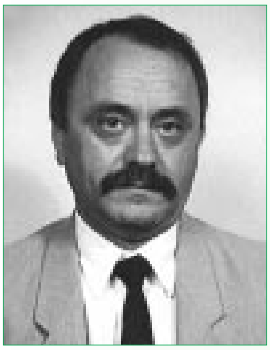

UNCERTAINTY

Qualifying measuring instruments based on the quadratic approach of the *Guide to the Expression of Uncertainty in Measurement*

DR. PÉTER BÖLÖNI, Sensor Metrology Ltd., Budapest, Hungary

Abstract

Qualifying a measuring instrument involves both a comparison and a decision:

- *Comparison of the metrological parameter(s) of the instrument as determined by (or evaluated based on) the results of the calibration and the required (or supposed) values of these parameters derived from technological or safety requirements (or tolerance in the general sense) or the manufacturer's specification of the instrument. This comparison usually results in a decision to accept or reject the instrument for use;*
- *The decision is influenced both by the first order (α-type or type I)* and the second order (β-type or type II)* errors of the decision and also by the uncertainty of the value measured or reproduced by the standard. A quadratic evaluation of the measured deviations reduces the risk of the incorrect decisions being made in both cases.*

Introduction

The qualification of a measuring instrument is based on the d_i deviations between the measured y_i values provided by the measuring instrument to be qualified and the respective z_i reference (or conventional) values of the measurand reproduced by or measured with the standard:

$$d_i = y_i - z_i \qquad (1)$$

* *Note:* Throughout the text, the uniform denominations "α-type" and "β-type" are used.

D is a random variable since both Y and Z are random variables and Y can itself be considered as the sum of the two random variables.

The first variable is the expected value of the results deviating in a random manner from the true or definitive value of the measurand in the range of the measurements (these deviations are called *systematic errors* and are estimated with the biases).

The second variable is the classical random error (or deviation), i.e. the deviation of the measured values from the expected value or from the average of many results in practice.

The metrologist's task is either to characterize the measuring instrument with an s(Y) standard deviation (or a multiple thereof) or to judge if the estimated standard deviation is less (but not more) than a value that is (or ought to be) specified for the measuring instrument to be qualified.

Note: The instrument's specification does not usually define the term *accuracy* itself, so the user should consider it either as a multiple of a standard deviation or as certain limits for the maximum permissible error (mpe).

The s(Y) standard deviation has to be calculated or estimated from the results of the calibration and from the specification of the measurement standard.

As the variance of D is equal to the sum of the variances of Y and Z, the variance of Y is the difference of the variances D and Z respectively. The variance of D can be estimated from the results of the calibration by eq. (12) according to the so-called or noted type A estimation of the standard deviation (or uncertainty) on the basis of the experimental or relative frequency-based concept of the probability. In this case:

$$s(Y) = \sqrt{s^2(D) - s^2(Z)} \qquad (2)$$

To use this approach the s(Z) standard deviation of Z has to be estimated from the calibration certificate of the measurement standard used for the calibration.

s(Z) is known this way, at least in principle. Attributing probability distribution functions to Y and Z, the probability distribution function of D can be derived and the probability of the event of:

$$P[|d| > k \cdot s(Y)] = p \quad (3)$$

can be calculated for any value of d. If the value d_b is found to be out of the limits of the mpe with a low probability of $p(d_b)$ then one might reject even an acceptable instrument. This is the so-called α-type error (concluding a hypothesis H_1 when H_0 is true). And as the absolute values of a few measured deviations can easily be less (or not more) than the critical value for d one can obtain acceptable values for d even in the case of measuring instruments having greater errors or standard deviation than the allowed value. Accepting an "unacceptable" instrument, i.e. accepting the H_0 hypothesis when another value of H_1 is true is called the β-type error.

For an acceptable instrument, a measured deviation can fall outside the tolerance limits because of a large (but rare) random error **or** because of the unknown (and therefore not considered) error of the measurement standard, **or both**. Similarly a measured deviation can fall within the acceptable range even for an unacceptable instrument because of the random nature of the errors **or** because of the influence of the unknown error of the measurement standard, or both. To consider or to reduce the chances of incorrect decisions being made, at least three different principles or rules of qualification are applied and one additional principle is suggested below.

