# General Linking Methodology with Application to CCT-K3 and EUROMET Project 280

Alan Steele, Barry Wood and Rob Douglas
NRC CANADA

**Abstract**

This paper describes the methodology (introduced at the most recent meeting of the CCEM Working Group on Key Comparisons) for linking two or more comparisons such that the bilateral degrees of equivalence are determined for all participants. Particular attention is given to consideration of the underlying assumptions and constraints, as well as some of the consequences of violating those constraints. The methodology is applied to the comparisons at the triple point of mercury summarized in CCT-K3 and EUROMET Project 280; a worked example can be found in the section 'Linking CCEM-K3 and EU 280'.

## Introduction

The CIPM Mutual Recognition Arrangement (MRA) is a world-wide project to describe the relationships between different, and yet supposedly equivalent, measurement standards realized in various national metrology institutes. The underlying technical basis for the MRA is a set of comparisons chosen by the Consultative Committees called Consultative Committee Key Comparisons (CCKC). These comparisons are usually conducted at the lowest possible level of uncertainty, and involve a small number of institutes drawn from the membership of the relevant CC. Linking the measurement capabilities of all interested signatories of the MRA to the results of each CCKC requires numerous additional comparisons - this is the only practical means of relating this large number of participants.

During the early development of the MRA, it was envisioned that these additional comparisons would take the form of either bilateral key comparisons (involving one CCKC participant and a laboratory to be linked), or Regional Metrology Organization (RMO) key comparisons (involving multiple laboratories to be linked from the region, and at least one CCKC participant). In either case, the joint participants in the CCKC and the subsequent key comparison play the role of 'linking laboratories', and their results in the multiple comparisons are to be used to extend the network of MRA degrees of equivalence around the world.

In the field of thermometry, we are now at the stage where some CCT KC's have been completed and accepted, and one has been approved for inclusion in Appendix B. As well, some bilateral comparisons have been initiated and some RMO comparisons are underway, and will be considered for acceptance. The task of linking these comparisons to determine the complete table of bilateral degrees of equivalence between all pairs of participants remains incomplete.

In this paper, we use real experimental data at the triple point of mercury taken from the recently-completed CCT-K3 comparison of fixed points of the ITS-90, and from the already-published Euromet Project 280 to construct an example which clearly illustrates the multiplicity of different degrees of equivalence and still creates a complete table of bilateral degrees of equivalence that links results from two separate comparisons.

**Identification of the Starting Assumptions**

In the mercury portion of CCT-K3 and EUROMET 280, which we shall now simply refer to as EU 280, there are four labs that participated in both comparisons, including both pilot laboratories, NIST and BNM-INM. The degree of equivalence (DoE) between any two laboratories as defined in the MRA is simply the difference between their respective reported mean values and an equivalence uncertainty generally determined as the quadrature sum (RSS) of each laboratory's claimed uncertainty. The uncertainty of the difference is intended to include any unaccounted for common uncertainties, all calculated in a manner to compensate for any correlation effects. [The authors recommend to the reader that he not yet focus on the complications of the calculation of the equivalence uncertainty. This will be considered in detail later. The authors also advise that not everything about the linking problem should be considered as a further complication in the equivalence uncertainty calculation.]

The calculation of the DoE of the non-linking labs of the two comparisons through the four linking labs is a mathematically over-determined problem. The normal approach to such a problem is to minimize a particular parameter, such as the variance of the linking labs. We have attempted this approach under various circumstances. A numerical result was achieved in each case but there were also unresolved problems that were created just as in other approaches considered by the CCEM. These problems were generally of the following nature:

Treating labs differently,
Having to 'go back in time' and change or increase all possible DoEs,
Consequences about all future comparison results

Eventually we focused on accepting a set of seemingly obvious assumptions. We found, since the problem is mathematically over-determined, that it was essential to consider the other constraints before worrying about how the mathematics and statistics were to be formulated. We have also identified some of the consequences of violating these constraints.

**Three Types of Comparison**

There are three possible types of comparison for which linking is needed.

