

PAPER • OPEN ACCESS

Analysis of a regional metrology organization key comparison: model-based unilateral degrees of equivalence

To cite this article: Maurice Cox and Katsuhiro Shirono 2023 *Metrologia* **60** 055014

View the [article online](#) for updates and enhancements.

You may also like

- [A kinetic study of the local field approximation in simulations of AC plasma display panels](#)
P J Drallos, V P Nagorny and W Williamson Jr
- [Evaluation of regional key comparison data: two approaches for data processing](#)
I A Kharitonov and A G Chunovkina
- [Final report on the key comparison EUROMET.AUV.V-K1](#)
Hans-Jürgen von Martens, Clemens Elster, Alfred Link et al.

Analysis of a regional metrology organization key comparison: model-based unilateral degrees of equivalence

Maurice Cox^{1,*}  and Katsuhiro Shirono² 

¹ National Physical Laboratory, Teddington TW11 0LW, United Kingdom

² National Metrology Institute of Japan (NMIJ), National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki 3050051, Japan

E-mail: maurice.cox@npl.co.uk

Received 3 March 2023, revised 14 August 2023

Accepted for publication 7 September 2023

Published 22 September 2023



Abstract

A statistical testing method is developed to analyze a Regional Metrology Organization (RMO) key comparison (KC) with linking to the corresponding KC conducted by the International Committee of Weights and Measures (CIPM). We establish a statistical model to which a generalized least-squares method is applied to determine a linking invariant, ensuring that the CIPM KC reference value is unchanged by the analysis. The proposed method generally gives different unilateral and bilateral degrees of equivalence (DOEs) from those assessed by other methods. The approach accounts for uncertainty information provided by all participating laboratories and correlation information provided by the linking laboratories. Since decisions are made on the basis of calculated DOEs, we emphasize that it is valuable to have adequate knowledge of the properties of methods used for analyzing the results from an RMO KC.

Keywords: CIPM and regional key comparisons, key comparison reference value, degrees of equivalence, statistical model, generalized least squares, uncertainty propagation

(Some figures may appear in colour only in the online journal)

1. Introduction

The International Committee of Weights and Measures (CIPM) Mutual Recognition Arrangement (MRA) [1] is technically supported by key comparisons (KCs), for which purpose the CIPM Consultative Committees (CCs) provide KCs in many metrology disciplines. The KCs are implemented to establish the degree of equivalence of national measurement

standards maintained by national metrology institutes (NMIs) and to provide for the mutual recognition of calibration and measurement certificates issued by NMIs. The calibration and measurement capabilities (CMCs) of participating NMIs must be consistent with the results derived from the KCs.

Although there is no official guideline for CIPM KC data evaluation, many analyses are conducted based on peer-reviewed publications such as those given in the references to this paper to determine the unilateral and bilateral degrees of equivalence (DoEs) for the laboratories participating in a KC. It is noted that DoEs consist of both value and uncertainty components, while the value components are often misleadingly referred to as *the* unilateral or bilateral DoEs.

These DoEs are measures indicating the extent to which the conducted measurements are consistent with a

* Author to whom any correspondence should be addressed.



Original Content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

consensus value, the KC reference value (KCRV), and among themselves.

The following is noted in the MRA document:

‘Participation in a CIPM key comparison is open to laboratories having the highest technical competence and experience, normally the member laboratories of the appropriate Consultative Committee.’ [1, clause 6.1]

To establish global metrological traceability, frameworks in addition to the CIPM KCs are needed. The MRA document [1] explicitly requires Regional Metrology Organization (RMO) KCs and RMO supplementary comparisons to take place. These RMO comparisons must be linked to the results in the CIPM KCs to show the degree of consistency of the measurements among laboratories participating in the RMO and CIPM comparisons. In this regard, the MRA document states

‘RMO key comparisons must be linked to the corresponding CIPM key comparisons by means of joint participants. The degree of equivalence derived from an RMO key comparison has the same status as that derived from a CIPM key comparison.’ [1, clause 3.2]

‘The results of the RMO key comparisons are linked to key comparison reference values established by CIPM key comparisons by the common participation of some institutes in both CIPM and RMO comparisons.’ [1, annex T.4].

An approach is therefore needed to assess the unilateral DoE that is directly related to the KCRV in the CIPM KC, which is referred to as the CIPM KCRV in the present study.

Similarly to the CIPM KCs, there is no official guideline for the analysis of the RMO KCs but procedures given in papers such as [2–6] can be applied. A concern is that in our understanding no statistical (hypothesis) testing approach has been suggested in these studies. Generally, testing and estimation are the two main approaches to decision making from data.

There are some studies where no statistical model is specified so that it is difficult to say which of the two approaches is employed. For example, Elster *et al* [3] describes such a method for which we can naturally derive the same procedure as theirs by using an estimation approach as shown in appendix C. Further, Kharitonov and Chunovkina [4] report two methods to give unilateral DoEs without adopting a statistical model. One of the two methods, called procedure C in their paper, is essentially incorporated in the method proposed by Decker *et al* [2] to develop a practical procedure.

Thus, as far as we know, there has been no method in which a statistical testing approach based on a specific statistical model is explicitly given. One possible drawback of the simple application of such a model may be that the measurand estimated by the CIPM KCRV would be re-estimated through the RMO KC analysis if special treatment were not given, as explained in section 3.

That the CIPM KCRV is not to be changed by the results of an RMO comparison is supported by [7, chapter 4] and [8] and especially by the statements

‘... RMO key comparisons extend the metrological equivalence established by the CIPM key comparisons to a greater number of national metrology institutes ...’ [1, annex T.9]

‘The linkage does not modify the value and the uncertainty of the master CIPM key comparison reference value, which remains unique and non-altered for the whole family [the CIPM KC and the corresponding RMO KCs]. It simply extends the matrix of equivalence and the graph of equivalence in order to give evidence on the comparability between institutes that have participated ...’ [9]

In this study, we develop a statistical data analysis of RMO KCs in which the measurand in the CIPM KC is not re-estimated, and consequently obtain unilateral and bilateral DoEs as statistics in a standard testing approach. Although re-estimation is largely inconsistent with the objectives of the MRA, it possesses some technical advantages in that with a large set of RMO participants, it would be expected that the resulting consensus value would often be better defined.

Moreover, to show explicit linking between the unilateral and bilateral DoEs and the CIPM KCRV, we introduce a ‘linking invariant’, which is also considered in the study by Decker *et al* [2]. To determine the linking invariant, a generalized least-squares (GLS) method is developed in which the estimate, the KCRV, of the measurand in the CIPM KC is fixed. Then, we propose unilateral and bilateral DoEs to be clearly used as statistics in statistical testing using the linking invariant. These unilateral and bilateral DoEs are generally different from those in previous studies. We will characterize those differences in this paper through the application to an actual example and simulations.

Literature searches indicate that quantitative information on the correlation coefficients, which is important information in RMO KC analyses, is often not available. There are exceptions. In the fluid flow example in section 5, the correlation coefficients were assessed as 0.8 [10]. In a KC of lamp spectral irradiance [11] for wavelengths from 250 nm to 2500 nm, participating laboratories were requested to separate uncertainties due to correlated and uncorrelated effects. These effects depend on the lamp being compared and the wavelength used. Based on that information, for low wavelengths, correlation coefficients for one laboratory were ≈ 0.8 , whereas for high wavelengths they were ≈ 0.4 . On the other hand, for a vibration accelerometer comparison [12], it is stated that, since no information about correlations is available, the data are treated as being uncorrelated. The last case appears to be quite common despite the frequent availability of uncertainty budgets from which the correlation coefficients could potentially be determined [13]. In the present study, we handle only cases where the correlation coefficients between the reported

values from identical laboratories are quantitatively reported or assessed.

This paper is organized as follows. Section 2 gives assumptions on the analysis, shows mathematical forms for the value components of unilateral and bilateral DoEs, and discusses general principles of the statistical testing approach we use as an extension of the analysis of a CIPM KC. Section 3 provides a GLS approach to estimate the linking invariant and makes comparison with other available studies. Section 4 provides the uncertainty evaluation of the DoEs. Section 5 applies the analysis to an actual example from the fluid flow area. We discuss the features of our proposal in section 6 through simulations. Section 7 gives a brief summary of this study.

In this work, as in the GUM [14], for economy of notation the same symbol is used for the random variable that represents a quantity and a realization of the quantity. For improved visualization, zero elements in some matrices are replaced by blanks, such as in expression (26).

All calculations given in this paper were made using MATLAB R2022a General Release (9.12.0.1884302) using an Intel $\times 64$ -based processor with an i5-5300U core, a 2.30 GHz CPU and a 64 bit operating system.

2. Available information and statistical model

2.1. Available information

Let $I_L = \{1, \dots, L\}$ denote the indices of the linking laboratories in the CIPM and RMO KCs, $I_x = \{L+1, \dots, M\}$ those of the non-linking laboratories in the CIPM KC, and, finally, $I_y = \{L+1, \dots, N\}$ those of the non-linking laboratories in the RMO KC.

