Measurement 46 (2013) 3715–3719

Contents lists available at SciVerse ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

Critical observations on rules for comparing measurement results for key comparisons

Frank Härtig *, Karin Kniel

Physikalisch-Technische Bundesanstalt Braunschweig und Berlin, Germany

ARTICLE INFO

Article history:
Available online 29 May 2013

Keywords:
Intercomparison
Evaluation rule
Precision metrology
Comparability
E_n value
Normalized error

ABSTRACT

International comparison measurements form the backbone for validating the competence of metrology institutes among one another. An informative interpretation of the results, however, requires intensive studies of the individual reports and great expertise on the part of the readers. This is due to the lack of a clear procedure and guidance for the assessment of comparison measurements which leads to different values which are difficult to compare. The article explains the most significant reasons.

1. Introduction

The worldwide unification of measurements and their traceability to the International System of Units (SI) is coordinated by the highest authority in the field of metrology: the Bureau International des Poids et Mesures (BIPM) which is headquartered in Paris [1]. On 14 October 1999, 39 metrology institutes (NMI) and two international organizations signed a Mutual Recognition Arrangement (MRA) [2] with the aim of mutually recognizing measurement standards and calibration and measurement certificates and facilitating their trade relations among one another. By May 2012, the number of signatories had increased to 87. Key comparisons (KC) of single measurands strengthen mutual confidence and provide, at the same time, information about the competence of the members. On behalf of the Comité International des Poids et Mesures (CIPM), they are organized by the Consultative Committees (CCs). The Calibration and Measurement Capabilities (CMCs) of a participant are stored in the Key Comparison Data Base (KCDB) of the BIPM and are publicly available [3]. By 21 May 2012, 792 key comparisons were entered there [4]. Even if the organizational realization of a KC has been determined in great detail, there are great liberties regarding the calculation of a reference value (RV), the selection of the measurement uncertainty for the evaluation and the exclusion of participants whose measurement results do not appear suitable. As a consequence, the results of a KC can be assessed – and thus interpreted – correctly only when the reports at hand are known in detail.

2. Estimation of reference values

For the calculation of a RV, different evaluation methods are available [5–10]. The most commonly used are the simple mean, the weighted mean and the median. In addition, the RV value can be calculated with the aid of other methods which will not, however, be discussed here in detail. In the following the coverage factor $k = 1$ is representing a confidence level of 66% for the measurement uncertainty u which is similar to the expression $U(k = 1)$ whereas $k = 2$ is representing the expanded measurement $U(k = 2)$ at a confidence level of 95%.

2.1. Simple mean

In this calculation, all measurement values enter into the RV equally weighted and measurement uncertainties

* Corresponding author. Address: Bundesallee 100, 38116 Braunschweig, Germany. Tel.: +49 531 592 5300.
E-mail address: frank.haertig@ptb.de (F. Härtig).

0263-2241/$ - see front matter
http://dx.doi.org/10.1016/j.measurement.2013.04.079

are not taken into account. The simple mean is calculated in accordance with the following equation:

$$x_s = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{and} \quad U_s(k=2) = 2 \cdot \frac{1}{n}\sqrt{\sum\nolimits_{i=1}^{n} u_i^2} \qquad (1)$$

x_s is the simple mean, n is number of measurement results, x_i measurement value of participant i, U_s is expanded measurement uncertainty of simple mean, k is coverage factor, u_i is the standard measurement uncertainty of participant i.

Laboratories with strongly deviating measurement values may "draw" the simple mean – in spite of the indication of a large measurement uncertainty – into a "wrong" direction (Fig. 1).

2.2. Weighted mean

In this calculation, the measurement uncertainty associated with the measurement value is additionally taken into account in addition. Here, the measurement values of laboratories with small measurement uncertainty contribute to the formation of the mean value with larger weight. The weighted mean value is calculated in accordance with Eq. (2).

The procedure requires a conscientious indication of the measurement uncertainties by the participants; otherwise, measurement uncertainties which have, for example, been estimated as too small, may lead to a biased RV.

$$x_w = \sum_{i=1}^{n} \frac{x_i}{u_i^2} \cdot \frac{1}{\sum_{i=1}^{n} \frac{1}{u_i^2}} \quad \text{and} \quad U_w(k=2) = 2 \cdot \frac{1}{\sqrt{\sum_{i=1}^{n} \frac{1}{u_i^2}}} \qquad (2)$$

x_w is the weighted mean, n is number of measurement results, x_i is measurement value of participant i, U_w is expanded measurement uncertainty of the weighted mean, k is coverage factor, u_i is the measurement uncertainty of participant i.

2.3. Median

In the case of this calculation, the measurement values are sorted by their numerical value. If the number of measurement values is odd, the RV is the measurement value at the middle position of the list. If the number of the measurement values is even, the RV is calculated from the simple mean of the two measurement values at the middle position of the list (Table 1).

Table 1
Ordered measurement values.

−2.5	−2.0	−1.5	−1.5	−0.5	1.5
		x_m = −1.5			

Table 2
Measurement results of the example, x_i indicates the measurement values whereas $U(k=2)$ indicates the expanded measurement uncertainty.

Participant	x_i in µm	$U(k=2)$ in µm
A	−1.5	1.5
B	1.5	3.0
C	−2.5	0.4
D	−2.0	0.5
E	−0.5	2.0
F	−1.5	0.7

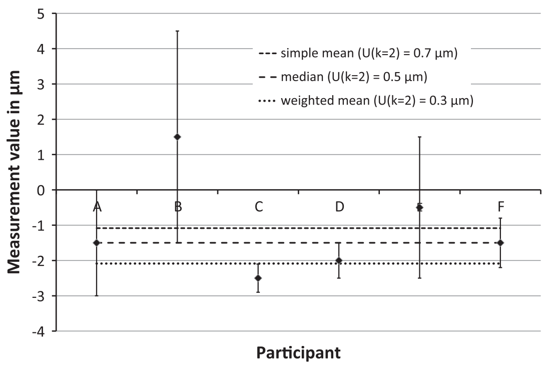

Fig. 1. Example for different RV caused by the weighted mean, the simple mean and the median calculation.

The measurement uncertainty of the median x_m can be determined via Monte Carlo simulations. This requires a corresponding evaluation software. In the example given in Fig. 1, the uncertainty of the median was calculated with an evaluation software for key comparisons [11,12]. An important property of the median is the robustness against outliers.

In the following, the three mean values are compared on the basis of an example. The measurement results used as a basis are shown in the list in Table 2. The data correspond to a realistic distribution from gear metrology. Fig. 1 shows the measurement values with their expanded measurement uncertainties. In addition, all three measurement values have been calculated and represented. The respective uncertainties of the mean values are shown in the legend.

The three mean values differ by up to more than 1 µm. In spite of the large measurement uncertainty, the simple mean is influenced by participant B, whereas the weighted mean is drawn towards participant C due to the very low measurement uncertainty. The selection of the RV from these three possibilities will in any case affect the further evaluation, as will be shown in the following.

3. E_n values

The agreement of a measurement result with the reference result may be checked on the basis of the so-called E_n value, i.e. the normalized error. The basis here is the measured values and their measurement uncertainties. If $|E_n| \leqslant 1$, this means that the observed measurement value x_i and the RV x_{ref} are comparable, provided that the respective measurement uncertainties are taken into account. For correlated quantities in accordance with [13], the E_n value is calculated as follows:

$$E_n = \frac{1}{k}\frac{x_i - x_{ref}}{\sqrt{u_i^2 - u_{ref}^2}} \qquad (3)$$