



# Evaluation of inconsistent data of key comparisons of measurement standards

A. Chunovkina, N. Zviagin, N. Burmistrova

*D.I.Mendeleyev Institute for Metrology, 19, Moskovsky pr. St Petersburg, Russia*

## ABSTRACT

This paper deals with evaluation of inconsistent data obtained in key comparisons of national measurement standards. The concept “equivalence of measurement standards” is discussed in the context of metrological compatibility of measurement results. Applications of different methods for evaluating inconsistent data are briefly considered. An explicit practical approach is proposed for evaluating inconsistent data. It is illustrated by an analysis of the results of CCQM-K5.

**Section:** RESEARCH PAPER

**Keywords:** measurement model, inconsistent data, metrological compatibility, measurement uncertainty, degree of equivalence.

**Citation:** A.Chunovkina, N.Zviagin, N.Burmistrova, Evaluation of inconsistent data of key comparisons of measurement standards, Acta IMEKO, vol. 2, no. 2, article 6, December 2013, identifier: IMEKO-ACTA-02 (2013)-02-06

**Editor:** Paolo Carbone, University of Perugia

**Received** February 22<sup>th</sup>, 2013; **In final form** December 12<sup>th</sup>, 2013; **Published** December 2013

**Copyright:** © 2013 IMEKO. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

**Funding:** This work was done in the framework of the project “Kompetenzdreieck Optische Mikrosysteme – Spitzenforschung und Innovation in den neuen Ländern”, which is supported by the German Ministry of Education and Research (FKZ: 16SV5473).

**Corresponding author:** A. Chunovkina, e-mail: A.G.Chunovkina@vniim.ru

## 1. INTRODUCTION

Since the signing of the CIPM MRA [1], many different approaches have been proposed for analysing key comparison data. The consistency of measurement data and the models applied is a crucial point in choosing a correct method of data evaluation. In general, all methods can be divided into two large groups (see Figure 1).

The first group comprises methods for evaluating consistent data. These methods do not require any additional information apart from measurement results and associated uncertainties provided by participants. As a rule, the key comparison reference value (KCRV) is calculated as a weighted mean [2-3], and a degree of equivalence of measurement standards is established as the deviation of a measurement result from the KCRV and associated uncertainty of this deviation in full agreement with the CIPM MRA.

If detailed information about an uncertainty budget is available, the bias estimates for results of participating laboratories can be obtained [4]. These estimates should not be regarded as the degree of equivalence and no further corrections for systematic effects are implied. This additional information about systematic biases obtained from the joint evaluation of all comparison data can be used for improving a

measurement procedure used in every laboratory.

The second group comprises methods for evaluating inconsistent data. It should be stressed that all these methods are based on some additional assumptions [5-13], the validity of which for concrete comparison data should be analysed in every particular case of their usage.

The present paper deals with evaluating inconsistent data of a CIPM key comparison. An algorithm for calculating the KCRV and the degree of equivalence (DoEs) is suggested. The paper is divided into 4 main sections. In Section 2 some general consideration concerning the concept of equivalence of measurement standards is given. Section 3 presents a brief analysis of different models and algorithms which are used for the inconsistent data evaluation. In Section 4 the algorithm suggested is discussed. The application of the algorithm for analysing the CCQM-K5 data and the comparable analysis of this algorithm with other approaches are given in Section 5

## 2. EQUIVALENCE OF MEASUREMENT STANDARDS

The CIPM MRA does not explicitly determine the concept “equivalence of measurement standards” and therefore different interpretations of the concept are available. The MRA gives the explicit definition of a measure of equivalence, namely the degree of equivalence. According to the MRA “The degree

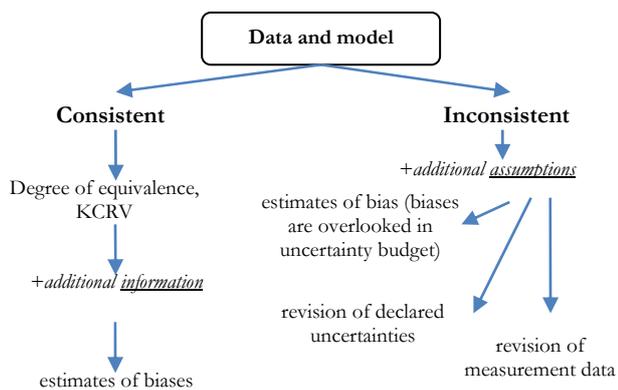


Figure 1. Two approaches for evaluating measurement data of key comparisons of national measurement standards.

