urn:nbn:de:gbv:ilm1-2011imeko-036:0

Joint International IMEKO TC1+ TC7+ TC13 Symposium
August 31st – September 2nd, 2011, Jena, Germany
urn:nbn:de:gbv:ilm1-2011imeko:2

LARGEST CONSISTENT SUBSETS IN INTERLABORATORY COMPARISONS: PREFERENCE AGGREGATION APPROACH

Sergey V. Muravyov[1] *and Irina A. Marinushkina*[1]

[1]Department of Computer-aided Measurement Systems and Metrology,
Tomsk Polytechnic University, Tomsk, Russian Federation
[2]Department of Manufacturing Engineering, Karaganda State Technical University,
Karaganda, Republic of Kazakhstan

Abstract – Interlaboratory comparisons need a reference value of the measurand to be assigned. It is necessary to have some procedure that allows to determine the reference value at a maximum number of participating laboratories results to be included into the determination and, at the same time, unreliable laboratory results must be disregarded. It is shown in the paper that this procedure can be implemented using a preference aggregation approach.

Keywords: interlaboratory comparisons, preference aggregation, consensus relation

1. INTRODUCTION

When organizing interlaboratory comparisons for proficiency testing (see, for example, [1]) the main task is usually determination of the *reference value* x_{ref} and its uncertainty range. Let the participating in comparisons laboratories be measuring nominally the same quantity Y, that is

$$Y = X_i,\ i = 1,\ \ldots,\ N, \tag{1}$$

where X_i is the quantity measured by the i-th laboratory and N is a number of laboratories participating in a comparison. The aim of the comparisons is to determine an estimate y of Y and the associated uncertainty $u(y)$ in terms of estimates x_i of the X_i provided by the laboratories and their associated standard uncertainties $u(x_i)$.

The estimate y is typically calculated as a weighted mean by formula

$$y = \sum_{i=1}^{N'} \frac{x_i}{u(x_i)^2} \Bigg/ \sum_{i=1}^{N'} \frac{1}{u(x_i)^2} \tag{2}$$

and corresponding uncertainty is

$$u(y) = \sqrt{1 \Big/ \sum_{i=1}^{N'} \frac{1}{u(x_i)^2}}, \tag{3}$$

where N' is the number of laboratories, results of which deem to be reliable.

It is follows from expressions (2) and (3) that a procedure of determination of the reference value must provide a highest possible consistency of the participating laboratories measurement results x_i, that is to allow shaping a subset of maximal possible power N' of laboratories providing reliable results (so called *largest consistent subset*, LCS). Therewith, this procedure must facilitate identifying unreliable results and subsequent elimination of corresponding laboratories from the set of comparison participants. This decision takes place, if the following condition is valid:

$$E_n = \frac{|\, x_i - x_{\text{ref}}\,|}{\sqrt{u(x_i)^2 + u(y)^2}} > 1. \tag{4}$$

Different ways of the problem solving are described in many publications. For example, in [2] statistical criteria testing the consistency assumption are analyzed in conjunction with full enumeration when building the LCS(s). In paper [3] an algebraic approach has been proposed giving the procedure LCS determination of polynomial complexity. In [4], it was proposed to consider the uncertainty range $u(x_i)$ as the rectangular distribution and to deem that each participant gives one vote to each value within its uncertainty range and no votes for values outside this range. This produces a robust algorithm that is insensitive to outliers, i.e. results with the uncertainty considerably lower than those of other participants. The algorithm successful application has been described in [5].

This paper is devoted to the problem solving in terms of *preference aggregation*.

2. PREFERENCE AGGREGATION APPROACH

Designate the uncertainty range $[-u(x_i), u(x_i)]$ obtained by i-th laboratory through I_i. Define the *range of actual values* of measured quantity as algebraic union of the uncertainty ranges obtained by each laboratory:

$$U = \bigcup_{i=1}^{N} [-u(x_i), u(x_i)] = \bigcup_{i=1}^{N} I_i. \tag{5}$$

Let us partition this range into equal intervals in such a way that their number ensures *enough accuracy* of the measured quantity values representation. Then we have n values of the measured quantity $\{a_1, a_2, \ldots,$

$a_n\} = A$, corresponding to the interval bounds that will play part of alternatives (candidates) in the consensus relation determination problem. The laboratories will play part of voters in the problem.

