Paper presented at the
IMEKO 20th TC3, 3rd TC16 & 1st TC22 International Conference,
November 27th to December 1st, 2007, Merida , Mexico

Suggestions on how to include the effect of reproducibility in the evaluation of simplified calibration procedures

Georg Wegener, Hans W. Werner,
Hottinger Baldwin Messtechnik GmbH, Im Tiefen See 45, Darmstadt, Germany
E-mail: georg.wegener@hbm.com, hans-werner.werner@hbm.com

Abstract

The presented study suggests a method that allows estimating the reproducibility error in torque calibrations without the need of performing several measurement series with the torque transducer being mounted in different mounting positions. The aim is to enable uncertainty estimations for simplified calibration procedures, too. The approach is to use statistical data gained from previous calibrations of transducers of the same type. The study shows how to select and classify past calibrations in order to allow prognoses for future calibrations. Furthermore it introduces the method of statistic evaluation and gives special considerations about the way in which the reproducibility error influences the overall measurement uncertainty in the case of a simplified calibration.

Keywords

Torque calibration, simplified calibration, reproducibility, statistical estimate, type testing, VDI 2646

1. Introduction

The generally accepted method for determining the reproducibility effect in calibration of force and torque transducers is three separate measurement series. For each of these series the transducer is mounted in a different angular mounting position. Procedures for this approach are the well-known procedures for high requirement calibrations. Examples are the German standard DIN 51309 [1], the European EA-10/14 [2] and the force calibration standard ISO 376 [3].

In the last few years there has been a growing demand to have uncertainty estimations for calibrations with lower requirements, too. For these calibrations, three different mounting positions would be too time- and cost-consuming. Therefore different approaches are made to provide an estimation of the reproducibility effect from a-priori information. A common approach is to base the estimation on statistical information, namely the data from a number of more comprehensive calibrations in the past. This approach is suggested in the recent German guidelines VDI 2646 for torque [4], and DKD R 3-3 [5]. However the torque guideline [4] does not fully explain about the detailed mathematical procedure for gaining such an estimation. This was the task of the study presented in this paper. Some of the considerations presented here have already been taken into account for in the present form of the force guideline DKD R 3-3 of the German Calibration Service (DKD) [5].

2. Mathematical description of the reproducibility effect

In calibration literature and guidelines there are two different concepts for defining a quantitative parameter for the reproducibility effect. Both are based on performing several independent measuring series. For each measuring series, the transducer is dismounted (or partly dismounted, depending on the type of calibration machine), rotated by a given angle and then re-mounted. The most common concept is to have mounting positions of 0°, 120° and 240°. Comparing the measurement signals obtained for each individual step in the different mounting positions allows deriving a quantitative parameter b (reproducibility in different mounting positions) for the description of the reproducibility. This parameter has to be determined separately for each calibration step.

The first concept for defining such a quantitative parameter is the so-called reproducibility error *b* to be the span between the maximum and the minimum measured value (only measurements during load series with increasing load are taken into account):

$$b(M_K) = \left| X_{\max}(M_K) - X_{\min}(M_K) \right| \tag{1}$$

For the subsequent implementation of this parameter in the estimation of the uncertainty of measurement method B according to GUM [6] is applied. The assumed distribution function is typically a rectangular one or a U-shaped one. The half width *a* of this function equals half the reproducibility error *b* as defined by equation (1). This is the method used by the DIN 51309 standard [1] and also by the guideline which gave the initiation for this study, the guideline VDI 2646 [4].

The second concept for defining a quantitative parameter for the reproducibility is based on the standard deviation of the different measured values for the respective load step according to the GUM method A. This method is used for example by the EA-10/14 standard [2].

It is not subject of the presented work to discuss which one of the two above concepts for the description of the reproducibility error should be used. The task was to develop a method for deriving a suitable estimate for *b* to be used for simplified calibrations according to the new guideline VDI 2646 [4]. Since that guideline uses the first one of the concepts described above (*b* as a span, combined with a type B evaluation), the further considerations are also based on this concept. Nevertheless, the main part of the considerations (section 3 and most of section 4) can also be used when aiming at a derivation of a standard uncertainty according to method A. Only the considerations on integrating the contribution of the reproducibility error into the total uncertainty budget (section 5) cannot be applied in the same way.