of equivalence is taken to mean the degree to which these standards are consistent with reference values determined from the key comparisons and hence are consistent with one another. The degree of equivalence of each measurement standard is expressed quantitatively by two terms: its deviation from the key comparison reference value and the uncertainty of this deviation (at a 95% level of confidence)". Thus, in the definition of a degree of equivalence the basic accent is made on the definition of the key comparison reference value (KCRV).

Before discussing methods for determining the KCRV the authors would like to clarify the following issues:

- concept of the equivalence of measurement standards,
- meaning of the KCRV,
- equivalence of measurement standards in case of inconsistent data.

The authors share statements [2-3] that the equivalence of measurement standards means the equivalence of measurement results obtained in national metrology institutes (NMI's) participating in a key comparison. In this context the KCRV is understood as an estimate of a measurand based on measurement results provided by NMI's. Therefore, each NMI provides an estimate of the measurand and, if these data are consistent, the KCRV is calculated as a superior estimate of the same measurand.

To our opinion, the equivalence of measurement standards (or of measurement results) can be defined as the metrological compatibility of the following set of measurement results:  $\{(x_1, u_1), \dots, (x_n, u_n); (x_{ref}, u_{ref})\}$ . According to the VIM3 [14], the metrological compatibility is "a property of a set of measurement results for a specified measurand, such that the absolute value of the difference of any pair of measured quantity values from two different measurements results is smaller than some chosen multiple of standard measurement uncertainty of that difference". Often, in case of a key comparison, for a confidence level of 95% the multiple equals to 2. So, the equivalence of measurement standards means that for any  $i$  and  $j$  the following equations are satisfied:

$$|x_i - x_j| \leq 2u(x_i - x_j) \text{ and } |x_i - x_{ref}| \leq 2u(x_i - x_{ref}). \quad (1)$$

The above equations are similar to the equation used in an  $E_n$  criterion applied in the analysis of key comparison data. If there is no covariance between the results obtained in the  $i$ -th and  $j$ -th laboratory, then the uncertainty of the difference

between these results is given by the formula

$$u^2(x_i - x_j) = u^2(x_i) + u^2(x_j). \quad (2)$$

In case when the measurement of each laboratory is realized independently of the other institutes' measurements and the value is calculated via the weighted mean of measured values obtained by other laboratories, using the inverses of squares of the associated standard uncertainties as the weights, the uncertainty of the difference between a measurement result and a reference value is given by

$$u^2(x_i - x_{ref}) = u^2(x_i) - u^2(x_{ref}). \quad (3)$$

For equivalent measurement results the degree of equivalence is expressed by two terms:  $d_i = x_i - x_{ref}$  and  $u(x_i - x_{ref})$ . The authors would like to stress that the second term is even more informative than the first one, since it is to a great extent based on the uncertainty associated with the measurement results of a particular laboratory and characterises the dispersion of possible deviations of the measurement results of this laboratory from the reference value.

From the above consideration the authors can conclude the following:

- equivalence of measurement standards can be interpreted as the compatibility of a set of measurement results, which comprises the results obtained by NMI's and the KCRV with an associated uncertainty
- the KCRV is regarded as the estimate of a value of a specified measurand, which is based on consistent measurement results provided by laboratories
- equivalence of measurement standards implies the consistency of corresponding data. It directly follows from equations (1).

In general, there could be several reasons for the data inconsistency: instability or drift of a measurement standard, underestimating the measurement uncertainty by some participants or significant systematic hidden biases in some results. Methods for analysing the key comparison data in the presence of a linear drift of a travelling standard is beyond the scope of this paper [6].