Let us compose the preference profile Λ which will consist of rankings describing the uncertainty ranges of each laboratory. Each k-th rankings is the union of binary relations of strict order and equivalence possessing the following properties:

i) $a_i \succ a_j$, if $a_i \in I_k$ and $a_j \notin I_k$

ii) $a_i \sim a_j$, if $a_i, a_j \in I_k$ or $a_i, a_j \notin I_k$,

whence it follows that each ranking includes one pair of alternatives with the strict order relation and $n - 2$ pairs with the equivalence relation.

Our aim is to determine a single preference relation that would give an integrative characterization of the alternatives. On the basis of a Kemeny distance (see, for instance, [6]) between the rankings, we can define the distance $D(\beta, \Lambda)$ from β to the profile Λ and then formulate a consensus relation determination problem as

$$\beta = \arg\min_{\lambda \in \Pi} D(\lambda, \Lambda), \quad (6)$$

where $D(\lambda, \Lambda) = \sum_{k=1}^{m} d(\lambda, \lambda_k)$ is Kemeny distance; $\{\lambda_1, ..., \lambda_m\} = \Lambda$ is the preference profile consisting of rankings shaped by the laboratories λ_k; Π is a set of all $n!$ linear orders $\succ$ on A; β is the *consensus ranking* (Kemeny median). Each linear order corresponds to one of permutations of first n natural numbers N_n.

We will use the recursive branch and bound algorithm for finding a solution of the problem (6) that was described in [7]. In the present version, the algorithm is modified in such a way to be able to find not only one but *all* solutions of the problem (6). In this case, the multiple optimal solutions must be convolute into a single one. In each individual situation the problem can be resolved on the basis of some rational considerations. The situations are demonstrated below by particular examples.

3. EXAMPLES

Consider how the proposed procedure works by the examples taken from [3-5].

3.1. Example 1

This example [4] represents some model data where results of a reference object length measurements made by 9 laboratories are shown in Fig. 1. The range of actual length values of is reduced to Table 1.

Rankings corresponding to uncertainty ranges of the nine laboratories produce the preference profile as follows:

λ_1: $a_8 \sim a_9 \succ a_1 \sim a_2 \sim a_3 \sim a_4 \sim a_5 \sim a_6 \sim a_7 \sim a_{10} \sim a_{11} \sim a_{12}$

λ_2: $a_3 \sim a_4 \sim a_5 \sim a_6 \sim a_7 \sim a_8 \sim a_9 \sim a_{10} \sim a_{11} \succ a_1 \sim a_2 \sim a_{12}$

λ_3: $a_8 \sim a_9 \succ a_1 \sim a_2 \sim a_3 \sim a_4 \sim a_5 \sim a_6 \sim a_7 \sim a_{10} \sim a_{11} \sim a_{12}$

λ_4: $a_6 \sim a_7 \sim a_8 \succ a_1 \sim a_2 \sim a_3 \sim a_4 \sim a_5 \sim a_9 \sim a_{10} \sim a_{11} \sim a_{12}$

λ_5: $a_2 \sim a_3 \sim a_4 \sim a_5 \sim a_6 \sim a_7 \sim a_8 \sim a_9 \sim a_{10} \succ a_1 \sim a_{11} \sim a_{12}$

λ_6: $a_4 \sim a_5 \sim a_6 \sim a_7 \sim a_8 \sim a_9 \sim a_{10} \sim a_{11} \succ a_1 \sim a_2 \sim a_3 \sim a_{12}$

λ_7: $a_8 \sim a_9 \succ a_1 \sim a_2 \sim a_3 \sim a_4 \sim a_5 \sim a_6 \sim a_7 \sim a_{10} \sim a_{11} \sim a_{12}$

λ_8: $a_6 \sim a_7 \sim a_8 \sim a_9 \sim a_{10} \succ a_1 \sim a_2 \sim a_3 \sim a_4 \sim a_5 \sim a_{11} \sim a_{12}$

λ_9: $a_3 \sim a_4 \sim a_5 \sim a_6 \sim a_7 \sim a_8 \sim a_9 \sim a_{10} \succ a_1 \sim a_2 \sim a_{11} \sim a_{12}$.