Input data for the statistical evaluation

For the evaluation a database of full calibrations was available, which had been gained by the calibration laboratory of HBM over several years. Calibrations to evaluate were selected according to the following rules:

- each size/measuring range of each transducer model was examined separately
- each calibration step was examined separately
- calibration of both, brand-new and used transducers was included, but only one calibration of each individual
- statistical results can be transferred to future calibrations only if both are performed on the same calibration machine
- outliers were eliminated only if they fulfilled two criteria: the purely mathematical criterion of Grubb's test for outliers and a plausible technical reason that the transducer was defective

In order to save evaluations and in order to increase the number of available calibrations for the statistic, only the calibration for clockwise torque was evaluated. It is technically reasonable to assume that statistically the behaviour of a given transducer type is identical with clockwise and with anti-clockwise torque.

3. Estimating the reproducibility from the statistical data

As has been explained above, the aim of the work was to derive an estimate for the parameter *b*, as described in equation (1). The method which seems the most obvious would be to take the mean value of this parameter from all the past calibration certificates. But statistical analysis of these values caused serious doubts about statistically correct results.

Therefore, a different approach was chosen, based on the measured values of all individual measuring series. In order to isolate the reproducibility effect from other differences among the calibrated transducers or among calibration conditions, the data had to be normalized in a suitable way: For each measured value, first the deviation from the mean value was calculated. The mean value in this context is the mean of all measurement values obtained during the respective calibration for the load step considered. As the second step, this value was expressed as a percentage of the mean value.

$$X_{k,j} = \frac{X_{k,j,\mathrm{raw}} - \overline{X}_{k,\mathrm{raw}}}{\overline{X}_{k,\mathrm{raw}}} \qquad \text{with} \qquad \overline{X}_{k,\mathrm{raw}} = \frac{1}{3}\sum_{j=1}^{3} X_{k,j,\mathrm{raw}} \qquad (2)$$

with

k = the consecutive number of the calibration step under consideration

j = the consecutive number of the measuring series (1, 2, 3 for the three different angular positions)

The result are positive and negative values, typically in the range between zero and ±0.05 %. Statistical analysis has shown that the resulting distribution is a normal distribution with an expected value of zero.

The next step in the derivation of the estimate for *b* is to estimate the maximum and minimum measured values that can be expected for the respective transducer type at the respective load step (again in the same normalized form as used above). In order to derive a prognosis from the statistic evaluation, we have to take into account that the values are based on a limited sample size. In order to obtain a defined confidence interval, Student's factor *t* has to be taken into account. Student's factor is depending on the desired confidence interval and on the sample size *n*. The confidence interval was chosen as 95.45 % (two σ rule). The relevant sample size *n* in this context is not the number of measurement points, but the number of individual calibrations included in the statistic.

From technical a priori knowledge about torque transducers, we know that the maximum and minimum values for each individual torque transducer can be assumed symmetrical with respect to the mean value for this particular transducer. In other words: Let us think of a transducer which shows a deviation between the mean value and the maximum value which is equal to the limit of the 95.45% interval. Due to the symmetry of the reproducibility, it follows that its deviation between mean and minimum value will also be equal to the limit of the 95.45% interval. Therefore the 95.45% value for the reproducibility error *b* (defined as a span as in equation (1)) is actually the difference between the 95.45% values for the minimum and maximum measured values. Thus we obtain:

$$b_k = 2 \cdot s(Z_k) \cdot t\big|_{n-1;95\%} \qquad (3)$$

with

$s(Z_k)$ = the empirical standard deviation of all evaluated measured data associated with the respective calibration step Z_k

$t|_{n-1;\,95\%}$ = Student's factor for the degree of freedom *n-1*, associated with a confidence interval of 95.45 %

n = the number of individual calibrations included in the statistic for the considered type of transducer

The steps of this evaluation can be seen in the schematic diagrams in figure 1 and figure 2. The examples in the diagrams show the evaluation for two different measuring ranges / sizes of the same type of transducer. A quantitative comparison of the estimation values for the parameter *b* shows that it is extremely important to do the evaluation for each measuring range separately.