The stability of the travelling standard is assumed in this paper. In practice the travelling standard stability is investigated by a pilot laboratory before sending the travelling standard to other comparison participants. The pilot laboratory often performs repeated measurements during the time of comparison with the aim to check the stability of the travelling standard.

Actually both the underestimated measurement uncertainty and hidden systematic errors of values measured can be regarded as a result of the incorrect uncertainty evaluation. In order to separate these two reasons, it is required to get additional information about a meaningful value of measurement uncertainty.

There are several approaches and a lot of algorithms for evaluating inconsistent data [5-13]. In the next section the authors will briefly discuss the models used in these approaches.

### 3. MODELING

This paper considers the evaluation of inconsistent data of key comparisons.

The basic model applied for the key comparison data

analysis is as follows. Each laboratory measures the same measurand  $X$ :

$$X_i = X. \quad (4)$$

and provides the measurement result  $x_i$  and associated standard uncertainty  $u_i, i=1, \dots, n$ . It is assumed that all laboratories have introduced corrections for known systematic biases and have provided a combined standard uncertainty.

If model (4) and data  $\{x_i, u_i\}$  are consistent, the conventional approach is applied [2-3]. In case of inconsistent data, the first that can be done is to form the largest consistent subset [3]. But sometimes, there could be difficulties in forming a single subset. A drawback of this approach is that for several laboratories (not included in the subset) the DoEs are not established. If the number of these laboratories is significant, the key comparison actually has to be repeated. There are two possibilities to remove the data inconsistency. The first one is to propose a more complicated measurement model, and the other one is to modify the measurement uncertainties associated with the results provided by NMIs.

Usually the following model is considered for processing inconsistent data:

$$X_i = X + B_i, \quad (5)$$

where  $B_i$  is the laboratory effect.

Under the fixed effects model  $B_i$  is treated as the laboratory systematic bias. Estimation of  $X, B_i$  requires additional information or some additional assumptions [10-12]. In [10] it is assumed that some laboratories measure without biases. Actually only the number of these laboratories should be specified, the identity of the laboratories is not presumed to be known. This assumption and application of Bayesian model averaging allows the estimates of the KCRV and those of the biases of the rest laboratories to be received. It should be noted that application of the fixed effect model for the key comparison data analysis is still an issue to be discussed because the relation between the DoEs and biases estimates is not explicitly identified.

Another application of model (5) (random effect model) consists in treating  $B_i$  as random variables [8, 13]. Usually they are assumed to be normally distributed with the zero mean and variance  $\sigma^2$ . The KCRV is calculated as the weighted mean,

$$X_{ref} = \frac{\sum \frac{x_i}{u_i^2 + \sigma^2}}{\sum \frac{1}{u_i^2 + \sigma^2}}. \quad (6)$$

The estimate of  $\sigma^2$  is based on measurement results provided by laboratories and can be obtained numerically. The DoEs can be calculated using simulation in order to take into account the correlation between the measurement data and estimate of  $\sigma^2$ .

The authors would like to discuss possible interpretations of the biases  $B_i$  in model (5) when it is applied for evaluating the KC data. They can describe the instability of the travelling standard. This instability is assumed to be significantly less than the measurement uncertainties declared by participants. So, usage of repeated measurements in a pilot laboratory seems more preferable for estimating  $\sigma^2$  than the usage of all results provided by the participants. Another interpretation of  $B_i$  is

considered in [8], where each laboratory underestimates the measurement uncertainty due to the fact that one and the same random error is hidden by all laboratories. Actually, this means that the laboratories have to recalculate the measurement uncertainties during the comparison in order to take into account the additional uncertainty component  $\sigma^2$ . In fact, it contradicts the CIPM KC rules [15], so in practice any increase of measurement uncertainties is not applied. On the other hand, in similar cases the uncertainty associated with the KCRV is actually enlarged when it is taken as the uncertainty of the weighted mean (6). Sometimes, when the measurement uncertainties seem to be underestimated by participants, the KCRV uncertainty is calculated as a sample standard deviation of the simple mean. It means that the uncertainties declared by participants are not used in establishing the KCRV and the associated uncertainty. Therefore, in such cases a confirmation of the declared uncertainties by the results of key comparisons becomes questionable.