For the given profile, the branch and bound algorithm found the following eight optimal solutions of equal worth:

$a_8 \succ a_9 \succ a_6 \succ a_7 \succ a_{10} \succ a_4 \succ a_5 \succ a_3 \succ a_{11} \succ a_2 \succ a_1 \succ a_{12}$

$a_8 \succ a_9 \succ a_6 \succ a_7 \succ a_{10} \succ a_4 \succ a_5 \succ a_3 \succ a_{11} \succ a_2 \succ a_{12} \succ a_1$

$a_8 \succ a_9 \succ a_6 \succ a_7 \succ a_{10} \succ a_5 \succ a_4 \succ a_3 \succ a_{11} \succ a_2 \succ a_1 \succ a_{12}$

$a_8 \succ a_9 \succ a_6 \succ a_7 \succ a_{10} \succ a_5 \succ a_4 \succ a_3 \succ a_{11} \succ a_2 \succ a_{12} \succ a_1$

$a_8 \succ a_9 \succ a_7 \succ a_6 \succ a_{10} \succ a_4 \succ a_5 \succ a_3 \succ a_{11} \succ a_2 \succ a_1 \succ a_{12}$

$a_8 \succ a_9 \succ a_7 \succ a_6 \succ a_{10} \succ a_4 \succ a_5 \succ a_3 \succ a_{11} \succ a_2 \succ a_{12} \succ a_1$

$a_8 \succ a_9 \succ a_7 \succ a_6 \succ a_{10} \succ a_5 \succ a_4 \succ a_3 \succ a_{11} \succ a_2 \succ a_1 \succ a_{12}$

As, in the obtained solutions, the relations $a_4 \succ a_5$ and $a_5 \succ a_4$ occur the same number of times we conclude that $a_5 \sim a_4$. Similar considerations give $a_6 \sim a_7$ and $a_1 \sim a_{12}$. Then the final consensus relation is

$$\beta = \{a_8 \succ a_9 \succ a_6 \sim a_7 \succ a_{10} \succ a_4 \sim a_5 \succ a_3 \succ a_{11} \succ a_2 \succ a_1 \sim a_{12}\}.$$

The winner is the alternative a_8 which corresponds to the value 1.77395 inches (red line in Fig. 1). It is selected as the reference value. It is clear from Fig. 1 that this value provides maximal consistency of the laboratory results. This result coincides with the outcome of Nielsen's analysis in [4] implemented by him using so called "Value Voted Most Likely To Be Correct" algorithm.

3.2. Example 2

Second example is based on the data of microwave power measurement exercise with aim to find if there are laboratories needing a revision of their measurement techniques and methods [5]. The twelve laboratories presented their results of the calibration factor measurements collected for a standard power sensor at the frequency 1 GHz as shown in Fig. 2.

Table 1

Twelve alternatives of the actual length values range, inches

a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}
1.77430	1.77425	1.77420	1.77415	1.77410	1.77405	1.77400	1.77395	1.77390	1.77385	1.77380	1.77375

