Harmonised Values and Reference Functions in Key Comparisons

Wolfram Bremser[1], and Bodo Mickan[2]
[1] Federal Institute for Materials Research and Testing (BAM), D-12200 Berlin, Germany
[2] Physikalisch-Technische Bundesanstalt (PTB), D-38116 Braunschweig,

Abstract: Key comparisons in the field of flow measurement have some peculiarities with impact on data evaluation and performance assessment, in particular on the caluclation of the key comparison reference value (KCRV). The normal KC approach which assumes that all measurements in all laboratories are made at prescribed nominal sampling points faces problems in practical implementation. Furthermore, it is normally assumed that values submitted by participants are independent of each other. It is well-known that this assumption may not hold in specific cases since some of the factors which influence the uncertainty may be the same in all or at least a couple of laboratories. The paper discusses the influence of the above factors. The problems are handled by calculating key comparison reference functions (KCRF) instead of series of isolated KCRV, and by correctly evaluating, and accounting for all correlations present within the data sets.

Keywords: flow measurement, key comparisons, reference functions, correlations, measurement uncertainty.

1. Introduction

Bilateral, regional or global laboratory intercomparisons are key instruments for ensuring the performance and reliability of, and the comparability between, the participating facilities. In the field of flow measurement (liquids or gases at different pressure levels), such intercomparisons have some peculiarities with impact on data evaluation and performance assessment, in particular on the caluclation of the key comparison reference value (KCRV).

Firstly, the normal approach which assumes that all measurements in all laboratories are made at prescribed nominal sampling points faces problems in practical implementation. Although technical performance of the test facilities is high, it remains difficult to achieve that all nominal sampling points are exactly met in all laboratories, at least at the claimed level of accuracy.

Secondly, it is normally assumed that values submitted by participants are independent of each other [e.g. 1, 2]. It is well-known that this assumption (if correspondingly scrutinised) may not hold in specific cases since some of the factors which influence the uncertainty may be the same in all or at least a couple of laboratories. With the establishment of the harmonised cubic meter of natural gas, correlations between adhering laboratories will automatically occur due to the common reference.

The paper will discuss the influence of the above factors. The first problem is handled by calculating key comparison reference functions (KCRF) instead of series of isolated KCRV, the second by correctly evaluating and accounting for all correlations which occur, in particular due to the harmonisation procedure. Recipes for calculating these correlations and acccounting for them in the KC evaluation will be given. Both approaches will be described in detail, and their performance demonstrated on real data from intercomparisons.

2. Harmonisation procedures as a source of correlation

2.1. Basic assumptions

It is assumed that all laboratories which adhere to the harmonisation procedure will have there own, independent calibration chains enabling them to convert a measured (raw) response signal of a gas meter into a meter deviation and a corresponding uncertainty. This will normally be done by using an instrument response function (or other standardised procedure) providing a (converted to standard reference conditions) deviation De(Re, lab #) and a corresponding (standard) uncertainty u(Re, lab #). The values and uncertainties thus derived in one laboratory may be considered independent of all values derived in other laboratories.

This assumption holds independently of the type of the instrument response function used and the homo- or heteroscedasticity of the assigned or calculated uncertainties, and of the stage of the calibration chain at which the particular measurement is carried out. On the other hand, it should be mentioned that the deviations obtained at different sampling points (different Re numbers) in one laboratory may well be correlated due to the use of the same instrument response function, but this specific issue will not be considered here.

Furthermore, it is assumed that the reference object (RO) used for the determination of the harmonisation correction is well distinguished from the transfer package (TP) of the KC and thus independent.

2.2. Harmonisation procedure

The procedure for determining the harmonisation correction is described as follows:

Step 1. On the reference object, measurement are taken in the *i*th (out of *n*) laboratories at individual sampling points providing data tables of the form

independent variable:	Re(1,i)	Re(2,i)	... Re(k,i)
dependent variable:	De(1,i)	De(2,i)	... De(k,i)

Note that the indepenedent variable may also be the mass flow. Since such a selection will not change the principles of data treament as described below, all further description troughout this paper will refer to Reynolds numbers. Data tables may be different in length for each of the *n* laboratories.