The KC aims are to establish DoEs and to provide a foundation for calibration and measurement capabilities of the NMI. If the measurement uncertainties are not confirmed in the KC, the NMI can enlarge the uncertainties presented in the calibration and measurement capabilities so that the measurement results can be consistent with the KCRV [1].

Sometimes initial measurement results are mutually non-compatible because several participants have underestimated the measurement uncertainties rather than that a few results can be regarded as outliers. In such cases the increasing measurement uncertainties of these results seems to be reasonable. The procedure proposed below addresses exactly these cases.

The idea is to increase the measurement uncertainties associated with several results in order to make the results compatible with each other and with the KCRV too. Two questions arise:

- to what extent the measurement uncertainties may be increased;
- what uncertainties are to be enlarged.

The dispersion of the measured values reported by the participants contains valuable information. It should be noted that one cannot rely on the declared measurement uncertainties because the measurement data are inconsistent. The authors would like to distinguish the data having significant systematic biases from those of which the uncertainties are underestimated. Here the authors propose to use the criterion  $Z_n$  for constructing the homogeneous set of measured values. This set is indicated as reference group. The meaningful measurement uncertainty is calculated as sample standard deviation of values included in this set. The data beyond this group are regarded as having a significant systematic bias. Nevertheless they can be consistent with the KCRV if the associated measurement uncertainties are large enough. Otherwise these measurement results are not equivalent to the others and the DoEs are not established for them. Measurement uncertainties are enlarged only for those results from the reference group which are non-compatible with the KCRV determined initially as the weighted mean of all measurement results.

#### 4. EVALUATION PROCEDURE

The proposed method implies usage of the basic model (4).

All laboratories report measured values and standard measurement uncertainties  $\{x_i, u_i\}$ .

### Step 1: Analysis of metrological compatibility

The evaluation starts with analysis of compatibility of the set of measurement results:

$$\{(x_1, u_1), \dots, (x_n, u_n); (x_{ref}, u_{ref})\}.$$

Initially the reference value is calculated as the weighted mean of measurement results provided by all laboratories:

$$X_{ref} = \frac{\sum \frac{x_i}{u^2(x_i)}}{\sum \frac{1}{u^2(x_i)}}, \quad u(X_{ref}) = \left( \sum \frac{1}{u^2(x_i)} \right)^{-1}. \quad (7)$$

If the measurement results are compatible with each other and with the KCRV the conventional approach can be applied for calculating the DoEs. If the measurement results are not compatible one comes to step 2.

### Step 2: Determining the reference group

The reference group comprises the measurement results which pass the criterion  $Z_n$ :

$$Z_n = \frac{|x_i - \bar{x}|}{2S}, \quad \bar{x} = \frac{1}{m} \sum_{i=1}^m x_i, \quad S^2 = \frac{1}{m-1} \sum_{i=1}^m (x_i - \bar{x})^2. \quad (8)$$

The results with  $Z_n < 1$ , are included into the reference group;  $m$  ( $m < n$ ) is the number of results in the reference group.

### Step 3: Extending the measurement uncertainties

The sample standard deviation obtained using measured values from the reference group serves as recommended value of uncertainty for measurement results that are non-compatible with the reference value (Step 1). After enlarging measurement uncertainties for several results the compatibility of a set of modified data is analysed. The KCRV is recalculated as the weighted mean of measurement results from the reference group. New weights are used because some measurement uncertainties have been changed.

If compatibility with the reference value still has not been achieved for the measurement results from the reference group it might be reasonable not to calculate the KCRV. The results of KC can be reported by a matrix of pair wise degree of equivalence.

## 5. EXAMPLE

The proposed approach is illustrated by its application for CCQM-K5 data evaluation. The initial data are shown in Figure 2. The KCRV was calculated as the simple mean with associated uncertainty taken as sample standard deviation of the mean. All measurement results (except number 10) were used for calculation of the KCRV and associated uncertainty. The authors can conclude that the KCRV uncertainty is large compared with measurement uncertainties of some participants. It is explained by the fact that the KCRV uncertainty is not calculated using the measurement uncertainties reported by the participants.