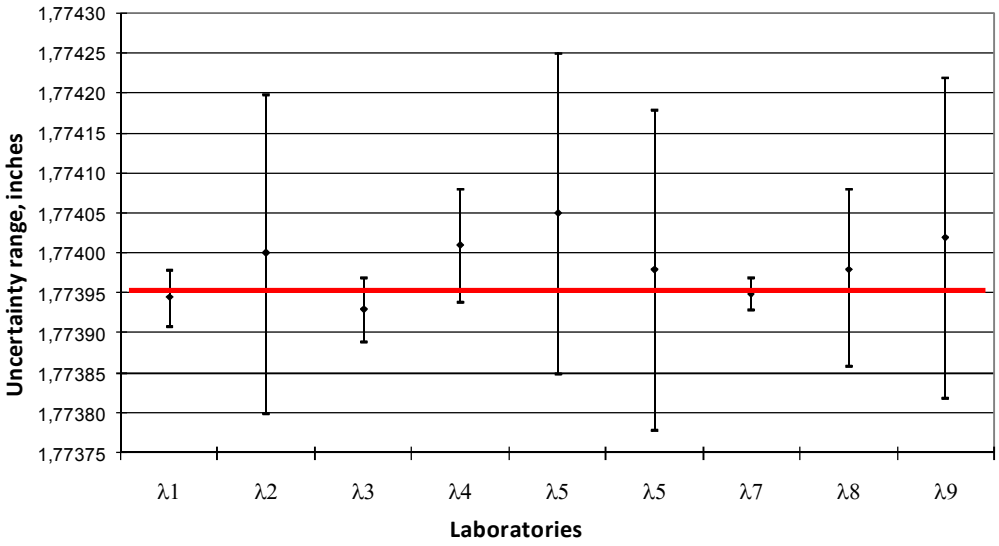

Рис. 1. Measurement results of 9 laboratories for proficiency testing

Rankings corresponding to uncertainty ranges of the 12 laboratories for the calibration factor of the travelling standard produce the preference profile as follows:

λ_1: $a_5 \sim a_6 \sim a_7 \succ a_1 \sim a_2 \sim a_3 \sim a_4 \sim a_8 \sim a_9 \sim a_{10}$

λ_2: $a_5 \sim a_6 \succ a_1 \sim a_2 \sim a_3 \sim a_4 \sim a_7 \sim a_8 \sim a_9 \sim a_{10}$

λ_3: $a_5 \sim a_6 \sim a_7 \succ a_1 \sim a_2 \sim a_3 \sim a_4 \sim a_8 \sim a_9 \sim a_{10}$

λ_4: $a_3 \sim a_4 \sim a_5 \sim a_6 \sim a_7 \sim a_8 \sim a_9 \succ a_1 \sim a_2 \sim a_{10}$

λ_5: $a_5 \sim a_6 \sim a_7 \succ a_1 \sim a_2 \sim a_3 \sim a_4 \sim a_8 \sim a_9 \sim a_{10}$

λ_6: $a_4 \sim a_5 \sim a_6 \sim a_7 \sim a_8 \sim a_9 \succ a_1 \sim a_2 \sim a_3 \sim a_{10}$

λ_7: $a_5 \sim a_6 \sim a_7 \sim a_8 \succ a_1 \sim a_2 \sim a_3 \sim a_4 \sim a_9 \sim a_{10}$

λ_8: $a_4 \sim a_5 \sim a_6 \sim a_7 \succ a_1 \sim a_2 \sim a_3 \sim a_8 \sim a_9 \sim a_{10}$

λ_9: $a_6 \sim a_7 \succ a_1 \sim a_2 \sim a_3 \sim a_4 \sim a_5 \sim a_8 \sim a_9 \sim a_{10}$

λ_{10}: $a_4 \sim a_5 \sim a_6 \sim a_7 \succ a_1 \sim a_2 \sim a_3 \sim a_8 \sim a_9 \sim a_{10}$

λ_{11}: $a_2 \sim a_3 \sim a_4 \succ a_1 \sim a_5 \sim a_6 \sim a_7 \sim a_8 \sim a_9 \sim a_{10}$

λ_{12}: $a_4 \sim a_5 \sim a_6 \sim a_7 \succ a_1 \sim a_2 \sim a_3 \sim a_8 \sim a_9 \sim a_{10}$.