1) We start with the Consultative Committee Key Comparison (CCKC), which has the following features:

the methodology and results of all pairwise DoEs is defined and accepted;
the CCKC is formally published, accepted by the CC and accepted for entry into the MRA Appendix B database.

2) We also have bilateral comparisons, which are specifically mentioned in the MRA as a method of including labs that miss the CCKC or other large-scale comparisons. Furthermore, bilateral comparisons are an acknowledged mechanism for a laboratory to improve disappointing comparison results and to reflect improvements in its metrology. In order to fully realize these advantages, the bilateral DoE must be linked to the rest of the CCKC participants. We note that this process has not yet been completed even for the BIPM bilateral comparisons, some of which have been accepted into Appendix B in other fields such as voltage.

3) Finally, we have multilateral comparisons usually conducted under the auspices of a Region Metrology Organization (RMO). EUROMET Project 280 is such a comparison involving an RMO and several other participants. This comparison has been published, but it has not yet

been considered as a key comparison or linked to the appropriate CCKC. It is possible to treat it as an example, however, alongside CCT-K3, since it is representative of what we expect to occur for "real" RMO key comparisons.

**Issues concerning the MRA**

Let us state some (but not all) of the implications of the MRA, and in particular about Appendix B.

1. Pairwise bilateral DoEs between all participants are required as an essential aspect of the MRA. These DoEs are a major part of the scientific support for the elimination of trade barriers involving metrological concerns.

2. Pairwise bilateral DoE between all participants within a single comparison can always be mathematically determined without reference to a KCRV or its uncertainty. The MRA does not say this explicitly but it is a simple mathematical consequence of the input data consisting of values and uncertainties from each participant. The pairwise bilateral DoEs can always be determined even if the KCRV is unknown, not calculated or even not defined.

3. The CCs have already tended to present the best DoE that can be obtained from the data (at least within a single comparison) in Appendix B. It may not be obvious that this is the case, but we will demonstrate using examples from existing MRA entries that this is already being done.

4. Bilateral comparisons are essential to the implementation of the MRA. They are specifically mentioned in the MRA and bilateral comparisons with BIPM have already been accepted into Appendix B. Their pairwise bilateral DoEs are uniquely and unambiguously determined by the difference of the participants' means and the RSS of their uncertainties, corrected for any known correlation effects.

5. Almost all comparisons can be considered as a set of bilateral comparisons, including CCT K3 and EUROMET 280. In both of these cases the value reported for each laboratory has been determined (among other ways) as the difference between the laboratory's result and the pilot laboratory's results. Each result is determined just as it would be in a series of bilateral comparisons, although the analysis required to do this was often complicated by the particulars of the experimental protocol.

6. A new bilateral comparison, between say NRC and NIST, both of whom have already participated in CCT-K3, generates a new and second DoE. It does not replace the first DoE from CCT-K3; rather the new information supplements the existing result. There will be two DoEs in the database. Perhaps the database will present the most recent DoE first but both DoEs exist. This idea should not be difficult to accept, given the fact that there are already a number of "overlapping" degrees of equivalence at several temperatures explored in more than one CCT Key Comparison.

7. If the CCT decides to present some sort of average of the two (or more) DoEs generated by the scenario described in item 6, then this average constitutes a new, different DoE and its existence does not eliminate either of the previously existing two DoEs. We expect that older DoEs will be 'retired' as the MRA database evolves; they may still be available, but will not

be considered current with respect to the NMI's present capabilities. The time scale for this 'retirement' is likely to be of the order of the CCKC repetition rate, 10 years or more.

8. In **all** comparisons it must be possible for the participants to improve their DoEs, especially with their trading partners. Without the possibility to improve their DoE, NMIs will eventually realize that there is little or no advantage (and perhaps significant risk and expense) to participate in a second comparison. Such a consequence will greatly discourage many NMIs from acting as linking laboratories and hamper the development of the MRA. If only CC key comparisons establish the 'best' DoEs, then significant improvements in a NMI's capabilities can not be demonstrated until the next CC key comparison. A ten-year wait is too long to correct a poor comparison result or to demonstrate advances in metrology that generates the smallest DoEs. The CCT has already established a policy of encouraging and performing bilateral key comparisons for participants whose CCKC results were compromised in some way. This policy explicitly supports the notion that NMIs want the best possible results put forward to represent their calibration and measurement capabilities.