Step 2. A common reference value De_ref(j) at a certain Re(j) is formed as a weighted average according to

$$De\_ref(j) = \sum_{i=1}^{n} De(j,i) * w(j,i) \qquad (1)$$

being w(j,i) the weights

$$w(j,i) = [\sum_{l=1}^{n} u^2(j,i)/u^2(j,l)]^{-1} \qquad (2)$$

This approach faces problems in practical implementation since normally deviations are not measured at the very same sampling point (say Re(j)) in all of the laboratories. Although technical performance of the test facilities is (extremely) high, it is difficult to achieve that Re(j,i) = Re(j,l) for ∀ i,l (at least at the level of accuracy claimed by the labs).

Thus, another approach must be chosen, namely weighted regression of an appropriately selected model function [3]. This approach applies both to harmonisation (i.e. the determination of a harmonised reference function) and to any subsequent KC (i.e. the determination of a KC reference function).

Choosing of this model function may be based upon different criteria which may also be discussed controversially. However, polynomials of the order four are common for gas meters. A modified sequential F test may be used to test for a possible reduction in the number of parameters (the order) of the polynomial used as the reference function. The number of parameters is reduced as long as the test does not indicate significant deterioration of the fit quality (compatibility of data with the fit within the stated uncertainties). The sequential F test is modified in the sense that normal test refers to residual scatter in the data. Since all values come with stated uncertainties, the modified test must refer to the sum of standardised (by the corresponding uncertainties) residuals.

Thus, the minimisation problem is

$$(\mathbf{De} - \mathbf{f\_ref}(Re))^T * \mathbf{VC}^{-1} * (\mathbf{De} - \mathbf{f\_ref}(Re)) = min \qquad (3)$$

where **De** is a vector consisting of all De(j,i) measured by the *n* labs at the *k* (individual) Re numbers, **f_ref** the vector formed of the model function values at the Re(j,i), and **VC** the variance-covariance matrix of all De(j,i) taken into consideration. Note that the minimisation criterion may take up also probably existing covariances between the De(j,i) (wherever they may originate from), but under the basic assumptions made here, **VC** will be diagonal.

Minimisation of the above expression will result in a vector of model function parameters $\mathbf{p}_{ref}$ and a corresponding variance-covariance matrix $\mathbf{M}_{ref}$ for the reference function f_ref. (Note that all calculations described will normally be carried out on the logarithm of Re; for simplicity, the lg is omitted here). $\mathbf{M}_{ref}$ may be calculated as the inverted normal equations' matrix of the above minimisation problem.

Step 3. From the De(j,i), deviations $d^{feed}(j,i)$ of the individual result from the reference curve are calculated at Re(j,i) as

$$d^{feed}(j,i) = De(j,i) - f\_ref(Re(j,i)) \qquad (4)$$

and individual so-called feeback functions f_fed_i determined by normal, straight-line OLS regression on $d^{feed}(j,i)$ the for each laboratory. The $d^{feed}(j,i)$ for ∀ i,j have a variance-covariance matrix $VC(d^{feed})$ according to

$$VC(d^{feed}) = \mathbf{J} * VC(De) * \mathbf{J}^T \qquad (5)$$

where **J** is the Jacobian consisting of the partial derivatives of $d^{feed}(j,i)$ with respect to all De(h,l). Again, uncertainty propagation is used to calculate the VC matrix of the feedback function parameters from $\mathbf{VC}(d^{feed})$:

$$VC(p) = \mathbf{J} * VC(d^{feed}) * \mathbf{J}^T \qquad (6)$$

where **J** is the Jacobian consisting of the partial derivatives of the feedback function parameters intercept p_0 and slope p_1 with respect to all $d^{feed}(j,i)$.

With these feedback functions, each laboratory is able to calculate (at each Re) the correction $f\_fed_i(Re)$ needed to trace back any measured value to the common, harmonised reference. The conversion is

$$\delta = De - f\_fed_i(Re) \qquad (7)$$

with De being the measured, and δ being the harmonised value.