For the given profile, we have again eight optimal solutions:

$a_6 \succ a_5 \succ a_7 \succ a_4 \succ a_8 \succ a_3 \succ a_9 \succ a_2 \succ a_1 \succ a_{10}$

$a_6 \succ a_5 \succ a_7 \succ a_4 \succ a_8 \succ a_3 \succ a_9 \succ a_2 \succ a_{10} \succ a_1$

$a_6 \succ a_5 \succ a_7 \succ a_4 \succ a_8 \succ a_9 \succ a_3 \succ a_2 \succ a_1 \succ a_{10}$

$a_6 \succ a_5 \succ a_7 \succ a_4 \succ a_8 \succ a_9 \succ a_3 \succ a_2 \succ a_{10} \succ a_1$

$a_6 \succ a_7 \succ a_5 \succ a_4 \succ a_8 \succ a_3 \succ a_9 \succ a_2 \succ a_1 \succ a_{10}$

$a_6 \succ a_7 \succ a_5 \succ a_4 \succ a_8 \succ a_3 \succ a_9 \succ a_2 \succ a_{10} \succ a_1$

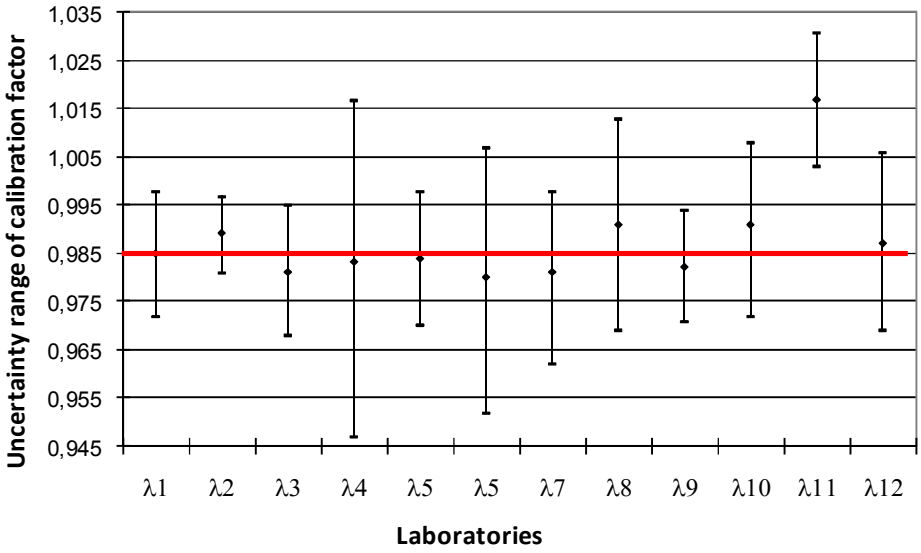

Рис. 2. Measurement results of 12 laboratories for the calibration factor of the travelling standard

$a_6 \succ a_7 \succ a_5 \succ a_4 \succ a_8 \succ a_9 \succ a_3 \succ a_2 \succ a_1 \succ a_{10}$

$a_6 \succ a_7 \succ a_5 \succ a_4 \succ a_8 \succ a_9 \succ a_3 \succ a_2 \succ a_{10} \succ a_1$.

Applying an analysis similar to that in section 3.1, we have the final consensus relation:

$\beta = \{a_6 \succ a_5 \sim a_7 \succ a_4 \succ a_8 \succ a_3 \sim a_9 \succ a_1 \sim a_{10}\}$.

The reference value is equal to a_6 = 0.985. It is evident from Fig. 2 that the only laboratory λ_{11} will be not included into the LCS as its uncertainty range is not compatible with other results.

Reference [5] authors used Nielsen's algorithm for determination of the LCS and obtained the same outcome.

3.3. Example 3

Finally, let us consider a third example taken from [3] where, among others, a length key comparison involving eleven National Metrological Institutes (NMIs) has been considered. The measurement results of a tungsten carbide gauge block of nominal length 1 mm using interferometry are shown in Fig. 3.