1 Spreading the risk of an incorrect qualification

A traditional qualification method is to compare all the measured deviations with the limits of the mpe's derived from the accuracy specification of the instrument:

$$d_l \le d_i \le d_u \quad (4)$$

where:

- d_l is the lower mpe limit (usually negative and may be a function of the measured value);
- d_u is the upper mpe limit (usually positive and may be a function of the measured value); and
- d_i is the i^{th} measured deviation.

The measuring instrument will be accepted or qualified as being "acceptable" if the condition in eq. (4) is met for all the d_i values. The decision might however be the subject of both an α-type or a β-type error, since a measured deviation can fall outside the tolerance limits even for an acceptable instrument because of the occurrence of a large but rare random error or because of the unknown (and therefore not considered) error of the measurement standard, or both.

On the other hand several consecutive measured deviations can fall within the acceptable range even for an unacceptable instrument because of the random nature of the errors and the influence of the unknown error of the measurement standard, or both. The probabilities of these two incorrect decisions being made are often considered to be equal or at least similar in value and neglected for this reason. This practice is often used in legal metrology although the owner and user of incorrectly rejected and incorrectly accepted instruments is not necessarily one and the same.

Furthermore, since alternative H_1 hypotheses to describe the behavior of unacceptable instruments are usually not proved, the probability of accepting an unacceptable instrument can hardly be ascertained. Having proved hypotheses for the probabilities of the D deviations, the probability of α-type errors can be calculated:

$$P(\alpha) \approx 1 - \int_{d_l}^{d_u} \varphi(D) \cdot d(D) \quad (5)$$

where $\varphi(D)$ is the probability density function of the deviations with the estimation $s^2(D)$ for the variance. Chebisev's equation can be used in cases where no proved hypothesis is available (i.e. when the distribution of the sample deviates significantly from the supposed one). The original form of the equation is:

$$P[|D - M(D)| > \varepsilon] = \frac{\mathrm{var}D}{\varepsilon^2} \quad (6)$$

where:

- M(D) is the expected value of D, which is zero in the present case; and
- ε is a small positive number.

Let $|d_l| = d_u \cong k \cdot s(D) = \varepsilon$ where k can be the well-known and widely used coverage factor. In this case:

$$P[|D| \ge k \cdot s(D)] \le \frac{1}{k^2} \quad (7)$$

The uncertainty of the z_i reference values contributes generally to the chance of the incorrect decision being made, but this contribution can often be neglected after reducing it to below one tenth of the $|d_l| = d_u$ values.

2 Qualification for maximum confidence of operation

Another traditional qualification method is to compare all the measured deviations with the "tightened" mpe limits:

$$d_l + U_z \le d_i \le d_u - U_z \quad (8)$$

where:

- d_l is the lower mpe limit;
- d_u is the upper mpe limit;
- d_i is the i^{th} measured deviation; and
- $U_z = k \cdot s(Z)$ is the uncertainty of the reproduction or measurement of the reference or conventional value of the measurand.

The measuring instrument will now be accepted or qualified as being "good" if the condition in eq. (8) is met for all the d_i values. An unacceptable instrument (i.e. one that failed to meet the specifications) will hardly be qualified as "good" in this way.

This decision might however more often be the subject of an α-type error than in the case of the "shared risk", since a measured deviation can fall outside the tightened tolerance limits with somewhat more probability even for an acceptable instrument. The reason for this can be the occurrence of a large and less rare random error or because of the unknown (and therefore not considered) error of the measurement standard, or both. Accepting $\varphi(D)$ for the probability density function of the differences with the estimation $s^2(D)$ for the variance, the probability of the α-type error is:

$$P(\alpha) = 1 - \int_{d_l + U_z}^{d_u - U_z} \varphi(D) \cdot d(D) \quad (9)$$

and the formula in eq. (10) can be used if no suitable $\varphi(D)$ probability density function is available for the differences:

$$P\left[\,|d| > k \cdot s(d) - U_z\right] \le \frac{s^2(D)}{(k \cdot s(D) - U_z)^2} \quad (10)$$

3 Avoiding the rejection of an instrument that meets the specifications

One more traditional qualification method is to compare all the measured deviations with the extended limits of the mpe's:

$$d_l - U_z \le d_i \le d_u + U_z \quad (11)$$

where:

- d_l is the lower mpe limit;
- d_u is the upper mpe limit;
- is the i^{th} measured deviation; and
- $U_z = k \cdot s(Z)$ is the uncertainty of the reproduction or measurement of the reference or conventional value of the measurand.