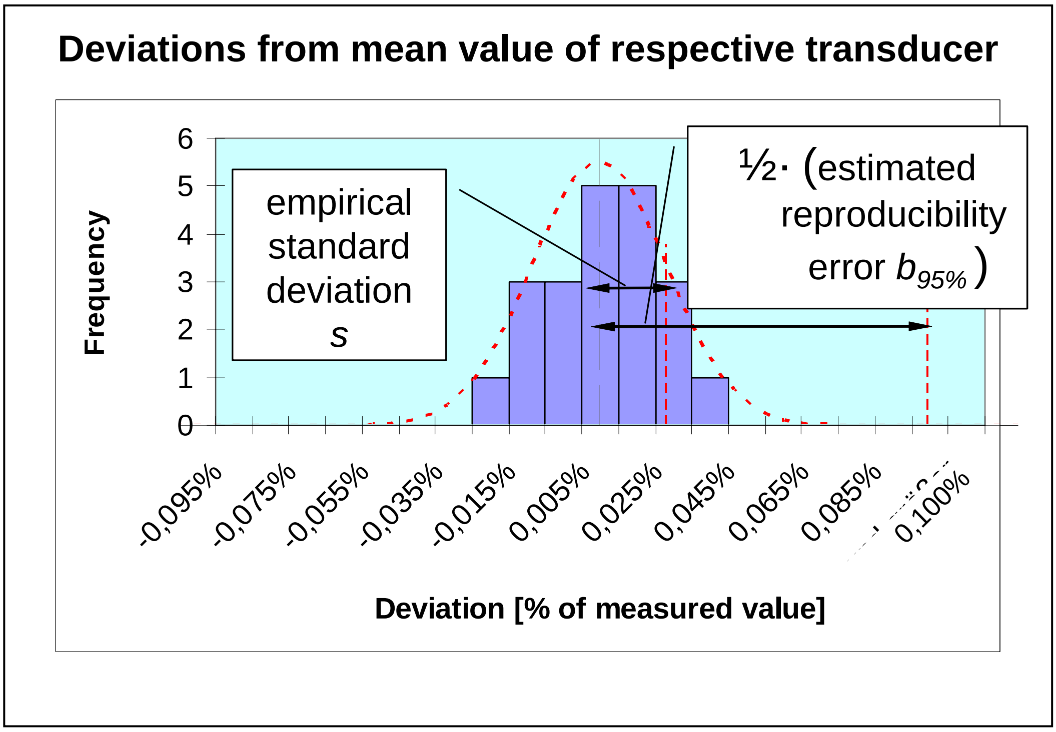

Fig. 1: Transducer type A, measuring range I, calibration step 60% of full range
7 individual transducers calibrated, Student factor for 95.45%: 2.52,
empirical std. deviation*:* 0.016%, estimated reproducibility error *b*=0.081%