Application of the conventional approach [2-3] reveals the inconsistency of measurement data. Table 1 contains the values characterizing the metrological compatibility of measurement

results,  $\left| \frac{x_i - x_j}{\sqrt{U_i^2 + U_j^2}} \right|$ , and their compatibility with the KCRV,

$\left| \frac{x_i - x_{ref}}{\sqrt{U_i^2 + U_{ref}^2}} \right|$ . If the measurement results are compatible, the

corresponding values should be less than 1. Result 10 can be identified as non-compatible with all others and with the KCRV (Table 1).

Below the authors present an analysis of the CCQM.K-5 data using the approach proposed. The reference group comprises all measurement results except that of number 10. The sample standard deviation equals  $\bar{u} = 0,026$ .

The expanded uncertainties associated with measurement results 1, 3, 5, 6, 8 and 10 were replaced by  $2\bar{u}$ . Then the KCRV was recalculated. Table 2 presents the compatibility of the modified measurement results. Most of the data revised demonstrate the compatibility of measurement results, i.e., with each other and with the reference value. However, the result of laboratory number 10 is non-compatible with others. Therefore, the degree of equivalence was not established for measurement results obtained in this laboratory.

Figure 3 presents the modified data after extending the measurement uncertainties for measurement results 1, 3, 5, 6, 8 and 10.

The approach given is compared with a random effect model as well as with the method used in the CCQM K-5 report. Comparison of different methods applied for the analysis of the CCQM.K-5 data is presented in Table 3 and Table 4. The approach proposed provides the smallest measurement uncertainty associated with the reference value.

Table 1. Metrological pair-wise compatibility of the measurement results. The last column presents compatibility with the KCRV. The bold figures indicate the pairs of measurement results which are non-compatible with each other or the KCRV.

	1	2	3	4	5	6	7	8	9	10	En
1		<b>1.07</b>	<b>1.72</b>	0.07	0.71	0.03	0.93	0.64	<b>1.37</b>	<b>3.90</b>	0.49
2	<b>1.07</b>		<b>1.01</b>	0.49	<b>2.27</b>	0.42	0.14	2.07	0.45	<b>3.55</b>	0.45
3	<b>1.72</b>	<b>1.01</b>		0.89	<b>2.58</b>	0.75	0.69	<b>2.46</b>	0.63	<b>1.69</b>	<b>1.22</b>
4	0.07	0.49	0.89		0.20	0.05	0.52	0.18	0.63	<b>1.70</b>	0.29
5	0.71	<b>2.27</b>	<b>2.58</b>	0.20		0.34	<b>1.66</b>	0.05	2.50	5.53	1.25
6	0.06	0.85	1.50	0.10	0.68		0.79	0.62	<b>1.13</b>	<b>3.35</b>	0.38
7	0.93	0.14	0.69	0.52	<b>1.66</b>	0.39		<b>1.57</b>	0.19	<b>2.43</b>	0.47
8	0.64	<b>2.07</b>	<b>2.46</b>	0.18	0.05	0.31	<b>1.57</b>		<b>2.31</b>	<b>5.19</b>	<b>1.16</b>
9	<b>1.37</b>	0.45	0.63	0.63	2.50	0.56	0.19	<b>2.31</b>		<b>2.87</b>	0.78
10	<b>3.90</b>	<b>3.55</b>	<b>1.69</b>	<b>1.70</b>	<b>5.53</b>	<b>1.68</b>	<b>2.43</b>	<b>5.19</b>	<b>2.87</b>		<b>3.23</b>

Table 2. Metrological compatibility of the modified measurement results. The uncertainties associated with measurement results 1, 3, 5, 6, 8 and 10 were extended.

