XIX IMEKO World Congress
Fundamental and Applied Metrology
September 6–11, 2009, Lisbon, Portugal

METROLOGICAL INSIGHTS FROM INTERNATIONAL COMPARISON DATA

Alan G Steele [1], *A Peruzzi* [2], *J E Decker* [1], *R J Douglas* [1]

[1] National Research Council of Canada - Institute for National Measurement Standards (NRC-INMS), Ottawa, Canada, alan.steele@nrc-cnrc.gc.ca
[2] VSL, Dutch Metrology Institute, Delft, Netherlands, aperuzzi@nmi.nl

Abstract – The full suite of tools for the evaluation of statistical consistency of metrological data is applied to the data set compiled from two recent international comparisons: CCT-K7 and EURAMET.T-K7. This paper illustrates how insights into lab measurement characteristics such as repeatability, reproducibility, secular uncertainties and laboratory equivalence statements can be obtained in a statistically-rigorous manner.

Keywords: comparison analysis, linking, statistical methods

1. INTRODUCTION

Amongst the most important principle techniques in thermometry is the triple point of water, therefore considerable effort is devoted to understanding metrological characteristics of the cells and measurement techniques. International comparisons CCT-K7 [1] and EURAMET.T-K7 [2] are both devoted to triple point of water determinations. Eight labs participated in both comparisons, ensuring strong linking and offering a unique opportunity to explore consistency through large redundancy. Results from 5 of the common participants were selected for performing linking.

Conventional methods of measurement rely on comparing with a reference which has smaller uncertainty, typically 4 times smaller. At the highest-accuracy frontiers of metrology it is a challenge to have a similar type of a validation of measurement results. A valuable role of measurement science consists of developing methods with which to treat cases where no reference is available. In this regard, *unmediated* comparisons via pairs of peer measurements can be evaluated and aggregated with statistically-rigorous variations on familiar tools: E_n and χ^2. Monte Carlo simulation can rigorously extend these tools into regions where departures from the traditional analytic approximations of probabilities are anticipated. These methods also enable aggregation of rich data sets in order to obtain straightforward statements about physically significant “consensus invariants” (i.e. quantities that should be expected to be consistent within claimed uncertainties) [3], and whether they have been demonstrated to be equivalent by this direct peer-to-peer comparison.

This paper explores application of testing for statistical consistency to sets of comparison data. These exercises illustrate how insight may be gained into quantities such as: laboratory measurement repeatability, stability in time of measurement capabilities (secular stability), reproducibility of results, and bilateral equivalence of laboratory measurements. Open-source NRC Toolkit software [4] for the realization of these evaluations is demonstrated.

2. COMPARISON RESULTS

Measurement results of linked key comparisons are plotted together in Figure 1. Data from the 5 labs selected for linking the comparisons are collected in the centre of the plot highlighting the linking mechanism and the chance to explore consistency.

2.1. Testing Statistical Consistency

An *all-pairs-difference chi-squared* [5, 6]

$$\chi^2_{APD} = \frac{1}{N(N-1)} \sum_{j=1,\neq i}^{N} \sum_{i=1}^{N} \frac{\left(T_i - T_j\right)^2}{u_i^2 + u_j^2 - 2r_{ij}u_i u_j} \qquad (1)$$

where N represents the number of participants with submitted measurement result T_i and standard uncertainty u_i, tests the linked comparison for overall consistency *as if the linked comparisons were a single comparison*. In the case of a KCRV, the same method would be used to report degrees of equivalence relative to the KCRV, but more care can be needed in accounting for covariances between the KCRV and the artefact-linking invariant since they make use of some of the same measured values. The mean-squared average of differences of a scalar consensus invariant, in this case the triple point of water, obtained between pairs of labs is normalized to the expected combined standard uncertainty in their difference. This test of statistical consistency is analogous in principle to testing a zero order fit – namely assuming all the results agree with each other. Pair equivalence statistics, described for example by equation (1), can be used for testing the pair-wise agreement of all participants while remaining independent of any specific choice or choices of key comparison reference value.

The underlying concept of pair equivalence can easily accommodate longer linking paths and multiplicities of a CIPM KC with links to one or more RMO comparisons. In a conventional CIPM or RMO KC, when there is more than one artefact circulated to all participants the overall comparison can be tested for statistical consistency via aggregation of the results of all artefacts together. This pair-wise aggregation provides information on the statistical consistency of the overall comparison results relative to the measurement results and uncertainties claimed by the participants. The data shown in Figure 1 demonstrates linked CCT-K7 and EUROMET.T-K7 where linking is performed by evaluation of the weighted mean of each of the linking labs' difference between CIPM and RMO measurements (see [3] for detail). Measurement results of the five linking labs are demonstrated to be statistically consistent. Statistical consistency of the aggregated comparisons will be discussed.