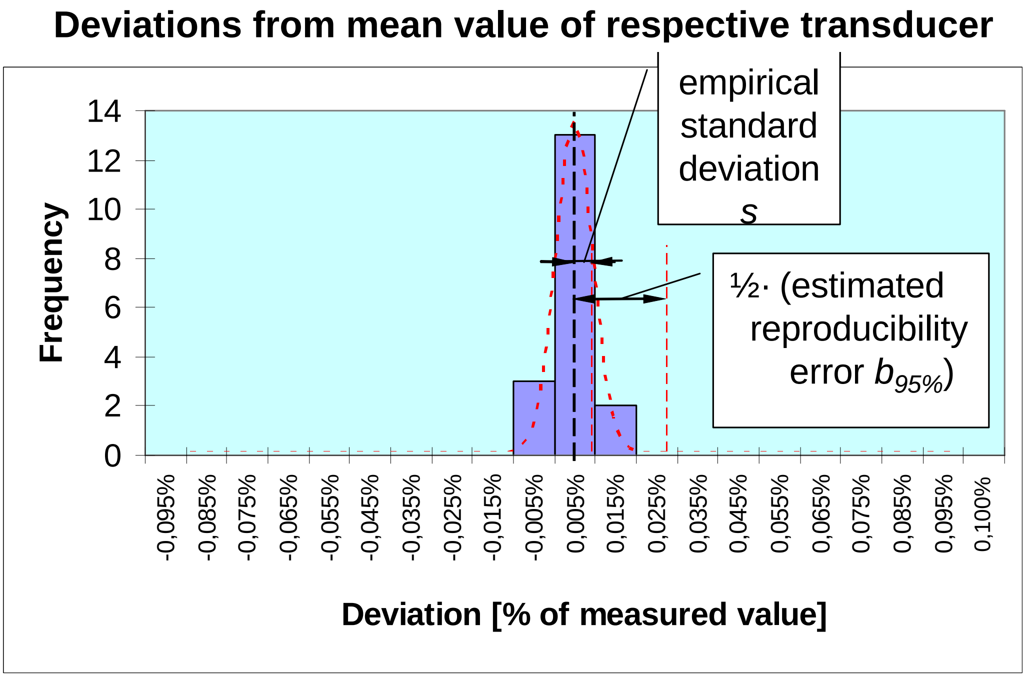

Fig. 2: Transducer type A, measuring range II, calibration step 60% of full range
6 individual transducers calibrated, Student factor for 95.45%: 2.65,
empirical std. deviation*:* 0.005%, estimated reproducibility error *b*=0.025%

4. The influence of the reproducibility error in the total uncertainty budget

When comparing the case of the simplified calibration method to the case of a calibration with three mounting positions, there is also a difference in the way the reproducibility error contributes to the uncertainty of measurement. This difference is illustrated in figures 3 and 4. We assume that the real uncertainty behaviour of the transducer (ignoring all other influences but the reproducibility effect) for a given load step T_k can be expressed by the mean value of all possible measuring signals for this torque step $\overline{X}_k$ and the reproducibility error b_k as follows:

$$X_k = \overline{X}_k \pm \frac{b_k}{2} \qquad (4)$$

The calibration result of classical calibration with three mounting positions gives a reasonable estimate for both, the mean value and the reproducibility error.

From the perspective of the user of the calibrated transducer, the main question concerning measurement uncertainty is: How big is the range of output signals that have to be expected at this particular torque step? Given the mean value $\overline{X}_{k123}$ and the estimated reproducibility span b_{k123} from the calibration, equation (4) can be approximated as

$$X_k \approx \overline{X}_{k123} \pm \frac{b_{k123}}{2} \qquad (5)$$

Figure 3 illustrates this consideration.

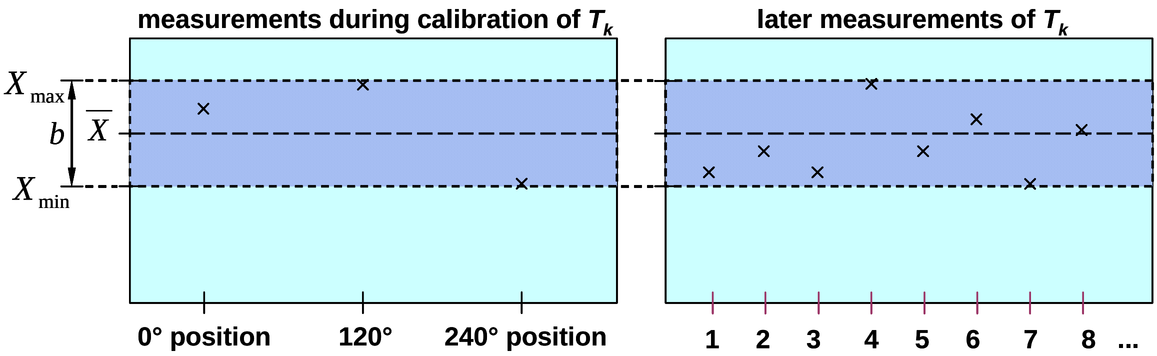

Fig. 3: Effect of reproducibility on uncertainty for transducer calibrated in three mounting positions.
left: during calibration, right: during later measurements by the user

For the user who only has data of the simplified calibration procedure available, the value of $\overline{X}_k$ is not known. He only knows the measured value X_{k1} of one single mounting position. If we assume that the reproducibility error b_k is known from statistical knowledge, the user knows that the correct mean value must be in the interval

$$\overline{X}_k \approx X_{k1} \pm \frac{b_{k,\text{statist}}}{2} \qquad (6)$$

Considering this, equation (4) results in

$$X_k \approx X_{k1} \pm b_{k,\text{statist}}\,. \qquad (7)$$

Figure 4 illustrates this consideration. Furthermore, the schematic also illustrates that the deviation is (at least partly) a systematic one. Therefore it would be incorrect to treat it as a random deviation according to the GUM. Nevertheless, this systematic deviation cannot be compensated with the information available from the simplified calibration procedure. The suggestion of the authors is to treat it in the same manner as other un-corrected systematic deviations.