The following information is assumed to be available:

1. Data $x_i, u(x_i)$, $i = 1, \dots, M$, representing values and associated standard uncertainties, respectively, provided by the M participants in the CIPM KC;
2. Analogously, $y_i, u(y_i)$, $i = 1, \dots, N$, provided by the N participants in the corresponding RMO KC;
3. Correlation coefficients ρ_i , $i = 1, \dots, L$, provided by the L laboratories that participated in both comparisons, the *linking* laboratories;
4. The CIPM KCRV x_{ref} , taken here as the weighted mean (WM) of the data in point 1.

Regarding point 4, we consider only cases where the CIPM KCRV x_{ref} is obtained as the WM of the x_i , $i \in I_L \cup I_x$, as given later in expression (4). Some of the reported data might not be used in actual CIPM KCs in the computation of the WM. We will give some remarks on that possibility in appendix B. Further, although our approach can potentially be adapted to choices other than the WM, we do not consider that option here.

2.2. Linking invariant and DOEs

In our analysis, the linking of the data in an RMO KC to a CIPM KC is expressed explicitly. For this purpose, we

introduce a linking invariant h_{link} as do Decker *et al* [2]. The value $y_i + h_{\text{link}}$ may be interpreted as what would have been reported by laboratory i in the RMO KC had it actually participated in the CIPM KC.

In this subsection, we show the computation and discuss the interpretation of unilateral and bilateral DoEs using the linking invariant. Here, the *value component* of the unilateral DoE is a statistic that expresses the difference between the reported value and the CIPM KCRV after adjustment by a linking invariant. Similarly, the value component of the bilateral DoE expresses the difference between the reported values from two laboratories after adjustment by a linking invariant.

The value component of the unilateral DoE for non-linking laboratory $j \in I_y$ in an RMO KC using the linking invariant h_{link} is

$$d_j^{(\text{RMO})} = y_j + h_{\text{link}} - x_{\text{ref}}. \quad (1)$$

The bilateral DoE is the difference between the unilateral DoEs after the adjustment. When values from the two KCs are to be compared, the value component of the bilateral DoE between laboratory $j \in I_y$ in the RMO KC and laboratory $\ell \in I_L \cup I_x$ in the CIPM KC is

$$d_{j,\ell}^{(\text{RMO}, \text{CIPM})} = d_j^{(\text{RMO})} - d_\ell^{(\text{CIPM})}. \quad (2)$$

Because the CIPM KCRV remains unchanged, the unilateral DoE of a linking laboratory is also not adjusted based on the results of an RMO KC, and expression (2) is employed for the relation between a non-linking laboratory in the RMO KC and a linking laboratory. The value components of the bilateral DoEs between non-linking laboratories $j \in I_y$ and $\ell \in I_y$ in the RMO KC are similarly given by

$$d_{j,\ell}^{(\text{RMO}, \text{RMO})} = d_j^{(\text{RMO})} - d_\ell^{(\text{RMO})}. \quad (3)$$

2.3. Interpretation of the unilateral DoE in a CIPM KC analysis

The value components of the DoEs in section 2.2 are regarded as test statistics. This interpretation can be considered to be a natural extension of the practice of a typical CIPM KC analysis. How we can interpret the unilateral DoEs in a CIPM KC is thus explained in this subsection before showing the statistical approach for RMO KC analyses.

The CIPM KCRV x_{ref} and its associated standard uncertainty $u(x_{\text{ref}})$ are given by Cox [15]:

$$x_{\text{ref}} = u^2(x_{\text{ref}}) \sum_{i \in I_L \cup I_x} \frac{x_i}{u^2(x_i)}, \quad u^2(x_{\text{ref}}) = \left[\sum_{i \in I_L \cup I_x} \frac{1}{u^2(x_i)} \right]^{-1}, \quad (4)$$

and the unilateral DoE $(d_j^{(\text{CIPM})}, U_j^{(\text{CIPM})})$ for laboratory $j \in I_L \cup I_x$ is

$$d_j^{(\text{CIPM})} = x_j - x_{\text{ref}}, \quad U_j^{(\text{CIPM})} = ku \left(d_j^{(\text{CIPM})} \right), \quad (5)$$

where

$$u^2 \left(d_j^{(\text{CIPM})} \right) = u^2(x_j) - u^2(x_{\text{ref}}). \quad (6)$$

and k is the coverage factor, which is equal to 1.96 under the assumption of normality.

Candidates for a statistical model for the unilateral DoEs (5) in a CIPM KC include

$$x_i \sim N(\mu_x, u^2(x_i)) \quad \text{for all } i \in I_L \cup I_x, \quad (7)$$

with which we can derive the model

$$d_j^{(\text{CIPM})} \sim N(0, u^2(d_j^{(\text{CIPM})})). \quad (8)$$

Here μ_x denotes a value of the measurand and $N(\mu, \sigma^2)$ the normal distribution with mean μ and standard deviation σ . It is noted that while x_{ref} is an estimate of the measurand, μ_x is the true value of it. Bayesian hypothesis testing for model (7) has been suggested by Kacker *et al* [16], in which the unilateral DoEs (5) are given as test statistics.

For statistical testing, the following score, called the E_n score, for the j th participant is computed from the unilateral DoE:

$$E_j^{(\text{CIPM})} = \frac{d_j^{(\text{CIPM})}}{U_j^{(\text{CIPM})}}.$$

The E_n score suggests ‘satisfactory’ performance when its absolute value is no greater than unity and ‘unsatisfactory’ otherwise. The probability of ‘unsatisfactory’ performance for a specific laboratory when model (7) holds is 5 %, that is, on average once in 20 KCs. Performance evaluation using the E_n score can hence be interpreted as a statistical test with a significance level of 5 %. Note that the statistical test is not for the consistency of the values in aggregate like the χ^2 test in [15] but that of the value provided by a specific laboratory. In other words, $E_j^{(\text{CIPM})}$ is not used as a component of multiple tests for model (7) but a sole statistic for it with high power of test against the bias from μ_x in the distribution of x_j . When the main interest of a CC Working Group is whether a participant has a bias or not, the hypothesis testing approach using this model can be used.

Our interpretation of unilateral DoEs as test statistics is explained here. Performance is satisfactory when the value component of the unilateral DoE is not significantly different from zero. When performance is unsatisfactory, it is expected that investigation to find causes of the inconsistency is conducted. It is not generally expected that the computed value parts of unilateral DoEs are used for compensation purposes.

Alternative interpretations are possible. For instance, there are some studies where the value components of unilateral DoEs are interpreted as estimators of hidden biases. Details are given in appendix C.1. It is worth noting that the bias estimation approach gives the physical meaning of the DoEs as hidden biases and their associated uncertainties. Based on this idea, once the non-zero hidden biases are discovered in the KCs, they may need to be corrected regardless of the performance evaluations of the E_n scores (because they are physically meaningful biases). Our impression is that since such corrections are not often implemented in many metrology fields, most CIPM CCs regard the DoEs as non-physical but statistical.

However, since both approaches can give the same unilateral DoEs in a CIPM KC as shown in the present subsection and appendix C.1, the choice does not influence the computation. Further, in an RMO KC analysis, there could be differences in the computed DoEs provided by these two approaches.

2.4. Statistical model to be validated in an RMO KC analysis

The value components of the DoEs specified in section 2.2 are considered as test statistics for statistical testing. Such testing is implemented in this study to check the validity of statistical models.

Define

$$\mathbf{z}_i = \begin{bmatrix} x_i \\ y_i \end{bmatrix}, \quad \mathbf{V}_i = \begin{bmatrix} u^2(x_i) & \rho_i u(x_i) u(y_i) \\ \rho_i u(x_i) u(y_i) & u^2(y_i) \end{bmatrix}, \quad i \in I_L,$$

where $\mathbf{V}_i$ is the covariance matrix associated with $\mathbf{z}_i$. We assume the following statistical model for the generation of the data in the CIPM KC and those reported by the linking laboratories in the RMO KC:

$$\begin{aligned} \mathbf{z}_i &\sim N(\boldsymbol{\mu}, \mathbf{V}_i), & i \in I_L, \\ x_i &\sim N(\mu_x, u^2(x_i)), & i \in I_x \end{aligned} \quad (9)$$

where $\boldsymbol{\mu} = [\mu_x, \mu_y]^\top$, $N(\boldsymbol{\mu}, \mathbf{V})$ denotes the bivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance $\mathbf{V}$, and μ_y denotes a value of the measurand in the RMO KC. Defining the difference between the values of the CIPM- and RMO-KC measurands as $\eta = \mu_x - \mu_y$, the estimate of η is the linking invariant h_{link} . The definition of η is given so that h_{link} is consistent with the linking invariant defined by Decker *et al* [2].

It should be noted that model (9) is not checked by the value components of the DoEs specified in section 2.2 because that is confirmed by the interpretation in section 2.3. Although the correlation information is not validated in the CIPM KC, we assume it is reliable, being based on information provided in carefully checked uncertainty budgets, for example.