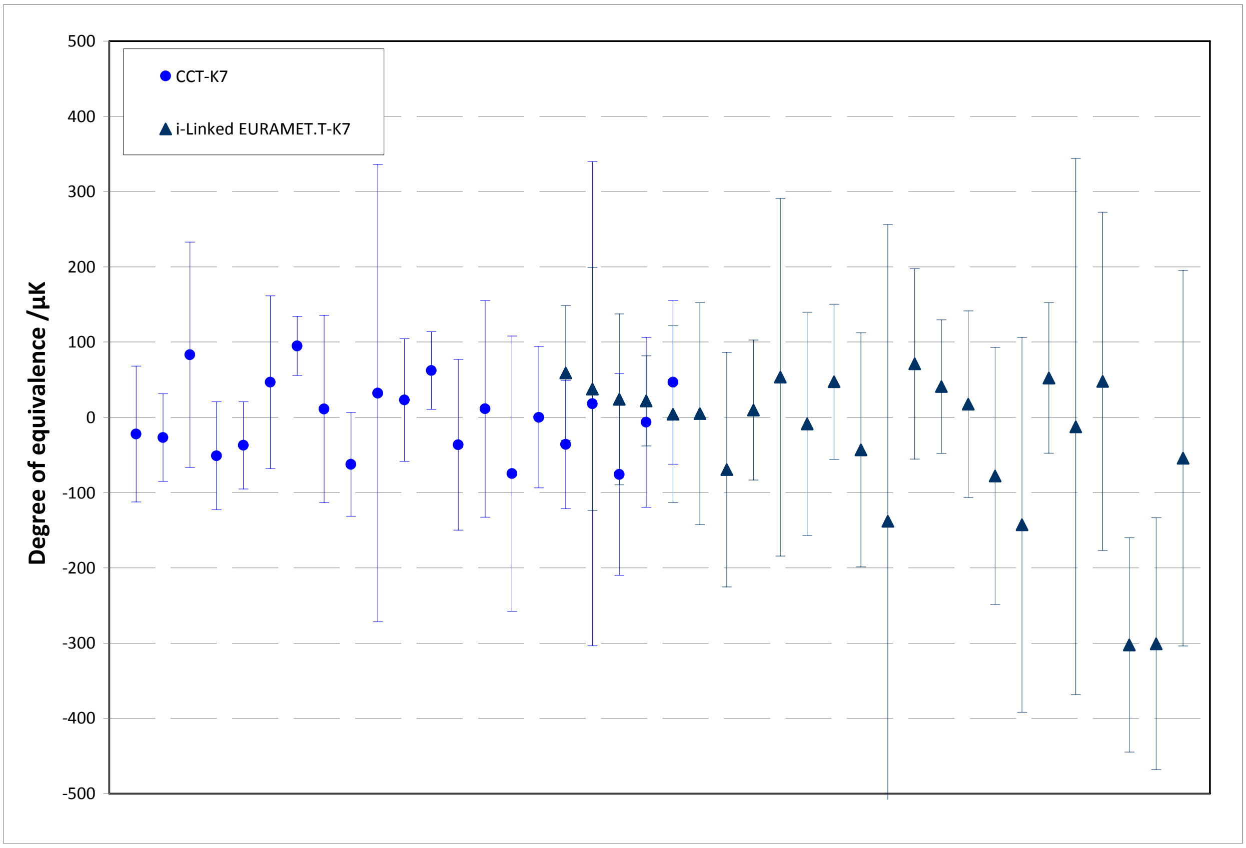

Fig. 1. Plot of measurement data from CCT-K7 (left side, filled circles) and EURAMET.T-K7 (right side, filled triangles) linked via data from labs participating in both comparisons (those results with two data points coinciding). Linking employs the mechanism described in [3]. Error bars represent expanded uncertainty (k=2). EURAMET.T-K7 data error bars include the linking uncertainty but do not account for correlated uncertainty components. The zero line is the CCT-K7 Key Comparison Reference Value (KCRV), the simple average of the filled circle data symbols.

4. CONCLUSIONS

Metrological aspects of participant measurement capability, and the comparison analysis for participants of CCT-K7 and EUROMET.T-K7 are discussed. An advantage of participating in international comparison experiments as a linking lab is that it offers unique insights regarding the consistency of in-house measurement processes, consistency amongst joint participants and those participants in the linked comparisons. Statistical methods for the comparison analysis presented here are based on unmediated statistics which offer an advantage when a metric for central tendency (such as the weighted mean) is demonstrated to be unreliable. In addition, these unmediated statistical methods applied to the linking process exploit the simple and straightforward basis assumption that measurement capabilities of the national metrology institutes are stable. Illustrations of chi-squared consistency testing demonstrate the NRC Toolkit software.

Precise knowledge of laboratory stability and consistency as demonstrated by the experimental data provided by key comparisons is an important result which in turn provides fundamental support to the premise of the equivalence of the national metrology institutes.

REFERENCES

[1] M Stock et al., “Final Report on CCT-K7: Key comparison of water triple point cells”, *Metrologia*, vol 43 Tech Suppl 03001, 2006.

[2] A. Peruzzi et al., “Final Report on EUROMET.T-K7: Key comparison of water triple point cells”, *Metrologia*, vol 46 Tech Suppl 03001, 2009.

[3] J. E. Decker, A. G. Steele and R. J. Douglas, “Measurement science and the linking of CIPM and regional key comparisons”, *Metrologia*, vol 45 pp. 223-232, 2008.

[4] Toolkit website: http://inms-ienm.nrc-cnrc.gc.ca/qde

[5] A. G. Steele and R. J. Douglas “Extending chi-squared statistics for key comparisons in metrology,” *Journal of Computational and Applied Mathematics*, **192**, pp. 51-58, (2006)

[6] J. E. Decker, A. J. Lewis, M. G. Cox, A. G. Steele and R. J. Douglas, “Evaluating results of international comparisons: worked example of CCL-K2 comparison of long gauge block calibration”, *XVIII IMEKO World Congress*, pp. 1-5, Rio de Janeiro, Brazil, Sept. 2006.