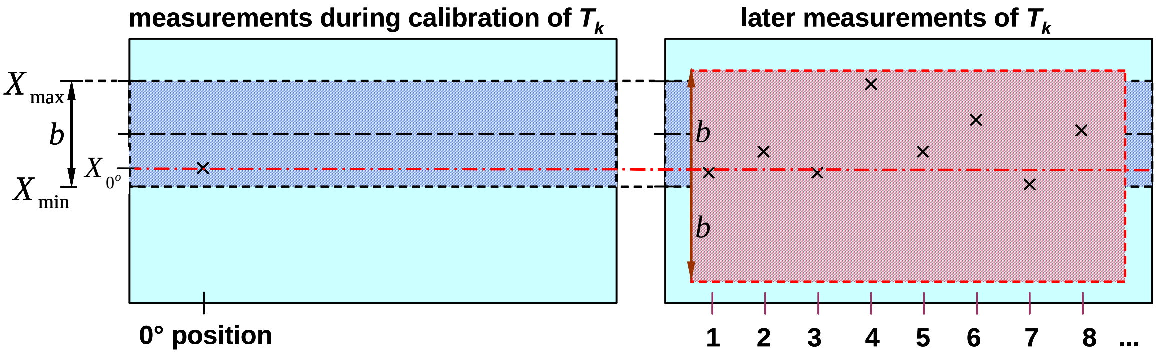

Fig. 4: Effect of reproducibility on uncertainty for transducer calibrated in only one mounting position.
left: during calibration, right: during later measurements by the user

An example for this kind of deviations is the treatment of the deviation of indication f_a in the torque calibration standard DIN 51309 [1] for the case of a linear approximation function.

$$W'(M_K) = |f_a(M_K)| + |b_{\text{statist}}(M_K)| + k \cdot \sqrt{w^2_{\text{standard}}(M_K) + \sum_j w_j^2(M_K)} \qquad (8)$$

This method was chosen although it is not conform with the VDI guideline. The guideline ignores the special situation and treats the contribution of the reproducibility in the same way as in a calibration with three mounting positions.

5. Conclusion

A consistent method has been derived and explained for calculating a statistical estimate of the reproducibility error *b* from a statistics gained from previous calibrations of the same transducer type. However, this method cannot replace full calibration. Even if statistical data was very good and the reproducibility error could be predicted very accurately, there is always a lack of information about the mean value when a calibration is performed only in one mounting position. For this reason the uncertainty of measurement will always be bigger than with several mounting positions.

Furthermore, the method requires a huge database since only transducers of the same type and measuring range / size can be used for the prediction. Typically, such huge databases can only be created by manufacturers calibrating their own torque transducers.

An alternative approach for providing information for the reproducibility would be to take reproducibility data from a previous calibration of the same individual transducer. This would require that the first calibration of an individual has to be done with several mounting positions and all future calibrations can be done with a simplified procedure. This method is (among others) suggested in the latest guideline for a simplified calibration of force transducers by the German calibration service DKD, guideline DKD R 3-3 [5].

6. References

[1] DIN (German Standards Institute), DIN 51309, Werkstoffprüfmaschinen – Kalibrierung von Drehmomentmessgeräten für statische Drehmomente (Materials Testing Machines – Calibration of static torque measuring devices), 2005.

[2] EA (European co-operation for Accreditation), EA-10/14, EA Guidelines on the Calibration of Static Torque Measuring Devices, 2000.

[3] ISO (International Organization for Standardization), ISO 376, Metallic materials – Calibration of force-proving instruments used for the verification of uniaxial testing machines, 2004.

[4] VDI/VDE (Society of German Engineers), Guideline VDI/VDE 2646, Drehmomentmessgeräte – Mindestanforderungen an Kalibrierungen (Torque measuring devices – Minimum requirements in calibrations), 2006.

[5] DKD (German Calibration Service), Guideline DKD R 3-3 Kalibrierung von Kraftmessgeräten (Calibration of Force Measuring Devices), 2007.

[6] ISO, IEC, BIPM et al, Guide to the Expression of Uncertainty in Measurement ("GUM"), 1993.