The statistical model to be validated using the DoEs specified in section 2.2 is

$$y_i \sim N(\mu_x - \eta, u^2(y_i)), \quad i \in I_y. \quad (10)$$

Model (10) is regarded as the statistical model suggested by laboratory i under the assumption of normality. Based on this statistical model, we investigate the distributions of the $d_j^{(\text{RMO})}$, $d_{j,\ell}^{(\text{RMO, CIPM})}$ and $d_{j,\ell}^{(\text{RMO, RMO})}$. Then, the E_n scores are used for statistical testing.

3. Linking invariant as a GLS solution

We consider first the best linear unbiased estimators (BLUEs) of μ_x and η having the GLS solutions x_{blue} and h_{blue} , respectively, using the data in model (9).

The solution vector for the GLS problem would be

$$\begin{aligned} [x_{\text{blue}}, h_{\text{blue}}]^\top &= \arg \min_{x, h} f(x, h), \\ f(x, h) &= \sum_{i \in I_L} \mathbf{e}_i^\top \mathbf{V}_i^{-1} \mathbf{e}_i + \sum_{i \in I_x} \frac{(x_i - x)^2}{u^2(x_i)}, \end{aligned} \quad (11)$$

where

$$\mathbf{e}_i = \mathbf{e}_i(x, y) = [x_i - x, y_i + h - x]^\top. \quad (12)$$

We do not include in expression (11) a sum involving $y_j \in I_y$ because the linking invariant must be determined using only the most reliable information, namely, that from the laboratories in the CIPM KC (including the linking laboratories, of course). By checking the validity of the statistical model for y_j with $j \in I_y$ after determining the linking invariant, we can avoid the discussion of possible outliers in y_j with $j \in I_y$.

In the solution so provided, h_{blue} would be the linking invariant, as required, and x_{blue} would be a *new* CIPM KCRV. More information would have been used by the approach in determining this new CIPM KCRV (compared with the data originally used in the CIPM comparison only), which implies that it could be an improved value. However, despite this potential benefit, any change to the KCRV would be inconsistent with the MRA as emphasized in section 1: the results from the RMO comparison must be expressed in terms of the KCRV provided by the CIPM KC.

Accordingly, a solution that preserves the CIPM KCRV x_{ref} can simply be obtained after setting $x = x_{\text{ref}}$ in expression (11). By defining h_{link} as the linking invariant for this purpose, the problem simplifies as follows:

$$h_{\text{link}} = \arg \min_h g(h), \quad g(h) = \sum_{i \in I_L} \mathbf{e}_{i, \text{ref}}^\top \mathbf{V}_i^{-1} \mathbf{e}_{i, \text{ref}}, \quad (13)$$

where

$$\mathbf{e}_{i, \text{ref}} = \mathbf{e}_i(x_{\text{ref}}, y) = [x_i - x_{\text{ref}}, y_i + h - x_{\text{ref}}]^\top, \quad (14)$$

because the difference between $f(x_{\text{ref}}, h)$ and $g(h)$, namely, the second sum in expression (11) with $x = x_{\text{ref}}$, is independent of h . It should be noted that although x in expression (11) is essentially fixed to x_{ref} in this formulation, μ_x in model (9) is not fixed.

Since, using symmetry,

$$\begin{aligned} \frac{dg(h)}{dh} &= \sum_{i \in I_L} \begin{bmatrix} 0 \\ 1 \end{bmatrix}^\top \mathbf{V}_i^{-1} \mathbf{e}_{i, \text{ref}} + \sum_{i \in I_L} \mathbf{e}_{i, \text{ref}}^\top \mathbf{V}_i^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ &= 2 \sum_{i \in I_L} \mathbf{e}_{i, \text{ref}}^\top \mathbf{V}_i^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \end{aligned}$$

at the minimum of g ,

$$\sum_{i \in I_L} \mathbf{e}_{i, \text{ref}}^\top \mathbf{V}_i^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 0.$$

After some algebra, we obtain the following expression:

$$h_{\text{link}} = -\frac{1}{Q} \sum_{i \in I_L} [p_i (x_i - x_{\text{ref}}) + q_i (y_i - x_{\text{ref}})], \quad (15)$$

where

$$\begin{bmatrix} p_i \\ q_i \end{bmatrix} = \mathbf{V}_i^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \frac{1}{(1 - \rho_i^2) u(x_i) u(y_i)} \begin{bmatrix} -\rho_i \\ u(x_i)/u(y_i) \end{bmatrix} \quad (16)$$

and (also defining P , which is used later)

$$P = \sum_{i \in I_L} p_i, \quad Q = \sum_{i \in I_L} q_i. \quad (17)$$

The linking invariant h_{link} can be expressed as a linear combination of the uncertain inputs x_i and y_i as shown in expressions (15) and (16). In the uncertainty evaluation of h_{link} , we can hence apply the law of propagation of uncertainty [14, 17], which would be exact in these circumstances.

4. Uncertainty component of DOEs

4.1. Unilateral DOEs

The standard uncertainty $u(d_j^{(\text{RMO})})$ associated with the value component $d_j^{(\text{RMO})}$ of the unilateral DoE for non-linking laboratory $j \in I_y$ in the RMO KC is considered. Since the KCRV x_{ref} and the linking invariant h_{link} are independent of the reported value y_j , expression (1) and the uncertainty propagation law [14, clause 5.2] give

$$u^2(d_j^{(\text{RMO})}) = u^2(y_j) + u^2(h_{\text{link}}) + u^2(x_{\text{ref}}) - 2u(x_{\text{ref}}, h_{\text{link}}). \quad (18)$$

We consider only unilateral DoEs for non-linking laboratories in the RMO KC, since the unilateral DoEs for the linking laboratories have already been evaluated in the CIPM KC.

The variance $u^2(h_{\text{link}})$ and covariance $u(x_{\text{ref}}, h_{\text{link}})$ are

$$u^2(h_{\text{link}}) = \frac{1}{Q} + \frac{(P+Q)^2}{Q^2} u^2(x_{\text{ref}}), \quad (19)$$

$$u(x_{\text{ref}}, h_{\text{link}}) = \frac{P+Q}{Q} u^2(x_{\text{ref}}), \quad (20)$$

the derivations of which are given in appendix A. Hence, expressions (18) to (20) give

$$u^2(d_j^{(\text{RMO})}) = u^2(y_j) + \frac{1}{Q} + \frac{P^2}{Q^2} u^2(x_{\text{ref}}). \quad (21)$$

The uncertainty component of the unilateral DoE is

$$U_j^{(\text{RMO})} = ku \left(d_j^{(\text{RMO})} \right),$$

where k is the coverage factor equal to 1.96 in this study as a consequence of the assumed normality and the coverage factor of 95 %. The E_n score for laboratory j is computed using

$$E_j^{(\text{RMO})} = \frac{d_j^{(\text{RMO})}}{U_j^{(\text{RMO})}}.$$

4.2. Bilateral DOEs

The standard uncertainty $u(d_{j,\ell}^{(\text{RMO}, \text{CIPM})})$ associated with the value component $d_{j,\ell}^{(\text{RMO}, \text{CIPM})}$ of the bilateral DoE between the non-linking laboratory j in an RMO KC and laboratory ℓ in a CIPM KC, that is, $j \in I_y$ and $\ell \in I_L \cup I_x$, is considered. Expression (2) and the uncertainty propagation rule [14, clause 5.2] give

$$u^2 \left(d_{j,\ell}^{(\text{RMO}, \text{CIPM})} \right) = u^2(y_j) + u^2(h_{\text{link}}) + u^2(x_\ell) - 2u(x_\ell, h_{\text{link}}). \quad (22)$$

The standard uncertainty $u(h_{\text{link}})$ of the linking invariant is given by expression (19) and the covariance as follows:

$$u(x_\ell, h_{\text{link}}) = \frac{P+Q}{Q} u^2(x_{\text{ref}}), \quad (23)$$

the derivation of which is given in appendix A.

For the standard uncertainty associated with the value component $d_{j,\ell}^{(\text{RMO}, \text{CIPM})}$ of the bilateral DoE between the non-linking laboratories $j \in I_y$ in an RMO KC and the laboratories $\ell \in I_L \cup I_x$ in the CIPM KC, the use of expressions (6) and (21) yields

$$\begin{aligned} u^2 \left(d_{j,\ell}^{(\text{RMO}, \text{CIPM})} \right) &= \left[u^2(y_j) + \frac{1}{Q} + \frac{P^2}{Q^2} u^2(x_{\text{ref}}) \right] \\ &\quad + [u^2(x_\ell) - u^2(x_{\text{ref}})] \\ &= u^2 \left(d_j^{(\text{RMO})} \right) + u^2 \left(d_\ell^{(\text{CIPM})} \right). \end{aligned}$$

It should be noted that this computation is valid only when the CIPM KCRV is computed as the WM of the reported values including x_ℓ . For a case where x_ℓ is not used in the computation, the uncertainty evaluation is given in appendix B.