The measuring instrument will now be accepted or qualified as being "good" if the condition in eq. (11) is met for all the d_i values. Practically all of the "good" instruments will be accepted but the chance of unacceptable instruments being accepted (i.e. those that fail to meet the specifications) will be increased this way.

This decision might more often be the subject of a β-type error than in the case of the "shared risk", since a measured deviation can fall within the extended tolerance limits with somewhat more probability even for an unacceptable instrument because of the occurrence of a few consecutive small and (unlikely but possible) random errors or because of the unknown (and therefore not considered) error of the measurement standard, or both.

4 A quadratic approach

Reducing the probability of unacceptable instruments being accepted by tightening the limits increases the probability of good instruments being rejected; reducing the probability of good instruments being rejected by extending the limits of acceptance increases the probability of unacceptable instruments being accepted. The probability of making incorrect decisions can be reduced:

- using measurement standards with low measurement uncertainty or reproduction when U_s is not more than 1/10 of the mpe (this is expensive and affects only one of the incorrect decision sources);
- using higher values for the coverage factor than k = 2 (though the demand for higher confidence does not aid in intuitive thinking but allows only likely tendencies or facts to be stated or recognized); or
- using the quadratic estimation of the *Guide* for the qualification as well.

Perhaps the accuracy or the uncertainty of the measuring instrument can be characterized either with the s(Y) standard deviation or with a multiple thereof. For this the experimental standard deviation of the results of the calibration shall first be calculated according to the so-called A type evaluation of the results:

$$s(D) = \sqrt{\frac{\sum_{i=1}^{n} d_i^2}{n}} \tag{12}$$

Substituting in eq. (2) the value of s(D) as calculated in eq. 12, one can compute a standard deviation which characterizes the measuring instrument with an experimentally determined expanded uncertainty that conforms to the *Guide*:

$$U = 2\sqrt{\frac{\sum_{i=1}^{n} d_i^2}{n} - s^2(Z)} \tag{13}$$

This can be used directly for uncertainty calculations, is not sensitive to any large individual deviation which can itself decide the qualification of the instrument, and is not affected by the limited accuracy of the measurement standard used for the calibration.

Extended tests have shown that above a certain low limit in the number of measured deviations, this second moment or non-central variance based on the s(Y) parameter can well describe the performance of the measuring instrument to be qualified and this s(Y) parameter can be interpreted according to existing international metrological normative documents.

Summary and conclusion

Traditional linear principles of qualification cannot exclude the possibility of incorrect decisions being made. The probabilities of this occurring can be reduced by applying the quadratic evaluation of the deviations and by considering the standard deviation of the reproduction or measurement of the reference value. This approach was presented to the 49th General Assembly of CECIP (*Comité European des Constructeurs d'Instruments de Pesage*) for further consideration. ■

Bibliography

[1] *International Vocabulary of Basic and General Terms in Metrology*, BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, OIML, 1993.

[2] *Guide to the Expression of Uncertainty in Measurement*, BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, OIML, Corrected and reprinted, 1995.

[3] P. Bölöni: *Connections between various interpretations of measurement uncertainty*, OIML Bulletin, Vol. XXXVIII No. 4, October 1997, pp. 21–23.

[4] W.H. Emerson: *The new VIM (Vocabulary for metrologists)*, OIML Bulletin Vol. XXV, No. 2, April 1994, pp. 30–34 and *Personal reflections about the VIM*, pages 35–37.

Note: At the time of going to press, the OIML TC 3 *Metrological control* meeting (1–3 June, Paris) has not yet taken place; one of the topics to be discussed at this meeting is measurement uncertainty in legal metrology. Information on the outputs of the meeting will be given in the October 1999 issue of the Bulletin.