	1	2	3	4	5	6	7	8	9	10	En
1		0.50	0.76	0.06	0.24	0.03	0.53	0.23	0.68	<b>1.47</b>	0.52
2	0.50		0.54	0.49	0.83	0.46	0.14	0.82	0.45	<b>1.50</b>	0.03
3	0.76	0.54		0.74	<b>1.01</b>	0.73	0.43	0.99	0.35	0.71	0.57
4	0.06	0.49	0.74		0.16	0.08	0.52	0.15	0.63	<b>1.37</b>	0.50
5	0.24	0.83	<b>1.01</b>	0.16		0.27	0.84	0.01	<b>1.00</b>	<b>1.71</b>	0.87
6	0.03	0.46	0.73	0.08	0.27		0.50	0.26	0.64	<b>1.44</b>	0.48
7	0.53	0.14	0.43	0.52	0.84	0.50		0.83	0.19	<b>1.32</b>	0.17
8	0.23	0.82	0.99	0.15	0.01	0.26	0.83		0.99	<b>1.70</b>	0.85
9	0.68	0.45	0.35	0.63	<b>1.00</b>	0.64	0.19	0.99		<b>1.30</b>	0.69
10	<b>1.47</b>	<b>1.50</b>	0.71	<b>1.37</b>	<b>1.71</b>	<b>1.44</b>	<b>1.32</b>	<b>1.70</b>	<b>1.30</b>		<b>1.58</b>

Table 3. Comparison of the approaches applied for the analysis of the CCQM K-5 data. Calculation KCRV.

CCQM-K5		Given approach		Random effect model	
KCRV	Uncertainty	KCRV	Uncertainty	KCRV	Uncertainty
1.513	0.023	1.5247	0.0083	1.5213	0.0243

Table 4. Comparison of the approaches applied for the analysis of the CCQM K-5 data. . Calculation DoEs.

	CCQM-K5		Given approach		Random effect model	
	D <sub>i</sub>	U(D <sub>i</sub> )	D <sub>i</sub>	U(D <sub>i</sub> )	D <sub>i</sub>	U(D <sub>i</sub> )
1	-0.015	0.029	-0.0267	0.0514	<b>-0.0233</b>	<b>0.0227</b>
2	0.012	0.025	0.0003	0.0113	0.0037	0.0129
3	<b>0.041</b>	<b>0.032</b>	0.0293	0.0514	<b>0.0328</b>	<b>0.0195</b>
4	-0.020	0.067	-0.0317	0.0635	-0.0282	0.0602
5	<b>-0.033</b>	<b>0.025</b>	-0.0447	0.0514	<b>-0.0413</b>	<b>0.0128</b>
6	-0.013	0.031	-0.0247	0.0514	<b>-0.0213</b>	<b>0.0189</b>
7	0.016	0.033	0.0043	0.0246	0.0077	0.0262
8	<b>-0.032</b>	<b>0.026</b>	-0.0437	0.0514	<b>-0.0402</b>	<b>0.0165</b>
9	0.022	0.026	0.0103	0.0148	0.0137	0.0170
10	<b>0.093</b>	<b>0.026</b>	<b>0.0813</b>	<b>0.0514</b>	<b>0.0848</b>	<b>0.0151</b>

The KCRV values inherent in these three methods are compatible. But the DoEs significantly differ depending on the method used. The approach given provides the compatibility of the measurement results and reference value for all data except result 10. As stated in the CCQM.K-5 report the significant

systematic bias was revealed in this result. The random effect model method and the method used in the CCQM.K-5 report do not provide mutually compatibility of the results. It is clearly seen by the fact that the corresponding DoEs exceed the values of the associated expanded uncertainties.

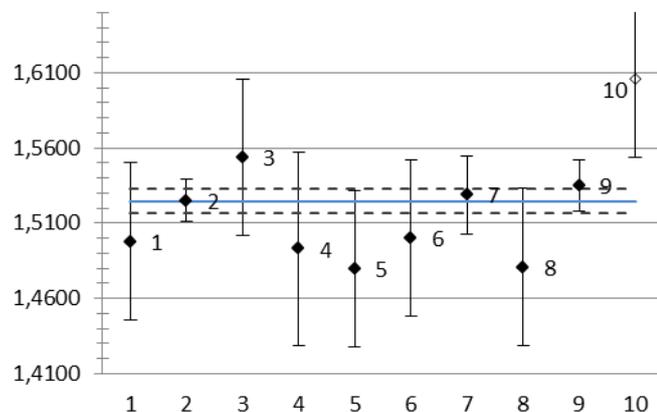


Figure 3. Modified data. The error bar shows the expanded uncertainty. The solid line indicates the recalculated KCRV and the dashed lines indicate the expanded uncertainty associated with the KCRV.