The standard uncertainty $u(d_{j,\ell}^{(\text{RMO}, \text{RMO})})$ associated with the value component $d_{j,\ell}^{(\text{RMO}, \text{RMO})}$ of the bilateral DoE between the non-linking laboratories j and ℓ in an RMO KC, that is, for $j, \ell \in I_y$, is discussed. From expression (3),

$$d_{j,\ell}^{(\text{RMO}, \text{RMO})} = y_j - y_\ell.$$

Since no correlation between y_j and y_ℓ is assumed, the uncertainty propagation rule [14, clause 5.1] yields

$$u^2 \left(d_{j,\ell}^{(\text{RMO}, \text{RMO})} \right) = u^2(y_j) + u^2(y_\ell).$$

The uncertainty components of the bilateral DoEs in both cases are

$$\begin{aligned} U_{j,\ell}^{(\text{RMO}, \text{CIPM})} &= ku \left(d_{j,\ell}^{(\text{RMO}, \text{CIPM})} \right), \\ U_{j,\ell}^{(\text{RMO}, \text{RMO})} &= ku \left(d_{j,\ell}^{(\text{RMO}, \text{RMO})} \right), \end{aligned} \quad (24)$$

where k is as before. The E_n scores based on these bilateral DoEs are formed using

$$E_{j,\ell}^{(\text{RMO}, \text{CIPM})} = \frac{d_{j,\ell}^{(\text{RMO}, \text{CIPM})}}{U_{j,\ell}^{(\text{RMO}, \text{CIPM})}}, \quad E_{j,\ell}^{(\text{RMO}, \text{RMO})} = \frac{d_{j,\ell}^{(\text{RMO}, \text{RMO})}}{U_{j,\ell}^{(\text{RMO}, \text{RMO})}}. \quad (25)$$

5. Application to a fluid flow RMO comparison

5.1. Computation of linking invariant

We use the data reported in APMP.FF-K4 [10], which is an RMO KC in the area of fluid flow. Two artefacts were circulated in this comparison with nominal volumes of 20 l and 100 ml. The data for the artefact with nominal volume of 20 l are employed as the actual example in the present study. The results of APMP.FF-K4 were linked to those of CCM.FF-K4 [18]. Two laboratories participating in the comparisons had been nominated as the linking laboratories.

In CCM.FF-K4, three artefacts with nominal volume 20 l identified as TS 710-04, TS 710-05 and TS 710-06 were circulated. The WM in expression (4) was used only for the data for TS 710-06. For TS 710-04 and TS 710-05, another computational method was employed to avoid influence from possible outliers. The CIPM KCRV was computed as the mean of the three values. Here, only the data for TS 710-06 are used. Therefore, it should be noted that the actual analysis cannot be directly compared to our results. In APMP.FF-K4, only one artefact was circulated, and the data on it are used in our analysis.

The reported values and their associated standard uncertainties in both KCs are shown in table 1 and illustrated in figure 1. For CCM.FF.K4, only the data for TS 710-06 are shown in the table. Laboratories 1 and 2 are the linking laboratories: $I_L = \{1, 2\}$. The values reported by laboratory 1 for the CCM and the APMP comparisons are correlated due to a common reference standard and the use of the same instrument. A similar statement can be made for laboratory 2. Both correlation coefficients ρ_1 and ρ_2 were assessed as 0.8 in [10]. The laboratory identifiers are not identical to those in the final report of those KCs.

Eight laboratories participated in CCM.FF-K4. All reported values were used in the computation of the WM of the data for TS 710-06, so $I_x = \{3, \dots, 8\}$. The WM for the CIPM KC data and its associated standard uncertainty are computed as

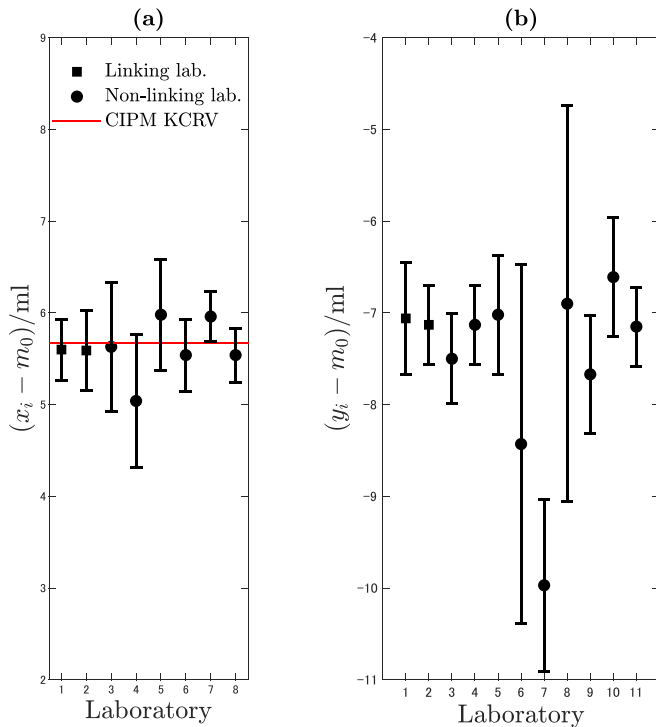
$$x_{\text{ref}} - m_0 = 5.670 \text{ ml}, \quad u(x_{\text{ref}}) = 0.071 \text{ ml},$$

where $m_0 = 20\,000$ ml is the nominal volume.

Table 1. APMP.FF-K4 [10] and CCM.FF-K4 [18] as an example of linking. The linking laboratories are laboratories 1 and 2. For simplicity of expression, the reported values are offset by the nominal value $m_0 = 20\,000$ ml.

CCM.FF-K4			APMP.FF-K4		
Lab. i	$(x_i - m_0)/\text{ml}$	$u(x_i)/\text{ml}$	Lab. i	$(y_i - m_0)/\text{ml}$	$u(y_i)/\text{ml}$
1	5.60	0.17	1	-7.06	0.31
2	5.59	0.22	2	-7.13	0.22
3	5.63	0.36	3	-7.50	0.25
4	5.04	0.37	4	-7.13	0.22
5	5.98	0.31	5	-7.02	0.33
6	5.54	0.20	6	-8.43	1.00
7	5.96	0.14	7	-9.97	0.48
8	5.54	0.15	8	-6.90	1.10
			9	-7.67	0.33
			10	-6.61	0.33
			11	-7.15	0.22

Correlation: $\rho_1 = \rho_2 = 0.8$

**Figure 1.** Reported data in (a) CCM.FF-K4 (the CIPM KC) [18] and (b) APMP.M.FF-K4 [10] (the RMO KC), as given in table 1. Vertical bars show the expanded uncertainty for $k = 1.96$. The legend in (a) also applies to (b). The y-axis ranges are different but the scale intervals are identical.

The covariance matrix for z_1 is

$$\begin{aligned} V_1 &= \begin{bmatrix} u^2(x_1) & \rho_1 u(x_1)u(y_1) \\ \rho_1 u(x_1)u(y_1) & u^2(y_1) \end{bmatrix} \\ &= \begin{bmatrix} 0.0289 & 0.0422 \\ 0.0422 & 0.0961 \end{bmatrix} \text{ml}^2. \end{aligned}$$

Thus, using expression (16),

$$\begin{aligned} \begin{bmatrix} p_1 \\ q_1 \end{bmatrix} &= \begin{bmatrix} -42.2 \\ 28.9 \end{bmatrix} \text{ml}^{-2}, \\ \begin{bmatrix} p_2 \\ q_2 \end{bmatrix} &= \begin{bmatrix} -45.9 \\ 57.4 \end{bmatrix} \text{ml}^{-2}, \end{aligned}$$

and $Q = 86.3 \text{ ml}^{-2}$, using the right-hand expression (17). Thus, expression (15) yields the linking invariant

$$h_{\text{link}} = -\frac{1}{Q} \sum_{i=1}^2 [p_i (x_i - x_{\text{ref}}) + q_i (y_i - x_{\text{ref}})] = 12.700 \text{ ml}.$$

We can compare this result with those of previous studies. Decker *et al* [2] suggested a linking invariant by using the statistical method proposed by Kharitonov and Chunovkina [4]. In their method, the linking invariant is obtained through the following computation:

$$h_{\text{link}} = \left[\sum_{i=1}^2 \frac{1}{u^2(x_i - y_i)} \right]^{-1} \sum_{i=1}^2 \frac{x_i - y_i}{u^2(x_i - y_i)} = 12.701 \text{ ml}.$$

Moreover, while the linking invariant is not given in the method proposed by Elster *et al* [3], it is naturally derived by comparing expression (1) in the present study with expression (27) in their paper. For a matrix A with (i, j) element $A_{i,j} = u(d_i^{(\text{CIPM})} - y_i, d_j^{(\text{CIPM})} - y_j)$ for $i, j \in I_L$, the linking invariant is given as

$$\begin{aligned} h_{\text{link}} &= \left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}^\top \Lambda^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right)^{-1} \left(\Lambda^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right)^\top \begin{bmatrix} x_1 - y_1 \\ x_2 - y_2 \end{bmatrix} \\ &= 12.704 \text{ ml}. \end{aligned}$$

This approach could be interpreted as an extension of the study given by Sutton [5]. More is shown in appendix C.2.