Using equations (4), (5) and (7), also the uncertainties of the individual values δ_i can be calculated by propagating the corresponding uncertainties to the final result. Normally, each laboratory will report its own harmonised values plus the corresponding uncertainty to the customer. If these values are intented to be used in a subsequent KC, the same procedure may be used for calculating existing covariances between a δ_i generated in lab *i* and a δ_i generated in lab *k*. Thus, the complete variance/covariance matrix of the minimisation criterion (3) can be evaluated. For a detailed description of the procedure see [4].

3. Harmonised values in key comparisons

3.1. Data

In key comparisons where a certain cluster of laboratories adhere to a reference harmonisation procedure but others do not, the latter may be tempted to suspect a direct or indirect impact on (i.e. a distorsion of) the KCRV which might shift it into a direction unfavorable for these laboratories. It will be shown that - for both unweighted and weighted KCRV calculation - the KCRV remains unchanged even if some participating laboratories adhere to a harmonisation procedure. Thus, the "distance" of a not adhering laboratory to the KCRV does not depend on whether or not other laboratories adhere to a procedure as described in part I. Since the uncertainty statements of I. Since the uncertainty statements of a non adhering laboratory also do not depend on the other labs, the same will be true for the degree of equivalence calculated for the non adhering labs.

For a better transparency of the following calculations, it will be assumed that all labs managed to meet the prescribed sampling point within a negligible tolerance interval, i.e. that for the sampling point under consideration univariate data evaluation is possible. The impact of a certain scatter in the sampling points (which is inevitable and does not depend on whether harmonised values are or are not included in the KC) will be discussed in clause 3.4.

The following symbol convention for the variables under consideration will be used: Let *k* be the number of labs adhering to harmonisation, and *l* the number of labs which do not adhere (altogether $n = k + l$ labs). The *l* labs submit (non-harmonised) values De_i and corresponding uncertainties u_i for the KC, the *k* labs - harmonised values δ_i and uncertainties u_i. In a harmonisation round conducted before the KC, the *k* labs produced (on a certain transfer standard) values $De_i^{(h)}$ and corresponding uncertainties $u_i^{(h)}$. It is common practice for the labs adhering to harmonisation that the stated uncertainties are based on a sound investigation of the uncertainty budgets of the particular test facilities and do not change from measurement to measurement (since they include more than a bare repeatability), thus $u_i^{(h)} = u_i$ holds. Let the harmonised reference value be De_{ref}, d_i^{feed} the feedback value (since the univariate case is considered here), and let <X> be the average (i.e. the normal or weighted mean according to the specific procedure) of a certain number of single values.

Summarising the above conventions, the laboratories produce the following data:

Procedure	property	adhering labs *1... k*	non adhering labs *k+1... n*
Harmonisation	values	$De_i^{(h)}$	-
	uncertainties	$u_i^{(h)}$	-
KC	values	δ_i	De_i
	uncertainties	$u_i^{(h)}$	u_i

3.2. Unweighted and weighted KCRV

3.2.1. Unweighted KCRV

Harmonised reference value and KCRV are calculated according to the standard procedure using the unweighted mean as the best estimate. Within this procedure, the harmonised reference value and the feedback are

$$D_{ref} = \sum_{i=1}^{k} De_i^{(h)}/k \qquad d_i^{feed} = De_i^{(h)} - D_{ref} \qquad (8)$$

The values provided by the labs adhering to harmonisation are

$$\delta_i = De_i - d_i^{feed} = De_i - De_i^{(h)} + D_{ref} \qquad (9)$$

The unweighted KCRV (indexed by $^{(h)}$) formed on the data provided by both the adhering and non adhering labs is

---

$$<De>^{(h)} = [\sum^{k} \delta_i + \sum^{l} De_i]/(k + l) = [\sum^{k} De_i - \sum^{k} De_i^{(h)} + \sum^{k} D_{ref} + \sum^{l} De_i]/(k + l) \qquad (10)$$

$$<De>^{(h)} = [\sum^{k} De_i - \sum^{k} De_i^{(h)} + \sum^{k} De_i^{(h)} + \sum^{l} De_i]/(k + l) = [\sum^{k} De_i + \sum^{l} De_i]/(k + l) \qquad (11)$$

$$<De>^{(h)} = \sum^{k+l} De_i/(k + l) = <De> \qquad (12)$$

---

due to the fact that $\Sigma D_{ref} = k * D_{ref} = \Sigma De_i^{(h)}$.