## 6. CONCLUSION

The paper is devoted to a discussion of issues concerning the evaluation of inconsistent data of key comparisons of measurement standards. The approach proposed is based on interpreting the equivalence of measurement standards as the compatibility of a set of measurement results, which comprises the data provided by NMI's and the reference value. The approach implies the extension of measurement uncertainties for several results provided by participants, which do not show the compatibility with the reference value.

Extension of measurement uncertainties during the processing of results is always a questionable point and requires a sound foundation. The authors consider that a significant argument is that the key comparisons are used to confirm the calibration and measurement capabilities of NMI's. These capabilities should be in agreement with the KC results.

Otherwise, this NMI should modify the measurement uncertainties declared in its calibration and measurement capabilities (after comparison) to make them consistent with the results of the comparison. If the measurement data reported by KC participants show the non-compatibility and no outliers can be identified unambiguously, the extension of measurement uncertainties for several measurement results seems to be reasonable.

Advantages of the method proposed are the following. The KC data analysis results in obtaining a set of compatible measurement results. The results are compatible with each other and with the KCRV calculated using these results. The interpretation of measurement standard equivalence as the metrological compatibility of a set of measurement results is explicit. This interpretation corresponds to the practical usage of KC results as objective foundation for mutual recognition of measurement results obtained in the participating labs.

The comparison of the procedure proposed with other approaches for evaluating inconsistent data is discussed and is illustrated by the analysis of CCQM.K-5. It is important to stress that there is no single method for analysing inconsistent data of KC's. Application of any method requires a preliminary consideration of the data reported by participants and a clear interpretation of a model to be chosen for a particular KC.

## REFERENCES

- [1] CIPM MRA. Bureau International des Poids et Mesures 1999.
- [2] Cox M G "The evaluation of key comparison data: An introduction" 2002 *Metrologia* 39, pp. 587-8.
- [3] Cox M G "The evaluation of key comparison data" 2002 *Metrologia* 39, pp. 589-95.
- [4] Chunovkina A. G., Elster C., Lira I., Woeger W. "Analysis of key comparison data and laboratory biases", 2008 *Metrologia* 45, pp. 211-6.
- [5] Lira I. "Combining inconsistent data from interlaboratory comparisons", 2007 *Metrologia* 44, pp. 415-21.
- [6] Zhang N F, Liu H, Sedransk N and Strawderman W E "Statistical analysis of key comparisons with linear trends" 2004 *Metrologia* 41, pp. 231-7.
- [7] The authorsise K. and Wöger W. "Removing model and data non-conformity in measurement evaluation", 2000, *Meas. Sci. Technol.* 11, pp. 1649-58.
- [8] J. Willink R. "Statistical determination of a comparison reference value using hidden errors", 2002 *Metrologia* 39, pp. 343-54.
- [9] Cox M. G. "The evaluation of key comparison data: determining the largest consistent subset", 2007 *Metrologia* 44, pp. 187-200.
- [10] Elster C., Toman B. "Analysis of key comparisons: estimating laboratories' biases by a fixed effects model using Bayesian model averaging", 2010, *Metrologia* 47, pp. 113-119.
- [11] White D R "On the analysis of measurement comparisons" 2004 *Metrologia* 41, pp. 122-31.
- [12] Sutton C M "Analysis and linking of international measurement comparisons" 2004 *Metrologia* 41, pp. 272-7.
- [13] Toman B, Possolo A "Laboratory effects models for interlaboratory comparisons", 2009 *Accreditation and Quality Assurance* 14, pp. 553-563.
- [14] International Vocabulary of Metrology – Basic and General Concepts and Associated Terms, 3rd edition, 2008 version with minor corrections, BIPM, JCGM 200, 2012.
- [15] CIPM MRA-D-05 Measurement comparisons in the CIPM MRA (<http://kcdb.bipm.org/>).