Table 2. Unilateral degrees of equivalence for non-linking laboratories and E_n scores computed using the data in table 1. $E_j^{(RMO)} = d_j^{(RMO)} / U_j^{(RMO)}$ is formed for performance evaluation.

Lab. j	Our proposal			Kharitonov and Chunovkina [4]/Decker <i>et al</i> [2]		Elster <i>et al</i> [3]/Sutton [5]	
	$d_j^{(RMO)}/\text{ml}$	$U_j^{(RMO)}/\text{ml}$	$E_j^{(RMO)}$	$d_j^{(RMO)}/\text{ml}$	$U_j^{(RMO)}/\text{ml}$	$d_j^{(RMO)}/\text{ml}$	$U_j^{(RMO)}/\text{ml}$
3	−0.47	0.55	−0.85	−0.47	0.56	−0.47	0.56
4	−0.10	0.50	−0.20	−0.10	0.51	−0.10	0.51
5	0.01	0.69	0.01	0.01	0.70	0.01	0.70
6	−1.40	1.98	−0.71	−1.40	1.98	−1.40	1.98
7	−2.94	0.97	−3.02	−2.94	0.98	−2.94	0.98
8	0.13	2.17	0.06	0.13	2.17	0.13	2.17
9	−0.64	0.69	−0.92	−0.64	0.70	−0.64	0.70
10	0.42	0.69	0.60	0.42	0.70	0.42	0.70
11	−0.12	0.50	−0.24	−0.12	0.51	−0.12	0.51

The differences among these values are insignificant compared to possible statistical errors in this example. The standard uncertainty $u(h_{\text{link}})$ associated with h_{link} is 0.108 ml using expression (19). However, it is shown that there are differences between the analysis methods. Section 6 will offer a synthetic data set to show that these differences can have a serious influence on the results.

5.2. Computation of DOEs

To exemplify the computation of the DOEs, we again apply the linking of APMP.M.FF-K4 with a part of the data in CCM.FF-K4. Specifically, we take the unilateral DoE of laboratory 10 in APMP.M.FF-K4:

$$d_{10}^{(RMO)} = y_{10} + h_{\text{link}} - x_{\text{ref}} = 0.42 \text{ ml.}$$

Using the computations in section 5.1 and $P = p_1 + p_2 = -88.1 \text{ ml}^{-2}$,

$$u(d_{10}^{(RMO)}) = \left[u^2(y_{10}) + \frac{1}{Q} + \frac{P^2}{Q^2} u^2(x_{\text{ref}}) \right]^{1/2} = 0.35 \text{ ml.}$$

The E_n score for the performance of laboratory j is computed as

$$E_j^{(RMO)} = \frac{d_j^{(RMO)}}{U_j^{(RMO)}}.$$

The E_n score suggests ‘satisfactory’ when $|E_j^{(RMO)}| \leq 1$ as explained in section 2.3. For laboratory 10,

$$E_{10}^{(RMO)} = \frac{d_{10}^{(RMO)}}{U_{10}^{(RMO)}} = 0.60.$$

Table 2 shows all the computed unilateral DOEs for the data in table 1. Only $|E_7^{(RMO)}|$ exceeds unity, implying that laboratory 7 reported an extreme value, an extreme uncertainty or both. The proposed analysis for an RMO KC, however, is unaffected

by the existence of possible outliers from non-linking laboratories, because the linking invariant is determined only using the data in a CIPM KC and the data from linking laboratories.

For comparison, we show the results from the other methods mentioned in section 5.1. It is found that almost the same results are reported by the three approaches. In this example, strong correlations are considered between the two values of the linking laboratories, and $|d_1^{(CIPM)}|$ and $|d_2^{(CIPM)}|$ are much smaller than $U_1^{(CIPM)}$ and $U_2^{(CIPM)}$, respectively. These facts result in the small difference between the methods employed as explained in section 6, and a synthetic case where we can find significant differences is shown there.

As an instance of a bilateral DoE, we focus on the relationship between laboratory 10 in APMP.M.FF-K4 and laboratory 4 in CCM.FF-K4. The value counterpart $d_{10,4}^{(RMO,CIPM)}$ of the bilateral DoE is assessed as

$$d_{10,4}^{(RMO,CIPM)} = d_{10}^{(RMO)} - d_4^{(CIPM)} = 1.05 \text{ ml.}$$

The standard uncertainty associated with $d_{10,4}^{(RMO,CIPM)}$ is computed as

$$u(d_{10,4}^{(RMO,CIPM)}) = \left[u^2(d_{10}^{(RMO)}) + u^2(d_4^{(CIPM)}) \right]^{1/2} = 0.51 \text{ ml.}$$

For the performance evaluation, the E_n score,

$$E_{10,4}^{(RMO,CIPM)} = \frac{d_{10,4}^{(RMO,CIPM)}}{ku(d_{10,4}^{(RMO,CIPM)})} = 1.1$$

with $k = 1.96$ is also considered. We see an ‘unsatisfactory’ result in this case. Thus, an ‘unsatisfactory’ performance in a bilateral DoE can arise even when performances using the unilateral DOEs of the two concerning laboratories are evaluated to be ‘satisfactory’.

Table 3 shows the bilateral DOEs computed for laboratory 10 in the RMO KC. We find some ‘unsatisfactory’ performances where $|E_{10,\ell}^{(RMO,CIPM)}| > 1$ or $|E_{10,\ell}^{(RMO,RMO)}| > 1$. Actions to be taken when such results occur depend on decisions taken

Table 3. Assessed bilateral DoEs involving laboratory 10 in the RMO KC based on the data in table 1, where (R,C) $\equiv$ (RMO,CIPM) and (R,R) $\equiv$ (RMO,RMO).

CC.FF-K4				APMP.FF-K4			
Lab. ℓ	$d_{10,\ell}^{(R,C)}/\text{ml}$	$U_{10,\ell}^{(R,C)}/\text{ml}$	$E_{10,\ell}^{(R,C)}$	Lab. ℓ	$d_{10,\ell}^{(R,R)}/\text{ml}$	$U_{10,\ell}^{(R,R)}/\text{ml}$	$E_{10,\ell}^{(R,R)}$
1	0.49	0.76	0.6	1	—	—	—
2	0.50	0.81	0.6	2	—	—	—
3	0.46	0.98	0.5	3	0.89	0.81	1.1
4	1.05	0.99	1.1	4	0.52	0.78	0.7
5	0.11	0.91	0.1	5	0.41	0.91	0.4
6	0.55	0.79	0.7	6	1.82	2.06	0.9
7	0.13	0.73	0.2	7	3.36	1.14	2.9
8	0.55	0.74	0.7	8	0.29	2.25	0.1
				9	1.06	0.91	1.2
				10	—	—	—
				11	0.54	0.78	0.7

by the CC Working Group. It should be noted that the data in table 1 is taken as a numerical example, and a different approach was implemented in the actual analyses [10].

6. Demonstration of the proposed method using synthetic data

No serious difference is found in the linking invariants and unilateral DoEs by the three approaches for APMP.M.FF-K4 as shown in section 5. In general, these approaches may differ appreciably depending on the data. In this section, the difference between methods is shown using a synthetic data set. Consequently, it can be said that an essential difference can arise only when the correlations between two values from the linking laboratories are weak. To explain clearly the reasons for the difference, a deliberately simple synthetic data set is used in which the correlation coefficient of the single linking laboratory is varied.

The synthetic example is given in table 4 and figure 2. It is assumed that five laboratories participate in the CIPM KC. Laboratory 1 is the only linking laboratory. The correlation coefficient is varied from 0 to 1. The assumed CIPM KC data are shown in figure 2(a). The CIPM KCRV x_{ref} is given as the WM, -0.65 , of $x_1, \dots, x_5$ with $u(x_{\text{ref}}) = 0.35$.

The data for the synthetic example in table 4 relate to (a) the CIPM KC and (b) the RMO KC in figure 2. The results obtained are portrayed in figure 2 as (c) the linking invariant h_{link} and (d) the E_n score $E_2^{(\text{RMO})}$.

It should be noted that when there is only a single linking laboratory, the linking invariant given by Decker's method [2] and that derived through Elster's method [3] are identical. We cannot find any differences either in the set of assessed DoEs with these two methods. We refer to this identical approach as *the reference method* in this section.

We focus on only a single non-linking laboratory in an RMO KC for simplicity of discussion. When assessing the unilateral DoE of a laboratory, data from non-linking laboratories other than the concerned laboratory have no effect. Therefore, there is no need to consider multiple non-linking laboratories

Table 4. Data employed in the simulation in section 6. The only linking laboratory is laboratory 1.