This demonstrates that the KCRV which includes harmonised values δ_i submitted by a subgroup of laboratories is exactly the same as in the case when all laboratories submit their own, non harmonised De_i.

3.2.2. Weighted KCRV

Harmonised reference value and KCRV are calculated according to the weighing procedure recommended by the BIPM using the inverse-variance weighted mean as the best estimate. This is the actual procedure which is followed by the laboratories adhering to the harmonised cubic meter [5, 6]. Within this procedure, the harmonised reference value and the feedback are

$$D_{ref} = [\sum_{i=1}^{k} De_i^{(h)}/u_i^{(h)2}] / S_k \qquad (13)$$

$$d_i^{feed} = De_i^{(h)} - D_{ref} \qquad (14)$$

using the symbol $S_k = \Sigma\ 1/u_i^{(h)2}$ for the sum of inverse variances. The values provided by the labs adhering to harmonisation is

$$\delta_i = De_i - d_i^{feed} = De_i - De_i^{(h)} + D_{ref}$$

The weighted KCRV (indexed by $^{(h)}$) formed on the data provided by both the adhering and non adhering labs is

---

$$<De>^{(h)} = [\sum^{k} \square_i/u_i^{(h)2} + \sum^{l} De_i/u_i^2]/(S_k + S_l) \qquad (15)$$

$$<De>^{(h)} = [\sum^{k} De_i/u_i^{(h)2} - \sum^{k} De_i^{(h)}/u_i^{(h)2} + \sum^{k} D_{ref}/u_i^{(h)2} + \sum^{l} De_i/u_i^2]/(S_k + S_l) \qquad (16)$$

$$<De>^{(h)} = [\sum^{k} De_i/u_i^{(h)2} - \sum^{k} De_i^{(h)}/u_i^{(h)2} + \sum^{k} De_i^{(h)}/u_i^{(h)2} + \sum^{l} De_i/u_i^2]/(S_k + S_l) \qquad (17)$$

$$<De>^{(h)} = [\sum^{k} De_i/u_i^{(h)2} + \sum^{l} De_i/u_i^2]/(S_k + S_l) = [\sum^{k+l} De_i/u_i^2]/S_{k+l} \qquad (18)$$

$$<De>^{(h)} = <De>$$

---

due to the fact that
$\Sigma\ D_{ref}/u_i^{(h)2} = D_{ref} * \Sigma\ 1/u_i^{(h)2} = D_{ref} * S_k = \Sigma\ De_i^{(h)}/u_i^{(h)2}$.

This demonstrates that the weighted KCRV which includes harmonised values δ_i submitted by a subgroup of laboratories is exactly the same as in the case

when all laboratories submit their own, non harmonised De_i.

3.3. KCRV standard uncertainty

As already mentioned, the harmonised values are partially correlated. These correlations must be taken into account in order to correctly determine both the KCRV uncertainty and the uncertainties of the values submitted by the adhering labs. This may transparently be illustrated for the case considered in clause 3.2.1. (unweighted procedure) by looking at the standard deviations obtained for the adhering laboratories based on the (actually) provided δ_i and on the (classically) provided De_i. It holds:

$$(k - 1) * s^2(De) = \Sigma (De_i - \langle De \rangle)^2 \quad (19)$$

$$(k - 1) * s^2(\delta) = \Sigma (\delta_i - \langle \delta \rangle)^2 = \Sigma (\delta_i - \langle De \rangle)^2 \quad (20)$$