**Linking Comparisons**

Given two comparisons in which more than one laboratory has jointly participated, the problem of linking the results is mathematically over-determined. That is to say that there are multiple solutions or pathways to evaluate the bilateral DoEs between laboratories.

Let us introduce a nomenclature to specifically identify the DoE being discussed. It will be of the form comparison label {lab i, lab j} = ($m_i$-$m_j$ ± $U_{i,j}$). Thus CCTK3Hg {NRC, NIST} = (0.22 ± 0.18) describes the CCT-K3 comparison degree of equivalence between NRC and NIST at the triple point of mercury. This corresponds to the NRC row and NIST column entry of the table in the Report and its Appendix B proposal. A similar notation can be used when describing the comparison degree of equivalence between an NMI and the KCRV (where one exists) if that becomes desirable.

The above nomenclature is useful but does obscure one important fact: even within a single comparison, such as CCT-K3, there are actually multiple DoE 'paths' linking any pair of laboratories, and the analysis for that comparison includes at least two such paths - "direct" and "indirect" - for each bilateral DoE explicitly.

Of course, other paths to calculating the DoE exist, and one can imagine creating a link between any two laboratories that "cycles through" every other participant. In general, such an extreme technique produces a degree of equivalence with a very large uncertainty even when correlations are properly accounted for, since it includes at least the reproducibility uncertainty of every other lab. We are not trying to suggest that the reported DoEs in CCT-K3 are in any way incorrect; instead we are trying to point out that proposals for Appendix B reporting look first to the "best" DoEs that can be generated by an optimized analysis of the input data.

**Expectations of What the Database Will Look Like**

At this point it is probably best to outline how we think the database will look. This is just our opinion, but we do believe that a consensus of what the database will look like it can simplify the problem.

- Each comparison is separately entered in the database.

- ♦ CCKC comparisons may have groups of three pages similar to CCT-K2 listing bilateral pairwise DoEs, a KCRV and DoEs with the KCRV and a hyperlink to further details.
- ♦ Bilateral comparisons may be just a single bilateral pairwise DoE entry with a hyperlink to further details. Many bilateral comparisons could be on one page if that is convenient.
- ♦ RMO comparisons will be a table of bilateral pairwise DoEs and a hyperlink to further details

- All of the NMIs in all of the above comparisons will be listed once (only) in a single Summary Table of DoEs at a given comparison temperature, similar to the CCT-K2 table of bilateral DoEs.

**What About the KCRV?**

The KCRV is defined only in the CCKC. Its definition and uncertainty are fixed and unaltered until the CCKC is repeated or re-analyzed and approved by the CC. Subsequent comparisons may alter a lab's DoE with the KCRV (even if they participated in its definition) but the KCRV remains unchanged. While there may be some linking paths that are mediated by the KCRV, and some linking strategies will propose or insist that these are necessary or obvious solutions to the problem, in general there is no need for a "reference value" when constructing the linked bilateral degrees of equivalence.

**Multiple Linking Labs**

In the case of bilateral comparisons with only one linking lab to the CCKC things are simple. There is a simple, single unambiguous bilateral pairwise DoE and an equally simple, single and unambiguously derived DoE with each of the other CCKC participants and the KCRV. There are details about the uncertainty calculation concerning the stability or reproducibility of the linking lab's reference but this is not the time to dwell on these issues. Derived pairwise bilateral DoEs can be calculated in a straightforward manner for all of the CCKC participants. A full N+1 Summary Table can be generated, which includes the results for the original N participants plus the single 'linked' laboratory.

In the case of "follow-up" bilateral comparisons between two labs, both of which are already linked to the CCKC, things are only somewhat more complicated. There are the original DoEs from the CCKC. From the bilateral comparison there is the additional bilateral pairwise DoEs and one DoE for each lab to the KCRV mediated by the other bilateral participant. The participants for inclusion into the Summary Table may select any of these new DoEs. Presumably the original CCKC DoEs will be updated with better DoEs, but this can be left to the discretion of the participants and the CC.