CIPM KC			RMO KC		
Lab.	x_i	$u(x_i)$	Lab.	y_i	$u(y_i)$
1	0.0	0.5	1	0.0	0.5
2	-1.3	1.0	2	1.9	1.0
3	-1.3	1.0			
4	-1.3	1.0			
5	-1.3	1.0			

to characterize a unilateral DoE. The assumed RMO KC data are shown in figure 2(b).

The linking invariant h_{link} is a function of ρ_1 in our proposal. Figure 2(c) shows the variation of h_{link} . It is found that in the reference method, h_{link} is a constant, because

$$h_{\text{link}} = x_1 - y_1$$

is given independently of the magnitude of ρ_1 .³ We can hence find the difference between these two linking invariants, with the difference at a maximum when $\rho_1 = 0$ in the interval $[0, 1]$.

Using our proposal when $\rho_1 = 0$, expression (15) simply gives

$$h_{\text{link}} = -\frac{1}{q_1} [p_1 (x_1 - x_{\text{ref}}) + q_1 (y_1 - x_{\text{ref}})] = x_{\text{ref}} - y_1.$$

It is thus found that

$$d_2^{(\text{RMO})} = y_2 + h_{\text{link}} - x_{\text{ref}} = y_2 - y_1.$$

In the case of using the reference method when $\rho_1 = 0$,

$$d_2^{(\text{RMO})} = y_2 + h_{\text{link}} - x_{\text{ref}} = y_2 - [y_1 - (x_1 - x_{\text{ref}})].$$

In other words, with our proposal, the amount to be compared with y_2 is given as y_1 . With the reference method, the value y_1

³ h_{link} is not a constant with respect to ρ_i for $i \in I_L$ when there are multiple linking laboratories.

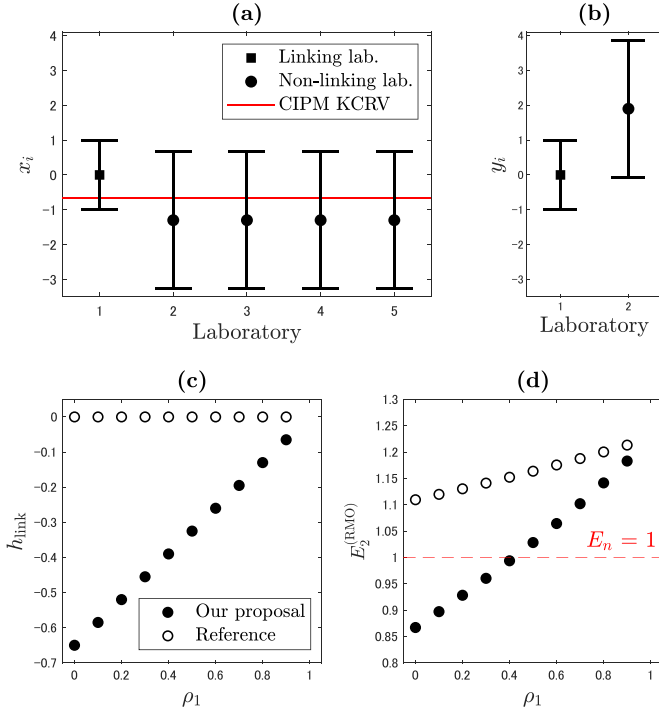


Figure 2. Data for the synthetic example in table 4 for (a) the CIPM KC and (b) the RMO KC, and results obtained for (c) the linking invariant h_{link} and (d) the E_n score $E_2^{(\text{RMO})}$. The legends in (a) and (c) apply also to (b) and (d), respectively.

in our proposal is compensated by the difference between x_1 and x_{ref} . This difference is evaluated as insignificant in the CIPM KC in accordance with section 2.3, because

$$|E_1^{(\text{CIPM})}| = \frac{|d_1^{(\text{CIPM})}|}{U_1^{(\text{CIPM})}} = \frac{0.65}{0.69} = 0.9 < 1.$$

In our proposal, the difference $x_1 - x_{\text{ref}}$ is neglected because no common bias in x_1 and y_1 is suggested through the information $\rho_1 = 0$. In the statistical testing concept shown in section 2.3, the purpose of the CIPM KC is to validate that the hidden bias is zero. When no significant bias is found for laboratory 1 in the CIPM KC and no correlation suggested, the zero bias may naturally reflect the qualitative conclusion obtained through the CIPM KC. In the reference method, the common bias is suggested even when no correlation is assumed. In other words, the distribution of the quantity for which y_1 is a realization has a mean to be estimated by the insignificant bias. Thus, we can find the conceptual and statistical difference between two methods.

This difference in the linking invariant can have practical effect in performance evaluation. For $\rho_1 = 0$, the proposed method gives

$$d_2^{(\text{RMO})} = 1.9, \quad U_2^{(\text{RMO})} = 2.2, \\ E_2^{(\text{RMO})} = \frac{d_2^{(\text{RMO})}}{U_2^{(\text{RMO})}} = 0.9.$$

The reference method gives

$$d_2^{(\text{RMO})} = 2.6, \quad U_2^{(\text{RMO})} = 2.3, \\ E_2^{(\text{RMO})} = \frac{d_2^{(\text{RMO})}}{U_2^{(\text{RMO})}} = 1.1.$$

That $|E_2^{(\text{RMO})}|$ is less than unity for the proposed method and greater than unity for the reference method may have significance in terms of any decisions made. Figure 2(d) shows the variation of $E_2^{(\text{RMO})}$ as a function of ρ_1 . Not only at $\rho_1 = 0$ but for $\rho_1 \leq 0.4$, the above discrepancy between the analysis methods arises.

The larger ρ_1 is, the smaller are the differences in the linking invariant and $E_2^{(\text{RMO})}$ between the two methods. When $\rho_1 = 1$, the unilateral DoEs given by the two methods are mutually consistent. The reason for the consistency is that expression (15) yields

$$h_{\text{link}} \rightarrow x_1 - y_1,$$

when $\rho_1 \rightarrow 1$. The value of h_{link} is obtained using $p_1/q_1 = -\rho_1 u(y_1)/u(x_1) = -\rho_1 \rightarrow -1$ for $\rho_1 \rightarrow 1$. This result means that the compensation given for the reference method is similarly applied for our proposal. The difference from the case of $\rho_1 = 0$ is that the common bias in x_1 and y_1 is implied through the correlation information provided by the linking laboratory. It can be said that our proposal gives minimal compensations based on the reliability of the uncertainty and correlation information provided by the linking laboratories. In contrast, the reference method may compensate biases as much as possible whenever the data support it.

In both methods, since h_{link} has no uncertainty when $\rho_1 = 1$, the unilateral DoE is

$$[d_2^{(\text{RMO})}, U_2^{(\text{RMO})}] = [y_2 + x_1 - y_1 - x_{\text{ref}}, k\sqrt{u^2(y_2) + u^2(x_{\text{ref}})}].$$

The uncertainty component of the unilateral DoE is determined to be smaller than that for the case of no correlation. In general, since using the correlation information can make the analysis more precise, the correlation information must be specified reliably.

The advantages of the proposed method and the reference method can be summarized as follows:

1. The presently proposed method compensates insignificant biases in a CIPM KC only when the possible biases are implied in the correlation information.
2. The reference method compensates insignificant biases in a CIPM KC even when the possible biases are not implied in the given information.

Since the methods have different features, the CC should choose an analysis method in accordance with its intention to implement an RMO KC. When linking laboratories report smaller uncertainties than non-linking laboratories in an RMO or the number of linking laboratories is large, the difference between these two methods can be marginal.

7. Summary

A statistical testing procedure is developed to analyze an RMO KC with linking to the corresponding CIPM KC. A statistical model in which biases are not considered is employed. This model can be implemented to check whether a bias exists statistically in a reported value. To determine a parameter of the linking invariant, we use the GLS method under the condition that the CIPM KCRV is fixed.

The proposed method generally gives different unilateral and bilateral DoEs from those assessed by other available methods. Compared to the other methods, our proposal has the advantage that insignificant biases found in a CIPM KC are compensated only when the possible biases are implied in the correlation information. In other words, our proposal is based on the reliability of the uncertainty and correlation information given by the linking laboratories. Since decisions are made on the basis of calculated DoEs, a conclusion is that it is valuable to have adequate knowledge of the properties of available linking methods for analyzing the results from an RMO KC.

Acknowledgments

The National Measurement Office of the UK government's Department for Science, Innovation and Technology (DSIT) supported the first author and also in part the second author as part of NPL's Data Science programme. The work of the second author was partly supported by JSPS KAKENHI Grant Number 17K18411.