For the substitution in equation (20) see clause 3.2.2. With $\delta_i = De_i - d_i^{feed} = De_i - De_i^{(h)} + D_{ref}$, this can be expanded to

$$(k - 1) * s^2(\delta) = \Sigma (De_i - De_i^{(h)} + D_{ref} - \langle De \rangle)^2 \quad (21)$$

$$= \Sigma [(De_i - \langle De \rangle) - (De_i^{(h)} - D_{ref})]^2 \quad (22)$$

$$= \Sigma (De_i - \langle De \rangle)^2 - 2 * \Sigma (De_i - \langle De \rangle) * (De_i^{(h)} - D_{ref}) + \Sigma (De_i^{(h)} - D_{ref})^2 \quad (23)$$

$$s^2(\delta) = s^2(De) - 2 * cov(De, De^{(h)}) + s^2(De^{(h)}) \quad (24)$$

Equation (24) illustrates that correlations must be taken into account for both introducing δ_i into the data ensemble of values submitted by non adhering labs ($s^2(De)$) and for calculating the degrees of equivalence ($s^2(\delta)$). Uncertainty propagation rules for the more complex case of determining a KCRF by regression is given in clause 2 and [4].

3.4. KCRF obtained from regression

As has already been mentioned above, it is inherent to this kind of KC that the participating laboratories will meet the envisaged (or prescribed) sampling points (i.e. the nominal flow rates) only with a certain scatter. The proposed way of determining a KCRF is regression in the "deviation vs Re number (or mass flow)" space. In the vicinity of the nominal sampling points, the regressed function will not exactly pass through the (weighted or unweighted) mean of submitted values (as calculated under the assumption that they all were taken at exactly the same sampling point). This property of the KC does **not** depend on whether or not laboratories adhering to a harmonisation procedure participate in the KC, and will distort the "classical" KCRV to a certain, but small extent.

However, the same applies to the harmonisation procedure as described above. Thus, at each sampling point, D_{ref} is replaced by a D_{ref}^* which is the estimate for the "mean" at this point taken from regression. In the equations above (which yielded exact coincidence between $\langle De \rangle^{(h)}$ and $\langle De \rangle$), this replacement leads to a small residual term ε. Since regression minimises the variances over the whole regression range (and not only at a certain sampling point), ε will follow a normal distribution with zero mean, i.e. $E(\varepsilon) = 0$, so that in a statistical sense, all conclusions drawn in clauses 3.2 and 3.3 remain valid.

4. Example

Data for the example presented here stem from a real intercomparison within the framework of the monitoring procedure for the European harmonised cubic meter. Three laboratories have been selected from the data set, and KCRF were determined for two cases, namely

i) on the basis of the original data from all three laboratories (see figure 1)

ii) with two (out of the three) laboratories having harmonised their data prior to the intercomparison, and one laboratory providing original data (see fig. 2).

Procedures for the harmonisation in case ii) were as described in clause 2. All values in fig. 1 and 2 are plotted against the independent variable, normally (but not necessarily) the Reynolds number or its logarithm.

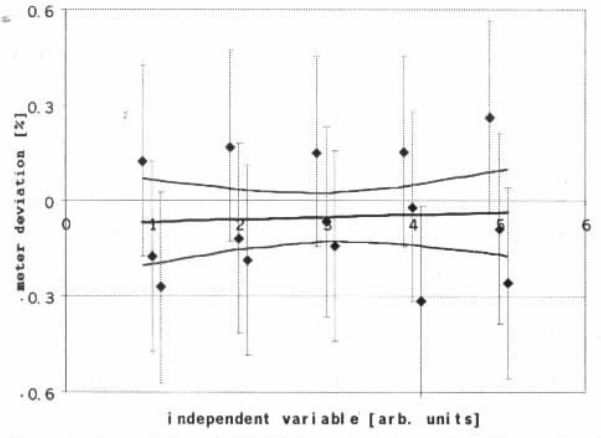

Figure 1: Determining the KCRF from the values originally measured. Original values (diamonds) and their uncertainty (k = 2), and the straight-line KCRF together with its confindence band.