In larger comparisons, such as RMO comparisons with multiple linking labs from the CCKC, things are further complicated but the basic idea for calculating DoEs remains the same. There are the original DoEs from the CCKC. There are all of the RMO pairwise bilateral DoEs. For each RMO participant there may be derived DoEs with the KCRV mediated by each linking lab. There are also derived pairwise DoEs generated between the non-linking labs of both comparisons mediated by each linking lab. There may also be other derived pairwise DoEs mediated by statistical combinations of the linking labs. This multiplicity illustrates just how over determined the problem is. The multitude of DoEs should not be of great concern. Selecting the 'best' DoE is quite straightforward and is computationally less difficult than recalculating a weighed mean for each comparison. Generating a unique Summary Table of the best complete set of DoEs may then be performed simply.

**Linking CCT-K3 and EU 280**

We introduce a nomenclature, which uses a subscript C for a CCT-K3 result and the subscript E for an EU 280 result. For simplicity, we consider the CCT-K3 summary with respect to the Pilot (NIST), but the methodology is equally applicable to the summaries with respect to any of the CCT-K3 participants. It is a particularly satisfying coincidence that this Key Comparison does not even define a KCRV, since our linking methodology is designed to work without one.
We define the following variables

- $D_{\mathrm{C\,i}}$ = the CCT-K3 value, with respect to NIST, of lab *i*. (i.e. $D_{\mathrm{i}}$ – NIST)
- $D_{\mathrm{E\,j}}$ = the EU 280 value, with respect to the EU 280 reference value, of lab *j*.
- $u_{\mathrm{C\,i}}$ = the expanded uncertainty, with respect to NIST, of the $i^{\mathrm{th}}$ lab in the CCT-K3 comparison.
- $u_{\mathrm{E\,i}}$ = the claimed expanded uncertainty of the $i^{\mathrm{th}}$ lab in the EU 280 comparison

It is worth noting at this point that we are proceeding with the already-established uncertainty statements for the two comparisons. If additional analysis is required or performed, the particular values in the $u_{\mathrm{Ci}}$ and $u_{\mathrm{Ei}}$ and their combinations may be modified when defining the linking uncertainties. The methodology remains unchanged, of course.

For each linking lab, *k*, we calculate the difference, $d_{\mathrm{k}} \pm u_{\mathrm{L,\,k}}$, between the linking lab's CCT-K3 result (that is its value – NIST's value) and its EU 280 result (i.e. its EU value – the EU reference). The linking uncertainty, $u_{\mathrm{L,k}}$, from lab *k* includes the transfer uncertainties of both comparisons and the reproducibility of the linking lab's result over the time period between the two comparisons, all combined in quadrature, and we use the *Q* operator to describe quadrature combinations.

$$d_{\mathrm{k}} = D_{\mathrm{C\,k}} - D_{\mathrm{E\,k}} \qquad\qquad u_{\mathrm{L,\,k}} = Q\,(\,t_{\mathrm{C}},\, r_{\mathrm{k}}\,,\, 2t_{\mathrm{E}}\,) \qquad\qquad \text{Eqn. 1}$$

All of the original EU 280 results are offset by $d_{\mathrm{k}} \pm u_{\mathrm{L,\,k}}$ creating a full set of linked input comparison data all with respect to a common reference.

$$\text{Link}\{\text{EU Lab}\,j,\,k\} = \text{EU280}\{\text{EU Lab}\,j,\,k\} - d_k \pm u_{\mathrm{L,\,k}} \qquad\qquad \text{Eqn. 2}$$

These data represents the input data as if the EU participants had all performed bilateral comparisons with the linking lab. The inferred DoEs of these labs with NIST is given by

$$\text{EU/NIST}\{\text{EU Lab}\,j,\,\text{NIST}\} = \text{Link}\{\text{EU Lab}\,j,\,k\} - \text{CCTK3Hg}\{k,\,\text{NIST}\} \qquad\qquad \text{Eqn. 3}$$

with the combined uncertainties explicitly given by $Q_{\mathrm{j,\,k}}\,(u_{\mathrm{E\,j}}\,,\, u_{\mathrm{C\,k}})$