Appendix A. Determination of elements in uncertainty evaluations

A.1. Covariances between reported values and the CIPM KCRV

The derivation of $u(x_i, x_{\text{ref}})$ and $u(y_i, x_{\text{ref}})$ for $i \in I_L$ is given. For this purpose, we use the exclusive WM $x_{\text{ex},i}$ for laboratory i , that is, the WM after excluding laboratory i 's data, and the associated standard uncertainty $u(x_{\text{ex},i})$, given by

$$x_{\text{ex},i} = u^2(x_{\text{ex},i}) \sum_{j \in I_L \cup I_X \setminus i} \frac{x_j}{u^2(x_j)},$$

$$u^2(x_{\text{ex},i}) = \sum_{j \in I_L \cup I_X \setminus i} \frac{1}{u^2(x_j)}.$$

As a consequence, using expressions (4), we can write

$$x_{\text{ref}} = u^2(x_{\text{ref}}) \sum_{j \in I_L \cup I_X} \frac{x_j}{u^2(x_j)} + u^2(x_{\text{ref}}) \frac{x_i}{u^2(x_i)}$$

$$= u^2(x_{\text{ref}}) \frac{x_{\text{ex},i}}{u^2(x_{\text{ex},i})} + u^2(x_{\text{ref}}) \frac{x_i}{u^2(x_i)}.$$

Using this result, by defining

$$z_{i,\text{ref}} = \begin{bmatrix} x_{\text{ref}} \\ x_i \\ y_i \end{bmatrix}, \quad C_{i,\text{ref}} = \begin{bmatrix} u^2(x_{\text{ref}})/u^2(x_{\text{ex},i}) & & \\ u^2(x_{\text{ref}})/u^2(x_i) & 1 & \\ & & 1 \end{bmatrix}, \quad (26)$$

we have

$$z_{i,\text{ref}} = C_{i,\text{ref}}^T \begin{bmatrix} x_{\text{ex},i} \\ x_i \\ y_i \end{bmatrix}.$$

Applying the rule for propagating covariances [17, clause 6.2], the covariance matrix associated with $z_{i,\text{ref}}$ is

$$V_{i,\text{ref}} = \begin{bmatrix} u^2(x_{\text{ref}}) & v_{i,\text{ref}}^T \\ v_{i,\text{ref}} & V_i \end{bmatrix} = C_{i,\text{ref}}^T \begin{bmatrix} u^2(x_{\text{ex},i}) & \\ & V_i \end{bmatrix} C_{i,\text{ref}}.$$

In particular, after explicitly evaluating this product,

$$v_{i,\text{ref}} = \begin{bmatrix} u(x_i, x_{\text{ref}}) \\ u(y_i, x_{\text{ref}}) \end{bmatrix} = \begin{bmatrix} 1 \\ \rho_i \frac{u(y_i)}{u(x_i)} \end{bmatrix} u^2(x_{\text{ref}}). \quad (27)$$

A.2. Relationship between concerned variables

The expression for $v_{i,\text{ref}}$ is used to express the covariance matrix for

$$s_{\text{link}}^T = [x_1 \ y_1 \ \dots \ x_L \ y_L \ x_{\text{ref}} \ x_{L+1}].$$

V_{link} denotes the covariance matrix for s_{link} given by

$$V_{\text{link}} = \begin{bmatrix} V_1 & & v_{1,\text{ref}} \\ & \ddots & \vdots \\ & & V_L & v_{L,\text{ref}} \\ v_{1,\text{ref}}^T & \dots & v_{L,\text{ref}}^T & u^2(x_{\text{ref}}) & u^2(x_{\text{ref}}) \\ & & & u^2(x_{\text{ref}}) & u^2(x_{L+1}) \end{bmatrix}, \quad (28)$$

because $u(x_{L+1}, x_{\text{ref}}) = u^2(x_{\text{ref}})$ since $u(x_i, x_{\text{ref}}) = u^2(x_{\text{ref}})$ for $i \in I_L$ as in expression (27). This result also extends to $u(x_i, x_{\text{ref}})$ by replacing x_{L+1} by x_i , for $i \in I_X$, in expression (28).

Define

$$t_{\text{link}}^T = [x_{\text{ref}} \ x_1 \ x_{L+1} \ h_{\text{link}}].$$

The relationship between t_{link} and s_{link} is given as follows:

$$t_{\text{link}} = C_{\text{link}}^T s_{\text{link}} = [c_{1,\text{link}} \ c_{2,\text{link}} \ c_{3,\text{link}} \ c_{4,\text{link}}]^T s_{\text{link}},$$

where

$$c_{1,\text{link}}^T = [0 \ \dots \ 0 \ 1 \ 0],$$

$$c_{2,\text{link}}^T = [1 \ 0 \ \dots \ 0],$$

$$c_{3,\text{link}}^T = [0 \ \dots \ 0 \ 1],$$

$$c_{4,\text{link}}^T = -\frac{1}{Q} [p_1 \ q_1 \ \dots \ p_L \ q_L \ -(P+Q) \ 0].$$

A.3. Covariance between the CIPM KCRV and the linking invariant

The rule for propagating covariances [17, clause 6.2] indicates that the covariance between the i th and the j th element in $\mathbf{t}_{\text{link}}$ is

$$\mathbf{c}_{i,\text{link}}^\top \mathbf{V}_{\text{link}} \mathbf{c}_{j,\text{link}}.$$

For example, the covariance between x_{ref} and h_{link} is given as

$$u(x_{\text{ref}}, h_{\text{link}}) = \mathbf{c}_{1,\text{link}}^\top \mathbf{V}_{\text{link}} \mathbf{c}_{4,\text{link}}.$$

This relationship results in expression (20).

A.4. Covariance between the reported values in the CIPM KC and the linking invariant

The covariance between x_1 and h_{link} is given as

$$\begin{aligned} u(x_1, h_{\text{link}}) &= \mathbf{c}_{2,\text{link}}^\top \mathbf{V}_{\text{link}} \mathbf{c}_{4,\text{link}} \\ &= -\frac{1}{Q} \begin{bmatrix} 1 \\ 0 \end{bmatrix}^\top \mathbf{V}_1 \begin{bmatrix} p_1 \\ q_1 \end{bmatrix} + \frac{P+Q}{Q} u^2(x_{\text{ref}}). \end{aligned}$$

Since

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix}^\top \mathbf{V}_1 \begin{bmatrix} p_1 \\ q_1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}^\top \mathbf{V}_1 \mathbf{V}_1^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 0,$$

we obtain

$$u(x_1, h_{\text{link}}) = \frac{P+Q}{Q} u^2(x_{\text{ref}}). \quad (29)$$

Moreover, we obtain the covariance $u(x_{L+1}, h_{\text{link}})$ as

$$u(x_{L+1}, h_{\text{link}}) = \mathbf{c}_{3,\text{link}}^\top \mathbf{V}_{\text{link}} \mathbf{c}_{4,\text{link}} = \frac{P+Q}{Q} u^2(x_{\text{ref}}). \quad (30)$$

Expressions (29) and (30) can be generalized without any changes in their right sides for the cases of x_i for $i \in I_L$ and $i \in I_X$, respectively. Thus, expression (23) is obtained.

A.5. Variance of the linking invariant

The variance of h_{link} is

$$u^2(h_{\text{link}}) = \mathbf{c}_{4,\text{link}}^\top \mathbf{V}_{\text{link}} \mathbf{c}_{4,\text{link}}.$$

Following some algebra, paying regard to symmetry we obtain

$$\begin{aligned} u^2(h_{\text{link}}) &= \frac{1}{Q^2} \sum_{j \in I_L} [p_j \ q_j] \mathbf{V}_j \begin{bmatrix} p_j \\ q_j \end{bmatrix} - 2 \frac{P+Q}{Q^2} \sum_{j \in I_L} \mathbf{v}_{j,\text{ref}}^\top \begin{bmatrix} p_j \\ q_j \end{bmatrix} \\ &\quad + \frac{(P+Q)^2}{Q^2} u^2(x_{\text{ref}}). \end{aligned}$$

Now, using expressions (16), (17) and (27),

$$[p_j \ q_j] \mathbf{V}_j \begin{bmatrix} p_j \\ q_j \end{bmatrix} = q_j, \quad \mathbf{v}_{j,\text{ref}}^\top \begin{bmatrix} p_j \\ q_j \end{bmatrix} = 0.$$

Hence, expression (19) is obtained.

Appendix B. Bilateral degree of equivalence with the values unused in the computation of the reference value

In the body of the paper, we assume that all reported values in the CIPM KC are used in the computation of the CIPM KCRV, which is the WM of those values. However, in practice, values regarded as outliers by the CC Working Group are not used in the computation. In this section, we show the evaluation of the bilateral DoE between a laboratory reporting one of the outliers in the CIPM KC and an RMO KC participant.