The two laboratories which adhere to harmonisation are the ones with the meter deviation values below the zero line.

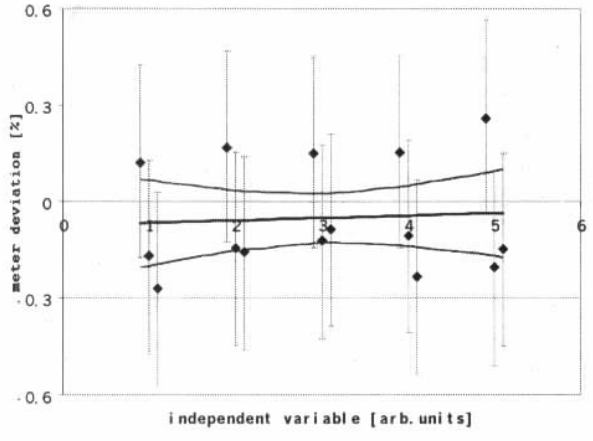

Figure 2: Determining the KCRF from the values after harmonistaion. The straight-line KCRF together with their confindence bands are plotted for the case of harmonisation (bold lines) and without harmonisation (from fig. 1, dotted bold lines).

Note that in fig. 2, the KCRF for both cases are plotted. As can easily be seen, the resolution of the graph (even if it would be enlarged) is insufficient to visualise the tiny difference between the KCRF of case i) and ii) occuring for the reasons discussed in clause 3.4. This difference can only be unveiled from the parameters of the KCRF and their corresponding uncertainties (see table 1).

	harmonised	classical
intercept	-0.075290766	-0.074674481
slope	0.00758797	0.007448605
u(inter)	0.181650285	0.181411959
u(slope)	0.054766507	0.054681196
cov	-0.008996203	-0.0089701

Table 1: KCRF parameters and their uncertainties for cases i) (classical approach) and ii) including harmonised values.

Obviously, all of the differences in the KCRF parameters are not only insignificant but negligible. This demonstrates that the conclusions drawn under clause 3 are equally valid for KCRV and KCRF.

5. Summary

Key comparisons in the field of flow measurement have peculiarities with impact on the caluclation of the key comparison reference (KCRV). It has been shown that KC reference functions instead of series of isolated KCRV should be calculated, and recipes are given for the corresponding data evaluation procedure. The latter equally applies to any harmonisation procedure.

Harmonisation between laboratories creates correlation in the harmonised values. So, the normal assumption of mutual independence of values provided by laboratories participating in a KC does not hold.

On the other hand, it has been shown here that reporting a harmonised instead of a stand-alone value in a KC does not affect the KC reference function nor the corresponding confidence interval band if only correlations between the values provided are correctly evaluated and taken into account.

This holds also for cases where a group of participating laboratories adhere to a harmonisation, but others do not.

References

[1] ISO 5725:1994-2002: Accuracy (trueness and precision) of measurement methods and results. ISO, Geneva.

[2] Guidelines for CIPM key comparisons (Appendix F to the "Mutual recognition of national measurements standards and of measurement certificates issued by national metrology institutes" (MRA)), CIPM, 1999.

[3] M.G. Cox: A discussion of approaches for determining a reference value in the analysis of key-comparison data. NPL Report CISE 42/99.

[4 Final Technical Protocol for CCM.FF.KC5ab: Technical Measurement Protocol for BIPM/CIPM High-Pressure Key Comparisons Natural Gases (5a) and air + Nitrogen (5b). PTB/NMI, July 2004.
[5] Cox, M. G.: The evaluation of key comparison data. Metrologia, 2002, 39, 589-595
[6] B. Mickan, R. Kramer, H.-J. Hotze, D. Dopheide: PIGSAR- the extended test facility and new german national primary standard for high-pressure natural gas. Proceedings of the 11th FLOMEKO, Groningen, The Netherlands, 12 - 14 May 2003.