Rankings corresponding to uncertainty ranges of the 11 laboratories give the following preference profile:

λ_1: $a_4 \sim a_5 \succ a_1 \sim a_2 \sim a_3 \sim a_6 \sim a_7 \sim a_8$

λ_2: $a_3 \sim a_4 \sim a_5 \succ a_1 \sim a_2 \sim a_6 \sim a_7 \sim a_8$

λ_3: $a_2 \sim a_3 \sim a_4 \succ a_1 \sim a_5 \sim a_6 \sim a_7 \sim a_8$

λ_4: $a_3 \sim a_4 \sim a_5 \succ a_1 \sim a_2 \sim a_6 \sim a_7 \sim a_8$

λ_5: $a_3 \sim a_4 \succ a_1 \sim a_2 \sim a_5 \sim a_6 \sim a_7 \sim a_8$

λ_6: $a_7 \succ a_1 \sim a_2 \sim a_3 \sim a_4 \sim a_5 \sim a_6 \sim a_8$

λ_7: $a_7 \succ a_1 \sim a_2 \sim a_3 \sim a_4 \sim a_5 \sim a_6 \sim a_8$

λ_8: $a_2 \sim a_3 \succ a_1 \sim a_4 \sim a_5 \sim a_6 \sim a_7 \sim a_8$

λ_9: $a_4 \sim a_5 \succ a_1 \sim a_2 \sim a_3 \sim a_6 \sim a_7 \sim a_8$

λ_{10}: $a_5 \sim a_6 \succ a_1 \sim a_2 \sim a_3 \sim a_4 \sim a_7 \sim a_8$

λ_{11}: $a_4 \succ a_1 \sim a_2 \sim a_3 \sim a_5 \sim a_6 \sim a_7 \sim a_8$.

For the given profile, eight optimal solutions are as follows:

$a_4 \succ a_3 \succ a_5 \succ a_2 \succ a_7 \succ a_6 \succ a_8 \succ a_1$

$a_4 \succ a_3 \succ a_5 \succ a_2 \succ a_7 \succ a_8 \succ a_6 \succ a_1$

$a_4 \succ a_3 \succ a_5 \succ a_7 \succ a_2 \succ a_6 \succ a_8 \succ a_1$

$a_4 \succ a_3 \succ a_5 \succ a_7 \succ a_2 \succ a_8 \succ a_6 \succ a_1$

$a_4 \succ a_5 \succ a_3 \succ a_2 \succ a_7 \succ a_6 \succ a_8 \succ a_1$

$a_4 \succ a_5 \succ a_3 \succ a_2 \succ a_7 \succ a_8 \succ a_6 \succ a_1$

$a_4 \succ a_5 \succ a_3 \succ a_7 \succ a_2 \succ a_6 \succ a_8 \succ a_1$

$a_4 \succ a_5 \succ a_3 \succ a_7 \succ a_2 \succ a_8 \succ a_6 \succ a_1$

After apllication of the convolution procedure we have the final consensus relation:

$\beta = \{a_4 \succ a_3 \sim a_5 \succ a_2 \sim a_7 \succ a_6 \sim a_8 \succ a_1\}$.

One can see that a_4 = 20 nm is the winner and should be assigned to be the reference value. Measurement results of NMIs λ_6 and λ_7 should be excluded due to condition (4). This outcome is similar to that obtained by M. Cox in [3].

4. CONCLUSION

In the paper a procedure of the largest consistent subset determination for interlaboratory comparisons has been proposed on the basis of preference aggregation approach. Experimental verification of the algo-