The reported value from the laboratory participating in the CIPM KC is denoted by x_ℓ . The variance of the unilateral DoE $d_\ell^{(\text{CIPM})} = x_\ell - x_{\text{ref}}$ is not given by expression (6) but by

$$u^2(d_j^{(\text{CIPM})}) = u^2(x_j) + u^2(x_{\text{ref}}). \quad (31)$$

The value component of the bilateral DOEs is given by formula (2). To derive the uncertainty component, while formula (22) always holds, formula (23) does not apply when x_ℓ is not used in the computation of the KCRV. Instead,

$$u(x_\ell, h_{\text{link}}) = 0 \quad (32)$$

because x_ℓ is not used in the assessment of h_{link} . Expressions (19), (22), (31) and (32) give

$$\begin{aligned} u^2(d_{j,\ell}^{(\text{RMO, CIPM})}) &= u^2(y_j) + \left[\frac{1}{Q} + \frac{(P+Q)^2}{Q^2} u^2(x_{\text{ref}}) \right] + u^2(x_\ell) \\ &= \left[u^2(y_j) + \frac{1}{Q} + \frac{P^2}{Q^2} u^2(x_{\text{ref}}) \right] + [u^2(x_\ell) + u^2(x_{\text{ref}})] \\ &\quad + 2 \frac{P}{Q} u^2(x_{\text{ref}}) \\ &= u^2(d_j^{(\text{RMO})}) + u^2(d_\ell^{(\text{CIPM})}) + 2 \frac{P}{Q} u^2(x_{\text{ref}}). \end{aligned}$$

$U_{j,\ell}^{(\text{RMO, CIPM})}$ and $E_{j,\ell}^{(\text{RMO, CIPM})}$ are computed by applying formulae (24) and (25) using the obtained standard uncertainty.

Appendix C. Unilateral degree of equivalence as an estimated bias

C.1. Bias estimation approach in a single CIPM KC

It is sometimes considered that the value component $d_i^{(\text{CIPM})}$ of a unilateral DoE in expressions (5) is an estimate of the bias for laboratory i . Based on that concept, White [6] developed the model in which biases Δ_i are assumed as follows:

$$x_i \sim N(\mu_x + \Delta_i, u^2(x_i)) \quad \text{for all } i \in I_L \cup I_X. \quad (33)$$

To obtain estimates of Δ_i , a constraint is needed since there are only m data ($x_1, \dots, x_m$) and $m+1$ quantities ($\Delta_1, \dots, \Delta_m, \mu_x$) to be estimated. One possible constraint [6] is

$$\sum_{i=1}^m \frac{\Delta_i}{u^2(x_i)} = 0. \quad (34)$$

White [6] gives expression (5) with variance (6) as least-squares solutions with using model (33) and (34). Several studies previously used this concept [19].

C.2. Extension of the bias estimating approach to an RMO KC

The model of expression (33) may be extended to the analysis of an RMO KC as follows:

$$\begin{aligned} z_i &\sim N\left([\mu_x + \Delta_i, \mu_y + \Delta_i]^\top, V_i\right), & i \in I_L, \\ x_i &\sim N\left(\mu_x + \Delta_i, u^2(x_i)\right), & i \in I_x, \\ y_i &\sim N\left(\mu_y + \Delta_i^{(\text{RMO})}, u^2(y_i)\right), & i \in I_y, \end{aligned} \quad (35)$$

with constraint (34). The term $\Delta_i^{(\text{RMO})}$ is the bias for laboratory i in the RMO KC, while Δ_i is that for laboratory i in the CIPM KC, which can be a linking laboratory. To handle the linking invariant explicitly, $\mu_x - \eta$ is employed in expression (10) instead of μ_y in expression (35). As far as we are aware, there has been no publication in which model (35) with constraint (34) is straightforwardly employed. However, there are some studies in which an equivalent model is employed.

Sutton [5] proposed a statistical model with explicit biases. The model is generally given and details are not specifically suggested in his paper. We developed the details in the proposed model on the assumption used in this study, and derive the model that is equivalent to

$$\begin{aligned} z_i &\sim N\left([\mu_x + \Delta_i, \mu_y + \Delta_i]^\top, V_i\right), & i \in I_L, \\ y_j &\sim N\left(\mu_y + \Delta_j^{(\text{RMO})}, u^2(y_j)\right), & i \in I_y, \\ x_{\text{ref}} &\sim N\left(\mu_x, u^2(x_{\text{ref}})\right), \end{aligned}$$

the covariances $u(x_i, x_{\text{ref}})$ and $u(y_i, x_{\text{ref}})$ being given by expression (27). In [5], a constant as large as the CIPM KCRV is defined as K , and the distributions of $x_i - K$ and $K - x_{\text{ref}}$ are considered rather than x_i and x_{ref} (if our understanding is correct). Specifically, $K - x_{\text{ref}} \sim N(K - \mu_x, u^2(x_{\text{ref}}))$ is considered. Giving the distributions of x_i and x_{ref} is however statistically equivalent to giving those of $x_i - K$ and $K - x_{\text{ref}}$.

The analysis with this statistical model can be consequently identical to that with the model with expression (35) with the constraint with expression (34), when the WM is employed as the CIPM KCRV. If we apply model (33) and take the WM as the CIPM KCRV x_{ref} , the distribution of x_{ref} has the population mean of

$$\left(\sum_{i=1}^m \frac{1}{u^2(x_i)}\right)^{-1} \left(\sum_{i=1}^m \frac{\Delta_i}{u^2(x_i)}\right) + \mu_x.$$

When expression (34) is used as a constraint, the mean is given as μ_x . Thus, expression (34) is implicitly applied in this method.

Moreover, the unilateral DoE $d_i^{(\text{RMO})}$ derived as the GLS estimate of $\Delta_i^{(\text{RMO})}$ based on the model of expression (35) and the bilateral DoEs obtained through expressions (2) and (3) are identical to those suggested by Elster *et al* [3] when the CIPM KCRV is given as the WM in expression (4). Although those authors developed the procedure in the absence of a specific model, we interpret their study as corresponding to that developed that by Sutton [5] and clarifying some mathematical expressions in it.

It should be noted that through the bias estimation method explained in this appendix, the re-estimation of μ_x is implemented, which can be interpreted as the re-determination of the KCRV purposely avoided in our proposal in the main manuscript.

ORCID iDs

Maurice Cox  <https://orcid.org/0000-0002-6342-7840>
Katsuhiro Shirono  <https://orcid.org/0000-0002-4930-799X>

References

- [1] BIPM 1999 Mutual recognition of national measurement standards and of measurement certificates issued by national metrology institutes (MRA) *Technical Report* Bureau International des Poids et Mesures, Sèvres, France
- [2] Decker J E, Steele A G and Douglas R J 2008 Measurement science and the linking of CIPM and regional key comparisons *Metrologia* **45** 223
- [3] Elster C, Chunovkina A G and Wöger W 2010 Linking of a RMO key comparison to a related CIPM key comparison using the degrees of equivalence of the linking laboratories *Metrologia* **47** 96
- [4] Kharitonov I A and Chunovkina A G 2006 Evaluation of regional key comparison data: two approaches for data processing *Metrologia* **43** 470
- [5] Sutton C M 2004 Analysis and linking of international measurement comparisons *Metrologia* **41** 272
- [6] White D R 2004 On the analysis of measurement comparisons *Metrologia* **41** 122
- [7] Kellermann M 2019 *Ensuring Quality to Gain Access to Global Markets: A Reform Toolkit* (The World Bank)
- [8] Steele A, Wood B and Douglas R 2001 General linking methodology with application to CCT-K3 and EUROMET Project 280, CCT/01-38 *Technical Report* NRC Canada
- [9] Thomas C 2005 The BIPM key comparison database (KCDB): linkage of key comparison results *Technical Report* Bureau International des Poids et Mesures, Sèvres, France Rapport BIPM-05/06
- [10] Man J, Arias R, Terao Y, Jae Lee Y, Ligong G, Tulasombut V, Kin Chan T, Hong Thai N, Steyn R and Sampath H L I S 2011 Final report on APMP key comparison of volume of liquids at 20 l and 100 ml: APMP.M.FF-K4 *Metrologia* **48** 07001
- [11] Woolliams E R, Fox N P, Cox M G, Harris P M and Harrison N J 2006 Final report on CCPR K1-a: spectral irradiance from 250 nm to 2500 nm *Metrologia* **43** 02003
- [12] von Martens H-J, Elster C, Link A, Wöger W and Allisy P J 2004 Linking the results of the regional key comparison

- APMP.AUV.V-K1 to those of the CIPM key comparison CCAUV.V-K1 *Metrologia* [41 09002](#)
- [13] Cox M G and Harris P M 2012 The evaluation of key comparison data using key comparison reference curves *Metrologia* [49 437](#)
- [14] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML 2008 Evaluation of measurement data—guide to the expression of uncertainty in measurement *Technical Report* Joint Committee for Guides in Metrology JCGM: 100
- [15] Cox M G 2002 The evaluation of key comparison data *Metrologia* [39 589–95](#)
- [16] Kacker R N, Forbes A, Kessel R and Sommer K-D 2008–09 Bayesian posterior predictive p -value of statistical consistency in interlaboratory evaluations *Metrologia* [45 512](#)
- [17] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML 2011 Evaluation of measurement data—supplement 2 to the ‘guide to the expression of uncertainty in measurement’ — models with any number of output quantities *Technical Report* Joint Committee for Guides in Metrology JCGM 102
- [18] Arias R *et al* 2006 Results of the key comparison CCM.FF-K4 for volume of liquids at 20 l and 100 ml *Metrologia* [43 07005](#)
- [19] Willink R 2002 Statistical determination of a comparison reference value using hidden errors *Metrologia* [39 343](#)