperature coefficient measurements. This is treated as a type B uncertainty contribution. If the temperature coefficient is measured over a 10°C range and the relative uncertainty of the voltage measurement is 0.2ppm, the temperature coefficient uncertainty will be 0.02ppm/°C. Where the temperature coefficient is small, the effects of the temperature uncertainty will usually be insignificant. It is important to record the local ambient temperature whenever voltage measurements are made. Allowances must be made for the effects of stacking instruments or placing them in racks - often the temperature under such conditions will be several degrees higher than the local ambient air temperature. This, in turn, will affect the uncertainty.

Noise and Short Term Stability

Whenever measurements are made, there will be variations in the values recorded due to electrical noise and short-term perturbations. Repeating the measurement or taking multiple readings allow a more precise value to be obtained. If the noise is entirely random (unlikely with Zener voltage standards), simple statistical methods can be used in its evaluation. However, usually there will be a small amount of periodic or "meandering" component present in the output that means that knowledge of the instrument is essential when determining the uncertainties attributable to this kind of behavior. This "non-randomness" can be due to the characteristics of the Zener itself, the temperature control circuits, or external influences such as short-term temperature fluctuations. For this reason, it is important to ensure that the measurements are repeated at different times in order to make a good estimate of the short-term variations under "typical" operating conditions. Sometimes, the best measurements can only be obtained under battery power, but care must be taken to ensure that the output voltage of the device does not slowly change because of internal temperature changes resulting from disconnecting the line supply and battery charging circuits. In the Fluke laboratory most measurements are made under line power and its affect and any additional noise are included in the uncertainty budget. For convenience, the sample size chosen for measurements is usually greater than 20. This means that a standard coverage factor for k can be used. Under these conditions, noise and short term stability contributions are treated as type A components.

Combining Uncertainty Contributions

There are three steps to this process. The first involves the estimation of the standard uncertainty for type A contributions based on the standard deviation of a sample of >20 measurements (note that in the case of a DC voltage standard, there may only be one type A contribution). The second stage is to estimate a standard uncertainty for the type B contributions. The final stage involves the combination of the type A and B standard uncertainties to give an expanded uncertainty at a specified minimum confidence level. In the case of NAMAS Accredited laboratories, this will not be less than 95%.

Standard Uncertainty of Type A Contributions

This is normally obtained from the root-sum-of-squares of the individual standard deviations (equation 3), however, as mentioned earlier, there may only be one contribution of this type where a transfer between voltage standards is performed.

$$A = \sqrt{s_1{}^2 + s_2{}^2 + s_3{}^2 + \ldots.. s_m{}^2} \qquad (3)$$

Standard Uncertainty of Type B Contributions

One of the type B contributions will be the calibration uncertainty. This is shown as a_1 in equation 4. If this is from another organization, it will, by international agreement, already be at a 95% minimum confidence level and must be divided by k (the *coverage factor*) before combining it with the other type B contributions. a_2 ..... a_m are assumed to be rectangularly distributed and the sum-of-squares of these contributions divided by (root) three effectively gives a standard deviation that may be combined as an RSS with a_1. The calculation reflects the central limit theorem[4] which recognizes the fact that in electrical measurements, when a number of contributions of <u>any</u> distribution are combined, the resultant probability distribution tends to be normal. The method shown is, in effect, estimating a standard deviation for the systematic contributions.

$$B = \sqrt{\frac{a_1{}^2}{k^2} + \left(\frac{a_2{}^2 + a_3{}^2 + \ldots.. a_m{}^2}{3} \right)} \qquad (4)$$

Expanded Uncertainty

Uncertainties should not be reported without a qualifying confidence level. This is usually expressed as a minimum confidence level, meaning that the confidence probability in not <u>less</u> than that specified, although it may be higher. The expanded uncertainty is calculated by multiplying the root-sum-of-squares of the type A (guassian) and type B (non-guassian) components by the coverage factor k, where k is equal to 2 for a confidence level of approximately 95% (equation 5). The expanded uncertainty is reported on the calibration certificate as the measurement uncertainty together a statement of the confidence probability, e.g. "The reported expanded uncertainty of ±0.5ppm is based on a standard uncertainty multiplied by a coverage factor of k=2, providing a level of confidence of approximately 95%".

$$U = k\sqrt{A^2 + B^2} \qquad (5)$$

NAMAS Audits

The claimed capability of the laboratory will be regularly tested by measurement audits. These are of two types: The routine audit via traceability mentioned earlier, and the so called "black box" audits, where the value is not known by the laboratory being audited. This audit uses standards owned by the NAMAS Executive. In order to try and minimize the possibility of laboratories maintaining a history of the audit standards (and therefore predicting what its value might be), the NAMAS audit officer tries to ensure that the laboratory receives a different standard each time. The result of the audit must be within one half of the accredited best measurement capability of the laboratory and is calculated by dividing the error in the laboratory's result by the root-sum-of-squares of the combined uncertainties (equation 6). This is called the *En* ratio.

$$En = \frac{E_{lab} - E_{ref}}{\sqrt{U_{lab}^2 + U_{ref}^2}} \qquad (6)$$

If *En* is greater than 0.5, the laboratory may be requested to increase its accredited uncertainty. At the present time, the best NAMAS Accredited uncertainty for 10V DC, based on the importation of voltage traceability using a transportable standard, is ±0.4ppm[5] at a minimum confidence level of 95%. The uncertainty given by the National Physical Laboratory for the calibration of this standard is ±0.25ppm at a minimum confidence level of 95%.

Other Applications

The techniques described may also be used with great effect for other types of standard including resistance standards, long-scale DMMs, in fact anything where data accumulated from the measurement process over period of time and under defined conditions can be subjected to this kind of analysis. However, it is essential that any corrections that can be applied, should be applied and that regular audits are performed.

References

[1] M3003 "The Expression of Uncertainty and Confidence in Measurements with Particular Reference to Electrical Measurements" Edition 1, December 1997. Published by the UKAS.

[2] ISO "Guide to the Expression of Uncertainty in Measurement" prepared by TAG4/WG3, Edition 1, 1993. Published by ISO.

[3] NAMAS M10 Accreditation Standard, Published by the UKAS.

[4] Uncertainty, Calibration and Probability, Edition 2 1990, Author C.F. Dietrich, Published by Adam Hilger, London. Chapter 4 discusses the combination of rectangular probability distribution and compares results with normal distributions of equal standard deviation.

[5] NAMAS Calibration Accreditation No 0183 (Permanent Laboratory) Fluke Precision Measurement Ltd., UK.