The rest of the table of bilateral equivalences including those with the NIST are easily completed as:

$$\text{EU/CCT-K3}\{\text{EU Lab}\,j,\,\text{K3 Lab}\,i\} = \text{Link}\{\text{EU Lab}\,j,\,k\} - \text{CCTK3Hg}\{k,\,\text{K3 Lab}\,i\} \qquad \text{Eqn. 4}$$

with the combined uncertainties explicitly given by $Q_{\mathrm{j,\,i}}\,(u_{\mathrm{E\,j}}\,,\; u_{\mathrm{C\,k}}\,,\, -\mathrm{u}_{\mathrm{NIST}}\,,\; u_{\mathrm{C\,i}})$

Note that the $-\mathrm{u}_{\mathrm{NIST}}$ term implies that this term is subtracted in quadrature and corrects for the correlation of the uncertainty of the definition of the NIST result, since it does not apply to the bilateral DoEs.

There are 12 participants in CCT-K3 at the mercury triple point, and 11 in EU 280. The CCT-K3 comparison generates a 12×12 table of bilateral equivalances and the EU 280 generates an 11×11 table of bilateral equivalances. The above process completes the full 23×23 bilateral equivalence matrix. Note that there are two row and column entries for each linking lab, one referring to its results in each comparison. There are four possible linking labs and each is used in turn to generate a unique and complete table.

The Summary Table is a separate 23×23 table of bilateral equivalances as well as a 23 element column of equivalances with respect to the CCT-K3 NIST value. Note that in cases where a KCRV exists, this column would contain the degrees of equivalence with respect to that quantity. The 23 rows and columns list each laboratory only once. Each cell of the summary table may be filled with the values of **any** one of the corresponding cells of the seven complete tables or the alternative indirect DoE generated only by the linking labs. As indicated earlier, we expect the CCT to select the best equivalence for its Key Comparisons. In any case, the Summary Table should be approved by the CCT.

The final step in the process is the selection of one, and preferably the best, equivalence for each cell of the Summary Table. This is easily done by selecting the smallest QDE interval generated by each of the four cells associated with a particular Summary Table cell, mediated by the different choices of linking laboratory. There are, of course, other choices for selecting the "best" DoE for the final summary table. It might be preferable to select the DoE with the smallest uncertainty; some may prefer a "composite" selection, using a weighted mean of the various DoEs; some may wish to explore minimization techniques to incorporate as much of the information as possible into a single summary DoE. The use of the smallest QDE interval has several advantages. This approach is a variation of the kind of "order statistics" that give rise to the median, for example, and thus all of the information generated in the individual linking tables is being used. This approach also provides a "quantified equivalence" table that mirrors the degree of equivalence table as a natural part of the calculations.

The full analysis used to link CCT-K3 and EU 280 is available in an Excel spreadsheet; the resultant Summary Table is shown below. Because the comparison uncertainties in EU 280 are generally smaller than their counterparts in CCT-K3, many of the bilateral degrees of equivalence among the linking laboratories have been improved over what was demonstrated in CCT-K3. These improvements are reflected in a number of bilateral CCT-K3 DoEs for these particular laboratories. In order for these advances to be formally recognized within the MRA, the CCT would first have to consider and approve EU 280 as a Regional Key Comparison, and authorize the linking methodology.

**Concluding concepts:**

- Only the CCKC generates a KCRV and it is unaltered by subsequent comparisons.
- Each comparison is separately presented and listed in the Appendix B
- All comparisons (CC, Bilateral and RMO) generate pairwise bilateral DoEs with all other summary participants and the KCRV when it exists.
- Any DoE (but most probably the best) may be selected for entry into the Summary Table by the participant with the approval of the CC.
- Some record of the date, and derivation methodology of each DoE should be considered. This will be required no matter what method is utilized to perform the linking and update the MRA Appendix B.