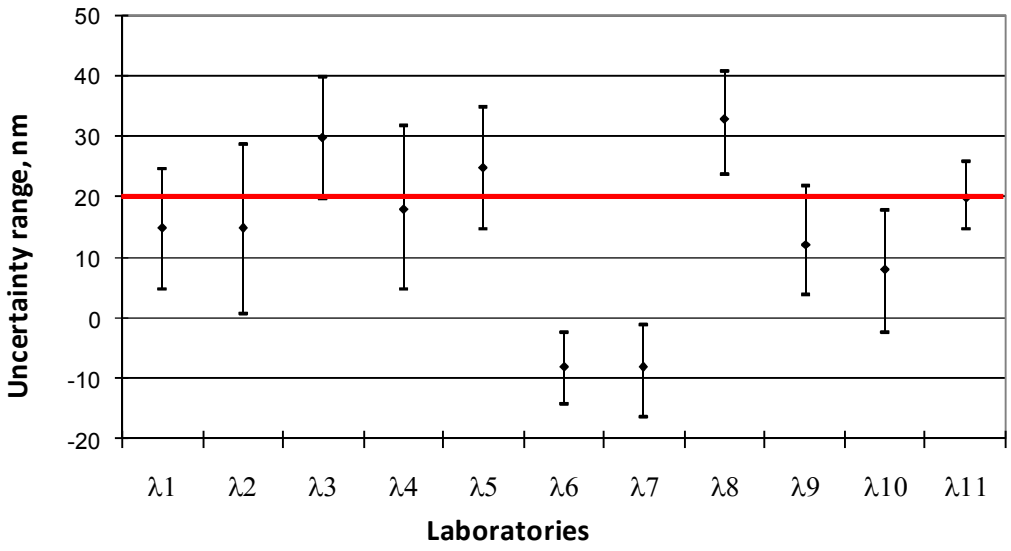

Рис. 3. Measurement results of 11 NMIs for the deviation from nominal length of a gauge block

rithm on comparison data given in various publications [3-5] confirms its correctness. In contrast to other approaches, it provides additional information about interrelations of values in the actual values range. For instance, from the solution in Example 1 (section 3.1) one can conclude that values a_6 and a_7 are equivalent to each other from the view point of their contribution into the final reference value and, at the same time, they both are less important for the reference value than the value a_8. This information, somehow or other, can be taken into account when analyzing the comparison results. The paper outcomes demonstrate the possibility of useful application of preference aggregation methods in metrological practice.

ACKNOWLEDGEMENT

This work was supported in part by the Ministry of Education and Science of Russian Federation under grant No. NK-566P/13 and by the Russian Foundation for Basic Research under grant No 10-08-90900-mob_sng_st.

REFERENCES

[1] M.A. Collett, M.G. Cox, T.J. Esward, P.M. Harris, J.A. Sousa, "Aggregating measurement data influenced by common effects", *Metrologia*, 44 (2007), pp. 308–318.

[2] H.K. Iyer, C.M. Wang, D.F. Vecchia, "Consistency tests for key comparison data", *Metrologia*, 41 (2004), pp. 223–230.

[3] M.G. Cox, "The evaluation of key comparison data: determining the largest consistent subset", *Metrologia*, 44 (2007), pp. 187–200.

[4] H.S. Nielsen, "Determining consensus values in interlaboratory comparisons and proficiency testing", *Proceedings of NCSL International Workshop and Symposium*, August 17-18, Tampa, Florida, USA, 2003, pp.1-16.

[5] L. Brunetti, L. Oberto, M. Sellone, P. Terzi, "Establishing reference value in high frequency power comparisons", *Measurement*, 42, (2009), pp. 1318–1323.

[6] S.V. Muravyov, V. Savolainen, "Special interpretation of formal measurement scales for the case of multiple heterogeneous properties", *Measurement*, Vol. 29, 2001, pp. 209-223.

[7] S.V. Muravyov, "Rankings as ordinal scale measurement results", *Metrology and Measurement Systems*, 2007, Vol. 13, No. 1, pp. 9-24.

Authors:

Sergey V. Muravyov, Department of Computer-aided Measurement Systems and Metrology, Tomsk Polytechnic University, Pr. Lenina, 30, Tomsk, 634050, Russian Federation, phone: +7-3822-417527; fax: +7-3822-420449, e-mail: muravyov@camsam.tpu.ru

Irina A. Marinushkina, Department of Manufacturing Engineering, Karaganda State Technical University, Bulvar Mira, 56, Karaganda, 100027, Republic of Kazakhstan, phone/fax: +7-7212-560328, e-mail: irma31@inbox.ru